UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

DIPARTIMENTO DI INGEGNERIA ”ENZO FERRARI” DI MODENA
Corso di Dottorato di Ricerca in ”Information and Communication Technologies (ICT)”

CICLO XXXIII

SIMULATION AND MODELING METHODS FOR PREDICTING PERFORMANCE AND RELIABILITY LIMITS OF 21ST-CENTURY ELECTRONICS

Candidato:
Dott. Nicolò Zagni

Relatore:
Prof. Paolo Pavan

Correlatore:
Prof. Francesco Maria Puglisi

Coordinatore Dottorato:
Prof. Sonia Bergamaschi

Anno Accademico 2019–2020

Davanti c'era la menzogna comprensibile,

e dietro, l'incomprensibile verità

M. Kundera

Acknowledgments

There are quite a few people to thank for helping me successfully reaching the end of the PhD journey.

- My deepest gratitude goes to professor Paolo Pavan, who introduced me to the research process as my Master's Thesis advisor first and as my PhD supervisor then. His valuable advices and — most of all — his guidance were very important to me to reach this point, that I could not reach otherwise.
- I thank professor Francesco Maria Puglisi, and I gladly acknowledge the many hours spent in the lab with him doing experiments or discussing on many different research topics. His expertise and enthusiasm were very inspiring to me.
- Many thanks are devoted to professor Giovanni Verzellesi, who shared with me his vast knowledge of semiconductor physics and other (not always technical) topics, helping me most of all in understanding and appreciating the nuances of device simulations.
- I acknowledge the fruitful collaboration with professor Alessandro Chini, that offered me his challenging perspective on Gallium Nitride devices and on the issues dealing with 'real devices' compared to those encountered in the simulations.
- I deeply thank professor Muhammad Ashraful Alam from Purdue University, whom I started working with as 'visiting scholar' in summer 2018 and 2019. He is a great advisor and one of the most recognized mentors in the electron devices field. Working with him allowed me to grow substantially and made me capable of bringing new perspectives to my work.

- I thank prof. Susanna Reggiani (University of Bologna) and prof. David Esseni (University of Udine) for evaluating and reviewing this Thesis.

I also gladly acknowledge my fellow PhD students in Modena, whom I shared with many hours in the lab. They all made my experience more enjoyable. I also thank all the (past and present) Alam's group members at Purdue University that I was very lucky to meet, in the hope that our paths will cross again.

On the personal side, I want to thank my girlfriend Giulia for her support and understanding even when we were physically far away. Needless to say, it was a painful process and of countless hours spent in the lab, taken away from friends and family. I appreciated their patience and understanding for the many occasions that I sacrificed spending time with them to work. This thesis is dedicated to all of you.

Abstract

In recent years, a plethora of novel semiconductor devices have started emerging as worthy heirs of Silicon-based transistors – giving rise to the 'post-Moore' era. Traditional electronics is mostly based on Si devices, – from logic to memory, to high frequency/power and sensing applications – but this paradigm is changing thanks to the developments in different fields ranging from physics and semiconductor materials, to processing techniques and computing architectures. In this hectic new scenario, before even considering a new technology as a replacement of the existing ones, the limiting factors to its performance and reliability need to be well-understood and engineered for. In this sense, simulations and physics-based modeling represent critical tools to make sure that newly conceived technologies stand up to the requirements of 21st century electronics.

In this thesis, state-of-the-art simulation and compact modeling tools are exploited to analyze the performance and reliability limits of several emerging technologies. Specifically, this dissertation is focused on four application scenarios and the relative candidate technologies that aim to providing enhanced performance/reliability compared to Si-based counterparts. These are: *i)* III-V MOSFETs for logic/digital circuits, *ii)* resistive-RAMs and ferroelectric-FETs for non-volatile memory and in-memory computing, *iii)* GaN-based high-speed transistors for power applications, and *iv)* negative capacitance transistors for biosensing.

Sommario

Negli ultimi anni si è visto emergere una pleteora di dispositivi a semiconduttore innovativi, candidati a essere degni eredi dei transistor basati su Silicio e dando via alla cosidetta era 'post-Moore'. Infatti, nonostante l'elettronica tradizionale sia basata su dispositivi in Silicio, – dai circuiti logici alle memorie, dalle applicazioni ad alta frequenza/potenza alla sensoristica – questo paradigma sta mutando grazie agli sviluppi provenienti da diversi campi di ricerca, dalla fisica e dai materiali a semiconduttore sino alle tecniche di fabbricazione e le architetture dei calcolatori. In questo nuovo, frenetico scenario, i potenziali fattori limite delle performance e dell'affidabilità di nuovi dispositivi devono essere ben compresi in una fase relativamente preliminare dello sviluppo affinché essi possano essere effettivamente presi in considerazione come potenziali sostituti della tecnologia esistente. In questo senso, la simulazione e la modellizzazione fisica rappresentano degli strumenti fondamentali per la comprensione e dunque la progettazione di nuove tecnologie per soddisfare i requisiti dell'elettronica del XXI secolo.

In questa tesi vengono presentati dei metodi di simulazione numerica e di modellizzazione compatta allo stato dell'arte per analizzare i limiti delle performance e dell'affidabilità di differenti tecnologie emergenti. Nello specifico, questa tesi affronta quattro differenti scenari applicativi e le tecnologie corrispondenti candidate come possibili sostituti delle controparti in Silicio. Queste sono: *i)* III-V MOSFETs per circuiti logici/digitali, *ii)* resitive-RAMs e ferroelectric-FETs per memorie non-volatili e elaborazione in-memory, *iii)* transistor ad alta velocità basati su GaN per applicazioni di potenza, e *iv)* transistor a capacità negativa per biosensori.

Contents

Acronyms

k · p **k · p** Method

$C-V$ Capacitance-Voltage Characteristics

$I-V$ Current-Voltage Characteristics

$P-E$ Polarization-Electric Field Relationship

2DEG Two-Dimensional Electron Gas

AI Artificial Intelligence

BE Bottom Electrode

BGF Band Gap Fluctuation

BGOS Back-Gating OFF-State Stress

BioFET FET-based (nano-) biosensor

BLER Body Line Edge Roughness

BT Border Trap

BTE Boltzmann Transport Equation

BTO Barium Titanate

CB Conduction Band

CF Conductive Filament

CMOS Complementary Metal-Oxide-Semiconductor

CVD Chemical Vapor Deposition

DAE Differential Algebraic Equation

DD Drift-Diffusion

DFT Density Functional Theory

DG-UTB Dual-Gate Ultra-Thin-Body

DOS Density of States

DRAM Dynamic Random Access Memory

DUT Device-Under-Test

FeFET Ferroelectric FET

FeRAM Ferroelectric Random Access Memory

FET Field-Effect Transistor

FGOS Front-Gating OFF-State Stress

FinFET Tri-Gate FET

FOM Figure-Of-Merit

GAA Gate-All-Around

GaAs Gallium Arsenide

GCA Gradual Channel Approximation

GLER Gate Line Edge Roughness

HD Hydro-Dynamic

HEMT High Electron-Mobility Transistor

HRS High-Resistance State

InAs Indium Arsenide

InGaAs Indium Gallium Arsenide

InP Indium Phosphide

IRDS International Roadmap for Devices and Systems

ISFET Ion-Sensitive FET

IT Interface Trap

ITF Interface Traps Fluctuation

ITRS International Technology Roadmap for Semiconductors

KCL Kirchhoff's Current Law

KVL Kirchhoff's Voltage Law

LER Line Edge Roughness

LIM Logic-In-Memory

LKE Landau-Khalatnikov Equation

LRS Low-Resistance State

LT Landau Theory

MC Monte Carlo

MFIS Metal-Ferroelectric-Insulator-Semiconductor

MFMIS Metal-Ferroelectric-Metal-Insulator-Semiconductor

MFS Metal-Ferroelectric-Semiconductor

MGG Metal Gate Granularity

MIM Metal-Insulator-Metal

MIS-HEMT Metal-Insulator-Semiconductor HEMT

MLDA Modified Local Density Approximation

MNA Modified Nodal Analysis

MOSFET Metal-Oxide-Semiconductor Field-Effect Transistor

MOVPE Metal Organic Vapor Phase Epitaxy

MSMC Multi Subband Monte Carlo

NBTI Negative Bias Temperature Instability

NC Negative Capacitance

NC-BioFET NC-BioFET

NC-BioFET Negative Capacitance FET-based biosensor

NCFET Negative Capacitance FET

NEGF Nonequilibrium Green's Function

NP-EMA Nonparabolic Effective Mass Approximation

NVM Non-Volatile Memory

NW Nanowire

OTF On-the-Fly

PCRAM Phase-Change Random Access Memory

PUF Physical Unclonable Function

PZT Lead Zirconium Titanate

QA Quantum Approach

QDD Quantum Drift-Diffusion

QHD Quantum Hydro-Dynamic

RDD Random Discrete Dopants

RDF Random Dopant Fluctuation

RNG Random Number Generator

RRAM Resistive Random Access Memory

RTN Random Telegraph Noise

SBT Strontium Bismuth Tantalate

S-IFM Statistical Impedance Field Method

SCA Semi-Classical Approach

SDT Source-Drain Tunneling

SIMS Secondary Ion Mass Spectroscopy

SNR Signal-to-Noise Ratio

SPE Suraface Potential Equation

SR Surface Roughness

SRAM Static Random Access Memory

SRH Shockley-Read-Hall

SRS Surface Roughness Scattering

STT-MRAM Spin-Transfer Torque Magnetic Random Access Memory

TAT Trap-Assisted Tunneling

TB Tight Binding

TCAD Technology Computer-Aided-Design

TE Top Electrode

TEM Transmission Electron Microscopy

TMD Transition Metal Dichalcogenide

UID Unintentionally Doped

VB Valence Band

WBG Wide Band Gap Semiconductor

WFF Work Function Fluctuation

WFP Wave-Function Penetration

List of Symbols

k_B Boltzmann Constant ($8.671 \times 10^{-5}\,\mathrm{eV/K}$)

q Unit of Charge ($1.602 \times 10^{-19}\,\mathrm{C}$)

$\hbar$ Reduced Planck Constant ($1.054 \times 10^{-34}\,\mathrm{J\,s}$)

ε_0 Vacuum Permittivity ($8.854 \times 10^{-12}\,\mathrm{F/m}$)

V_{th} Thermal Voltage ($25.85\,\mathrm{mV}$, at $T = 300\,\mathrm{K}$)

E_F Fermi Energy (eV)

E_i Intrinsic Fermi Energy (eV)

$E_{F,n}$ Electron Quasi-Fermi Energy (eV)

$E_{F,p}$ Hole Quasi-Fermi Energy (eV)

$\boldsymbol{J}_{n,p}$ Electron (Hole) Current Density ($\mathrm{A/m^2}$)

$G_{n,p}$ Electron (Hole) Generation Rate ($\mathrm{cm^{-3}\,s^{-1}}$)

$R_{n,p}$ Electron (Hole) Recombination Rate ($\mathrm{cm^{-3}\,s^{-1}}$)

n, p Free-Electron (Hole) Density ($\mathrm{cm^{-3}}$)

n_i Intrinsic Carrier Density ($\mathrm{cm^{-3}}$)

N_D, N_A Donor (Acceptor) Impurity Density ($\mathrm{cm^{-3}}$)

$\boldsymbol{E}$ Electric Field (V/m)

ρ Space-Charge Density (cm^{-3})

ε_s Semiconductor Dielectric Constant (F/m)

κ_s Semiconductor Relative Dielectric Constant (1)

μ_n, μ_p Electron (Hole) Mobility ($\mathrm{cm}^2/\mathrm{Vs}$)

D_n, D_p Electron (Hole) Diffusion Coefficient (cm^2/s)

ψ Electrostatic Potential (V)

D_{IT} Interface Trap Density of States ($\mathrm{eV}^{-1}\,\mathrm{cm}^{-2}$)

I_D Drain Current (A/metre)

V_{DD} Supply Voltage (V)

I_{ON} ON-State Current (A/metre)

I_{OFF} OFF-State Current (A/metre)

E_C Conduction Band Minimum (eV)

E_V Valence Band Maximum (eV)

E_g Band Gap (eV)

m_n^* Effective Electron Mass (kg)

v_{inj} Injection Velocity (m/s)

C_G Gate Capacitance ($\mathrm{F/m}^2$)

C_{ox} Gate Oxide Capacitance ($\mathrm{F/m}^2$)

C_{inv} Inversion Layer Capacitance ($\mathrm{F/m}^2$)

C_S Semiconductor Capacitance ($\mathrm{F/m}^2$)

V_t Threshold Voltage (V)

SS Sub-threshold Slope (mV/dec)

$\vec{k}$ Wave Vector (m^{-1})

α_p Nonparabolocity Factor (eV^{-1})

W_G Gate Width (m)

L_G Gate Length (m)

V_{GS} Gate to Source Voltage (V)

V_{DS} Drain to Source Voltage (V)

Q_n Mobile Electron Charge (C/cm^2)

t_{ox} Gate Oxide Thickness (m)

EOT Equivalent Oxide Thickness (m)

t_{ch} Channel Thickness (m)

$\mu(V_t)$ Threshold Voltage Average Value (V)

$\sigma(V_t)$ Threshold Voltage Standard Deviation (Variability) (V)

$\sigma_{TOT}(V_t)$ Total Threshold Voltage Variability (V)

N_{SD} Source/Drain Doping (cm^{-3})

Δ_{rms} RMS Amplitude of LER-induced Variations (nm)

Λ_{LER} Correlation Length of LER-induced Variations (nm)

x In Mole Fraction ($In_xGa_{1-x}As$)

N_{trap} Trapped Charge Density (cm^{-2})

N_{inv} Inversion Charge Density (cm^{-2})

t_{CF} Conductive Filament thickness (m)

r_{CF} Conductive Filament radius (m)

S_{CF} Conductive Filament area (m^2)

ρ_{CF} Restivity of Conductive Filament (Ωm)

R_{CF} Restance of Conductive Filament (Ω)

t_B Barrier thickness (m)

R_B Barrier resistance (Ω)

T_B Barrier temperature (K)

I_C Current Compliance (A)

R_{LRS} LRS Resistance (Ω)

V_{RESET} Reset Voltage (V)

V_{READ} Read Voltage (V)

P_s Saturation Polarization (C/cm^2)

P_r Remnant Polarization (C/cm^2)

E_C Coercive Field (V/m)

α Landau 1^{st}-Order Parameter (m/F)

β Landau 3^{rd}-Order Parameter (m^5/FC^2)

γ Landau 5^{th}-Order Parameter (m^9/FC^4)

ρ Landau Damping Parameter (Ωm)

ε_{FE} Ferroelectric dielectric constant (F/cm)

V_{FE} Ferroelectric potential (V)

t_{FE} Ferroelectric layer thickness (m)

C_{FE} Ferroelectric layer capacitance ($\mathrm{F/cm^2}$)

N_A Acceptor doping concentration ($\mathrm{cm^{-3}}$)

μ_{eff} Effective Electron Mobility ($\mathrm{cm^2/Vs}$)

V_{ins} Insulator Voltage (V)

Q_{sw} Switching Charge ($\mu\mathrm{C/cm^2}$)

V_{sw} Switching Voltage (V)

$V_{t,on}$ On-Threshold Voltage (V)

$V_{t,off}$ Off-Threshold Voltage (V)

MW Memory Window (V)

$|V_{P/E}|$ Program/Erase Pulse Amplitude (V)

$t_{P/E}$ Program/Erase Pulse Duration (s)

η Power Conversion Efficiency

R_{ON} ON-Resistance ($\Omega\,\mathrm{mm}$)

V_{BD} Breakdown Voltage (V)

E_{crit} Critical Electric Field Strength for Avalanche Breakdown ($\mathrm{V/m}$)

Φ_B Schottky Barrier Height (V)

Q_{π}^{net} Net Polarization Charge at AlGaN/GaN interface ($\mathrm{C/m^2}$)

ε_{AlGaN} Relative AlGaN Dielectric Constant

t_{bar} AlGaN Barrier Thickness (m)

t_{2DEG} 2DEG Centroid Distance From Interface (m)

n_s 2DEG Electron Density (cm^{-3})

L_{GD} Gate-to-Drain Spacing (m)

E_A Activation Energy (eV)

V_{SUB} Substrate Voltage (V)

ψ_s Surface Potential (V)

Q_s Semiconductor Charge Density (C/cm^2)

Q_{surf} Surface Charge Density (C/cm^2)

Q_{dl} Double-Layer Charge Density (C/cm^2)

ψ_0 Double-Layer Potential (V)

ψ_e Stern Layer Potential (V)

β_s Surface Buffer Capacity (cm^2)

PSD Power Spectral Density (V^2/Hz or A^2/Hz)

V_{INT} Internal gate potential (V)

V_{FB} Flat-band Voltage (V)

A_V Voltage gain

Q_{MOS} MOSFET charge (C/cm^2)

C_{MOS} MOSFET capacitance (F/cm^2)

Q_{BIO} Biomolecules Charge Density (C/cm^2)

λ_{2D} Natural Length of the Device (nm)

N_{2D} 2D Effective Density of States ($eV^{-1}\,cm^{-2}$)

N_d Net Impurity Density (cm^{-2})

N_{BIO} Biomolecules Concentration (cm^{-2})

S_I Current Sensitivity

Chapter 1

Introduction

A common inside joke in the semiconductors community goes roughly as this: "the number of people predicting the demise of Moore's law doubles every two years". Behind the irony, there is a certain amount of truth regarding the ever-increasing challenges in keeping up with the strict demands of the original Moore's law. Since the early days of the semiconductor industry, Moore's law empirically predicted the doubling of transistors count on a chip every two years. The issues related to keeping up with Moore's law by relying on the traditional scaling rules defined by Dennard and co-workers in 1974 [1] became evident already at the end of the previous century, with International Technology Roadmap for Semiconductors (ITRS) forecasting a major roadblock already in 2005. Nonetheless, the number of transistors in a chip kept growing, see Fig. 1.1, and the semiconductor industry reached an outstanding global revenue of $481 billion in 2018 [2], [3].

How was it then possible for the industry to keep growing despite predictions forecast a totally different scenario? Simply put, engineers and researchers were able to devise new strategies to continue scaling the Complementary Metal-Oxide-Semiconductor (CMOS) technology (i.e., the most widespread technology in integrated circuits) and circumvent the issues encountered with Dennard's scaling rules. This is certified by the Deloitte 2019 report that states: "The global semiconductor industry is set to continue its robust growth well into the next decade due to emerging technologies such as autonomous driving, artificial intelligence (AI), 5G and Internet of Things [...]" [3]. The plethora of new applications for the electronic devices stems from the

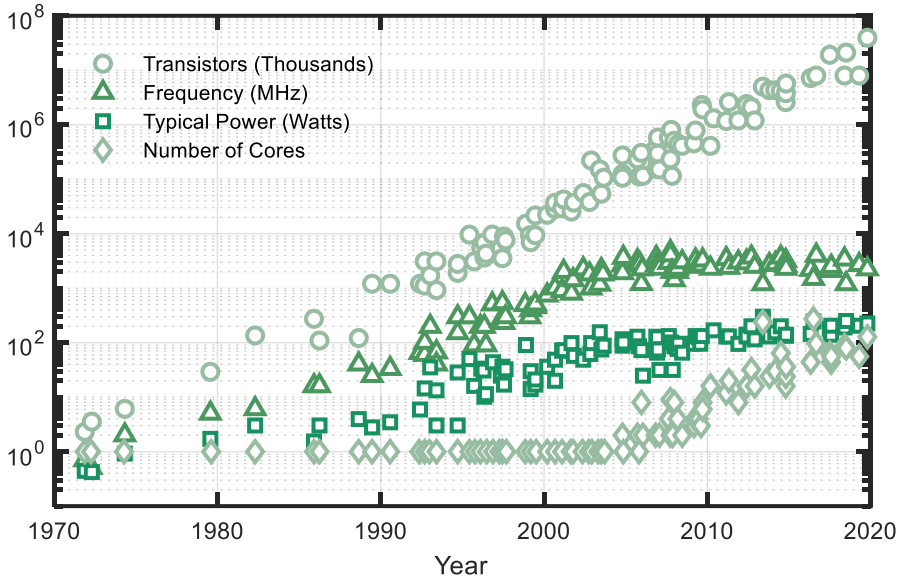


Figure 1.1: Performance indicators vs year of several commercial microprocessor from the early seventies up to year 2019. Circles: thousands of transistors per processor; triangles: operating frequency; squares: power consumption; diamonds: number of logical cores. After [4].

sparkling of novel technologies enabled by the developments in different research fields ranging from physics and semiconductor materials, to processing techniques and computing architectures. Correspondingly, a new international roadmap called International Roadmap for Devices and Systems (IRDS) was established in 2016, reflecting the shift to a top-down, application-oriented development paradigm. As illustrated in Fig. 1.2, IRDS bases its predictions and future directions on a scenario where heterogeneous integration of different functionalities on a single chip will enable the scaling of the technology to achieve even higher efficiency, speed, reliability (at the technology level) and augment the range of functions such as sensing, energy harvesting, energy conversion, analog computing (at the application level). Summarizing, the era of 21st-Century Electronics will be characterized by a plethora of novel technologies that will pair up with (or replace) Silicon-based transistors depending on the requirements of Automotive, Internet of Things, Machine Learning, Computing, Wireless (5G) Communications and many other applications.

In this hectic context, it is vital for technology manufacturers to be able to predict performance and reliability limits of novel solutions in order to speed up the time-to-market and increasing

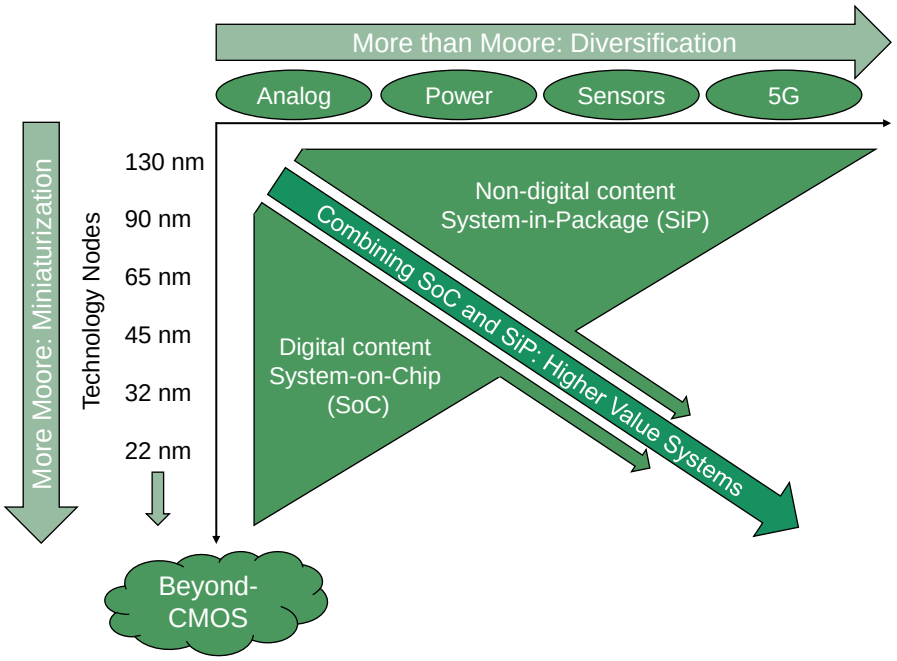


Figure 1.2: More Moore (MM) and More than Moore (MtM) scaling if combined together provide both dimensional and functional scaling that are essential for the continuing development of semiconductor industry. Beyond-CMOS includes non-traditional devices for continuing dimension scaling. Adapted from [5].

process yield. These goals can only be met by recurring to simulation and modeling tools that allow gaining insights into device behavior and observing phenomena that could not be measured on real devices (due to issues related to production costs and timelines) [6]. Mimicking devices (or circuits) behavior with simulations to reproduce experimental results has actually been common practice since the late 1970's. However, with the recent emergence of novel technologies either at the nanometre scale, of different semiconductor materials, as well as of different physical principles to encode information (e.g., spin-based devices), simulations have become fundamental for understanding – and thus engineering – these technologies. Goal of well-calibrated simulation models is to reproduce (and predict) device behavior under different conditions (that can be for instance geometrical parameter variations, bias, temperature, etc.) in a relatively short time compared to real experiments. One of the main concerns of the device physicist/engineer is the trade-off between accuracy and simulation time. Thus, any good simulation set-up should strive to be as comprehensive as possible to consistently reproduce and predict measurement results in

the least (computational) time possible.

In this thesis, state-of-the-art simulation and compact modeling tools are exploited to analyze the performance and reliability limits of several different emerging technologies. Specifically, in this dissertation we address four application scenarios and the corresponding candidate technologies to provide enhanced performance compared to Si-based counterparts. These are: *i)* III-V Metal-Oxide-Semiconductor Field-Effect Transistors (MOSFETs) for logic/digital circuits, *ii)* Resistive Random Access Memories (RRAMs) and Ferroelectric FETs (FeFETs) for non-volatile memory and in-memory computing, *iii)* GaN-based High Electron-Mobility Transistors (HEMTs) for power applications, and *iv)* Negative Capacitance FETs (NCFETs) for biosensing. In the context of Fig. 1.2, III-V MOSFETs belong to the 'More Moore' chapter of the IRDS roadmap, RRAMs and FeFETs belong to the 'Beyond CMOS' chapter [7], GaN HEMTs to the 'More than Moore', and Negative Capacitance FET-based biosensors (NC-BioFETs) make use of a 'Beyond CMOS' technology (i.e., NCFETs) to improve a 'More than Moore' device such as the nano-biosensor.

In the remainder of this chapter we briefly introduce some general concepts about the simulation paradigms that are at the basis of any theoretical analysis concerning semiconductor devices. Then, we will discuss process and device simulations that are used to understand in detail the fabrication processes and physical mechanisms influencing the single device operation. Finally, we discuss SPICE simulations that are employed to evaluate the single (or multiple) device behavior in the context of circuit applications.

1.1 Semiconductor Device Simulation and Modeling

Due to the complexity of device's geometry and the amount of physical processes involved, semiconductor device simulations are carried out at three different levels of abstraction [8]. This allows to effectively understand the influence of single processing steps or physical models in the general behavior of the device. This differentiation gives rise to a hierarchy of abstraction levels in which the devices are considered. The schematic of the general hierarchy of semiconductor device simulation is depicted in Fig. 1.3. The first tier is represented by the so-called process simulations

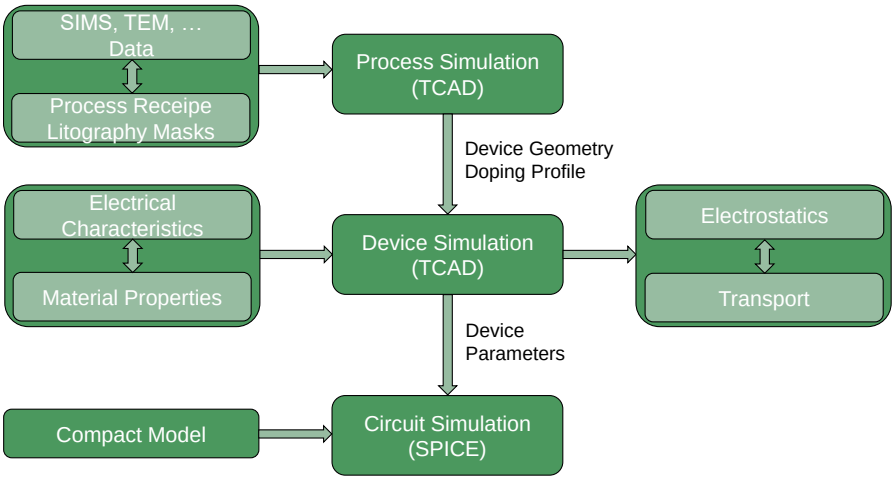


Figure 1.3: General Hierarchy of Semiconductor Device Simulation.

that aims at virtually reproducing the fabrication steps performed in the white-room to realize a semiconductor device. The second tier is represented by the actual device simulations that aim at reproducing (and predicting) with physical models the electrical behavior of a device. The final tier is constituted by the circuit simulations that use device models to produce complex functions at circuit level (i.e., with many devices connected together). In the following we describe more in detail each of the hierarchy tiers.

1.1.1 Process Simulation

The models implemented in the process simulation tier range from the epitaxial growth of semiconductor layers to implantation and diffusion of doping atoms in the semiconductor regions to oxidation, etching and lithography. The simulation models used to reproduce device processes are normally included in softwares of Technology Computer-Aided-Design (TCAD) suites such the commercial Sentaurus Process (Synopsys, Inc.), and Victory Process (Silvaco, Inc.) as well as other open-sources alternatives. The inputs required by process simulation are the doping profiles as obtained from Secondary Ion Mass Spectroscopy (SIMS) measurements, topography provided by Transmission Electron Microscopy (TEM), the process recipe, and the lithography masks, see Fig. 1.3. These inputs are necessary in order to reproduce accurately the actual fabricated device geometry and doping profiles. The recreated device geometry and doping profiles are then used

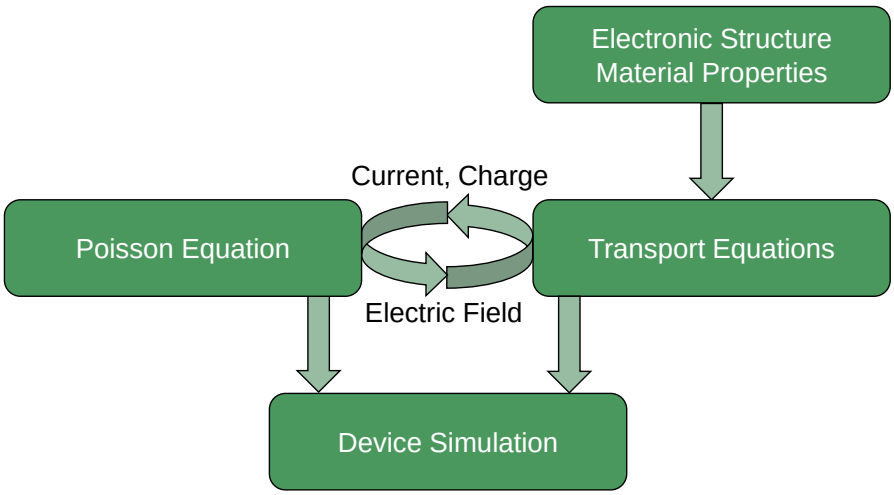


Figure 1.4: Schematic of the Device Simulation Sequence. Adapted from [6].

as input for the proper device simulations, described in the following section[1].

1.1.2 Device Simulation

After the device geometry and doping profiles have been generated by the process simulation, the device (physics) simulation can take place. As for the previous tier, device simulations are carried out by TCAD tools (e.g., the commercial Sentaurus Device and Victory Device) that include several physical models to compute the electromagnetic fields and the charges flowing inside of a device to determine its electrical behavior. To perform these simulations, the device is sub-divided into a discretized structure called *mesh*. Thus each differential equation of the system defining the device behavior is solved at each of the mesh vertices. Clearly, both accuracy and speed of the simulation heavily depend on the definition of the mesh: a fine mesh leads to more accurate results than a coarse one but at the cost of higher computational time.

As for the process simulation, device simulation requires input files obtained from the measurements of electrical characteristics (e.g., Current-Voltage Characteristics ($I-V$) and Capacitance-Voltage Characteristics ($C-V$)) and material properties obtained from either measurements or by sophisticated calculations performed at the microscopic level (e.g., Density

[1] We mentioned process simulations for the sake of completeness. In fact, for the rest of this thesis we will not make further use of them, rather we will use tools that allow defining both geometry and doping in a simplified manner than can be used as input to device simulations.

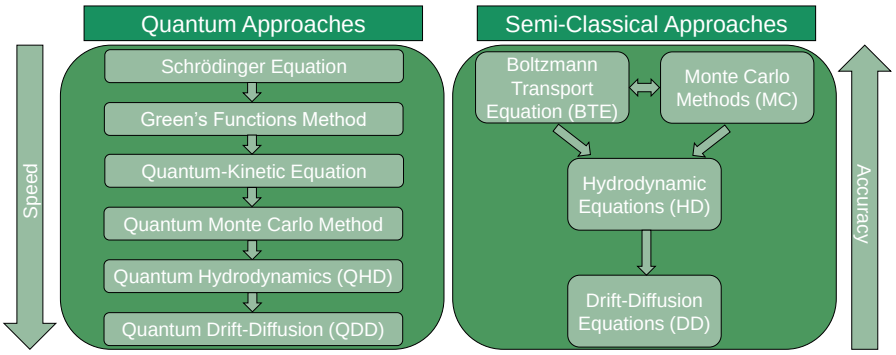


Figure 1.5: Hierarchy of Transport Models.

Functional Theory (DFT) simulations to determine the band-structure, Monte Carlo (MC) simulations to obtain the carriers' mobility profiles, etc.). This data is necessary in order to *calibrate* the results obtained by the simulations with the real device characteristics. The calibration not only guarantees that simulations represent the actual device behavior but also that they can be used to predict behavior when for instance device parameters are varied.

The schematic representation of the device simulation sequence is depicted in Fig. 1.4. Basically, a device simulation run computes the solution of the coupled electrostatic and transport equations. The electrostatics is generally determined by the Poisson equation [6] and allows to determine the electric field distribution across all regions and layers of a device. The field is used to determine the current and charge in the device from the carrier density profiles determined by the transport equations. The charge computed via the transport equations is fed-back to the Poisson equation and the final solution is found by a self-consistent iterative procedure. In this scheme, a list of parameters describing the electronic structure and material properties needs to be provided for each of the different layers constituting the device geometry to obtain meaningful results.

While the electrostatic problem is essentially determined by the solution of the Poisson equation, the transport problem can be approached at different levels of accuracy depending on the complexity of the geometry and on the dimension scale of the device [6]. The transport models are generally categorized depending on whether they are based on semi-classical or quantum approaches, as shown in Fig. 1.5.

Semi-Classical Approaches

The Semi-Classical Approaches (SCAs) treat electronic charges as particles and make use of either approximations of the Boltzmann Transport Equation (BTE) or solve exactly the problem by using MC methods to determine (statistically) the dynamics and scattering of each particle involved into the device behavior [9][2]. However, due to the complexity and computational burden of MC methods, BTE is generally approximated by using the so-called moment's method that creates a set of balance equations derived from the BTE itself by performing appropriate integrals and averages over the momentum space [9][3]. The models approximating the BTE that are mostly used are the Hydro-Dynamic (HD) and Drift-Diffusion (DD) equations.

Quantum Approaches

When the device dimensions scale down to the nanometer level, carrier transport behaves quite differently than how would SCAs predict. This is because the carriers' wave nature giving rise to phenomena like tunneling and limited (or absence of) scattering in their traveling range leads to more complex behavior. This is where Quantum Approaches (QAs) come into play. Fig. 1.5 summarizes the transport models based on QAs. The most general and also complex QA is based on the direct solution of the Schrödinger equation, to determine the wave function for each carrier involved in the device operation. This approach however is rather impractical unless for simplified problems where few carriers are considered. At the other end of the spectrum, there exist the Quantum Hydro-Dynamic (QHD) and Quantum Drift-Diffusion (QDD) equations that are the quantum equivalent of the relative SCA. These equations resemble basically the SCA ones,

[2]It is worth mentioning that much progress has been made in the last 10-15 years in developing the deterministic methods to solve the BTE for low dimensional systems, see for instance [10], [11] and [12] for deterministic solutions of the BTE for a 1D electron gas (e.g. Nanowire (NW) transistors) and for 2D systems, respecitvely.

[3]The main issue when approximating the BTE is the necessity to make an *a priori* assumption on the carrier distribution function, which is not required by MC methods. For instance, the well-known Fermi distribution function

$$f(E) = \frac{1}{1+\exp\left[\frac{(E-E_F)}{k_B T}\right]} \tag{1.1}$$

(where E is the energy of the carrier, and T the lattice temperature) is in fact only valid for near-equilibrium transport, which is often not the case [9].

including correction in the carrier density profile to take into account quantum confinement[4]. More details regarding each of these approaches can be found in [6].

1.1.3 Drift-Diffusion Equations

We now describe the general equations of the DD model as they constitute the simplest treatment of semiconductor device behavior. The DD model of transport is the most simplified view of how actually a device works. Nevertheless, this paradigm was employed for a long time in simulations (and still is) thanks to its simplicity and ease of implementation in numerical solvers, constituting the starting point of the understanding of most semiconductor devices. The DD equations describe the carriers motion in a semiconductor crystal with applied external electric field as a combination of acceleration and scattering events through the paths of the carriers that determine the current flows [13]. The DD equations are also known as the Semiconductor Equations and are summarized as follows (for both electrons and holes)

$$\frac{\partial n}{\partial t} = -\nabla \cdot \left(\frac{\boldsymbol{J}_n}{q} \right) + G_n - R_n \tag{1.2a}$$

$$\frac{\partial p}{\partial t} = -\nabla \cdot \left(\frac{\boldsymbol{J}_p}{q} \right) + G_p - R_p \tag{1.2b}$$

$$\nabla \cdot (\varepsilon_s \boldsymbol{E}) = \rho. \tag{1.2c}$$

Eqs. (1.2a) and (1.2b) are the *continuity* equations for electrons and holes, respectively. They basically express a balance equations for electron (hole) density over time, taking into account the in/out flux (expressed by the current $\boldsymbol{J}_{n,p}$) and the generation/recombination rates (expressed by $G_{n,p}$ and $R_{n,p}$, respectively). Eq. (1.2c) is the local-form of Maxwell's equation resembling the Poisson equation in space. To solve Eqs. (1.2a) to (1.2c) one needs to write three additional equations to determine $\boldsymbol{J}_{n,p}$ and $\boldsymbol{E}$.

The electron and hole current densities ($\boldsymbol{J}_{n,p}$) are determined by the so-called *current* equa-

[4]This aspect will be discussed more in detail when analyzing the behavior of ultra-scaled III-V MOSFETs in Chapter 2.

tions, that read

$$\boldsymbol{J}_n = q\mu_n n\boldsymbol{E} + qD_n\nabla n \tag{1.3a}$$

$$\boldsymbol{J}_p = q\mu_p p\boldsymbol{E} - qD_p\nabla p. \tag{1.3b}$$

Both mobility and diffusion coefficient depend on electric field in general; however, under non-degenerate conditions (i.e., the Fermi level is *inside* the band-gap) Einstein relationship can be shown to hold true

$$D_{n,p} = \mu_{n,p}\left(\frac{k_B T}{q}\right) = \mu_{n,p} V_{th}. \tag{1.4}$$

The free carrier's densities at equilibrium can be simply determined by using Boltzmann approximation of Fermi-Dirac's statistics, obtaining [13]

$$n = n_i \exp\left[\frac{(E_F - E_i)}{k_B T}\right], \quad p = n_i \exp\left[\frac{(E_i - E_F)}{k_B T}\right]. \tag{1.5}$$

Under equilibrium condition, $np = n_i^2$ ($n_i = 1 \times 10^{10}\,\mathrm{cm}^{-3}$ for Silicon). Under non-equilibrium conditions, it is no longer possible to define a single Fermi energy for the whole semiconductor. It is thus useful to write *quasi* Fermi levels for electrons ($E_{F,n}$) and holes ($E_{F,p}$) so that the two pools of carriers result under equilibrium within themselves. Now we have

$$n = n_i \exp\left[\frac{(E_{F,n} - E_i)}{k_B T}\right] \tag{1.6a}$$

$$p = n_i \exp\left[\frac{(E_i - E_{F,p})}{k_B T}\right] \tag{1.6b}$$

$$np = n_i^2 \exp\left[\frac{(E_{F,n} - E_{F,p})}{k_B T}\right]. \tag{1.6c}$$

Thus finally by using Eqs. (1.4), (1.6a) and (1.6b) to rewrite Eqs. (1.3a) and (1.3b) we obtain

$$\boldsymbol{J}_n = -\mu_n n\nabla E_{F,n} \tag{1.7a}$$

$$\boldsymbol{J}_p = -\mu_p p\nabla E_{F,p}. \tag{1.7b}$$

Eqs. (1.7a) and (1.7b) are the form of current equations commonly employed in simulators.

The electric field ($\boldsymbol{E}$) is obtained through ρ, which can be solved for by applying the charge-neutrality condition in the bulk of the semiconductor

$$\rho = -q\left(n - p + N_A - N_D\right). \tag{1.8}$$

By substituting Eq. (1.8) in Eq. (1.2c), coupled self-consistently with Eqs. (1.7a) and (1.7b), $\boldsymbol{E}$ can be determined and, consequently, the electrostatic potential (ψ) in the device as $\boldsymbol{E} = -\nabla\psi$.

1.1.4 Circuit Simulation and Compact Modeling

According to the scheme in Fig. 1.3, the last tier of semiconductor device modeling is represented by circuit simulation augmented by the use of compact models. Circuit simulations are most commonly carried out with a SPICE simulator, that constructs a mathematical model of the circuit comprising the device and the external components, the topological network of the circuit itself as well as the parasitic elements (resistors, capacitors, inductors) [14]. A circuit simulator basically constructs a system of coupled non-linear Differential Algebraic Equations (DAEs) of the form

$$\boldsymbol{F}(\boldsymbol{x}, \dot{\boldsymbol{x}}, t) = 0 \tag{1.9}$$

where $\boldsymbol{x}$ are the unknowns of the system. Based on the Modified Nodal Analysis (MNA), the system of Eq. (1.9) can be augmented to solve for the branch voltages and nodal currents for each element of the circuit including the external sources [14]. The problem is solved by employing iterative algorithms that allow to find the self-consistent solution of the system (if it exists). Note that a circuit simulator does not solve for Maxwell's equations but rather for Kirchhoff's Voltage Laws (KVLs) and Kirchhoff's Current Laws (KCLs), thus the wires are treated as ideal conductors and hence the external fields are assumed not to affect each element internal behavior. This in turn implies that each element must be charge neutral within itself [14].

Compact models represent the mathematical abstraction of complex device behavior, written commonly in Verilog-A [14] in such a way to be solved fast, accurately and robustly by the circuit simulator. Compact models are so widely employed that there exist industry-standard models that are employed by companies to mimic the behavior of their own products (e.g., BSIM,

PSP, EKV for transistors[5]). As for process and device simulations, compact models need to be calibrated in order to reproduce (and predict) accurately the behavior of real components. The calibration in this case is done by extracting the device parameters of interest from the output of device TCAD simulations. This way one guarantees that the compact model is able to reproduce a device behavior in a circuit application.

1.2 Outline

Having defined the simulation and modeling frameworks that will be used in this thesis, we now briefly summarize the content discussed in detail in the next chapters

1. Chapter 2 discusses ultra-scaled III-V MOSFETs for future nodes of Moore's law. TCAD simulations are employed to investigate the influence of traps in the mobility extraction method and to study variability issues.

2. Chapter 3 treats the modeling of noise and endurance in RRAMs and FeFET, respectively.

3. Chapter 4 investigates the stability and reliability issues in GaN HEMTs for power applications. TCAD simulations are employed to reveal the role of Carbon doping in the buffer on breakdown voltage, on-resistance degradation and recovery transients and bidirectional threshold voltage shifts.

4. Chapter 5 presents the concept of a potentiometric nano-biosensor based on the 2D semiconductor MoS_2 exploiting the *negative capacitance* effect to improve sensitivity and Signal-to-Noise Ratio (SNR).

Finally, Chapter 6 draws the conclusions of this work and offers an outlook on the future of TCAD simulations augmented with Artificial Intelligence (AI) tools.

[5]A more exhaustive list of transistor compact models is available at https://en.wikipedia.org/wiki/Transistor_model#Popular_models.

References – Chapter 1

[1] R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. R. Leblanc, "Design of Ion-Implanted MOSFET's With Very Small Physical Dimensions," *IEEE Journal of Solid-State Circuits*, vol. 9, no. 5, pp. 256–268, Oct. 1974. DOI: 10.1109/JSSC.1974.1050511.

[2] *Semiconductor industry - Wikipedia*. [Online]. Available: https://en.wikipedia.org/wiki/Semiconductor_industry (visited on 10/14/2020).

[3] Deloitte, "Semiconductors-the Next Wave Opportunities and winning strategies for semiconductor companies," Tech. Rep., 2019.

[4] K. Rupp, *40 Years of Microprocessor Trend Data*. [Online]. Available: https://www.karlrupp.net/2015/06/40-years-of-microprocessor-trend-data/ (visited on 11/13/2020).

[5] IEEE, "International Roadmap for Devices and Systems (IRDS) 2020 - More than Moore White Paper," Tech. Rep., 2020. [Online]. Available: https://irds.ieee.org/editions/2020/more-than-moore.

[6] D. Vasileska, S. M. Goodnick, and G. Klimeck, *Computational Electronics*, 1st ed. Boca Raton (FL, USA): CRC Press, 2010. DOI: 10.1201/b13776.

[7] IEEE, "International Roadmap for Devices and Systems (IRDS) 2020 - Beyond CMOS," 2020. [Online]. Available: https://irds.ieee.org/editions/2020/beyond-cmos.

[8] A. Gehring, "Simulation of Tunneling in Semiconductor Devices," PhD Dissertation, TU Wien, 2003. [Online]. Available: https://www.iue.tuwien.ac.at/phd/gehring/diss.html.

[9] D. Esseni, P. Palestri, and L. Selmi, *Nanoscale MOS Transistors*, 1st ed. Cambridge (UK): Cambridge University Press, 2011. DOI: 10.1017/CBO9780511973857.

[10] M. Lenzi, P. Palestri, E. Gnani, S. Reggiani, A. Gnudi, D. Esseni, L. Selmi, and G. Baccarani, "Investigation of the transport properties of silicon nanowires using deterministic and Monte Carlo approaches to the solution of the Boltzmann transport equation," *IEEE Transactions on Electron Devices*, vol. 55, no. 8, pp. 2086–2096, 2008. DOI: 10.1109/TED.2008.926230.

[11] O. Badami, E. Caruso, D. Lizzit, P. Osgnach, D. Esseni, P. Palestri, and L. Selmi, "An Improved Surface Roughness Scattering Model for Bulk, Thin-Body, and Quantum-Well MOSFETs," *IEEE Transactions on Electron Devices*, vol. 63, no. 6, pp. 2306–2312, 2016. DOI: 10.1109/TED.2016.2554613.

[12] C. Jungemann, J. A. Pham, B. Meinerzhagen, C. Ringhofer, and M. Bollhöfer, "Stable discretization of the Boltzmann equation based on spherical harmonics, box integration, and a maximum entropy dissipation principle," *Journal of Applied Physics*, vol. 100, no. 2, 2006. DOI: 10.1063/1.2212207.

[13] W. Fichtner, D. J. Rose, and R. E. Bank, "Semiconductor Device Simulation," *IEEE Transactions on Electron Devices*, vol. 30, no. 9, pp. 1018–1030, 1983. DOI: 10.1109/T-ED.1983.21256.

[14] C. C. McAndrew, G. J. Coram, K. K. Gullapalli, J. R. Jones, L. W. Nagel, A. S. Roy, J. Roychowdhury, A. J. Scholten, G. D. J. Smit, X. Wang, and S. Yoshitomi, "Best Practices for Compact Modeling in Verilog-A," *IEEE Journal of the Electron Devices Society*, vol. 3, no. 5, pp. 383–396, Sep. 2015. DOI: 10.1109/JEDS.2015.2455342.

Chapter 2

Traps and Variability Issues in III-V MOSFETs

Abstract — In this chapter we investigate the issues related to traps and variability in III-V MOSFETs.

First, we perform a combined experimental/simulation analysis of InGaAs planar MOSFETs to evaluate the effects of Interface Traps (ITs) and Border Traps (BTs) on the electrical characteristic and on the accuracy of split-CV mobility measurements. We find that BTs induce hysteresis in the quasi-static $I-V$ characteristics and lead to frequency dispersion in the $C-V$ curves. Moreover, we find that traps affect mobility measurements due to the spurious contributions of trapped charge to the total gate charge.

Second, we investigate the statistical variability of threshold voltage in InGaAs ultra-thin-body MOSFETs. InGaAs devices have comparatively higher variability than Si counterparts as a result of stronger quantization effects in the ultra-thin channel. The most relevant variability sources are the gate work function fluctuation and line edge roughness. Dimension scaling enhances V_t sensitivity to t_{ch} variations, configuring the control over t_{ch} as the most critical design issue. At extremely scaled nodes, the variability due to interface traps plays a significant role that could ultimately contrast the trend of reduced variations with smaller dimensions, as opposite to Si devices.

2.1 Scaling of Logic Transistors — III-Vs to the rescue

As mentioned in the Introduction, the strict requirements demanded by Moore's law to keep up with its predictions led to the exploration of novel technologies to extend the scaling of CMOS technology. One pathway in the quest towards enhanced performance at smaller dimensions was towards replacing the Si channel material with III-V compound semiconductors. III-V's offer enhanced mobility compared to Si leading to uncompromised performance when reducing supply voltage [1]. One of such III-V materials in Indium Gallium Arsenide (InGaAs), that has an outstanding bulk (electron) mobility of $10\,000\,\mathrm{cm}^2/\mathrm{Vs}$ against $1400\,\mathrm{cm}^2/\mathrm{Vs}$ of Si (i.e., about a factor of 10 improvement) [2]. Since the speed of an integrated CMOS chip is proportional to the output current I_D (for a given supply voltage, V_{DD}) and, in turn, to the mobility μ_n [4], then clearly a mobility boost would directly translate to speed enhancement. Although several ultra-scaled and high-speed III-V MOSFETs have already been experimentally demonstrated [5]–[7], these devices still suffer from parasitic effects that reduce their potential benefits over their Si counterparts [41], [48], [55]. In particular, these issues are relative to: *i)* the high density of defects in the gate oxide (i.e., BTs) and at the interface with the channel (i.e., ITs), *ii)* the statistical variability that causes identical devices to have different performance, *iii)* the low Density of States (DOS) reducing the maximum achievable electron density, and *iv)* the band-to-band tunneling leading to high off-state leakage [8, Ch. 1]. These issues still need to be fully understood — and consequently reduced — through proper engineering solutions to fully exploit the superior, intrinsic channel material properties of InGaAs.

In this chapter we will present the research work on the detrimental role of ITs and BTs on the mobility and on the variability of III-V MOSFETs. The findings in this chapter were obtained by performing TCAD device simulations calibrated either on experimental results or on more sophisticated simulation models, as discussed later. Before discussing these results, we will provide some background knowledge by reviewing relevant prior work on this topic.

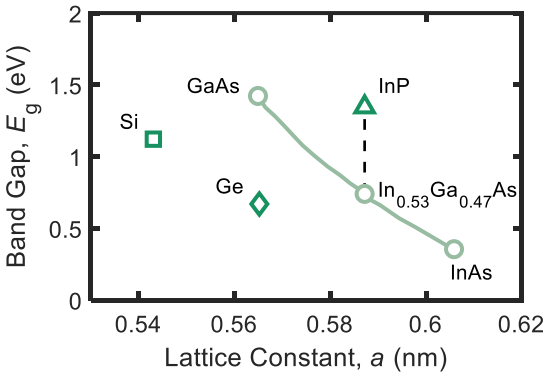


Figure 2.1: Energy bandgap E_g vs lattice constant a of different semiconductor materials, showing that $In_{0.57}Ga_{0.43}As$ has the same lattice constant as InP. Adapted from [3].

2.2 Background

2.2.1 InGaAs Properties

As mentioned before, InGaAs is a compound material combining the properties of column III and V semiconductors to achieve high speed and low power consumption. Basically, InGaAs is a mixture of the binary compounds Indium Arsenide (InAs) and Gallium Arsenide (GaAs), where the molar fraction x of In with respect to Ga defines the electrical properties of the material. We focus on the $In_{0.57}Ga_{0.43}As$ compound as this particular stoichiometry gives a semiconductor that is lattice-matched to Indium Phosphide (InP), as shown in Fig. 2.1, allowing to obtain high quality epitaxial layers with either no tensile or compressive strain [9], [10]. InGaAs has a bulk mobility of about $10\,000\,\mathrm{cm}^2/\mathrm{Vs}$ and a band-gap E_g of about $0.74\,\mathrm{eV}$. Clearly the properties of InGaAs lay somewhere in between those of InAs and GaAs and represent a good compromise to guarantee both high I_{ON} (through mobility) and low I_{OFF} (through bangap) [4].

The high mobility of InGaAs can be attributed to the low effective mass of this material compared to that of Si. In the classical picture in fact, the velocity of an electron in a semiconductor crystal in presence of an external electric field can be simply determined as

$$v_n = \frac{p_n}{m_n^*} = -\frac{q\tau_n}{m_n^*}E_x = -\mu_n E_x \tag{2.1}$$

where v_n is the electron velocity, $p_n = -q\tau_n E_x$ its momentum (with τ_n being the mean time

between scattering and E_x the electric field component along the arbitrary 1D direction), and m_n^* is the electron effective mass. From this simple relationship it becomes clear that as m_n^* is reduced, μ_n is increased[1]. The low effective mass of InGaAs ($m_t = 0.041\,\mathrm{m}_0$ [2]) compared to that of Si ($m_t = 0.19\,\mathrm{m}_0$) not only leads to a mobility improvement but also to a reduction of the available states in the Conduction Band (CB) (i.e., the so-called DOS bottleneck) [12] and in an increase of quantization effects [11]. Strong quantization and low DOS lead on the one hand to a reduction in electron available for conduction and hence reduced I_{ON}, and, on the other hand, to an increased probability of source-drain tunneling hand hence higher I_{OFF} [12]. Thus, design of InGaAs devices need to careful take into account the material properties in order to achieve higher performance compared to Si counterparts. Previous theoretical analysis considering the trade-offs between the aforementioned aspects concluded that III-V semiconductors could offer enhanced performance for CMOS applications only for low power operation [11].

2.2.2 Quantum Correction to Carrier Density in Ultra-Scaled Devices

In this section we briefly discuss how quantization effects can be accounted for in TCAD simulations to correctly model the electron density in the channel. In classical MOSFET theory, it is generally assumed that an infinitely small inversion layer is formed between the source and drain (i.e., the channel), leading to the conclusion that above threshold the gate capacitance C_G is equal to C_{ox} [13]. Actually, the inversion layer has a finite thickness because the electrons crowding near the interface with the insulator have a peak in their distribution (called centroid) which is physically distant from the interface, giving a finite inversion layer thickness and hence finite capacitance C_{inv}. As device dimensions reduce, the inversion layer thickness becomes important (because the relative ratio between C_{inv} and C_{ox} decreases) as C_G is defined as the series of C_{ox} and C_S (i.e., the semiconductor capacitance). In addition, it can be shown that C_S is directly related to the DOS of the semiconductor [4], thus for materials like InGaAs the C_G reduction due to quantum confinement is more prominent than for Si.

[1] Generally, carrier transport for InGaAs devices (especially at nanometre scale) requires a more detailed treatment due to the limits of the classical theory and the appearance of quantum mechanical effects. Revision of the classical MOSFET theory led to the definition of a quantity called injection velocity v_{inj} upon which a direct relationship with I_{ON} can be established (rather than on mobility) [4]. Nonetheless, it can still be shown that v_{inj} inversely depends on m_n^*, keeping the argument about speed improvement with m_n^* reduction valid [4].

In order to evaluate correctly the carrier density in the channel (and C_G consequently) a quantum mechanical treatment is required. Taking into consideration the bandgap widening induced by the discreteness of the CB available levels (for electrons), the classical carrier density becomes

$$n = N_C \exp\left(\frac{E_{F,n} - E_C - \Lambda_n}{k_B T}\right) \tag{2.2}$$

where N_C is the conduction band effective density of states [13], and Λ_n is the first available energy level (i.e., the first subband) in the conduction band [4]. This approach is indeed the one followed by TCAD simulation softwares [14].

2.2.3 InGaAs MOSFET Architectures

Before discussing InGaAs MOSFET architectures, it is important to stress on their operation mode, which differs from classical Si MOSFETs. As mentioned previously in fact, in classical n-type MOSFETs the channel charge is composed of electrons that are the minority carriers in the semiconductor body region (p-type). This is why the channel layer is also called the *inversion* layer and the device 'inversion-type'. InGaAs MOSFETs on the other hand, are called *accumulation-type* because the body is normally undoped or lightly doped (to compensate for the native conductivity).

As for conventional Si-based MOSFETs, several device architectures were explored for InGaAs, ranging from planar [5], [18] to Dual-Gate Ultra-Thin-Bodys (DG-UTBs) [10], Tri-Gate FETs (FinFETs) [6], [7], [19], and Gate-All-Around (GAA) NWs [20], [21]. The cross-section of the planar MOSFET discussed in [18] (and used for the analysis in Section 2.3) is shown in Fig. 2.2. Depending on the particular device architecture, several critical aspects related to either quantum confinement or tunneling become more or less severe. This is true also for the effect of traps on the device electrostatic behavior and transport. In Section 2.3 and Section 2.4 we will present simulation results for the planar MOSFET and the DG-UTB architecture, respectively.

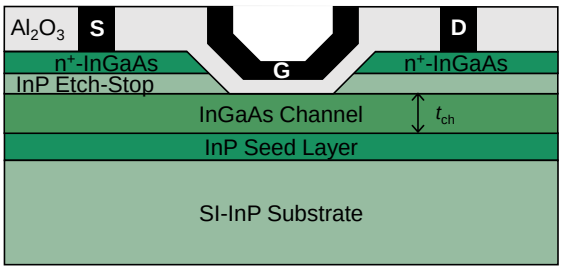


Figure 2.2: Sketch of the cross-section of the simulated InGaAs planar MOSFET.

2.2.4 Simulation Approaches

InGaAs is a direct bandgap material, meaning that the CB minimum and Valence Band (VB) maximum are aligned in the $k-p$ space, thus making it an excellent material also for photonics applications (e.g. LEDs, lasers). Its band structure is quite different from that of Si [2] in the sense that it exhibits strong non-parabolicity in the vicinity of CB minimum (and VB maximum). Conventional TCAD simulators based on the DD equations and on the near-equilibrium transport[2] treat the band structure in a simplified manner such that it can be approximated as a (quasi-) parabolic function near its local minimum/maximum. Thus, the energy of carriers in the vicinity of the CB minimum can be simply written as

$$E(\vec{k})[1+\alpha_p E(\vec{k})] = E_C + \frac{(\hbar\vec{k})^2}{2m_n^*} \tag{2.3}$$

where α_p is the so-called nonparabolicity factor (for $\alpha_p = 0$ we obtain a parabolic energy dispersion) that accounts for the deviation from parabolic shape when moving away from the CB minimum. InGaAs exhibits strong nonparabolicity that affects effective mass [23] as well as DOS calculation[3]. It is therefore of paramount importance to determine the band structure accurately and thus extract the material parameters necessary for device simulations. Different methods can be used to determine the band structure, for instance the atomistic DFT, the empirical

[2]Near-equilibrium is a condition for which the carrier density can be accurately determined by Fermi-Dirac statics, see Eq. (1.1). This condition does not generally apply for ultra-scaled devices where strong lateral fields are present and dimensions are comparable to the mean-free-path for scattering [4]. In this case, 'far-from-equilibrium' transport models are required.

[3]Actually, m_n^* is estimated from Eq. (2.3) itself, thus an accurate description of energy dispersion relation is crucial for obtaining correct simulation results.

Tight Binding (TB), the $\mathbf{k}\cdot\mathbf{p}$ Method ($\mathbf{k}\cdot\mathbf{p}$), and Nonparabolic Effective Mass Approximation (NP-EMA) [22], [24]. The $\mathbf{k}\cdot\mathbf{p}$ method offers an intermediate level of approximation with a computational burden between DFT, TB and the least accurate NP-EMA. Thus, $\mathbf{k}\cdot\mathbf{p}$ if often used for full quantum transport simulations and the NP-EMA method for the semiclassical transport approaches [22].

As we mentioned previously, due to their smaller m_n^*, InGaAs MOSFETs are more severely impacted by tunneling effects (leading to leakage currents) than their Si counterparts. A rigorous treatment of Source-Drain Tunneling (SDT) (which is expected to be the most prominent leakage contribution [22]) can be carried out only with full quantum simulations, that are commonly solved with the Nonequilibrium Green's Function (NEGF) method based on either TB, $\mathbf{k}\cdot\mathbf{p}$, or NP-EMA for the band structure calculation [22]. On the other hand, semiclassical methods based either on Multi Subband Monte Carlo (MSMC) [9] or QDD (such as that used by TCAD simulators) have severe limitations when it comes to SDT modeling [22].

As far as mobility is concerned, the most accurate way to extract it accounting for all the scattering mechanisms is by using MSMC simulations. It has been shown in [26] that short devices still operate quite far from the ballistic limit, thus semiclassical approaches can accurately describe transport. The main limiting factor to mobility in the on-state (when there are no interface traps included [26]) is the Surface Roughness Scattering (SRS) [22].

On the TCAD side, a methodology to account for quasi-ballistic transport in DG-UTB devices was developed in [25]. The so-called ballistic mobility was then combined to the scattering-limited (semiclassical) mobility via a Matthiessen-like rule. In Section 2.4 we will present a different approach for mobility estimation in TCAD simulations that requires calibration with MSMC simulations to account for additional scattering due to ITs [55].

As we will see in Section 2.3, mobility measurements of InGaAs transistors via the split $C-V$ method is affected by the trapped charge that overestimates the actual charge available for conduction [30]. Thus, simulation strategies are required in order to extract the intrinsic mobility of the devices [41].

2.2.5 Effects and Signatures of ITs and BTs

One of the most pressing challenge to the development of fast and reliable III-V MOSFETs is represented by ITs and BTs[4]. This is because unlike Si, III-Vs do not normally have a native oxide such as SiO_2, therefore the phenomenon called "Fermi-level pinning" occurs [15], leading to a limited accumulation charge in the channel. ITs and BTs belong to the so-called "switching states" [16] that communicate with the channel (by exchanging charge) on a time scale comparable to that of measurement conditions. As the name suggests, ITs are those traps located at the interface between the semiconductor and the gate insulator whereas BTs are located in the oxide itself near the interface (few nanometers away) and interact with the semiconductor by quantum tunneling [17]. Both ITs and BTs lead to several drawbacks such as frequency dispersion in $C-V$ measurements [17], [22], hysteresis in both $I-V$ and $C-V$ characteristics [18], [41], mobility reduction [22], [26], and threshold voltage (V_t) and sub-threshold slope (SS) degradation [27].

All in all, the reason why InGaAs MOSFETs are feasible is not because D_{IT} could be reduced to levels similar to that of Si/SiO_2 interface (i.e., $\sim 1\times10^{10}\,\mathrm{cm}^{-2}$) but rather thanks to the interface properties itself, that exhibits a charge neutrality level that is close to the E_C of InGaAs [27], thus rendering the traps in the band gap of the semiconductor donor-type and those *in* the conduction band itself acceptor-type[5]. Experimental data from [29], [30] showing trap states both in the energy range corresponding to InGaAs E_g and in its CB is collected in Fig. 2.3. According to the assumption of the charge neutrality level being located approximately at the conduction band minimum, E_C [27], the D_{IT} is modeled as a donor distribution in the band gap and an acceptor trap one in the conduction band, see Fig. 2.3. As mentioned previously, ITs and BTs give rise to a series of issues that reduce the performance gain of InGaAs MOSFETs over Si counterparts. As we will analyze in Section 2.3, one of the key signature of the presence of high

[4]Traps are electrically active defects that interact with the semiconductor by exchanging charge and that effectively limit the performance of MOSFETs.

[5]From basic semiconductor physics [28], when a trap energy state is located *below* the charge neutrality level it acts as a donor, i.e., it is positively charged when empty and neutral when occupied (by an electron). Conversely, a trap state *above* the neutrality level acts as an acceptor, i.e., it is neutral when empty and negatively charged when occupied (by an electron). Traps inducing a reduction in I_{ON} for n-type devices are the acceptors. This is why having a charge neutrality level far into the CB is important to avoid performance reductions.

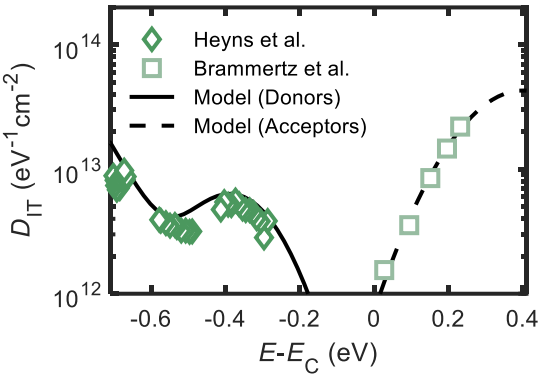


Figure 2.3: D_{IT} distribution between InGaAs and Al_2O_3. Trap states below (above) E_C are modeled as donor-type (acceptor-type) [27]. Data is taken from [29], [30].

trap density is the frequency dispersion in $C-V$ characteristics [17], [18] and also hysteresis, i.e., the difference between the upwards and downwards sweep when tracing either the $I-V$ or $C-V$ curves.

2.2.6 Variability in Nanoscale MOSFETs

With the scaling of MOSFETs dimension down to the nanometer scale, issues related to the intrinsic parameter fluctuations caused either by the discreteness of matter (e.g., number of dopants per unit volume), or by the imperfections in the definition of geometric sizes by lithography [31]. These issues are commonly referred to as statistical variability, which causes identically drawn devices to have different characteristics and performance. For instance, a given transistor technology (InGaAs NW MOSFETs in this case [33]), can have very different I_{ON}/I_{OFF} ratio as shown in Fig. 2.4. Clearly variability leads to circuit design issues that need to be tackled [32]. The reason why scaling has made variability a more and more pressing issue is because of its intrinsic nature that (for different reasons) makes it not scale with dimension reduction [31] (one cannot reduce the atom size!). For this reason, a number of simulation/experimental studies were carried out to predict/assess the relative role of variability on critical device parameter variations (such as V_t, SS, I_{ON}, and I_{OFF}) at each technology generation, in both Si and InGaAs devices [34]–[38]. Several statistical variability sources play a role [31], [38] in determining the total variation (expressed as standard deviation) from the mean value (obtained from an idealized

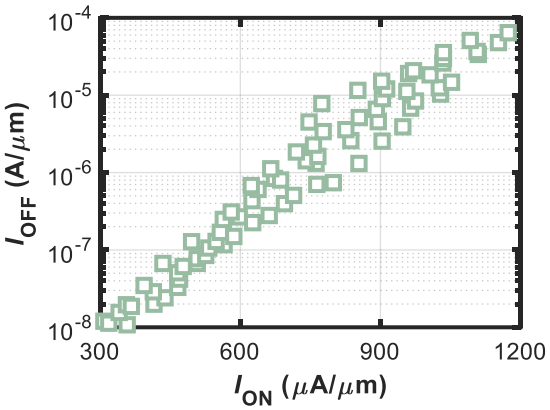


Figure 2.4: I_{OFF} vs I_{ON} plot for a given InGaAs technology showing variability, i.e., identically drawn devices have different current values and thus performance. Data from [33].

device with all nominal dimensions and no discreteness effects). These are

- Random Dopant Fluctuation (RDF) or Random Discrete Dopants (RDD) [31]
- Work Function Fluctuation (WFF) or Metal Gate Granularity (MGG) [36]
- Line Edge Roughness (LER) or Surface Roughness (SR) [31]

Besides these, other variability sources can also play a role. For instance, in Section 2.4 we will analyze the effect of the variation in molar composition of InGaAs, giving rise to Band Gap Fluctuation (BGF) [8, Ch. 1], and the variability introduced by the variation in the concentration of traps [39], i.e., Interface Traps Fluctuation (ITF).

2.3 Contribution: Mobility Measurements Inaccuracies due to ITs and BTs in planar InGaAs MOSFETs

As we discussed previously, mobility measurements in InGaAs MOSFETs are highly affected by ITs and BTs. This is because the trapped charge contributes to the gate capacitance C_G, leading to an underestimation of the carriers' mobility. To understand why this is the case, we need first to discuss how mobility is extracted through the split $C-V$ measurements. The drain current (I_D) of a MOSFET in the linear region can be expressed as (ignoring the contribution due to diffusion) [4]

$$I_D = W_G\,(Q_n \times v_n) = W_G\left(Q_n \times \frac{\mu_n V_{DS}}{L_G}\right) \tag{2.4}$$

from which mobility can be extracted as

$$\mu_n = \frac{L_G}{W_G}\frac{I_D}{Q_n V_{DS}}. \tag{2.5}$$

The transistor dimensions W_G, L_G are known by design, whereas V_{DS} is chosen to bias the transistor in the linear region (i.e., V_{DS} is equal to a few times V_{th}). The electron mobile charge Q_n needs to be determined indirectly by measuring the gate to channel capacitance, that is

$$Q_n = \int_{-\infty}^{V_{GS}} C_G(V)dV. \tag{2.6}$$

C_G is determined by $C-V$ measurements thus the name of the technique "split CV" [40]. Because the trapped charge adds up to C_G the actual Q_n results to be larger than that given actually by mobile charges, thus measured mobility through this method is lower than the intrinsic one which can be assessed by Hall measurements. However, Hall mobility requires dedicated samples in order to perform the measurement, thus it is useful to explore alternative methods to know the actual mobility. One such technique consists in the use of calibrated TCAD simulations that not only allow to extract the intrinsic mobility but also to quantify the contribution due to traps.

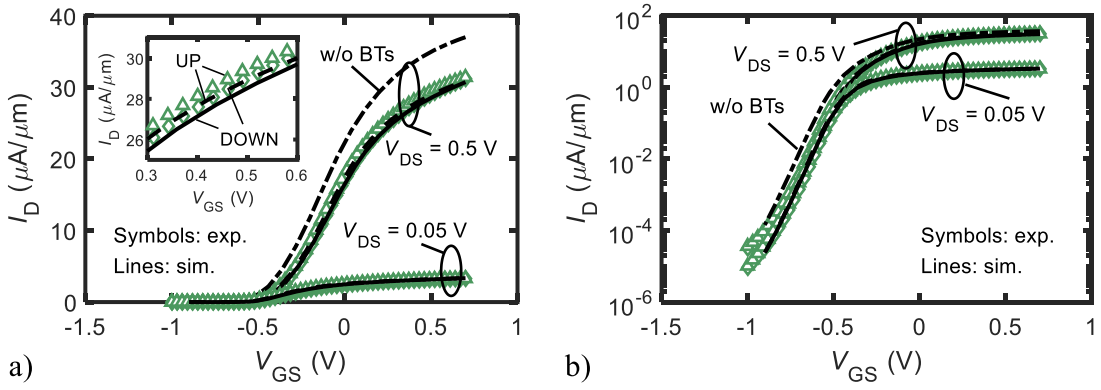


Figure 2.5: Comparison between experimental (symbols) and simulated (lines) $I-V$ characteristics of the InGaAs planar MOSFET in (a) linear and (b) log scale at $V_{DS} = \{0.05, 0.5\}$ V. The inset in (a) shows the hysteresis between the up- and down- sweep caused by BTs. Adapted from [41].

2.3.1 Calibration of *I–V* and *C–V* characteristics

The device under investigation is a planar InGaAs MOSFET with the cross-section depicted in Fig. 2.2, that mimics the device realized in [18]. L_G is 10 µm. The high-κ gate dielectric is Al_2O_3 with thickness (t_{ox}) of 10 nm (EOT of 4.9 nm). The channel thickness (t_{ch}) is 15 nm, thus electron density in the channel needs to be computed with an appropriate quantum-corrected model. In [41] the "density-gradient" model was adopted to calculate Λ_n in Eq. (2.2) [14].

The $I-V$ characteristics of the device in [18] are shown for two different V_{DS} in Fig. 2.5 in both the (a) linear and (b) log scale. Along with the experimental data (symbols) are shown the results of the calibrated simulations (lines). The trap distribution used in the simulation to fit the experimental data is the one shown in Fig. 2.3 (lines). Coherently with the discussion in Section 2.2.5, ITs are considered to be donor-like traps in the InGaAs bandgap whereas BTs are considered to be acceptor-like traps above E_C. The BTs are spread throughout the gate oxide in a 5 nm-layer that is 1 nm away from the interface. Non-local tunneling from the Al_2O_3/InGaAs interface into BTs allows them to capture electrons in the channel as the device is driven into on-state conditions. To account for this effect in the simulations under both dynamic voltage sweeps and $C-V$ simulations the tunneling-into-traps model [14] was activated. A first glance of the effects of BTs can be appreciated by the dashed line representing the simulation results taken

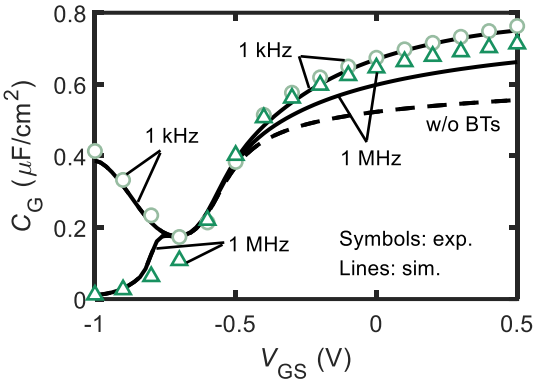


Figure 2.6: Comparison between experimental (symbols) and simulated (lines) $C-V$ characteristics of the InGaAs planar MOSFET. Both experiments and simulations show frequency dispersion at 1 kHz and 1 MHz. The dispersion at low V_{GS} can be attributed to ITs, whereas the one at high V_{GS} (i.e., above threshold) is due to BTs only, as demonstrated by the simulated curves w/o BTs (dashed lines). Adapted from [41].

without the BTs. Clearly in this case I_D above threshold is much larger than the experimental one and the SS is closer to the ideal lower limit of 60 mV/dec. Moreover, the hysteresis (i.e., the difference between the UP and DOWN sweep highlighted in the inset of Fig. 2.5(a)) is absent in the simulation w/o BTs indicating that indeed traps are the cause for it. We conclude then that BTs are at the origin of $I-V$ hysteresis (as also found in previous reports [42]) and that the latter can be used as an effective signature for BTs [41].

The comparison between measured and simulated $C-V$ curves is shown in Fig. 2.6 at two different frequencies, namely 1 kHz and 1 MHz. As it can be appreciated from Fig. 2.6, simulations yield $C-V$ curves in reasonably good agreement with experiments without further adjustment of input parameters with respect to the ones adopted for $I-V$ simulations. Frequency dispersion between 1 kHz and 1 MHz can also be captured with the simulations with the distribution shown in Fig. 2.3. As pointed out in Section 2.2.5, frequency dispersion in the $C-V$ curves can be attributed to the presence of traps. Particularly, in the subthreshold region ITs are generally held responsible for dispersion. On the contrary, in the above-threshold region (or accumulation region) BTs are to be blamed. The simulation result obtained without BTs included confirms this, see the dashed line in Fig. 2.6, as in this case no dispersion is found in the accumulation region whereas it is still present in the subthreshold region (which is due to ITs). It is worth pointing out

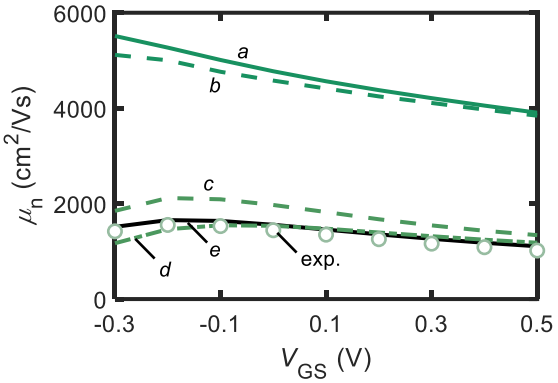


Figure 2.7: Channel electron mobility μ_n vs V_{GS}. Experimental data curve (symbols) is compared against simulation results (lines). Curve *a* is the μ_n given as input to the simulations, curve *b* is the calculated mobility with Eq. (2.5) including the free electron charge only, curves *c*, *d* include also charge due to BTs and BTs+ITs, respectively. Curve *e* is calculated by using Eqs. (2.5) and (2.6) (i.e., the same methodology used to extract experimental data). Adapted from [41].

that trapped charge by BTs adds up to the accumulation capacitance somewhat compensating for the quantization effects occurring in the channel, as far as their impact on gate capacitance is concerned. Quantization effects in the narrow channel in fact tend to decrease C_G (due to the shift in charge centroid, as discussed in Section 2.2.2) while trapped charge tend to increase it. This aspect is insidious for technologists since it may lead to $C-V$ data misinterpretation. Interestingly, the TCAD simulations discussed here give the same qualitative results as given by other models that consider BTs as a distributed RC network [17], [43].

2.3.2 Effects of ITs and BTs on Mobility Extraction

Having calibrated the simulations on experimental results, we are now ready to assess the effects of ITs and BTs on the mobility. Coherently with the experimental procedure, we extract mobility from the simulations by using Eqs. (2.5) and (2.6) obtaining the results shown in Fig. 2.7. The experimental mobility versus V_{GS} data (symbols) is compared to several different simulated mobility curves (labelled *a*-*e*) to highlight key aspects as detailed in the following.

Curve *e* is the one computed with Eqs. (2.5) and (2.6) and, not surprisingly, is in good agreement with measured mobility demonstrating the effectiveness of the calibration procedure shown in Figs. 2.5 and 2.6. Curve *a* represents the mobility that was used as an input to simulations

including the contribution due to surface scattering and acoustic phonon scattering summed up to the bulk mobility (used as fitting parameters and set to $6500\,\mathrm{cm}^2/\mathrm{Vs}$) via the Matthiessen rule, following the Lombardi model [14]. Impurities scattering due to doping in the channel was neglected because the channel is assumed to be unintentionally doped. The large discrepancy between curves *e* and *a* indicates that mobility extracted with the split $C-V$ method is affected by large errors. A $3\times$ larger mobility was actually assumed in simulations compared to extracted values (compare curve *a* and symbols in Fig. 2.7). Similar results were obtained in [27], finding the same issues with split $C-V$ method for a different InGaAs MOSFET technology.

Curve *b* in Fig. 2.7 represents mobility values obtained by using Eq. (2.5) but with Q_n not obtained with Eq. (2.6) but by integrating the channel electron density only, obtained as an output of the simulations. As can be noted, in this case mobility almost completely coincides with curve *a*, confirming that the mobility extraction inaccuracies are mainly related to how Q_n is obtained. Finally, curves *c* and *d* are the mobilities computed by adding to the channel electron charge the trapped charge by BTs only (curve *c*) and that by both BTs and ITs (curve *d*), respectively. Mobility of curve *c* drops significantly compared to curves *a* and *b*, indicating that most of the contribution to mobility measurement inaccuracy comes from the trapped charge by BTs, as a consequence of the spurious increase in C_G. Curve *d* including also trapped charge by ITs drops further closely approaching measured data, indicating that also IT contribute (to a lesser extent than BTs) to mobility extraction inaccuracy.

The validity of the approach in determining the contribution due to both ITs and BTs was further confirmed by the agreement between simulations and experimental data for the same MOSFETs but with narrower channels, i.e., $t_{ch} = 10, 5\,\mathrm{nm}$ (not shown) [41]. It is worth pointing out that in order to extract the intrinsic mobility for the narrowest channel device (5 nm) not only the correction of Q_n to purify from traps contribution is required, but also an additional one to exclude the electron charge present in the InP confinement layer at the bottom of the channel. In fact, as indicated by $\mathbf{k}\cdot\mathbf{p}$ calculations in [18], thinning of the channel layer causes the effective mass to increase — which can also be interpreted as deconfinement of electrons from the InGaAs channel to the bottom InP layer — thus further lowering the mobility.

2.3.3 Conclusions

We presented the results of a combined experimental/simulation analysis of InGaAs planar MOSFETs addressing the effects of interface and border traps on the electrical characteristic and on the accuracy of split-CV mobility measurements in these devices. We showed how both ITs and BTs contribute to determine hysteresis in the double-sweep $I-V$ characteristics and to frequency dispersion in $C-V$ characteristics. We then pointed out that mobility measurement by the split-CV method lead to drastic underestimations of actual InGaAs channel mobility, as a result of the appreciable spurious contributions of ITs and BTs to the total gate charge. Moreover, in the case of very narrow channel devices, electrons deconfinement into the underlying InP buffer layer also contributes to mobility measurement inaccuracies.

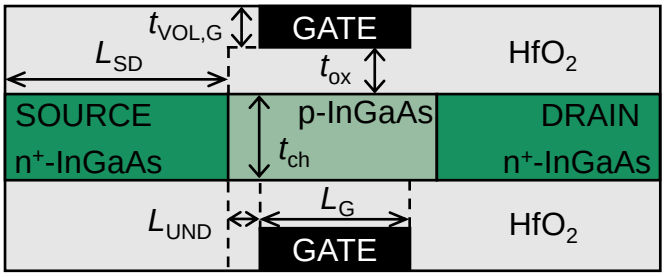


Figure 2.8: Sketch of the cross-section of the simulated InGaAs DG-UTB MOSFET.

2.4 Contribution: Statistical Variability in Ultra-scaled Double-Gate InGaAs MOSFETs

In this section, we analyze the impact of both local and global variations (i.e., variability and sensitivity, respectively) on the V_t of InGaAs MOSFETs. Two ultra-scaled technological nodes are analyzed, namely ($L_G = 15\,\text{nm}$, $V_{DD} = 0.63\,\text{V}$) and ($L_G = 10.4\,\text{nm}$, $V_{DD} = 0.59\,\text{V}$). This section is divided in two parts. In the first one, we present results on WFF, LER, RDF, and BGF to assess their relative impact on the total V_t variability, quantified as the standard deviation of the nominal V_t, i.e., $\sigma(V_t)$. A comparison between InGaAs devices and their Si equivalent at the node ($L_G = 15\,\text{nm}$, $V_{DD} = 0.63\,\text{V}$) is also included for benchmarking purposes. In the second part, we present results on InGaAs devices with the inclusion of ITs to evaluate their impact on variability. To this extent, we present a TCAD simulation methodology that allows correctly reproducing MSMC results on the same device when including ITs. We end the section by presenting a comparison between total variability for latest Intel Si technology nodes (available in the open literature [34]) and variability obtained from our simulations on InGaAs DG-UTB devices.

2.4.1 Assessing Variability in TCAD with the Statistical Impedance Field Method

Before discussing the results related to variability, we first give a brief description of the Statistical Impedance Field Method (S-IFM) [14], [44]. S-IFM is an efficient and reasonably accurate tool useful for assessing the impact of several variability sources independently of each other (i.e., no interplay between random sources is considered). The S-IFM models variability as small perturbations of the response of a 'nominal' device. This 'nominal' response is obtained from

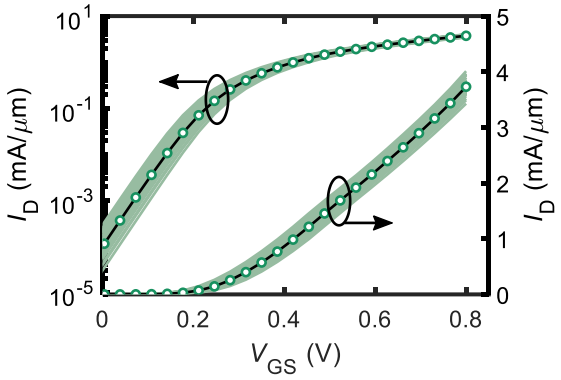


Figure 2.9: Distribution of $I-V$ characteristics obtained with S-IFM by including WFF as variability source in the 15-nm node InGaAs DG-UTB device. The black line with symbols is the reference curve computed from the self-consistent QDD simulation. The light green curves are computed as small linear perturbations of the reference curve. Adapted from [48].

the full, self-consistent QDD simulation of the device under investigation. Variability is then obtained by computing the altered response (in terms of contact current) for a large number M of randomized device realizations (in our case, $M = 10\,000$) by performing a linearization of the device perturbations via a Green's function-based approach [14]. From the computed contact current fluctuations a set of varied $I-V$ can be constructed, as shown for example Fig. 2.9 for WFF-induced variability, from which the V_t distribution — and, consequently, $\sigma(V_t)$ — is extracted. Obtaining variability as small, linearized perturbations of the nominal device characteristics represents an approximation (that assumes the operating point is not largely affected by variability) compared to the more accurate atomistic simulations. Nevertheless, S-IFM it is a sufficiently accurate method that has the clear advantage (over sophisticated simulations) of efficiency and speed, as the self-consistent device simulation needs to be solved only once.

2.4.2 Variability Sources Modeling

All the main microscopic variability sources of relevance for ultrascaled MOSFETs [45] have been taken into account, namely: RDF, gate-metal WFF, body- surface or fin-sidewall roughness (Body Line Edge Roughness (BLER)), and Gate Line Edge Roughness (GLER). For InGaAs, also BGF and ITF have been considered. In the following, we briefly illustrate the modelling approach for each variability source:

- RDF, caused by inherent discreteness of impurity atoms, is modeled assuming spatially uncorrelated doping both in the channel and source/drain regions. The number of impurity atoms due to doping in a given volume follows the Poisson distribution with an average value equal to the nominal number of dopants in that volume [14].

- WFF is caused by intrinsic granularity of the gate metal and is modeled by assuming TiN as the gate metal, which is characterized by two major grain orientations with a work function difference of 0.2 eV and a relative probability of 0.4/0.6 for the lowest/highest WF, respectively [36], see Section 2.4.2. The number of metal grains in the gate region — each with different shape and size — follows a Poisson distribution with an average value obtained by using the average grain size [14]. Here, we consider an average grain size of 5 nm for both grain orientations as representative of the typical value achievable in present gate-first processes with TiN as the gate metal [48].

- LER, inherent to the litography patterning steps of CMOS technology, leads to local L_G variations (labelled as GLER) and to body (fin) thickness variations in DG-UTB (FinFET) devices, (labelled as BLER). These effects are modeled by means of a Gaussian autocorrelation function [31], with rms. roughness amplitude (Δ_{rms}) and correlation length (Λ_{LER}) set in accordance with ITRS projections for the nodes under consideration [46], see Section 2.4.2.

- BGF stems from the random variations of indium content (i.e., mole fraction) in $In_{0.53}Ga_{0.47}As$ which cause the channel volume to be subject to random bandgap variations. BGF modeling considers the shifts of both VB and CB edges when calculating the total δE_g according to the simple model shown in Section 2.4.2, that relates it to electron affinity variations $\delta\chi$ [2]. The value of 10 meV of $\delta\chi$ is chosen in accordance to [47] by considering the expected experimental mole fraction variations range. The inhomogeneity due to variations of In content is modeled via a granular approach similar to that adopted for WFF. Specifically, the channel volume is divided into "grains" of different types and sizes each associated with a different E_g with size randomized around an average value [52].

- ITF reflects the random number of traps present at the interface between the gate oxide

Table 2.1: Variability Sources Parameters.

WFF	**BLER, GLER 15-nm (10.4-nm)**	**BGF**
Avg. Grain Size = 5 nm	$\Delta_{rms} = 1.8\,(1.2)\,\text{nm}$	$\delta\chi = 10\,\text{meV}$
		$\Lambda_{BGF} = 300\,\text{nm}$
$P_{WF1} = 60\,\%$	$\Lambda_{LER} = 15.5\,(8.3)\,\text{nm}$	$\alpha = -1.3$
$P_{WF2} = 40\,\%$		$\delta E_g = \alpha\delta\chi$

Table 2.2: DG-UTB Design Parameters.

Parameter	**15-nm Node**	**10.4-nm Node**
$L_G\,(\text{nm})$	15	10.4
$W_G\,(\text{nm})$	26	14
$t_{ch}\,(\text{nm})$	7	4
$t_{ox}\,(\text{nm})$	3.84	3.3
$L_{SD}\,(\text{nm})$	24.8	22
$t_{VOL,G}\,(\text{nm})$	3	
$L_{UND}\,(\text{nm})$	2	
$N_{ch}\,(\text{cm}^{-2})$	1×10^{17}	
$N_{SD}\,(\text{cm}^{-2})$	5×10^{19}	

and the channel, thus it is modeled alike RDF [55]. That is, along the interface the random number of traps is calculated from a Poisson distribution with an average equal to the nominal number of traps determined from the D_{IT}.

2.4.3 Threshold Voltage Variability and its Sensitivity to Critical Design Parameters in InGaAs DG-UTB MOSFETs

The device architecture considered for the variability analysis is the DG-UTB, sketched in Fig. 2.8. The geometrical dimensions as well as doping densities are summarized in Section 2.4.3 for both technological nodes, which we will label as "15-nm" and "10.4-nm" for brevity. Although the structure has a 2D geometry, the self-consistent QDD simulations is performed on a 3D mesh

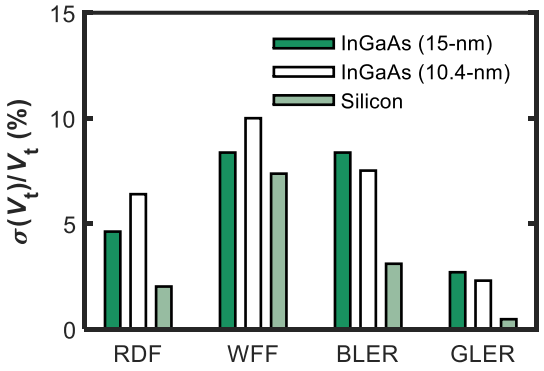


Figure 2.10: Comparison of $\sigma(V_t)$ due to different variability sources (namely, RDF, WFF, BLER, GLER) for InGaAs 15-nm and 10.4-nm nodes and Si counterparts. Adapted from [48].

obtained by extruding the device into the third dimension by an amount equal to the gate width, W_G. This was done to accurately evaluate variability as the correct number of random dopants, for instance, needs to be evaluated over the entire gate volume. Prior to performing the variability simulations on the InGaAs devices, the QDD deck was calibrated against more sophisticated simulations results obtained either from MSMC or NEGF results [48]. The numerical device simulations of these structures are carried out using the QDD transport model with quantum effects accounted for by means of the Modified Local Density Approximation (MLDA) model [14]. Calibration of the Si devices (for benchmarking) is performed by comparing simulations with the experimental data available in the open literature for a Si FinFET with $L_G = 20\,\text{nm}$ [34]. The device architecture is then adapted to that of a DG-UTB to perform a fair variability comparison [48].

Comparison of Variability in InGaAs and Si DG-UTBs

Bar plots in Fig. 2.10 provide a quantitative assessment of variability for both InGaAs and Si devices, showing extracted $\sigma(V_t)$ value for each variability source. Note that, for the sake of comparison, $\sigma(V_t)$ values are normalized to the corresponding nominal V_t. The variability sources are considered separately from each other to quantify and compare their individual contributions. From Fig. 2.10 it is clear that InGaAs devices show higher $\sigma(V_t)$ due to RDF, BLER, and GLER than the Si counterpart. On the other hand, WFF has comparable effects on InGaAs and Si

devices due to the same device geometries. Differences in BLER and GLER effects for the two technologies can be understood by considering that InGaAs has a smaller effective mass than Si, resulting in: *i)* larger quantum confinement effects, making V_t more prone to t_{ch} fluctuations (i.e., larger BLER-induced $\sigma(V_t)$) [49], and *ii)* reduced inversion capacitance, that makes V_t more sensitive to L_G modulation and associated short-channel effects (i.e., larger GLER-induced $\sigma(V_t)$). It is found that RDF-induced $\sigma(V_t)$ is affected by the lower InGaAs DOS, leading to a non-negligible dependence of V_t (and $\sigma(V_t)$) on N_{SD} (as further discussed later on). Another important point about $\sigma(V_t)$ data shown in Fig. 2.10 is that WFF and BLER have comparable and predominant variability effects on V_t of InGaAs devices, whereas WFF is the only predominant source of variability in Si.

Bar plots in Fig. 2.10 also provide a quantitative comparison of InGaAs $\sigma(V_t)$ at the 15- and 10.4-nm nodes. Notably, the relative V_t variability is not dramatically affected by L_G scaling. More specifically: *i)* RDF- and WFF-induced $\sigma(V_t)$ slightly increase at the 10.4-nm node mostly due to the small nominal V_t reduction and *ii)* BLER- and GLER-induced $\sigma(V_t)$ actually decreases with node scaling due to the reduction in both Δ_{rms} and Λ_{LER} parameters, see Section 2.4.2. In this scenario, the detrimental impact of BLER on V_t variability evident at the 15-nm node is attenuated, leaving to WFF the role of dominant variability source (akin to Si devices). However, this encouraging scaling perspective holds true only if the enhanced LER control forecast by ITRS is effectively attained [46].

Combined Variability and Sensitivity Analysis

The overall picture regarding $\sigma(V_t)$ needs to be confirmed when considering also the effect of systematic design parameter variations. To this end, we evaluate how $\sigma(V_t)$ and the nominal V_t (indicated also by the average value of the distribution, $\mu(V_t)$) are affected by the systematic variations. In addition, we evaluate the effect of critical parameters affecting WFF and BGF modeling, namely the average metal grain size for the former and the relative In mole fraction (x) variation for the latter. Fig. 2.11 shows the variation of WFF-, BLER-, RDF-, and BGF-induced $\sigma(V_t)$ when considering metal grain size, t_{ch}, N_{SD}, and x variations. Fig. 2.11(a) shows the effect of different metal grain size (in the range of 2–20 nm, [48] and references therein) on the

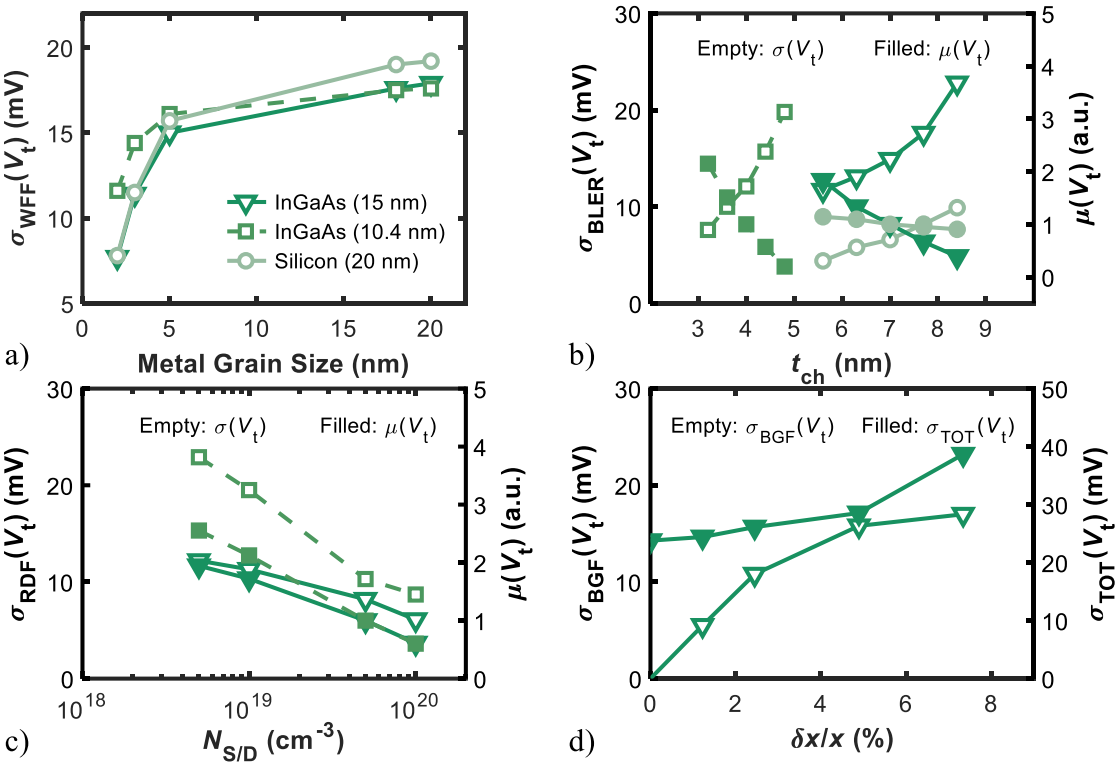


Figure 2.11: V_t variability ($\sigma(V_t)$) due to different sources for InGaAs and Si DG-UTB devices (see legend). (a) WFF vs metal grain size; (b) BLER vs t_{ch} and corresponding nominal V_t (right axis); (c) RDF vs N_{SD} for InGaAs devices only; (d) BGF vs relative Indium mole fraction variation $\delta x/x$ for InGaAs 15-nm node only. (a), (b) adapted from [48]. (c) adapted from [50]. (d) adapted from [52].

WFF-induced $\sigma(V_t)$. Results indicate that for grain size >5 nm (i.e., the value taken as reference, [48]) $\sigma(V_t)$ increases slightly. On the other hand, grain size <5 nm causes WFF variability to drop considerably. Nevertheless, this reduction can only be achieved in real-case scenarios with processes preventing the crystallization of the metal, i.e., gate-last flow, or with different materials such as TaSiN that remain amorphous even in gate-first flows, [48] and references therein.

Fig. 2.11(b) shows a larger BLER-induced $\sigma(V_t)$ and a larger V_t due to t_{ch} scaling. In addition, scaling to the 10.4-nm node renders sensitivity to t_{ch} even more critical for InGaAs devices. In [48] sensitivity to t_{ch} was found to be the predominant variation effect (with respect to other geometrical parameter variations such as t_{ox} or L_G) in both the InGaAs and the Si devices. The observed larger $\mu(V_t)$ sensitivity to t_{ch} variations in InGaAs devices at the 15-nm node compared

to Si ones raises concerns about the scaling potential of InGaAs devices [48]. This picture is further worsened by the fact that sensitivity to t_{ch} variations of BLER-induced $\sigma(V_t)$ increases with node scaling for InGaAs, as shown by Fig. 2.11(b). Eventually, the combination of the increased sensitivity to t_{ch} for both $\mu(V_t)$ and $\sigma(V_t)$ can be a showstopper for this technology, highlighting the critical role of channel thickness control [48].

Fig. 2.11(c) highlights the high influence of N_{SD} systematic variations on RDF-induced $\sigma(V_t)$ and $\mu(V_t)$ [50], which is further increased at the scaled node (10.4-nm), at least for $N_{SD} \leq 5 \times 10^{19}\,\mathrm{cm}^{-3}$. In fact, results in Fig. 2.11(c) show that for $N_{SD} \geq 5 \times 10^{19}\,\mathrm{cm}^{-3}$ scaling has small effects on both $\mu(V_t)$ and $\sigma(V_t)$ sensitivity to N_{SD}. However, the maximum achievable N_{SD} in InGaAs MOSFETs is limited to $5 \times 10^{19}\,\mathrm{cm}^{-3}$, as confirmed by both experimental evidences and theoretical calculations that blame the dopant vacancy-complexing mechanism in heavily Si-doped InGaAs for this [51]. Thus, results obtained by employing realistic N_{SD} values leads to increased $\mu(V_t)$, see Fig. 2.11(c). In addition, RDF-induced $\sigma(V_t)$ is also increased when N_{SD} decreases. Particularly, if N_{SD} is as low as $1 \times 10^{19}\,\mathrm{cm}^{-3}$ RDF-induced $\sigma(V_t)$ can reach 20 mV at the 10.4-nm node, that is comparable or higher than that due to the other variability sources [48], potentially making RDF the main variability source.

Finally, Fig. 2.11(d) (empty symbols) shows the variation of BGF-induced $\sigma(V_t)$ vs varying In mole fraction with respect to the nominal value $x = 0.53$ (i.e., the relative variation $\delta x/x$). Mole fraction variations have been computed by correlating them to E_g variations obtained from the model of Section 2.4.2 and by linearizing the E_g vs x relationship near the nominal value $x = 0.53$ [52]. Clearly, larger $\delta x/x$ lead to increased $\sigma(V_t)$ because of the increased inhomogeneity of the channel and of its electrical behavior. Interestingly, for $\delta\chi = 10\,\mathrm{meV}$ (corresponding to $\delta x/x \approx 4\%$), BGF-induced $\sigma(V_t)$ is less than WFF- and BLER-induced $\sigma(V_t)$, i.e., the most impacting variability sources in InGaAs [48]. Although this result gives the positive indication that BGF should not be the first concern regarding V_t variability, it is instructive to analyze more in detail the impact of BGF-induced $\sigma(V_t)$. This is done by evaluating the relative BGF-induced $\sigma(V_t)$ (normalized to the nominal V_t) at different x variations and then by computing the $\sigma_{TOT}(V_t)$ as the square sum of the individual contributions from all sources (i.e., assumed to be independent), see Fig. 2.11(d) (filled symbols). From this plot we find that the projected

worst-case relative $\sigma_{TOT}(V_t)$ is $\sim 23\%$, corresponding to the worst-case $\delta x/x$ of about 9% as reported in [47]. Conversely, the the projected relative $\sigma_{TOT}(V_t)$ is $\sim 15\%$ for the average x variations as also reported in [47].

2.4.4 Impact of Traps on Variability

In order to evaluate the effects of random ITs concentration (i.e., ITF) on V_t variability we establish a systematic modeling approach for TCAD that allows achieving calibration with more sophisticated MSMC simulations even at the 10.4-nm node. At this node in fact, since $t_{ch} = 4\,\text{nm}$ (see Section 2.4.2), the electron wave-function has a finite penetration length into the gate oxide that cannot however be accurately captured with TCAD simulations. Obtaining the correct electron density distribution in the channel is necessary for accurately assessing the influence of ITs on the device electrostatics (and transport). Thus, in [55] a systematic modeling approach was developed that provides a consistent agreement between QDD and MSMC simulations not only in terms of the $I-V$ curve but also in terms of inversion and trapped- charge density in the channel, as well as the mobility.

MSMC and QDD Simulation Approach

The MSMC set up employs a NP-EMA model for determining both quantization and transport. Interface traps have been introduced in the solution of the coupled Schrödinger and Poisson equations as a sheet of charge at the interface between the channel and the gate dielectric. Therefore, in this case we make no distinction between ITs and BTs but instead consider all traps to be lumped at the semiconductor and dielectric interface. MSMC simulations account for scattering mechanisms, such as elastic intravalley and inelastic intervalley phonons, remote phonons from the high-κ dielectric, local polar phonons, Coulomb, alloy, and surface roughness scattering. The modeling of the device electrostatics with the inclusion of ITs was carried out by: *i)* employing a D_{IT} matching experimentally measured data, shown in Fig. 2.3, and *ii)* preserving the device geometry, particularly by not increasing the channel thickness as done in previous works [53]. The latter is a crucial point, since t_{ch} strongly influences the variability of the device, especially at ultrashort L_G [48]. In other words, to avoid compromising the dependability of the

variability analysis, unaltering t_{ch} is mandatory. The reference MSMC simulation setup in this work is taken from [26].

To reproduce the MSMC results with the TCAD simulator, we employed a QDD model that considers both the increased confinement of carriers in the channel and the quasi-ballistic transport at very short channel length. The quantum-correction is implemented by the MLDA model that requires no calibration parameter [14]. Since the QDD, differently from the MSMC, does not consider the electron Wave-Function Penetration (WFP) in the gate oxide (stemming from the strong geometric confinement due to thin t_{ch}), the WFP was turned off in the MSMC as well [25]. Thus, to maintain the same results as the MSMC with (w/) WFP, in the MSMC without (w/o) WFP and QDD setups we employed a modified trap distribution model from the one shown in Fig. 2.3. Such a distribution was obtained by translating the original model toward higher energies to compensate for the difference in the first available electron energy level that arises when not considering the WFP[6]. The energy shift allowed recovering the same relationship between the trapped charge density in the channel (N_{trap}) and the inversion charge density (N_{inv}) as compared to that obtained with the MSMC simulations including the original trap distribution and WFP. Thus, this approach allows preserving the electrostatic properties of the original device even without directly accounting for the WFP. In addition, the surface roughness parameters in the simulation w/o WFP were modified with respect to the reference setup [26] to compensate for the different wave function shapes that influence the scattering rates and the electron mobility. After the re-calibration, similar $\mu_n - N_{inv}$ profile to that of the original device w/ WFP was achieved, thus preserving the transport properties as well [55].

To calibrate the QDD over MSMC simulations, we employed the same trap distribution as the one used in the MSMC setup and calibrated the mobility versus carrier density curve in TCAD. The model employed to reproduce the MSMC mobility curve is the University of Bologna Model [54] available in SDevice [14]. The calibration of the mobility model of the QDD versus MSMC was performed on a long-channel device ($L_G = 100\,\text{nm}$) biased at low V_{DS} (i.e., 25 mV) to avoid the influence of short-channel effects, quasi-ballistic transport and velocity-saturation effects.

[6]When the carriers' wave function penetrates into the gate oxide, it is as if the quantization is reduced (because carriers are deconfined into the insulator). Thus, in the setup w/o WFP the quantization is stronger than in the case w/ WFP and this is why an energy compensation is required.

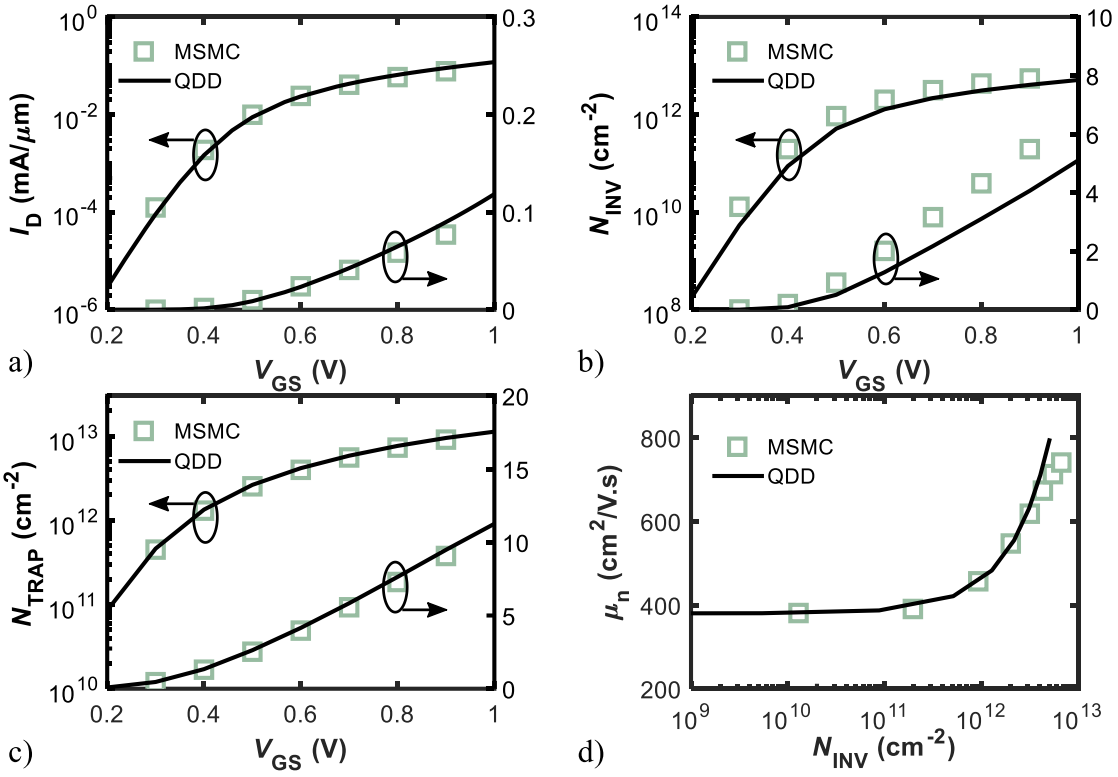


Figure 2.12: Calibration of the QDD (lines) over MSMC (symbols) simulations in terms of (a) $I_D - V_{GS}$, (b) $N_{inv} - V_{GS}$, (c) $N_{trap} - V_{GS}$, and (d) $\mu_n - N_{inv}$. The simulated device in this case has a long-channel ($L_G = 100\,$nm) and low bias ($V_{DS} = 25\,$mV) to avoid quasi-ballistic transport and velocity saturation, respectively. Adapted from [55].

Besides the calibration of the mobility model, no further parameter adjustment was required to obtain the agreement between MSMC and QDD, shown in Fig. 2.12. The calibrated QDD setup was then used to simulate the ultrascaled DG-UTB 10.4-nm InGaAs device and to compare with the MSMC results obtained with the same device geometry. Results in terms of $I - V$ curves (not shown) for two different V_{DS} biases (for the linear and saturation regimes, respectively) show an excellent agreement [55]. The agreement was achieved by using the same models (with the same parameter values) used for the calibration of the long channel device and with the addition of an empirical ballistic mobility model and the Canali model accounting for velocity saturation effects, both available in SDevice [14].

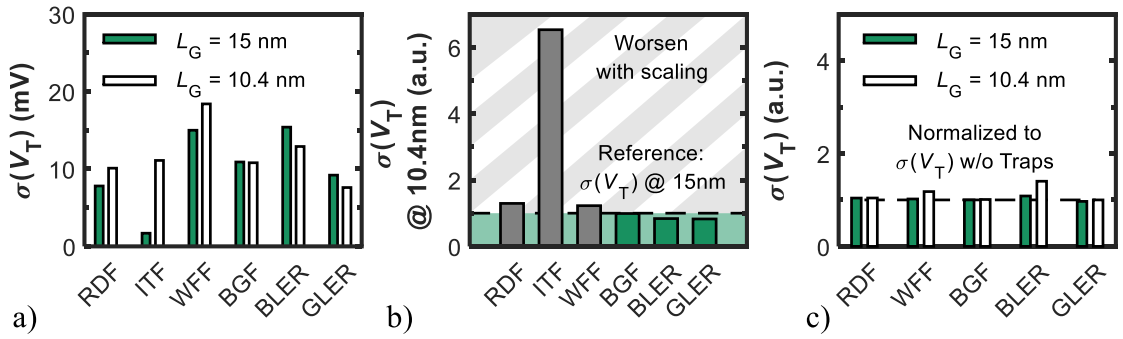


Figure 2.13: $\sigma(V_t)$ induced by the six variability sources considered for InGaAs devices (see horizontal axis labels). (a) Comparison between the $\sigma(V_t)$ calculated from devices at the 15-nm (dark green bins) and 10.4-nm node (white bins). (b) $\sigma(V_t)$ results for the 10.4-nm node normalized to the 15-nm one. Grey (green) bins represent the variability induced by the respective sources that worsen (improve) with scaling [i.e., increase (reduce) the induced $\sigma(V_t)$]. (c) $\sigma(V_t)$ for both 15- and 10.4-nm nodes normalized to the one obtained for devices at the same node but without interface traps. Adapted from [55].

Variability Analysis

The calibrated 10.4-nm device QDD simulation was then used as a reference from whch obtaining the varied curves with the S-IFM method as explained previously. Again, variability is quantified as the V_t standard deviation, $\sigma(V_t)$.

The results of the variability analysis are shown in Fig. 2.13. Fig. 2.13(a) shows a comparison between the $\sigma(V_t)$ obtained for the InGaAs devices at the 15 and 10.4-nm nodes. Variability results for the 15-nm device were calculated from QDD simulations calibrated on MSMC with an effective D_{IT} obtained from the model of Fig. 2.3. The effective distribution was obtained following the same procedure as that described previously for the 10.4-nm node, to preserve the N_{trap} vs N_{inv} relationship (not shown). From Fig. 2.13(a) it can be seen that, upon scaling, both RDF- and WFF-induced $\sigma(V_t)$ increase, BLER and GLER decrease (due to reduction of the associated amplitude, Δ_{rms}, following ITRS, see [46] and Section 2.4.2) and BGF-induced $\sigma(V_t)$ shows a negligible variation. Notably, ITF-induced $\sigma(V_t)$ has a $\sim 6\times$ increase, that negatively impacts the total $\sigma(V_t)$ variations. This increase is a consequence of the higher trapped charge density for the scaled device [55]. In fact, since the ITs are Poisson-distributed, if their mean value increases (proportional to N_{trap}), then also the standard deviation does, thus explaining the

ITF increment shown in Fig. 2.13(a) [39].

The ITF trend is more clearly shown in Fig. 2.13(b),where the $\sigma(V_t)$ at the 10.4-nm node for each source is normalized to that of the respective source at the 15-nm node. These results indicate that scaling strongly worsens ITF. Similar trends of increased trap-induced variability due to higher trapped charges were found also for Si multi-gate MOSFETs [39]. However, conversely to InGaAs, in Si devices, the increased N_{trap} was attributed to the effects of aging rather than to increased quantum confinement (which is stronger in InGaAs due to its low effective mass).

The effect of interface traps on the $\sigma(V_t)$ induced by all other variability sources (i.e., other than ITF) can be appreciated with the aid of Fig. 2.13(c). It shows the $\sigma(V_t)$ for both the 15- and 10.4-nm nodes normalized to the $\sigma(V_t)$ obtained for the devices at the same nodes without including ITs (The reference devices without interface traps were calibrated to MSMC simulations as well, as reported in [52].). Remarkably, for both nodes, the other variability sources appear not to be significantly affected by the presence of interface traps, except for BLER. This result indicates that the device with ITs is more sensitive to BLER, especially at the 10.4-nm node. This is clear from the fact that electrostatic integrity is reduced by the presence of traps, and, since BLER acts as a random variation of the t_{ch}, V_t fluctuations consequently increase.

Scaling of Total Threshold Voltage Variability

The total $\sigma(V_t)$ is calculated as the quadrature sum of the contribution of all the different sources, assuming each source to be independent of each other for simplicity. Results for Si-based MOSFETs (taken from available Intel technology data [34]) and both Si and InGaAs devices (from our simulations [48], [52], [55]) are shown in Fig. 2.14. The comparison shows that V_t variability in InGaAs devices is higher than that regarding Si (at the 14-nm node) [34], and that is further increased when including ITs. Moreover, the scaling of the InGaAs technology further increases the total $\sigma(V_t)$, thereby contrasting the scaling trend of improved performance with reduced dimensions observed in Si technology. The fact that $\sigma(V_t)$ obtained from QDD simulations match experimental data at the same node prove the dependability of the variability analysis based on the S-IFM. Due to the lack of variability data for InGaAs devices (for statistically meaningful data sets) this comparison can only be performed for Si devices.

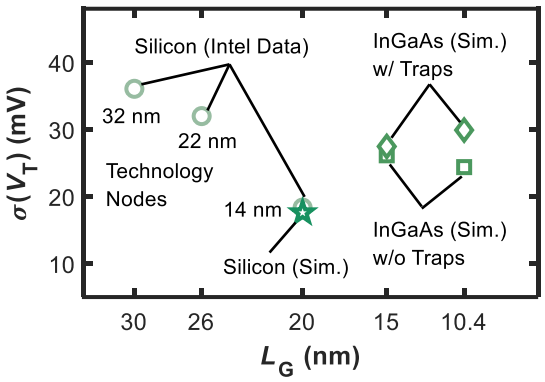


Figure 2.14: Comparison of the scaling behavior of total V_t variability, $\sigma(V_t)$, for InGaAs with- (diamonds) and without- (squares) ITs, from [52], Silicon Intel Data (circles), from [34], and Si QDD simulations (star), from [48]. Adapted from [55].

The result showing increases $\sigma_{TOT}(V_t)$ with scaling indicates that the significant increase of variability due to ITs could be a serious bottleneck for the adoption of InGaAs for ultra-scaled nodes for general-purpose digital applications.

2.4.5 Conclusions

We investigated the statistical variability of V_t in InGaAs DG-UTB MOSFETs by means of calibrated TCAD simulations. We included the most relevant variability sources known to affect performance, compared results with Si counterpart devices, and analyzed the effects of dimension scaling. The most relevant findings are summarized as follows:

- InGaAs devices have comparatively higher $\sigma(V_t)$ than Si counterparts as a result of stronger quantization effects in the ultrathin device channel.
- The most relevant variability sources at the 15nm node are WFF and BLER
- Dimension scaling affects V_t sensitivity to t_{ch} variations, configuring the control over t_{ch} as the most critical issue
- RDF-induced variability worsens due to scaling, being dominated by dopant fluctuations in the source/drain regions, especially for nominal N_{SD} lower than 5×10^{19} cm^{-3}

- At the 10.4nm node, although WFF remains the most detrimental variability source — if LER is kept under control by following ITRS prescriptions — the contribution due to ITF plays a significant role on the total V_t variability. The enhanced ITF contrasts the trend of reduced variability with smaller dimensions, as opposite to Si devices.

References – Chapter 2

[1] J. A. Del Alamo, "Nanometre-scale electronics with III-V compound semiconductors," *Nature*, vol. 479, no. 7373, pp. 317–323, 2011. DOI: 10.1038/nature10677.

[2] Ioffe Institute, *NSM Archive - Physical Properties of Semiconductors*. [Online]. Available: http://www.matprop.ru/ (visited on 10/29/2020).

[3] H. Brech, *Optimization of GaAs based High Electron Mobility Transistors by Numerical Simulations*, 1998. [Online]. Available: https://www.iue.tuwien.ac.at/phd/brech/diss.htm (visited on 12/20/2020).

[4] M. Lundstrom, *Fundamentals of Nanotransistors*, ser. Lessons from Nanoscience: A Lecture Notes Series. WORLD SCIENTIFIC, Sep. 2015, vol. 06. DOI: 10.1002/0471142905.hg1113s83.

[5] J. Lin, X. Cai, Y. Wu, D. A. Antoniadis, and J. A. Del Alamo, "Record maximum transconductance of 3.45 mS/μm for III-V FETs," *IEEE Electron Device Letters*, vol. 37, no. 4, pp. 381–384, Apr. 2016. DOI: 10.1109/LED.2016.2529653.

[6] W. Lu, Y. Lee, J. Murdzek, J. Gertsch, A. Vardi, L. Kong, S. M. George, and J. A. Del Alamo, "First Transistor Demonstration of Thermal Atomic Layer Etching: InGaAs FinFETs with sub-5 nm Fin-width Featuring in situ ALE-ALD," in *IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA: IEEE, Dec. 2018, pp. 39.1.1–39.1.4. DOI: 10.1109/IEDM.2018.8614536.

[7] C. Convertino, C. B. Zota, D. Caimi, M. Sousa, and L. Czornomaz, "InGaAs FinFETs 3-D Sequentially Integrated on FDSOI Si CMOS with Record Performance," *IEEE Journal*

of the Electron Devices Society, vol. 7, pp. 1170–1174, 2019. DOI: 10.1109/JEDS.2019.2928471.

[8] IEEE, "International Roadmap for Devices and Systems (IRDS) 2020 - More Moore," 2020. [Online]. Available: https://irds.ieee.org/editions/2020/more-moore.

[9] S. J. Sweeney, T. D. Eales, and A. R. Adams, "The impact of strained layers on current and emerging semiconductor laser systems," *Journal of Applied Physics*, vol. 125, no. 8, 2019. DOI: 10.1063/1.5063710. [Online]. Available: http://dx.doi.org/10.1063/1.5063710.

[10] A. Vardi, J. Lin, W. Lu, X. Zhao, A. Fernando-Saavedra, and J. A. Del Alamo, "A Si-Compatible Fabrication Process for Scaled Self-Aligned InGaAs FinFETs," *IEEE Transactions on Semiconductor Manufacturing*, vol. 30, no. 4, pp. 468–474, Nov. 2017. DOI: 10.1109/TSM.2017.2753141.

[11] S. Takagi, T. Irisawa, T. Tezuka, T. Numata, S. Nakaharai, N. Hirashita, Y. Moriyama, K. Usuda, E. Toyoda, S. Dissanayake, M. Shichijo, R. Nakane, S. Sugahara, M. Takenaka, and N. Sugiyama, "Carrier-transport-enhanced channel CMOS for improved power consumption and performance," *IEEE Transactions on Electron Devices*, vol. 55, no. 1, pp. 21–39, Jan. 2008. DOI: 10.1109/TED.2007.911034.

[12] N. Neophytou, T. Rakshit, and M. S. Lundstrom, "Performance analysis of 60-nm gate-length III-V InGaAs HEMTs: Simulations versus experiments," *IEEE Transactions on Electron Devices*, vol. 56, no. 7, pp. 1377–1387, Jul. 2009. DOI: 10.1109/TED.2009.2021437.

[13] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, 2nd. Cambridge University Press, 2009, p. 680, ISBN: 9781107635715. DOI: 10.1017/CBO9781139195065.

[14] Synopsys, *Sentaurus SDevice Manual (O-2018.06)*, 2018.

[15] N. Goel, P. Majhi, C. O. Chui, W. Tsai, D. Choi, and J. S. Harris, "InGaAs metal-oxide-semiconductor capacitors with HfO2 gate dielectric grown by atomic-layer deposition," *Applied Physics Letters*, vol. 89, no. 16, pp. 100–103, 2006. DOI: 10.1063/1.2363959.

[16] D. M. Fleetwood, P. S. Winokur, R. A. Reber, T. L. Meisenheimer, J. R. Schwank, M. R. Shaneyfelt, and L. C. Riewe, “Effects of oxide traps, interface traps, and "border traps" on metal-oxide-semiconductor devices,” *Journal of Applied Physics*, vol. 73, no. 10, pp. 5058–5074, 1993. DOI: 10.1063/1.353777.

[17] Y. Yuan, L. Wang, B. Yu, B. Shin, J. Ahn, P. C. McIntyre, P. M. Asbeck, M. J. Rodwell, and Y. Taur, “A distributed model for border traps in Al2O3 - InGaAs MOS devices,” *IEEE Electron Device Letters*, vol. 32, no. 4, pp. 485–487, 2011. DOI: 10.1109/LED.2011.2105241.

[18] A. Alian, M. A. Pourghaderi, Y. Mols, M. Cantoro, T. Ivanov, N. Collaert, and A. Thean, “Impact of the channel thickness on the performance of ultrathin InGaAs channel MOSFET devices,” in *IEEE International Electron Devices Meeting (IEDM)*, IEEE, Dec. 2013, pp. 437–440. DOI: 10.1109/IEDM.2013.6724644.

[19] V. Djara, V. Deshpande, M. Sousa, D. Caimi, L. Czornomaz, and J. Fompeyrine, *IEEE Electron Device Letters*, vol. 37, no. 2, pp. 169–172, Feb. 2016. DOI: 10.1109/LED.2015.2514080.

[20] N. Waldron, C. Merckling, L. Teugels, P. Ong, S. A. U. Ibrahim, F. Sebaai, A. Pourghaderi, K. Barla, N. Collaert, and A. V. Y. Thean, “InGaAs gate-all-around nanowire devices on 300mm Si substrates,” *IEEE Electron Device Letters*, vol. 35, no. 11, pp. 1097–1099, Nov. 2014. DOI: 10.1109/LED.2014.2359579.

[21] X. Zhao, J. Lin, C. Heidelberger, E. A. Fitzgerald, and J. A. del Alamo, “Vertical nanowire InGaAs MOSFETs fabricated by a top-down approach,” in *IEEE International Electron Devices Meeting (IEDM)*, IEEE, Dec. 2013, pp. 28.4.1–28.4.4. DOI: 10.1109/IEDM.2013.6724710.

[22] L. Selmi, E. Caruso, S. Carapezzi, M. Visciarelli, E. Gnani, N. Zagni, P. Pavan, P. Palestri, D. Esseni, A. Gnudi, S. Reggiani, F. M. Puglisi, and G. Verzellesi, “Modelling nanoscale n-MOSFETs with III-V compound semiconductor channels: From advanced models for band structures, electrostatics and transport to TCAD,” in *IEEE International Electron Devices Meeting (IEDM)*, IEEE, Dec. 2017, pp. 13.4.1–13.4.4. DOI: 10.1109/IEDM.2017.8268384.

[23] Y. Liu, N. Neophytou, G. Klimeck, and M. S. Lundstrom, "Band-Structure Effects on the Performance of III–V Ultrathin-Body SOI MOSFETs," *IEEE Transactions on Electron Devices*, vol. 55, no. 5, pp. 1116–1122, May 2008. DOI: 10.1109/TED.2008.919290.

[24] G. Zerveas, E. Caruso, G. Baccarani, L. Czornomaz, N. Daix, D. Esseni, E. Gnani, A. Gnudi, R. Grassi, M. Luisier, T. Markussen, P. Osgnach, P. Palestri, A. Schenk, L. Selmi, M. Sousa, K. Stokbro, and M. Visciarelli, "Comprehensive comparison and experimental validation of band-structure calculation methods in III-V semiconductor quantum wells," *Solid-State Electronics*, vol. 115, pp. 92–102, 2016. DOI: 10.1016/j.sse.2015.09.005.

[25] S. Carapezzi, E. Caruso, A. Gnudi, P. Palestri, S. Reggiani, and E. Gnani, "TCAD Mobility Model of III-V Short-Channel Double-Gate FETs Including Ballistic Corrections," *IEEE Transactions on Electron Devices*, vol. 64, no. 12, pp. 4882–4888, Dec. 2017. DOI: 10.1109/TED.2017.2759420.

[26] M. Rau, E. Caruso, D. Lizzit, P. Palestri, D. Esseni, A. Schenk, L. Selmi, and M. Luisier, "Performance projection of III-V ultra-thin-body, FinFET, and nanowire MOSFETs for two next-generation technology nodes," in *IEEE International Electron Devices Meeting (IEDM)*, Dec. 2017, pp. 30.6.1–30.6.4. DOI: 10.1109/IEDM.2016.7838515.

[27] L. Morassi, A. Padovani, G. Verzellesi, D. Veksler, I. Ok, and G. Bersuker, "Interface-Trap Effects in Inversion-Type Enhancement-Mode InGaAs/ZrO_2 N-Channel MOSFETs," *IEEE Transactions on Electron Devices*, vol. 58, no. 1, pp. 107–114, Jan. 2011. DOI: 10.1109/TED.2010.2086461.

[28] S. Sze and K. K. Ng, *Physics of Semiconductor Devices*, 3rd. Hoboken, NJ, USA: John Wiley & Sons, Inc., Oct. 2006. DOI: 10.1002/0470068329.

[29] M. Heyns, C. Adelmann, G. Brammertz, D. Brunco, M. Caymax, B. De Jaeger, A. Delabie, G. Eneman, M. Houssa, D. Lin, K. Martens, C. Merckling, M. Meuris, J. Mittard, J. Penaud, G. Pourtois, M. Scarrozza, E. Simoen, S. Sioncke, and W. E. Wang, "Ge and III/V devices for advanced CMOS," in *International Conference on ULtimate Integration of Silicon, ULIS*, IEEE, 2009, pp. 83–86. DOI: 10.1109/ULIS.2009.4897544.

[30] G. Brammertz, H. C. Lin, M. Caymax, M. Meuris, M. Heyns, and M. Passlack, "On the interface state density at In0.53 Ga0.47 As /oxide interfaces," *Applied Physics Letters*, vol. 95, no. 20, pp. 1–4, 2009. DOI: 10.1063/1.3267104.

[31] A. Asenov, S. Kaya, and A. R. Brown, "Intrinsic parameter fluctuations in decananometer MOSFETs introduced by gate line edge roughness," *IEEE Transactions on Electron Devices*, vol. 50, no. 5, pp. 1254–1260, May 2003. DOI: 10.1109/TED.2003.813457.

[32] S. Dongaonkar, S. P. Mudanai, and M. D. Giles, "From Process Corners to Statistical Circuit Design Methodology: Opportunities and Challenges," *IEEE Transactions on Electron Devices*, vol. 66, no. 1, pp. 19–27, 2019. DOI: 10.1109/TED.2018.2860929.

[33] C. B. Zota, F. Lindelow, L. E. Wernersson, and E. Lind, "InGaAs nanowire MOSFETs with ION = 555 μa/μm at IOFF = 100 nA/μm and VDD = 0.5 v," in *Symposium on VLSI Technology*, 2016, pp. 10–11. DOI: 10.1109/VLSIT.2016.7573418.

[34] "A 14nm Logic Technology Featuring 2 nd -Generation FinFET Interconnects , Self-Aligned Double Patterning and a 0.0588 μm2 SRAM cell size," in *IEEE International Electron Devices Meeting (IEDM)*, Dec. 2014, pp. 3.7.1–3.7.3. DOI: 10.1109/IEDM.2014.7046976.

[35] N. Agrawal, Y. Kimura, R. Arghavani, and S. Datta, "Impact of Transistor Architecture (Bulk Planar, Trigate on Bulk, Ultrathin-Body Planar SOI) and Material (Silicon or III-V Semiconductor) on Variation for Logic and SRAM Applications," *IEEE Transactions on Electron Devices*, vol. 60, no. 10, pp. 3298–3304, Oct. 2013. DOI: 10.1109/TED.2013.2277872.

[36] A. R. Brown, N. M. Idris, J. R. Watling, and A. Asenov, "Impact of metal gate granularity on threshold voltage variability: A full-scale three-dimensional statistical simulation study," *IEEE Electron Device Letters*, vol. 31, no. 11, pp. 1199–1201, Nov. 2010. DOI: 10.1109/LED.2010.2069080.

[37] A. Nidhi, V. Saripalli, V. Narayanan, Y. Kimura, R. Arghavani, and S. Datta, "Will strong quantum confinement effect limit low V CC logic application of III-V FINFETs?" In

Device Research Conference - Conference Digest, DRC, IEEE, 2012, pp. 231–232. DOI: 10.1109/DRC.2012.6256968.

[38] X. Wang, B. Cheng, A. R. Brown, C. Millar, J. B. Kuang, S. Nassif, and A. Asenov, "Interplay between process-induced and statistical variability in 14-nm CMOS technology double-gate SOI FinFETs," *IEEE Transactions on Electron Devices*, vol. 60, no. 8, pp. 2485–2492, 2013. DOI: 10.1109/TED.2013.2267745.

[39] A. R. Brown, V. Huard, and A. Asenov, "Statistical simulation of progressive NBTI degradation in a 45-nm technology pMOSFET," *IEEE Transactions on Electron Devices*, vol. 57, no. 9, pp. 2320–2323, Sep. 2010. DOI: 10.1109/TED.2010.2052694.

[40] C. G. Sodini, T. W. Ekstedt, and J. L. Moll, "Charge accumulation and mobility in thin dielectric MOS transistors," *Solid State Electronics*, vol. 25, no. 9, pp. 833–841, 1982. DOI: 10.1016/0038-1101(82)90170-8.

[41] P. Pavan, N. Zagni, F. M. Puglisi, A. Alian, A. V. Y. Thean, N. Collaert, and G. Verzellesi, "The impact of interface and border traps on current–voltage, capacitance–voltage, and split-CV mobility measurements in InGaAs MOSFETs," *Physica Status Solidi A*, vol. 214, no. 3, p. 1 600 592, Mar. 2017. DOI: 10.1002/pssa.201600592.

[42] J. Lin, Y. Y. Gomeniuk, S. Monaghan, I. M. Povey, K. Cherkaoui, É. O'Connor, M. Power, and P. K. Hurley, "An investigation of capacitance-voltage hysteresis in metal/high-k/In 0.53Ga0.47As metal-oxide-semiconductor capacitors," *Journal of Applied Physics*, vol. 114, no. 14, p. 144 105, 2013. DOI: 10.1063/1.4824066.

[43] X. Sun, S. Cui, A. Alian, G. Brammertz, C. Merckling, D. Lin, and T. P. Ma, "AC Transconductance Dispersion (ACGD): A Method to Profile Oxide Traps in MOSFETs Without Body Contact," *IEEE Electron Device Letters*, vol. 33, no. 3, pp. 438–440, Mar. 2012. DOI: 10.1109/LED.2011.2181318.

[44] F. Bonani, G. Ghione, M. R. Pinto, and R. Kent Smith, "An efficient approach to noise analysis through multidimensional physics-based models," *IEEE Transactions on Electron Devices*, vol. 45, no. 1, pp. 261–269, Jan. 1998. DOI: 10.1109/16.658840.

[45] N. Agrawal, H. Liu, R. Arghavani, V. Narayanan, and S. Datta, "Impact of variation in nanoscale silicon and non-silicon FinFETs and tunnel FETs on device and SRAM performance," *IEEE Transactions on Electron Devices*, vol. 62, no. 6, pp. 1691–1697, Jun. 2015. DOI: 10.1109/TED.2015.2406333.

[46] *2013 ITRS - International Technology Roadmap for Semiconductors*, 2013. [Online]. Available: http://www.itrs2.net/2013-itrs.html (visited on 11/11/2020).

[47] P. R. Pinsukanjana, J. M. Marquis, J. Hubbard, M. A. Trivedi, R. F. Dickey, J. M. Tsai, S. P. Kuo, P. S. Kao, and Y. C. Kao, "InGaAs composition monitoring for production MBE by in situ optical-based flux monitor (OFM)," *Journal of Crystal Growth*, vol. 251, no. 1-4, pp. 124–129, 2003. DOI: 10.1016/S0022-0248(02)02293-5.

[48] N. Zagni, F. M. Puglisi, G. Verzellesi, and P. Pavan, "Threshold Voltage Statistical Variability and Its Sensitivity to Critical Geometrical Parameters in Ultrascaled InGaAs and Silicon FETs," *IEEE Transactions on Electron Devices*, vol. 64, no. 11, pp. 4607–4614, Nov. 2017. DOI: 10.1109/TED.2017.2754323.

[49] J. P. Colinge, "Multiple-gate SOI MOSFETs," *Solid-State Electronics*, vol. 48, no. 6, pp. 897–905, 2004. DOI: 10.1016/j.sse.2003.12.020.

[50] N. Zagni, F. F. M. Puglisi, P. Pavan, and G. Verzellesi, "Random dopant fluctuation variability in scaled InGaAs dual-gate ultra-thin body MOSFETs: Source and drain doping effect," in *IEEE International Integrated Reliability Workshop Final Report*, 2018, pp. 1–4. DOI: 10.1109/IIRW.2017.8361227.

[51] A. G. Lind, H. L. Aldridge, C. Hatem, M. E. Law, and K. S. Jones, " Review—Dopant Selection Considerations and Equilibrium Thermal Processing Limits for n + -In 0. 53Ga 0.47 As," *ECS Journal of Solid State Science and Technology*, vol. 5, no. 5, Q125–Q131, 2016. DOI: 10.1149/2.0201605jss.

[52] N. Zagni, F. M. Puglisi, P. Pavan, and G. Verzellesi, "Effects of mole fraction variations and scaling on total variability in InGaAs MOSFETs," *Solid-State Electronics*, vol. 159, pp. 135–141, Mar. 2019. DOI: 10.1016/J.SSE.2019.03.048.

[53] P. Aguirre, M. Rau, and A. Schenk, "2D and 3D TCAD simulation of III-V channel FETs at the end of scaling," *Solid-State Electronics*, vol. 159, pp. 123–128, Sep. 2019. DOI: 10.1016/j.sse.2019.03.043.

[54] S. Reggiani, E. Gnani, A. Gnudi, M. Rudan, and G. Baccarani, "Low-field electron mobility model for ultrathin-body SOI and double-gate MOSFETs with extremely small silicon thicknesses," *IEEE Transactions on Electron Devices*, vol. 54, no. 9, pp. 2204–2212, Aug. 2007. DOI: 10.1109/TED.2007.902899.

[55] N. Zagni, E. Caruso, F. M. Puglisi, P. Pavan, P. Palestri, and G. Verzellesi, "Systematic Modeling of Electrostatics, Transport, and Statistical Variability Effects of Interface Traps in End-of-the-Roadmap III–V MOSFETs," *IEEE Transactions on Electron Devices*, vol. 67, no. 4, pp. 1560–1566, Apr. 2020. DOI: 10.1109/TED.2020.2974966.

Chapter 3

Compact Modeling of Noise and Aging in Emerging Non-Volatile Memories

Abstract — In this chapter we develop compact analytical models to evaluate noise and aging in emerging Non-Volatile Memory (NVM) technologies.

First, we derive an analytical formulation for Random Telegraph Noise (RTN) in RRAMs. The model is derived from refined physics-based simulations by applying geometrical simplifications and is validated on a wide set of experimental data. The model successfully captures the trends of resistance variation for different oxide materials. The RTN model can be seamlessly integrated into existing RRAM compact models, and its potential is exemplified by designing a simple RTN-based Random Number Generator (RNG).

Second, we evaluate the endurance of FeFETs by using an analytical expression of the Memory Window, MW. The analytical MW formulation allows to quickly estimate the generated trap concentration as a function of number of writing cycles (or time) without recurring to numerical simulations. With the aid of the analytical model, we find that for typical program/erase pulse amplitudes and duration, endurance has a weak dependence on writing conditions. The proposed endurance characterization technique based on the MW can be helpful to accelerate development of next-generation FeFETs.

3.1 NVMs to Tear Down the Memory Wall

In Chapter 1 we have discussed how recently the halt of scaling of *logic* devices started the quest towards finding alternative routes to keep the electronic industry growing. Besides improving speed and efficiency of logic devices, another bottleneck to the performance of computing systems is the so-called *memory-wall* [1]–[3]. Memory is a key component of any computing system that — depending on its specific requirements — requires both high capacity and low latency, i.e., the time taken to access and retrieve data.

The memory-wall is the performance gap between speed improvement rate of logic devices and that of Dynamic Random Access Memories (DRAMs), i.e., the basic circuit blocks of which the memory is composed of. This problem can be understood simply by considering the average time to access memory as follows [3]

$$t_{avg} = p \times p_c + (1 - p) \times t_m \tag{3.1}$$

where p_c (p_m) is the cache (DRAM) access time and p is the probability that the requested data is in the cache and that DRAM memory does not need being accessed to. Clearly, even by assuming that cache (or Static Random Access Memories (SRAMs)) speed scales as the microprocessor speed for simplicity and that data is often available in the cache (i.e., $p << 1$), the probability of accessing the DRAM is never going to be null. This inevitably makes t_{avg} increasing, thus leading to performance degradation and hitting of the memory-wall.

Tearing down the memory-wall is a critical issue that demands technology breakthroughs to be overcome [1], [2]. There are two main courses of action sought for by engineers to overcome the limitation of existing SRAM and DRAM technology. These involve [1]:

- physically locating microprocessors and memory blocks as close as possible to each other, thus enabling high-bandwidth data traffic between the two; and
- overcoming the boundary between logic and memory devices by merging the two functions together such that computation can be embedded within the memories themselves.

However, achieving either of the above-mentioned goals would not be sufficient to sustain

performance improvement because of the inefficiency of current memory cells. In fact, both SRAMs and DRAMs are volatile thus require power all the time to hold the stored information.

The solid-state device community therefore came with the solution of adopting novel embedded NVMs, that offer cell size and speed similar to DRAMs at zero stand-by power. On top of that, NVMs offer the possibility to co-integrate them with high performance logic devices to achieve Logic-In-Memory (LIM) functionality [1], [2]. The most classic example of an NVM technology is the Flash memory, that is composed of a conventional MOSFET with a floating gate layer (that enables charge storage) between the metal itself and the gate insulator. Despite Flash memories are fairly established, they are mostly employed for storage as the high voltage required for programming/erasing the modules and their poor endurance (i.e., the ability to withstand many writing cycles without degradation) makes them unfit for fast, reliable and power-saving applications (as required to bringing down the memory-wall).

Recently, several emerging technology options have been investigated to realize fast, scalable memory chips at ever decreasing cost. These are mainly: *i)* Spin-Transfer Torque Magnetic Random Access Memories (STT-MRAMs) [4], [5], *ii)* Ferroelectric Random Access Memories (FeRAMs) [6], [7], *iii)* FeFETs [7], [33], *iv)* Phase-Change Random Access Memories (PCRAMs) [8], and *v)* RRAMs [2], [15]. Several key device properties and performance metrics of these technologies are summarized in Table 3.1, along with the SRAM, DRAM and Flash memories for comparison.

In this chapter, we will focus on two of the novel NVM technologies listed above, namely HfO_2-based RRAMs and FeFET. Although their working principle is quite different, both RRAM and FeFET share a common important feature, i.e., the possibility of integration with current CMOS process. This feature is enabled by the use of HfO_2 as the insulating material, which is already widely employed as high-κ oxide layer in CMOS process, and it is responsible for conduction in RRAMs and for polarization switching in FeFETs. We will analyze both these technologies with the aid of relatively simple, compact analytical models that were derived to determine the role of traps on peculiar properties and metrics of the devices, namely, RTN in RRAMs and endurance in FeFETs.

Table 3.1: Key device parameters and performance metrics of volatile and non-volatile memories (F is the feature size). Adapted from [1].

	SRAM	**DRAM**	**Flash**	**STT-MRAM**	**FeRAM**	**FeFET**	**PCRAM**	**RRAM**
Cell Size (F^2)	120-150	10-30	10-30	10-30	10-30	10-30	10-30	10-30
Non-Volatility	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Write Voltage	$< 1\,V$	$< 1\,V$	$\approx 10\,V$	$< 1.5\,V$	$< 3\,V$	$< 4\,V$	$< 3\,V$	$< 3\,V$
Write Energy	$\approx 1\,fJ$	$\approx 10\,fJ$	$\approx 100\,pJ$	$\approx 1\,pJ$	$\approx 0.1\,pJ$	$\approx 0.1\,pJ$	$\approx 10\,pJ$	$\approx 1\,pJ$
Standby Power	High	Medium	Low	Low	Low	Low	Low	Low
Write Speed	$\approx 1\,ns$	$\approx 10\,ns$	$\approx 0.1-1\,ms$	$\approx 5\,ns$	$\approx 10\,ns$	$\approx 10\,ns$	$\approx 10\,ns$	$\approx 10\,ns$
Read Speed	$\approx 1\,ns$	$\approx 3\,ns$	$\approx 10\,ns$	$\approx 5\,ns$	$\approx 10\,ns$	$\approx 10\,ns$	$\approx 10\,ns$	$\approx 10\,ns$
Endurance	10^{16}	10^{16}	10^4-10^6	10^{15}	10^{14}	$> 10^5$	$> 10^{12}$	$> 10^7$

3.2 Background – Elements of RRAM Physics

RRAMs are two-terminal devices exhibiting either a high- or low-conductivity state determined by the diffusion and migration of defects that can be controlled electrically. The basic structure of the RRAM is the Metal-Insulator-Metal (MIM) configuration is depicted in Fig. 3.1(a), and consists of an oxide layer (in this case HfO_x[1]) sandwiched between two metal layers, namely the Top Electrode (TE) and Bottom Electrode (BE). Defects in the oxide layer are responsible for the conduction in both the LRS and HRS state and can be induced by different sources, such as oxygen vacancies in the filament (as in the case of HfO_x) or metallic ions injected by the electrodes [2]. As shown in Fig. 3.1(b), LRS is induced by applying a positive voltage to the TE leading to a controlled breakdown of the oxide itself [2], [10]. HRS instead, is induced by a

[1] x depends on the stoichiometry of the oxide. For instance, in HfO_2 two atoms of hafnium are present for each oxygen atom.

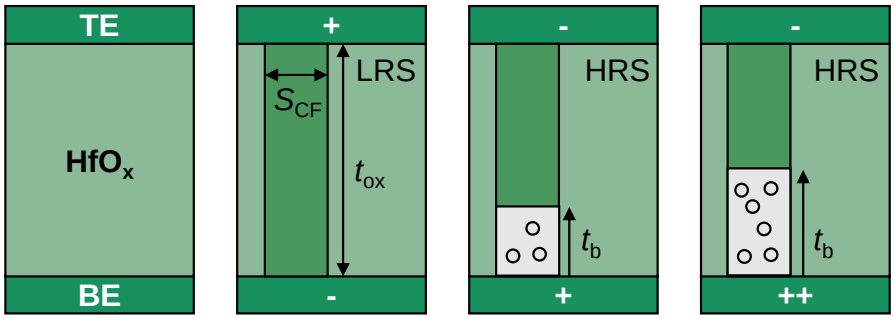


Figure 3.1: (a) Sketch of a RRAM in the MIM configuration. TE and BE are also indicated. (b) A Conductive Filament (CF) is formed by applying a positive voltage on the TE, determining the Low-Resistance State (LRS). (c) High-Resistance State (HRS) is induced by applying a positive voltage to the BE, which partially oxidizes the CF due to the formation of a barrier. The barrier thickness, t_B, is controlled by the reset bias as shown by (c) and (d). Adapted from [13].

.

positive voltage applied on the BE and consists in the partial oxidation of the CF due to field- and temperature-driven movement and recombination of oxygen ions with both oxygen vacancies and metallic Hf atoms in the CF [10]. A barrier forms due to the partial oxidation of the CF and its thickness (t_B) depends on the magnitude of the reset bias, compare Fig. 3.1(c) and (d).

The non-volatile memory property in RRAM is encoded in its resistance, which can be electrically switched between the LRS and HRS. A compact model describing RRAM behavior in both states is described in [14] and its main features as described in the following[2]. The model takes inspiration from the physical mechanism occurring in RRAM devices and introduces some approximations to enable a compact implementation. Specifically, the equations reported below consider the CF and the dielectric barrier contributions that determine the total resistance of the

[2]The Verilog-A code for the NCFET model described here is available at the following link: https://nanohub.org/publications/289/1.

device

$$R_{LRS} = \frac{\rho \times t_{CF}}{S_{CF}} \tag{3.2a}$$

$$\rho = \rho_0(1 + \alpha(T_{CF} - T_0)) \tag{3.2b}$$

$$R_{CF} = R_{LRS}\frac{t_{CF} - t_B}{t_{CF}} \tag{3.2c}$$

$$R_B = R_{LRS}\beta\left(e^{t_B/k} - 1\right)\exp\frac{E_A}{k_B T_B} \tag{3.2d}$$

Eq. (3.2a) models the resistance of the device in the LRS. Eq. (3.2b) models the influence of the CF temperature (T_{CF}) with the thermal coefficient α as a fitting parameter. Eq. (3.2c) models the CF resistance scaling with the increasing of barrier thickness (t_B). Eq. (3.2d) models R_B by considering Trap-Assisted Tunneling (TAT) as the dominating conduction mechanism through the barrier. Specifically, k is the typical tunneling length [10], [12], E_A is the activation energy of the HRS current (extracted from its Arrhenius plot) and β is a proportionality constant that can be estimated from experimental measurements. The overall device current (I_{RRAM}) is computed by considering the series connection of R_{CF} and R_B as follows

$$I_{RRAM} = \frac{V_{0,LRS}}{R_{CF}}\sinh\left(\frac{V_{CF}}{V_{0,LRS}}\right) \tag{3.3a}$$

$$V_B = V_{0,HRS}\sinh^{-1}\left(I_{RRAM}\frac{R_B}{V_{0,HRS}}\right) \tag{3.3b}$$

$$V_B = V_{RRAM} - V_{CF} \tag{3.3c}$$

where $V_{0,LRS}$ and $V_{0,HRS}$ model the non-linear conduction in the CF and in the barrier, respectively. t_B and T_B are the states variables of the model, whose evolution over time needs to be computed to correctly reproduce the switching dynamics. The differential equations governing the full dynamics are reported in [14]. To show the capability of the model, Fig. 3.2 shows the agreement between the experimental DC $I-V$ characteristics (symbols) for two different technologies and those calculated with the model (solid lines). As the compact model described here is physics-based, most of the model parameters are material related. Therefore, these parameters were from the literature and/or from repositories of material properties [14].

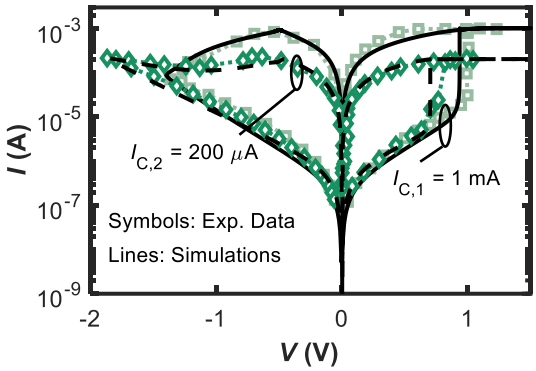


Figure 3.2: Experimental (symbols) and simulated (curves) $I-V$ characteristics covering the full switching cycle from set to reset. The devices (belonging to two different technologies) were formed at two different compliance currents, $I_{C,1} = 1\,\text{mA}$ (light green squares, black solid curves) and $I_{C,1} = 200\,\mu\text{A}$ (dark green diamonds, black dashed curves). Adapted from [14].

.

3.3 Contribution: Exploiting Noise in RRAMs to Design Advanced Circuit Applications

RTN is a phenomenon associated with charge trapping/de-trapping occurring at discrete defect sites present in the gate oxide of MOSFETs or in the oxide layer of RRAMs, to name a few. While RTN can represent a serious reliability concern in both cases [19], [28], its intrinsic stochastic nature can be actually exploited in order to enable novel circuit applications such as Physical Unclonable Function (PUF) [20] or RNG [21]. Therefore, modeling of RTN effects in both the HRS and LRS is imperative in order to quantify the detrimental impact of RTN on the read margins of RRAMs [27], [28] and to design novel applications such the ones mentioned earlier. Goal of this section is to present a compact model for RTN in RRAMs that can be seamlessly integrated into existing RRAM models thus harnessing the effects of stochastic noise for circuit simulations. The proposed compact model for RTN in RRAM: *i)* is valid in both HRS and LRS, correctly captures *ii)* the statistics of the RTN properties (i.e., variations in amplitude and transition times), and *iii)* the variability in the number of defects contributing to the RTN. The model is implemented in Verilog-A, and its effectiveness is demonstrated by using it to design the building block of a RNG circuit exploiting the RTN randomness as an entropy source.

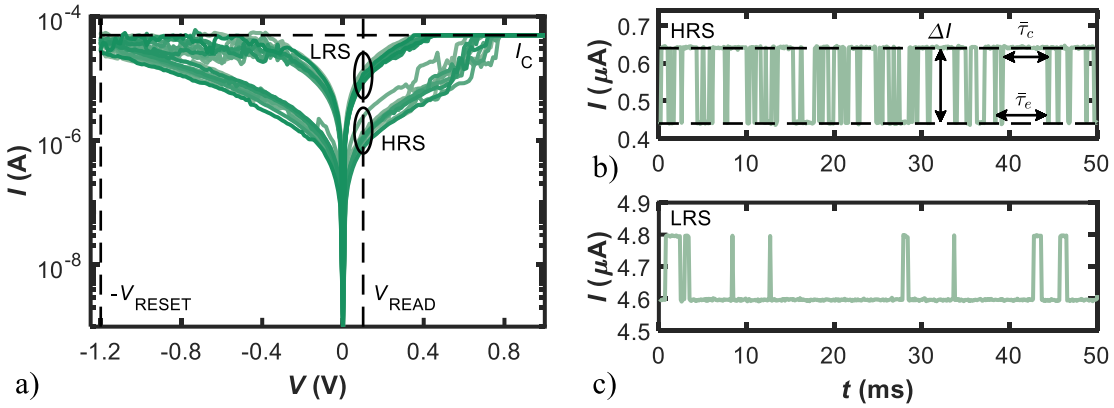


Figure 3.3: (a) Experimental $I-V$ curves measured at 20 consecutive switching cycles on a TiN/Ti/HfO_2/TiN device using $I_C = 50\,\mu A$ and $V_{RESET} = 1.2\,V$, displaying cycle-to-cycle variability. (b), (c) Two-level RTN, related to a single trap, as detected in HRS and in LRS, respectively. The statistical RTN parameters are evidenced (i.e., the average capture time, $\bar{\tau}_c$, the average emission time, $\bar{\tau}_e$, and the fluctuation amplitude, ΔI). Adapted from [28].

3.3.1 RTN Compact Model in HRS and LRS

A representative $I-V$ characteristic of the RRAM device exhibiting both RTN and cycling variability is shown in Fig. 3.3. Developing a physics-based compact model for RRAM devices that includes the effects of RTN requires discussing the underlying mechanisms dominating charge transport in the two resistive states. In fact, in general the RTN is a temporary alteration of the current flow [17], determined by different physical mechanisms in either HRS and LRS [11], [17], [22]. Therefore, we discuss the charge transport mechanism and its alterations separately for the two states.

RTN Statistical Model in HRS

The reset operation, driving the device in HRS, leads to the partial reoxidation of the CF, creating a dielectric barrier [11]–[13], [22], [27], as shown in Fig. 3.4(a)-(c). The dielectric barrier thickness, t_B, is determined by the reset conditions, i.e., V_{RESET} and temperature (T). The HRS current is dominated by the electron TAT at positively charged oxygen vacancy defects (Vo^+) in the reoxidized tip of the CF, i.e., in the dielectric barrier [11]–[13], [17], [23].

Although the actual physical mechanism responsible for RTN in HRS is still under debate (see

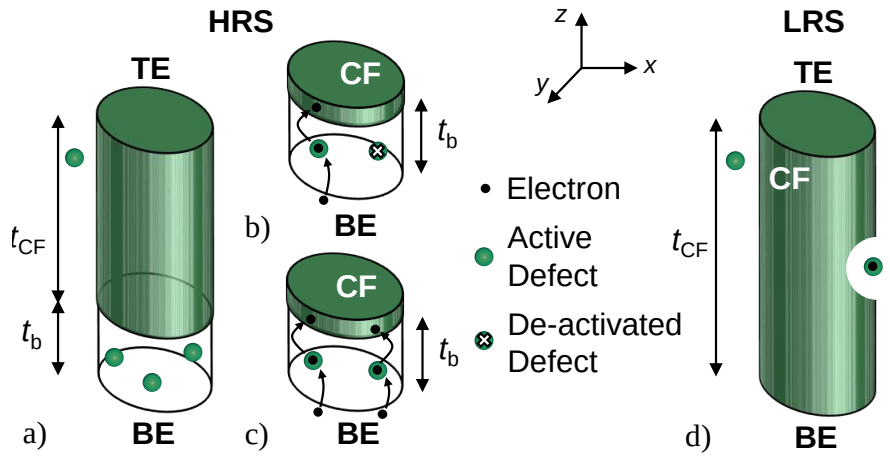


Figure 3.4: (a) Schematic picture of the device and of the RTN mechanism in HRS. (b), (c) show the defects de-activation and activation, respectively. (d) shows the RTN mechanism in LRS, indicating the electrostatic coupling between a charged defect and the CF. Adapted from [27], [28].

[28] and references therein), it is widely believed that it is fact due to the temporary (de)-activation of Vo^+ defects, distributed in the barrier that assist charge transport, as sketched in Fig. 3.4(b),(c). The mechanisms leading to (de)-activation of defects is also due to charge (de)-trapping in additional slow traps that do not participate to charge transport (supposedly oxygen interstitial atoms). Physics-based simulations performed in [28] that only a small fraction of all the Vo^2 defects in the barrier is responsible for almost the whole HRS current (i.e., >95%), regardless of t_B (determined by reset conditions), V_{READ}, and T. This is because the current assisted by a defect is maximized when it has small, comparable $\bar{\tau}_c$ and $\bar{\tau}_e$, which is typical for defects in the middle of the barrier. As such, since RTN is due to the random (de)-activation of these defects, the average $\Delta I/I_{HRS}$ (or equivalently $\Delta R/R_{HRS}$)[3] is expected to be constant in every operating condition. Devising a simple average formulation for the RTN amplitude in HRS requires estimating the average ΔI related to the (de)-activation of individual defects that strongly depends on their location, energy, cross section, as well as on the local voltage and temperature profile. Results as shown in Fig. 3.5 indicate that the probability distribution of $\Delta I/I_{HRS}$ appears lognormal and agnostic to the operating conditions, in agreement with experimental data [22].

[3] ΔI is defined as the current contribution given by an individual Vo^+ defect when active [22], see Fig. 3.4(b),(c), while the overall current I is the sum of all the current contributions.

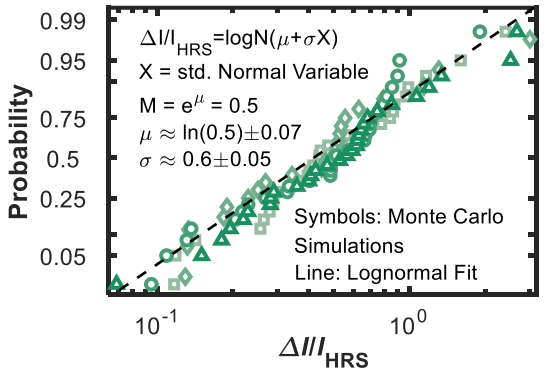


Figure 3.5: Distribution of simulated $\Delta I/I$ in HRS obtained using kinetic MC simulations. The I vs time traces related to a $10 \times 10\,\mathrm{nm}^2$ MIM device with TiN electrodes with a 1nm-thick barrier are simulated, including the effect of Vo^+ defect activation and deactivation. Vo^+ are randomly distributed in the barrier with a density of $N_T = 2 \times 10^{21}\,\mathrm{cm}^{-3}$. Simulations are performed in different conditions (V_{READ}, t_B, and temperature) and are repeated on 10 different device realizations by randomizing defect positions and energies (a total of 40 simulation runs). Adapted from [27], [28].

This lets writing

$$M\left(\frac{\Delta R}{R_{HRS}}\right) \approx \frac{1}{2}, \quad \sigma\left(\frac{\Delta R}{R_{HRS}}\right) \approx 0.6 \tag{3.4}$$

Hence, despite the significant intrinsic variability and the complexity of the RTN mechanism, it is possible to conceive a simple statistical description of RTN fluctuations amplitude in HRS in every operating condition.

RTN Statistical Model in LRS

In LRS, the device is characterized by the presence of a CF shunting the two electrodes, see Fig. 3.4(d). The CF, created during the preliminary forming operation, is generated by tightly packed oxygen vacancy defects, and its size is controlled by the current compliance I_C used in the forming and set operations [11]–[13], [22], [27]. In this resistive state, charge transport is due to the drift of de-localized electrons in the CF, consistently with the ohmic-like behavior generally observed in LRS. The ohmic-like transport is attributed to the formation of a conductive subband in the CF itself [11].

In LRS, RTN is commonly attributed to electron trapping and detrapping at individual defect sites in the proximity of the CF, i.e., within one electron Debye length, λ (see [28] and references

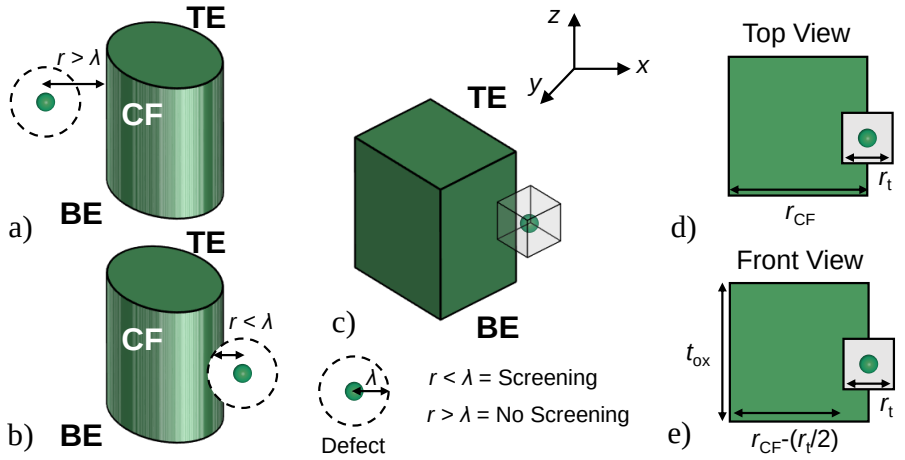


Figure 3.6: Schematic of RTN mechanism in LRS. Defects close to the CF ($r < \lambda$) cause a screening effect on a portion of the CF, inducing a resistance change. (a) No screening ($r > \lambda$). (b) Screening on a wide-area CF. (c) Simplified geometrical framework with its (d) top view and (e) front view. Adapted from [28].

therein). The trapped charge perturbs the potential in its surroundings causing a screening effect on the portion of the CF close to the defect, which induces a resistance change as shown in Fig. 3.4(d) and schematically repeated in Fig. 3.6. Unlike for the HRS case, the relative current/resistance change *depends* on the CF geometry, i.e., its radius, r_{CF} (assuming a cylindrical CF), and its thickness, t_B. For a large CF (i.e., $r_{CF} >> \lambda$) the CF section screened by a defect is much smaller than the CF area, $S_{CF} = \pi r_{CF}^2$, causing a relatively small resistance change (see [28] and references therein). As the CF gets smaller, the screened portion of the CF gets comparatively wider, with a larger impact on the average relative resistance change. Nonetheless, while the average effect is dependent on the CF size, the deviations from the average are CF size independent, as they only depend on the distance between the defect and the CF edge. This results in the lognormal $\Delta I/I_{LRS}$ distribution having a slope, i.e., σ, of about 0.3 that is invariant with the CF size (see [28] and references therein). Conversely, the median $\Delta I/I_{LRS}$ value changes with the CF size

$$M\left(\frac{\Delta R}{R_{LRS}}\right) = f(S_{CF}), \quad \sigma\left(\frac{\Delta R}{R_{LRS}}\right) \approx 0.3 \tag{3.5}$$

Therefore, once the relation between S_{CF} and the median $\Delta R/R_{LRS}$ is found, it is possible to estimate the ΔI of any RTN current fluctuation regardless of the resistive state and of the operating

conditions. To write $f(S_{CF})$, we exploit the CF geometry to calculate the relative resistance change, $\Delta R/R_{LRS}$ induced in the CF by a charge trapped by a close defect site. Still, $\Delta R/R_{LRS}$ depends on many variables, namely

- the distance between the defect and the CF edge,
- the local CF cross section and its composition
- the effective screening length of an electron in a complex medium, and
- the coupling between the electric field of the trapped electron and a portion of the CF.

Therefore, some reasonable simplifying assumptions need to be made.

We assume an ohmic-like charge transport in the CF, justified by the linear $I-V$ characteristics typically observed in LRS, see Fig. 3.3(a), and a uniform CF composition, i.e., a constant CF resistivity (ρ_{CF}) in space. In addition, we consider the CF geometrically uniform (i.e., a prism-shaped CF with arbitrary base profile). Though the exact shape of the CF is currently under debate in the scientific community, this assumption is extremely useful in easing calculations and is frequently adopted [12], [13], [24]. To simplify the problem, here, we assume an individual defect site placed at the edge of the CF (i.e., we assume zero distance between the defect and the CF edge), as shown in Fig. 3.6. This defect site, when filled with an electron, exhibits the maximum screening effect on the CF. Moreover, we consider that the trapped charge produces uniform screening in a spherical region around itself, and no screening beyond this region. In addition, we consider the screened portion of the CF to be completely unavailable to charge transport (i.e., we consider its resistance to be infinite). The full CF resistance with no screening effect, R_{LRS}, and the relative resistance change induced in the CF by the trapped charge ($\Delta R/R_{LRS}$) can be written as

$$R_{LRS} = \frac{\rho_{CF} \times t_{CF}}{S_{CF}}, \quad \frac{\Delta R}{R_{LRS}} = \frac{R^*_{LRS} - R_{RLS}}{R_{LRS}} \tag{3.6}$$

where R^*_{LRS} is the resistance of the CF when the screening effect occurs.

Here, to simplify calculations we consider the CF to be a prism with a square base and cross section S_{CF}, i.e., with side r_{CF}, see Fig. 3.6(c)-(e). This allows rewriting $R^0_{LRS} = \rho_{CF} \times t_{CF}/r^2_{CF}$. In addition, we simplify the spherical electric field of the trapped charge to a cube with side r_t, a

parameter accounting for the effective screening length of the trapped charge (related to λ [15], [25]). It is worth noting that r_t can be used as a fitting parameter, effectively compensating for the assumption of zero distance between the CF and the trapped charge. Under these conditions, the $\Delta R/R_{LRS}$ in Eq. (3.6) can be written in terms of r_t and of the CF properties (t_{CF}, r_{CF}, and r_{CF}) as follows

$$\frac{\Delta R}{R_{LRS}} = \frac{\left[\frac{\rho_{CF}(t_{CF}-r_t)}{r_{CF}^2} + \frac{\rho_{CF}(r_t)}{\left(r_{CF}^2 - r_t^2/2\right)}\right] - R_{LRS}^0}{R_{LRS}^0}. \tag{3.7}$$

With some algebra, Eq. (3.7) can be rewritten as

$$\frac{\Delta R}{R_{LRS}} = \frac{r_t^3}{2t_{CF} \times \left(r_{CF^2} - \frac{r_t^2}{2}\right)} \approx \frac{r_t^3}{2t_{CF} \times r_{CF}^2} = \frac{r_t^3}{2t_{CF} \times S_{CF}}, \tag{3.8}$$

where the last passage is carried out by assuming $r_{CF}^2 \gg r_t^2/2$. Eq. (3.8) is the expression used in the compact model to estimate the relative resistance change (or equivalently, relative current change) induced in the CF by a trapped charge at a close defect site.

Capture and Emission Times

The last step involves the calculation of $\tau_{c,e}$ to determine the complete statistical model for each defect contributing to RTN. This is necessary to reproduce the RTN signals over time, i.e., to perform accurate transient simulations with the proposed compact model. Since the physical mechanism causing RTN is associated with charge (de)-trapping in both resistive states, we calculate for each defect the related $\bar{\tau}_c$ and $\bar{\tau}_e$ using the multiphonon TAT compact formalism adopted in [12], [13], [26]. These formulas, described in [26] require defining V, T, and the defect distance from the electrodes (d). Their basic form in fact reads

$$\tau_c \propto \exp\left(\frac{d}{\lambda_c}\right)\exp\left(\frac{E_{A,c}(V)}{k_B T}\right), \quad \tau_e \propto \exp\left(\frac{d}{\lambda_e}\right)\exp\left(\frac{E_{A,e}(V)}{k_B T}\right). \tag{3.9}$$

The distance from the electrodes d of each defect is determined by their vertical position within the dielectric barrier when the device is in HRS, or alongside the CF within the whole insulating layer when the device is in LRS. Since charge carriers can be captured from (and emitted to) either

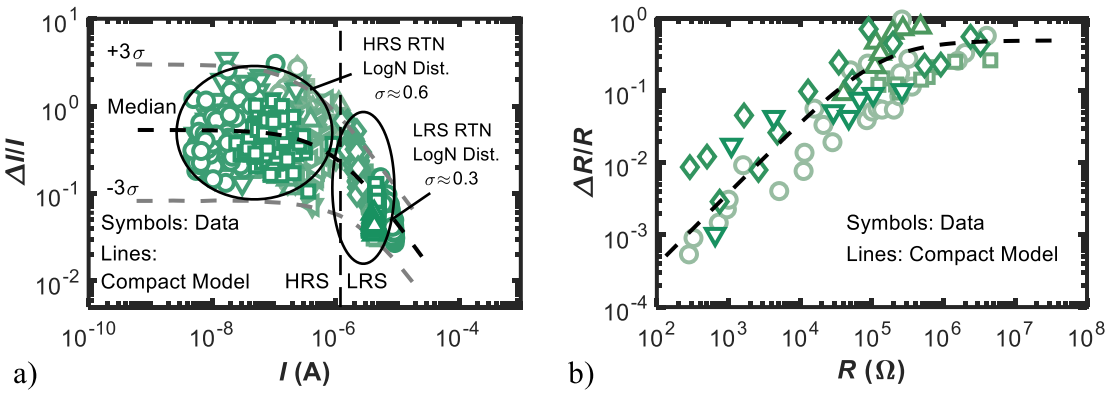


Figure 3.7: (a) Experimental $\Delta I/I$ vs I (symbols) in HRS and LRS at different operating conditions (I_C, V_{RESET}, and T) for five devices (different symbols) with a 5-nm HfO_2 layer. (b) Compact model predictions plotted as $\Delta R/R$ vs R (dashed line) agree with experimental data (symbols) in HRS and LRS taken from [15]–[17]. Adapted from [28].

the top or the bottom electrodes, $\bar{\tau}_c$ and $\bar{\tau}_e$ are calculated for both cases (two possible capture and two possible emission processes), and the minimum value is considered, as it is representative of the most likely event between the two. The defect properties are effectively lumped in the typical capture and emission lengths, λ_c and λ_e, respectively, and in the capture and emission activation energies, $E_{A,c}$ a $E_{A,e}$, respectively. These parameters only depend on the defect typology [12], [13], [23], [26] (i.e., oxygen ions in HRS and Vo_+ in LRS). These parameters are calculated exploiting *ab-initio* calculations (e..g, DFT or molecular dynamics) [12], [13], [23], [26].

3.3.2 Model Validation

The compact model for RTN developed here was validated by comparison with an extensive experimental data set, see Fig. 3.7. Five devices with a 5-nm HfO_2 layer were formed at different I_C values and then cycled through HRS and LRS for 100 times to collect RTN data in both resistive states, also accounting for the effects of cycle-to-cycle and device-to-device variability. Model predictions (dashed lines) are in excellent agreement with actual data. The extracted σ values of the normal distributions associated with the lognormal $\Delta I/I$ distributions in HRS and LRS agree with the values extracted from simulations and used in the compact model. This is further confirmed by the fact that almost the entire experimental data falls within the $\pm 3\sigma$

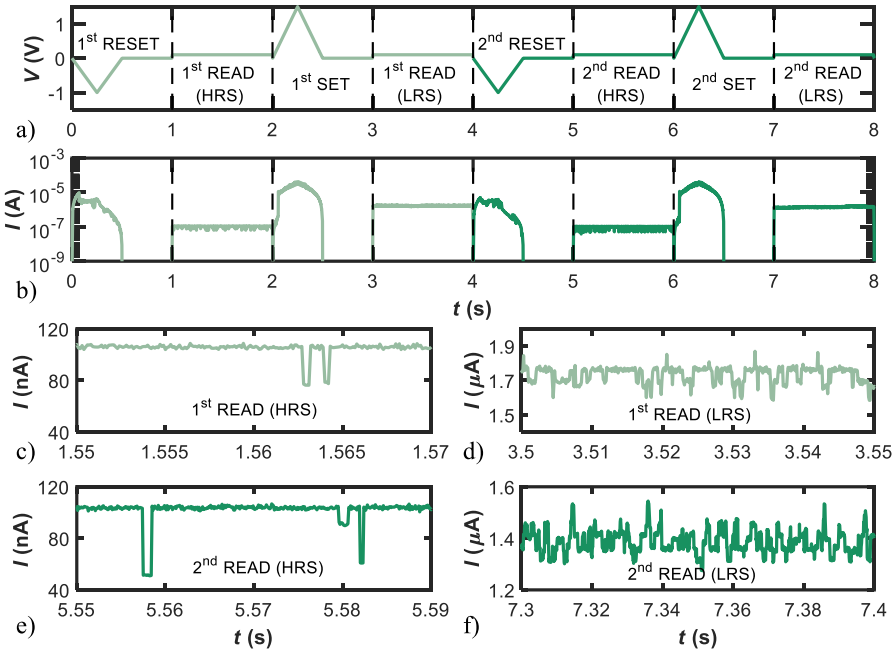


Figure 3.8: Simulations of resistance switching and RTN readout in HRS and LRS. A RRAM with $R_{LRS} = 15\,\mathrm{k}\Omega$ was considered. The voltage waveform in (a) is applied to perform the reset ($V_{RESET} = 1\,\mathrm{V}$), the set ($V_{SET} = 1.5\,\mathrm{V}$), and the readout operation ($V_{READ} = 100\,\mathrm{mV}$) for two consecutive cycles. (b) Readout current shows RTN signals in both states and (c)-(f) are zoomed-in views in either HRS or LRS. The cycle-to-cycle variability of RTN is correctly accounted for, as highlighted by different RTN patterns obtained in the same resistive states in two consecutive cycles. Adapted from [28].

predictions (gray-dashed lines) in Fig. 3.7(a). Moreover, the compact model prediction is also in excellent agreement with data taken from [15]–[17] for HfO_2 RRAMs and for devices made of different materials (e.g., Cu and NiO). This is consistent with the fact that the model is agnostic to the ρ_{CF}, and suggests that the same (or a similar) RTN mechanism may hold for devices employing different materials.

The results of RTN readout simulations in HRS and LRS are reported in Fig. 3.8. These results were obtained by performing SPICE simulations with a RRAM compact model [12], [13] integrated with the RTN module here proposed. Details of its actual implementation are described in [27].

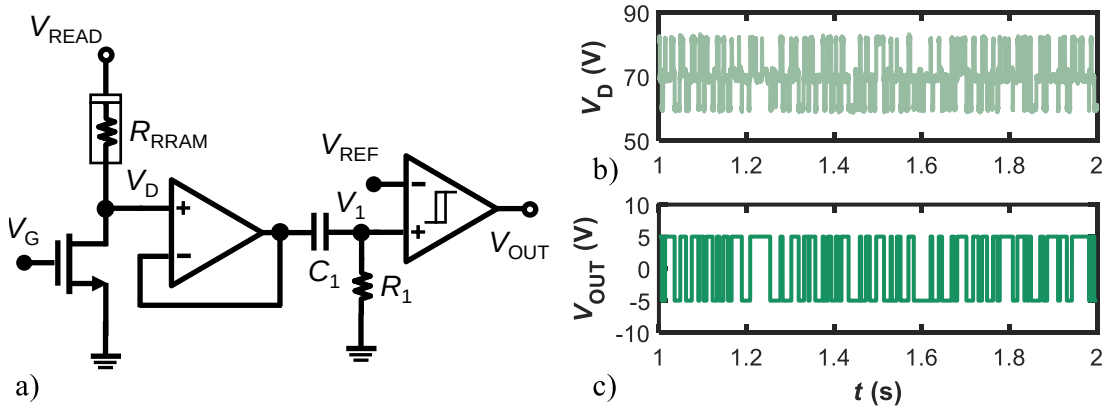


Figure 3.9: (a) Schematic of the RNG circuit based on the RTN in the RRAM device. (b) RTN voltage fluctuations as detected at the transistor drain terminal. (c) Random RTN pattern is reproduced at the circuit output, giving a random stream with an excellent randomness (50.8%). Adapted from [28].

3.3.3 Application: Random Number Generator

The proposed model allows considering the effect of RTN in circuit simulations, which is advantageous in the design of circuits for many applications [28]. Here, we discuss the design of a RTN-based RNG.

Recently, the presence of RTN signals in RRAM devices has drawn attention for their inherent randomness can be exploited as a possible source of entropy in RNG circuits [21]. Nevertheless, practical design of such circuits can be attained only by introducing suitable compact models, as the one proposed in this paper. The circuit topology we explore, shown in Fig. 3.9(a), finds implementation in high reliability systems for the generation of truly random numbers. The circuit, composed of the RRAM device in HRS and a series transistor, a buffer with a high-pass filter, and a Schmitt comparator, has been implement in Cadence Virtuoso to perform SPICE simulations. We included an RRAM device initialized in HRS, with $t_b = 1\,\text{nm}$. The results of a representative transient simulation are shown in Fig. 3.9(b) and (c).The application of a constant voltage V_{READ} to the top electrode of the RRAM causes voltage RTN fluctuations to appear at the buffer input Fig. 3.9(b). The RTN pattern is transferred to the high-pass filter (to get rid of unwanted DC and low-frequency components) and is then fed to the Schmitt comparator. The reference voltage V_{ref} regulates the amount of hysteresis that the comparator can withstand. The

randomness in the RTN signal produces a random bit stream at the output of the comparator, see Fig. 3.9(c), achieving a randomness value of nearly 50%. The simulation of this circuit topology shows the potential of the proposed model for advanced circuit design.

3.3.4 Conclusions

We developed a compact model for RTN in RRAM — valid in both HRS and LRS — by combining refined physics-based simulations with some geometrical simplifications. The model correctly captures the complex physical mechanisms at the basis of RTN in both states. Its predictions are validated on a wide experimental data set retrieved by measuring many devices in different operating conditions. Moreover, the model successfully captures the trends reported in the literature for different oxide materials. The RTN model can be seamlessly integrated into existing RRAM compact models. The potential of the model as a tool for advanced circuit design is exemplified by designing a simple RTN-based RNG circuit.

3.4 Background – Basics of FeFET Physics

Ferroelectric-based memories are a class of devices that exploit the hysteretic polarization-field loop of ferroelectric materials to encode information [7]. Ferroelectrics are characterized by two stable polarization states at zero electrical field that can be switched from one value to the other by applying an electrical field that is larger than the coercive field (i.e., the field that needs to be overcome to obtain polarization switching). FeFETs integrate a ferroelectric layer sandwiched between the gate metal and gate insulator to store the information as either positive or negative polarization, that effectively shifts the V_t of the underlying MOSFET. Because of the V_t shift, the $I-V$ characteristics depends on the polarization of the ferroelectric layer that retains itself even when no power is applied.

The first demonstration of an FeFET dates back to 1963 [29] but for more than 50 years there was no major leap towards the integration of FeFETs in existing technology. The main culprit to the integration of traditional FeFETs in CMOS process (which effectively simplifies design and reduces fabrication costs) was related to the use of inorganic perovskite-based ferroelectric materials such as Lead Zirconium Titanate (PZT), Barium Titanate (BTO) or Strontium Bismuth Tantalate (SBT). These materials exhibit coercive fields in the range of 30-50 $\mathrm{kV/cm}$ and thereby require great thickness to have appreciable memory window (MW) [7], which goes in contrast to scaling rules. In fact, high ferroelectric thickness (t_{FE}) in FeFETs also limits the scaling of the underlying MOSFET thus requiring large chips to guarantee appreciable memory capacity. The problem is exemplified by Fig. 3.10, showing how perovskite (or organic) ferroeletric-based FeFET simply could not match the scaling rate dictated by ITRS roadamp (and later IRDS). Things changed dramatically with the discovery of ferroelectricity in HfO_2 [30], [31]. When doped with Zirconium or Silicon, HfO_2 can exhibit ferroelectricity in its orthorhombic phase with coercive field in the range of 0.8-2 $\mathrm{MV/cm}$, which is significantly higher than the previously discussed materials. On top of this, its compatibility with existing CMOS process makes HfO_2 based-FeFET a game-changer technology [7] with ultra-scaled demonstrations already present in the literature [32], [33].

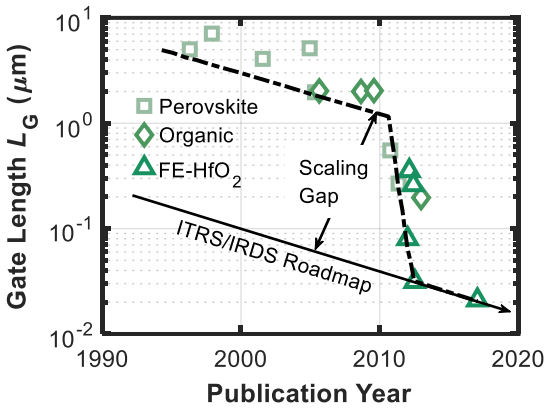


Figure 3.10: Physical gate length scaling of different FeFET technologies compared to the requirements of ITRS/IRDS roadmap. Adapted from [33].

3.4.1 Ferroelectric Polarization Loop Description with Landau Theory

The most common FeFET configuration is the Metal-Ferroelectric-Insulator-Semiconductor (MFIS) one, as shown in Fig. 3.11(a), and it integrates the ferroelectric layer between the gate metal and gate insulator. While for a conventional dielectric material the polarization charge induced on its terminals depends linearly on the applied voltage (through the dielectric constant), a ferroelectric material presents an hysteresis in its $P-E$ loop, as shown in Fig. 3.11(b). Hysteresis originates from displacement of ions or atoms inside the material which in its most simplified picture presents two stable, symmetric points at which $|P| > 0$ [7]. Switching occurs between these two stable points by applying an electric field that is larger than the coercive field of the ferroelectric, whereas applying a smaller bias leads to no switching.

The most common model to describe the $P-E$ hysteretic loop of ferroelectrics is the Preisach model that reads [34], [14, Ch. 2]

$$P(E) = P_s \tanh\left(\frac{E \pm E_C}{2\delta}\right) \tag{3.10a}$$

$$\delta = E_C \left[\ln\frac{P_s + P_r}{P_s - P_r}\right]^{-1} \tag{3.10b}$$

where P_s, P_r, E_C are the ferroelectric saturation (remnant) polarization and coercive field, respectively. Here however, we will make use of LT to describe the polarization loop of ferroelectrics.

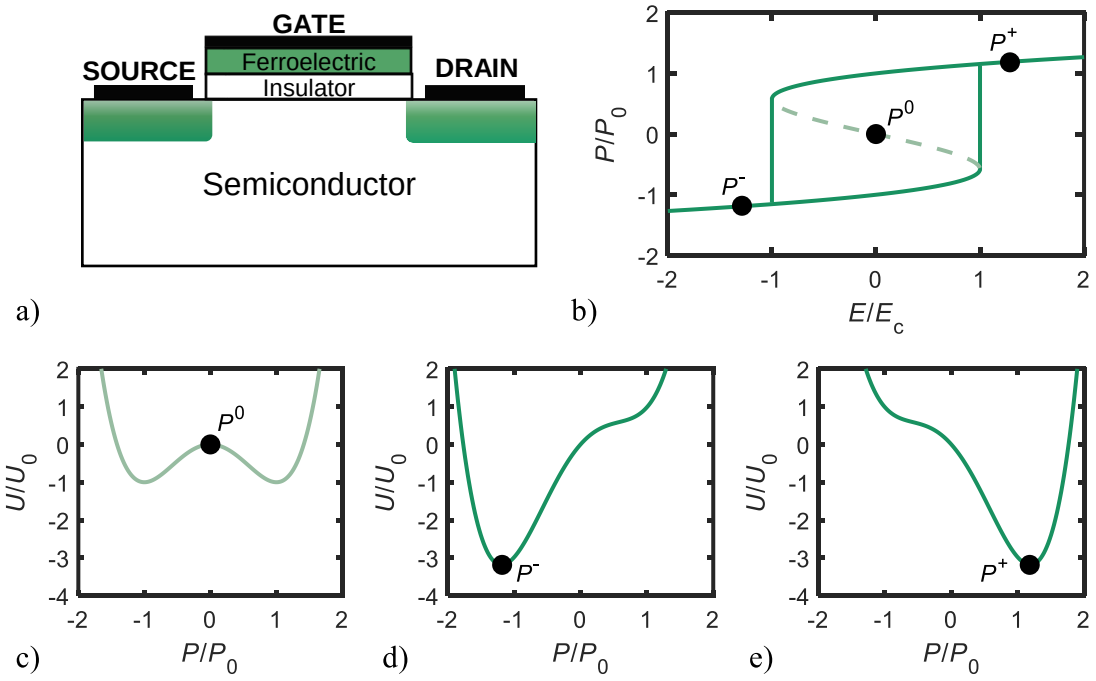


Figure 3.11: (a) Sketch of a FeFET in the MFIS configuration. (b) Stand-alone ferroelectric Polarization-Electric Field Relationship ($P-E$) loop as obtained from Landau Theory (LT), showing the negative (dashed line) and positive (solid lines) capacitance region. The corresponding energy landscapes for the three points denoted in (b) are shown in (c)-(e). Landau parameters are for SBT: $\alpha = -6.5\times10^7\,\mathrm{m/F}$, $\beta = 3.75\times10^9\,\mathrm{m^5/FC^2}$, $\gamma = 0\,\mathrm{m^9/FC^4}$, $\rho = 0\,\Omega\mathrm{m}$ [36].

.

This approach allows us to derive simple analytical expression for the *MW* of FeFETs that can also take into account the effects of aging to assess endurance [59], as discussed in detail in Section 3.5. LT allows writing the $E-P$ relationship as follows [35]

$$E = 2\alpha P + 4\beta P^3 + 6\gamma P^5 + \rho\frac{\partial P}{\partial t} \tag{3.11}$$

where α, β, γ, ρ are material specific parameters. Fig. 3.11(b) illustrate the $P-E$ loop for SBT. The point P^0 corresponds to the so-called *negative capacitance* region (see the corresponding energy in Fig. 3.11(c), which can be accessed only under specific conditions (this region will be further discussed in Chapter 5). On the other hand, Fig. 3.11(d), Fig. 3.11(e) show the two stable polarization states (P^+, P^-) that correspond to energy minimum states. According to LT, since these stable states are separated by an energy barrier that can be overcome by applying $|E| > E_C$,

then hysteresis occurs.

3.4.2 MFS Structure Analysis

In this section, we discuss the fundamental aspects of the physics of FeFETs by analyzing a simplified configuration, namely the Metal-Ferroelectric-Semiconductor (MFS). This configuration is an idealized version of the MFIS one, because due to difficulties in creating a good ferroelectric/semiconductor interface (that would lead to severe reliability issues) an intermediate oxide layer is always present [7]. However, it allows identifying the main parameters that control the operation of the device and that in particular the on- and off-V_t, which in turn define the MW.

The analysis starts by writing the static Landau-Khalatnikov Equation (LKE) for the ferroelectric layer from Eq. (3.11) as follows[4]

$$V_{FE} = E \times t_{FE} = t_{FE}(2\alpha P + 4\beta P^3). \tag{3.12}$$

Since the ferroelectric layer is in series with the semiconductor body, they both share the same charge Q_s; thus, the polarization in the ferroelectric can be considered on a first approximation as $P = Q_s$ [37]. The electrostatic behavior of an FeFET can be described akin to that of a regular MOSFET by writing the Suraface Potential Equation (SPE) (at the source side) as follows [37], [38]

$$V_{GS} - V_{FB} = V_{FE} + \psi_s. \tag{3.13}$$

ψ_s depends non-linearly on the semiconductor charge Q_s. The general expression for $Q_s(\psi_s)$ valid in the accumulation, depletion and inversion regimes is written as [13, Ch. 2]

$$Q_s = \pm\sqrt{2\varepsilon_{Si}k_BTN_a}\left[\left(e^{-\frac{\psi_s}{V_{TH}}} + \frac{\psi_s}{V_{TH}} - 1\right) + \left(\frac{n_i}{N_a}\right)^2\left(e^{-\frac{V}{V_{TH}}}\left(e^{\frac{\psi_s}{V_{TH}}} - 1\right) - \frac{\psi_s}{V_{TH}}\right)\right]^{1/2} \tag{3.14}$$

where V is the quasi-Fermi potential at a point in the channel (the sign of Q_s corresponds to that of ψ_s). By coupling Eqs. (3.12) to (3.14), one can solve for Q_s, ψ_s for each $V_{GS} - V_{FB}$ value.

[4] Since we focus on the static analysis the $\rho\partial P/\partial t$ term can be neglected. Moreover, to derive simple analytical expressions in the following we will consider $\gamma = 0$. Nonetheless, this simplification does not cause a loss of generality because β and the higher-order terms are all positive [35].

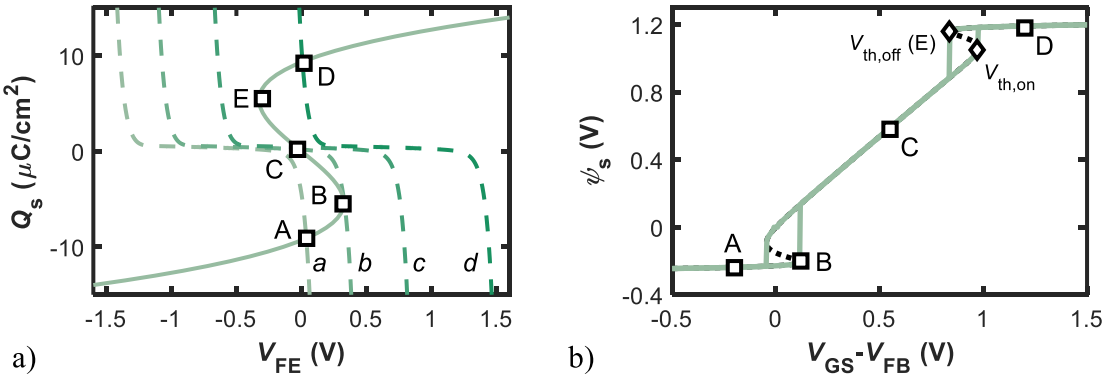


Figure 3.12: (a) $Q_s - V_{FE}$ curves for four different $V_{GS} - V_{FB}$ values ($\{-0.2; 0.12; 0.55; 1.2\}$ V, dashed lines *a-d*). For each $V_{GS} - V_{FB}$ the intersection point with the S-shaped Landau curve (solid line) represents the operating point (shown as the squared symbols A-E). (b) $\psi_s - (V_{GS} - V_{FB})$ curve showing the off-to-on ($V_{t,on}$) and on-to-off ($V_{t,off}$) thresholds and the ψ_s corresponding to the operating points in (a). The dashed line represents the region in the S-shaped curve where $\partial Q_s / \partial V_{FE} < 0$ (i.e., the so-called *negative capacitance* region). Because this region is unstable, the device switches from one stable state to the other giving rise to hysteresis. Adapted from [37].

The solution of Eqs. (3.12) to (3.14) for a MFS configuration with SBT as the ferroelectric layer is shown in Fig. 3.12 (the simulation parameters are given in Table 3.2). In Fig. 3.12(a) the $Q_s - V_{FE}$ curves (calculated with Eq. (3.12) with $P = Q_s$) are obtained for four different $V_{GS} - V_{FB}$ values (see caption) to highlight the different operating points of the device. $Q_s > 0$ ($Q_s < 0$) corresponds to the inversion (accumulation) charge; the region where the slope of $Q_s - V_{FE}$ is small is the depletion region. For negative $V_{GS} - V_{FB}$ values, the device is in the accumulation region (corresponding to the A and B operating points in Fig. 3.12). Increasing further V_{GS} causes the device to enter the depletion region (point C) intersecting the *negative capacitance* portion of the S-shaped curve. For $Q_s > 0$ the device switches abruptly from this unstable operating region to the inversion regime and stabilizes again (see for instance point D). When sweeping V_{GS} in the opposite direction, the device snaps back in the depletion region at a lower V_{GS} than before, at point E, leading to hysteresis. The corresponding $\psi_s - (V_{GS} - V_{FB})$ curve is shown in Fig. 3.12(b) along with the operating points A-E. The on- and off- threshold ($V_{t,on}$ and $V_{t,off}$) for an n-channel are also indicated by the diamond symbols (note that point E corresponds to the condition at which $V_{t,off}$ is obtained, as explained in the following).

The next step is to compute the drain current (I_D) to visualize the hysteretic behavior of the

Table 3.2: Parameters used to generate Fig. 3.12 and Fig. 3.13.

Symbol	**Value**
L	$1\,\mu m$
W	$1\,\mu m$
V_{FB}	$-0.8\,V$
N_A	$1\times10^{18}\,cm^{-3}$
μ_n	$200\,cm^2/Vs$
V_{DS}	$0.28\,V$
t_{FE}	$70\,nm$
α	$-6.5\times10^{7}\,m/F$
β	$3.75\times10^{9}\,m^5/F/C^2$

FeFET. I_D can be determined by calculating the Pao-Sah double-integral [37]

$$I_D = \mu_{eff}\frac{W}{L}\int_0^{V_D} -Q_i(V)dV = q\mu_{eff}\frac{W}{L}\int_0^{V_{DS}}\left(\int_{\varepsilon}^{\psi_s}\frac{\frac{n_i^2}{N_a}e^{\frac{(\psi-V)}{V_{TH}}}}{\mathscr{E}(\psi,V)}d\psi\right)dV \tag{3.15}$$

where $\mathscr{E}(\psi,V)$ is the electric field in the channel[5], and $\varepsilon << kT$ is the bulk potential[6]. The solution of Eq. (3.15) obtained for the forward and backward sweep is shown in Fig. 3.13.

We are now ready to present the analytical expressions for $V_{t,on}$, $V_{t,off}$, and MW. The full derivation is here omitted for brevity and can be found in [37]. The main assumption in this derivation is that the switching conditions can be ascribed to inversion charge only, thus simplifying Eq. (3.14) as follows

$$Q_s \approx Q_i = \sqrt{\frac{2\varepsilon_{Si}k_BTn_i^2}{N_A}}e^{\psi_s/2V_{TH}} = Q_0e^{\psi_s/2V_{TH}}. \tag{3.17}$$

[6] $\mathscr{E}(\psi,V)$ is a two-dimensional vector that depends on the potential from the surface to the bulk of the semiconductor (i.e., the y-direction) and from source to drain (i.e., the x-direction). Its expression can be obtained from Gauss's law (applying the Gradual Channel Approximation [13, Ch. 2])

$$\mathscr{E}(y) = -\frac{Q}{\varepsilon_{Si}} \tag{3.16}$$

where Q is the charge at a position $x > 0$ in the channel (where $x = 0$ is the source-end of the channel) and is obtained from Eq. (3.14) by substituting ψ_s with ψ to account for the spatial variation of the potential in the y-direction.

[6] $\varepsilon > 0$ is required for numerical reasons to avoid divergence of the inner integral [13, Ch. 2].

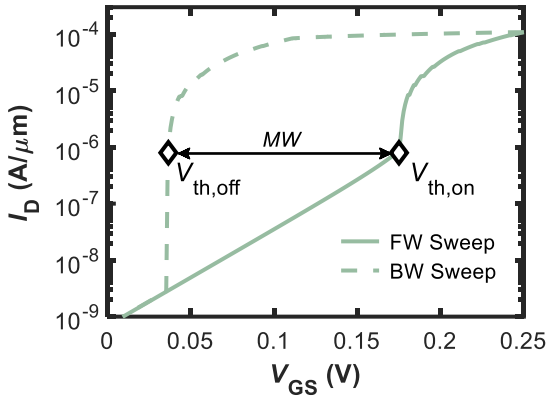


Figure 3.13: $I-V$ curve of the FeFET when performing a forward (solid line) and a backward (dashed line) sweep. The points corresponding to $V_{t,on}$ and $V_{t,off}$ are highlighted. MW is also shown, defined as $V_{t,on}-V_{t,off}$. Adapted from [37].

As clarified by Fig. 3.12, $V_{t,on}$ corresponds to the gate bias where the device exits the unstable region (where the ferroelectric capacitance is negative) and enters the stable inversion region. Its expression thus reads

$$V_{t,on} = V_{FB} + V_{FE,on} + \psi_{s,on} = V_{FB} - 2V_{TH} + 2V_{TH} \ln\left(-\frac{V_{TH}}{\alpha t_{FE} Q_0}\right). \tag{3.18}$$

$V_{t,off}$ can be identified in Fig. 3.12 as the gate voltage at which the device sweeps back from inversion to the unstable region of operation, i.e., point 'E'. This point corresponds to the condition $\partial Q_s/\partial V_{FE} = 0$ (with $Q_s > 0$). As already discussed, because of hysteresis this point is not the same as the previous one (i.e., $V_{t,on}$) and a lower gate bias is required to switch back off the FeFET. We thus find Q_s at the switching point as

$$\frac{\partial Q_s}{\partial V_{FE}} = 0 \Leftrightarrow Q_s = Q_{sw} = \sqrt{\frac{-\alpha}{6\beta}} \tag{3.19}$$

and the corresponding switching voltage, V_{sw}

$$V_{sw} = \frac{4}{3}\alpha t_{FE} Q_{sw}. \tag{3.20}$$

$V_{t,off}$ can thus be written as

$$V_{t,off} = V_{FB} + V_{sw} + 2V_{TH} \ln\left(\frac{Q_{sw}}{Q_0}\right) \tag{3.21}$$

As shown in Fig. 3.13, the MW is simply defined as $MW \equiv V_{t,on} - V_{t,off}$; thus, we obtain

$$MW = -(2V_{TH} + V_{sw}) + 2V_{TH} \ln\left(-\frac{V_{TH}}{\alpha t_{FE} Q_{sw}}\right) = -(2V_{TH} + V_{sw}) + 2V_T \ln\left(-\frac{4}{3}\frac{V_{TH}}{V_{sw}}\right). \tag{3.22}$$

The first (second) term in the MW expression is the difference between V_{FE} (ψ_s) at on- and off-threshold conditions. Interestingly, MW does not depend on substrate doping N_A nor on V_{FB}: these terms in fact cancel out because they are present both in $V_{t,on}$ and $V_{t,off}$ expressions.

In Section 3.5 we will revise and generalize the analytical expressions derived here to account for: *i)* the insulator layer effects in the MFIS configuration; *ii)* the dielectric response of the ferroelectric layer; and *iii)* the interface and oxide traps generated during prolonged operation.

3.5 Contribution: Endurance Limits Evaluation in FeFETs with Memory Window Analytic Expression

As mentioned earlier, the discovery of ferroelectricity in binary oxides such as HfO_2 and ZrO_2 constituted a major breakthrough in FeFET development thus reviving the interest of both industrial and academic communities in the technology[7], [33], [39], [40]. As highlighted by Table 3.1, FeFETs offer a wide range of improvements in terms of nonvolatility, scaling potential, read-write speed, and power dissipation with respect to either SRAM, DRAM, and Flash memory [39]. However, this technology still presents a number of issues at device level that limit its retention, and - mostly - endurance. These involve [7]

- providing a clean interface to the semiconductor in the MFS structure,
- avoiding a large depolarization field to realize ten years of data retention since the large dielectric permittivity (up to hundreds for perovskite ferroelectrics) in series with the depletion layer (and also with the interface layer in MFIS structure) of the semiconductor is unfavorable, and
- realizing a high endurance in a one-transistor-based memory array.

While HfO_2-based FeFETs have lessened the issue about retention thanks to their lower dielectric permittivity and depolarization over coercive field ratio compared to their perovskite-based counterparts [45] endurance still represents a major issue. The high coercive field of ferroelectric HfO_2 in this case does not help, as higher E_C requires pulses of high magnitude (3-4 V) to program the memory, exacerbating the reliability issues [7].

Recent reports showed that endurance typical range is $\sim 10^4 - 10^6$ writing cycles [44], [46]; this is far from meeting the International Roadmap for Devices and Systems (IRDS) requirements of 10^{12} cycles [7, 0, Ch. 1]. Nonetheless, the interest in a ultra-scaled CMOS-compatible FeFET has not faded out and potential innovative applications based on this technologies have already been demonstrated, such as: *i)* LIM circuits [48]; *ii)* artificial neural networks (ANNs) [47], [49]; and *iii)* Ternary Content Addressable Memories (TCAMs) [47], [48].

Thus, the development of characterization tools able to quantify and identify the limiting factors to endurance of FeFETs would facilitate and speed up their development [31]. Here, we develop a simple characterization tool to evaluate endurance based on an analytical expression of MW. The MW expression allows quantifying the impact of oxide and interface traps generated over time during writing cycles. In addition, from the MW expression it is possible to estimate the generated trap concentration without the need for numerical TCAD simulations. We find also that, under specific assumptions regarding the program/erase pulse amplitude and duration, endurance is weakly affected by the writing conditions.

3.5.1 Derivation of the Analytical Model

The MW is expressed as the difference between the on- and off-threshold voltage ($V_{t,on}$, and $V_{t,off}$) that correspond to the right and left path followed by the FeFET $I-V$ characteristics, respectively, see Fig. 3.13. These paths differ because of polarization switching and represent the logic binary states "0" and "1" of the memory. The expressions for $V_{t,on}$, and $V_{t,off}$ are the generalized version of the ones discussed previously in Section 3.4 that apply to the MFIS configuration illustrated in Fig. 3.11(a). To describe the electrostatic behavior of the MFIS structure requires modifying Eq. (3.13) to account for the interface oxide layer as follows

$$V_{GS} - V_{FB} = V_{FE} + V_{ox} + \psi_s = V_{ins} + \psi_s \tag{3.23}$$

V_{ins} includes the contribution from the ferroelectric and oxide interlayer and can be expressed as follows

$$V_{ins} = Q_s \left(\frac{1}{C_{FE}} + \frac{1}{C_{ox}} \right), \tag{3.24}$$

where $C_{ox} = \varepsilon_{ox}/t_{ox}$ is the oxide capacitance (SiO_2 here) and C_{FE} is the ferroelectric capacitance.

The latter can be written according to LT theory as follows

$$C_{FE} = \frac{1}{t_{FE}(2\alpha + 12\beta Q_s^2)} + \frac{\varepsilon_{FE}}{t_{FE}} \tag{3.25}$$

ε_{FE} is the parameter accounting for the dielectric response of the ferroelectric layer [50], [51]. [7] The two terms in Eq. (3.25) reflect the contributions to the displacement of the electric field (D) obtained from the spontaneous polarization and the applied electric field [50], i.e., $D = \varepsilon_{FE}E + P$. The closed-form expressions for $V_{t,on}$, $V_{t,off}$ and MW read

$$V_{t,on} = V_{FB} + 2V_{TH}\ln\left(\frac{2V_{TH}}{|a|Q_0}\right) - 2V_{TH} \tag{3.27a}$$

$$V_{t,off} = V_{FB} + 2V_{TH}\ln\left(\frac{Q_{sw}}{Q_0}\right) - V_{sw} \tag{3.27b}$$

$$MW = 2V_{TH}\ln\left(\frac{2V_{TH}}{|a|Q_{sw}}\right) + (V_{sw} - 2V_{TH}). \tag{3.27c}$$

The expression for Q_{sw} and V_{sw} can be derived by solving a quadratic expression (as discussed in [59]). However, simplified expressions can be written by considering $\varepsilon_{FE} = 0$ in Eq. (3.25) as follows

$$Q_{sw} = \sqrt{\frac{|a|}{3b}} \tag{3.28a}$$

$$V_{sw} = -\left(aQ_{sw} + bQ_{sw}^3\right) = \frac{2}{3}|a|Q_{sw}. \tag{3.28b}$$

The main difference of these expressions with respect to the ones of the MFS structure is encapsulated in the a and b parameters, i.e.,

$$a \equiv 2\alpha t_{FE} + \frac{1}{C_{ox}}, \quad b \equiv 12\beta t_{FE} \tag{3.29}$$

that account for the voltage divider between the ferroelectric and oxide layer. Basically for a

[7] Because C_{FE} as written in Eq. (3.25) is a *differential* capacitance (i.e., $C \sim (\partial V/\partial Q)^{-1}$), in general V_{ins} should be written as

$$V_{ins} = \int_{-\infty}^{Q_s}\left(\frac{1}{C_{FE}} + \frac{1}{C_{ox}}\right)dQ = \int_{-\infty}^{Q_s}\frac{1}{C_{FE}}dQ + \frac{Q_s}{C_{ox}} \tag{3.26a}$$

$$= \frac{t_{FE}}{\varepsilon_{FE}}\left[Q_s - \frac{1}{1+2\alpha\varepsilon_{FE}}\sqrt{\frac{12\beta\varepsilon_{FE}}{1+2\alpha\varepsilon_{FE}}}^{-1}\tan^{-1}\left(Q_s\sqrt{\frac{12\beta\varepsilon_{FE}}{1+2\alpha\varepsilon_{FE}}}\right)\right] + \frac{Q_s}{C_{ox}}. \tag{3.26b}$$

Here, for the sake of simplicity and to avoid complex expressions, we assume C_{FE} to be the integral capacitance of the ferroelectric. However, more extensive analysis to check the validity of this approximation and to compare with the results of the accurate calculation needs to be done and will be subject of future investigation.

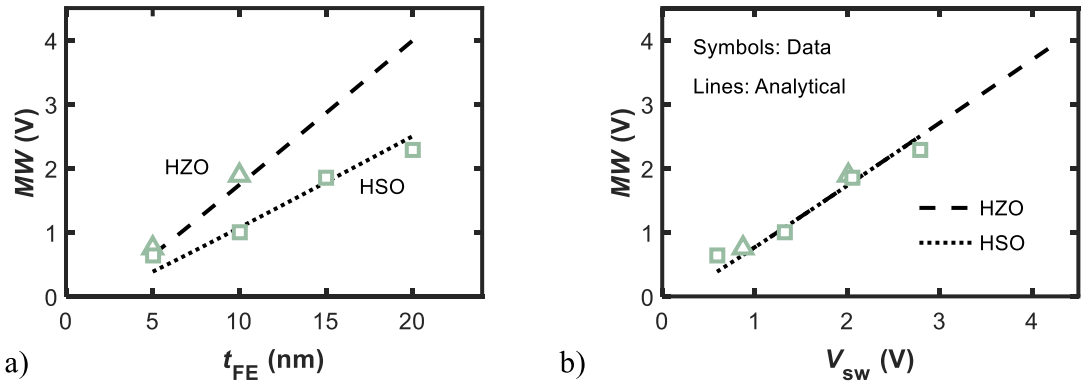


Figure 3.14: Comparison of MW calculated with (3.27c) and experimental data from [40] for Zr- and Si-doped HfO_2-based FeFETs plotted vs (a) t_{FE} and (b) switching voltage, V_{sw}. Adapted from [58].

given t_{FE}, MW reduces when $t_{ox} > 0$ and the fact that $\varepsilon_{ox} << \varepsilon_{FE}$ increases the field on the oxide layer leading to reliability issues limiting endurance [43]. Note that since $V_{sw} \propto t_{FE}$, and $MW \sim V_{sw} \propto t_{FE}$, Eq. (3.27c) correctly anticipates the theoretical (and experimentally observed) linear MW-t_{FE} relationship [34], [40], [58]. This result is shown in Fig. 3.14, where MW is plotted vs both t_{FE} and V_{sw} for Zr- and Si-doped HfO_2-based FeFETs [40]. Eq. (3.27a)-Eq. (3.29) suggest that MW reduces with increasing t_{ox}, as a consequence of reduced a magnitude. However, recent experimental results seem to indicate the opposite behavior as documented for instance in [41]. This behavior could be explained by the decrease of compensation of the polarization charge by traps at the ferroelectric/dielectric interface due to reduced leakage with increasing t_{ox} [42]. Further investigation is required on this aspect to model consistently the behavior of MW vs t_{ox} and compensating charges in the whole MFIS structure.

3.5.2 Modeling of traps effects

We now discuss how to include the effects related to charge trapping and trap generation limiting endurance of FeFETs. Here we define endurance as the time (or total number of cycles) taken during repeated program/erase operations before the "0" and "1" states of the memory become indistinguishable (i.e., $MW \approx 0$). The main limiting factor to FeFET endurance is the trap generation in the oxide layer between the ferroelectric and the semiconductor body [43]. Other

limiting factors are related to ferroelectric aging that might lead to additional V_t's shifts, premature breakdown due to formation of percolation paths [52], [53] and remnant polarization degradation [43]. Here we focus only on the former limiting factor and neglect the latter ones. Moreover, we do not take into account the fast MW decay due to depolarization fields and trapping/detrapping because it is expected to mainly influence retention [43], [45].

During repeated cycling tests that stress the devices to probe endurance, the high electric field induced in the gate stack by the program/erase pulses accelerates trap generation. The electric field mainly concentrates on the oxide layer rather than in the ferroelectric because of its lower dielectric constant [54] thus, to a first approximation, generation can be assumed to occur only at the oxide/semiconductor interface and in the oxide itself. The effect of generated defects is modeled by adding to Eq. (3.23) the following contributions [55]

$$V_{ot} \equiv -\frac{q\Delta N_{ot}}{C_{eq}} \quad V_{it} \equiv \frac{q\Delta D_{it}}{C_{eq}}(\psi_s - \phi_b), \tag{3.30}$$

where ΔN_{ot} is the generated trap concentration in the oxide interface layer (cm^{-2}), ΔD_{it} is the generated interface trap density of states ($\mathrm{cm}^{-2}\,\mathrm{eV}^{-1}$), and ϕ_b is the body potential ($C_{eq}^{-1} = C_{FE}^{-1} + C_{ox}^{-1}$). We assume that the charge neutrality level for the interface traps is located at Si mid-gap [55]. Stress is induced by positive and negative pulses applied on the gate to erase and program the FeFET, respectively. Hence, $V_{t,on}$ tends to decrease and $V_{t,off}$ to increase, as observed in [56]. The concentration of generated defects during writing of the memory is in general different depending on the sign of the writing pulse, therefore the shifts in $V_{t,on}$ and $V_{t,off}$ are not symmetric. This is reflected in the different symbols used to indicate the generated defects during program and erase cycles, namely, $\Delta N_{ot,P/E}$ and $\Delta D_{it,P/E}$ for oxide and interface traps, respectively.

The expressions for $\Delta V_{t,on}$, $\Delta V_{t,off}$ can be easily obtained by combining Eq. (3.23) and Eq. (3.30) as follows

$$V_{GS} - V_{FB} = V_{ins} - \frac{q\Delta N_{ot,P/E}}{C_{eq}} + \frac{q\Delta D_{it,P/E}}{C_{eq}}(\psi_s - \phi_b) + \psi_s. \tag{3.31}$$

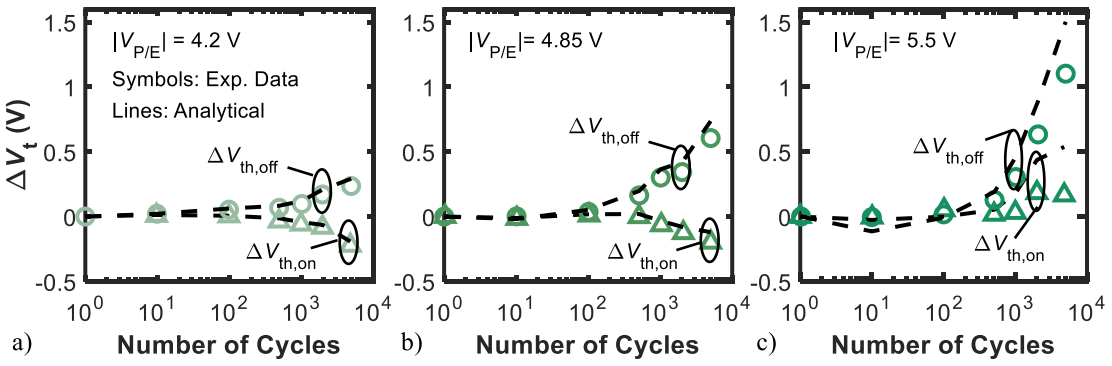


Figure 3.15: Calculated (dashed lines) and experimental data from [56] (symbols) $\Delta V_{t,on}$ and $\Delta V_{t,off}$ vs program/erase cycles. The different panels show different program/erase pulse amplitude, $|V_{P/E}|$: (a) 4.2 V, (b) 4.85 V, and (c) 5.5 V, respectively. Adapted from [59].

The final expressions read

$$\Delta V_{t,on} = 2V_{TH}\ln\left(1+\frac{q\Delta D_{it,P}}{C_{eq}}\right)\times\left(1+\frac{q\Delta D_{it,P}}{C_{eq}}\right) - \frac{q}{C_{eq}}\times\left\{\Delta N_{ot,P}-\Delta D_{it,P}\left[2V_t\ln\left(\frac{2V_{TH}}{|a|Q_0}\right)-2V_{TH}-\phi_b\right]\right\} \tag{3.32a}$$

$$\Delta V_{t,off} = -\frac{q}{C_{eq}}\left\{\Delta N_{ot,E}-\Delta D_{it,E}\left[2V_{TH}\ln\left(\frac{Q_{sw}}{Q_0}\right)-\phi_b\right]\right\} \tag{3.32b}$$

$$\Delta MW = 2V_{TH}\ln\left(1+\frac{q\Delta D_{it,P}}{C_{eq}}\right)\times\left(1+\frac{q\Delta D_{it,P}}{C_{eq}}\right) - \frac{q}{C_{eq}}\left\{(\Delta N_{ot,P}-\Delta N_{ot,E})-2V_t\Delta D_{it,P}\left[\ln\left(\frac{2V_{TH}}{|a|Q_0}\right)-1\right] + 2V_{TH}\Delta D_{it,E}\ln\left(\frac{Q_{sw}}{Q_0}\right)+(\Delta D_{it,P}-\Delta D_{it,E})\phi_b\right\}. \tag{3.32c}$$

3.5.3 Results

To assess the accuracy of the above expressions, we compared the results obtained with Eqs. (3.32a) and (3.32b) with experimental data of endurance tests from [56]. The results in terms of $\Delta V_{t,on}$ and $\Delta V_{t,off}$ for three different values of program/erase pulse amplitude ($|V_{P/E}|$, see legend) are shown in Fig. 3.15. The parameters $\alpha = -3\times10^{9}\,\mathrm{m/F}$, $\beta = 2\times10^{11}\,\mathrm{m^5/F/C^2}$, $\varepsilon_{FE} = 8\varepsilon_0$ were set to match the experimental data trends. $t_{FE} = 10\,\mathrm{nm}$ and $t_{ox} = 3\,\mathrm{nm}$ [56]. The duration of both program and erase pulse for each $|V_{P/E}|$ is $t_{P/E} = 100\,\mathrm{ns}$, thus the time it takes for a single writing

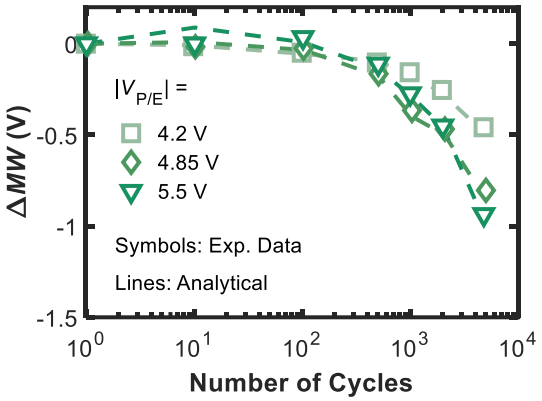


Figure 3.16: Calculated (dashed lines) and experimental data from [56] (symbols) ΔMW vs program/erase cycles for different $|V_{P/E}|$ (see legend). Adapted from [59].

cycle is $t_{cycle} = 200\,\text{ns}$ [56]. The variation in $V_{t,on}$ and $V_{t,off}$ in turn affects the MW, as shown in Fig. 3.16 for the same $|V_{P/E}|$ values of Fig. 3.15.

Because degradation primarily occurs in the insulator layer, the trend of the degraded V_{th}'s and MW is fully captured by V_{ot} and V_{it} only. From this observation, simplified $\Delta V_{t,on}$, $\Delta V_{t,off}$ and ΔMW expressions can be obtained as follows (i.e., by neglecting V_{ins} variations)

$$\Delta V'_{t,on} \sim \frac{-q}{C_{eq}} \left\{ \Delta N_{ot,P} - \Delta D_{it,P} \left[2V_{TH} \ln\left(\frac{2V_{TH}}{|a|Q_0} \right) - \phi_b \right] \right\} \tag{3.33a}$$

$$\Delta V'_{t,off} \sim \frac{-q}{C_{eq}} \left\{ \Delta N_{ot,E} - \Delta D_{it,E} \left[2V_{TH} \ln\left(\frac{Q_{sw}}{Q_0} \right) - \phi_b \right] \right\} \tag{3.33b}$$

$$\begin{aligned} \Delta MW' \sim \frac{-q}{C_{eq}} \Bigg\{ & (\Delta N_{ot,P} - \Delta N_{ot,E}) - \Delta D_{it,P} \left[2V_t \ln\left(\frac{2V_{TH}}{|a|Q_0} \right) \right] \\ & + \Delta D_{it,E} \left[2V_{TH} \ln\left(\frac{Q_{sw}}{Q_0} \right) \right] + (\Delta D_{it,P} - \Delta D_{it,E})\, \phi_b \Bigg\}. \end{aligned} \tag{3.33c}$$

Note that $\Delta V'_{t,on}$, $\Delta V'_{t,off}$, and $\Delta MW'$ are proportional to the variation introduced by the generation of both oxide and interface defects. The surface potential ψ_s [corresponding to the logarithmic terms in square brackets in Eq. (3.33a)-Eq. (3.33b)] is calculated differently according to the two threshold conditions. As intuition suggests, if the degradation were symmetric, i.e., the generated defects were giving equal and opposite V_{ot} and V_{it}, the MW variation would be $\sim -2q/C_{eq} \times [\Delta N_{ot} - \Delta D_{it}\,(\psi_s - \phi_b)]$.

The good agreement between analytical and experimental results in Figs. 3.15 and 3.16

was obtained by extracting the generated oxide and interface trap concentrations from $\Delta V_{t,on}$ and $\Delta V_{t,off}$ data in [56] following the approach described in [57]. That is, N_{ot} and D_{it} were extracted by separating the threshold voltage shifts due to oxide (ΔV_{mg}) and interface traps (ΔV_{it}) separately. The former is obtained from the mid-gap voltage, V_{mg}, that correlates with N_{ot}-induced V_t drifts as at $V_G = V_{mg} \Rightarrow \psi_s = \phi_b$ and $\Delta V_{it} = 0$, see Eq. (3.31); the latter is obtained by letting $\Delta V_{it} = \Delta V_{th} - \Delta V_{ot}$ [56], [57].

To summarize, Eq. (3.32a)-Eq. (3.32c) directly connect the FeFET parameters to the stress-dependent oxide and interface trap generation. As such, Eq. (3.32c) represents the proposed MW-based characterization tool for extracting oxide and interface defects. This could serve either as an alternative to traditional techniques, or as a stand-alone method to characterize defect densities under a variety of stress conditions. For instance, notice that when only N_{ot} generation affects MW degradation then it is possible to estimate the *net* generated traps from Eq. (3.33c) asd follows

$$\Delta N_{ot,net} \equiv \Delta N_{ot,P} - \Delta N_{ot,E} \approx -\Delta MW' \frac{C_{eq}}{q} \tag{3.34}$$

This expression allows to simply and directly correlate MW measurements with generated traps.

Writing-Conditions Agnostic Endurance

In the following we show that the endurance extrapolated from the analytical equations — under specific conditions — is weakly influenced by the writing conditions (in terms of $|V_{P/E}|$ and $t_{P/E}$). With the N_{ot} and D_{it} data extracted from Fig. 3.16, it is possible to extrapolate the generated trap concentration for an arbitrary number of writing cycles. For simplicity and clarity of presentation, we assume that the MW degradation is induced by oxide traps only [56] and neglect the generation of interface traps. The generated oxide trap density, N_{ot} is shown in Fig. 3.17(a), Fig. 3.17(b) for both program and erase operation that set $V_{t,on}$ and $V_{t,off}$, respectively. By fitting the experimental data in Fig. 3.17 it is found that generated N_{ot} follows a power law with respect to writing time (with $t_{cycle} = 200\,\text{ns}$), i.e.,

$$\Delta N_{ot} \sim N_0 \times (t_{cycle})^{\beta_s}. \tag{3.35}$$

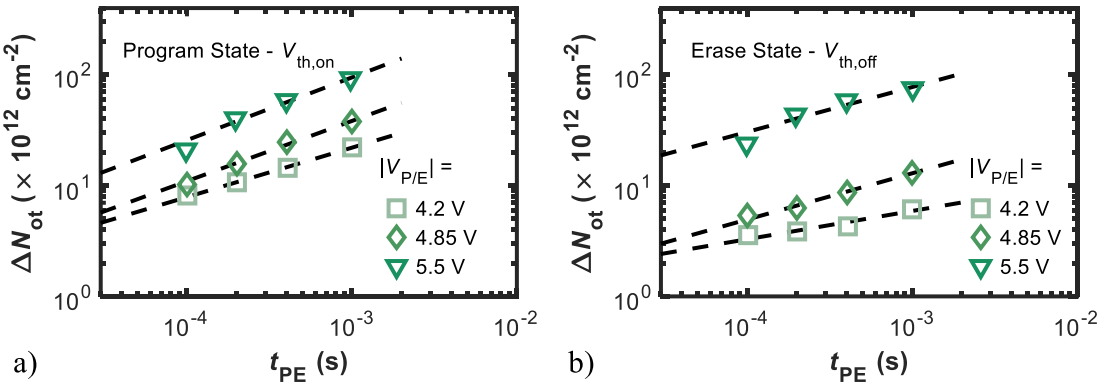


Figure 3.17: Generated oxide traps, ΔN_{ot}, vs program/erase time for different $|V_{P/E}|$ (see legend) determining (a) $V_{t,on}$ and (b) $V_{t,off}$ degradation. Black dashed lines are the fitting of experimental data from [56] (symbols) with Eq. (3.35). Adapted from [59].

Table 3.3: Coefficients of the power law in Eq. (3.35).

	Program ($V_{th,on}$)		**Erase ($V_{th,off}$)**	
$V_{P/E}$ (V)	N_0 (cm^{-2})	β_s	N_0 (cm^{-2})	β_s
4.2	9.6×10^{13}	0.45	4.6×10^{12}	0.25
4.85	3.28×10^{14}	0.54	3.1×10^{13}	0.41
5.5	9.5×10^{14}	0.54	1.7×10^{14}	0.41

The values of N_0 and β_s coefficients are collected in Table 3.3 for different writing conditions. The extrapolated MW degradation obtained by using the predicted ΔN_{ot} from the generation model is shown in Fig. 3.18(a), (b) for different $|V_{P/E}|$ and $t_{P/E}$ values, respectively. Note that MW values are normalized to the respective initial value for a fair comparison with different writing conditions. For the sake of argument, the arbitrary threshold set as the 20% of the initial MW is chosen to evaluate the endurance, see Fig. 3.18. Interestingly, notice from Fig. 3.18(a) that $|V_{P/E}|$ increment does not degrade endurance significantly (at least for the range of values as in [56]). This is because higher $|V_{P/E}|$ leads to higher initial MW [40] but also higher ΔN_{ot}, see Fig. 3.17. Similarly, Fig. 3.18(b) shows that increasing the pulse duration negligibly influences endurance. Note that in this case it was assumed that $t_{P/E}$ increase leads to the same increase in MW and initial N_{ot} to that caused by $|V_{P/E}|$. This was done for the specific purpose of illustrating

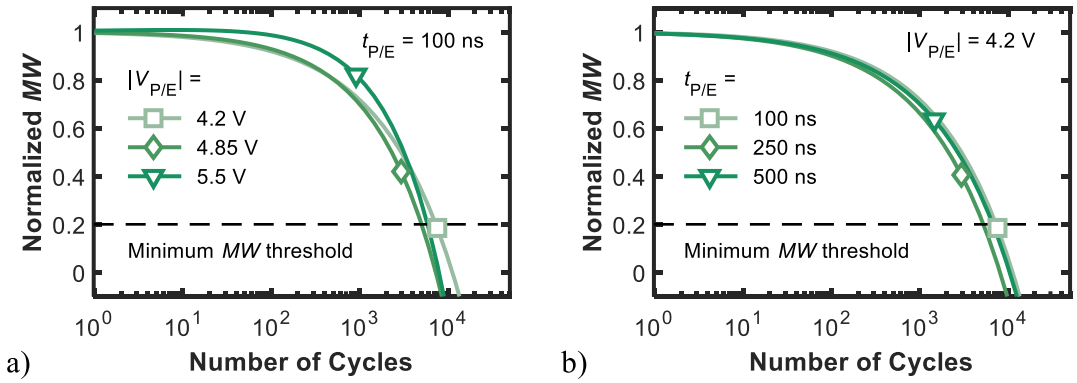


Figure 3.18: Normalized MW degradation calculated from $\Delta MW'$ with only the contribution of ΔN_{ot} extrapolated from Fig. 3.17. (a) and (b) show the dependence for different $|V_{P/E}|$ and $t_{P/E}$ values, respectively. An arbitrary minimum MW threshold defines endurance. Adapted from [59].

that if both MW and initial N_{ot} increase with program conditions, then the combined effect leads to negligible variation in endurance. However, if the assumption regarding MW and N_{ot} increase with $|V_{P/E}|$ (or $t_{P/E}$) is not satisfied, then the endurance limit will be affected by the writing conditions.

3.5.4 Conclusions

We evaluated the endurance of FeFETs by using an analytical expression of the Memory Window, MW, for the conventional MFIS structure. The MW expression takes into account the contribution from generated interface and oxide traps and was validated against experimental data. We find that:

- MW can be used to extract oxide and interface traps being generated during endurance tests, see Eq. (3.34);
- the generation trend follows a power-law with time exponent $\sim 0.3 - 0.5$, see Eq. (3.35); and
- under specific assumptions, the endurance limit is essentially independent of writing conditions, see Fig. 3.18.

These considerations drawn from the simple analytical formulation can be helpful to develop

next-generation FeFET with improved endurance.

References – Chapter 3

[1] S. Salahuddin, K. Ni, and S. Datta, "The era of hyper-scaling in electronics," *Nature Electronics*, vol. 1, no. 8, pp. 442–450, Aug. 2018. DOI: 10.1038/s41928-018-0117-x.

[2] D. Ielmini and H. S. Wong, "In-memory computing with resistive switching devices," *Nature Electronics*, vol. 1, no. 6, pp. 333–343, 2018. DOI: 10.1038/s41928-018-0092-2.

[3] W. A. Wulf and S. A. McKee, "Hitting the memory wall," *ACM SIGARCH Computer Architecture News*, vol. 23, no. 1, pp. 20–24, 1995. DOI: 10.1145/216585.216588.

[4] H. Noguchi, K. Ikegami, K. Kushida, K. Abe, S. Itai, S. Takaya, N. Shimomura, J. Ito, A. Kawasumi, H. Hara, and S. Fujita, "A 3.3ns-access-time 71.2μW/MHz 1Mb embedded STT-MRAM using physically eliminated read-disturb scheme and normally-off memory architecture," in *IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 58, IEEE, Feb. 2015, pp. 136–137. DOI: 10.1109/ISSCC.2015.7062963.

[5] Q. Dong, Z. Wang, J. Lim, Y. Zhang, Y. C. Shih, Y. D. Chih, J. Chang, D. Blaauw, and D. Sylvester, "A 1Mb 28nm STT-MRAM with 2.8ns read access time at 1.2V VDD using single-cap offset-cancelled sense amplifier and in-situ self-write-termination," in *IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 61, 2018, pp. 480–482. DOI: 10.1109/ISSCC.2018.8310393.

[6] D. Takashima, "Overview of FeRAMs: Trends and perspectives," in *Non-Volatile Memory Technology Symposium (NVMTS)*, IEEE, 2011, pp. 36–41, ISBN: 9781457714313. DOI: 10.1109/NVMTS.2011.6137107.

[7] T. Mikolajick, U. Schroeder, and S. Slesazeck, "The Past, the Present, and the Future of Ferroelectric Memories," *IEEE Transactions on Electron Devices*, vol. 67, no. 4, pp. 1434–1443, 2020. DOI: 10.1109/TED.2020.2976148.

[8] R. Annunziata, P. Zuliani, M. Borghi, G. De Sandre, L. Scotti, C. Prelini, M. Tosi, I. Tortorelli, and F. Pellizzer, "Phase change memory technology for embedded non volatile memory applications for 90nm and beyond," in *IEEE International Electron Devices Meeting (IEDM)*, IEEE, 2009, pp. 97–100. DOI: 10.1109/IEDM.2009.5424413.

[9] L. Vandelli, A. Padovani, G. Bersuker, D. Gilmer, P. Pavan, and L. Larcher, "Modeling of the forming operation in HfO2-based resistive switching memories," in *IEEE International Memory Workshop (IMW)*, IEEE, May 2011, pp. 6–9. DOI: 10.1109/IMW.2011.5873224.

[10] L. Vandelli, A. Padovani, L. Larcher, G. Broglia, G. Ori, M. Montorsi, G. Bersuker, and P. Pavan, "Comprehensive physical modeling of forming and switching operations in HfO 2 RRAM devices," in *IEEE International Electron Devices Meeting (IEDM)*, IEEE, 2011, pp. 421–424. DOI: 10.1109/IEDM.2011.6131574.

[11] G. Bersuker, D. C. Gilmer, D. Veksler, P. Kirsch, L. Vandelli, A. Padovani, L. Larcher, K. McKenna, A. Shluger, V. Iglesias, M. Porti, and M. Nafría, "Metal oxide resistive memory switching mechanism based on conductive filament properties," *Journal of Applied Physics*, vol. 110, no. 12, 2011. DOI: 10.1063/1.3671565.

[12] F. M. Puglisi, L. Larcher, G. Bersuker, A. Padovani, and P. Pavan, "An empirical model for RRAM resistance in low-and high-resistance states," *IEEE Electron Device Letters*, vol. 34, no. 3, pp. 387–389, 2013. DOI: 10.1109/LED.2013.2238883.

[13] F. M. Puglisi, P. Pavan, A. Padovani, and L. Larcher, "A compact model of hafnium-oxide-based resistive random access memory," in *ICICDT 2013 - International Conference on IC Design and Technology, Proceedings*, IEEE, 2013, pp. 85–88. DOI: 10.1109/ICICDT.2013.6563309.

[14] T. Zanotti, F. M. Puglisi, and P. Pavan, "Multi-input Logic-in-Memory for Ultra-Low Power non-von Neumann Computing," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. Under Rev. Pp. 1–12, 2021.

[15] S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy, and D. Ielmini, "Understanding switching variability and random telegraph noise in resistive RAM," in *IEEE International Electron Devices Meeting (IEDM)*, 2013, pp. 782–785. DOI: 10.1109/IEDM.2013.6724732.

[16] N. Raghavan, R. Degraeve, A. Fantini, L. Goux, S. Strangio, B. Govoreanu, D. J. Wouters, G. Groeseneken, and M. Jurczak, "Microscopic origin of random telegraph noise fluctuations in aggressively scaled RRAM and its impact on read disturb variability," in *IEEE International Reliability Physics Symposium Proceedings*, IEEE, 2013, pp. 10–16. DOI: 10.1109/IRPS.2013.6532042.

[17] D. Veksler, G. Bersuker, L. Vandelli, A. Padovani, L. Larcher, A. Muraviev, B. Chakrabarti, E. Vogel, D. C. Gilmer, and P. D. Kirsch, "Random telegraph noise (RTN) in scaled RRAM devices," in *IEEE International Reliability Physics Symposium (IRPS)*, 2013, pp. 15–18. DOI: 10.1109/IRPS.2013.6532101.

[18] F. M. Puglisi, L. Larcher, P. Pavan, A. Padovani, and G. Bersuker, "Instability of HfO2 RRAM devices: Comparing RTN and cycling variability," in *IEEE International Reliability Physics Symposium (IRPS)*, 2014, pp. 5–9. DOI: 10.1109/IRPS.2014.6861160.

[19] M. Si, N. J. Conrad, S. Shin, J. Gu, J. Zhang, M. A. Alam, and P. D. Ye, "Low-Frequency Noise and Random Telegraph Noise on Near-Ballistic III-V MOSFETs," *IEEE Transactions on Electron Devices*, vol. 62, no. 11, pp. 3508–3515, 2015. DOI: 10.1109/TED.2015.2433921.

[20] A. Chen, "Utilizing the variability of resistive random access memory to implement reconfigurable physical unclonable functions," *IEEE Electron Device Letters*, vol. 36, no. 2, pp. 138–140, 2015. DOI: 10.1109/LED.2014.2385870.

[21] T. Figliolia, P. Julian, G. Tognetti, and A. G. Andreou, "A true Random Number Generator using RTN noise and a sigma delta converter," in *IEEE International Symposium on Circuits and Systems (ISCS)*, 2016, pp. 17–20. DOI: 10.1109/ISCAS.2016.7527159.

[22] F. M. Puglisi, L. Larcher, A. Padovani, and P. Pavan, "A complete statistical investigation of RTN in HfO2-based RRAM in high resistive state," *IEEE Transactions on Electron Devices*, vol. 62, no. 8, pp. 2606–2613, 2015. DOI: 10.1109/TED.2015.2439812.

[23] L. Vandelli, A. Padovani, L. Larcher, R. G. Southwick, W. B. Knowlton, and G. Bersuker, "A physical model of the temperature dependence of the current through SiO2/HfO2 stacks," *IEEE Transactions on Electron Devices*, vol. 58, no. 9, pp. 2878–2887, 2011. DOI: 10.1109/TED.2011.2158825.

[24] P. Y. Chen and S. Yu, "Compact Modeling of RRAM Devices and Its Applications in 1T1R and 1S1R Array Design," *IEEE Transactions on Electron Devices*, vol. 62, no. 12, pp. 4022–4028, 2015. DOI: 10.1109/TED.2015.2492421.

[25] F. M. Puglisi, P. Pavan, L. Larcher, and A. Padovani, "Statistical analysis of random telegraph noise in HfO<inf>2</inf>-based RRAM devices in LRS," *Solid-State Electronics*, vol. 113, pp. 132–137, 2015. DOI: 10.1016/j.sse.2015.05.027.

[26] A. Padovani, L. Larcher, G. Bersuker, and P. Pavan, "Charge transport and degradation in HfO2 and HfOx dielectrics," *IEEE Electron Device Letters*, vol. 34, no. 5, pp. 680–682, 2013. DOI: 10.1109/LED.2013.2251602.

[27] F. M. Puglisi, N. Zagni, L. Larcher, and P. Pavan, "A new verilog-A compact model of random telegraph noise in oxide-based RRAM for advanced circuit design," in *European Solid-State Device Research Conference (ESSDERC)*, 2017, pp. 204–207. DOI: 10.1109/ESSDERC.2017.8066627.

[28] ——, "Random Telegraph Noise in Resistive Random Access Memories: Compact Modeling and Advanced Circuit Design," *IEEE Transactions on Electron Devices*, vol. 65, no. 7, pp. 2964–2972, Jul. 2018. DOI: 10.1109/TED.2018.2833208.

[29] J. L. Moll and Y. Tarui, "A New Solid State Memory Resistor," vol. 10, no. 5, p. 338, 1963. DOI: 10.1109/T-ED.1963.15245.

[30] T. S. Böescke, J. Müller, D. Bräuhaus, U. Schröder, and U. Böttger, "Ferroelectricity in hafnium oxide: CMOS compatible ferroelectric field effect transistors," in *IEEE Interna-*

tional Electron Devices Meeting (IEDM), 2011, pp. 547–550. DOI: 10.1109/IEDM.2011.6131606.

[31] J. Müller, T. S. Böscke, U. Schröder, S. Mueller, D. Bräuhaus, U. Böttger, L. Frey, and T. Mikolajick, "Ferroelectricity in simple binary ZrO 2 and HfO 2," *Nano Letters*, vol. 12, no. 8, pp. 4318–4323, 2012. DOI: 10.1021/nl302049k.

[32] M. Trentzsch, S. Flachowsky, R. Richter, J. Paul, B. Reimer, D. Utess, S. Jansen, H. Mulaosmanovic, S. Muller, S. Slesazeck, J. Ocker, M. Noack, J. Muller, P. Polakowski, J. Schreiter, S. Beyer, T. Mikolajick, and B. Rice, "A 28nm HKMG super low power embedded NVM technology based on ferroelectric FETs," pp. 11.5.1–11.5.4, 2016. DOI: 10.1109/IEDM.2016.7838397.

[33] S. Dünkel, M. Trentzsch, R. Richter, P. Moll, C. Fuchs, O. Gehring, M. Majer, S. Wittek, B. Müller, T. Melde, H. Mulaosmanovic, S. Slesazeck, S. Müller, J. Ocker, M. Noack, D. A. Löhr, P. Polakowski, J. Müller, T. Mikolajick, J. Höntschel, B. Rice, J. Pellerin, and S. Beyer, "A FeFET based super-low-power ultra-fast embedded NVM technology for 22nm FDSOI and beyond," in *IEEE International Electron Devices Meeting (IEDM)*, Oct. 2017, pp. 19.7.1–19.7.4. DOI: 10.1109/IEDM.2017.8268425.

[34] H. T. Lue, C. J. Wu, and T. Y. Tseng, "Device modeling of ferroelectric memory field-effect transistor (FeMFET)," *IEEE Transactions on Electron Devices*, vol. 49, no. 10, pp. 1790–1798, Oct. 2002. DOI: 10.1109/TED.2002.803626.

[35] P. Chandra and P. B. Littlewood, "A Landau Primer for Ferroelectrics," in *Physics of Ferroelectrics - A Modern Perspective*, K. M. Rabe, C. H. Ahn, and J.-M. Triscone, Eds., 1st, vol. 105, New York: Springer, 2007, ch. 3, pp. 69–116.

[36] A. Jain and M. A. Alam, "Stability constraints define the minimum subthreshold swing of a negative capacitance field-effect transistor," *IEEE Transactions on Electron Devices*, vol. 61, no. 7, pp. 2235–2242, Jul. 2014. DOI: 10.1109/TED.2014.2316167.

[37] H. P. Chen, V. C. Lee, A. Ohoka, J. Xiang, and Y. Taur, "Modeling and design of ferroelectric MOSFETs," *IEEE Transactions on Electron Devices*, vol. 58, no. 8, pp. 2401–2405, Aug. 2011. DOI: 10.1109/TED.2011.2155067.

[38] G. Pahwa, T. Dutta, A. Agarwal, and Y. S. Chauhan, "Compact model for ferroelectric negative capacitance transistor with MFIS structure," *IEEE Transactions on Electron Devices*, vol. 64, no. 3, pp. 1366–1374, Mar. 2017. DOI: 10.1109/TED.2017.2654066.

[39] M. Okuyama, "Features, Principles and Development of Ferroelectric–Gate Field-Effect Transistors," in *Ferroelectric-Gate Field Effect Transistor Memories: Device Physics and Applications*, B.-E. Park, H. Ishiwara, M. Okuyama, S. Sakai, and S.-M. Yoon, Eds., 1st, Springer Netherlands, 2016, ch. 1, pp. 3–18. DOI: 10.1007/978-94-024-0841-6.

[40] T. Ali, P. Polakowski, S. Riedel, T. Büttner, T. Kämpfe, M. Rudolph, B. Pätzold, K. Seidel, D. Löhr, R. Hoffmann, M. Czernohorsky, K. Kühnel, X. Thrun, N. Hanisch, P. Steinke, J. Calvo, and J. Müller, "Silicon doped hafnium oxide (HSO) and hafnium zirconium oxide (HZO) based FeFET: A material relation to device physics," *Applied Physics Letters*, vol. 112, no. 22, p. 222 903, Jun. 2018. DOI: 10.1063/1.5029324.

[41] J. Li, Y. Qu, M. Si, X. Lyu, and P. D. Ye, "Multi-Probe Characterization of Ferroelectric/Dielectric Interface by C-V, P-V and Conductance Methods," in *VLSI Technology Symposium*, 2020, pp. 2020–2021. DOI: 10.1109/VLSITechnology18217.2020.9265069.

[42] Y. Qu, J. Li, M. Si, X. Lyu, and P. D. Ye, "Quantitative Characterization of Interface Traps in Ferroelectric/Dielectric Stack Using Conductance Method," *IEEE Transactions on Electron Devices*, vol. 67, no. 12, pp. 5315–5321, 2020. DOI: 10.1109/TED.2020.3034564.

[43] E. Yurchuk, S. Mueller, D. Martin, S. Slesazeck, U. Schroeder, T. Mikolajick, J. Muller, J. Paul, R. Hoffmann, J. Sundqvist, T. Schlosser, R. Boschke, R. Van Bentum, and M. Trentzsch, "Origin of the endurance degradation in the novel HfO2-based 1T ferroelectric non-volatile memories," in *IEEE International Reliability Physics Symposium (IRPS)*, 2014, 2E.5.1–2E.5.5. DOI: 10.1109/IRPS.2014.6860603.

[44] E. Yurchuk, J. Muller, J. Paul, T. Schlosser, D. Martin, R. Hoffmann, S. Mueller, S. Slesazeck, U. Schroeder, R. Boschke, R. Van Bentum, and T. Mikolajick, "Impact of scaling on the performance of HfO2-based ferroelectric field effect transistors," *IEEE Transactions on Electron Devices*, vol. 61, no. 11, pp. 3699–3706, Nov. 2014. DOI: 10.1109/TED.2014.2354833.

[45] N. Gong and T. P. Ma, "Why Is FE-HfO2 More Suitable Than PZT or SBT for Scaled Nonvolatile 1-T Memory Cell? A Retention Perspective," *IEEE Electron Device Letters*, vol. 37, no. 9, pp. 1123–1126, Sep. 2016. DOI: 10.1109/LED.2016.2593627.

[46] H. Mulaosmanovic, E. T. Breyer, T. Mikolajick, and S. Slesazeck, "Ferroelectric FETs With 20-nm-Thick HfO 2 Layer for Large Memory Window and High Performance," *IEEE Transactions on Electron Devices*, vol. 66, no. 9, pp. 3828–3833, Sep. 2019. DOI: 10.1109/ted.2019.2930749.

[47] A. Aziz, E. T. Breyer, A. Chen, X. Chen, S. Datta, S. K. Gupta, M. Hoffmann, X. S. Hu, A. Ionescu, M. Jerry, T. Mikolajick, H. Mulaosmanovic, K. Ni, M. Niemier, I. O'Connor, A. Saha, S. Slesazeck, S. K. Thirumala, and X. Yin, "Computing with ferroelectric FETs: Devices, models, systems, and applications," in *Design, Automation and Test in Europe Conference and Exhibition (DATE)*, Mar. 2018, pp. 1289–1298. DOI: 10.23919/DATE.2018.8342213.

[48] X. Yin, A. Aziz, J. Nahas, S. Datta, S. Gupta, M. Niemier, and X. S. Hu, "Exploiting ferroelectric FETs for low-power non-volatile logic-in-memory circuits," in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2016, pp. 1–8. DOI: 10.1145/2966986.2967037.

[49] M. Jerry, P. Y. Chen, J. Zhang, P. Sharma, K. Ni, S. Yu, and S. Datta, "Ferroelectric FET analog synapse for acceleration of deep neural network training," in *IEEE International Electron Devices Meeting (IEDM)*, Oct. 2017, pp. 6.2.1–6.2.4. DOI: 10.1109/IEDM.2017.8268338.

[50] M. Hoffmann, M. Pešić, S. Slesazeck, U. Schroeder, and T. Mikolajick, "On the stabilization of ferroelectric negative capacitance in nanoscale devices," *Nanoscale*, vol. 10, no. 23, pp. 10 891–10 899, 2018. DOI: 10.1039/c8nr02752h.

[51] M.-Y. Y. Kao, A. B. Sachid, Y.-K. K. Lin, Y.-H. H. Liao, H. Agarwal, P. Kushwaha, J. P. Duarte, H.-L. L. Chang, S. Salahuddin, and C. Hu, "Variation Caused by Spatial Distribution of Dielectric and Ferroelectric Grains in a Negative Capacitance Field-Effect

Transistor," *IEEE Transactions on Electron Devices*, vol. 65, no. 10, pp. 4652–4658, Oct. 2018. DOI: 10.1109/TED.2018.2864971.

[52] S. Migita, "(Invited) Apprehension of Large Coercive Field in Ferroelectric Hf-Zr-O Thin Films," *ECS Transactions*, vol. 86, no. 2, pp. 13–19, Jul. 2018. DOI: 10.1149/08602.0013ecst.

[53] H. Mulaosmanovic, E. T. Breyer, T. Mikolajick, and S. Slesazeck, "Recovery of Cycling Endurance Failure in Ferroelectric FETs by Self-Heating," *IEEE Electron Device Letters*, vol. 40, no. 2, pp. 216–219, Feb. 2019. DOI: 10.1109/LED.2018.2889412.

[54] K. Ni, P. Sharma, J. Zhang, M. Jerry, J. A. Smith, K. Tapily, R. Clark, S. Mahapatra, and S. Datta, "Critical Role of Interlayer in Hf 0.5 Zr 0.5 O 2 Ferroelectric FET Nonvolatile Memory Performance," *IEEE Transactions on Electron Devices*, vol. 65, no. 6, pp. 2461–2469, Jun. 2018. DOI: 10.1109/TED.2018.2829122.

[55] I. S. Esqueda and H. J. Barnaby, "A defect-based compact modeling approach for the reliability of CMOS devices and integrated circuits," *Solid-State Electronics*, vol. 91, pp. 81–86, Jan. 2014. DOI: 10.1016/j.sse.2013.10.008.

[56] B. Zeng, M. Liao, J. Liao, W. Xiao, Q. Peng, S. Zheng, and Y. Zhou, "Program/Erase Cycling Degradation Mechanism of HfO 2 -Based FeFET Memory Devices," *IEEE Electron Device Letters*, vol. 40, no. 5, pp. 710–713, May 2019. DOI: 10.1109/LED.2019.2908084.

[57] D. K. Schroder, "Oxide and Interface Trapped Charges, Oxide Thickness," in *Semiconductor Material and Device Characterization*, 3rd, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2005, ch. 6, pp. 319–387. DOI: 10.1002/0471749095.

[58] N. Zagni, P. Pavan, and M. A. Alam, "A Memory Window Expression to Predict the Scaling Trends and Endurance of FeFETs," *arXiv preprint*, Mar. 2020. arXiv: 2003.12743.

[59] N. Zagni, P. Pavan, and M. A. Alam, "A memory window expression to evaluate the endurance of ferroelectric FETs," *Applied Physics Letters*, vol. 117, no. 15, p. 152 901, Oct. 2020. DOI: 10.1063/5.0021081.

Chapter 4

Role of Carbon Doping in the Operation of Lateral GaN Power Devices

Abstract — In this chapter we investigate the role of Carbon doping on the operation of lateral GaN-based power HEMTs.

First, we analyze the off-state, three-terminal, lateral breakdown of AlGaN/GaN HEMTs with different L_{GD} and with either Unintentionally Doped (UID) or C-doped buffer layers. Breakdown occurs due to a combination of gate electron injection and source-drain punch-through current in the former case, while in the latter case due to a combination of avalanche breakdown triggered by gate electron injection.

Second, we presented a 'hole redistribution' model explaining the R_{ON} stress/recovery transients that solves the puzzle of the activation energy being the same in both phases and explains the full recovery of the R_{ON}.

Third, concerning the influence of buffer traps on V_t instability experimental/simulation results under relatively high negative gate bias stress reveal a positive V_t shift. This behavior is ascribed to the increased negative ionized acceptor trap density in the buffer and the recombination of electrons injected from the gate terminal with holes emitted from C-related traps.

4.1 Wide-Band Gap Devices for New Generation Power Electronics

The concept of power electronics, first introduced in 1975, entails the conversion of electrical energy using power semiconductor devices and circuits [1]. The energy conversion can be of the following types: DC-DC (voltage conversion), DC-AC, AC-DC, AC-AC (voltage or frequency conversion). Regardless of the particular type of converter, the main performance parameter is the conversion efficiency (η) simply defined as

$$\eta = \frac{P_{out}}{P_{in}}. \tag{4.1}$$

The typical efficiency of power conversion with current Si technology is around 85%-95% which is not high enough, as about 10% of the input electric power is wasted as heat during each conversion cycle [1]. In general, the efficiency of power electronics is limited by the non-idealities of its components, namely of the semiconductor switch (either a power transistor or a diode) and of the lumped elements (capacitors, inductors) storing the energy. In recent years, a research field emerged to explore the so-called Wide Band Gap Semiconductors (WBGs) to replace Si in power devices. The usefulness of WBGs in power electronics stems from their wide E_g that allows operating devices at high voltage and temperature with lower losses compared to Si [1]–[3]. The most important (and mature) WBGs are Silicon Carbide (SiC) and Gallium Nitride (GaN); other emerging WBGs are Gallium Oxide (Ga_2O_3) and Diamond (C).

Among these materials GaN has emerged as a very promising candidate for power electronics thanks not only to its outstanding intrinsic properties (which will be discussed in SEC. XX) but also to the maturity of its manufacturing process, that makes it possible to build GaN devices on large diameter Si wafers (leading to high yield and reduced costs) [2], [3]. Despite the huge interest in the technology by both academia and industry [4], there are several challenges that need to be solved for further development of GaN electronics. Particularly, stability and reliability of these devices is one of the most pressing concerns [3]. A mature technology in fact should not only provide outstanding performance, but should also guarantee stability over time and long-term operation. In this chapter, we will focus on the role of Carbon doping of lateral GaN HEMTs in determining their breakdown capability and the stability of key parameters (i.e., threshold voltage,

V_t, and on-state resistance, R_{ON}).

4.2 Background

4.2.1 WBGs vs Si: Power Electronic Devices Metrics

Power Electronic Devices require high breakdown voltage (V_{BD}) capability, as during operation they are subject to large voltage that should not cause the device to fail. A simple definition of V_{BD} can be given considering a reverse-biased *pn* junction (with the weakly doped n-type layer being positively biased with a large voltage) in which the depletion layer width reaches the full length of the n-type layer at breakdown [5]. Under this assumption, the electric field in the n-type layer has a triangular shape with peak (at breakdown) equal to E_{crit}, i.e., the critical field strength for avalanche breakdown (which depends on material properties). Thus, simply

$$V_{BD} = \frac{E_{crit} \times W_D}{2}, \tag{4.2}$$

where W_D is the width of the depletion layer. E_C is a material property that is proportional to E_g of a semiconductor, i.e., the larger the bandgap, the larger E_C and hence V_{BD}. For this reason, WBG semiconductors have been explored as replacement for Si in power electronics. It is common to evaluate ideal material performance by means of specific Figures-Of-Merit (FOMs) that compare their electrical parameters to understand which material performs better. One such FOM is the Baliga's FOM [5], that relates the material V_{BD} to the specific on-resistance, R_{ON}, i.e., $R_{ON,sp} = R_{ON} \times A$ (where A is the device area) as follows

$$B-FOM = \frac{V_{BD}^2}{R_{ON,sp}} = \frac{\varepsilon_s \mu_n E_{crit}^3}{4}. \tag{4.3}$$

The B-FOM calculated from Eq. (4.3) is shown in Fig. 4.1 for several different WBG (compared to Si). The material parameters used for the calculation are collected in TAB and were taken from [6]. Thanks to the B-FOM one can clearly see the advantage of using WBG instead of Si for power applications. An important takeaway message from B-FOM is the trade-off between V_{BD} and R_{ON}, as for a given technology increasing the former leads to an increase in the latter. In

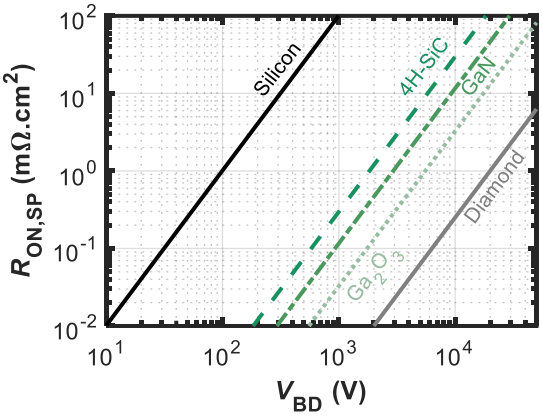


Figure 4.1: Comparison of Baliga FOM (B-FOM) for Si and four different WBG semiconductors, namely 4H-SiC, GaN, Diamond and β-Ga_2O_3.

Table 4.1: Si and WBGs Material Parameters. From [6].

	Si	**4H-SiC**	**GaN**	**Diamond**	**β-Ga_2O_3**
E_g (eV)	1.1	3.3	3.4	5.5	4.5-4.9
μ_n (cm^2/Vs)	1.4×10^3	1×10^3	1.2×10^3	2×10^3	300
E_{crit} (MV/cm)	0.3	2.5	3.3	10	8
ε_s	11.8	9.7	9.0	5.5	10
$B-FOM$	1 (Reference)	340	870	24664	3444

Section 4.2.4 we will see how incorporating impurities (such as Carbon) in the buffer of lateral GaN HEMTs leads to an increase in V_{BD} at the expense of R_{ON}.

4.2.2 Elements of Physics of AlGaN/GaN HEMTs

Unlike a conventional Si MOSFET, where the channel conducting current is formed due to the inversion of the body induced by the electrostatic potential of the gate terminal, a GaN HEMT bases its high carrier concentration and conductivity on the modulation of a two-dimensional electron gas (Two-Dimensional Electron Gas (2DEG)) forming at the heterojunction between AlGaN and GaN. A conventional GaN HEMTs in fact, is formed by stacking a layer of GaN with $Al_xGa_{1-x}N$ (with some molar fraction $x > 0$) giving rise to a polarization charge that attracts electrons at the interface between the two semiconductors. The polarization charge, under some

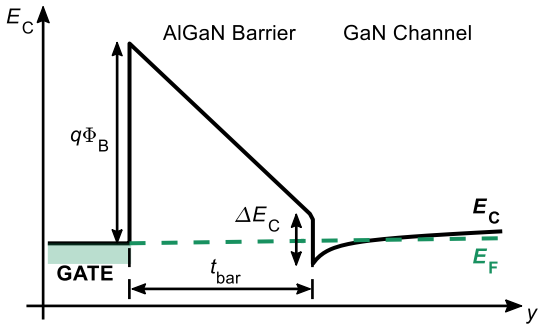


Figure 4.2: Conduction Band profile at the AlGaN/GaN interface at equilibrium ($V_G = 0\,\mathrm{V}$).

constraints, forms naturally due to the polar nature of the nitride semiconductors [7]. A sketch of the conduction band profile at equilibrium ($V_G = 0\,\mathrm{V}$) is shown in Fig. 4.2. The E_C profile between AlGaN and GaN has an abrupt discontinuity at the heterojunction given by the difference in electron affinity, $\Delta\chi = \Delta E_C$. Interestingly, at $V_G = 0\,\mathrm{V}$ E_C in the GaN channel layer is below E_F thus giving rise to strong confinement for electrons. This confinement creates a so-called "quantum-well" that is generally few nanometers thick and is conventionally approximated as a two-dimensional layer in which electrons are able to move with high mobility (due to the absence of impurities in the channel). This is the origin of the 2DEG. As the 2DEG layer is present even at zero bias, AlGaN/GaN-based HEMTs are normally-on devices (i.e., $V_t < 0\,\mathrm{V}$). A simple expression for the V_t of these devices can be derived from the band-diagram in Fig. 4.2. The threshold condition is set as the gate voltage necessary to deplete completely the 2DEG below the gate. Thus, we obtain [7]

$$V_t^{HEMT} = -\frac{Q_\pi^{net} t_{bar}}{\varepsilon_{AlGaN}} + \Phi_B - \frac{\Delta E_C}{q}. \tag{4.4}$$

Q_π^{net} is the *net* polarization charge forming at the AlGaN/GaN interface; for 2DEG to form, $Q_\pi^{net} > 0$ (because of electrostatic attraction). The electron density in the 2DEG layer, n_s, can be obtained as follows [7]

$$n_s = \frac{Q_\pi^{net} + \varepsilon_{AlGaN}\left[V_{GS} - (\Phi_B - \Delta E_C/q)\right]}{q(t_{bar} + t_{2DEG})} = \frac{\varepsilon_{AlGaN}}{q(t_{bar} + t_{2DEG})}(V_{GS} - Vt) \tag{4.5}$$

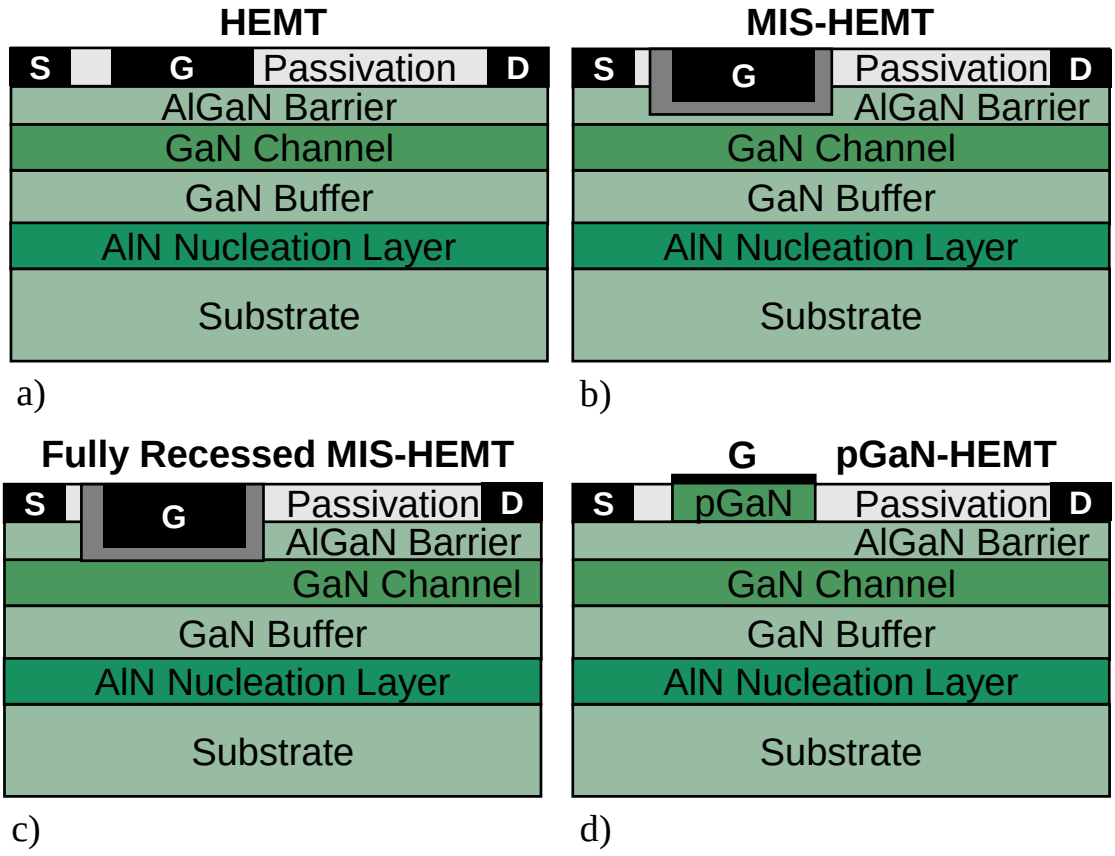


Figure 4.3: Sketch of the AlGaN/GaN HEMT and of the different gate stacks with (a) Schottky-gate HEMT (HEMT), (b) insulator (Metal-Insulator-Semiconductor HEMT (MIS-HEMT)) (c) fully recessed insulator (Fully Recessed MIS-HEMT), and (d) p-GaN layer (pGaN-HEMT)

.

which resembles the expression for the inversion charge density in a classical MOSFET (noting that $\varepsilon_{AlGaN}/(t_{bar}+t_{2DEG})$ has the dimension of capacitance per unit area).

4.2.3 Architectures of GaN Lateral Devices

AlGaN/GaN lateral devices can have different gate stacks, that allow them to be dived into four main categories. These are depicted in Fig. 4.3. The first one is the HEMT with the gate metal contact forming a Schottky diode with the barrier, i.e., it directly contacting the semiconductor [8], see Fig. 4.3(a). The second one has an insulator between the barrier and gate metal contact, and is called MIS-HEMT [9], see Fig. 4.3(b). Both HEMT and MIS-HEMT are normally-on

transistors. The two remaining structures instead are normally-off transistors, i.e., $V_t > 0\,\mathrm{V}$. The third one, see Fig. 4.3(c), is the fully recessed MIS-HEMT, with the insulator directly contacting the GaN channel thus removing the polarization charge beneath and pinching-off the 2DEG at equilibrium. The fourth and last structure includes a p-type GaN layer (pGaN) in the gate stack forming a p-i-n diode with the underlying AlGaN (i) and GaN (n) layers, thus depleting the 2DEG at equilibrium [10]. The pGaN HEMT is shown in Fig. 4.3(d).

Although AlGaN/GaN lateral devices can have different gate stacks they all share basically the same epitaxial structure [8]–[10], [12]. This starts with a Si or SiC substrate (depending on whether to use the transistor for power or Radio Frequency applications [3]) and with an AlN nucleation layer, that aims at reducing strain and dislocations to the above active layers. In addition, it isolates the active device from the substrate thus reducing vertical leakage. Above the Nucleation layer stands the GaN Buffer (which can be separated from the Nucleation layer by several layers, called Transition Layers, again used to reduce strain and growth imperfections) above which the actual GaN channel layer is grown. The 2DEG forms in this layer at the interface with the AlGaN barrier layer. The basic device structure is then completed by the passivation layer (typically SiN_x) that is then selectively etched to deposit the metal contacts.

4.2.4 Breakdown Capability and Buffer Doping

Being devised for power applications, AlGaN/GaN HEMTs need to sustain large V_{BD} to avoid premature failure. GaN grown by Metal Organic Vapor Phase Epitaxy (MOVPE)/Chemical Vapor Deposition (CVD) exhibit a low n-type conductivity (on the order of $1 \times 10^{16}\,\mathrm{cm}^{-3}$) due to incorporation of either Si or Oxygen impurities during growth [11]. This native conductivity can cause severe issues during OFF-state conduction (i.e., for $V_{GS} < V_t$ and $V_{DS} \gg 0\,\mathrm{V}$) as high leakage currents in the buffer can build up due to the high electric fields [12]. The leakage current, I_{OFF}, can be due to punch-through between either source/drain or gate/drain contacts, depending on the design. These leakage components can be effectively suppressed by incorporating either Iron (Fe) or Carbon (C) impurities in the GaN buffer. In fact, these impurities act as acceptor levels in the bandgap that deplete the channel from electrons thereby increasing its resistivity [12]–[14]. Basically, the effect of these trap levels in the buffer is to deplete the 2DEG in the

access regions (i.e., the ungated regions between the source and gate and between the gate and drain) when the device is biased in the OFF-state thus increasing V_{BD}. As mentioned earlier, this comes at the price of increased resistivity, i.e., higher R_{ON}, compared to undoped buffer devices [12]. Moreover, depending on their energy location in the bandgap buffer traps can give rise to dynamic effects that cause dispersion in the $I-V$ characteristics of the devices. The trapping mechanisms along with their effects are briefly summarized in Section 4.2.5.

A third leakage component that limits the high-voltage capability of AlGaN/GaN HEMTs is the vertical current that comes up from the substrate. This leakage component plays a fundamental role in well-isolated lateral devices and represents the ultimate limit to breakdown for large area devices [12].

4.2.5 Trapping Mechanisms and Relative Detrimental Effects

The presence of trap states in the bandgap of GaN and its alloys poses challenges to both the stability and reliability of devices based on this technology. Trap levels are the root-cause for slow charge capture/emission processes [15] leading to dispersion effects during their typical operating conditions in power RF and power switching applications. In this section we briefly describe the most important trapping mechanisms affecting GaN transistors and then review the resulting trap effects.

Trapping Mechanisms

Traps influence the electrical behavior of transistors since they are characterized by relatively long capture and emission times — thus charging or discharging (under fast bias changes) more slowly than device capacitances governing the "prompt" device response — causing dispersion.

Regardless of the nature of traps, i.e. whether they are associated to intrinsic crystallographic defects, unintentional or intentional impurities or defect-impurity complexes [15], only those traps that change their state during operation through capture or emission of mobile carriers can effectively cause issues to the device performance. Given this, trapping mechanisms can be conveniently categorized as the combination of trap location (along with the associated charging/discharging path) and type of involved mobile carriers. Traps can be virtually located

in any semiconductor or dielectric layer (as well as at interfaces) across the epitaxial structure, see Fig. 4.3. However, the locations that have more frequently been associated with harmful effects include the device surface within the gate-drain access region, the barrier, the buffer, the surface passivation layer, and, when present, the gate dielectric (including the interface with the underlying semiconductor) or the pGaN layer.

On the other hand, the sources of mobile carrier involved with trapping/emission events include all the leakage currents from the device terminals (gate, source, drain, substrate), the 2DEG at the barrier/channel interface in HEMTs and unrecessed or partially-recessed MIS-HEMTs or at the dielectric/GaN interface in fully-recessed MIS-HEMTs, the floating C-doped buffer, as well as high-field charge generation mechanisms like impact ionization and Zener trapping. One or more of these carrier sources can be activated depending on the applied device bias, that induces an *increase in the negative charge* associated to traps, i.e. either *electron capture* or *hole emission* (the opposite processes take place when the bias is removed or reversed). Table 4.2 lists the major trapping mechanisms that have been reported in the literature classified in terms of trap location and of the corresponding source (and type) of involved mobile carriers. In this chapter, we will mainly discuss buffer traps associated with Carbon doping and analyze their associated effects.

Trapping Effects

Here we briefly review the major charge trapping effects observed in GaN transistors and associate them to the mechanism(s) put into evidence in the previous section. It is important to observe that since GaN transistors for practical applications are n-type transistors, traps can induce a reduction in the source-to-drain channel conductivity and, consequently, I_D if the negative trapped charge increase (or if the positive one decreases). These changes can result either from electron capture or hole emission. On the other hand, an increase in the channel conductivity and I_D can be promoted by electron emission or hole capture. We also point out that trapping effects taking place in the device portion under the gate directly influence V_t, whereas those occurring in the access regions (source-to-gate and gate-to-drain) impact the associated access resistances and the transconductance peak [40], [41].

Table 4.2: Major trapping mechanisms in GaN transistors classified in terms of trap location and source (type) of involved mobile carriers.

Trap Location	**Source (Type) of Involved Carriers**	**References**
Surface in the G-D Access Region	Gate Contact (Electrons)	[16], [17]
Surface in the G-D Access Region	Gate 2DEG (Electrons)	[18], [19]
Barrier	Gate Contact (Electrons)	[20], [21]
Barrier in the G-D Access Region	2DEG (Electrons)	[18], [22]
Barrier/Channel Interface	Barrier Traps (Electrons)	[23]
Barrier	2DEG (Electrons)	[3], [38]
pGaN/Barrier Interface	pGaN Layer (Holes)	[3], [39]
Buffer	Source Contact (Electrons)	[12], [24]
Buffer Under the Gate	Gate Contact (Electrons)	[17], [18], [25]
Buffer in the G-D Access Region	Substrate Contact (Electrons)	[12], [26]
Buffer in the G-D Access Region	2DEG (Electrons)	[22], [27]
Buffer and Barrier/Buffer Interface[†]	Zener Trapping (VB Electrons)	[28], [29]
Barrier/Buffer Interface[†]	Impact Ionization (Electrons)	[30]
Buffer	"Leaky-Dielectric" Buffer	[31], [32]
Buffer	C doping (Holes)	[60], [61]
Gate Insulator	Gate Contact (Electrons)	[33], [34]
Insulator/Semiconductor Interface	2DEG (Electrons)	[29], [35]–[37]
Insulator/Semiconductor Interface	C-related Acceptors (Holes)	[68]

[†] In this case the buffer layer is undoped and it is not distinguished from the channel.

The main trapping effects are listed as follows.

- *RF Current Collapse*. This effect results in a reduction in the maximum I_D and simultaneous increase in the minimum V_{DS} to achieve saturation, thus reducing the explorable range by the operating point during the RF sweep [16], [42], [43]. The increase in the minimum V_{DS} is also referred to as *knee-voltage walk-out*. Responsible mechanisms for current collapse are related to electron trapping taking place during the OFF-state part of the RF cycle. Optimization of the technology through passivation of the access regions and the use of field plates to mitigate electric field peaks [12] left buffer traps as the major contributors to current collapse, especially in devices where compensating impurities (like Fe) are used to increase the breakdown voltage [12]–[14].

- *Dynamic R_{ON} Increase*. When GaN transistors are used in power switching converters, the most detrimental trap-related effect is the increase in the dynamic R_{ON} compared to its DC value [19], [28], [41], [44], resulting in an undesirable increase in power losses (and consequent efficiency reduction). This effect is related to modulation of the channel conductivity, reduced by the increase in negative trapped charge during OFF-state and that cannot be restored promptly as the device is driven to ON-state, leading to the dynamic R_{ON} increase. Possible underlying mechanisms include all the electron trapping processes inducing also RF current collapse, that take place during the OFF-state phase of the switch-mode operation. In addition to these mechanisms, there are trapping processes specific to power transistors (because of the higher biases employed) that involve the charging and discharging of buffer traps [31], [41], [55], [60]. This will be subject of investigation in Section 4.4.

- *V_t Instability*. In GaN transistor types employing a gate insulator, i.e. the MIS-HEMT or fully recessed MIS-HEMT, see Fig. 4.3(b)-(c), also the gate insulator and the interface with the underlying semiconductor can be trap sites causing V_t instability (i.e., drift in either the positive or negative direction with respect to the value of the pristine device). These instabilities are typically analyzed by applying either negative or positive gate bias stress voltages with $V_{DS} = 0\,\mathrm{V}$, with the aim of isolating V_t instability effects from the drain access-resistance ones [3]. For normally-on devices (i.e., partially recessed MIS-HEMTs) with large negative threshold, V_t stability under negative gate stress is the most critical aspect. On the contrary, in normally-off devices (i.e., fully recessed MIS-HEMTs), V_t stability under positive gate stess is instead of major concern. However, in the latter devices, assessing V_t stability under negative gate stress can be important as well, since a negative V_{GS} can be applied to switch off the transistor, in order to prevent false turn-on and ensure safe operation against voltage spikes on the gate [45]. In any case, negative gate stress measurements are a proxy for the off state operation, as similar, large values of drain-gate voltage can be achieved with both biasing conditions. This effect will also be investigated in Section 4.4.

- *Kink Effect*. The "kink" effect is an operational instability emerging during a V_{DS} sweep, by which I_D in the saturation region is initially compressed and then increases (over a relatively narrow voltage range) to a higher value [23]. This behavior is detrimental especially in transistors for RF amplifiers because it can result in transconductance compression and output conductance increase. Most recent works on this issue proposed that the kink effect can be caused by: *i)* impact ionization of traps in the channel or the barrier emitting electrons [30], *ii)* trapping of holes generated by trap-assisted tunneling into C-related traps [46], *iii)* field-enhanced ionization of AlGaN barrier traps under the gate and near the GaN/AlGaN interface [23].

4.3 Contribution: Mechanisms Underlying Three-Terminal Breakdown in Carbon-Doped GaN Power HEMTs

As discussed previously, in AlGaN/GaN HEMTs for power switching applications, the three-terminal OFF-state breakdown voltage (V_{BD}) is typically extended up to the vertical breakdown limit by compensating the unintentional conductivity in the buffer through Carbon (C) doping and by increasing the lateral gate-to-drain spacing (L_{GD}) [12]. V_{BD} is typically found to scale about linearly with L_{GD} with slope that is smaller than the critical field for avalanche ($E_{crit} \approx 3.4-3.9\,\mathrm{MV/cm}$ [6], [47]). This is often considered to be an indication that avalanche generation should be ruled out as the phenomenon determining V_{BD}. This misleading interpretation however is based on the tacit assumption of a quite idealized, constant electric-field distribution throughout the access region between gate and drain. The field distribution, however, is two-dimensional and, above all, characterized by intense accumulation spots at the drain-end of the gate, under the field-plate end (if present), and at the drain contact border [48]. Moreover, it is strongly influenced by the intrinsic or doping-related traps in the buffer. For these reasons, numerical device simulation is probably the only means by which the role played by avalanche generation in determining the OFF-state breakdown in lateral AlGaN/GaN HEMTs can be clarified. The purpose of this section is to provide physical insights into the OFF-state, three-terminal, lateral breakdown in AlGaN/GaN HEMTs for power switching applications and to highlight the role of avalanche generation and the other possible breakdown limiting phenomena. The numerical analysis presented in this section is based on the devices realized in [12] that are taken as reference because said work is one of the very few examples in the open literature providing a systematic analysis of the V_{BD} vs L_{GD} dependence with and without C doping in the GaN buffer. The sketch of the simulated structures, i.e., without and with C doping in the buffer are shown in Fig. 4.4(a) and (b), respectively. Gate-to-source distance (L_{GS}), gate length (L_G) and gate field plate overhang (L_{FP}) are for both structures 1, 0.7, 0.6 µm, respectively. The substrate is semi-insulating SiC. The substrate contact was left floating in the simulations to reproduce measurement conditions in [12], so that vertical breakdown is not expected to play a role for the voltage range and L_{GD} under investigation. The full list of geometric parameters used in the simulations is available in [54].

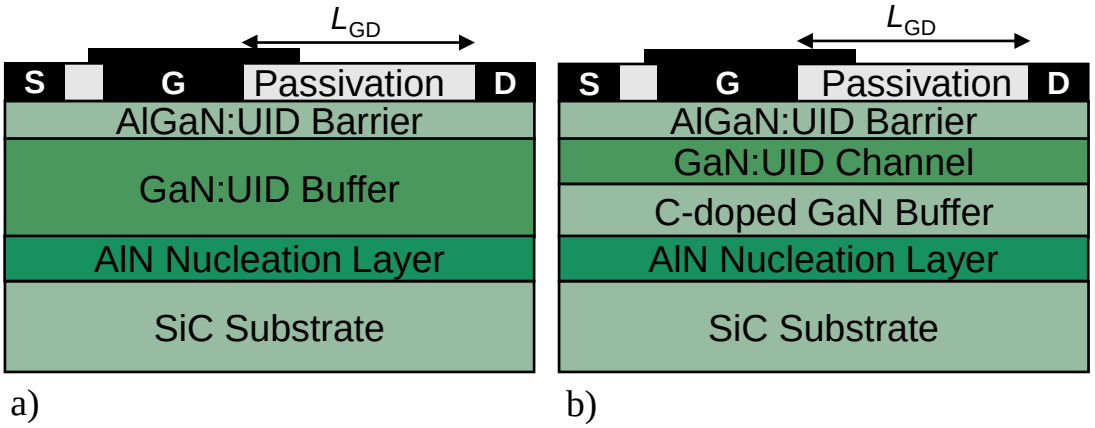


Figure 4.4: Sketch of the cross-section of the AlGaN/GaN HEMTs used for the numerical breakdown analysis, (a) without and (b) with C doping in the buffer.

Before discussing the simulation results, we briefly describe the simulation setup used to reproduce the experimental data in [12] — focusing in particular on the modeling of the C-doped buffer (that will be used throughout the rest of this chapter).

4.3.1 TCAD Simulation Setup

Device simulations were carried with SDevice simulator (Synopsys) [14, Ch. 2]. Carrier distribution was modeled with the Fermi-Dirac statistic, Shockley-Read-Hall (SRH) recombination was included as well as mobility degradation due to doping and high field. Piezoelectric polarization was included by using the strain model included in the simulator. A fully dynamic trap modeling approach was adopted, with one SRH trap-balance equation for each distinct trap level included, describing the dynamics of trap occupation without any quasi-static approximation. Both gate and source/drain contacts were modeled as Schottky contacts with proper barrier/workfunction. Electron tunneling was activated at the contacts to properly reproduce leakage current (at the gate) and to mimic Ohmic contacts (at the source and drain). A more detailed description of the modeling approach to describe device physics in AlGaN/GaN based HEMTs can be found in [49].

In the simulations, all process and geometrical parameters were set to their nominal values stated in [12]. Cynoweth's law was used to calculate impact-ionization generation rate for both

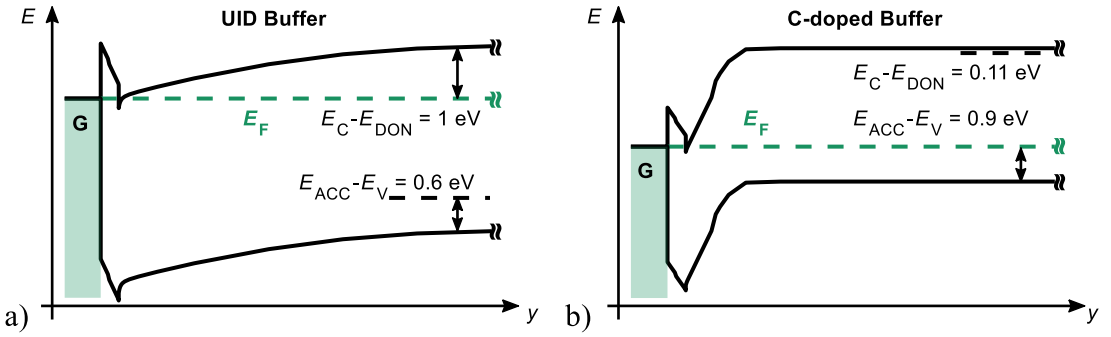


Figure 4.5: Band diagram for the HEMT with (a) UID and (b) C-doped buffer, showing the trap levels in both cases. In the UID buffer case, donors are the dominant level (rendering the buffer weakly n-type) whereas in the C-doped buffer case acceptors are dominant (rendering the buffer weakly p-type).

electrons and holes, with coefficients set in agreement with recent Monte-Carlo calculations [50]. The full list of parameters for calibration of the simulations is available in [54].

Modeling of UID and C-doped Buffer Traps

For the UID GaN buffer, traps in the bandgap were modeled by assuming a pair of intrinsic donor-acceptors as in [13], i.e., the donor level is 1 eV below E_C whereas the acceptor level is 0.6 eV above E_V, as shown in Fig. 4.5(a). In this case, because the donor trap concentration is higher than that of acceptors, i.e., $N_D = 1 \times 10^{16}\,\mathrm{cm}^{-3}$ vs $N_A = 5 \times 10^{15}\,\mathrm{cm}^{-3}$, donors are the dominant traps and E_F is pinned at 1 eV below E_C, thus rendering the buffer weakly *n-type*.

Conversely, in the C-doped buffer, the dominant traps are the acceptors ($N_A = 8 \times 10^{17}\,\mathrm{cm}^{-3}$) with level at 0.9 eV above E_V [53] — partially compensated with shallow donors 0.1 eV below E_C [11], [52], $N_D = 4 \times 10^{17}\,\mathrm{cm}^{-3}$ — rendering the buffer weakly *p-type*. The band diagram for the C-doped buffer device is shown in Fig. 4.5(b). The weak p-type conductivity of the C-doped buffer creates a depletion region at the interface with UID channel in OFF-state conditions thus increasing the blocking voltage capability [12], [13], [51]. As it will become clearer from the results in Section 4.4, buffer design requires V_{BD} optimization to be traded off with detrimental trap effects. In the case of doped buffers, this is typically achieved by switching off the impurity flow during growth at a designed distance from the 2DEG [12], [51].

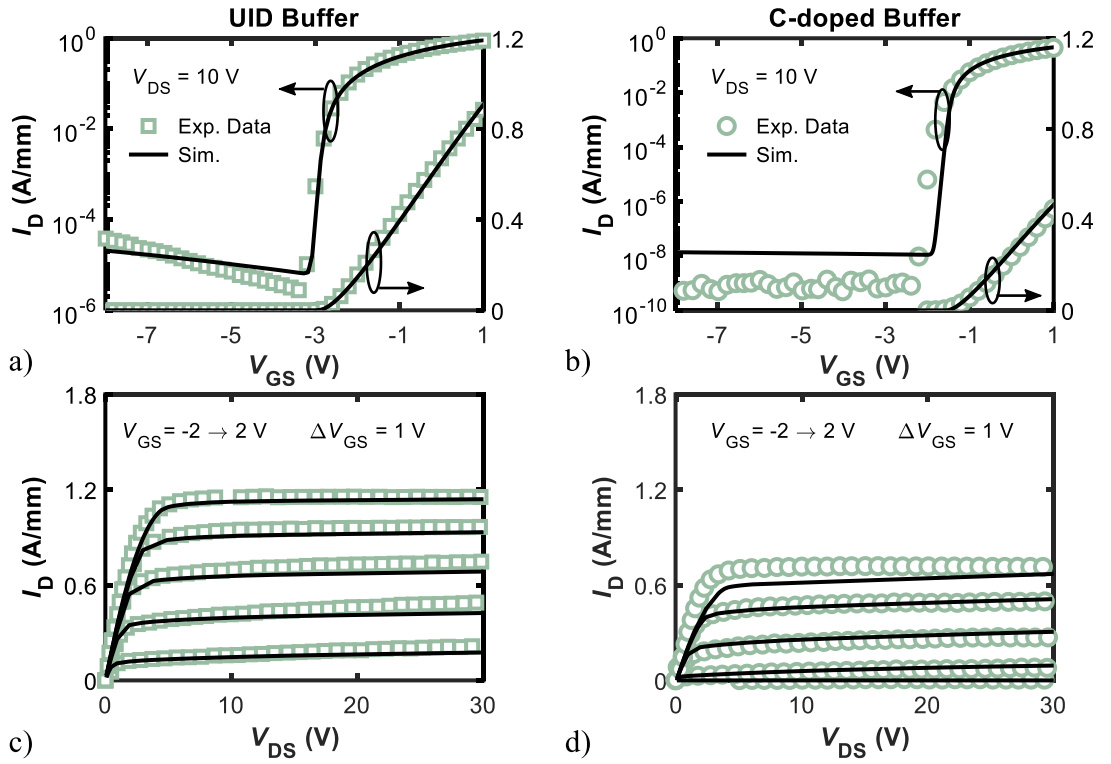


Figure 4.6: $I_D - V_{GS}$ (transfer) and $I_D - V_{DS}$ (output) curves for (a), (b) UID and (c), (d) C-doped buffer devices. For the transfer characteristic, V_{DS} is set to 10 V. For the output characteristic, V_{GS} is swept between -2 and 2 V at 1 V steps. Good overall agreement was found between experimental data [12] (symbols) and simulations (lines). Adapted from [54].

4.3.2 Breakdown Analysis

Simulations were first calibrated against experimental transfer and output $I-V$ curves for both UID and C-doped devices. The outcomes are illustrated in Fig. 4.6, showing the experimental and simulated transfer and output characteristics for UID and C-doped buffer devices. As it can be noticed, a satisfactory agreement was achieved in all cases. The calibrated simulations were then used to analyze the V_{BD} vs L_{GD} scaling. The results are reported in Fig. 4.7(a) and (b), respectively, showing the experimental and simulated off-state I_D–V_{DS} curves for UID and C-doped buffer devices with different L_{GD} values and the corresponding V_{BD} vs L_{GD} plots, respectively. Consistently with [12], V_{BD} is defined as the V_{DS} bias for which I_D reaches 1 mA/mm. As it can be noticed from Fig. 4.7, an overall reasonable agreement is achieved between simulated and experimental data. In particular, the simulations are able to fully capture

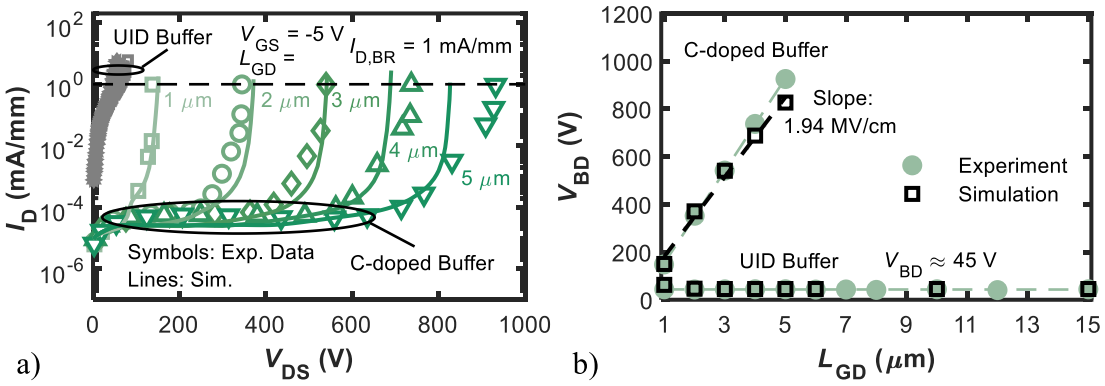


Figure 4.7: (a)Experimental (from [12]) (symbols) and simulated (lines) off-state I_D–V_{DS} curves for UID and C-doped buffer devices for different L_{GD} values. (b) Experimental (dots) and simulated (squares) V_{BD} as a function of gate-drain spacing L_{GD} for both UID and C-doped buffer devices. Adapted from [54].

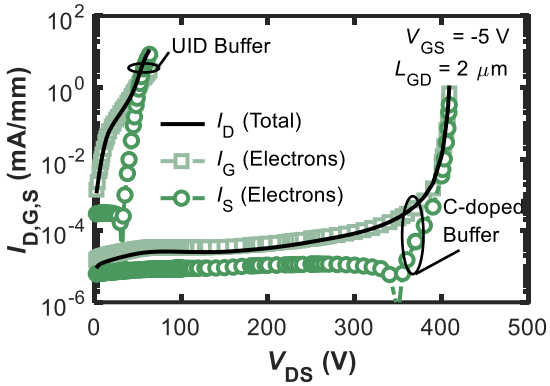


Figure 4.8: Simulated total drain (black solid lines), gate electron (light green squares), and source electron (green dots) currents vs V_{DS} for the UID and C-doped buffer devices with $L_{GD} = 2\,\mu$m. Adapted from [54].

the completely different behavior exhibited by UID and C-doped buffer devices in terms of V_{BD} vs L_{GD} relationship. Namely, V_{BD} shows no appreciable dependence on L_{GD} in the UID buffer HEMT, whereas it scales almost linearly with L_{GD} in the C-doped buffer device. We limited the comparison of simulation results with experimental devices for $L_{GD} < 6\,\mu$m because measurements were limited to $V_{DS} \leq 1000\,$V [12], and no breakdown occurred in this range for the longer devices.

Fig. 4.8 shows the simulated I_D up to V_{BD} for the UID and C-doped buffer devices with $L_{GD} = 2\,\mu$m, along with the corresponding electron currents entering the device from the gate and

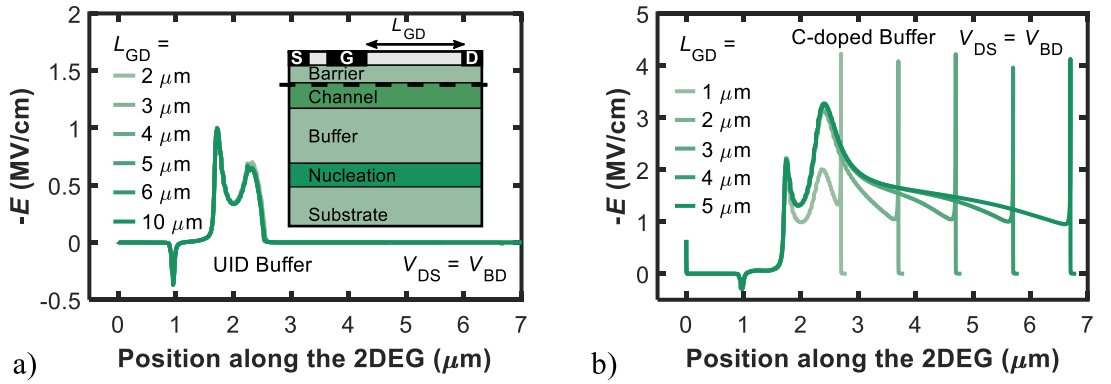


Figure 4.9: Parallel component of the electric-field as a function of position along the AlGaN/GaN interface for different L_{GD} values (at $V_{DS} = V_{BD}$) in the (a) UID buffer and (b) C-doped buffer device. Adapted from [54].

the source, I_G and I_S respectively. In both devices, I_D is approximately equal to the gate electron current up to V_{BD}. At this point, the source electron current becomes comparable with, and, in the case of the UID buffer device, even larger than I_G in the breakdown regime. In the device with C-doped buffer, instead, C-related acceptor traps effectively suppress leakage in terms of both gate electron injection and source-drain punch-through, so that breakdown occurs at much higher V_{DS} (at the same L_{GD}) than for the UID buffer device.

The electric-field distribution at V_{BD} along the AlGaN-GaN interface with varying L_{GD} is shown in Fig. 4.9(a) for the UID buffer and (b) for the C-doped buffer device, respectively. As discussed previously, in the UID buffer device the breakdown I_D limit of 1 mA/mm is reached due to the combination of gate-injected electron and source-drain punch-through currents. As shown in Fig. 4.9(a), the electric field peaks at the drain-end of the gate while it is negligibly small in the region between the end of the field plate and the drain contact. The peak of the field is smaller than the critical field for avalanche and, more importantly, is negligibly impacted by L_{GD}. For these reasons, V_{BD} is almost insensitive to L_{GD} as seen in Fig. 4.7 and Fig. 4.8.

For the C-doped buffer device instead, see Fig. 4.9(b), the high field region is effectively distributed throughout the gate-drain access region, and the field is non-negligible even in the region between the field-plate end and the drain contact. For this reason, increasing L_{GD} shifts the breakdown to larger voltages. However, the electric field distribution is non-uniform, with distinct peaks at three positions along the AlGaN-GaN interface, namely in correspondence of the

drain end of the gate, the field-plate edge, and the drain contact. As a result, the V_{BD}/L_{GD} slope is smaller than E_{crit} even though breakdown is induced by avalanche generation. The critical field for breakdown is specifically reached at the drain contact for all L_{GD} values considered.

4.3.3 Conclusions

We have analyzed the off-state, three-terminal, lateral breakdown of AlGaN/GaN HEMTs for power switching applications, by comparing two-dimensional numerical device simulations with experimental data from devices with different L_{GD} and with either UID or C-doped buffer layers. In the former case (i.e., UID buffer devices), the breakdown voltage is insensitive to L_{GD}, while in the latter case V_{BD} increases linearly with L_{GD} with a slope of $\approx 2\,\mathrm{MV/cm}$. These aspects are successfully captured by the simulations and are attributed to the different breakdown mechanisms in the two devices, namely: *i)* a combination of gate electron injection and source-drain punch-through current in UID buffer devices; and *ii)* avalanche breakdown triggered by gate electron injection in C-doped buffer HEMTs. Our TCAD device model can be useful for designers to predict the voltage handling capabilities of HEMTs during the device optimization loop, depending on the buffer doping employed. Moreover, it can also be adopted as an aid in the interpretation of failure modes during robustness and off-state step-stress tests.

4.4 Contribution: Influence of Buffer Traps on Dynamic ON- Resistance and Threshold Voltage Instabilities in C-doped GaN Power MIS-HEMTs

As mentioned earlier, Carbon doping of the buffer in GaN transistors is a widespread technology solution to reduce buffer conductivity and increase breakdown voltage for power applications [12], [13], [51]. The introduction of acceptor traps associated with C-doping, however, leads to an enhancement of current collapse [12], [13], dynamic R_{ON} degradation [32], [60] and V_t instabilities [29], [63], [68]. In this regard, several physical aspects on the role of C-doping in determining both dynamic R_{ON} degradation and V_t instability occurring when performing off-to-on switching still need to be fully understood. The purpose of this section is to elucidate the role of C-related traps in determining: *i)* the R_{ON} degradation and recovery transients after OFF-state stress with same activation energy, $E_A = 0.9\,\text{eV}$; and *ii)* the bidirectional V_t shifts after Negative Bias Temperature Instability (NBTI) stress.

4.4.1 Hole Redistribution To Explain ON-Resistance Stress/Recovery Experiments in AlGaN/GaN MIS-HEMTs

As discussed in Section 4.2.5, buffer doping induces an increase of charge trapping during stress experiments that lead into the manifestation of dispersion effects such as dynamic R_{ON} degradation. Experimentally, "stress" tests are carried out by applying bias conditions that are known to accelerate dynamic R_{ON} degradation, i.e., the R_{ON} increase during the typical pulse-mode operation of power transistors in power switching converters. Stress in this case is performed either: *i)* by applying a negative V_{GS} and a large positive V_{DS} (with source and substrate contacts grounded), or *ii)* by applying a negative bias to the substrate contact (V_{SUB}) (with all other contacts grounded). We will refer to the above stress conditions as to Front-Gating OFF-State Stress (FGOS) and Back-Gating OFF-State Stress (BGOS), respectively. During either FGOS or BGOS experiments, the C-doped buffer can be the site for peculiar trapping mechanisms, that are mainly governed by the C_N acceptor state at $E_V + 0.9\,\text{eV}$. It has been actually shown that both R_{ON} increasing transients under either FGOS or BGOS and subsequent R_{ON} recovery transients

are all thermally activated with the same E_A of 0.9 eV [56], clearly indicating a correlation with C-related acceptor traps. While this peculiar behavior was previously explained by correlating the thermally activated stress with the temperature-induced increase of buffer leakage [41], [56], here we provide numerical simulation results showing that C-related traps are responsible for R_{ON} stress/recovery transients being thermally activated with the same E_A. The model described here, labelled 'hole redistribution' model for brevity, assumes that when the device is in the OFF state holes are being emitted by C-related acceptor traps within the gate-drain access region, leading to an increase in the density of negatively-ionized traps and therefore to an increase of R_{ON}. Conversely, during recovery transients, previously emitted holes (and re-trapped in a different location of the buffer) get re-emitted and are captured by the same traps that emitted them in the first place [60], [61].

FGOS/BGOS Stress and Recovery Transients

The simulated device, resembling the actual samples on which stress/recovery measurements were carried out [56], is a partially recessed AlGaN/GaN MIS-HEMT, whose cross-section was shown in Fig. 4.3b. The comparison between experimental data and simulation results is shown in Figs. 4.10 and 4.11, with stress and recovery conditions applied as follows. *i)* FGOS and recovery: $(V_{GS}, V_{DS}, V_{SUB}) = (-8, 25, 0)$ V and $(V_{GS}, V_{DS}, V_{SUB}) = (0, 0.5, 0)$ V, respectively; *ii)* BGOS and recovery: $(V_{GS}, V_{DS}, V_{SUB}) = (0, 0, -25)$ V and $(V_{GS}, V_{DS}, V_{SUB}) = (0, 0.5, 0)$ V, respectively[1]. The chosen experimental FGOS and BGOS conditions represent "intermediate" OFF-state bias conditions, i.e., with V_{DS} that is large enough to have appreciable dynamic R_{ON} degradation but, at the same time, low enough not to promote significant electron injection through the C-doped buffer due to lateral source-drain punch-through or vertical leakage current. During stress simulations, R_{ON} values were obtained by fast sweeping the device bias to measurement conditions $(V_{GS}, V_{DS}) = (0, 0.5)$ V in 10 ms to mimic On-the-Fly (OTF) measurements [56]. During recovery,

[1]Note that both stress conditions (i.e., FGOS and BGOS) bias the device in the subthreshold region. However, the BGOS setup is useful to rule out surface trapping effects — which can be present during FGOS instead — thus allowing us to attribute the observed phenomena to buffer traps only [57], [58]. Under BGOS tests in fact, the formed 2DEG channel screens the superficial layers from the field effect induced by back-gating, so surface effects should be negligible [57]. The fact that a similar kinetics was found for FGOS and BGOS is an indication that buffer traps are mainly involved [41], [56]. Moreover, since we only included buffer traps in our simulation setup, both FGOS and BGOS conditions modify the state of C-related traps only.

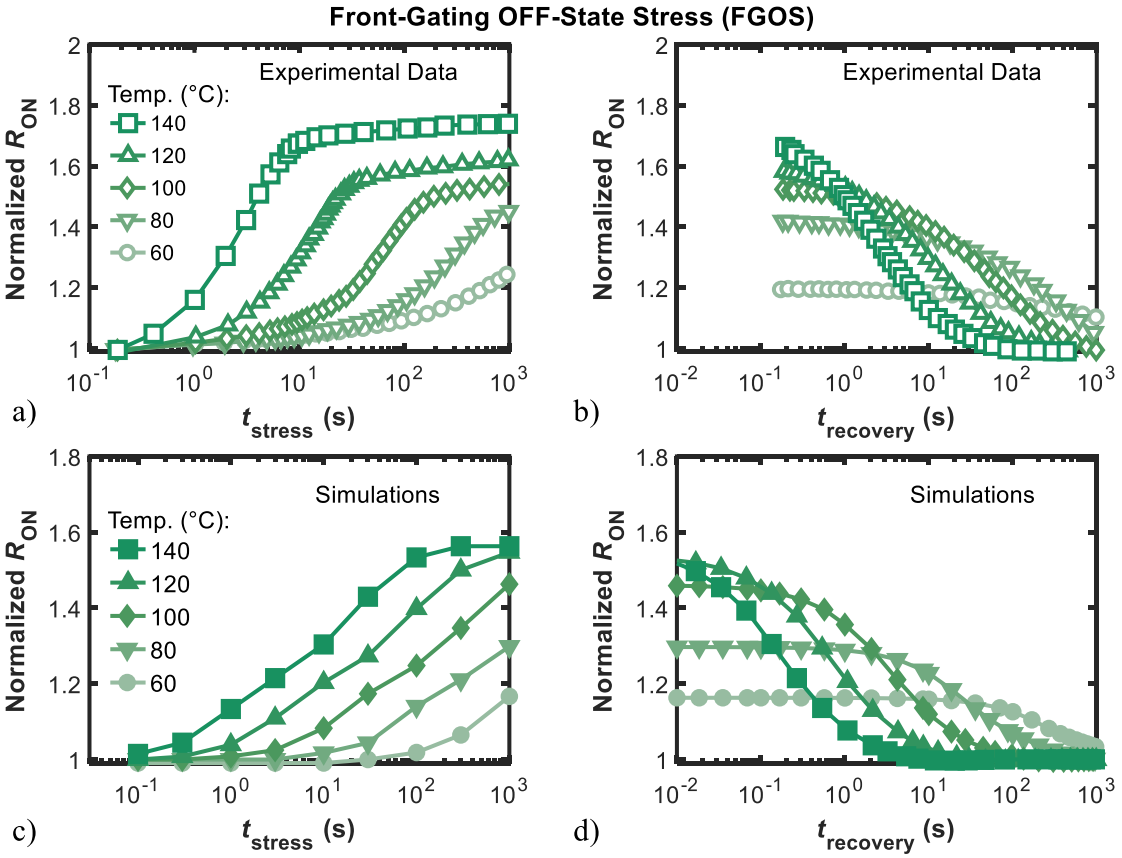


Figure 4.10: R_{ON} variations (normalized w.r.t. the pre-stress value) during FGOS (a), (c) and (b), (d) consequent recovery carried out at different temperatures (see legend). Stress/Recovery conditions are $(V_{GS}, V_{DS}, V_{SUB}) = (-8, 25, 0)$ V, and $(V_{GS}, V_{DS}, V_{BS}) = (0, 0.5, 0)$ V, respectively. Experimental Data is taken from [56]; simulation results are from [60], [61].

instead, R_{ON} was monitored throughout the simulation as recovery and measurement conditions were the same. R_{ON} results in Figs. 4.10 and 4.11 are normalized with respect to the fresh value at each temperature to purify results from the RON degradation induced by mobility reduction. Recovery tests were performed immediately after the stress phase was completed. Since R_{ON} measurement takes about 50 ms [56] no measurement data points were acquired for recovery time less than 100 ms. For both FGOS and BGOS, simulation results can reproduce reasonably well the essential features shown by the experimental results taken at different temperatures. That is, simulations capture the thermally activated processes at the basis of both R_{ON} degradation and

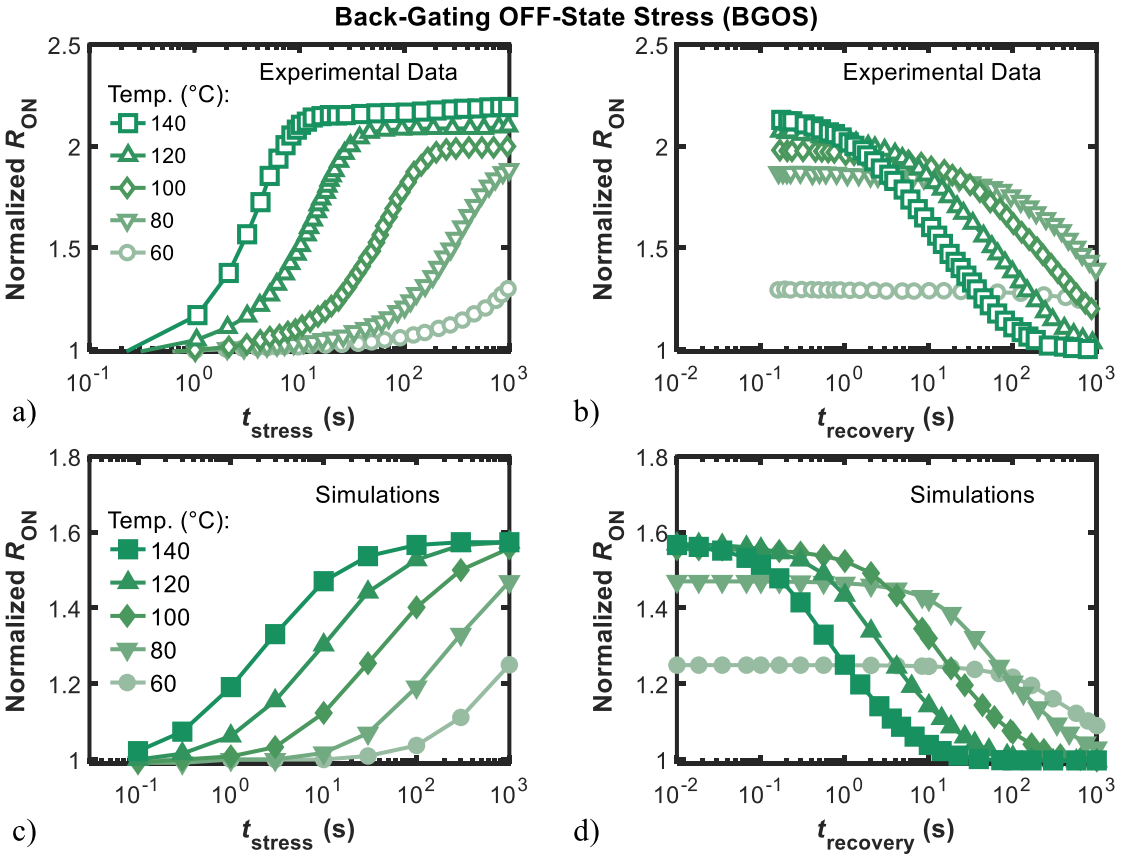


Figure 4.11: R_{ON} variations (normalized w.r.t. the pre-stress value) during BGOS (a), (c) and (b), (d) consequent recovery carried out at different temperatures (see legend). Stress/Recovery conditions are $(V_{GS}, V_{DS}, V_{SUB}) = (0, 0, -25)$ V, and $(V_{GS}, V_{DS}, V_{BS}) = (0, 0.5, 0)$ V, respectively. Experimental Data is taken from [56]; simulation results are from [60], [61].

recovery, as well as the time constant ranges. As shown in [56], stress and recovery transients are found to be thermally activated with similar E_A in the range 0.84-0.95 eV, irrespective of the stress condition. We report the experimental Arrhenius plots shown in [56] (for stress only) in Fig. 4.12 for both FGOS and BGOS conditions and compare them with simulation results (in this case showing both stress and recovery). The time constants at each temperature for both experiments and simulations were extracted by fitting the curves with the stretched exponential method [59]. As it can be noted, the Arrhenius signature of the stress process is well reproduced by our simulations and, more importantly, simulations predict the same $E_A \approx 0.9$ eV for both

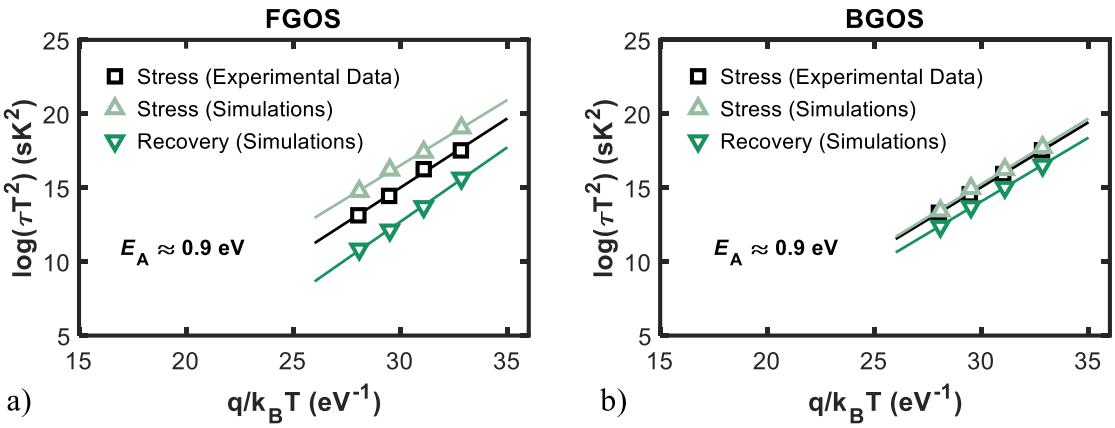


Figure 4.12: Arrhenius plot for the simulated (a) FGOS and (b) BGOS transients as well as the relative recovery processes. Experimental data from [56] of stress processes are reported for comparison. Lines are the linear fitting of the data showing that both experiments and simulations are characterized by the same $E_A \approx 0.9\,\text{eV}$. Adapted from [61].

stress and recovery in either FGOS/BGOS conditions. The fact that both emission and capture processes are thermally activated and, more importantly, exhibit the E_A is not trivial. Indeed, while carrier emission is always a thermally activated process, carrier capture can only be thermally activated in traps that feature a capture barrier, although the associated E_A is generally different (and smaller) than the emission one [41]. In the case of the devices considered here, the extracted E_A (for both stress and recovery) correlates very well with the transition energy of the dominant acceptor level (C_N) related to Carbon in GaN.

Hole Emission, Trapping and Re-Emission in the Buffer

We now show that the R_{ON} experiments described previously can be explained with a hole emission, redistribution, and re-trapping model ('hole redistribution' in short). The model applies to both FGOS and BGOS thus we focus only on the latter case for simplicity of data presentation, as under back-gating conditions the buffer is uniformly exposed to the backside bias and all internal quantities are characterized by an almost one-dimensional distribution (along the device depth direction).

We explain the processes occurring during stress and recovery with the aid of the plots of the

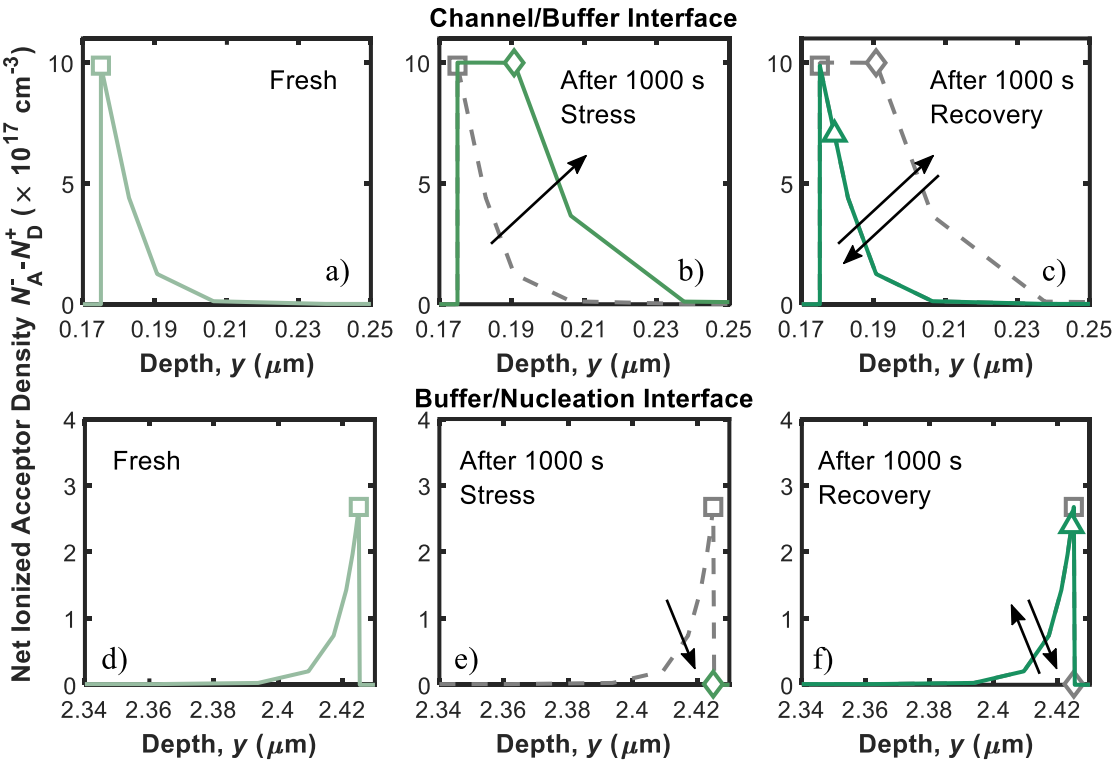


Figure 4.13: Net ionized acceptor trap density, $(N_A^- - N_D^+)$, along the vertical direction in the Gate-to-Drain access region at different conditions, namely: fresh (a), (d), after 1000 s stress (b), (e), and after 1000 s recovery (c), (f) near the channel/buffer interface (a)- (c) and near the buffer/nucleation interface (d)- (f). BGOS conditions were applied ($T = 100\,°\text{C}$). Adapted from [61].

net ionized acceptor trap density, $(N_A^- - N_D^+)$, and of the free hole density, p, shown in Fig. 4.13. These plots are taken along a cutline parallel to the device depth drawn in the middle of the gate-to-drain access region and zoomed in at the top and bottom regions of the buffer, i.e., at the channel/buffer and nucleation/buffer interface, respectively. During stress, $(N_A^- - N_D^+)$ (i.e., net negative charge) in the top region of the buffer close to the channel increases because holes are being emitted from the C-related acceptor traps at $E_V + 0.9\,\text{eV}$. This correlates with the observed R_{ON} increase during stress. The variation in $(N_A^- - N_D^+)$ close to the channel is evident by comparing the cases before and after stress in Fig. 4.13(a) and (b). The hole emission process is thermally activated with $E_A = 0.9\,\text{eV}$, see Fig. 4.12(b). The holes being emitted at the top region of the buffer (compare Fig. 4.14(a) and (b)), drift towards the bottom edge of the buffer

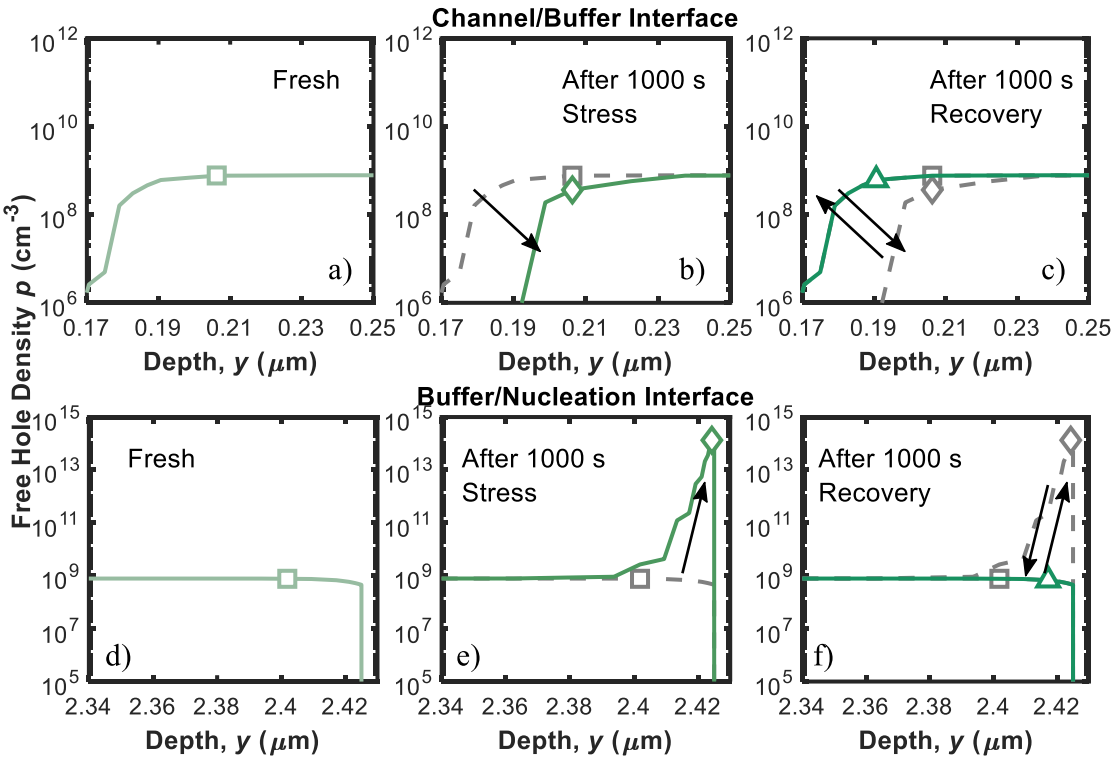


Figure 4.14: Free hole density, p, along the vertical direction in the Gate-to-Drain access region at different conditions, namely: fresh (a), (d), after 1000 s stress (b), (e), and after 1000 s recovery (c), (f) near the channel/buffer interface (a)- (c) and near the buffer/nucleation interface (d)- (f). BGOS conditions were applied ($T = 100\,°\text{C}$). Adapted from [61].

attracted by the negative V_{SUB}, and accumulate at the buffer/nucleation layer interface, compare Fig. 4.14(d) and (e). The free holes get partially trapped and thus discharge the negatively charged acceptor traps at the same interface, compare Fig. 4.13(d) and (e). During recovery, all processes described above are inverted. Holes are emitted from C-related acceptors in the bottom region of the buffer, see Fig. 4.14(f), and drift back towards the channel/buffer interface, see Fig. 4.14(c), where they get re-trapped by the acceptor states from which they were emitted during stress. This process decreases $(N_A^- - N_D^+)$ at the channel/buffer interface as shown in Fig. 4.13(c), thus explaining the R_{ON} decrease during recovery. The hole re-trapping process during recovery is also thermally activated with $E_A = 0.9\,\text{eV}$, since the re-trapped holes in the upper part of the buffer need to be emitted from the C-related acceptor traps in the bottom region of the buffer,

see Fig. 4.13(f). After 1000 s of recovery, the state prior to stress is fully restored (as testified by the results in Fig. 4.11 and consequently the *p* peak at the bottom of the buffer disappears, see Fig. 4.14(f). The recovery phase fully restores the pre-stress value of R_{ON} because the holes emitted during stress do not leak out from the device contacts or recombine with electrons.

4.4.2 Bidirectional Threshold Voltage Instability After Negative Gate Bias Stress in Carbon-doped Fully Recessed AlGaN/GaN MIS-HEMTs

Fully recessed MIS-HEMTs devices aim at reduing gate leakage current and allow for normally-off (i.e., $V_t > 0$) operation. However, the presence of a gate oxide and the associated defect-prone oxide/semiconductor interface leads to V_t instability, severely impacting device stability and operation [3]. Assessing the V_t stability after NBTI stress is important even for normally-off MIS-HEMTs, since $V_{GS} \ll V_t$ is required to avoid false turn-on when devices are pulsed at large drain voltages [45]. Moreover, if large $V_{GS} < 0$ are adopted, NBTI tests constitute a proxy for OFF-state conditions (experienced by the device under pulse-mode operation) as similar, large values of drain-gate voltage can be achieved with both stress setups. The advantage of NBTI stress setups is that V_t instability can be separated from other dispersion effects (such as dynamic R_{ON} degradation).

Different V_t instability trends after NBTI stress were observed in the literature depending on the sign of the observed V_t shifts (ΔV_t). Three main categories can be identified as follows: *i)* $\Delta V_t < 0\,\mathrm{V}$ [37], [62], [63]; *ii)* $\Delta V_t > 0\,\mathrm{V}$ [34], [45], [64]; and *iii)* ΔV_t of both signs, depending on device type or stress conditions [29], [33]. $\Delta V_t < 0\,\mathrm{V}$ following NBTI is generally attributed to the emission of electrons from traps in the gate oxide or at the interface with the semiconductor. On the other hand, the observed $\Delta V_t > 0\,\mathrm{V}$ as a consequence of high negative gate bias stress still lacks of a univocal interpretation. Moreover, $\Delta V_t > 0\,\mathrm{V}$ was also observed in off-state operating conditions, attributed to a hole generation mechanism [64], [65]. For this reason, it is worthwhile performing an in-depth analysis of V_t stability after NBTI stress by exploring a wide range of stress conditions to provide a comprehensive explanation for the observed results.

In this section, we present experimental data on V_t stability in fully-recessed AlGaN/GaN MIS-HEMTs after NBTI under different stress conditions, namely: *i))* gate bias ($V_{GS,STR}$); *ii)* stress time

(t_{STR}); and *iii)* temperature (T). Depending on the $V_{GS,STR}$ magnitude, we observed bidirectional ΔV_T (i.e., negative and positive) thermally activated with different E_A [70]. Particularly, we observed $\Delta V_T < 0\,\mathrm{V}$ (with $E_A \approx 0.5\,\mathrm{eV}$) under moderate stress (i.e., $V_{GS,STR} = -7\,\mathrm{V}$) and $\Delta V_T > 0\,\mathrm{V}$ (with $E_A \approx 0.9\,\mathrm{eV}$) with high stress (i.e., $V_{GS,STR} = -17\,\mathrm{V}$). In accordance with the literature, we attribute the former behavior (i.e., $\Delta V_T < 0\,\mathrm{V}$) to the emission of electrons from traps located in the gate oxide and/or at the interface with the semiconductor [29], [37], [62]. We attribute the latter behavior (i.e., $\Delta V_T > 0\,\mathrm{V}$) to the ionization of buffer acceptor traps and the consequent recombination of holes (emitted by these traps) with electrons injected from the gate terminal. We also present TCAD simulation results to support the above interpretation.

Experimental Results

The Devices-Under-Test (DUTs) were fully recessed AlGaN/GaN MIS-HEMTs grown by MOVPE on a Si(111) p-type substrate with 1 µm gate length. The schematic view of such devices is shown in Fig. 4.3c. The channel is composed of a Low-Carbon-doped (LC, $\sim 1 \times 10^{16}\,\mathrm{cm}^{-3}$) GaN layer of 0.3 µm. The GaN buffer layer instead is 4.7 µm thick and is highly C-doped ($\sim 1 \times 10^{18}\,\mathrm{cm}^{-3}$). The AlGaN barrier layer is 20 nm with 25% Al concentration. The gate dielectric (20 nm Al_2O_3) is fully recessed and penetrates 2 nm into the LC GaN.

The NBTI OTF characterization was carried out by applying a finite gate bias while keeping all other contacts grounded. DUT temperature was set by means of a thermal chuck. During stress, the device is periodically turned-on by means of a stairway-like gate signal to monitor the V_t evolution [63]. To this end, the gate is swept between $-2\,\mathrm{V}$ and $2\,\mathrm{V}$ with steps of $0.1\,\mathrm{V}$ each lasting $10\,\mu\mathrm{s}$, while the drain is fixed to $0.1\,\mathrm{V}$. In correspondence of each step, I_D and V_{GS} are acquired, allowing the reconstruction of the $I-V$ characteristics. Since the induced stress was monitored for up to $t_{STR} = 1000\,\mathrm{s}$, the gate stairway signal was applied to the DUT at logarithmically-spaced time instants to ensure a uniform distribution of V_t acquisitions [66]. The chosen $I-V$ sweep rate represents the best compromise between speed and accuracy, since the short time required for the V_t measurement minimizes the fast recovery between stress phases [63] but still provides enough samples for a correct acquisition. For each $I-V$ curve, V_t was extracted with the constant-current method (i.e., $I_D = 200\,\mu\mathrm{A}$). ΔV_t is then simply calculated

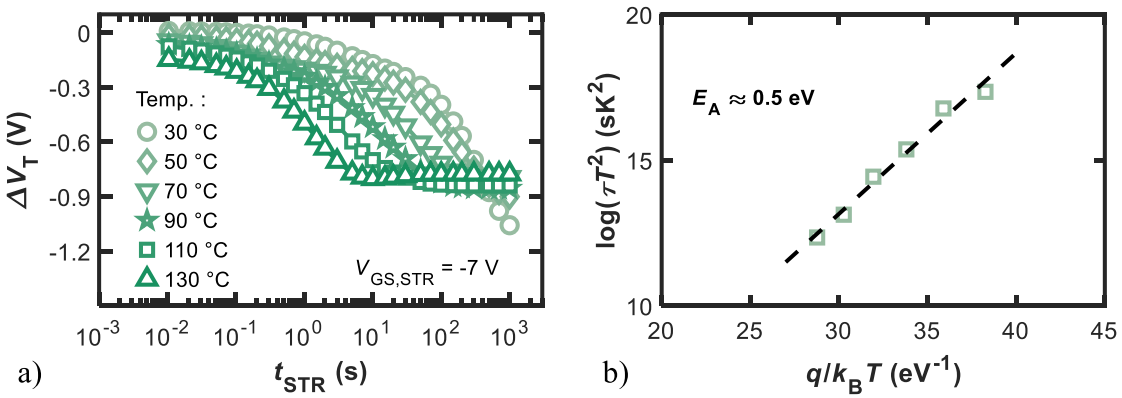


Figure 4.15: (a) Measured ΔV_t vs t_{STR} for different temperature (see legend) at $V_{GS,STR} = -7\,\mathrm{V}$. (b) Arrhenius plot of the stress data showing that $E_A \approx 0.5\,\mathrm{eV}$. Adapted from [70].

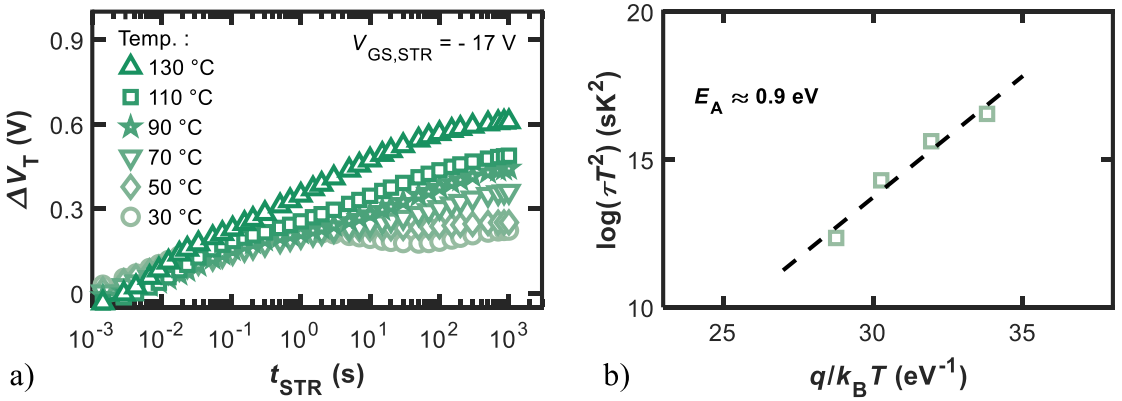


Figure 4.16: (a) Measured ΔV_t vs t_{STR} for different temperature (see legend) at $V_{GS,STR} = -17\,\mathrm{V}$. (b) Arrhenius plot of the stress data showing that $E_A \approx 0.9\,\mathrm{eV}$. Adapted from [70].

as $(V_t - V_{t,0})$ with the reference $V_{t,0}$ acquired before starting the stress phase from a fresh $I-V$ curve.

Two different stress conditions were investigated in this work, i.e., moderate ($V_{GS,STR} = -7\,\mathrm{V}$) and high stress ($V_{GS,STR} = -17\,\mathrm{V}$), observing ΔV_t of both signs. The measured ΔV_t evolution vs t_{STR} at different T is shown Figs. 4.15 and 4.16 for the moderate and high stress conditions, respectively. Fig. 4.15 shows that, for moderate stress, V_t drifts to the negative direction (i.e., becomes less than $V_{t,0}$) with a minimum ΔV_t in the order of $-0.8/-1\,\mathrm{V}$ at $t_{STR} = 1000\,\mathrm{s}$. This effect is widely reported in the literature and is attributed to the emission of electrons from traps

located in the gate oxide or at the interface with the semiconductor [29], [62], which is known to be a thermally activated process [37]. In the devices tested in this work, the E_A of this process was estimated to be $\approx 0.5\,\text{eV}$, see Fig. 4.15(b). This process was also related to the de-ionization of acceptor traps in the buffer (related to C-doping) as a consequence of the applied negative gate stress [63]. According to our simulations, this second possible contribution to the negative ΔV_t is small compared to the former in the devices considered here.

Conversely, Fig. 4.16 shows that the high stress condition causes V_t to drift in the positive direction (i.e., V_t becomes larger than $V_{t,0}$). The positive drift is found to be weakly dependent on temperature for t_{STR} in the range of 1 ms to 1 s, whereas it exhibits a clear temperature-dependent component for $t_{STR} > 1\,\text{s}$. The E_A of the latter process was found to be $\approx 0.9\,\text{eV}$, see Fig. 4.16(b). Note that because no additional V_t shift was observed for $t_{STR} > 1\,\text{s}$ at $T \leq 50\,^{\circ}\text{C}$, extraction of E_A from the linear fitting excluded data in this temperature range. The fast $\Delta V_t > 0\,\text{V}$ can be attributed to the capture of electrons injected by the gate terminal in the oxide [34], whereas the thermally activated increase is induced by a slower process. $E_A \approx 0.9\,\text{eV}$ suggests that this behavior can be attributed to buffer acceptor traps (related to C-doping) that are energetically located at about $E_V + 0.9\,\text{eV}$ [53]. In fact, buffer traps can increase their ionization rate under high $|V_{GS,STR}|$ due to the high recombination rate between electrons injected from the gate terminal and the holes emitted by the acceptor traps themselves [68].

Simulation Results and Physical Interpretation

To verify the validity of these claims, we performed numerical 2D device simulations with SDevice [14, Ch. 2]. The simulation deck was calibrated against results from the tested devices [70]. The same TCAD setup described in Section 4.3.1 (adapted to the measured DUTs) was used. Particularly, one SRH trap-balance equation was set for each distinct trap level included in the simulation. The dynamics of trap occupation was thus modelled without any quasi-static approximation. For simplicity, traps in the gate oxide were projected to a single energy level ($E_T = 0.5\,\text{eV}$ from E_C) at the interface with the semiconductor, with a concentration of $1 \times 10^{12}\,\text{cm}^{-2}$. Carbon-related traps in the GaN buffer were modeled with a dominant deep acceptor trap at $E_V + 0.9\,\text{eV}$ and a shallow donor trap at $E_C - 0.11\,\text{eV}$ [52]. The effective acceptor

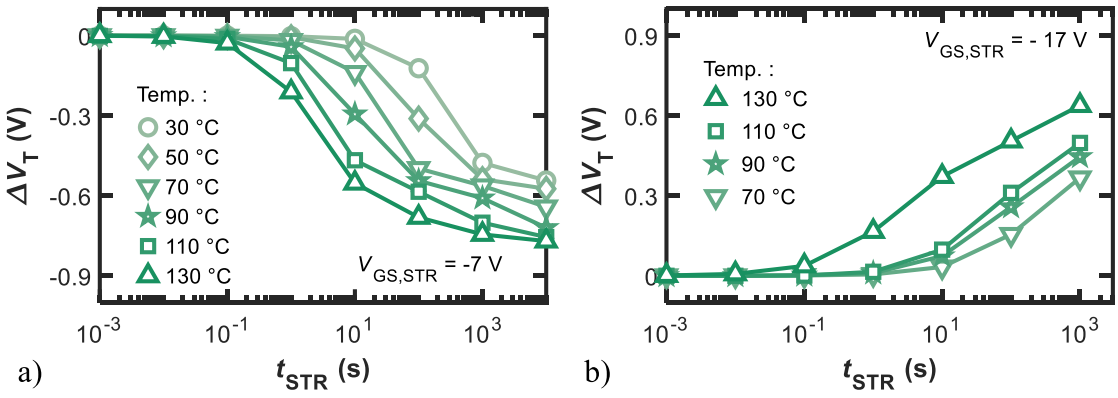


Figure 4.17: Simulated ΔV_t vs t_{STR} after (a) moderate and (b) high $V_{GS,STR}$ at different temperature (see legend). The negative and positive ΔV_t with different gate bias stress – as well as its temperature-dependent behavior – observed in the DUTs can be correctly captured. Adapted from [70].

density in the buffer was $3 \times 10^{16}\,\mathrm{cm}^{-3}$ (about 1% of the nominal concentration [67]).

Simulations results are shown in Fig. 4.17 for (a) moderate and (a) high $V_{GS,STR}$. Note that V_t was extracted with the same constant-current method as that used for the measured data. In the first case, negative ΔV_t is explained by simulations as the result of electron emission from interface traps. Note that the energy level of these traps was set to 0.5 eV in the simulations to reproduce the temperature dependence observed in the experiments as shown in Fig. 4.15. As mentioned previously, C-related traps contribute only to a small extent to the *negative* ΔV_t. However, the emission from these traps of holes attracted by the negative gate potential lead to a saturation of ΔV_t as also observed in the experiments. This saturation effect stems from the accumulation of holes under the gate oxide/channel interface that prevents the further increase of band bending and consequently of the negative V_t drift. In fact, we observed that when removing the buffer traps in the simulations, ΔV_t would saturate at much larger negative values, consistent with the hypothesis that accumulation of holes under the gate emitted by C-related traps prevent the further drift of V_t.

In the second case, seeFig. 4.17(b), *positive* ΔV_t could be reproduced in the simulations by assuming that electrons leaking through the gate dielectric would recombine with holes emitted from the buffer acceptor traps attracted to the surface by the large negative $V_{GS,STR}$. This

recombination leads to the increase in negatively charged traps in the buffer (under the gate), making V_t increase [68]. This is confirmed by simulation results obtained by assuming the gate insulator to be ideal (i.e., with no electron leakage) resulting in $\Delta V_t < 0$ even for the high stress conditions ($V_{GS,STR} = -17\,\mathrm{V}$).

We point out that the simulation results in Fig. 4.17(b) should only be compared to the experiments for $t_{STR} > 1\,\mathrm{s}$, see Fig. 4.16(a). This is because the observed weakly T-dependent $\Delta V_t > 0$ for $t_{STR} < 1\,\mathrm{s}$ is likely due to the capture of electrons into traps in the leaky gate dielectric [34]. Thus, the mechanism leading to the fast and weakly T-dependent $\Delta V_t > 0$ could be related to the conduction of electrons in the Al_2O_3 (with low $E_A \approx 0.045\,\mathrm{eV}$ [69]). On the other hand, the thermally activated process observed in the 100-1000 s range is reproduced by the simulations, suggesting that C-related traps in the buffer cause this behavior. Simulations results for $T = 30, 50\,^{\circ}\mathrm{C}$ are not shown because the slow T-dependent process is negligible in this case, as discussed previously. Summarizing, for $V_{GS,STR} \ll 0\,\mathrm{V}$, the thermally activated emission of holes from buffer traps and the removal of positive charge from the device due to hole recombination with gate-injected electrons gives rise to a net V_t increase.

As a final remark, it is worth pointing out that while we focused on NBTI to isolate ΔV_t from other possible effects (such as R_{ON} drift), the phenomena leading to $\Delta V_t > 0$ can also be induced by typical OFF-state stress conditions (i.e., $V_{GS,STR} < V_t$, $V_{DS,STR} \gg 0\,\mathrm{V}$) [64], [65] even though concentrated at the gate edge on the drain side.

4.4.3 Conclusions

Concerning the influence of buffer traps on dynamic R_{ON} behavior, we presented a 'hole re-distribution' model explaining the stress/recovery experiments in C-doped AlGaN/GaN power MIS-HEMTs. Within the framework of the model, stress is ascribed to hole emission from C-related acceptor traps close to the channel/buffer interface that redistribute and get trapped in the same type of traps in the bottom region of the buffer close to the buffer/nucleation interface. During recovery, the opposite process takes place: the previously trapped holes in the bottom part of the buffer are emitted and get re-trapped by the same traps at the top of the buffer that emitted them during stress. The proposed model: *i)* solves the puzzle of the activation energy

being the same in both stress/recovery phases, and *ii)* explains the full recovery of the R_{ON} after the complete stress-and-recovery cycle in the analyzed devices.

Concerning the influence of buffer traps on V_t instability, we performed an in-depth investigation of NBTI behavior in fully recessed AlGaN/GaN MIS-HEMTs. Results obtained by means of on-the-fly characterization under relatively low negative gate bias stress (i.e., $V_{GS,STR} = -7\,\mathrm{V}$) revealed conventional negative V_t drift associated to the thermally activated emission of electrons from interface and/or oxide traps. The positive V_t shift observed for $V_{GS,STR} = -17\,\mathrm{V}$ for long stress time was instead ascribed to the increased negative ionized acceptor trap density in the buffer and the recombination of electrons injected from the gate terminal with holes emitted from C-related traps.

References – Chapter 4

[1] T. Kimoto and J. A. Cooper, *Fundamentals of Silicon Carbide Technology: Growth, Characterization, Devices and Applications*. Wiley, 2014, pp. 1–538. DOI: 10.1002/9781118313534.

[2] K. J. Chen, O. Haberlen, A. Lidow, C. L. Tsai, T. Ueda, Y. Uemoto, and Y. Wu, "GaN-on-Si power technology: Devices and applications," *IEEE Transactions on Electron Devices*, vol. 64, no. 3, pp. 779–795, Mar. 2017. DOI: 10.1109/TED.2017.2657579.

[3] J. A. Del Alamo and E. S. Lee, "Stability and Reliability of Lateral GaN Power Field-Effect Transistors," *IEEE Transactions on Electron Devices*, vol. 66, no. 11, pp. 4578–4590, Nov. 2019. DOI: 10.1109/TED.2019.2931718.

[4] H. Amano, Y. Baines, E. Beam, M. Borga, T. Bouchet, P. R. Chalker, M. Charles, K. J. Chen, N. Chowdhury, R. Chu, C. De Santi, M. M. De Souza, S. Decoutere, L. Di Cioccio, B. Eckardt, T. Egawa, P. Fay, J. J. Freedsman, L. Guido, O. Häberlen, G. Haynes, T. Heckel, D. Hemakumara, P. Houston, J. Hu, M. Hua, Q. Huang, A. Huang, S. Jiang, H. Kawai, D. Kinzer, M. Kuball, A. Kumar, K. B. Lee, X. Li, D. Marcon, M. März, R. McCarthy, G. Meneghesso, M. Meneghini, E. Morvan, A. Nakajima, E. M. Narayanan, S. Oliver, T. Palacios, D. Piedra, M. Plissonnier, R. Reddy, M. Sun, I. Thayne, A. Torres, N. Trivellin, V. Unni, M. J. Uren, M. Van Hove, D. J. Wallis, J. Wang, J. Xie, S. Yagi, S. Yang, C. Youtsey, R. Yu, E. Zanoni, S. Zeltner, and Y. Zhang, "The 2018 GaN power electronics roadmap," *Journal of Physics D: Applied Physics*, vol. 51, no. 16, p. 163 001, 2018. DOI: 10.1088/1361-6463/aaaf9d.

[5] K. Shenai, "High-Density Power Conversion and Wide-Bandgap Semiconductor Power Electronics Switching Devices," *Proceedings of the IEEE*, vol. 107, no. 12, pp. 2308–2326, Dec. 2019. DOI: 10.1109/JPROC.2019.2948554.

[6] M. Higashiwaki, K. Sasaki, H. Murakami, Y. Kumagai, A. Koukitu, A. Kuramata, T. Masui, and S. Yamakoshi, "Recent progress in Ga2O3 power devices," *Semiconductor Science and Technology*, vol. 31, no. 3, p. 34 001, 2016. DOI: 10.1088/0268-1242/31/3/034001.

[7] U. K. Mishra and J. Singh, "Field Effect Transistors," in *Semiconductor Device Physics and Design*, Springer, 2008, pp. 356–432. DOI: 10.1007/978-1-4020-6481-4.

[8] M. Asif Khan, A. Bhattarai, J. N. Kuznia, and D. T. Olson, "High electron mobility transistor based on a GaN-AlxGa 1-xN heterojunction," *Applied Physics Letters*, vol. 63, no. 9, pp. 1214–1215, 1993, ISSN: 00036951. DOI: 10.1063/1.109775.

[9] M. Van Hove, S. Boulay, S. R. Bahl, S. Stoffels, X. Kang, D. Wellekens, K. Geens, A. Delabie, and S. Decoutere, "CMOS Process-Compatible High-Power Low-Leakage AlGaN/GaN MISHEMT on Silicon," *IEEE Electron Device Letters*, vol. 33, no. 5, pp. 667–669, May 2012. DOI: 10.1109/LED.2012.2188016.

[10] Y. Uemoto, M. Hikita, H. Ueno, H. Matsuo, H. Ishida, M. Yanagihara, T. Ueda, T. Tanaka, and D. Ueda, "Gate injection transistor (GIT) - A normally-off AlGaN/GaN power transistor using conductivity modulation," *IEEE Transactions on Electron Devices*, vol. 54, no. 12, pp. 3393–3399, 2007. DOI: 10.1109/TED.2007.908601.

[11] A. Fariza, A. Lesnik, J. Bläsing, M. P. Hoffmann, F. Hörich, P. Veit, H. Witte, A. Dadgar, and A. Strittmatter, "On reduction of current leakage in GaN by carbon-doping," *Applied Physics Letters*, vol. 109, no. 21, p. 212 102, 2016. DOI: 10.1063/1.4968823.

[12] E. Bahat-Treidel, F. Brunner, O. Hilt, E. Cho, J. Würfl, and G. Trankle, "AlGaN/GaN/-GaN:C back-barrier HFETs with breakdown voltage of over 1 kV and low RON× A," *IEEE Transactions on Electron Devices*, vol. 57, no. 11, pp. 3050–3058, Nov. 2010. DOI: 10.1109/TED.2010.2069566.

[13] M. J. Uren, J. Moreke, and M. Kuball, "Buffer design to minimize current collapse in GaN/AlGaN HFETs," *IEEE Transactions on Electron Devices*, vol. 59, no. 12, pp. 3327–3333, 2012. DOI: 10.1109/TED.2012.2216535.

[14] H. Y. Wang, K. D. Mai, L. Y. Peng, Y. H. Cheng, and H. C. Chiu, "Effects of the Fe-doped GaN buffer in AlGaN/GaN HEMTs on SiC substrate," in *IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC)*, vol. 7, IEEE, Jun. 2015, pp. 645–648. DOI: 10.1109/EDSSC.2015.7285198.

[15] D. K. Schroder, "Defects," in *Semiconductor Material and Device Characterization*, Hoboken, NJ, USA: John Wiley & Sons, Inc., Apr. 2005, pp. 251–317. DOI: 10.1002/0471749095.ch5.

[16] R. Vetury, N. Q. Zhang, S. Keller, and U. K. Misha, "The impact of surface states on the DC and RF characteristics of AlGaN/GaN HFETs," *IEEE Transactions on Electron Devices*, vol. 48, no. 3, pp. 560–566, Mar. 2001. DOI: 10.1109/16.906451.

[17] W. S. Tan, M. J. Uren, P. A. Houston, R. T. Green, R. S. Balmer, and T. Martin, "Surface leakage currents in SiNx passivated AlGaN/GaN HFETs," *IEEE Electron Device Letters*, vol. 27, no. 1, pp. 1–3, 2006. DOI: 10.1109/LED.2005.860383.

[18] G. Meneghesso, G. Verzellesi, F. Danesin, F. Rampazzo, F. Zanon, A. Tazzoli, M. Meneghini, and E. Zanoni, "Reliability of GaN high-electron-mobility transistors: State of the art and perspectives," *IEEE Transactions on Device and Materials Reliability*, vol. 8, no. 2, pp. 332–343, 2008, ISSN: 15304388. DOI: 10.1109/TDMR.2008.923743.

[19] I. Hwang, J. Kim, S. Chong, H. S. Choi, S. K. Hwang, J. Oh, J. K. Shin, and U. I. Chung, "Impact of channel hot electrons on current collapse in AlGaN/GaN HEMTs," *IEEE Electron Device Letters*, vol. 34, no. 12, pp. 1494–1496, 2013. DOI: 10.1109/LED.2013.2286173.

[20] E. Zanoni, F. Danesin, M. Meneghini, A. Cetronio, C. Lanzieri, M. Peroni, and G. Meneghesso, "Localized damage in AlGaN/GaN HEMTs induced by reverse-bias testing," *IEEE Electron Device Letters*, vol. 30, no. 5, pp. 427–429, 2009, ISSN: 07413106. DOI: 10.1109/LED.2009.2016440.

[21] J. Joh and J. A. Del Alamo, "Impact of electrical degradation on trapping characteristics of GaN high electron mobility Transistors," in *IEEE International Electron Devices Meeting (IEDM)*, 2008, pp. 5–8. DOI: 10.1109/IEDM.2008.4796725.

[22] M. Meneghini, A. Stocco, R. Silvestri, G. Meneghesso, and E. Zanoni, "Degradation of AlGaN/GaN high electron mobility transistors related to hot electrons," *Applied Physics Letters*, vol. 100, no. 23, pp. 1–4, 2012. DOI: 10.1063/1.4723848.

[23] M. Grupen, "Reproducing GaN HEMT Kink Effect by Simulating Field-Enhanced Barrier Defect Ionization," *IEEE Transactions on Electron Devices*, vol. 66, no. 9, pp. 3777–3783, 2019. DOI: 10.1109/TED.2019.2928536.

[24] M. J. Uren, K. J. Nash, R. S. Balmer, T. Martin, E. Morvan, N. Caillas, S. L. Delage, D. Ducatteau, B. Grimbert, and J. C. De Jaeger, "Punch-through in short-channel AlGaN/GaN HFETs," *IEEE Transactions on Electron Devices*, vol. 53, no. 2, pp. 395–398, Feb. 2006. DOI: 10.1109/TED.2005.862702.

[25] W. Saito, M. Kuraguchi, Y. Takada, K. Tsuda, I. Omura, and T. Ogura, "Influence of surface defect charge at AlGaN-GaN-HEMT upon Schottky gate leakage current and breakdown voltage," *IEEE Transactions on Electron Devices*, vol. 52, no. 2, pp. 159–164, 2005. DOI: 10.1109/TED.2004.842710.

[26] C. Zhou, Q. Jiang, S. Huang, and K. J. Chen, "Vertical leakage/breakdown mechanisms in AlGaN/GaN-on-Si structures," in *2012 24th International Symposium on Power Semiconductor Devices and ICs*, vol. 33, IEEE, 2012, pp. 245–248. DOI: 10.1109/ISPSD.2012.6229069.

[27] D. Jin and J. A. Del Alamo, "Impact of high-power stress on dynamic ON-resistance of high-voltage GaN HEMTs," *Microelectronics Reliability*, vol. 52, no. 12, pp. 2875–2879, 2012. DOI: 10.1016/j.microrel.2012.08.023.

[28] D. Jin, J. Joh, S. Krishnan, N. Tipirneni, S. Pendharkar, and J. A. del Alamo, "Total current collapse in high-voltage GaN MIS-HEMTs induced by Zener trapping," in *IEEE International Electron Devices Meeting (IEDM)*, IEEE, Dec. 2013, pp. 6.2.1–6.2.4. DOI: 10.1109/IEDM.2013.6724572.

[29] A. Guo and J. A. Del Alamo, "Negative-bias temperature instability of GaN MOSFETs," in *IEEE International Reliability Physics Symposium Proceedings*, 2016, 4A11–4A16. DOI: 10.1109/IRPS.2016.7574526.

[30] X. H. Ma, M. Lü, L. Pang, Y. Q. Jiang, J. Z. Yang, W. W. Chen, and X. Y. Liu, "Kink effect in current-voltage characteristics of a GaN-based high electron mobility transistor with an AlGaN back barrier," *Chinese Physics B*, vol. 23, no. 2, 2014. DOI: 10.1088/1674-1056/23/2/027302.

[31] M. J. Uren, M. Silvestri, M. Casar, G. A. M. Hurkx, J. A. Croon, J. Sonsky, and M. Kuball, "Intentionally carbon-doped algan/gan HEMTs: Necessity for vertical leakage paths," *IEEE Electron Device Letters*, vol. 35, no. 3, pp. 327–329, 2014. DOI: 10.1109/LED.2013.2297626.

[32] M. J. Uren, S. Karboyan, I. Chatterjee, A. Pooth, P. Moens, A. Banerjee, and M. Kuball, ""Leaky Dielectric" Model for the Suppression of Dynamic RON in Carbon-Doped AlGaN/GaN HEMTs," *IEEE Transactions on Electron Devices*, vol. 64, no. 7, pp. 2826–2834, Jul. 2017. DOI: 10.1109/TED.2017.2706090.

[33] F. Sang, M. Wang, C. Zhang, M. Tao, B. Xie, C. P. Wen, J. Wang, Y. Hao, W. Wu, and B. Shen, "Investigation of the threshold voltage drift in enhancement mode GaN MOSFET under negative gate bias stress," *Japanese Journal of Applied Physics*, vol. 54, no. 4, p. 044 101, Apr. 2015, ISSN: 13474065. DOI: 10.7567/JJAP.54.044101.

[34] J. He, M. Hua, Z. Zhang, and K. J. Chen, "Performance and VTH Stability in E-Mode GaN Fully Recessed MIS-FETs and Partially Recessed MIS-HEMTs with LPCVD-SiNx/PECVD-SiNx Gate Dielectric Stack," *IEEE Transactions on Electron Devices*, vol. 65, no. 8, pp. 3185–3191, 2018. DOI: 10.1109/TED.2018.2850042.

[35] P. Lagger, C. Ostermaier, G. Pobegen, and D. Pogany, "Towards understanding the origin of threshold voltage instability of AlGaN/GaN MIS-HEMTs," in *IEEE International Electron Devices Meeting (IEDM)*, IEEE, 2012, pp. 299–302. DOI: 10.1109/IEDM.2012.6479033.

[36] A. Guo and J. A. Del Alamo, "Positive-bias temperature instability (PBTI) of GaN MOSFETs," in *IEEE International Reliability Physics Symposium Proceedings*, 2015, pp. 6C51–6C57. DOI: 10.1109/IRPS.2015.7112770.

[37] M. Meneghini, I. Rossetto, D. Bisi, M. Ruzzarin, M. Van Hove, S. Stoffels, T.-L. Wu, D. Marcon, S. Decoutere, G. Meneghesso, and E. Zanoni, "Negative Bias-Induced Threshold Voltage Instability in GaN-on-Si Power HEMTs," *IEEE Electron Device Letters*, vol. 37, no. 4, pp. 474–477, Apr. 2016. DOI: 10.1109/LED.2016.2530693.

[38] Y. Shi, Q. Zhou, Q. Cheng, P. Wei, L. Zhu, D. Wei, A. Zhang, W. Chen, and B. Zhang, "Bidirectional threshold voltage shift and gate leakage in 650 v p-GaN AlGaN/GaN HEMTs: The role of electron-Trapping and hole-injection," *Proceedings of the International Symposium on Power Semiconductor Devices and ICs*, pp. 96–99, 2018. DOI: 10.1109/ISPSD.2018.8393611.

[39] E. Canato, F. Masin, M. Borga, E. Zanoni, M. Meneghini, G. Meneghesso, A. Stockman, A. Banerjee, and P. Moens, "μs-Range Evaluation of Threshold Voltage Instabilities of GaN-on-Si HEMTs with p-GaN Gate," in *2019 IEEE International Reliability Physics Symposium (IRPS)*, IEEE, Mar. 2019, pp. 1–6. DOI: 10.1109/IRPS.2019.8720549.

[40] M. Meneghini, N. Ronchi, A. Stocco, G. Meneghesso, U. K. Mishra, Y. Pei, and E. Zanoni, "Investigation of trapping and hot-electron effects in GaN HEMTs by means of a combined electrooptical method," *IEEE Transactions on Electron Devices*, vol. 58, no. 9, pp. 2996–3003, 2011. DOI: 10.1109/TED.2011.2160547.

[41] D. Bisi, M. Meneghini, F. A. Marino, D. Marcon, S. Stoffels, M. Van Hove, S. Decoutere, G. Meneghesso, and E. Zanoni, "Kinetics of Buffer-Related R_{ON}-Increase in GaN-on-Silicon MIS-HEMTs," *IEEE Electron Device Letters*, vol. 35, no. 10, pp. 1004–1006, Oct. 2014. DOI: 10.1109/LED.2014.2344439.

[42] I. Daumiller, D. Theron, C. Gaquière, A. Vescan, R. Dietrich, A. Wieszt, H. Leier, R. Vetury, U. K. Mishra, I. P. Smorchkova, S. Keller, N. X. Nguyen, C. Nguyen, and E. Kohn, "Current instabilities in GaN-based devices," *IEEE Electron Device Letters*, vol. 22, no. 2, pp. 62–64, Feb. 2001. DOI: 10.1109/55.902832.

[43] J. M. Tirado, J. L. Sánchez-Rojas, and J. I. Izpura, "Trapping effects in the transient response of AlGaN/GaN HEMT devices," *IEEE Transactions on Electron Devices*, vol. 54, no. 3, pp. 410–417, 2007. DOI: 10.1109/TED.2006.890592.

[44] F. Iucolano, A. Parisi, S. Reina, A. Patti, S. Coffa, G. Meneghesso, G. Verzellesi, F. Fantini, and A. Chini, "Correlation between dynamic Rdsou transients and Carbon related buffer traps in AlGaN/GaN HEMTs," *IEEE International Reliability Physics Symposium Proceedings*, pp. CD21–CD24, Sep. 2016, ISSN: 15417026. DOI: 10.1109/IRPS.2016.7574586.

[45] M. Hua, J. Wei, Q. Bao, Z. Zhang, Z. Zheng, and K. J. Chen, "Dependence of VTH Stability on Gate-Bias Under Reverse-Bias Stress in E-mode GaN MIS-FET," *IEEE Electron Device Letters*, vol. 39, no. 3, pp. 413–416, Mar. 2018. DOI: 10.1109/LED.2018.2791664.

[46] M. Singh, M. J. Uren, T. Martin, S. Karboyan, H. Chandrasekar, and M. Kuball, "'Kink' in AlGaN/GaN-HEMTs: Floating buffer model," *IEEE Transactions on Electron Devices*, vol. 65, no. 9, pp. 3746–3753, 2018. DOI: 10.1109/TED.2018.2860902.

[47] "High voltage and high current density vertical GaN power diodes," *Electronics Letters*, vol. 52, no. 13, pp. 1170–1171, 2016. DOI: 10.1049/el.2016.1156.

[48] V. Joshi, S. P. Tiwari, and M. Shrivastava, "Part I: Physical Insight into Carbon-Doping-Induced Delayed Avalanche Action in GaN Buffer in AlGaN/GaN HEMTs," *IEEE Transactions on Electron Devices*, vol. 66, no. 1, pp. 561–569, 2019. DOI: 10.1109/TED.2018.2878770. [Online]. Available: https://ieeexplore.ieee.org/document/8536449/.

[49] V. Joshi, A. Soni, S. P. Tiwari, and M. Shrivastava, "A Comprehensive Computational Modeling Approach for AlGaN/GaN HEMTs," *IEEE Transactions on Nanotechnology*, vol. 15, no. 6, pp. 947–955, Nov. 2016, ISSN: 1536-125X. DOI: 10.1109/TNANO.2016.2615645.

[50] E. Bellotti and F. Bertazzi, "A numerical study of carrier impact ionization in Al xGa 1-xN," *Journal of Applied Physics*, vol. 111, no. 10, p. 103 711, May 2012. DOI: 10.1063/1.4719967.

[51] C. Poblenz, P. Waltereit, S. Rajan, S. Heikman, U. K. Mishra, and J. S. Speck, "Effect of carbon doping on buffer leakage in AlGaN/GaN high electron mobility transistors," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, vol. 22, no. 3, p. 1145, 2004, ISSN: 0734211X. DOI: 10.1116/1.1752907.

[52] A. Armstrong, A. R. Arehart, D. Green, U. K. Mishra, J. S. Speck, and S. A. Ringel, "Impact of deep levels on the electrical conductivity and luminescence of gallium nitride codoped with carbon and silicon," *Journal of Applied Physics*, vol. 98, no. 5, 2005. DOI: 10.1063/1.2005379.

[53] J. L. Lyons, A. Janotti, and C. G. Van De Walle, "Effects of carbon on the electrical and optical properties of InN, GaN, and AlN," *Physical Review B - Condensed Matter and Materials Physics*, vol. 89, no. 3, pp. 1–8, Jan. 2014, ISSN: 10980121. DOI: 10.1103/PhysRevB.89.035204.

[54] N. Zagni, F. Puglisi, P. Pavan, A. Chini, and G. Verzellesi, "Insights into the off-state breakdown mechanisms in power GaN HEMTs," *Microelectronics Reliability*, vol. 100-101, Sep. 2019. DOI: 10.1016/j.microrel.2019.06.066.

[55] N. Zagni, A. Chini, F. M. Puglisi, P. Pavan, and G. Verzellesi, "The Role of Carbon Doping on Breakdown, Current Collapse, and Dynamic On-Resistance Recovery in AlGaN/GaN High Electron Mobility Transistors on Semi-Insulating SiC Substrates," *physica status solidi (a)*, vol. 217, Dec. 2019. DOI: 10.1002/pssa.201900762.

[56] G. Meneghesso, M. Meneghini, I. Rossetto, D. Bisi, S. Stoffels, M. Van Hove, S. Decoutere, and E. Zanoni, "Reliability and parasitic issues in GaN-based power HEMTs: a review," *Semiconductor Science and Technology*, vol. 31, no. 9, p. 093 004, Sep. 2016. DOI: 10.1088/0268-1242/31/9/093004.

[57] D. Bisi, M. Meneghini, M. Van Hove, D. Marcon, S. Stoffels, T.-L. L. Wu, S. Decoutere, G. Meneghesso, and E. Zanoni, "Trapping mechanisms in GaN-based MIS-HEMTs grown on silicon substrate," *physica status solidi (a)*, vol. 212, no. 5, pp. 1122–1129, May 2015. DOI: 10.1002/pssa.201431744.

[58] S. Yang, C. Zhou, Q. Jiang, J. Lu, B. Huang, and K. J. Chen, "Investigation of buffer traps in AlGaN/GaN-on-Si devices by thermally stimulated current spectroscopy and back-gating measurement," *Applied Physics Letters*, vol. 104, no. 1, p. 013 504, Jan. 2014. DOI: 10.1063/1.4861116.

[59] D. Bisi, M. Meneghini, C. De Santi, A. Chini, M. Dammann, P. Bruckner, M. Mikulla, G. Meneghesso, and E. Zanoni, "Deep-level characterization in GaN HEMTs-Part I: Advantages and limitations of drain current transient measurements," *IEEE Transactions on Electron Devices*, vol. 60, no. 10, pp. 3166–3175, 2013. DOI: 10.1109/TED.2013.2279021.

[60] N. Zagni, A. Chini, F. M. Puglisi, P. Pavan, M. Meneghini, G. Meneghesso, E. Zanoni, and G. Verzellesi, "Trap Dynamics Model Explaining the RON Stress/Recovery Behavior in Carbon-Doped Power AlGaN/GaN MOS-HEMTs," in *2020 IEEE International Reliability Physics Symposium (IRPS)*, IEEE, Apr. 2020, pp. 1–5. DOI: 10.1109/IRPS45951.2020.9128816.

[61] N. Zagni, A. Chini, F. M. Puglisi, M. Meneghini, G. Meneghesso, E. Zanoni, P. Pavan, and G. Verzellesi, "'Hole Redistribution' Model Explaining the Thermally Activated RON Stress/Recovery Transients in Carbon-Doped AlGaN/GaN Power MIS-HEMTs," *IEEE Transactions on Electron Devices*, 2021. DOI: 10.1109/TED.2020.3045683.

[62] A. Guo and J. A. del Alamo, "Unified mechanism for positive- and negative-bias temperature instability in gan mosfets," *IEEE Transactions on Electron Devices*, vol. 64, no. 5, pp. 2142–2147, 2017. DOI: 10.1109/TED.2017.2686840.

[63] A. Viey, F. Gaillard, R. Modica, F. Iucolano, M. Meneghini, E. Zanoni, G. Meneghesso, G. Ghibaudo, W. Vandendaele, M.-A. Jaud, J. Cluzel, J.-P. Barnes, S. Martin, A. Krakovinsky, R. Gwoziecki, and M. Plissonnier, "Investigation of nBTI degradation on GaN-on-Si E-mode MOSc-HEMT," in *IEEE International Electron Devices Meeting (IEDM)*, IEEE, Dec. 2019, pp. 4.3.1–4.3.4. DOI: 10.1109/IEDM19573.2019.8993588.

[64] M. Hua, J. Wei, Q. Bao, Z. Zheng, Z. Zhang, J. He, and K. J. Chen, "Hole-induced threshold voltage shift under reverse-bias stress in e-mode gan mis-fet," *IEEE Transactions on Electron Devices*, vol. 65, no. 9, pp. 3831–3838, 2018. DOI: 10.1109/TED.2018.2856998.

[65] M. Hua, S. Yang, J. Wei, Z. Zheng, J. He, and K. J. Chen, "Hole-induced degradation in e-mode gan mis-fets: Impact of substrate terminations," *IEEE Transactions on Electron Devices*, vol. 67, no. 1, pp. 217–223, 2020. DOI: 10.1109/TED.2019.2954282.

[66] A. Chini and F. Iucolano, "Evolution of on-resistance (ron) and threshold voltage (vth) in gan hemts during switch-mode operation," *Materials Science in Semiconductor Processing*, vol. 78, pp. 127–131, 2018. DOI: 10.1016/j.mssp.2017.10.029.

[67] A. Chini, G. Meneghesso, M. Meneghini, F. Fantini, G. Verzellesi, A. Patti, and F. Iucolano, "Experimental and numerical analysis of hole emission process from carbon-related traps in gan buffer layers," *IEEE Transactions on Electron Devices*, vol. 63, no. 9, pp. 3473–3478, 2016. DOI: 10.1109/TED.2016.2593791.

[68] N. Zagni, A. Chini, F. M. Puglisi, P. Pavan, and G. Verzellesi, "The effects of carbon on the bidirectional threshold voltage instabilities induced by negative gate bias stress in GaN MIS-HEMTs," *Journal of Computational Electronics*, vol. 19, no. 4, pp. 1555–1563, Dec. 2020. DOI: 10.1007/s10825-020-01573-8.

[69] Z. H. Liu, G. I. Ng, H. Zhou, S. Arulkumaran, and Y. K. T. Maung, "Reduced surface leakage current and trapping effects in algan/gan high electron mobility transistors on silicon with sin/al2o3 passivation," *Applied Physics Letters*, vol. 98, no. 11, p. 113 506, 2011. DOI: 10.1063/1.3567927.

[70] N. Zagni, M. Cioni, A. Chini, F. Iucolano, F. M. Puglisi, P. Pavan, and G. Verzellesi, "Mechanisms Underlying the Bidirectional VT Shift after Negative Bias Temperature Instability Stress in Carbon-Doped Fully Recessed AlGaN/GaN MIS-HEMTs," *IEEE Transactions on Electron Devices*, vol. Accepted, 2021. DOI: 10.1109/TED.2021.3063664.

Chapter 5

Negative Capacitance Effect to Beat Sensitivity Limits of Nano-Biosensors

Abstract — In this chapter we discuss the benefits of exploiting the negative capacitance effect to beat the sensitivity limits of nano-biosensors.

We thus propose the concept of NC-BioFET (NC-BioFET) as a novel class of devices that have improved SNR compared to FET-based (nano-) biosensors (BioFETs). We find that — upon the triggering of the non-linearity associated with the Negative Capacitance (NC) effect — the sensitivity of NC-BioFETs is significantly enhanced compared to conventional devices. At the same time, rejection of the Flicker noise of the underlying transistor increases due to the higher equivalent gate capacitance of the NCFET, leading to higher SNR. Finally, by properly choosing the ferroelectric material, it is possible to scale the ferroelectric thickness while still obtaining the instability required to improve SNR, hence proving the scalability of NC-BioFETs within current CMOS process.

5.1 Biosensors and Nano-biosensors

In Chapter 1 we discussed how the future of the electronic industry depends not only on the possibility to further scaling transistors but also on the augmentation of functionalities of the chips by resorting to new class of devices such as sensors. In recent years particular attention was devoted to Field-Effect Transistor (FET)-based nano-biosensors for their astounding potential in revolutionary applications such as early detection of diseases, personalized medicine, genome sequencing and many others.

Biosensors are a class of devices employed to detect biological molecules converting the sensed signal to another physical quantity by means of a physical or chemical interaction. The biological molecule (in short, biomolecule) to be measured can be related for example to the presence of a virus in the blood, the pH of a substance, or DNA. Common, everyday examples of biosensors are glucose meters, pH meters, stethoscopes.

The biological sensitive element of a biosensor is called cell receptor, an antibody molecule that reacts, combines or in anyway interacts with the analyte (i.e., the biomolecule to be measured). This interaction gives rise to a signal that is converted by the detector (or transducer) to the measured quantity. In this sense, biosensing conventionally involves an indirect type of measurement, where the quantity of interest is obtained from the actual measurement via well-known physical relations underlying the detection mechanism.

Although biosensing is fairly established, thanks to the advances in nanotechnology in recent years a new class of devices took place — generally referred to as *nano*-biosensors. As their name suggests, these nano devices aim at performing measurements at the nanoscale allowing to detect analytes of very low concentrations [1], potentially enabling a variety of new applications such as labs-on-chip [3], personalized medicine [2], early detection of diseases [4], and genome sequencing [5], [6]. The remarkable interest spurred by nano-biosensors stems from their ability of detecting very small concentrations of target analytse (down to the order of femto and pico molar[1]) thanks to their ultra-small dimensions [1]. More specifically, the increased *sensitivity*

[1]The measurement unit for molecules concentration in an aqueous solution is the *molar concentration*, that measures the number of moles of an atom/molecule for a liter of solution, i.e., $1\,\mathrm{M} = 1\,\mathrm{mol/L} = 1 \times 10^3\,\mathrm{mol/m^3}$. A mole indicates the substance quantity needed to have a number of atoms/molecules equivalent to Avogadro's number, i.e., $N_{AVG} = 6 \times 10^{23}\,\#/\mathrm{mol}$.

of nano biosensors with dimension scaling allowed improving detection of biomolecules while enabling the integration of many electronic components on a single-chip, increasing dramatically the ability to detect and process the bio-signals at ever reducing costs.

In this chapter we will discuss the benefits of combining the NCFET with the BioFET to obtain a biosensor with improved sensitivity and SNR at the same time. The BioFET under investigation is based on a transistor having as the semiconductor material a two-dimensional (2D) Transition Metal Dichalcogenide (TMD), i.e., MoS_2 [10]–[12]. The NCFET is integrated with the BioFET by adding a layer of Zr-doped HfO_2 exhibiting ferroelectricity [30, Ch. 3], [31, Ch. 3] and allowing to harness the *negative capacitance* behavior [13], [14], discussed more in detail in Section 5.2.2.

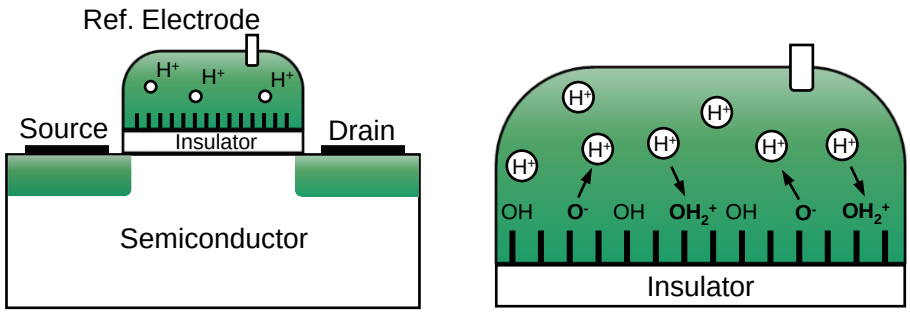
5.2 Background

5.2.1 Screening-limited response of FET-based Biosensors

A BioFET is composed of a MOSFET with the gate metal (or polysilicon) layer removed and replaced by the series of an electrolytic solution and a reference electrode. The electrolytic solution contains the biomolecule sample that needs to be measured and the electrode establishes a reference potential for the solution and allows to bias the transistor (in place of the metal layer). The insulator layer has two important functions: *i)* it serves as the sensing surface, and *ii)* it separates the semiconductor body from the solution avoiding possible contamination and reliability issues.

BioFETs operation relies on the modulation of the surface potential (ψ_s) of the MOSFET caused by the charged biomolecules adsorbed on the gate insulator. To understand more in detail the operating principles of BioFETs and the fundamental limits to their sensitivity we analyze the Ion-Sensitive FET (ISFET), a particular kind of BioFET that is sensitive to the pH concentration of the electrolytic solution under measurement [8]. The electrolytic solution is basically an aqueous solution containing salt that dissociates water (i.e., H_2O) into the ions that need to be measured by the ISFET to determine pH[2]. The sketch of the cross-section of a generic ISFET is

[2]Recall that pH is an indication of the concentration of H^+ ions in a solution, i.e., $pH = -\log_{10}[H^+]$. For instance,



a) b)

Figure 5.1: (a) Sketch of the cross-section of the ISFET and (b) magnification of the electrolytic solution, illustrating the protonation and deprotonation reactions between the solution and the gate insulator.

shown in Fig. 5.1(a). As mentioned earlier, the combination of electrolytic solution and reference electrode replaces the gate metal with the gate insulator acting as the sensing surface. The top surface of the insulator acts as the sensing layer because of the open dangling bonds that can bind to the positive H^+ ions (which are actually just protons) present in the electrolytic solution. Dangling bonds are incomplete oxygen bonds of the SiO_2 (or any other binary oxide acting as the insulator layer) that form because of the two-dimensional surface that breaks the structure and symmetry of the bulk oxide itself. Thus, when placed in contact with an electrolytic solution the oxygen dangling bonds get terminated by the dissociated hydrogen contained in the solution leading to the formation of OH amphoteric sites. These sites act as the receptors that collect the H^+ ions through protonation and deprotonation reactions, as depicted in Fig. 5.1(b), and allow for pH to be measured.

Several different transistor technology were employed to realize either ISFETs or BioFETs, such as dual-gate FETs [15], Silicon NW [1], [16], [17], and 2D-semiconductor FETs [10]–[12]. Regardless of the particular transistor technology employed, however, the sensitivity of ISFETs is fundamentally limited to be $\leq 60\,\mathrm{mV/pH}$ (the so-called Nernst Limit [8]). This limit depends on the diffusion of biomolecules to the insulator surface [19], surface charge screening and to the linear "charge-to-surface potential" relationship [1].

a solution with $pH = 7$ has a density of 1×10^{-7} M protons.

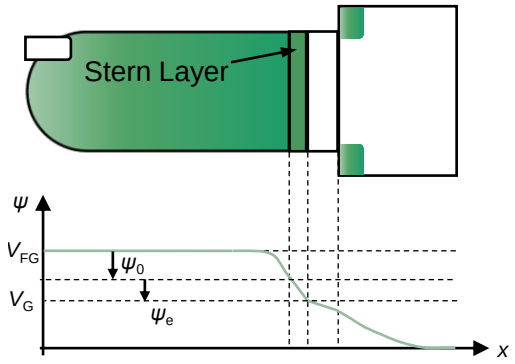


Figure 5.2: ISFET schematic 1D representation showing the potential distribution building up as a consequence the charge accumulating at the electrolytic and insulator layer.

ISFET Basic Equations

We now present the general equations governing the operation of ISFETs to understand the fundamental sensitivity limits of these devices. As shown in Fig. 5.1(b), the pH sensing occurs at the insulator layer surface in touch with the electrolytic solution thanks to the exchange of H^+ ions. These ions interact with the surface by either protonate or deprotonate the amphoteric OH sites residing on it. Basically, there exists a dynamic equilibrium between the surface and the rest of the solution that is determined by the pH of the solution itself. Equilibrium occurs when charge-neutrality of the system is reached, namely when the following occurs

$$Q_s + Q_{surf} + Q_{dl} = 0. \tag{5.1}$$

Q_s is the semiconductor charge, Q_{surf} is the surface charge due to exchange of ions, and Q_{dl} is the double-layer charge. Q_{dl} is due to the accumulation of ions at the electrolytic/insulator interface and it affects the sensor sensitivity, as discussed later. From Eq. (5.1) observe that a variation in pH, which either increases or decreases Q_{surf}, reflects on Q_s thus varying the current flowing into the channel. Thus the actual quantity monitored by the ISFET is the I_D of the MOSFET and not the pH itself, as typical of an *indirect* type of measurement. A schematic 1D representation of the system is shown in Fig. 5.2, alongside with the potentials building up due to the accumulation of charges. ψ_0 is the potential drop due to the double-layer, an equivalent layer representing the accumulation of positive and negative ions building up as a result of the protonation and

deprotonation reactions. ψ_e is an additional potential drop given by the so-called "Stern layer" which accounts for the finite dimension of ions limiting the accumulation of additional charges at the interface [7]. In practice, both the double-layer and Stern-layer partially screen the charges accumulating at the surface due to the binding with the OH groups causing potential drops that in turn limit the sensitivity of the device.

Q_{dl} depends on ψ_0 whereas Q_{surf} depends on ψ_e [7]. The relation between these two potentials is obtained from the Gouy-Chapman-Stern theory [7], and reads

$$\psi_0 = \psi_e + \frac{Q_{dl}}{C_{ST}} \tag{5.2}$$

with C_{ST} being the Stern-layer capacitance ($\approx 20\,\mu F/cm^2$ [7]). Note that both the double-layer and the Stern-layer can be seen as capacitive elements providing the charge Q_{dl} that sums up to Q_{surf} (this giving an additional capacitive term) to give Q_s, as stated by Eq. (5.1). Although this is a simplified picture of the real scenario (where basically the dimension of the electrolytic solution have been neglected), it is sufficient to understand the essential physics at the basis of ISFETs operation and becomes extremely useful for compact modeling purposes, as discussed in Section 5.2.1.

A further assumption that we will employ here is that to a first approximation Q_s can be neglected in Eq. (5.1). This assumption is based on the experimental observation revealing that the majority of the surface charge is screened by the double-layer and that only a small fraction of it is actually reflected on the transistor channel[3]. This assumption simplifies the analysis, allowing to write the following system of equations fully describing the operation of ISFETs (in

[3]For instance, simulations carried out to obtain Eq. (5.3) reveal that $Q_{dl} \approx 37\,\mu C/cm^2$, $Q_{surf} \approx -40\,\mu C/cm^2$, while $Q_s \approx 1\,\mu C/cm^2$ in inversion. This shows that indeed the approximation is reasonable. A further validation for this approximation is the agreement between the DC model and the transient model, that does not assume Q_s to be negligible [7].

DC conditions)

$$V_G = V_{FG} + \psi_e \tag{5.3a}$$

$$\psi_e = 2V_T \sinh^{-1}\left(\frac{Q_{surf}}{\sqrt{8k_BT\varepsilon_w n_0}}\right) + \frac{Q_{surf}}{C_{ST}} \tag{5.3b}$$

$$Q_{surf} = -2qN_{OH}\left\{\frac{\tanh\left(\frac{\psi_e}{V_{th}} + \ln 10(pH - pH_z)\right)}{2 + 10e^{\Delta pK/2}\sqrt{1 - \left[\tanh\left(\frac{\psi_e}{V_{th}} + \ln 10(pH - pH_z)\right)\right]^2}}\right\}. \tag{5.3c}$$

N_{OH} is the concentration of surface OH groups, $pH_z = (pK_A + pK_B)/2$ is the point of zero-charge (i.e., the pH value for which the surface charge is null), $\Delta pK = pK_B - pK_A$, and pK_A (pK_B) is the acid (base) dissociation constant. Eq. (5.3b) is obtained by imposing $Q_{dl} = -Q_{surf}$ (from Eq. (5.1) with $Q_s \approx 0$) and by substituting into Eq. (5.2) the following

$$Q_{dl} = -\sqrt{8k_BT\varepsilon_w n_0}\sinh\left(\frac{\psi_0}{2V_{th}}\right) \tag{5.4}$$

where $\varepsilon_w \approx 80\varepsilon_0$ is the dielectric permittivity of the aqueous solution and n_0 is the salt concentration (obtained from the ionic concentration i_0, i.e., $n_0 = N_{AVG} \times i_0$).

Sensitivity and Nernst Limit

We are now ready to find an expression for the sensitivity of the ISFET. Since pH varation modulate the V_{FB} and hence V_t it is convenient to express the sensitivity as follows

$$S_V \equiv \left|\frac{\Delta V_t}{\Delta pH}\right|. \tag{5.5}$$

This way, we can use the relations derived previously to obtain an expression that highlights the fundamental physical quantities limiting sensitivity. From Eq. (5.3a) we observe that ΔV_t (or equivalently, ΔV_G) corresponds to $\Delta\psi_e$, which can be related to pH variation through Q_{surf}. First, by using Eq. (5.1) we obtain

$$\frac{\partial Q_{surf}}{\partial \psi_e} = -\frac{\partial Q_{dl}}{\partial \psi_e} - \frac{\partial Q_s}{\partial \psi_e} = C_{diff} - \frac{\partial Q_s}{\partial \psi_g} = C_{diff} + C_{MOS} \tag{5.6}$$

where $C_{diff} = C_{dl} \parallel C_{stern}$[4]. Then, by using the chain-rule we can write

$$\frac{\partial \psi_e}{\partial pH} = \frac{\partial \psi_e}{\partial Q_{surf}} \frac{\partial Q_{surf}}{\partial pH} \tag{5.7}$$

where pH is the concentration of ions on the *insulator surface*. Actually, the quantity that needs to be measured is the *bulk* pH in the electrolytic solution, pH_B, determined simply as

$$pH_B = pH - \frac{\psi_e}{\ln(10) V_{th}}. \tag{5.8}$$

We then obtain

$$\frac{\partial \psi_e}{\partial pH} = \frac{\partial \psi_e}{\partial (pH_B + \psi_e / V_T)} = -\frac{q\beta_s}{C_{diff} + C_{MOS}} \tag{5.9}$$

where we have used the definition $\partial Q_{surf} / \partial pH = -q\beta_s$, with β_s being the surface buffer capacity (i.e., the efficiency of the insulator to take or give up protons [8]). By rearranging Eq. (5.9) we can now obtain the final expression for S_V as follows

$$S_V = \left| \frac{\partial \psi_g}{\partial pH_B} \right| = \ln(10) V_T \times \frac{1}{\frac{1}{C_{surf}} (C_{diff} + C_{MOS}) + 1} \tag{5.10}$$

where $C_{surf} = q\beta_s / (\ln(10) V_{th})$ [9].

Eq. (5.10) is an important relationship that reveals the dependence of sensitivity on key parameters of the sensor itself. Particularly, we can observe that for an ideal sensor, i.e., $\beta_s \to \infty$

$$S_V^{Nernst} = \ln(10) V_{th} \approx 60\,\mathrm{mV/dec} \tag{5.11}$$

at room temperature. This is is the so-called *Nernst limit* that upper bounds the maximum sensitivity. This fundamental limitation comes from the potential barrier building-up at the insulator surface as a consequence of accumulation of ions in the double-layer. Because of the similarity of the physics of the double-layer potential and the energy barrier at the virtual source of a MOSFET [4, Ch. 2], it should come as no surprise that the Nernst limit has the same value as

[4]The last passage in Eq. (5.6) is justified by the fact that Q_s has opposite sign with respect to the other charges since it is defined on the other side of the gate insulator.

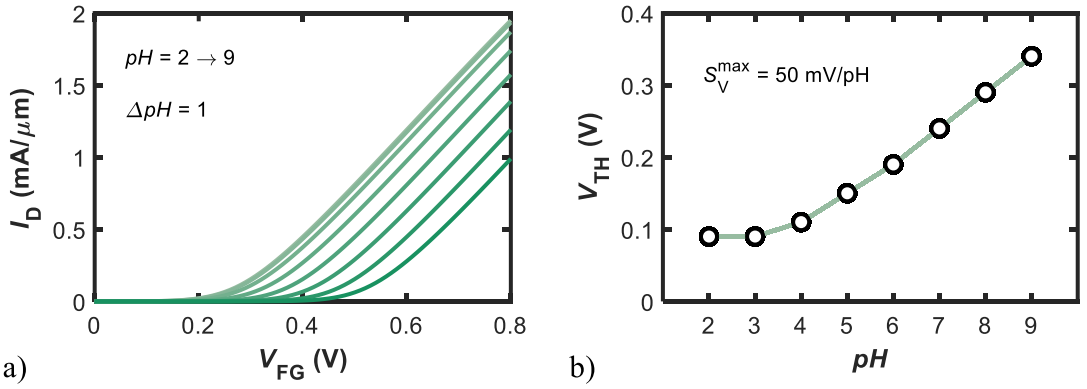


Figure 5.3: (a) Illustrative $I_D - V_{FG}$ characteristics of the ISFET with pH of the electrolytic solution varying between 2 and 9. (b) Corresponding sensitivity of the sensor with varying pH. Maximum sensitivity is well below the Nernst Limit due to charge-screening at solution/SiO_2 interface.

SS at room temperature, see Eq. (5.19). This is due to the thermodynamic nature of both processes leading to the building up of an energy barrier that needs to be overcome for the accumulation of ions (as in this case) or electrons/holes (as in the case of MOSFETs).

Illustrative $I_D - V_{FG}$ characteristics of the ISFET are shown in Fig. 5.3(a) (V_{FG} is the front-gate bias applied through the reference electrode). The characteristics were drawn for pH values varying in a range from 2 to 9 for sensor system including a conventional bulk MOSFET and SiO_2 as the insulator layer. The model employed for simulating the ISFET is built upon the MVS model for the MOSFET and the DC pH sensor model [7], described more in detail in Section 5.2.1. Fig. 5.3(b) shows the V_t[5] evolution when increasing the pH, which has a non-linear behavior near the point of zero charge. For SiO_2 in fact, pH_{zpc} corresponds to 2 (see model parameters in Table 5.1), and near this point Q_{surf} is very small, leading to small variations in ψ_e. When increasing further the pH and moving away from the zero-charge point, V_t variation with pH reaches 50 mV/pH value at maximum. This value is well below the theoretical Nernst limit, as anticipated by Eq. (5.5), because of the finite β_s that limits the ability of the sensor to capture and release protons.

[5]Here V_t is defined as the voltage at which $I_D = 1\,\mu A/\mu m$.

Table 5.1: Parameters for the electrolytic solution used in ISFET simulations.

Symbol	Value
pK_A	-2
pK_B	6
N_{OH}	$5 \times 10^{14}\,\mathrm{cm}^{-2}$
i_0	$0.1\,\mathrm{M}$
C_{ST}	$20\,\mu\mathrm{F/cm}^2$

ISFET Compact Model

The ISFET physics-based compact model used to derive the results included in this chapter was implemented in Verilog-A for SPICE circuit simulations. Depending on the type of simulation analysis (i.e., quasi-static, transient, or small-signal) the sensor is modeled via equivalent circuit elements differently [7]. We briefly analyze the DC, transient and AC (small-signal) equivalents of the ISFET. In the AC model we include also the noise elements relative to the electrolytic solution and the transistor itself that are needed in order to determine the SNR. In the case of Fig. 5.3 the MOSFET was modeled with the MIT-Virtual-Source model [4, Ch. 2] although in principle any compact model can be used [7] (in Section 5.3 we will make use of a compact model for a MoS_2 MOSFET).

The DC model shown in Fig. 5.4(a) treats the sensing unit as a pH-controlled voltage generator that determines ψ_e through Eqs. (5.2) and (5.3b). Despite the assumption of $Q_s \approx 0$ (i.e., negligible MOSFET charge) the model was found to be able to match experimental data for different transistor technologies in both above- and sub-threshold regions [7].

The transient model in Fig. 5.4(b) was derived by assuming quasi-static conditions applied for simplicity, i.e., the pH and ionic concentration variations are instantaneously sensed at the solution/insulator interface. This assumption allows the transient model of the electrolyte and electrolyte/insulator interface to be simply described in terms of three capacitors and one resistor. The capacitive elements are given by the double layer capacitance (C_{dl}), the Stern capacitance (C_{ST}), and the capacitance due to the charging/discharging of the surface groups (C_{surf}). The capacitor are calculated from their respective constitutive equations, i.e., $C \sim \partial\sigma/\partial\psi$ (where

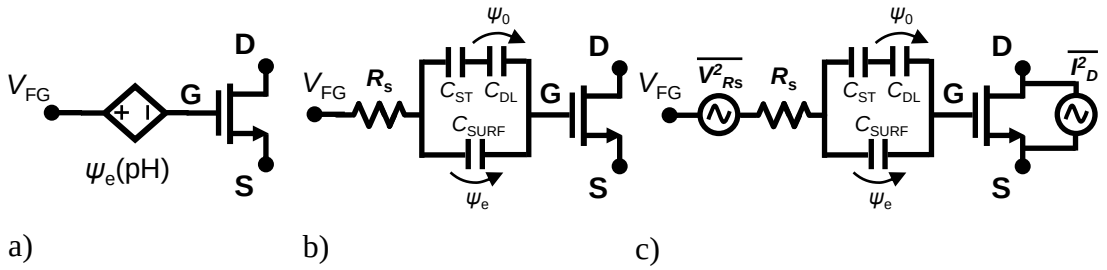


Figure 5.4: ISFET compact model for (a) DC, (b) transient, and (c) AC (with noise sources) analysis.

σ, ψ constitute the charge density and potential across each capacitor in Fig. 5.4(b)). The resistor (R_S) is given by the finite conductivity of the ions ions present in the solution calculated as

$$R_s = \frac{1}{q(\mu_{i,n} + \mu_{i,p})n_0}\sqrt{\frac{\pi}{A}} \tag{5.12}$$

valid for reference electrodes much larger than the sensor area (A) [7]. We assumed that the reference electrode is an ideal Faradaic electrode, i.e., it does not have any resistance to charge transfer and, therefore, there is no potential drop between the electrode and electrolyte [7]. Unlike the DC model, where Q_s is assumed negligible, the transient model solves the system self-consistently and appropriately sets the voltage at the gate oxide/electrolyte interface. Interestingly, in [7] it was found that results obtained with DC and transient models match remarkably well, particularly for high pH values (when the surface charge is relatively large), indicating that the simplifying assumption of MOSFET charge being negligible can be safely applied.

For the small-signal (AC) analysis the circuit in Fig. 5.4(c) is used. The small-signal currents flowing across the capacitors are expressed as $i \sim C\partial v/\partial t$, where each capacitance is obtained as for the transient model in terms of the respective charge and potential. The operating point is obtained from the DC solution of the circuit. The noise sources contributing to the ISFET noise figure are

1. Thermal Noise due to Electrolyte
2. Thermal Noise due to MOSFET
3. Flicker Noise due to MOSFET.

The thermal noise due to the electrolyte is modeled as a voltage generator in series with R_s as shown in Fig. 5.4(c), whose Power Spectral Density (PSD) in V^2/Hz reads

$$PSD^{elec} = 4k_BTR_s. \tag{5.13}$$

The MOSFET contributes with both thermal and Flicker ($1/f$) noise. The first comes from the finite resistivity of the channel and can be expressed as [3]

$$PSD_{I_D}^{ch} = 4k_BTR_{ch}g_m^2 = 4k_BT\frac{\mu_n(-Q_n)}{L^2}. \tag{5.14}$$

The second instead, comes from the fluctuations in the mobility and in the number of carriers flowing in the channel (due to trapping/de-trapping from traps in the oxide). These fluctuations give rise to $1/f$ noise (also called Flicker noise) with PSD in A^2/Hz as follows

$$PSD_{I_D}^{flicker} = \frac{q^2k_BT\lambda N_T}{fWLC_{ox}^2}(g_m + \alpha_C\mu_nC_{ox}I_D)^2 \tag{5.15}$$

where λ is the tunneling length, N_t is the oxide trap density, and α_C is the Coulomb scattering coefficient. The noise sources of the MOSFET are modelled as a current generator in parallel to the transistor itself as shown in Fig. 5.4(c). We will make use of the model for AC-noise analysis in Section 5.3.3 to show the beneficial effect of adding a ferroelectric layer to the gate stack in terms of SNR of the BioFET for biomolecules sensing.

5.2.2 Voltage Amplification with Negative Capacitance

Basics of NCFET Physics

The possibility of integrating a ferroelectric layer in the gate stack of a MOSFET was first discussed in [13]. In this seminal contribution, it was argued that if a ferroelectric layer with the "S-shaped" $P-E$ relation given by LT was integrated in the gate stack of a conventional MOSFET, then it would be possible to obtain a voltage step-up action (like in a transformer) leading to voltage amplification and thus to sub-60 mV/dec SS. The $P-E$ relationship according

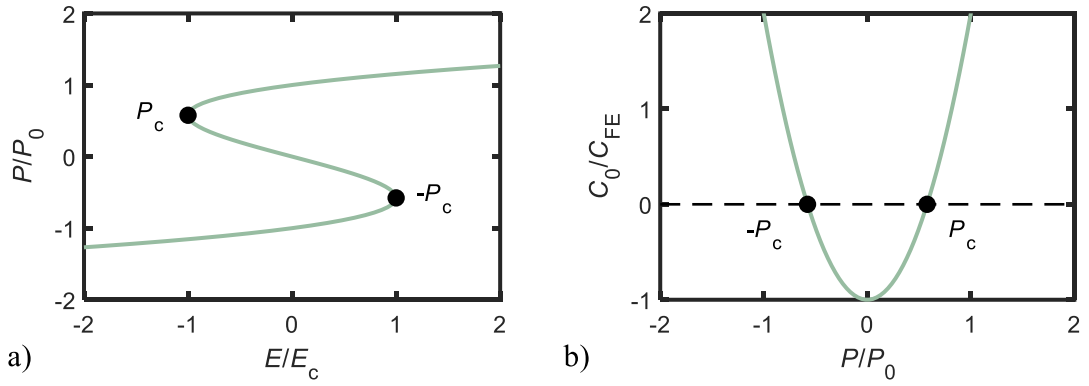


Figure 5.5: (a) $P-E$ and (b) $1/C-P$ characteristics of a SBT ferroelectric capacitor according to LT obtained with Eq. (5.16). The NC region corresponds to the region for which $|P| \leq P_C$. $P_0 = \sqrt{|\alpha|/2\beta}$, $P_C = \sqrt{|\alpha|/6\beta}$, $E_C = 4/3|\alpha|P_C$, $C_0 = 1/2|\alpha|$. $\alpha = -6.5\times10^7\,\mathrm{m/F}$, $\beta = 3.75\times10^9\,\mathrm{m^5/FC^2}$, $\gamma = 0\,\mathrm{m^9/FC^4}$, $\rho = 0\,\Omega\mathrm{m}$.

to LT was previously discussed for FeFETs in Section 3.4 and it is here repeated for convenience

$$E = 2\alpha P + 4\beta P^3 + 6\gamma P^5 + \rho\frac{\partial P}{\partial t}. \tag{5.16}$$

The $P-E$ loop corresponding to Eq. (5.16) for the SBT ferroelectric is shown in Fig. 5.5(a). The region for which $|P| \leq P_C$ (with $P_C = \sqrt{|\alpha|/6\beta}$ is defined as the point at which $\frac{\partial E}{\partial P} = 0$) is the NC region to which corresponds $C \equiv (\partial E/\partial P)^{-1} < 0$, as shown in Fig. 5.5(b).

In a stand-alone ferroelectric capacitor, the NC region of the $P-E$ loop is not stable [13] (because the energy of the system in this region has no stable minimum). However, if connected in series with a positive capacitor, say for instance the series connection of composed of C_{ox} and C_S in the gate stack of a MOSFET then the NC would be stabilized giving rise to an voltage amplification of the surface potential hence reducing *SS* [14], [20].

The structure of an NCFET in the Metal-Ferroelectric-Metal-Insulator-Semiconductor (MFMIS) configuration is shown in Fig. 5.6(a). Compared to a conventional MOSFET, an NCFET has an extra ferroelectric capacitor, C_{FE}, connected in series with the gate oxide capacitance, C_{ox}, see Fig. 5.6(b). Notice that for this particular structure, namely MFMIS, the metal layer in between the ferroelectric and oxide ensures that the potential along the channel direction is uniform (on

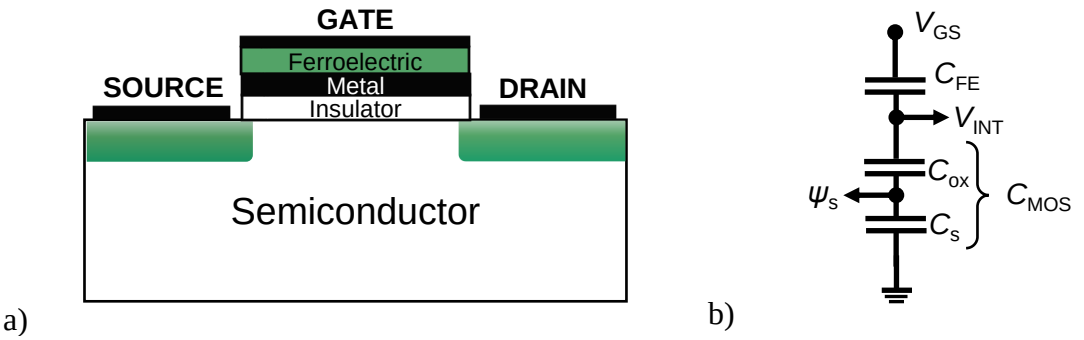


Figure 5.6: (a) Sketch of the NCFET in the MFMIS configuration. (b) Simplified circuit schematic of the NCFET.

the ferroelectric capacitor), hence simplifying the analysis and design of the NCFET[6]. With the aid of the simple capacitor network for the gate loop shown in Fig. 5.6(b) it is possible to derive the KVL as follows

$$V_{GS} - V_{FB} = V_{FE} + V_{INT} = V_{FE} + V_{ox} + \psi_s \tag{5.17}$$

where $V_{int} = V_{ox} + \psi_s$ is the internal gate potential. From (5.17) it is clear that for $V_{FE} < 0$ then $V_{INT} > V_{GS}$ (assuming that V_{FB} is constant). In more specific terms, a "well-tempered" NCFET guarantees that

$$A_V \equiv \frac{\partial V_{INT}}{\partial V_{GS}} = \frac{C_{FE}}{C_{FE} + C_{MOS}} = \frac{|C_{FE}|}{|C_{FE}| - C_{MOS}} > 1 \tag{5.18}$$

is greater than 1 [21]. The last passage in Eq. (5.18) is justified by the fact that $C_{FE} < 0$, see Fig. 5.5(b).

When $A_V > 1$, SS reduction takes place. This can be understood better from the expression of SS, that reads [14] (at room temperature)

$$\begin{aligned} SS &= \left(\frac{\partial \log I_D}{\partial V_{GS}}\right)^{-1} = \left(\frac{\partial \psi_s}{\partial V_{GS}} \times \frac{\partial \log I_D}{\partial \psi_s}\right)^{-1} \\ &= \underbrace{\left(\frac{\partial \psi_s}{\partial V_{GS}}\right)^{-1}}_{m} \times \underbrace{(2.3 V_T)}_{n} \approx m \times 60\,\mathrm{mV/dec}. \end{aligned} \tag{5.19}$$

When $A_V > 1$ in fact, $m < 1$ and thus $SS < 60\,\mathrm{mV/dec}$. The SS reduction associated with A_V

[6]It is also possible to realize an NCFET in the MFIS configuration [38, Ch.3] already discussed for FeFETs, see Chapter 3.

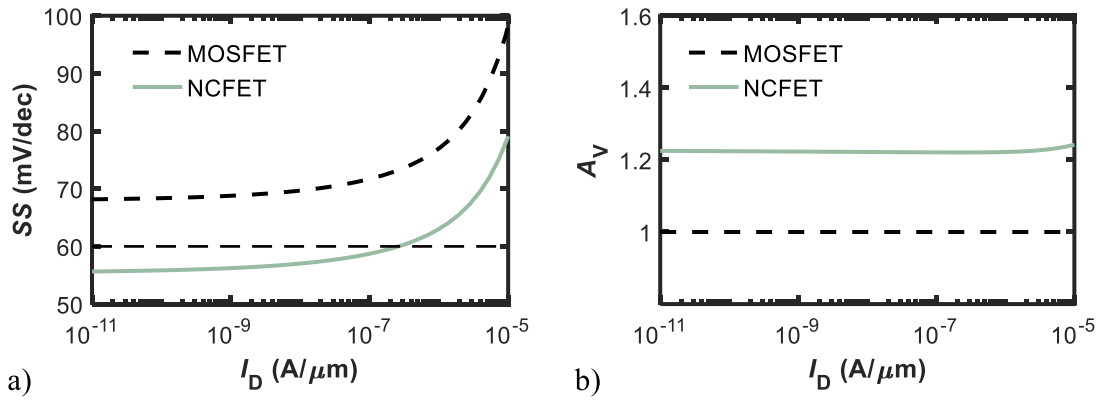


Figure 5.7: (a) SS and (b) A_V vs I_D comparison between the MOSFET (black dashed line) and NCFET (green solid lines) case.

increase is illustrated in Fig. 5.7, comparing the case for the MOSFET and NCFET. As it can be seen from Fig. 5.7(a), SS is reduced below the 60 mV/dec limit for the NCFET. This is confirmed by the increased A_V shown in Fig. 5.7(b). Notice that the SS ratio between the MOSFET and NCFET case corresponds to A_V, i.e.,

$$SS^{\text{NCFET}} = SS^{\text{MOSFET}} \times \frac{1}{A_V}. \tag{5.20}$$

Note as C_{FE} and C_{MOS} get closer in value (as obtained for instance when inreasing t_{FE}), then A_V increases, see Eq. (5.18), and SS reduces. The condition at which C_{FE} equals C_{MOS} is called the "Capacitance Matching" condition for which A_V reaches a maximum and SS is minized without incurring in hysteresis [20], [22].

The SS improvement can also be appreciated with the $I-V$ characteristics. Fig. 5.8 compares the $I-V$ curves of a MOSFET with that of a NCFET highlighting the two (alternative) advantages obtained thanks to SS reduction. The first consists of I_{OFF} reduction at constant I_{ON}, see Fig. 5.8(a); the second, viceversa, consists of I_{ON} increase at constant I_{OFF}, see Fig. 5.8(b). Both these improvements stem from the lower SS of NCFET compared to that of MOSFET. In fact, from the definition of I_{OFF} [4, Ch. 2], we can write

$$I_{OFF} \propto 10^{-V_t/SS}. \tag{5.21}$$

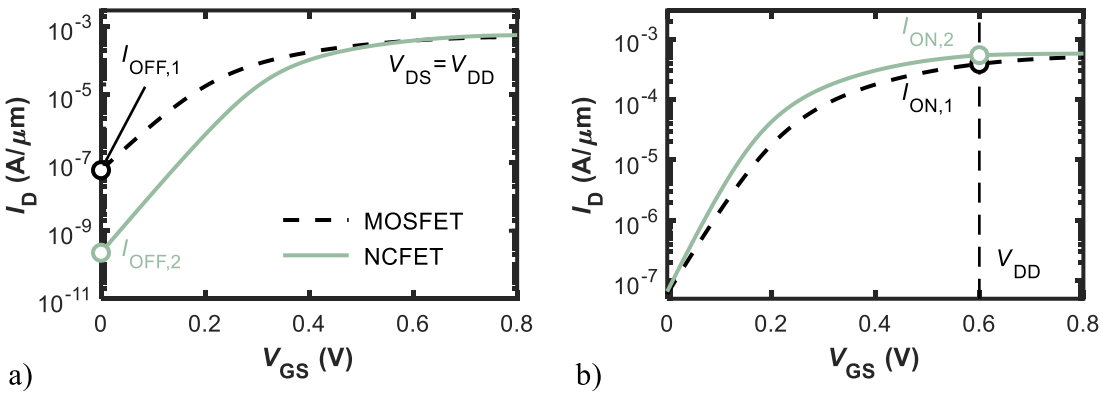


Figure 5.8: $I-V$ curve for MOSFET (black dashed line) and NCFET (green solid line) showing the improvement in I_{ON}/I_{OFF} ratio ($V_{DS}=V_{DD}=0.6\,\mathrm{V}$). (a) I_{OFF} reduction at constant I_{ON}. (b) I_{ON} increase at constant I_{OFF}.

Eq. (5.21) reveals that SS reduction leads to an exponential decrease in I_{OFF} (at constant V_t). By the same token, SS reduction leads to I_{ON} increase (at constant I_{OFF}). In fact, observe from Eq. (5.21) that keeping I_{OFF} constant with decreasing SS requires to decrease V_t as well. Thus, since I_{ON} depends linearly on V_t [4, Ch. 2] then V_t reduction leads to a linear increase in I_{ON}. This can be visualized in Fig. 5.8, showing that I_{ON} increase is linear while I_{OFF} reduction is exponential (with respect to SS reduction).

NCFET Compact Model

In this section we present the compact model for NCFETs that can be used in circuit simulations (SPICE). We focus the discussion on the compact model for the MFMIS structure, see Fig. 5.6(a), because it allows to decouple the ferroelectric layer modeling from that of the underlying transistor [21], [23]. This is useful in circuit analysis as one can rely on existing MOSFET compact models (combined with the appropriate ferroelectric model) to simulate the behavior of NCFETs.

Thus the basic assumption of NCFETs compact models is that the underlying MOSFET is decoupled from the ferroelectric model, simplifying the circuit analysis. In fact, from the structure depicted in Fig. 5.6(a), notice that the NCFET can be seen as MOSFET with an additional ferroelectric capacitor in series in the gate stack. Thus, the NCFET can be modeled by a capacitive divider network in the gate stack [21]. The non-linear $P-V$ relationship describing

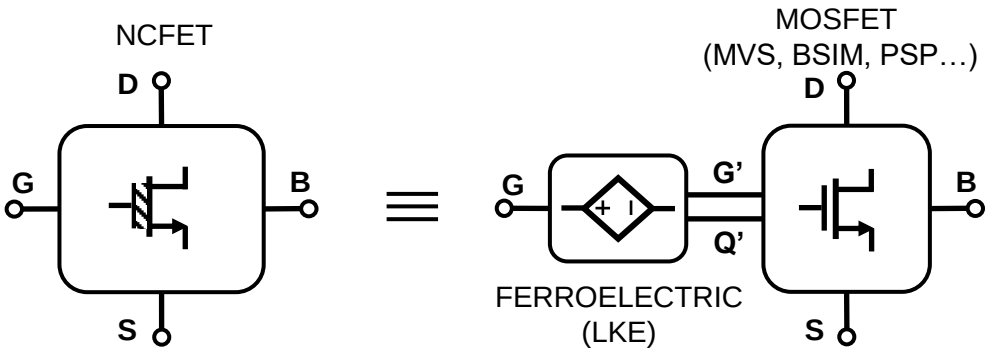


Figure 5.9: Schematic of the NCFET compact model for SPICE simulations.

the ferroelectric capacitor is the static LKE, i.e.,

$$V_{FE} = t_{FE}(2\alpha P + 4\beta P^3 + 6\gamma P^5). \tag{5.22}$$

This relationship can be simply modeled as a charge-dependent voltage source as schematically shown in Fig. 5.9[7]. The polarization charge is assumed for simplicity to be equal to Q_s, i.e., $Q_s = \varepsilon_0 E + P \approx P$ [21]. As shown in Fig. 5.9, the gate charge is computed internally by the MOSFET model and is fed back to the ferroelectric capacitor by the internal branch labeled as Q' [24]. The other shared branch G' corresponds to V_{INT} in Fig. 5.6(b) and is useful to compare the device performance with and without the ferroelectric layer. Notice in fact that seen from the G' terminal, the NCFET is actually the same as a MOSFET. The built-in solver of the SPICE simulator makes sure that the solution for each input bias is self-consistent and determines the potential and current flow at each terminal. The parameters used to draw Figs. 5.7 and 5.8 are collected in Table 5.2. The ferroelectric material chosen is Si-doped HfO_2, with Landau parameters set as in [14]. The Verilog-A code for the NCFET model described here is available at the following link: https://nanohub.org/publications/95/5.

[7] An equivalent approach is to calculate the charge feeding the ferroelectric capacitor via a 'dummy' feedback network, as done in [23].

Table 5.2: Parameters for the SPICE simulations in Section 5.2.2 with the NCFET compact model.

Symbol	Value
L	45 nm
W	1 µm
C_{ox}	$3.85\,\mu F/cm^2$
V_t	0.34 V
R_{DS}	$500\,\Omega\mu m$
t_{FE}	10 nm
α	-8.65×10^8 m/F
β	$1.92 \times 10^{10}\,m^5/F/C^2$
γ	$0\,m^9/F/C^4$

5.3 Contribution: Exploiting NCFETs to Beat Nernst Sensitivity Limit of Biosensors and Improving Signal-To-Noise Ratio

In this section, we introduce the concept of the NC-BioFET, combining the NCFET with a BioFET for biomolecule detection to obtain improved sensitivity and enhanced SNR. To this end, we develop an analytical model of the NC-BioFET to predict the performance limits of this sensor. Our results show that despite the fundamental limits of charged-based BioFETs [15], [18], the NC-BioFET could dramatically improve the limits of label-free detection of biomolecules. Our analysis, based on nontrivial integration of two well established technologies (e.g. NCFET and BioFET), provides the theoretical foundation necessary for the experimental demonstration of the NC-BioFET. Although here we focus on nano-biosensing, the principle of negative capacitance based detection is general and can improve, for example, the SNR of any potentiometric transistor-based sensor, such as gas sensors for instance [25].

The physical model of the proposed NC-BioFET relies on three fundamental mechanisms: (*i*) the screening of the electrolyte solution (representing the 'sensing' part of the device); (*ii*) the nonlinear response of the negative capacitor and; (*iii*) electrical response of the transistor. The modeling approach integrates the negative capacitor with the conventional MOSFET and with the electrolyte model, thereby simplifying the design of NC-BioFETs. We illustrate the

principle of operation of the NC-BioFET by using a 2D semiconductor Molybdenum Disulfide (MoS_2)-based FET. The choice of this particular class of transistors is motivated by their potential for integration in ultra-scaled chips [11], [26] and because of the experimental demonstration of both MoS_2-based NCFETs [26]–[28] and BioFETs [10]–[12].

5.3.1 NC-BioFET Compact Model

BioFET Model

The first block represents the BioFET sensor. Its working principle is similar to that of the ISFET already described in Section 5.2.1, with in addition of target charged analytes (e.g., DNA, protein, etc.) in the solution. The analytes are detected when they conjugate with the receptors sites on the insulator surface (that needs to be properly *functionalized* to be sensitive to biomolecules). As for the ISFET, in the BioFET the metal gate is replaced by the electrolyte and by the reference electrode. The effect of charged biomolecules, Q_{BIO}, is included in the charge neutrality equation Eq. (5.1) as follows

$$Q_{BIO} + Q_{MOS} + Q_{surf} + Q_{dl} = 0 \tag{5.23}$$

To a first-order approximation, Q_{MOS} can be neglected as Q_{surf} almost entirely neutralizes Q_{dl} and Q_{BIO}, alike for the ISFET case [7]. As discussed in Section 5.2.1, the charge screening effect is at the origin of the so-called "screening-limited" response of BioFETs that fundamentally limits sensitivity below the Nernst limit [1].

MoS_2 NCFET Model

The MoS_2 NCFET is modeled by considering the baseline FET and the negative capacitance separately (i.e., considering the MFMIS configuration). The combination of the two components is solved self-consistently by imposing KVL [29]

$$V_{GS} = V_{FE} + V_{INT}. \tag{5.24}$$

The ferroelectric voltage V_{FE} is obtained by applying the LKE, that reads

$$V_{FE} = t_{FE}\left(2\alpha Q_{MOS} + 4\beta Q Q_{MOS}^3 + 6\gamma Q Q_{MOS}^5 + \rho\frac{\partial Q_{MOS}}{\partial t}\right) \tag{5.25}$$

where Q_{MOS} is the total charge seen by the ferroelectric, written as [29]

$$Q_{MOS} = Q_s + \frac{Q_{P,1} + Q_{P,2}}{L} = Q_s + \frac{C_P V_{INT} + C_P(V_{INT} - V_{DS})}{L} \tag{5.26}$$

where C_P is the parasitic capacitance (per unit length) due to gate/source and gate/drain overlap [29].

The solution for the I_D and Q_s is found by applying Pao and Sah double-integral (imposing the Gradual Channel Approximation (GCA), valid for long-channel devices) and by using Ward-Dutton charge partitioning method [29], respectively. The equations of the analytical model are reported as follows

$$I_D = q\mu_{eff}\frac{W}{L}\left\{N_{2D}\left[\left(b + \frac{k_B T}{q}a\right)\psi - a\frac{\psi^2}{2}\right]\right\}_{\psi_s}^{\psi_d} \tag{5.27a}$$

$$Q_s = q\cdot WL\frac{N_{2D}\left\{\frac{ad}{3}(\psi_s^2 + \psi_s\psi_d + \psi_d^2) - \frac{1}{2}(b_1 d + ac)(\psi_s + \psi_d) + b_1 c\right\}}{\left[c - \frac{d}{2}(\psi_s + \psi_d)\right]} \tag{5.27b}$$

$$\psi(x) = \frac{b}{a} - \frac{k_B T}{q}W_0\left\{\frac{q}{ak_B T}\exp\left[\frac{q(b - aV_{ch})}{ak_B T}\right]\right\} \tag{5.27c}$$

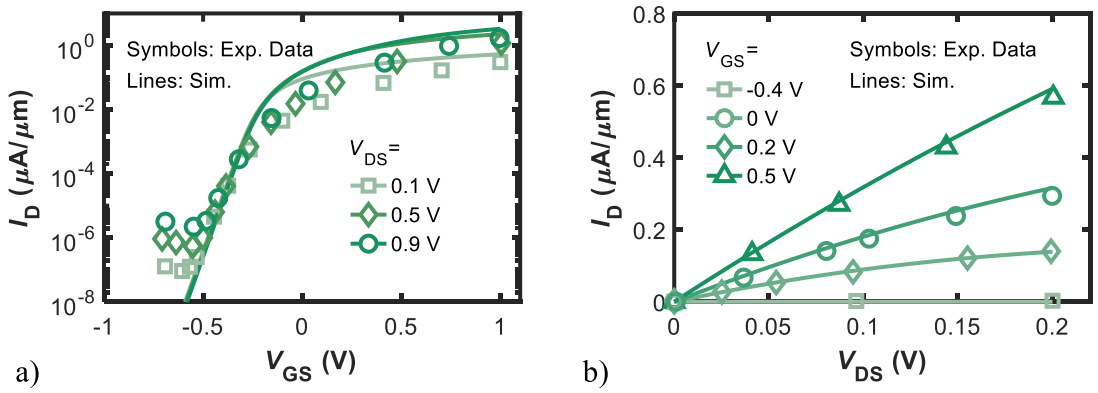


Figure 5.10: Calibration of the MoS_2 NCFET compact model with experimental data from [26]. (a) $I_D - V_{GS}$ and (b) $I_D - V_{DS}$ characteristics. Adapted from [36].

where

$$a = \frac{\varepsilon_{2D} t_{2D}}{q N_{2D} \lambda_{2D}^2} \tag{5.28a}$$

$$b = \frac{\left(\frac{\varepsilon_{2D} t_{2D} \xi}{q} + N_d\right)}{N_{2D}} \tag{5.28b}$$

$$b_1 = \frac{\varepsilon_{2D} t_{2D} \xi}{q N_{2D}} \tag{5.28c}$$

$$c = N_d + \frac{\varepsilon_{2D} t_{2D}}{q}\left(\xi + \frac{k_B T}{q \lambda_{2D}^2}\right) \tag{5.28d}$$

$$d = \frac{\varepsilon_{2D} t_{2D}}{q \lambda_{2D}^2} \tag{5.28e}$$

$$\xi = \frac{V_{INT} - V_{FB}}{\lambda_{2D}^2} \tag{5.28f}$$

Note that $W_0(z)$ in Eq. (5.27c) is the upper branch of the Lambert function and ψ_S, ψ_D are the boundary conditions for the potential calculated at the source and drain edge of the channel, respectively. The expression for I_D is derived starting from the Poisson Equation in the 2D channel, by assuming Boltzmann statistics for the carrier's distribution. For the full derivation, see [29].

The compact model of the MoS_2 NCFET was calibrated against experimental $I - V$ characteristics, as shown in Fig. 5.10. The parameter values used for fitting of the curves are collected in Table 5.3. The reasonable agreement between the simple analytical model and experimental

Table 5.3: Parameters for calibration of MoS_2 NCFET model with data from [26].

Symbol	Value
L	$2\,\mu m$
W	$1\,\mu m$
t_{ch}	$10\,nm$
t_{ox}	$2\,nm$
N_d	$1.5\,cm^{-2}$
C_p	$3.54\,fF/\mu m$
V_{FB}	$-0.2\,V$
μ_{eff}	$2\,cm^2/Vs$
t_{FE}	$20\,nm$
α	$-1.911\times10^8\,m/F$
β	$5.898\times10^9\,m^5/F/C^2$
γ	$0\,m^9/F/C^4$
ρ	$1.8\times10^{-3}\,\Omega\,m^{\dagger}$

† Determined by Fourier Infrared spectroscopy analysis for HfO_2 [30].

data in terms of both $I_D - V_{GS}$ and $I_D - V_{DS}$ characteristics shows that the essential features of the MoS_2 NCFET physics are successfully captured, which is sufficient for the purpose of demonstrating the possibility of enhancing the performance of NC-BioFET.

NCFET Noise Model

In addition to the MOSFET noise sources already discussed in Section 5.2.1, NCFETs have an additional thermal noise source introduced by the ferroelectric layer. This noise source stems from the dissipative process due to the damped ferroelectric switching [31], which is expressed by the last term in Eq. (5.25). The *PSD* of the ferroelectric thus reads

$$PSD^{ferro} = 4k_BT\rho t_{FE}. \tag{5.29}$$

Interestingly, in [26] it was found that Flicker noise associated with trapping/detrapping events at the semiconductor/insulator interface reduces with increasing ferroelectric thickness. This is due to the fact C_{FE} reduces with increasing t_{FE}, causing a better matching with C_{MOS} of

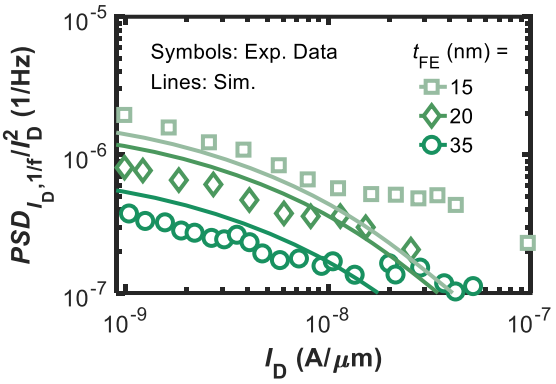


Figure 5.11: Comparison of Eq. (5.30) (lines) with experimental data (symbols) from [26]. The expression successfully reproduces the observed decreasing Flicker noise *PSD* trend with increasing t_{FE}. Adapted from [36].

the underlying transistor. In other words, the equivalent capacitance $C_{eq} = \left(C_{MOS}^{-1} - |C_{FE}|^{-1}\right)^{-1}$ *increases* with increasing t_{FE} thus reducing Flicker noise. The expression for Flicker noise of the NCFET taking into account this experimental observation can be obtained from Eq. (5.15)[8]as follows

$$PSD_{I_D}^{flicker,NCFET} = \frac{q^2 k_B T \lambda N_T}{fWLC_{eq}^2}(g_m + \alpha_C \mu_n C_{eq} I_D)^2. \tag{5.30}$$

The validity of this expression is verified by the comparison with data in [26] as shown in Fig. 5.11.

5.3.2 Sensitivity Improvement

Fig. 5.12(a) shows the NC-BioFET device under study. We consider the ferroelectric layer to be the insulator in touch with the electrolytic solution for simplicity[9], though in general this needs not be the case. In [37] a NCFET-bases ISFET was realized in which a ferroelectric capacitor is placed between the MOSFET and the sensing surface, leading to similar results (concerning sensitivity improvement) to those presented here. Fig. 5.12(b) shows the simplified circuit schematic of the NC-BioFET. The schematic highlights the fact that the single elements

[8]Note that the expression for the Flicker noise, Eq. (5.15) was originally derived for a long-channel bulk MOSFET considering both carriers' number and mobility fluctuations. Its validity for MoS_2 FETs was discussed in [33], and it was shown to describe well the intrinsic noise for this class of devices.

[9]Accordingly, site-binding model parameters for HfO_2 are taken [32].

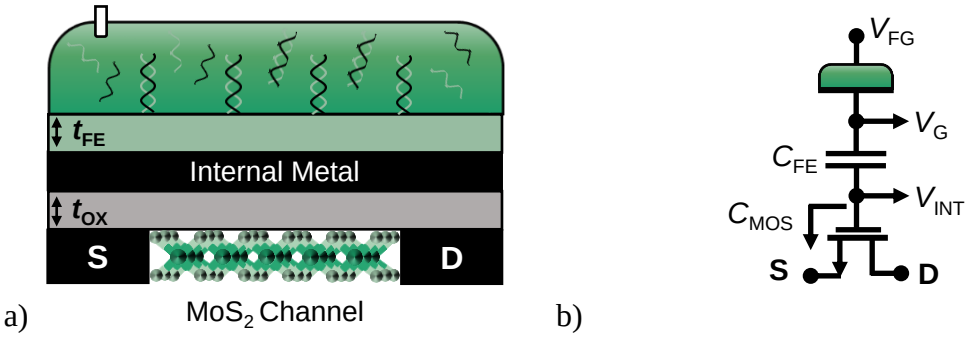


Figure 5.12: (a) Illustration of the cross-section of the NC-BioFET under study. Note that the electrolytic solution lies on the ferroelectric layer. (b) Simplified circuit schematic of the system. Adapted from [36].

are all modeled separately and solved self-consistently within the SPICE simulator used.

The $I-V$ characteristics of the NC-BioFET for different t_{FE} are illustrated in Fig. 5.13 along with the corresponding A_V (see Eq. (5.18)). As we will see, the increased sensitivity of the NC-BioFET (compared to the conventional BioFET, i.e., the case for which $t_{FE} = 0\,\text{nm}$ in Fig. 5.13) stems from the cancellation of the ferroelectric negative capacitance by the equivalent MOSFET capacitance, i.e., $C_{MOS} = |C_{FE}|$ [22]. At this condition, i.e., at the critical point shown in Fig. 5.13(a), A_V exhibits a strong peak, see Fig. 5.13(b). The peak occurs only if the negative capacitor layer is sufficiently thick to compensate the MOSFET capacitance C_{MOS}, indicated in the circuit schematic in Fig. 5.12(b).

The instability at the critical point giving rise to the abrupt current change is the cause for enhanced sensitivity, as shown in Fig. 5.14. Sensitivity is defined as the ratio between the current before and after the capture of biomolecules, i.e.,

$$S_I = \frac{I_{before}}{I_{after}}. \tag{5.31}$$

Here, we consider negatively charged biomolecules (e.g., DNA) that increase V_t of the NCFET and shift the $I-V$ characteristics to the right, see Fig. 5.13(a). The shift in $I-V$ characteristics due to increasing concentration of captured biomolecules (N_{BIO}) triggers the abrupt transition from inversion to sub-threshold regime, thereby producing a significant change in I_D, as shown in Fig. 5.14(b). Sensitivity in Fig. 5.14 is plotted at the critical point condition (determined for

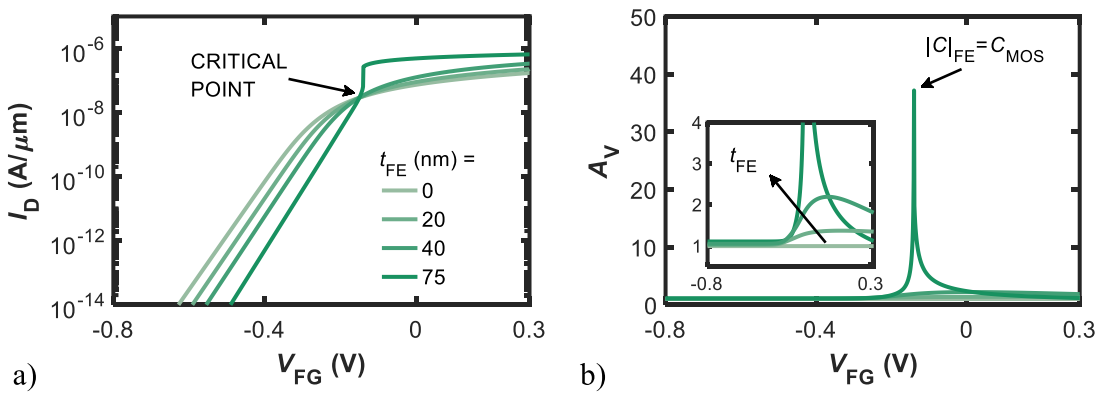


Figure 5.13: (a) $I_D - V_{FG}$ and (b) $A_V - V_{FG}$ characteristics for different t_{FE} (see legend) showing the an abrupt transition at the critical point occurring when the capacitance matching condition is reached ($C_{MOS} = |C_{FE}|$). V_{DS} is set to 0.1 V. The case with $t_{FE} = 0$ nm is equivalent to a conventional BioFET and serves as a reference for comparing NC-BioFET performance. Adapted from [36].

the $t_{FE} = 75$ nm case, as discussed later), confirming the fact that the steep sensitivity increase is obtained at the capacitance matching condition. Note that one must scan I_D by sweeping V_{FG} to achieve the improved gain promised by the NC-BioFET at the critical points. As expected, the response of the NC-BioFET is indistinguishable from that of a conventional BioFET as long as the instability is absent (i.e., for low t_{FE}), see the inset in Fig. 5.14(a). This can be understood by the fact that the screening remains the fundamental limiting factor to the response to the capture of biomolecules [18].

Note that both C_{ox} and C_{FE} must be tailored carefully (through proper choice of t_{ox} and t_{FE}) to control the degree of hysteresis present in the $I - V$ characteristics [13], [27] and to trade it off with the corresponding gain in the NC-BioFET sensitivity. An approximate closed-form expression for V_{FG} at the critical point can be written as follows [37, Ch. 3]

$$V_{G,CP} = \psi_e + V_{FB} + V_{th} \ln\left(-V_{th} \frac{1}{(2\alpha t_{FE} + 1/C_{ox})} \frac{1}{qN_{2D}}\right) - V_{th} - \frac{qN_d}{C_{ox}}. \tag{5.32}$$

5.3.3 Noise Rejection and SNR Improvement

We now discuss the noise rejection and SNR improvement of the NC-BioFETs. To evaluate the overall noise of the system, we consider the noise sources separately and obtain the output *PSD* by

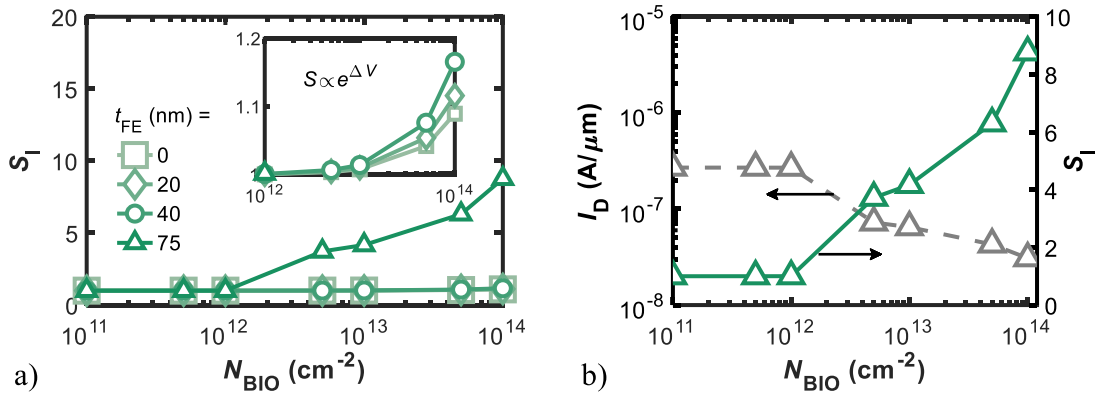


Figure 5.14: (a) Current sensitivity S_I vs biomolecule concentration N_{BIO} for different t_{FE} (see legend). The abrupt current variation at the critical point for the $t_{FE} = 75\,\text{nm}$ is shown in (b). The inset in (a) shows a magnification of the plot for low S_I values, indicating that if t_{FE} is not thick enough to achieve capacitance matching the sensitivity follows the conventional BioFETs exponential dependence on the potential shift. Adapted from [36].

summing up the single contributions (assuming statistical independence between the individual processes). The four different noise contributions considered in this work are shown in Fig. 5.15. Results indicate that the thermal noise coming from the conductive electrolyte, Fig. 5.15(a), and the ferroelectric, Fig. 5.15(b), increase as t_{FE} increases. The increase is explained by the fact that the negative capacitor noise is directly proportional to t_{FE}, see Eq. (5.29), whereas the electrolyte *PSD* increases due to the voltage amplification given by the negative capacitor. As expected, the MOSFET channel noise, see Fig. 5.15(c), does not vary with t_{FE}, as it is due to the conduction mechanism in the semiconductor that is not affected by the presence of the ferroelectric.

Note that all these noise contributions are negligible compared to the Flicker noise of the transistor, Fig. 5.15(d), hence they are not a limiting factor to the SNR of the system. Flicker noise is indeed found to be order of magnitudes higher than the other noise sources for the bias range of interest. We ascribe this to the relatively high defect density ($\approx 1 \times 10^{12}\,\text{eV}^{-1}\,\text{cm}^{-3}$) necessary to reproduce the experimental data in Fig. 5.11. The total drain current *PSD* thus coincides with the Flicker noise *PSD*, thus it exhibits the same decreasing trend with increasing t_{FE}. The reason for the *PSD* reduction was explained previously as due to the better capacitance matching with increasing t_{FE} that increases the *total gate capacitance*, C_{eff}, inversely proportional to $PSD_{I_D}^{flicker,NCFET}$, see Eq. (5.30). The noise reduction provided by the NCFET is particularly

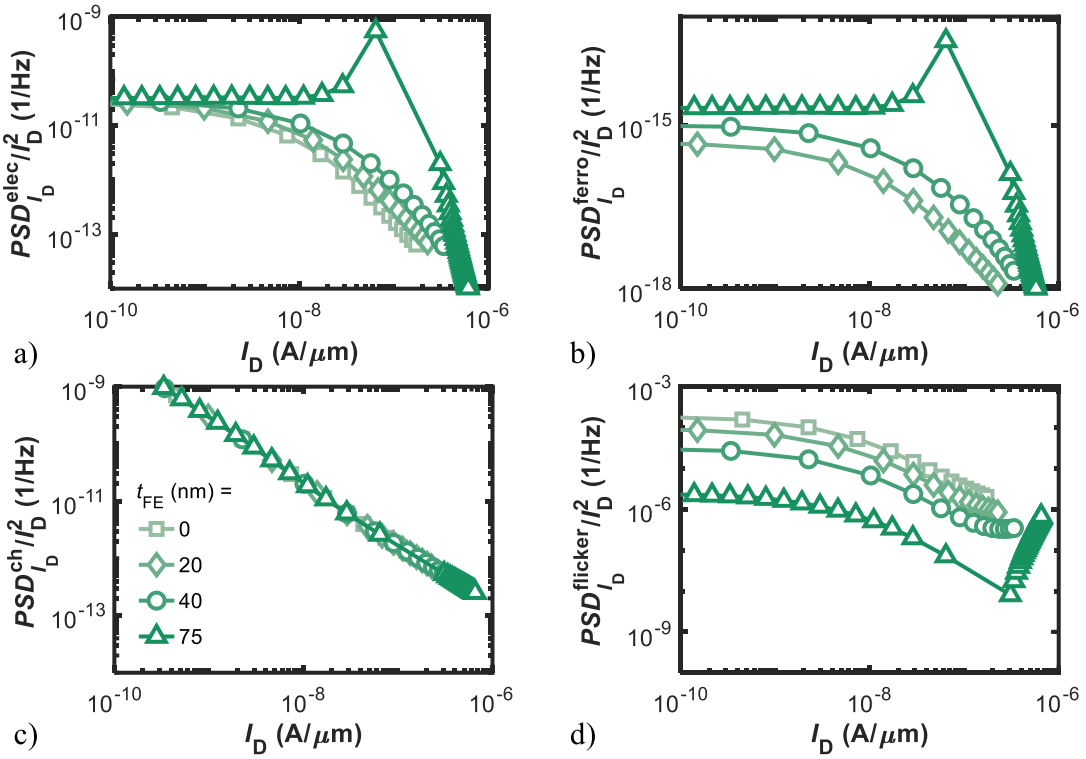


Figure 5.15: Noise contributions to the total power spectral density for different t_{FE} (see legend): (a) electrolyte, (b) ferroelectric, MOSFET (c) channel and (d) Flicker noise. While electrolyte and ferroelectric noise increase, channel noise remains unvaried and Flicker noise reduces with increasing t_{FE}. Adapted from [36].

important for BioFET given the difficulty in providing low noise amplifiers on small geometries due to the high value of the intrinsic Flicker noise in these devices [34].

The Flicker noise reduction, along with the sensitivity increase, leads to improved SNR defined as follows [35]

$$SNR = \frac{\Delta I_D}{\delta I_{D,n}} \tag{5.33}$$

where ΔI_D is the difference between the current before and after the capture of biomolecules, and $\delta I_{D,n}$ is the noise signal superimposed to the DC value, calculated from the total noise *PSD* as $\delta I_{D,n} = \sqrt{\int PSD_{I_D} df}$. Note that the SNR can be more conveniently expressed per unit voltage to have a quantitative expression for the intrinsic SNR that is independent of the input signal. The

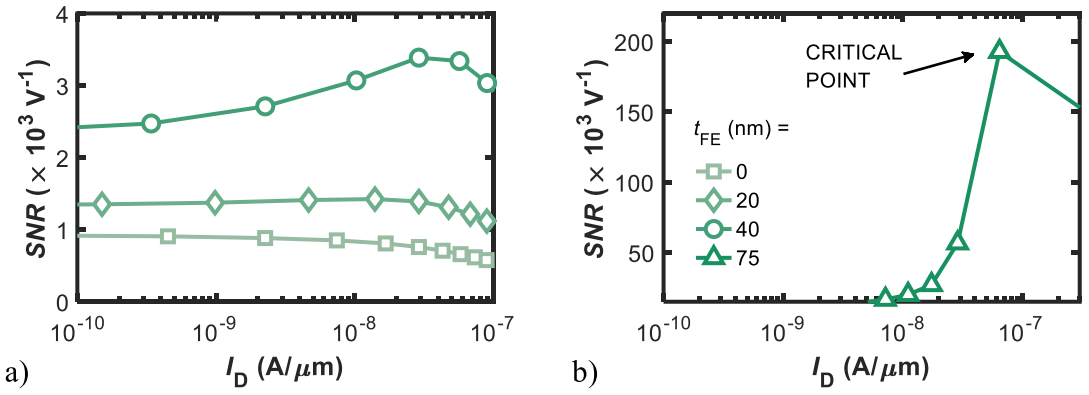


Figure 5.16: Intrinsic SNR trend of the NC-BioFET with different t_{FE}. (a) shows SNR for small t_{FE} and (b) for $t_{FE} = 75\,\text{nm}$ (i.e., the value triggering the abrupt current switching at the critical point). For $t_{FE} = 75\,\text{nm}$, the SNR peaks at the peak of g_m. Adapted from [36].

normalized SNR reads

$$SNR_n = \frac{1}{\Delta V_{FG}} \frac{g_m \Delta V_{FG}}{\delta I_{D,n}}. \tag{5.34}$$

Intrinsic SNR for the NC-BioFET with different t_{FE} is shown in Fig. 5.16. The SNR shows qualitatively an opposite behavior to the total *PSD* (even for small t_{FE}, shown in Fig. 5.16(a)) which can be attributed to the Flicker noise reduction, see Fig. 5.15(d). Fig. 5.16(b) shows that for $t_{FE} = 75\,\text{nm}$, SNR exhibits a peak corresponding to the g_m peak of the NCFET, in accordance with Eq. (5.34). The peak SNR of the NC-BioFET is about $2 \times 10^5\ \text{V}^{-1}$, which is almost two orders of magnitude higher than that of a Silicon NW BioFET [35].

Although this remarkably high sensitivity (and thus SNR) is achieved with a relatively thick ferroelectric (i.e., 75 nm) in this case, typical long channel biosensors can easily accommodate such a thick film [7], [11]. On the other hand, it is also possible to reduce the ferroelectric thickness in NC-BioFET and still obtain similar benefits in terms of SNR by properly doping the HfO_2 layer. For example, as illustrated in [36], by using Si instead of Zr as the doping atoms for HfO_2, one can use a ferroelectric with $t_{FE} = 16\,\text{nm}$ and still obtaining similar SNR peak value compared to the case with HZO. This result shows that the NC-BioFETs can effectively be scaled to realize nano-biosensors with enhanced SNR.

5.3.4 Conclusions

To summarize, we have proposed the concept of NC-BioFET as a novel class of devices exploiting the NC effect to improve SNR of BioFETs. We found that, upon the triggering of the non-linearity associated with the NC effect, the sensitivity of NC-BioFETs is significantly improved compared to that of a conventional BioFET. At the same time, rejection of the Flicker noise of the underlying transistor increases due to the higher equivalent gate capacitance of the NCFET, leading to higher SNR. Finally, by properly choosing the ferroelectric material, it is possible to scale t_{FE} while still obtaining the instability required to improve SNR, hence proving the scalability of NC-BioFETs within current CMOS process.

References – Chapter 5

[1] P. R. Nair and M. A. Alam, "Screening-limited response of NanoBiosensors," *Nano Letters*, vol. 8, no. 5, pp. 1281–1285, 2008. DOI: 10.1021/nl072593i.

[2] B. R. Dorvel, B. Reddy, J. Go, C. D. Guevara, E. Salm, M. A. Alam, and R. Bashir, "Silicon nanowires with high-k hafnium oxide dielectrics for sensitive detection of small nucleic acid oligomers," *ACS Nano*, vol. 6, pp. 6150–6164, 7 Jul. 2012. DOI: 10.1021/nn301495k.

[3] M. J. Deen, M. W. Shinwari, J. C. Ranuárez, and D. Landheer, "Noise considerations in field-effect biosensors," *Journal of Applied Physics*, vol. 100, 7 2006. DOI: 10.1063/1.2355542.

[4] P. Namdari, H. Daraee, and A. Eatemadi, "Recent advances in silicon nanowire biosensors: Synthesis methods, properties, and applications," *Nanoscale Research Letters*, vol. 11, 1 2016. DOI: 10.1186/s11671-016-1618-z.

[5] D. A. Wheeler and J. M. Rothberg, "The complete genome of an individual by massively parallel dna sequencing," *Nature*, vol. 452, pp. 872–876, 7189 Apr. 2008. DOI: 10.1038/nature06884.

[6] J. Go and M. A. Alam, "The future scalability of ph-based genome sequencers: A theoretical perspective," *Journal of Applied Physics*, vol. 114, p. 164 311, 16 Oct. 2013. DOI: 10.1063/1.4825119.

[7] P. Dak, W. Seo, B. Jung, and M. A. Alam, "A physics-based (Verilog-A) compact model for DC, quasi-static transient, small-signal, and noise analysis of MOSFET-based pH

sensors," *IEEE Transactions on Electron Devices*, vol. 64, no. 3, pp. 1277–1285, Mar. 2017. DOI: 10.1109/TED.2017.2651905.

[8] P. Bergveld, "Thirty years of ISFETOLOGY: What happened in the past 30 years and what may happen in the next 30 years," *Sensors and Actuators, B: Chemical*, vol. 88, no. 1, pp. 1–20, Jan. 2003. DOI: 10.1016/S0925-4005(02)00301-5.

[9] P. Dak, P. Nair, J. Go, and M. A. Alam, "Extended-gate biosensors achieve fluid stability with no loss in charge sensitivity," in *71st Device Research Conference*, vol. 20, IEEE, Jun. 2013, pp. 105–106. DOI: 10.1109/DRC.2013.6633815.

[10] J. Lee, P. Dak, Y. Lee, H. Park, W. Choi, M. A. Alam, and S. Kim, "Two-dimensional layered MoS2 biosensors enable highly sensitive detection of biomolecules," *Scientific Reports*, vol. 4, no. 1, p. 7352, May 2014. DOI: 10.1038/srep07352.

[11] D. Sarkar, W. Liu, X. Xie, A. C. Anselmo, S. Mitragotri, and K. Banerjee, "MoS2field-effect transistor for next-generation label-free biosensors," *ACS Nano*, vol. 8, no. 4, pp. 3992–4003, 2014. DOI: 10.1021/nn5009148.

[12] H. Nam, B. R. Oh, P. Chen, M. Chen, S. Wi, W. Wan, K. Kurabayashi, and X. Liang, "Multiple MoS2 transistors for sensing molecule interaction kinetics," *Scientific Reports*, vol. 5, no. 1, p. 10 546, Sep. 2015. DOI: 10.1038/srep10546.

[13] S. Salahuddin and S. Datta, "Use of negative capacitance to provide voltage amplification for low power nanoscale devices," *Nano Letters*, vol. 8, no. 2, pp. 405–410, 2008. DOI: 10.1021/nl071804g.

[14] K. Karda, A. Jain, C. Mouli, and M. A. Alam, "An anti-ferroelectric gated Landau transistor to achieve sub-60 mV/dec switching at low voltage and high speed," *Applied Physics Letters*, vol. 106, no. 16, p. 163 501, Apr. 2015. DOI: 10.1063/1.4918649.

[15] J. Go, P. R. Nair, and M. A. Alam, "Theory of signal and noise in double-gated nanoscale electronic pH sensors," *Journal of Applied Physics*, vol. 112, no. 3, p. 034 516, 2012. DOI: 10.1063/1.4737604.

[16] Y. Cui, Q. Wei, H. Park, and C. M. Lieber, "Nanowire Nanosensors for Highly Sensitive and Selective Detection of Biological and Chemical Species," *Science*, vol. 293, no. 5533, pp. 1289–1292, 2001. DOI: 10.1126/science.1062711.

[17] J.-i. Hahm and C. M. Lieber, "Direct Ultrasensitive Electrical Detection of DNA and DNA Sequence Variations Using Nanowire Nanosensors," *Nano Letters*, vol. 4, no. 1, pp. 51–54, Jan. 2004. DOI: 10.1021/nl034853b.

[18] J. Go, P. R. Nair, B. Reddy, B. Dorvel, R. Bashir, and M. A. Alam, "Coupled heterogeneous nanowire-nanoplate planar transistor sensors for giant (>10 V/pH) nernst response," *ACS Nano*, vol. 6, no. 7, pp. 5972–5979, 2012. DOI: 10.1021/nn300874w.

[19] P. R. Nair and M. A. Alam, "Performance limits of nanobiosensors," *Applied Physics Letters*, vol. 88, no. 23, p. 233 120, Jun. 2006. DOI: 10.1063/1.2211310.

[20] M. A. Alam, M. Si, and P. D. Ye, "A critical review of recent progress on negative capacitance field-effect transistors," *Applied Physics Letters*, vol. 114, no. 9, p. 090 401, Mar. 2019. DOI: 10.1063/1.5092684.

[21] G. Pahwa, T. Dutta, A. Agarwal, S. Khandelwal, S. Salahuddin, C. Hu, and Y. S. Chauhan, "Analysis and Compact Modeling of Negative Capacitance Transistor with High ON-Current and Negative Output Differential Resistance - Part I: Model Description," *IEEE Transactions on Electron Devices*, vol. 63, no. 12, pp. 4981–4985, Dec. 2016. DOI: 10.1109/TED.2016.2614432.

[22] ——, "Analysis and Compact Modeling of Negative Capacitance Transistor with High ON-Current and Negative Output Differential Resistance - Part II: Model Validation," *IEEE Transactions on Electron Devices*, vol. 63, no. 12, pp. 4986–4992, Dec. 2016. DOI: 10.1109/TED.2016.2614436.

[23] A. Aziz, S. Ghosh, S. Datta, and S. K. Gupta, "Physics-Based Circuit-Compatible SPICE Model for Ferroelectric Transistors," *IEEE Electron Device Letters*, vol. 37, no. 6, pp. 805–808, Jun. 2016. DOI: 10.1109/LED.2016.2558149.

[24] M. A. Alam, P. Dak, M. A. Wahab, and X. Sun, "Physics-based compact models for insulated-gate field-effect biosensors, landau-transistors, and thin-film solar cells," in *2015 IEEE Custom Integrated Circuits Conference (CICC)*, Sep. 2015, pp. 1–8. DOI: 10.1109/CICC.2015.7338406.

[25] T. Yoshizumi and Y. Miyahara, "Field-Effect Transistors for Gas Sensing," in *Different Types of Field-Effect Transistors - Theory and Applications*, Jun. 2017, ch. 8, pp. 149–163. DOI: 10.5772/intechopen.68481.

[26] S. Alghamdi, M. Si, L. Yang, and P. D. Ye, "Low frequency noise in MOS<inf>2</inf> negative capacitance field-effect transistor," in *2018 IEEE International Reliability Physics Symposium (IRPS)*, May 2018, P–TX.1–1–P–TX.1–5. DOI: 10.1109/IRPS.2018.8353696.

[27] M. Si, C. J. Su, C. Jiang, N. J. Conrad, H. Zhou, K. D. Maize, G. Qiu, C. T. Wu, A. Shakouri, M. A. Alam, and P. D. Ye, "Steep-slope hysteresis-free negative capacitance MoS2 Transistors," *Nature Nanotechnology*, vol. 13, no. 1, pp. 24–28, 2018. DOI: 10.1038/s41565-017-0010-1.

[28] F. A. Mcguire, Y.-C. C. Lin, K. Price, G. B. Rayner, S. Khandelwal, S. Salahuddin, and A. D. Franklin, "Sustained Sub-60 mV/decade switching via the negative capacitance effect in MoS2 Transistors," *Nano Letters*, vol. 17, no. 8, pp. 4801–4806, 2017. DOI: 10.1021/acs.nanolett.7b01584.

[29] C. Jiang, M. Si, R. Liang, J. Xu, P. D. Ye, and M. A. Alam, "A Closed Form Analytical Model of Back-Gated 2-D Semiconductor Negative Capacitance Field Effect Transistors," *IEEE Journal of the Electron Devices Society*, vol. 6, no. 1, pp. 189–194, 2018. DOI: 10.1109/JEDS.2017.2787137.

[30] K. Chatterjee, A. J. Rosner, and S. Salahuddin, "Intrinsic speed limit of negative capacitance transistors," *IEEE Electron Device Letters*, vol. 38, no. 9, pp. 1328–1330, Sep. 2017. DOI: 10.1109/LED.2017.2731343.

[31] A. I. Khan, K. Chatterjee, B. Wang, S. Drapcho, L. You, C. Serrao, S. R. Bakaul, R. Ramesh, and S. Salahuddin, "Negative capacitance in a ferroelectric capacitor," *Nature Materials*, vol. 14, no. 2, pp. 182–186, Feb. 2015. DOI: 10.1038/nmat4148.

[32] Ajay, R. Narang, M. Saxena, and M. Gupta, "Analytical Model of pH sensing Characteristics of Junctionless Silicon on Insulator ISFET," *IEEE Transactions on Electron Devices*, vol. 64, no. 4, pp. 1742–1750, Apr. 2017. DOI: 10.1109/TED.2017.2668520.

[33] J. Renteria, R. Samnakay, S. L. Rumyantsev, C. Jiang, P. Goli, M. S. Shur, and A. A. Balandin, "Low-frequency 1/f noise in MoS2 transistors: Relative contributions of the channel and contacts," *Applied Physics Letters*, vol. 104, no. 15, p. 153 104, 2014. DOI: 10.1063/1.4871374.

[34] T. C. Nguyen, W. Qiu, M. Altissimo, P. G. Spizzirri, L. H. W. van Beveren, and E. Skafidas, "CMOS-Compatible Nanowire Biosensors," in *Graphene, Carbon Nanotubes, and Nanostructures: Techniques and Applications*, 2013, pp. 265–278.

[35] N. K. Rajan, D. A. Routenberg, and M. A. Reed, "Optimal signal-to-noise ratio for silicon nanowire biochemical sensors," *Applied Physics Letters*, vol. 98, no. 26, p. 264 107, Jun. 2011. DOI: 10.1063/1.3608155.

[36] N. Zagni, P. Pavan, and M. A. Alam, "Two-dimensional MoS2 negative capacitor transistors for enhanced (super-Nernstian) signal-to-noise performance of next-generation nano biosensors," *Applied Physics Letters*, vol. 114, no. 23, p. 233 102, Jan. 2019. DOI: 10.1063/1.5097828.

[37] F. Bellando, C. K. Dabhi, A. Saeidi, C. Gastaldi, Y. S. Chauhan, and A. M. Ionescu, "Subthermionic negative capacitance ion sensitive field-effect transistor," *Applied Physics Letters*, vol. 116, no. 17, p. 173 503, 2020, ISSN: 0003-6951. DOI: 10.1063/5.0005411.

Chapter 6

Conclusions and Outlook

This thesis is devoted to the use of both device simulation tools and custom physics-based compact models to predict the performance and reliability of 21-st century electronics. Specifically, four different application scenarios and relative enabling technologies are discussed. These are: *i)* III-V MOSFETs for logic/digital circuits; *ii)* RRAMs and FeFETs for non-volatile memory and in-memory computing; *iii)* GaN-based HEMTs for power applications; and *iv)* NCFETs to enhance performance of nano-biosensors.

Regarding III-V MOSFETs, we found that traps affect mobility measurements significantly and give rise to hysteresis and frequency dispersion in the device electrical characteristics. These effects need to be taken into account for a correct evaluation of III-V MOSFETs performance and potential over conventional Si counterparts. Moreover, we discussed the intrinsic parameter fluctuations giving rise to the so-called *variability*, i.e., variations in the performance of identically drawn devices. We found that dimension scaling, conventionally employed to improve device performance, leads to higher variability especially due to interface traps, thus limiting III-V-based devices potential over Si-counterparts.

Regarding NVMs, we developed a compact model to account for noise in RRAMs in both resistive states that accurately matches experimental data distribution of resistance variations. This model can be effectively used to perform accurate transient simulations of RRAMs and in the design of novel applications such as random number generators. Reliability of FeFETs in terms of endurance was analyzed by means of an analytical model of the memory window. This simple

model is able to seamlessly account for the effect of generated traps on the *MW* closure leading to the limited endurance of this technology. By means of the developed model it is possible to perform early assessment of the endurance of novel FeFETs, accelerating development.

Concerning GaN-based power HEMTs, we investigated the role of buffer traps due to Carbon doping in determining *i* the off-state breakdown, *ii)* the R_{ON} degradation and recovery transients, and *iii)* the V_t bidirectional shifts. Simulations were calibrated with data from actual devices to validate the trap model provided for Carbon doping, based on a deep acceptor and a shallow donor trap. Particularly, the deep acceptor trap (0.9 eV above the VB edge) determines the breakdown voltage scaling with the lateral dimension and gives rise to dispersion effects (of both V_t and R_{ON}) leading to the so-called *current-collapse*. The developed model sheds light on the origin of these dispersion effects, which are due to emission/capture of holes in the VB by the deep acceptors (depending on the applied stress bias).

Finally, the intrinsic *Nernst-limit* to the sensitivity of FET-based nano-biosensors can be overcome by exploting the *negative capacitance* effect of properly stabilized ferroelectric layers. By means of a compact model, the performance of a MoS_2-based BioFET in the NCFET configuration (i.e., including a ferroelectric layer in the gate stack) were analyzed. It was found that by proper design of the NCFET, sensitivity could be dramatically improved while increasing the Flicker noise rejection at the same time. This combined sensitivity improvement and noise rejection leads to greater SNR compared to conventional BioFETs.

Outlook

New paradigms are emerging for accelerating the prediction of performance (and reliability) of novel electronic devices thanks to the developments in the field of AI algorithms (and technology). Within these novel frameworks, simulations — calibrated on experimental data — are used to generate the pool of data to train a machine to perform tasks of either inference or classification. In this way, both training and testing tasks can be accelerated and costs can be drastically reduced, as experiments get 'replaced' by simulations [1].

As machines get more evolved, by learning quickly and more accurately, one could get the

feeling that knowledge of device physics and development of predictive models could become consequently less and less important. However, in the future these two realms will more likely 'contaminate' rather than eliminate each other. This way, both machines' ability to learn and humans' to develop models will improve. However, the models developed in this new framework will likely be very different from the ones more customary today. These would arguably be more devoted to capture the 'essential' physics of a process (that could be validated by prior knowledge and physical acumen), leaving the detailed and cumbersome task of 'perfect matching' of experimental data to the trained machines.

The bottom line is, whatever new development scenario might emerge from the integration of novel tools to the modeling/simulation of semiconductor devices, understanding of the physics and of reliability aspects will still be crucial to the success and continuing improvement of novel technology.

References

[1] H. Y. Wong, M. Xiao, B. Wang, Y. K. Chiu, X. Yan, J. Ma, K. Sasaki, H. Wang, and Y. Zhang, "TCAD-Machine learning framework for device variation and operating temperature analysis with experimental demonstration," IEEE J. Electron Devices Soc., vol. 8, pp. 992-1000, 2020. DOI: 10.1109/JEDS.2020.3024669.

List of Publications

List of published and under review contributions covered in the PhD Thesis (in chronological order).

Journal Articles

- P. Pavan, N. Zagni, F. M. Puglisi, A. Alian, A. V. Y. Thean, N. Collaert, and G. Verzellesi, "The impact of interface and border traps on current–voltage, capacitance–voltage, and split-CV mobility measurements in InGaAs MOSFETs," Phys. Status Solidi A, vol. 214, no. 3, p. 1600592, Mar. 2017. DOI: 10.1002/pssa.201600592.

- N. Zagni, F. M. Puglisi, G. Verzellesi, and P. Pavan, "Threshold Voltage Statistical Variability and Its Sensitivity to Critical Geometrical Parameters in Ultrascaled InGaAs and Silicon FETs," IEEE Trans. Electron Devices, vol. 64, no. 11, pp. 4607–4614, Nov. 2017. DOI: 10.1109/TED.2017.2754323.

- F. M. Puglisi, N. Zagni, L. Larcher, and P. Pavan, "Random Telegraph Noise in Resistive Random Access Memories: Compact Modeling and Advanced Circuit Design," IEEE Trans. Electron Devices, vol. 65, no. 7, pp. 2964–2972, 2018. DOI: 10.1109/TED.2018.2833208.

- N. Zagni, F. M. Puglisi, P. Pavan, and G. Verzellesi, "Effects of mole fraction variations and scaling on total variability in InGaAs MOSFETs," Solid. State. Electron., vol. 159, pp. 135–141, Mar. 2019. DOI: 10.1016/J.SSE.2019.03.048.

- N. Zagni, P. Pavan, and M. A. Alam, "Two-dimensional MoS2 negative capacitor tran-

sistors for enhanced (super-Nernstian) signal-to-noise performance of next-generation nano biosensors," Appl. Phys. Lett., vol. 114, no. 23, p. 233102, Jun. 2019. DOI: 10.1063/1.5097828.

- N. Zagni, F. M. Puglisi, P. Pavan, A. Chini, and G. Verzellesi, "Insights into the off-state breakdown mechanisms in power GaN HEMTs," Microelectron. Reliab., vol. 100–101, p. 113374, Sep. 2019. DOI: 10.1016/j.microrel.2019.06.066.
- N. Zagni, A. Chini, F. M. Puglisi, P. Pavan, and G. Verzellesi, "The Role of Carbon Doping on Breakdown, Current Collapse, and Dynamic On-Resistance Recovery in AlGaN/GaN High Electron Mobility Transistors on Semi-Insulating SiC Substrates," Phys. status solidi, vol. 217, p. 1900762, Dec. 2019. DOI: 10.1002/pssa.201900762.
- N. Zagni, E. Caruso, F. M. Puglisi, P. Pavan, P. Palestri, and G. Verzellesi, "Systematic Modeling of Electrostatics, Transport, and Statistical Variability Effects of Interface Traps in End-of-the-Roadmap III–V MOSFETs," IEEE Trans. Electron Devices, vol. 67, no. 4, pp. 1560–1566, Apr. 2020. DOI: 10.1109/TED.2020.2974966.
- N. Zagni, P. Pavan, and M. A. Alam, "A memory window expression to evaluate the endurance of ferroelectric FETs," Appl. Phys. Lett., vol. 117, no. 15, p. 152901, Oct. 2020. DOI: 10.1063/5.0021081.
- N. Zagni, A. Chini, F. M. Puglisi, P. Pavan, and G. Verzellesi, "The effects of carbon on the bidirectional threshold voltage instabilities induced by negative gate bias stress in GaN MIS-HEMTs," J. Comput. Electron., vol. 19, no. 4, pp. 1555–1563, Dec. 2020. DOI: 10.1007/s10825-020-01573-8.
- N. Zagni, A. Chini, F. M. Puglisi, M. Meneghini, G. Meneghesso, E. Zanoni, P. Pavan, and G. Verzellesi, "' Hole Redistribution' Model Explaining the Thermally Activated RON Stress/Recovery Transients in Carbon-Doped AlGaN/GaN Power MIS-HEMTs," IEEE Trans. Electron Devices, 2021. DOI: 10.1109/TED.2020.3045683.
- N. Zagni, M. Cioni, A. Chini, F. Iucolano, F. M. Puglisi, P. Pavan, and G. Verzellesi, "Mechanisms Underlying the Bidirectional VT Shift after Negative Bias Temperature

Instability Stress in Carbon-Doped Fully Recessed MIS-HEMTs," IEEE Trans. Electron Devices, Accepted, 2021. DOI: 10.1109/TED.2021.3063664.

Conference Proceedings

- P. Pavan, N. Zagni, F. M. Puglisi, A. Alian, A. V. Y. Thean, N. Collaert, and G. Verzellesi, "Effects of border traps on transfer curve hysteresis and split-CV mobility measurement in InGaAs quantum-well MOSFETs," in 2016 Compound Semiconductor Week, CSW, 2016. DOI: 10.1109/ICIPRM.2016.7528574.
- N. Zagni, F. M. Puglisi, G. Verzellesi, and P. Pavan, "Combined variability/sensitivity analysis in III-V and silicon FETs for future technological nodes," in IEEE International Reliability Physics Symposium Proceedings, 2017, p. XT4.1-XT4.5. DOI: 10.1109/IRPS.2017.7936418.
- F. M. Puglisi, N. Zagni, L. Larcher, and P. Pavan, "A new verilog-A compact model of random telegraph noise in oxide-based RRAM for advanced circuit design," in European Solid-State Device Research Conference, 2017, pp. 204–207. DOI: 10.1109/ESSDERC.2017.8066627.
- N. Zagni, F. M. Puglisi, G. Verzellesi, and P. Pavan, "Variability and sensitivity to process parameters variations in InGaAs Dual-Gate Ultra-Thin Body MOSFETS: A scaling perspective," in 2017 27th International Symposium on Power and Timing Modeling, Optimization and Simulation, PATMOS 2017, 2017, pp. 1–5. DOI: 10.1109/PATMOS.2017.8106966.
- N. Zagni, F. M. F. M. Puglisi, P. Pavan, and G. Verzellesi, "Random dopant fluctuation variability in scaled InGaAs dual-gate ultra-thin body MOSFETs: Source and drain doping effect," in IEEE International Integrated Reliability Workshop Final Report, 2018, pp. 1–4. DOI: 10.1109/IIRW.2017.8361227.
- L. Selmi, E. Caruso, S. Carapezzi, M. Visciarelli, E. Gnani, N. Zagni, P. Pavan, P. Palestri, D. Esseni, A. Gnudi, S. Reggiani, F. M. Puglisi, and G. Verzellesi, "Modelling nanoscale n-MOSFETs with III-V compound semiconductor channels: From advanced models for band

structures, electrostatics and transport to TCAD," in IEEE International Electron Devices Meeting (IEDM), Dec. 2017, pp. 13.4.1-13.4.4. DOI: 10.1109/IEDM.2017.8268384.

- N. Zagni, F. M. Puglisi, P. Pavan, and G. Verzellesi, "On the impact of channel compositional variations on total threshold voltage variability in nanoscale InGaAs MOSFETs," 2018 Jt. Int. EUROSOI Work. Int. Conf. Ultim. Integr. Silicon, EUROSOI-ULIS 2018, pp. 1–4, Mar. 2018. DOI: 10.1109/ULIS.2018.8354745.
- N. Zagni, A. Chini, F. M. Puglisi, P. Pavan, M. Meneghini, G. Meneghesso, E. Zanoni, and G. Verzellesi, "Trap Dynamics Model Explaining the R ON Stress/Recovery Behavior in Carbon-Doped Power AlGaN/GaN MOS-HEMTs," in 2020 IEEE International Reliability Physics Symposium (IRPS), Apr. 2020, pp. 1–5. DOI: 10.1109/IRPS45951.2020.9128816.