Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
SCIENZE DELL'INGEGNERIA

XXI CICLO

COORDINATORE Prof. Stefano Trillo

MODELING AND RELIABILITY OF INNOVATIVE FLASH MEMORIES

Settore Scientifico Disciplinare ING-INF/01

Dottorando
Dott. ANDREA PADOVANI CHIMENTON

Tutore
Prof. ANDREA

Co-Tutore
Prof. LUCA LARCHER

Anni 2006/2008

Acknowledgments

These three years of work toward my Ph.D graduation have been a wonderful and intense period, during which I met, discussed with and had the pleasure to work with a lot of people from both academia and semiconductor industry. I am truly grateful to all of them: they all contributed to my personal and professional growth.

Above all, I am in debt with my Tutor, Prof. Andrea Chimenton, Università di Ferrara, and my co-Tutor, Prof. Luca Larcher, Università di Modena e Reggio Emilia. They have been great minds to work with, great teachers and also good friends.

I'm also really grateful to Prof. Paolo Pavan, Università di Modena e Reggio Emilia for his constant support. He has been fundamental for my personal and professional growth.

These three years would have not been the same without my true friend and colleague Riccardo Brama: sharing the office in Reggio Emilia with you was one of the best things of these three years. We also shared the experience of our first performance as speakers during the 2006 IRPS: that was definitely a bonding experience!

The last, but way more important thanks go to my family, Elena and Matteo, for being supportive and patient: if I reached what I have is only because of you.

To Elena and Matteo, my life

“We have a habit in writing articles published in scientific journals to make the work as finished as possible, to cover all the tracks, to not worry about the blind alleys or to describe how you had the wrong idea first, and so on. So there isn't any place to publish, in a dignified manner, what you actually did in order to get to do the work”

Richard P. Feynman, Nobel Lecture

Table of Contents

List of Acronyms

BGE – Band-Gap Engineered
BTBT – Band-To-Band-Tunneling
CB – Conduction Band
CBO – Conduction Band Offset
CG – Control Gate
CHE – Channel Hot Electron
CHEI – Channel Hot Electron Injection
CHISEL – CHannel Initiated Secondary ELectrons
CMOS – Complementary Metal Oxide Semiconductor
CP – Charge-Pumping
CT – Charge Trapping
DD – Drift-Diffusion
DT – Direct Tunneling
DV – Depletion Verify
ECC – Error Code Correction
ECU – Engine Control Unit
EOT – Equivalent Oxide Thickness
EV – Erase Verify
FE – Full-Erased
FG – Floating Gate
FN – Fowler-Nordheim
FeRAM – Ferroelectric RAMs
GCR – Gate Coupling Ratio
GIDL – Gate-Induced Drain Leakage
GPS – Global Positioning System
HDD – Hard-Disk Drives
HE – Half-Erased
HHI – Hot Hole Injection
IPD – Inter-Poly Dielectric
MC – Monte Carlo
MLC – Multi-Level Cell
MRAM – Magnetoresistive RAM
MTJ – Magnetic Tunnel Junction
NVM – Non-Volatile Memory
MC – Monte Carlo
ONO – Oxide-Nitride-Oxide
OP – Over Programming
PC – Personal Computer
PCM – Phase Change Memory
PDA – Personal Digital Assistant
PF – Poole Frenkel
PTAT – Phonon Trap-Assisted Tunneling
PV – Program Verify
RL – Read Level
SCE – Short-Channel Effects

SILC – Stress-Induced Leakage Current
SRH – Shockley-Read-Hall
SSD – Solid-State Drive
SSM – Subthreshold Slope Monitor
TAT – Trap Assisted-Tunneling
TBT – Trap-to-Band Tunneling
TE – Thermal Emission
TM – Temperature Monitor
TMR – Tunneling Magneto-Resistance
VARIOT – VARIable Oxide Thickness
WF – Work Function
WKB – Wentzel-Kramers-Brillouin

List of Figures

List of Tables

CHAPTER I

Introduction

SINCE their introduction in 1984, Flash memory devices have followed Moore's law keeping their structure basically unchanged. This era of *happy scaling* is now ending as Floating Gate (FG) devices – the mainstream Flash technology – cannot be further shrunk due to severe physical and technological limitations. As happened with the introduction of high-κ/metal gate structures for logic in late 2007, the time for innovation has come also for Flash memory devices.

Extensive research efforts are today devoted to the investigation of alternative non-volatile memory (NVM) devices able to overcome FG limits, thus enabling the scaling of Flash technology under the 30-nm technology node. Among the wide number of solutions proposed in the last years, nitride-based charge-trapping (CT) and bang-gap engineered (BGE) devices are two of the most promising technologies. This work describes part of the research on these topics author has been involved in during the XXI ciclo, Dottorato in Scienze dell'Ingegneria doctorate course.

Chapter II will give a thorough overview of FG Flash technology. After some considerations on current status and future trends of the NVM market, the main limits of the floating gate transistor will be discussed for both NOR and NAND architectures. Then, a quick overview of the most interesting innovative devices proposed for the post-Flash scenario will be given. Particular care will be devoted to charge-trapping and bang-gap engineered devices, as they are the topic of this thesis work. Their operating principles and expected improvements with respect to FG devices will be discussed and investigated.

In **Chapter III**, the statistical Monte Carlo (MC) simulator developed to reproduce leakage currents flowing through high-κ based layered structures will be introduced. We will show that simulations reproduce accurately experimental data measured on large area capacitors having both symmetric (SiO_2/high-κ/SiO_2) and asymmetric (SiO_2/high-κ) gate stacks, proving that the model catches correctly leakage current conduction mechanism

physics, thus being a valuable tool to investigate defect properties of high-κ composite dielectrics. Feasibility and optimization of band-gap engineered barriers for future NVM generations will be then investigated. Statistical simulations will be exploited to assess the real benefits of high-κ stacks as Flash memory tunnel dielectrics considering a 1Mb array of 65nm NAND Flash cells. We will show that the strong reliability improvements predicted by the adoption of BGE barriers disappear when trap-assisted contributions are included, warning on the possibility of replacing conventional tunnel oxides with the high-κ stacks

Author's research work on charge-trapping devices will be presented in the next three chapters. **Chapter IV** will present two tools allowing to profile program charge distributions in NROM devices based on I_D-V_{GS} sensitivity on local charge storage. Compact formulas to calculate length and density of the program charge distribution will be derived, and their accuracy will be tested for cells programmed at different levels and under different bias conditions. Tools accuracy and sensitivity will be investigated, and their limits when applied to erased NROM cells will be discussed.

As the tools presented in Chapter IV are sensitive only to the net charge above the channel, a new technique to profile hole distributions in erased NROM devices, which combines compact models, device simulations, and Gate-Induced Drain Leakage (GIDL), will be presented in **Chapter V**. Electron discharge effects are also taken into account, and the accuracy of the final charge scenario obtained will be verified by comparing drain and GIDL experimental currents to simulations. The technique will be used to monitor charge evolution after program and erase operations, allowing explaining some general mechanisms related to NROM reliability. First, it will be demonstrated that in cycled devices the amount of electrons in the nitride portion above the channel increases, because holes injected above the junction during erase shift the lateral field peak into the channel. Second, we will prove that V_{TR} drifts occurring in NROM cells left unbiased in the erased state are due to the lateral migration of trapped holes. The model presented allows explaining also the polarity dependence of the V_{TR} drift on the erase scheme adopted.

Chapter VI will describe a new physics-based model to simulate program transients of nitride-based future generation NAND devices (TANOS). Experimental results measured on TANOS devices with different oxide and nitride thicknesses will be reproduced with a great accuracy using an unique set of parameters, proving that the model catches correctly TANOS programming physics. In particular we will show that trapping process is independent from the energy of injected electrons, according to conventional Shockley-Read-Hall (SRH) theory, while electron de-trapping is dominated by tunneling. Finally, the evolution of the nitride charge during program will be investigated. We will show that the charge centroid is almost constant during the program transient and depends on the thickness of the nitride layer. These information are crucial for the optimization of TANOS memory cells.

Results will be summarized in **Chapter VII**.

CHAPTER II

Flash Memories: Past, Present and Future

"I do not see why somebody should need more than 64K of memory"

Bill Gates

The aim of this chapter is to provide a thorough overview of state-of-the-art Flash memory cells. Operating principles, charge injection mechanisms and scaling limits of Floating Gate devices are quickly reviewed. Then, an overview of the most interesting innovative devices proposed in the literature is presented, devoting particular attention to charge-trapping and bang-gap engineered devices, as they are the topic of this thesis work.

SINCE the introduction of the first device in 1984 [1] the Flash memory market has been continuously growing, driven by the increasing number of applications demanding for a flexible, low-cost and reliable solid-state memory. Historically, the three major markets for Flash memories have been related to Personal Computers (PC), wireless and telecom applications, and automotive electronics. More recently, Flash memory market has been driven by the large mass success of new portable electronic equipments like mp3 audio players, USB storage devices, smart cellular phones and digital cameras.

The evolution of the whole semiconductor and Non-Volatile Memory (NVM) markets are shown in Fig. II.1(a) and (b), respectively [2]. As can be seen, computers have been the main market driver for years. This means that devices such as DRAMs and Microprocessors have driven the largest sales and volumes of IC companies, see Fig.II.1(a)-(b). Nevertheless, NVM market share is constantly increasing, mainly because of the high demand for high-capacity non-volatile memory devices for portable applications, see Fig. II.1(b).

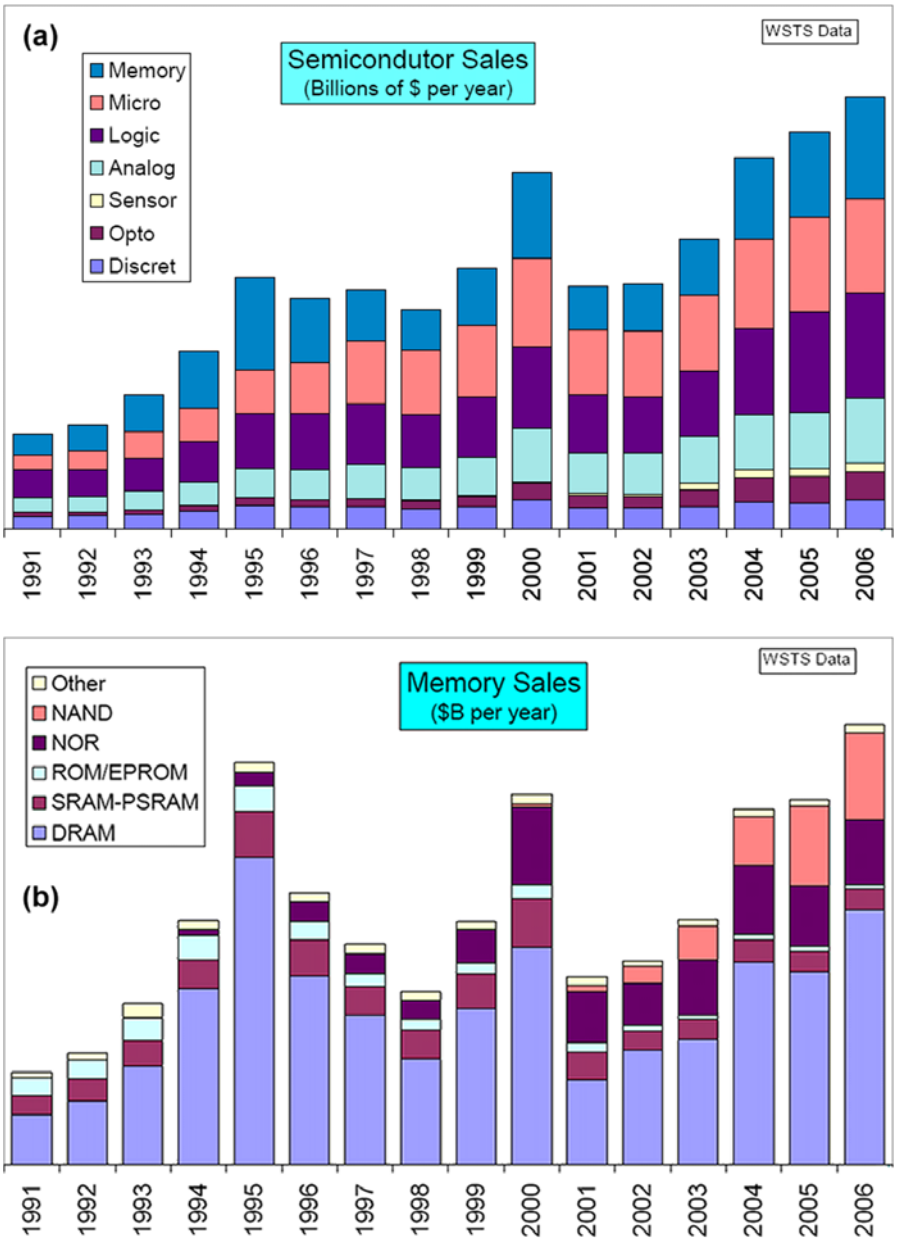

Figure II.1: *Recent evolution of (a) semiconductor and (b) memory markets.*

In 2006, the Complementary Metal Oxide Semiconductor (CMOS) memory market accounted for 23% of the total IC market with $58B sales, see Fig.II.1(a). Of these, $12B come from NAND Flash memories, while $9B were from NOR type Flash[1]. As can be noticed in Fig. II.1(b), the most growing semiconductor memory is Flash, especially NAND architecture, which sales increased exponentially since their introduction. On the contrary, NOR sales have been almost constant since 2001.

Presently, the Flash NVM market is in the range of $20 billions, but it is forecasted to grow with a higher average annual rate than DRAM and SRAM, reaching the $50 billions in 2011, see Fig. II.2.

[1] NOR and NAND architectures will be discussed in Paragraph II.3.2.

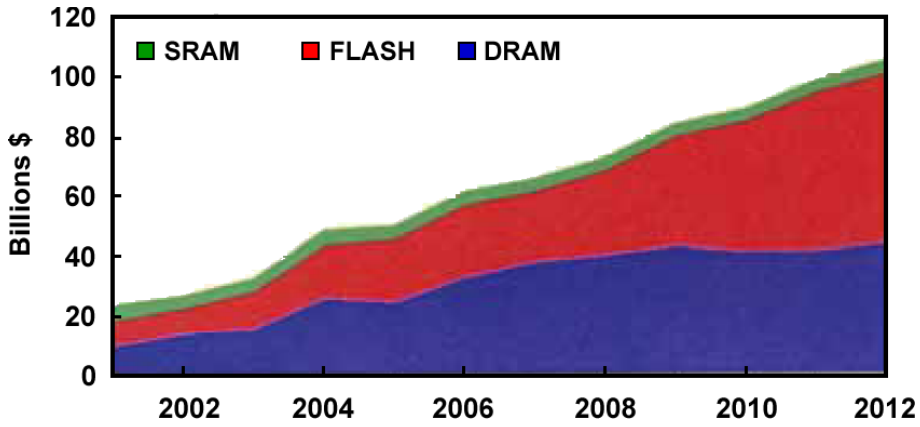

Figure II.2: *CMOS memory market evolution.*

II.1 Applications

Flash memories have two major applications: one is the *code application*, i.e. the possibility of nonvolatile memory integration in logic systems to allow software updates, store identification codes, or reconfigure the system on the field.

In this sense, Flash devices are widely used in several fields. In the computer environment they allow to store and update the operating system in PC BIOS and Hard-Disk Drives (HDDs), in almost all peripherals like printers and DVD-readers, and in most add-on boards like video and sound cards. On computer network equipments, they allow to quickly upgrade the software in modems, interface cards and network routers. In the automotive electronic field they are used in vital functions such as Engine Control Units (ECUs) and Global Positioning Systems (GPS). Finally, cellular phones are a key Flash marked driver in years, with their demand for an always increasing amount of reliable and low-power memory devices.

The second application, called *data application*, is to create storing elements like memory boards or solid-state hard disks, made by Flash memory arrays, which are configured to create large size memories.

In this field, besides the well established USB storage devices that have now reached the incredible size of 64GB, the growth and advancements in Flash technology have enabled a significant opportunity for Solid-State Drives (SSDs) to make tangible inroads into markets currently dominated by HDDs. For example, Toshiba recently announced (Dec. 2008) the first 512GB SSD, which uses 43nm Flash technology to fit into a standard 2.5-inch drive casing and is expected to undergo mass production in mid 2009.

II.2 Fundamentals of NVMs

To have a memory cell that can commute from one state to the other, and which can store the information independently of external conditions, the storing element needs to be a device whose conductivity can be changed in a non-destructive way. One solution is to have a transistor with the threshold voltage that can change repetitively from a high to a low state, corresponding to the two states of the memory cell, i.e. the binary values "1" and "0" of the stored bit.

The threshold voltage V_T of a MOS transistor can be written as [3]:

$$V_T = K - \frac{\overline{Q}}{C_{OX}} \tag{II.1}$$

where K is a constant that depends from gate oxide thickness, doping, gate and substrate materials, $\overline{Q}$ is the charge weighted with respect to its position in the gate oxide, and C_{OX} is the gate oxide capacitance. It is evident from above equation that the threshold voltage of the MOS transistor can be altered by changing the amount of charge present in the insulator. Two are the main solutions adopted.

The most common way is to store the charge in a conductive layer between the gate and the channel that is completely surrounded by insulator. Since this layer acts as a completely electrically isolated gate, this type of device is commonly referred to as a Floating Gate (FG) device [4]-[6]. These devices still represent the mainstream technology and will be discussed more in detail in the next paragraph.

An alternative solution is to store the charge in discrete trapping centers of an appropriate insulating layer. The most commonly used material is nitride [7]-[10], with new materials being currently investigated [11]-[13]. These devices are called charge-trapping (CT) devices and have been introduced almost simultaneously to the FG transistor [7]. Today they are considered one of the most promising alternative to FG devices for future technology nodes [9], [14]. They will be treated more in detail in Paragraph II.4.2.

II.3 The Floating Gate Device

Most of the modern non-volatile memory devices are based on the Floating Gate (FG) transistor, whose structure is depicted in Fig. II.3. The upper gate is the control gate (CG) and the lower one, a conductive layer completely surrounded by dielectric, is the FG.

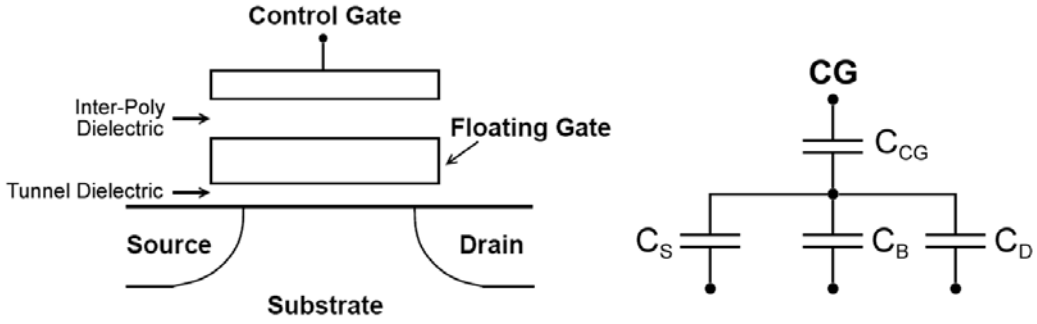

Figure II.3: *a) Schematic cross-section of a Floating Gate transistor; b) electrical model of a floating gate device.*

The FG is electrically isolated from source, drain and bulk regions through a high-quality thermal oxide called *tunnel oxide*, whereas CG and FG are separated by the Inter-Poly Dielectric (IPD), often called *control oxide*. The basic concepts and functionality of a FG device can be easily understood by determining the relationship between the FG potential, that physically controls the channel conductivity, and the control gate potential, controlled by external circuitry [15]. This can be done using the simple electrical model (known also as capacitive coupling coefficient model) shown in Fig. II.3(b). C_{CG}, C_S, C_D, and C_B are the capacitances between FG and CG, Source (S), Drain (D) and Body (B), respectively. If no charge is stored in the FG, i.e. Q = 0:

$$Q = 0 = C_{FG}\left(V_{FG} - V_{CG}\right) + C_S\left(V_{FG} - V_S\right) + C_D\left(V_{FG} - V_D\right) + C_B\left(V_{FG} - V_B\right) \qquad \text{(II.2)}$$

where V_{FG} is the potential on the floating gate, V_{CG} is the potential on the control gate, V_S,V_D, V_B are potentials on S, D and B, respectively. If we name $C_T = C_{FG}+C_D+C_S+C_B$ the total capacitance of the FG, and we define $\alpha_J = C_J/C_T$ as the coupling coefficient relative to the electrode J, where J can be one among CG, D, S, and B, the FG potential is given by

$$V_{FG} = \alpha_{CG}V_{CG} + \alpha_D V_D + \alpha_S V_S + \alpha_B V_B \qquad \text{(II.3)}$$

It is interesting to note that the Floating Gate voltage does not depend only on the control gate voltage, but also on source, drain and bulk potentials. Moreover, if source and body are both grounded Eq. (II.3) can be rearranged and reduces to

$$V_{FG} = \alpha_{CG}\left(V_{CG} + f \cdot V_{DS}\right), \qquad \text{where} \qquad f = \frac{\alpha_D}{\alpha_{CG}} = \frac{C_D}{C_{CG}} \qquad \text{(II.4)}$$

Device equations for the FG MOS transistor can be obtained from the conventional MOS transistor equations by replacing MOS gate voltage, V_{GS}, with V_{FG}, and transforming the device parameters, such as threshold voltage, V_T, and conductivity factor, β, to values measured with respect to the control gate: $V_T^{CG} = \alpha_{CG} \cdot V_T^{FG}$ and $\beta^{CG} = \beta^{FG}/\alpha_{CG}$ [15].

Thus, the current-voltage (I-V) equations of FG MOS transistor in both Triode Region (TR) (II.5) and in the Saturation Region (SR) (II.6), can be easily derived from that of a conventional MOS transistor [16].

TR:
$$I_{DS} = \beta^{CG}\left[\left(V_{CG} - V_T^{CG}\right)V_{DS} - \left(f - \frac{1}{2\alpha_{CG}}\right)V_{DS}^2\right]$$
$$\left|V_{DS}\right| < \alpha_G\left|V_{CG} + fV_{DS} - V_T^{CG}\right| \qquad \text{(II.5)}$$

SR:
$$I_{DS} = \frac{\beta^{CG}}{2}\alpha_{CG}\left(V_{CG} + fV_{DS} - V_T^{CG}\right)^2$$
$$\left|V_{DS}\right| \geq \alpha_{CG}\left|V_{CG} + fV_{DS} - V_T^{CG}\right| \qquad \text{(II.6)}$$

These equations underline some major differences between I-V characteristics of FG and conventional MOS transistor [16] that are mainly due to the capacitive coupling between drain and floating gate. Two of them are worth to be mentioned.

First, the FG MOS transistor can conduct current even when $|V_{CG}\text{-}V_S| < |V_T|$, because the channel can be turned on by the drain voltage through the $f \cdot V_{DS}$ term in (II.4).

Second, in the saturation region I_{DS} continues to rise as the drain voltage increases: no saturation will occur. Thus, I_{DS} results to be dependent on V_{DS}, on the contrary to what happens in conventional MOS transistors.

If some charge is stored in the FG, i.e. $Q \neq 0$, all the hypotheses made above hold true, although the following modifications in the V_{FG} and V_T calculation need to be included

$$V_{FG} = \alpha_{CG}V_{CG} + \alpha_D V_{DS} + \frac{Q}{C_T} \qquad \text{(II.7)}$$

$$V_T^{CG} = V_{T0}^{CG} - \frac{Q}{C_{CG}} \qquad \text{(II.8)}$$

V_{T0} is the threshold voltage when Q=0. From (II.8), we can see that the role of injected charge is to shift V_T, i.e. the I-V curves of the cell, by the amount $-Q/C_{CG}$. If the reading biases are fixed, the presence of charge greatly impacts the current level of the cell state. At this regard, Fig. II.4 shows two I-V curves: curve A represents the "1" state, whereas curve B the "0" state of the same cell.

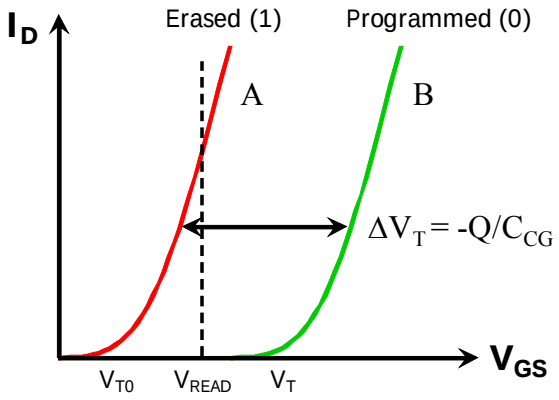

Figure II.4: *I-V curves of a FG devices with (B) and without (A) charge stored in the FG.*

II.3.1 Charge Injection Mechanisms

Among the several mechanisms that can be used to transfer charge from and into the FG, two are the ones currently used in industry-standard FG devices: Channel Hot Electron (CHE) injection [17], which is used for the program operation, and Fowler-Nordheim (FN) tunneling [18], used for both program and erase operations.

Channel Hot Electron Injection. The physical mechanism of CHE injection is relatively simple to understand qualitatively. An electron traveling from the source to the drain gains energy from the lateral electric field and looses energy to the lattice vibrations (acoustic and optical phonons). At low fields, this is a dynamic equilibrium condition, which holds until the field strength reaches approximately 100 KV/cm [19]. For fields exceeding this value, electrons are no longer in equilibrium with the lattice, and their energy begins to increase. Electrons are "heated" by the high lateral electric field and a small fraction of them gains enough energy to surmount the oxide barrier. For an electron to overcome this potential barrier, it must have kinetic energy higher than the potential barrier and velocity directed towards the FG [20].

A simple description of the CHE injection mechanisms can be given following the *lucky electron model* [17]. This model is based on the probability for an electron to be lucky enough to travel ballistically for a distance several times the mean free path without scattering, eventually acquiring enough energy to cross the potential barrier if a collision pushes it towards the Si/SiO_2 interface. Consequently, the probability of injection is the lumped probability of the following statistically independent events, see Fig. II.5: 1) the carrier is "lucky" enough to acquire the energy to overcome the oxide barrier and to retain this energy after the collision that redirects it towards the interface ($P_{\phi b}$); 2) carrier follow a collision-free path from the redirection point to the interface (P_{ED}); 3) the carrier can surmount the repulsive

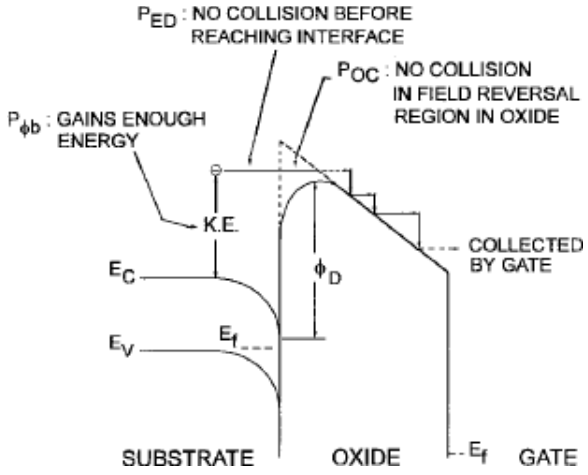

Figure II.5: *Schematic band diagram illustrating the "lucky electron model" [17].*

oxide field at the injection point without suffering an energy-robbing collision in the oxide (P_{OC}). CHE injection is a fast program mechanism, but has a poor efficiency and leads to a large power consumption.

Fowler-Nordheim Tunneling. The solution of the Schrödinger equation shows that a tunneling through a potential barrier is possible even for classically forbidden barriers [22]. The probability of electron-tunneling depends on the distribution of occupied states in the injecting material and on the shape, height and width of the potential barrier. Using a free-electron gas to model the electron population in the injecting material and the Wentzel-Kramers-Brillouin (WKB) approximation to calculate the tunneling probability [23], the well known expression for FN current density can be obtained [24]

$$J_{FN} = A_{FN} F_{OX}^2 \cdot \exp\left(-\frac{B_{FN}}{F_{OX}}\right) \tag{II.9}$$

where F_{OX} is the oxide field, T is the temperature and A_{FN} and B_{FN} are coefficients that can be calculated from MOS physical constants [24].

$$A_{FN} = \frac{q^3 m_{Si}}{16\pi^2 \hbar m_{OX} \Phi_0} \qquad B_{FN} = \frac{4\sqrt{2m_{OX}\Phi_0^3}}{3q\hbar} \tag{II.10}$$

q is the electron charge, and $\hbar$ is the reduced Plank's constant; Φ_0 is the oxide barrier height; m_{Si} and m_{OX} are the electron effective masses in the silicon and in the oxide layers, respectively.

The optimum thickness (about 7-8 nm) for FG memory using tunneling phenomenon is chosen trading off between performances constraints (programming speed, power consumption, …) which would require thin oxides, and reliability concerns, which would require thick oxides. Moreover, tunneling currents are also important for device-reliability at low fields. In the case of bad-quality tunnel oxides, or when thin oxides are stressed many times at high voltages, Trap Assisted-Tunneling (TAT) through bulk traps either present or generated in the oxide can strongly enhance the tunnel current. Therefore, oxide defects must be avoided to control program/erase characteristics and to have good reliability. This aspect is particularly important also for BGE devices and will be further discussed in Chapter III.

Although being slow with respect to CHE injection, FN tunneling is highly efficient: the only current flow produced into the cell is the tunneling current across the oxide, and all the transferred charge contributes to the modification of the FG charge state.

II.3.2 Flash Architectures

Two types of array organization are currently used for Flash devices: the NOR and the NAND architectures (see Fig. II.6). Both were invented by Fusjio Masuoka[2] (Toshiba) and respond to the different needs of code and data applications.

NOR architecture [1] is the most commonly used for code storage, since it guarantees the fast random access times required by this kind of applications. The NOR array is sketched in Fig. II.6 (a): all gates of the cells belonging to the same row are connected to the same wordline, whereas all drains of the cells in a column are connected to the same bitline; the sources of all the cells in the same sector are connected to a common source line. Programming is performed by means of CHE injection, whereas erasing is performed through FN tunneling. The thickness of the tunnel oxide is currently in the range of 7-8nm [26] and cannot be scaled below this value due to retention reliability issues (see paragraph II.3.3).

Beside common NOR, parallel architecture, Flash memories can also be organized in NAND arrays [27], by connecting 32 or even 64 cells in series between a bitline and the sourceline, as depicted in Fig. II.6(b). The main advantage of this solution is the achievement

[2] Masuoka worked on his ideas without permission from Toshiba. By 1980 he had already applied for the basic patents on NOR-type Flash memory, but it was not able to produce the first device until four years later, after a promotion (before, he was not senior enough to be allowed to go to the factory without permission and order them to make him some devices). Masuoka presented his Flash memory at the annual International Electron Device Meeting, in 1984. Intel immediately put more than 300 engineers to work full time on developing Flash memory, whereas Toshiba assigned only five engineers to help Masuoka on a part-time basis. It was not long before Intel completely dominated the market.
In 1987, again without permission, Masuoka made the first batches of his new NAND-type Flash memory. This time he was senior enough to be able to devote resources to the project to ensure that Toshiba gained an insurmountable lead in both patents and production technology. However, shortly after the first of the new memories went on the market in 1990, Toshiba began pressuring him to accept a "promotion" that, at the ripe old age of 47, would have put him in a job with no subordinates. In 1994 he quit Toshiba to become a professor at Tohoku University. *Adapted from [25].*

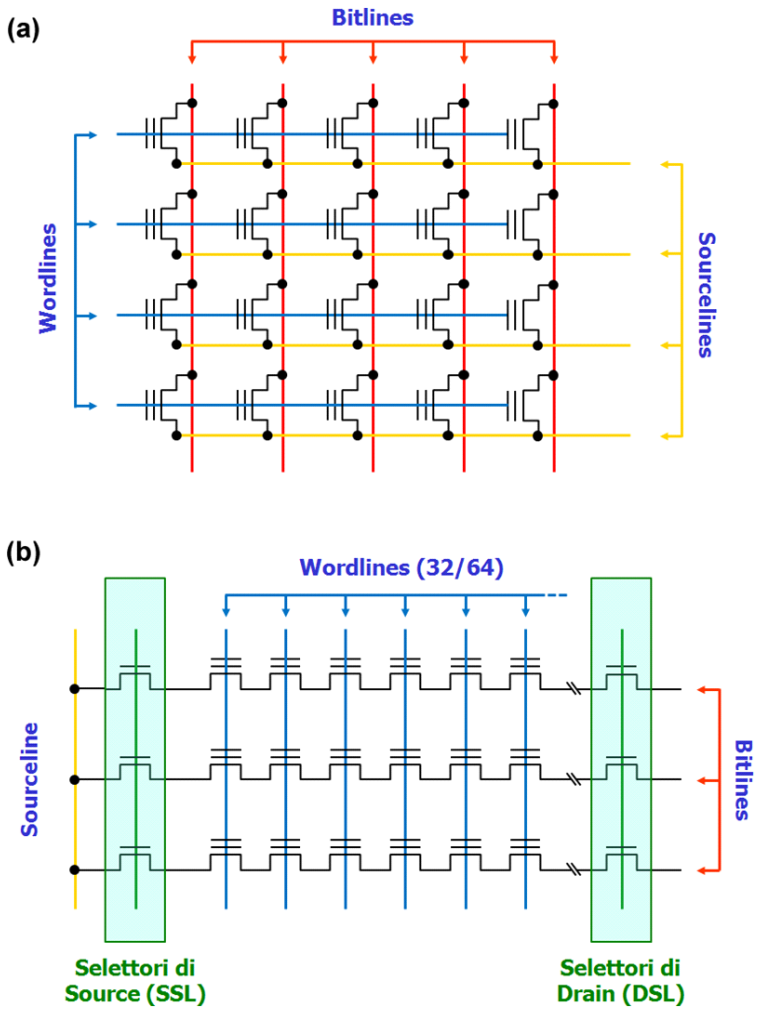

Figure II.6: *Schematic illustration of (a) NOR and (b) NAND architectures.*

of an higher integration density because of a decreased number of contacts, from one contact to the bitline for every cell as in NOR array, to one contact for 32/64 cells. On the other hand, the read operation is much slower with respect to NOR, as the whole string must be accessed. For these reasons, NAND Flash memories are limited to mass storage applications only[3]. Both program and erase operation are performed through FN tunneling. Although the scaling of the tunnel oxide for NAND Flash faces the same challenges as that for NOR Flash, the use of Error Code Correction (ECC) in NAND, together with the less stringent reliability requirements due to the target applications, allows thinner oxides to be used (6-7nm [26]).

A fundamental difference between NOR and NAND architectures is the threshold voltage (V_T) range adopted for programmed and erased distributions, see Fig. II.7. In NOR architectures the V_T working range is higher than 1.5V and lower than 8V. Reading is performed by biasing word and bitline of the selected cell to positive voltages, with the other

[3] This was absolutely true in the past. Recently, new products that aim to combine the high density of NAND with the fast read of NOR have been proposed, e.g. Samsung's OneNAND™.

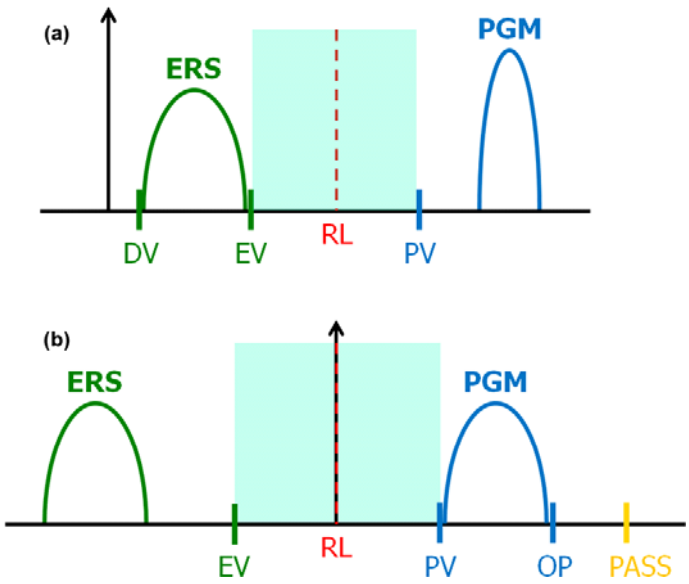

Figure II.7: *(a) NOR and (b) NAND program windows. DV=Depletion Verify level; EV=Erase Verify level; RL=Read Level; PV=Program Verify level; OP=Over-Programming level; PASS=PASS voltage.*

wordlines and bitlines grounded. Thus, the unselected cells sharing the bitline of the cell being read must have a $V_T > 0$ to have a correct sensing of the cell's state. In the NAND architecture, the V_T operating range is from -5V to +3V. In this case read is performed by grounding the wordline of the selected cell and positively biasing the bitline and the wordline of the unselected cells belonging to the same string. The unselected cells act as pass transistors and the current flowing in the string is determined only by the state of the selected cell. Therefore, the unselected cells must have a threshold voltage lower than the pass voltage to read properly the state of the selected device.

II.3.3 Scaling Limits of FG Devices

To achieve higher densities and reduce costs, a reduction of device dimensions is required. Unfortunately, the scaling of conventional FG devices below the 45nm technology node appears critical, due to severe technological limits and reliability issues.

Tunnel Oxide Scaling. Program and erase operations of Flash memories require high voltages, as they are based on physical mechanisms whose major parameters do not scale (3.1eV energy barrier for CHE injection and at least 10MV/cm for FN data alteration in 0.1s).

The simpler way to decrease writing voltages would be to reduce the thickness of the tunnel oxide. However, in order to guarantee at least 10 years of data retention, the tunnel oxide limit is fixed to 6nm by the direct tunneling mechanism, a value that needs to be

increased to 7-8nm to account for the anomalous conduction through the tunnel oxide due to trap-assisted-tunneling caused by oxide aging [28]. This current is commonly referred to as Stress-Induced Leakage Current (SILC) and leads to a dramatic increase of the charge loss from the FG during retention. This phenomenon has been deeply investigated in the past [28]-[30] and depends mainly on stress magnitude and oxide thickness. It is generally assessed that the tunnel oxide cannot be scaled below 7-8nm to control SILC effects. A reduction of this limit requires the replacement of conventional SiO_2 tunnel dielectric with BGE barriers, as will be discussed in the next paragraph.

Gate Coupling Ratio (GCR). One of the major stumbling blocks for future FG devices is the expected reduction of the coupling coefficient between CG and FG (α_G in Eq. II.4). This parameter determines the fraction of the voltage applied at the CG that is transferred to the FG: it must be higher than 0.6 in order to obtain reasonable writing voltages.

In current technologies, a high GCR is normally achieved by wrapping the control gate around the sidewalls of the floating gate[4]. However, below 40nm the spacing between FG may become too narrow for the IPD and CG to wrap around the FG, leading to a strong decrease of the GCR. To overcome this problem, innovative solutions are required, e.g. the replacement of ONO inter-poly dielectric with high-k materials [26].

Capacitance Coupling Between adjacent Cells (NAND). The continuous shrinking of NAND memories lead to an enhanced capacitance coupling between adjacent cells, which significantly increases the lateral fringing field disturbing NAND operations [31]. 3D TCAD simulations show that the threshold voltage shift (ΔV_T) induced by adjacent cells on the same bitline (BL) and wordline (WL) increases exponentially with technology scaling, tripling moving from the 57nm to the 32nm technology node [32]. To reduce the FG interference it is necessary to thin the FG and to adopt low-k dielectric materials. An alternative is represented by CT devices that does not suffer from this problem.

[4] Without this expedient, the required value of 0.6 could not be reached. The GCR is defined as the ratio between the CG to FG capacitance, C_{CG}, and the total capacitance of the FG cell, C_{TOT} (see paragraph II.3). Neglecting the contribution of C_S and C_D, it can be expressed as

$$\alpha_G = \frac{1}{1+\frac{t_{IPD}}{t_{TUN}}}$$

where t_{IPD} and t_{TUN} are the thicknesses of IPD and tunnel layers, respectively. As typical values of t_{IPD} and t_{TUN} are in the range of 7-8nm and 13-15nm, respectively [26], this leads to α_G ranging from 0.32 to 0.41 for current technologies (a little bit pessimistic as C_S and C_D have been neglected).

II.4 Innovative NVM Devices

Extensive research efforts are today devoted to the investigation of alternative non-volatile memory devices able to overcome the FG limits discussed in the previous paragraph, thus enabling the scaling of Flash technology under the 45-nm technology node. The solutions that have been proposed in the last years can be classified into *evolutionary* and *revolutionary* concepts. The first class of devices adopts new materials and/or new storage mechanisms to push the traditional FG structure beyond its limits. BGE and CT devices belong to this category. On the contrary, revolutionary devices are based on completely new concepts in which the FG transistor is replaced by new storage elements obeying to a completely different physic to store the information. Some examples are Ferroelectric RAMs (FeRAMs), Magnetoresistive RAMs (MRAMs) and Phase Change Memories (PCMs).

A brief description of some of the most promising revolutionary devices will be given in the following. BGE and CT devices will be discussed more in detail in the next pharagraphs.

Ferroelectric RAMs. The operation of these devices is based on the ferroelectric effect [33], i.e. the ability of a material to retain an electrical polarization in the absence of an applied electric field. This stable polarization results from the alignment of internal dipoles within the Perovskite crystal units in the ferroelectric material.

There are different approaches to implement these ferroelectric materials into a memory cell. A common one is to use a ferroelectric capacitor addressed by a transistor [34]. In order to sense the polarization state of the ferroelectric film in a capacitor, a switching of the polarization is required. Depending on the polarization state of the ferroelectric film, a small or a large amount of charge will flow in the circuit. However, a write back cycle is necessary in this cell concept in order to restore the initial read information. There are also concepts to realize non-destructive read-out cells, like the Ferroelectric Field Effect Transistor (FeFET) cell [35]. This cell type is a MOS transistor, where the gate oxide is replaced by a ferroelectric film thus giving the possibility of nonvolatile data storage in an extremely compact cell. The read-out is performed by sensing the source drain current, which is dependent on the threshold voltage given by the polarization state of the ferroelectric gate layer.

The main advantages of FeRAM devices are the fast read/write operation (around 100ns), high endurance (up to 10^{12} cycles) and low voltage write. However ferroelectric materials show some specific issues, like the decrease in switching polarity with cycling and the imprint phenomenon: the capacitor tends to prefer the state in which it has been for extended periods of time. The integration of ferroelectric materials in the CMOS process is also difficult.

Magnetoresistive RAMs. In magnetic random access memories the data are stored as

magnetization directions and the read-out is done by a resistance measurement. Among the different MRAM concepts, the cell based on Tunneling Magneto-Resistance (TMR) is the most promising [36]. These devices use a Magnetic Tunnel Junction (MTJ) obtained with two stacked ferromagnetic layers. One of them has a fixed magnetization direction and is used as a reference layer, while the second can be switched between two states, thus representing the storage layer. The two ferromagnetic layers are separated by a very thin dielectric layer. The current flow through the structure is limited by this thin dielectric barrier and depends on the magnetic state of the two ferromagnetic films.

The MRAM cells are written by current pulses through the bit-line and the word-line in order to generate a magnetic field, which is larger enough to switch one of the ferromagnetic layers, but not large enough to switch the second magnetic reference layer. The magnetic domain switching and, thus, the writing time of the cells is in the range of a few ns and there is no endurance limit expected (in contrast with ferroelectric materials), since the electronic spin flipping mechanism does not degrade with continuous cycling.

Also MRAMs show some specific issues. The integration of the magnetic stack is a very critical point, due to the exact thickness definition of the insulating tunnel layer, which is required to avoid shorting between the tunneling electrodes. Another problem is the one associated with scaling. As the critical switching field (and thus the current density of the word line) becomes even larger upon shrinking the width of the cell size, a method for scaling the world line current has to be engineered in order to avoid electromigration and cross talk of neighboring cells.

Phase Change Memories. Phase Change Memory is another interesting competitor in the class of nonvolatile memory contenders. This concept is based on the reversible phase change between the amorphous and the crystalline phase of a chalcogenide glass [37]. These two physical states of matter differ in their resistivity, as the conduction is generally much better in a crystalline chalcogenide than in an amorphous one due to the reduced scattering of charge carriers in films with atomic long range order. The transition from the crystalline to the amorphous state is performed by applying a very short electrical pulse to a resistive heater in contact with the phase change material thereby melting it (typical melting temperature is about 600°C) and thereafter rapidly cooling it to freeze the amorphous phase. In order to write the crystalline state into the cell a lower but a little bit longer pulse is applied, thus heating the material over the critical crystallization temperature (about 300°C in materials typically used) and leaving it in the low resistivity polycrystalline phase. The difference in resistivity of the two phases is about 2-3 orders of magnitude, which is considerably higher than in MRAM. The writing mechanism also allows the realization of Multi-Level Cell (MLC) data storage by programming the cell to intermediate resistance levels, thus yielding a lower fraction of the crystalline phase. Reading is accomplished by measuring resistance changes in the cell.

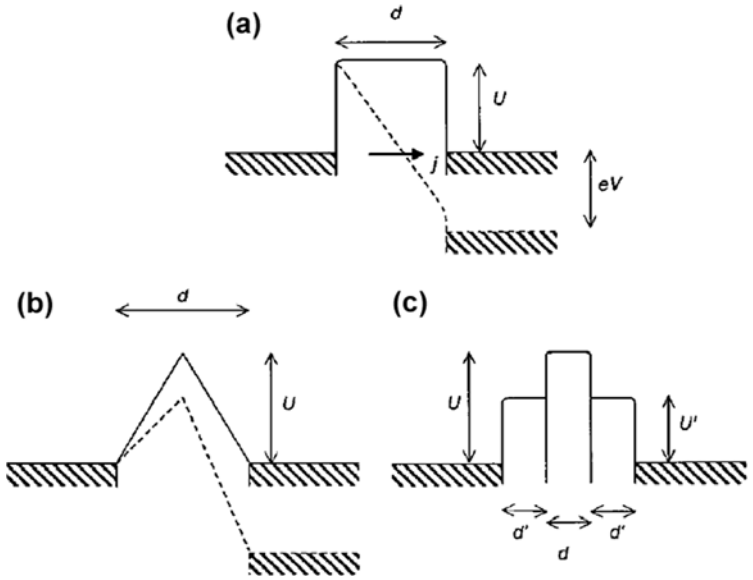

Figure II.8: *Schematic representation of the conduction band diagrams of various tunnel barriers: (a) conventional SiO_2 barrier; (b) ideal crested barrier; (c) real crested barrier [38].*

II.4.1 Band-Gap Engineered Devices

As discussed in paragraph II.3.3 the SiO_2 tunnel dielectric cannot be scaled below 6-7nm in order to guarantee data retention. This is clearly in contrast with the need to reduce program and erase voltages, which demands for a thinner tunnel oxide.

In order to scale the writing voltages while satisfying retention requirements, the thickness of the tunnel barrier should be reduced only during program and erase operations. In other words, the tunnel barrier should be **thick** in retention conditions and *thin* during program and erase operations, i.e. it should show an increased sensitivity on the applied voltage. This principle is at the basis of band-gap engineered barriers.

In BGE barriers, the conventional SiO_2 tunnel oxide is replaced with a dielectric stack incorporating high-k materials, i.e. materials having a relative dielectric constant higher than the one of SiO_2 (3.9). There are two possibilities: crested barriers [38], [39], that relies on the concept of barrier height modulation (ϕ-engineering), and VARIable Oxide Thickness (VARIOT) barriers [40], which are instead based on the modulation of the electric field (κ-engineering).

The basic operating principle of crested barriers is sketched in Fig. II.8. For the case of a conventional rectangular energy barrier as the one depicted in Fig. II.8(a), the tunneling current varies slowly by increasing the applied voltage. This slow dependence of the barrier transparency on the electric field is due to the fact that the part of the barrier close to the electron source is only weakly affected by the applied voltage, see dashed lines in Fig. II.8(a).

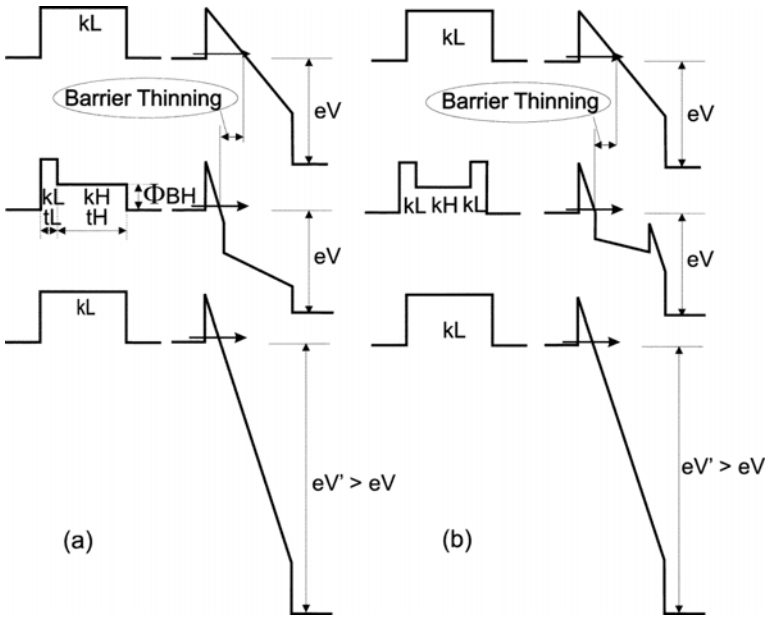

Figure II.9: *Band diagrams illustrating the VARIOT concept for the case of (a) two-layer (asymmetric) barrier and (b) three-layer (symmetric) barrier [40].*

To increase the sensitivity of the energy barrier on the applied voltage, a "crested" barrier like the one depicted in Fig. II.8(b), with the potential barrier height peaking in the middle and gradually decreasing toward the electrodes, should be used [39]. The current through such a barrier changes much faster with the applied voltage. The reason is that the highest part of the barrier is pulled down by the electric field very quickly, see dashed lines in Fig. II.8(b).

The implementation of crested barriers is straightforward in composite semiconductors [38], but not in the case of a conventional CMOS process, as the ideal shape in Fig. II.8(b) cannot be achieved. A more practical solution is the one represented in Fig. II.8(c), where the ideal barrier shown in Fig. II.8(b) is approximated by a staircase energy barrier obtained with a dielectric stack. This kind of barrier can be obtained using high-k dielectrics, which typically have conduction band (CB) offsets with respect to silicon lower than the 3.1eV of SiO_2 [41]. For example, the one depicted in Fig. II.8(c) is a tri-layered structure having a SiO_2 layer sandwiched between two high-k layers. It is usually called symmetric crested barrier (high-k/SiO_2/high-k) in contrast to the asymmetric one (high-k/SiO_2), in which only one high-k layer is used [39].

One of the main drawbacks of crested barriers is the requirement to have a high-k layer in direct contact with the silicon substrate. This leads to two major technological issues: i) an abrupt Si/high-k interface cannot be obtained, as a thin SiO_X layer is usually formed ii) the Si/high-k interface is usually of poor quality. These issues are not present in VARIOT engineered barriers, which can be considered an evolution of the crested barrier concept.

The operating principle of VARIOT BGE barriers is schematically represented in Fig. II.9. It relies on the fact that when a voltage is applied to the stack, the electric field

redistributes in its layers being always higher in the one with the lower-k value (as a consequence of Gauss' law). In addition, the high-k dielectric has typically a lower barrier height with respect to the SiO_2 low-k dielectric: this means that for voltages higher than a *critical* value electron tunneling (direct or FN) takes place only through the thin low-k layer, see Fig. II.9. These effects determine a strong variation of the stack barrier with the applied bias and the stack can be regarded as a VARIable Oxide Thickness dielectric [40].

The VARIOT concept is very attractive for NVM applications since it theoretically allows to achieve a higher programming speed (or a lower voltage programming at the same speed), while guaranteeing the same retention performance of a conventional tunnel oxide [42]. Alternatively, BGE barriers can be optimized to improve Flash memory retention with respect to conventional SiO_2, guaranteeing the same Program/Erase (P/E) performances.

Unfortunately, high-k materials feature very high bulk defect and interface state densities, and their theoretical advantages in retention improvement have to be weighed against their degraded parasitic trap-assisted leakage currents that lead to the undesired reduction of the threshold voltage. This aspect is very important in the assessment of real chances of high-k stacks to replace conventional SiO_2 tunnel layers, as will be shown in Chapter III.

II.4.2 Charge Trapping Devices

The possibility to use trap-rich dielectric layers as the storage medium in charge-trapping non-volatile memory devices was recognized early [7]. Today these devices are considered one of the most promising alternatives to the FG technology, especially for NAND applications.

CT devices have two main advantages with respect to the conventional FG transistor. First, they have an inherent immunity to retention loss mechanisms related to point defects in the tunnel oxide (e.g. SILC). When a trap or percolation path is formed in the tunnel oxide, only the charge trapped in the portion of the nitride that is directly above it will discharge toward the substrate. As a consequence, the thickness of the tunnel oxide can be reduced well below the 6-7nm limit of conventional FG devices.

The second advantage is related to the interferences between adjacent cells. As discussed in Paragraph II.3.3, the continuous scaling of NAND memories lead to an enhanced capacitance coupling between adjacent cells: the threshold voltage of a cell depends also on the state of the adjacent cells. CT devices does not suffer from this issue: as the charge is localized, cross talk with neighboring cells is strongly reduced.

Among the different charge trapping devices that have been proposed in the literature [7]-[14] two are particularly promising: the NROM [10] and the TANOS [14] devices.

NROM Device. The NROM cell is an n-channel MOSFET device where the gate dielectric is

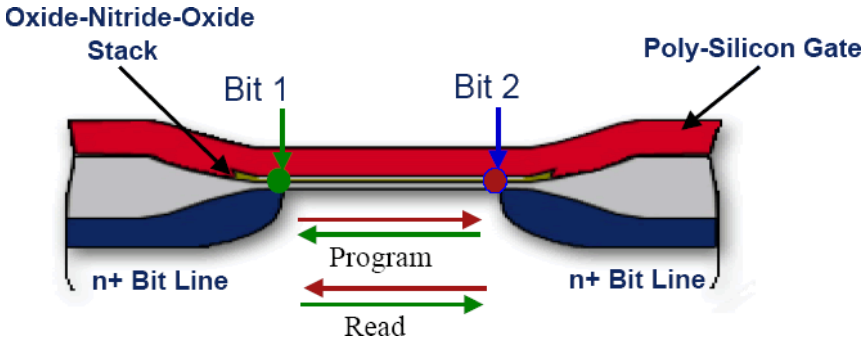

Figure II.10: *Cross section of a NROM device.*

replaced with a nitride layer sandwiched between two silicon dioxide layers (forming the so-called ONO – Oxide-Nitride-Oxide – layered structure), as illustrated in Fig. II.10. The program operation is performed by CHE injection, whereas Hot Hole Injection (HHI) has been identified as the dominant erase mechanism [43]. The injected charge is stored in the nitride layer above the channel, in a very narrow region above the metallurgic junctions. The localization of the trapped charge, together with the reverse read methodology, allows storing two bits per cell [44]. Optimized technology, accurate and fast program algorithm, no single bit failures and window sensing with moving reference as an error detection and correction scheme, allow 4-bit product [45].

As already mentioned, reliability concerns that are crucial for the floating gate technology are of little importance for NROM memory devices. Electrons (or holes) are trapped individually and the loss of one charge through a defect does not affect other trapped charges that are not directly over the point defect (in contrast to FG technology). The specific reliability issues of NROM technology are associated with the localized trapping and with the presence of both electron and hole distributions in the nitride layer. In particular, lateral charge redistribution in the nitride may occur over time due to thermally activated charge migration between traps [46-49], thus leading to a threshold voltage shift over time.

To improve cell reliability and guarantee a correct cell operation also in scaled NROM devices, a deep knowledge of the charge distribution after both program and erase operations is needed. Different techniques developed during the Ph.D. research activity to profile electron and hole distributions in NROM devices will be discussed in Chapter V and VI.

TANOS Device. The TANOS cell structure was proposed by Samsung in 2003 [14]. It is an evolution of the previously reported SANOS device [50], which in turn is an evolution of the SONOS device [8], [51].

Although allowing faster program and erase operations with respect to SONOS due to the introduction of the Al_2O_3 high-k dielectric as blocking oxide, original SANOS devices were characterized by a reduced program window due to the fast saturation of the threshold voltage

during erase [50]. This saturation is attributed to electrons that are injected from the gate, thus compensating the effect of holes injected from the substrate [14].

To overcome the above mentioned issue, the TANOS structure was proposed, in which the poly-silicon gate used in the SANOS devices was replaced by a TaN metal gate [14]. This allowed to reduce the erase saturation level of approximately 2V. Such a significant improvement of the erase characteristics can be explained by the higher energy barrier that the electrons injected from the gate see at the TaN/Al_2O_3 interface.

Today, TANOS devices are considered one of the most promising candidate for scaled NAND Flash technologies. However, a deep understanding of the physical mechanisms involved in its operation is still lacking. In this sense, accurate models of program, erase and retention are highly needed, as they can providing important insights on trap characteristics, on the evolution of trapped charge during program and erase and on their dependencies on the TANOS stack composition. Such information are vital to for the optimization of TANOS memory cells. An accurate model of TANOS program transients will be presented in Chapter VI.

CHAPTER III

Feasibility of BGE Barriers

"No exponential is forever… but we can delay forever"

Gordon Moore

In this chapter the feasibility of BGE barriers as Flash memory tunnel dielectrics in future Flash technologies is investigated. First of all, we describe the statistical Monte Carlo (MC) simulator developed to reproduce leakage currents flowing through symmetric and asymmetric high-k based layered structures. The simulator is validated against experimental data measured on large area capacitors having both symmetric (SiO_2/high-k/SiO_2) and asymmetric (SiO_2/high-k) gate stacks. Then, its statistical capabilities are exploited to assess the real benefits of the introduction of high-k stacks as Flash memory tunnel dielectrics in future Flash technologies. We simulate 1Mb array of 65nm NAND Flash cells in retention condition to extract the statistical distribution of the leakage current. We show that the strong reliability improvements predicted by BGE theory disappear when trap-assisted contributions are included. These results warn regarding the possibility to replace conventional tunnel oxides with the high-k stacks.

HIGH-K materials have been successfully introduced in the CMOS process to replace the conventional SiO_2 as gate dielectric in 45nm logic transistors [52] and beyond. They are also considered one of the most promising solution that has been proposed to overcome the limits of FG devices. Recently, they have been proposed both as storage layers in charge-trapping memories [11]-[13], and to replace conventional tunnel oxide to realize smarter band-gap engineered barriers [38]-[40].

In the framework of charge trapping memories, HfO_2, HfSiON Al_2O_3 and more exotic materials like Y_2O_3 [13] have been investigated as viable options to replace Si_3N_4 [10], showing better charge-storage capabilities due to the very high trap density and the high permittivity.

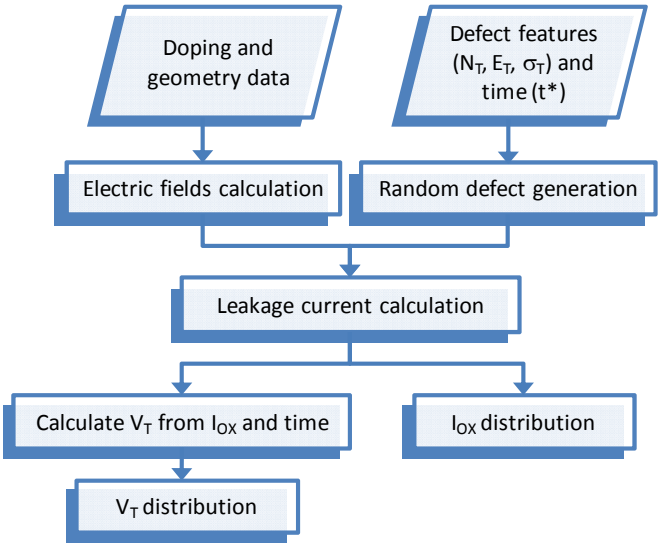

Figure III.1: *Flowchart of the statistical MC simulator.*

On the other hand, band-gap engineered potential barriers have been introduced more than ten years ago [38], [39]. As discussed in Paragraph II.4.1, they theoretically allow improving Flash memory retention compared to conventional tunnel oxide, while guaranteeing the same Program/Erase performances.

Unfortunately, high-κ materials feature very high bulk defect and interface state densities, and their theoretical advantages in retention improvement have to be weighed against their degraded parasitic trap-assisted leakage currents that lead to the undesired reduction of the threshold voltage. This aspect is very important in the assessment of real chances of high-κ stacks to replace conventional SiO_2 tunnel layers. In addition, when considering ultra-dense, ultra-scaled Flash memory devices, the feasibility of BGE barriers has to be investigated also at the array level, accounting for the statistical effects related to the random position of defects in the gate stack. In this context, simulations are indispensable, as extensive reliability measurements on large Flash memory arrays are very time-consuming and costly [53].

III.1 Simulation Model

To simulate the leakage currents flowing through generic SiO_2/high-κ dielectric stacks, we extended the statistical Monte Carlo (MC) simulator used in [54]. Compared to the simpler SiO_2 case, several issues related to the presence of a composite dielectric stack have to be accounted for, such as the calculation of the electric field in different materials and of the tunneling probability between traps located in different dielectrics.

The block diagram of the MC simulator we developed is shown in Fig. III.1. It consists of three main parts devoted to electric fields calculation, random defect generation and leakage

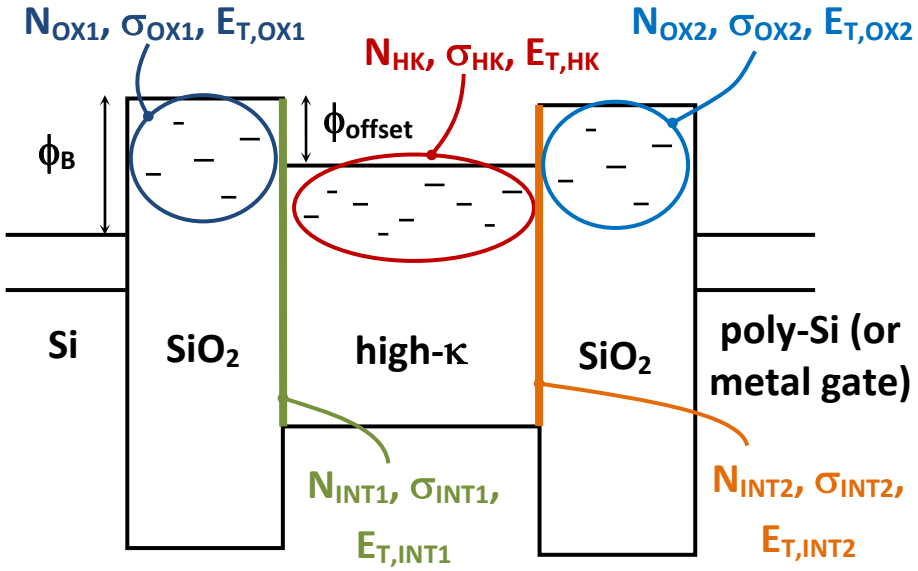

Figure III.2: *Schematic representation of a symmetric SiO_2/high-k/SiO_2 dielectric stack showing some key parameters used in simulations.*

current density (J_{LEAK}) calculation, respectively. When the model is used to simulate leakage currents flowing through the dielectric stacks of Flash memories, the device threshold voltage (V_T) changes induced by J_{LEAK} are calculated through an ad-hoc procedure, see Fig. III.1.

The electric fields across the dielectrics composing the stack are calculated with a model that takes into account poly depletion and charge quantization effects [55]. The presence of charge trapped in high-κ and oxide bulks and at the SiO_2/high-κ interface is also included.

Defects are randomly generated within oxide and high-κ dielectric and at their interfaces, according to device geometry and defect statistics. Different densities (N_T), cross sections (σ_T), and energy distributions (E_T) have been considered for different defects, as sketched in Fig. III.2 for a symmetric barrier. For each of these features, we can select fixed as well as random values generated considering various statistical distributions. Noticeably, we can consider whichever physical defect distribution we want, not only the uniform one, which is usually considered for simplicity. This is very important to simulate the leakage current through thin gate stacks (EOT≈1-2nm) [56], as experimental evidences of non-uniform spatial distribution of defects have been reported already [57].

Once traps have been generated, J_{LEAK} is calculated by summing both direct (J_{TUNN}) and defect-assisted (J_{PTAT}) tunneling contributions. The conduction model we adopted is described more in detail in the next section.

III.2 PTAT Conduction Model

The new leakage current model adopted by our statistical MC-like simulator is based on the multi-phonon trap-assisted tunneling model formerly presented in [58]-[60]. Differently from common TAT models presented in the literature (for the simpler SiO_2 case) [61]-[64], our model accounts for the coupling to oxide and high-k phonons, which result in a series of virtual states in the energy band-gap broadening the trap energy level, E_T.

The leakage current flowing through each randomly generated trap is automatically calculated checking if multi-trap conductive paths are formed within the stack. Under steady-state conditions and without charge buildup in any trap, the rate R of electrons passing through the n-trap conductive path is calculated as [60]

$$R = \frac{1}{\max_j\left(\tau_{c,j} + \tau_{e,j}\right)}, \tag{III.1}$$

where $\tau_{c,j}$ and $\tau_{e,j}$ are the time constants of the capture and the emission of electrons by and from the j^{th} trap, respectively. As can be seen from Eq. III.1, the rate the charge passes through faster traps is limited by the rate of the slowest trap [60].

Capture and emission time constants are calculated summing over the discrete energies $E_{j,n} = E_{C,j} + n \cdot \hbar\omega_0$ all the single phonon time constant contributions, $\tau_{c,j,n}$ and $\tau_{e,j,n}$, where $E_{C,j}$ is the conduction band edge for j=0 or j=trap number+1, or the j^{th} trap energy level $E_{T,j}$ for 0<j<trap number+1.

$$\tau_{c,j}^{-1} = \sum_n \tau_{c,j,n} = \sum_n N\left(E_{j-1,n}\right) \cdot f\left(E_{j-1,n}\right) \cdot P_T\left(E_{C,j} - E_{j-1,n}, F_{j-1,j}, D_{j-1,j}\right) \cdot Ca_{j,n} \tag{III.2}$$

$$\tau_{e,j}^{-1} = \sum_n \tau_{e,j,n} = \sum_n N\left(E_{j+1,n}\right) \cdot P_T\left(E_{C,j} - E_{j,n}, F_{j,j+1}, D_{j,j+1}\right) \cdot Em_{j,n} \tag{III.3}$$

$N(E_j)$ is the density of states at the cathode (j=0), in the trap states (0<j<trap number+1), and at the anode (j=trap number+1); f is the Maxwell-Boltzmann occupation probability; $Ca_{j,n}$ and $Em_{j,n}$ are the trap capture and the emission rates; P_T is the tunnel probability, where $D_{j,i}$ is the distance between the j^{th} and the i^{th} trap, and $F_{j,i}$ is the equivalent oxide field given by $F_{j,i}=F_{OX}(z_j-z_i)/D_{j,i}$ (z is the trap coordinate with respect to the axis perpendicular to the Si/SiO_2 interface, and F_{OX} is the oxide field) [60]. Particular care has been devoted to the calculation of the tunneling probability between traps located in different dielectrics.

Table III.1: *Main characteristics of the dielectric stacks used. The "Type" column refers to the type of dielectric stack: Symmetric (S) or Asymmetric (A).*

Sample	High-k	Type	EOT [nm]	t_{OX} [nm]	t_{HK} [nm]	t_{OX} [nm]
A	HfO_2	A	4.2	3.4	4	-
B	HfO_2	A	4.5	3.6	4.7	-
C	HfSiON	A	4.5	3.1	5.8	-
D	HfSiON	A	5.2	3.4	7.3	-
E	Al_2O_3	A	3.9	2.1	4.75	-
F	Al_2O_3	A	5.8	2.1	9.5	-
G	Al_2O_3	A	3.6	2	3.5	-
H	HfO_2	S	4.2	2	1.5	2

Table III.2: *SiO_2 and high-k parameters used in leakage simulations.*

Material	κ	CBO [eV]	m^*/m_0	E_G [eV]
SiO_2	3.9	3.1	0.5	8.9
HfO_2	19-21	1.8	0.2	5.8
HfSiON	15.6	1.8	0.2	5.7
Al_2O_3	9-10	2.8	0.1	8.8

Table III.3: *Trap parameters used in simulations. The indexes OX, INT and HK refer to traps in oxide, at the interface(s) and in the high-k, respectively.*

Sample	N_{OX} [cm^{-3}]	$E_{T,OX}$ [eV]	$\sigma_{T,OX}$ [cm^2]	N_{INT} [cm^{-3}]	$E_{T,INT}$ [eV]	$\sigma_{T,INT}$ [cm^2]	N_{HK} [cm^{-3}]	$E_{T,HK}$ [eV]	$\sigma_{T,HK}$ [cm^2]
A, B	$2\div5\cdot10^{17}$	1.5÷2.2	10^{-14}	10^{12}	1.3÷2.0	10^{-13}	-	-	-
C, D	$2.5\cdot10^{17}$	1.5÷1.9	10^{-14}	10^{12}	1.3÷2.0	10^{-13}	-	-	-
E, F, G	$1\div5\cdot10^{17}$	1.0÷1.6	10^{-14}	10^{12}	1.0÷1.4	10^{-13}	$0.8\div5\cdot10^{19}$	1.1÷1.7	10^{-14}
H	$5\cdot10^{17}$	1.5÷2.1	10^{-14}	$2\cdot10^{12}$	1.7÷2.2	$2\cdot10^{-14}$	$3\cdot10^{19}$	1.3÷1.6	10^{-14}

III.3 Simulation Results

To verify the accuracy of the simulator we developed, we started reproducing experimental leakage currents measured on relatively large area ($10^{-4}cm^2$) capacitors.

Film thicknesses of the dielectric stacks we considered are reported in Table III.1. We used both p-MOS & n-MOS capacitors (p & n-type Si (100) substrate). SiO_2 was thermally grown on top of Si, followed by ALD deposited Al_2O_3, HfO_2 or $HfSiON_x$. In $HfSiON_x$, nitrogen was introduced using a NH_3 anneal, the Hf/(Hf+Si) ratio is ~80% and the N concentration is around 7%. PVD Al gate (samples E and F), TiN gate (samples G and H) or TaN gate (samples A to D) was then deposited to complete the gate stack.

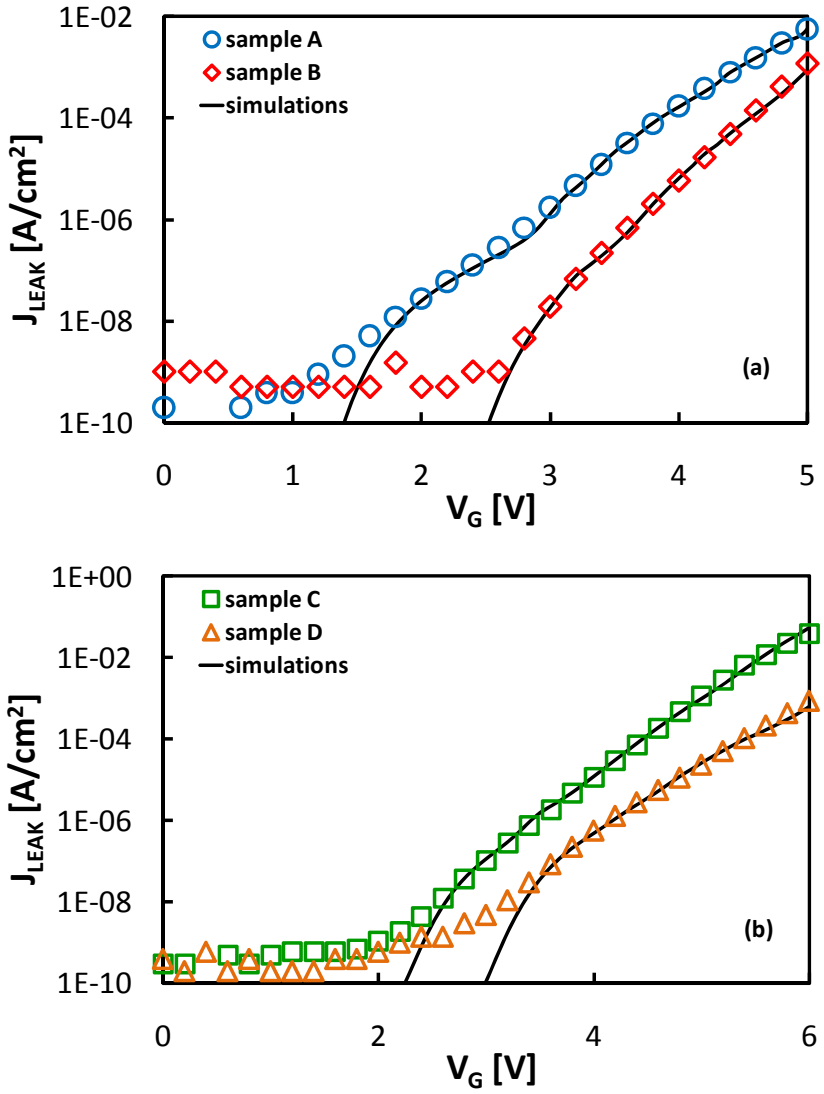

Figure III.3: *Comparison between experimental and simulated leakage currents across (a) SiO_2/HfO_2 and (b) SiO_2/HfSiON large area capacitors.*

The material parameters used in simulations are reported in Table III.2. Here, κ is the relative dielectric constant, CBO the conduction band offset with respect to Silicon, m* and m_0 are the effective mass and the free mass of electrons, respectively, and E_G is the energy gap. It is worth to notice that there is still some controversy in the literature about some of these values, that have not been univocally determined yet. Nevertheless, values used in simulations agree with the most reasonable ones [41], [66]-[69]. Work Functions (WFs) of 4.6eV, 4.5eV and 4.15eV were considered for TaN, TiN and Al gates, respectively [65].

The trap parameters extracted from simulations are reported in Table III.3 and agree well with other values reported in the literature [70], [71].

Figg. III.3(a)-(b) show J_{LEAK}-V_G curves simulated and measured on pMOS capacitors with different SiO_2/HfO_2 and SiO_2/HfSiON dielectric stacks, respectively. As shown, the agreement between measurements and simulations is excellent. Noticeably, simulations are

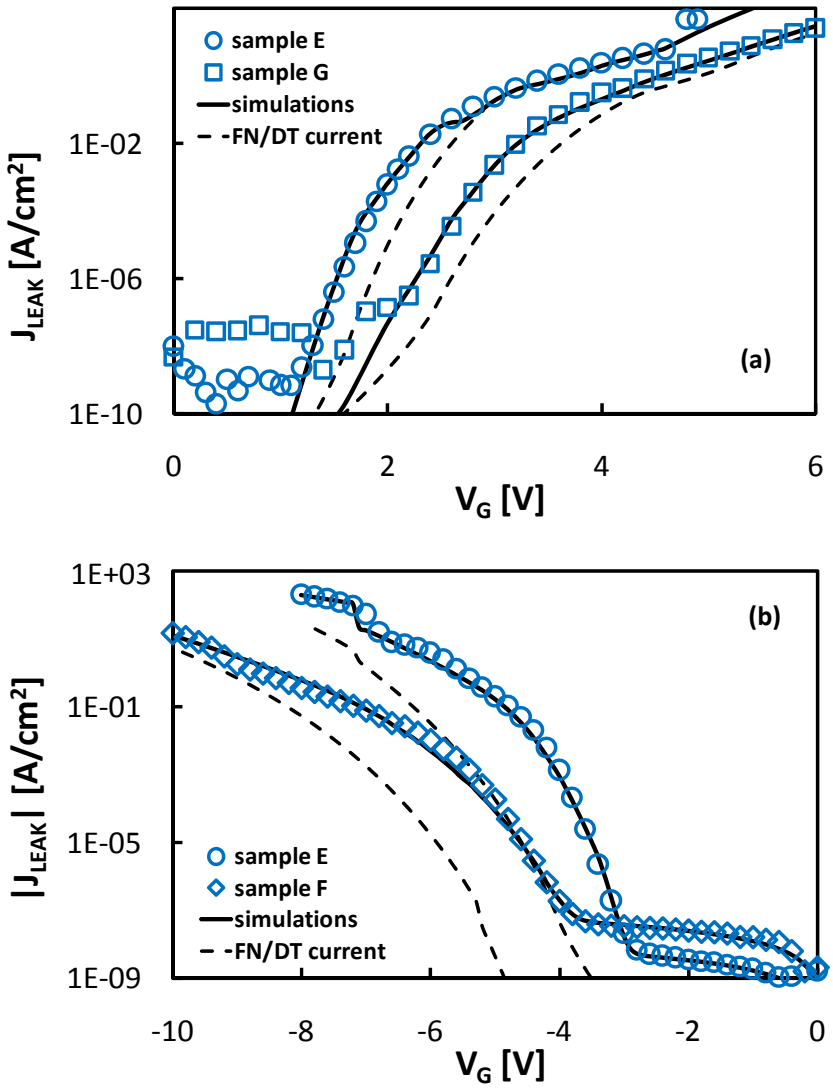

Figure III.4: *Comparison between experimental and simulated leakage currents across SiO_2/Al_2O_3 stacks under (a) substrate and (b) gate injection conditions.*

performed considering only oxide and SiO_2/high-κ interface defect contributions (see values in Table III.3), whereas HfO_2 and HfSiON bulk traps were found to affect negligibly J_{LEAK} curves when electrons are injected from the substrate, at least for the dielectric thicknesses considered here.

We obtained excellent results also for SiO_2/Al_2O_3 samples, as shown in Fig. III.4. Again, we found that Al_2O_3 bulk traps slightly affect the leakage current when electrons are injected from the substrate, whereas they dominate the trap-assisted conduction when electron injection is from the gate.

Interface traps are found to be critical especially at low gate voltages, i.e. in retention conditions, whereas bulk oxide traps dominate J_{LEAK} conduction mechanism at relatively higher voltages, i.e. in the program/erase regime. This is clearly shown in Fig. III.5 for sample

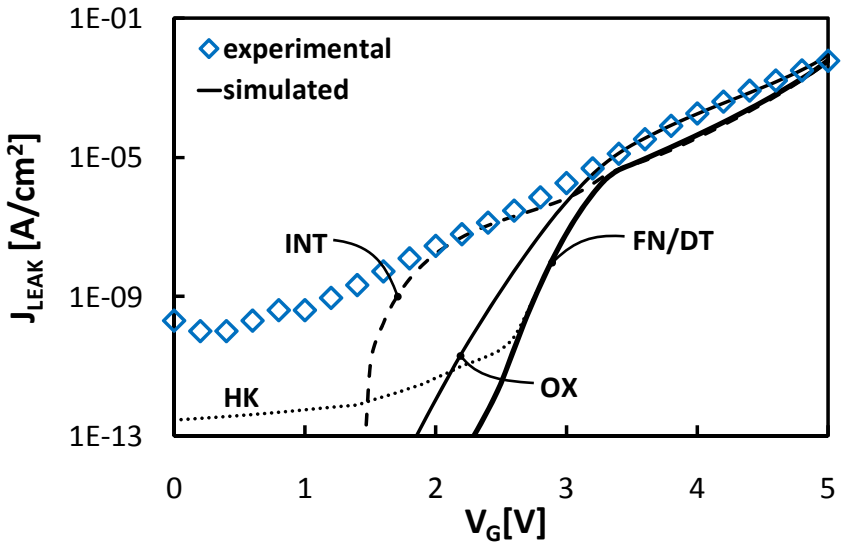

Figure III.5: *J_{LEAK}.vs V_G curve simulated considering interface defects (INT), oxide defects (OX), and high-k defects (HK), in addition to the FN/DT current contribution.*

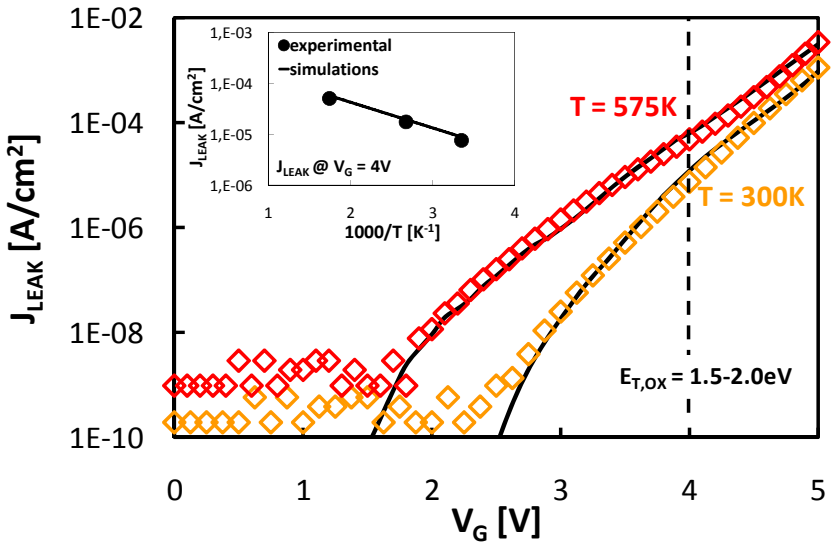

Figure III.6: *Leakage current simulated and measured on sample B at different temperatures. Symbols: experiments; lines: simulations. The inset shows the Arrhenius plot at V_G=4V.*

B, where roles played by interface, oxide and high-κ defects are highlighted. As expected, the contribution of bulk HfO_2 traps is negligible (i.e. below the noise level) when electrons are injected from the substrate, even if a very high trap density (10^{20} cm^{-3}) is considered.

As shown in Fig. III.6 for sample B, simulations reproduce accurately J_{LEAK} temperature dependence without adjusting additional fitting parameters, proving that the model catches correctly the J_{LEAK} conduction mechanism physics and is a valuable tool to investigate defect properties of high-κ composite dielectrics. Interestingly, the activation energy (E_A≈0.1 eV) estimated from the J_{LEAK} Arrhenius-like plot shown in the inset of Fig. III.6 does not match the defect energy (E_T≈1.5-2eV), demonstrating that tunneling via the traps with taking into

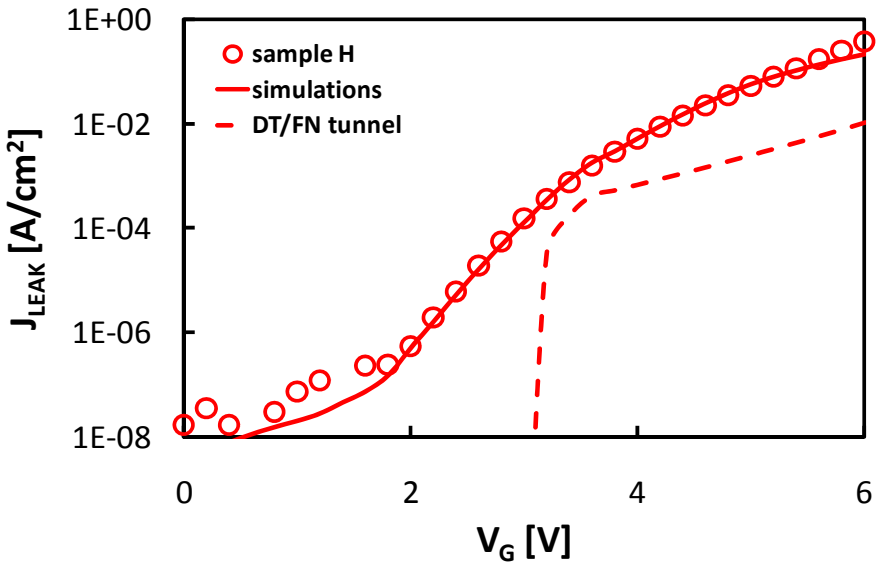

Figure III.7: *Comparison between experimental and simulated leakage currents across symmetric $SiO_2/HfO_2/SiO_2$ structure.*

account thermal population of the trap vibrational levels is the dominant conduction mechanism over thermally activated ones, like thermoionic emission.

Our model is also able to describe the leakage currents measured on large area capacitors with symmetric high-k dielectric stacks, as shown in Fig. III.7 for sample H. Noticeably, the parameters used are almost the same as for samples A and B, see Table III.3 confirming that the model correctly describes the physics involved in the transport through the dielectric stacks.

III.4 J_{LEAK} Statistical Simulations

Once the simulation capability of the model has been established, we used it to test the real feasibility of high-κ dielectric stacks to replace conventional tunnel oxides in future Flash memory generations. It is known that introducing high-κ composite dielectrics theoretically allows reducing by orders of magnitude the leakage current under retention conditions while maintaining the same P/E current performances. This is clearly demonstrated in Fig. III.8. Here, V_{RET} and V_{PROG} correspond to typical retention and program voltages (referred to the FG), whereas the thick horizontal line represents the maximum leakage current that can be sustained to guarantee 10 years data retention [40], [42]. As can be seen, a 4.5nm thick SiO_2 tunnel layer do not allows to satisfy retention requirements, whereas the introduction of the Hf-based dielectric stack with the same EOT reduces by orders of magnitude the retention current.

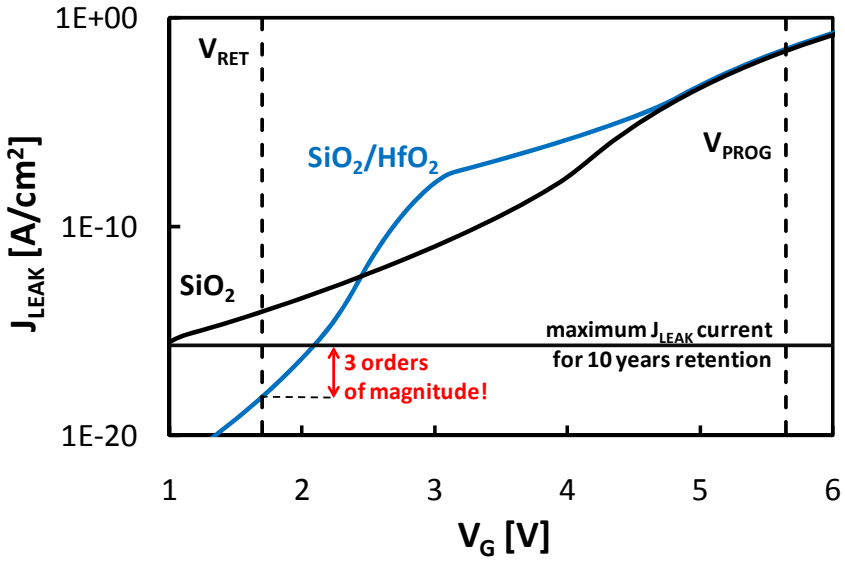

Figure III.8: *Comparison between ideal leakage currents simulated considering SiO_2/HfO_2 (sample B in Table III.1) stack and pure SiO_2 capacitors with the same EOT=4.5nm.*

Results shown in Fig. III.8 are true when the ideal J_{LEAK}-V_G curves, obtained by neglecting defect contributions in the relatively large area capacitors, are considered. Unfortunately, when typical oxide and interface defects are included, the leakage current flowing through high-k stacks increases by several orders of magnitude, see for example Figs. III.4 and III.7. Looking at these data, a question arises: are BGE barriers still able to guarantee retention when defects are considered?

To answer the above question, we performed statistical J_{LEAK} simulations considering a 1Mb NAND array of 65nm Flash memory cells having the same HfO_2 and HfSiON tunnel stacks as the ones of samples B and C in Table III.1 (EOT=4.5nm).

First, we calculated the retention voltage, i.e. the equivalent FG voltage forcing across the tunnel stack the same field calculated in a Flash memory cell in retention conditions, as [15]

$$V_{RET} = \alpha_G \left(V_{T,UV} - V_{T,PROG/ERS} \right) \tag{IV.1}$$

where typical values have been considered for the UV threshold voltage, $V_{T,UV} \approx 0$ and control gate coupling ratio, $\alpha_G \approx 0.6$. We evaluated V_{RET} considering both erased (V_{RET}=1.8V) and programmed (V_{RET}=-1.5V) memory cells, assuming that erase and program threshold voltages are $V_{T,ERS}$ = -3V and $V_{T,PROG}$ = 2.5V, respectively. To account for the worst case retention conditions, we considered V_{RET}=1.8V, which leads to highest fields within the tunnel high-κ stack. Therefore, we expect an higher leakage current compared to that obtained in the case of gate injection, which should be also investigated, given the asymmetrical barrier profile of Hf-based dielectric stacks considered in this case.

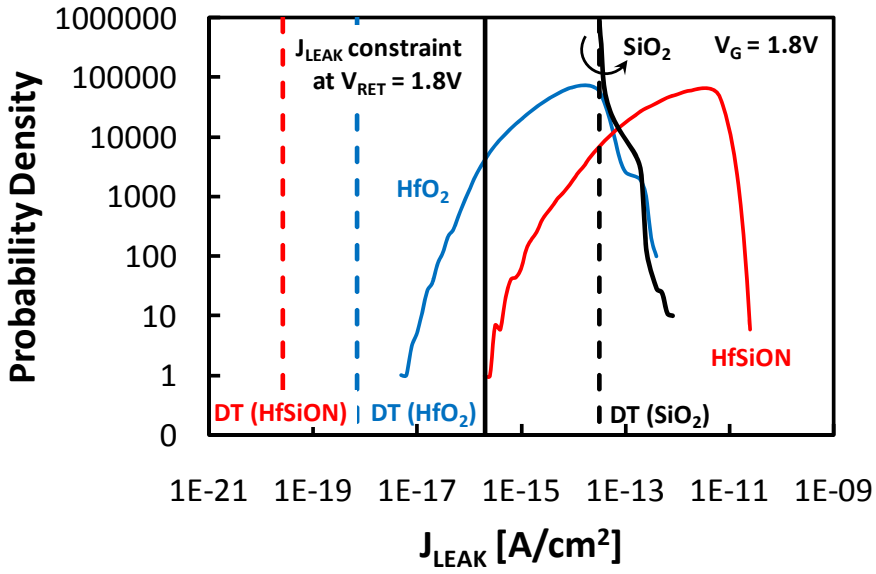

Figure III.9: *Statistical distributions of J_{LEAK} at V_{RET} for SiO_2/HfO_2 and SiO_2/HfSiON stacks and pure SiO_2 having the same EOT (4.5nm). Dashed lines depict DT currents.*

Second, we run statistical MC simulations assuming the same trap features extracted from capacitor J_{LEAK} simulations, see Table III.3. J_{LEAK} probability density curves simulated at V_{RET} are depicted in Fig. III.9 (solid lines) for both Hf-based high-κ dielectric stacks (samples B and C in Table III.3). Dashed lines depict the current simulated without defects, i.e. considering only the direct and Fowler-Nordheim tunneling contribution, whereas the thick line represents the maximum current which can be tolerated to satisfy Flash memory retention requirements [72].

As expected, when the defect contribution is not included (see dashed lines), leakage currents through HfSiON and HfO_2 stacks are ~4-5 orders of magnitude lower than that in SiO_2 according to the band-gap engineered dielectric stack concept [39]. In addition, leakage currents are significantly lower than the retention current limit, guaranteeing Flash memory retention. On the contrary, when trap-assisted contributions are included, significant reliability improvement associated with the Hf-based dielectric stack disappears. In fact, both HfSiON and HfO_2 stack leakage currents increase by orders of magnitude, and a significant part of their distributions is higher than the maximum retention current. Furthermore, the margin between the currents of the Hf-based and SiO_2 dielectrics disappears baffling in practice the VARIOT concept when implemented using the real Hf stacks. J_{LEAK} spreading is due to random locations and energies of the defects, which depends on trap densities and cell volume.

Simulation results shown in Fig. III.9 provide a warning regarding the possibility of replacing conventional tunnel oxides with the high-κ stacks. Of course, in order to gain a definitive understanding, more work needs to be done. Different t_{OX} and t_{HK} combinations

giving the same EOT have to be explored to find the optimum, and different structures (symmetrical) as well as high-κ materials should be considered. Nevertheless, it is worth noticing that statistical simulations play a crucial role for optimizing the high-κ dielectric stacks and assessing the possibility of the high-κ as tunnel dielectrics in future Flash memory generations.

III.5 Chapter Summary

In this Chapter, a statistical MC-like simulator developed to reproduce leakage currents flowing through a generic high-κ dielectric stack has been presented.

The simulation capability of the model has been assessed against experimental data measured on asymmetric and symmetric structures and a different temperatures. The excellent agreement between measurements and simulations confirms that the model correctly catches the J_{LEAK} conduction mechanism physics.

The model has been used to characterize defect properties of the high-κ dielectric stacks considered in this Chapter. These data have been used, together with the statistical capabilities of the model, to test the real feasibility of high-κ dielectric stacks as tunnel oxides in future Flash memory generations. It is found that the Hf-based asymmetric stacks considered do not allow to satisfy retention requirements. These results underline the need for an optimization of the engineered stack: different t_{OX} and t_{HK} combinations giving the same EOT have to be explored to find the optimum, and different structures (symmetrical) as well as high-κ materials should be considered. In this framework, statistical simulations play a crucial role.

CHAPTER IV

I_D-V_{GS} Based Tools to Profile Charge Distributions on NROM™ Memory Devices

This chapter presents two tools based on I_D-V_{GS} curves (i.e. subthreshold slope and I_D temperature effects) allowing to profile program charge distributions in NROM devices. Simple formulas to calculate length and density of the program charge distribution are derived and their accuracy is tested for cells programmed at different levels and under different bias conditions. Tools accuracy and sensitivity is investigated, and their limits when applied to erased NROM cells are discussed.

NROM cells are characterized by an inherent immunity to failure due to point defects in the gate dielectric. However, cycling-related concerns on scaling, endurance and retention are reported due to the presence of physically separated electron and hole distributions, and on the difficult control of their relative position and spread in the charge trapping material [73]-[75]. When NROM cells are erased, holes are trapped into the nitride and the final effect is to reduce the net stored charge and to change the shape of the final charge distribution. Thus, to improve cell reliability for guaranteeing a correct cell operation also in scaled memory devices, a deep knowledge of charge distribution after program and erase is needed.

Some techniques have been proposed in the literature to profile the charge distribution in localized charge trapping memory devices [76]-[82]. The Charge Pumping (CP) technique is used in [76]-[78] to characterize the lateral charge distribution in local charge-trapping memory devices, but this technique requires a non-trivial dedicated experimental setup and its accuracy has not been proved in this specific case. Other authors employ I_D-V_{GS} measurements (that do not require a dedicated experimental setup) to profile charge distribution, providing some estimates of the length (L_{CN}) and density (ρ_{CN}) of charge distribution [79]-[82]. Typically, threshold voltage and subthreshold slope (SS) are used as monitors to evaluate charge distribution effects on I_D-V_{GS} curves [79]-[82]. Occasionally, gate-induced drain leakage current measurements are employed along with two-dimensional device simulations to improve the estimate accuracy, especially when erased cells are considered [82]. Despite few differences, the above referenced I_D-V_{GS} based techniques allow investigating charge distribution features, although simple formulas to derive charge distribution length (L_{CN}) and density (ρ_{CN}) have not been reported so far in the literature.

This Chapter presents two tools to profile charge distribution in NROM devices based on I_D-V_{GS} sensitivity on local charge storage, which will provide compact formulas allowing estimating length and density of charge distributions in programmed memory cells. Unfortunately, the proposed methods do not allow to derive similar information on hole distributions in erased cells. To this purpose, a different technique will be presented Chapter V.

IV.1 Devices and Experiments

Samples used in this work are NROM cells manufactured in 0.5-μm technology. Their effective channel length is 0.32μm. Top and bottom oxides are 8 nm thick, whereas the interleaving nitride layer is 6.5 nm thick. Devices have an intrinsic threshold voltage of about 2V. NROM cells are read in the reverse direction, i.e. by interchanging the roles of source and drain electrodes with respect to programming conditions [79]. The reverse-mode threshold voltage, V_{TR}, is defined as the gate voltage at which the drain current I_D reaches the value of 1μA with V_{DS} = 0.1 V, whereas the corresponding subthreshold slope, denoted as SS_R, is defined as

$$SS_R = \frac{dV_{GS}}{dLog(I_{DR})} \tag{IV.1}$$

NROM cells used in this work have been programmed to three different levels (namely, B, C and D), whose V_{TR} is 0.5 V, 1.5 V, and 2.3 V higher than the threshold voltage of a fresh device (A). Further, some cells have been programmed to the same level (D) under different CHE program conditions (V_{DS} = 4.5V; V_{GS} varying from 8, to 9 and 10 V) and by using a

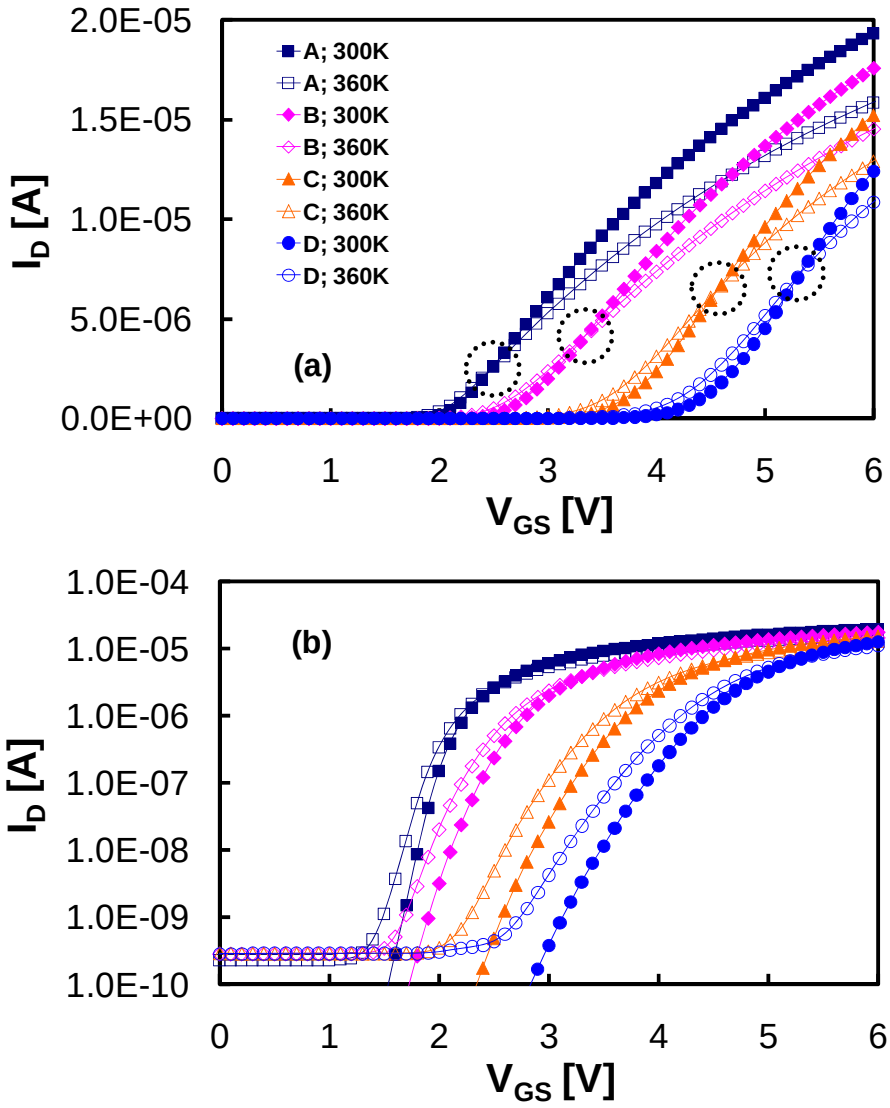

Figure IV.1: *(a) Linear and (b) logarithmic I_D-V_{GS} characteristics measured at 300K (filled symbols) and 360K (empty symbols) for a fresh cell (A), and for devices programmed to different levels (B, C and D, as in [80]).*

different program mechanism, that is the CHannel Initiated Secondary ELectrons (CHISEL) injection [83] (V_{GS} = 7V, V_{DS} = 1.5V, V_B = -5V). Erase is performed by using two different V_{GS} and V_{DS} combinations, called positive erase (V_{GS} = 0 V, V_{DS} = 5.5 V) and negative erase (V_{GS} = -8 V, V_{DS} = 3.3 V), respectively.

Typical I_D-V_{GS} curves measured on fresh and programmed NROM cells at both 300K and 360K are shown in Fig. IV.1(a)-(b) in linear and logarithmic scales, respectively. As expected from previous papers [79], [80], SS_R monotonically increases with the program level (from B to D), i.e. V_{TR}. Further, the temperature increase has three major effects on I_D-V_{GS} trans-characteristics: 1) threshold voltage reduction, and 2) subthreshold slope increase, which are related to the effective source-bulk barrier lowering, and 3) mobility reduction, which reduces the cell trans-conductance [3]. Interestingly, on increasing the program V_{TR}, i.e. the charge

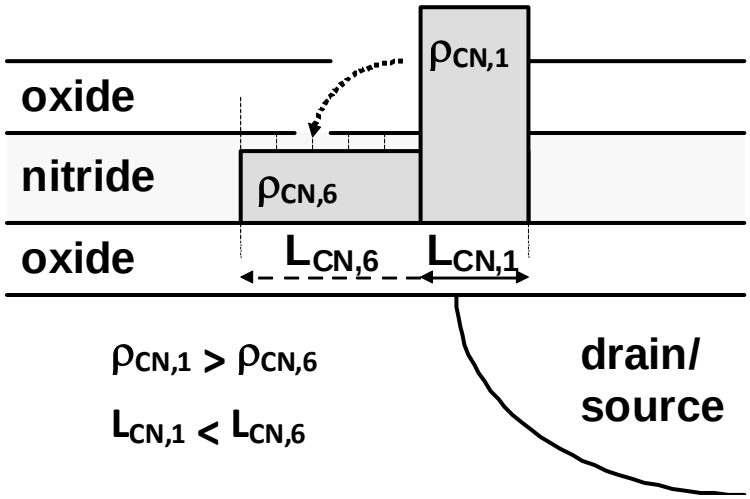

Figure IV.2: *Schematic cross section zoom of rectangular charge distributions defined into the nitride layer of the device. Only narrowest and largest L_{CN} cases are shown.*

stored in the nitride region above source/drain junctions, temperature effects on I_D-V_{GS} curves are amplified. In particular, on increasing the temperature, the $I_{D,300K}$ curve crosses the $I_{D,360K}$ one at higher I_D values (see dashed circle in Fig. IV.1(a)); the cell trans-conductance degrades less compared to non–programmed device; and the subthreshold slope increases much more than in a non-programmed device. The rise of I_D-V_{GS} temperature sensitivity when increasing V_{TR} is at the basis of the Temperature Monitor (TM), which will be described in Paragraph IV.3. SS_R can also be used for charge profiling (SSM), as described in Paragraph IV.4.

IV.2 Devices Simulations

To prove both charge profiling tools, TM and SSM, we employed two-dimensional device simulations carried out with the drift-diffusion code implemented by DESSIS (SYNOPSYS). We adopted the mobility model developed by Lombardi [84], simulating I_D-V_{GS} curves of a virgin NROM cell to calibrate it. Geometry and doping information fed to the device were obtained as output of a 2-D process simulation performed through TSUPREME-4.

To analyze the effects of length and density of the charge distribution on NROM I_D-V_{GS} curves, we considered four different program levels, whose reverse threshold voltages are 0.5V, 1V, 1.5V, and 2V higher than the threshold voltage of the virgin cell, $V_{TR,FRESH}$. For every program level, we defined six rectangular-shaped nitride charge distributions of different length, with ρ_{CN} adjusted to give the same V_{TR}, see schematic sketch in Fig. IV.2. The length of the portion of the charge distribution beyond the metallurgical junction was fixed to 10nm, since the charge stored more than 10 nm from the junction has a negligible effect on the electrical behaviour of the cell [80]. For every charge distribution considered, we

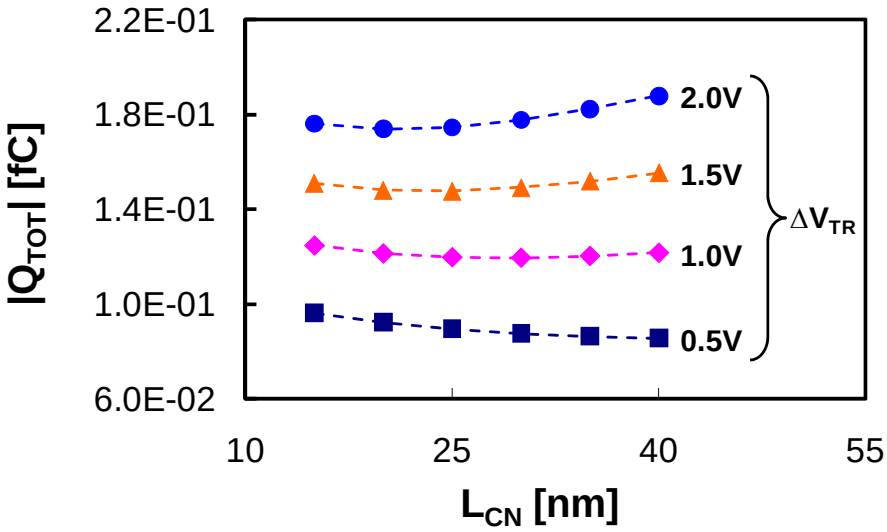

Figure IV.3: *Total charge as a function of L_{CN} for the considered rectangular charge distributions, see Table IV.1. The amount of charge needed to obtain the same V_{TR} is approximately constant with respect to L_{CN}.*

simulated the I_D-V_{GS} curve at both 300K and 360K.

Interestingly, to obtain the same V_{TR} shift, the same amount of negative charge (i.e. electrons for programmed cells) should be stored in the nitride regardless the length of the rectangular nitride charge region. This is clearly shown in Fig. IV.3, confirming that the total stored charge (Q_{TOT}) is not significantly sensitive to L_{CN} for the V_{TR} levels considered.

Thus, Q_{TOT} depends only on V_{TR} in the L_{CN} range considered. As expected from classical MOSFET equations [3], the Q_{TOT} dependence on the V_{TR} shift is linear, the V_{TR} shift being defined as $\Delta V_{TR}=V_{TR}-V_{TR,FRESH}$. Thus, the total charge can be easily estimated from ΔV_{TR} through (IV.2).

$$\left|Q_{TOT}\right| = k_1 \Delta V_{TR} + k_2 \qquad \text{(IV.2)}$$

k_1 and k_2 are constants and are determined once for a given technology using the simulation procedure discussed in this Section. The Q_{TOT} independence on L_{CN} can be explained by looking at the programmed NROM cell as the series of two devices [80]. The NROM device modeling the portion of the channel beneath the nitride charge region (whose V_{TR} is the same of the whole NROM cell one) has a very short channel length (L_{CN}), and therefore its behavior is strongly affected by Short-Channel Effects (SCEs) [80]. Thus, a higher nitride charge density is needed to obtain the same V_{TR} as L_{CN} reduces. The quasi-constant trend of Q_{TOT} when varying L_{CN} means that for the limited L_{CN} range considered, the V_T increase due to the $\rho_{CN}=Q_{TOT}/L_{CN}$ rise is compensated by an equal V_T reduction due to the SCE increase, almost proportional to $1/L_{CN}$.

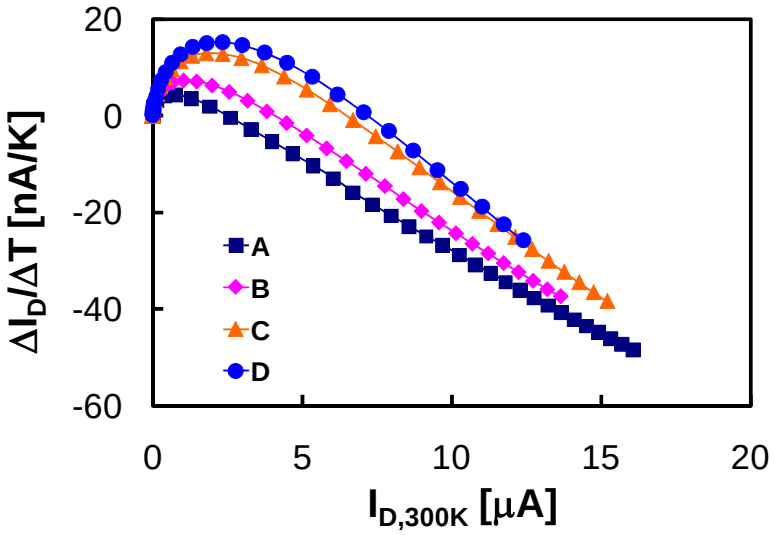

Figure IV.4: *Temperature Monitor curves corresponding to cells in condition A (initial) and programmed to B, C and D levels.*

IV.3 Temperature Monitor

The Temperature Monitor is a tool allowing profiling the local net charge distribution. It makes use of a graph, where the I_D difference due to temperature normalized to the temperature difference [$\Delta I_D/\Delta T=(I_{D,360K}-I_{D,300K})/(360-300)$] is plotted versus the drain current measured at room temperature ($I_{D,300K}$) [85]. Basically, this tool is based on the fact that the temperature effect on I_D-V_{GS} curves is very sensitive to the features, length and density, of the nitride charge distribution above source/drain junctions [86].

Figure IV.4 shows TM curves extracted from the I_D-V_{GS} characteristics depicted in Fig. IV.1. As shown, TM curves change according to V_{TR} level: on increasing V_{TR}, the TM maximum increases and moves to larger values, while the zero-cross point, that is defined as the I_D value where $I_{D,300K} = I_{D,360K}$, moves to larger drain current.

The TM graph responds to V_{TR}, SS_R and cell trans-conductance variations induced by temperature: TM peak (TM_{MAX}) and zero cross point increase with V_T and ΔSS (= SS_{360K}– SS_{300K}), while the slope of the TM curve after the peak reduces on increasing the cell trans-conductance. TM curves are shaped by two competing physical mechanisms, namely diffusion and drift, dominating I_D conduction mechanisms in sub-threshold and "above-threshold" regime, respectively. Thus, two regions can be identified on the TM graph. In the first one, which goes from zero to TM_{MAX}, TM increases since diffusion current rises with temperature and V_{TR} slightly reduces. In the second region beyond the TM peak, electron drift dominates the drain current conduction, which reduces with temperature because of the mobility reduction, thus causing the TM curve reduction.

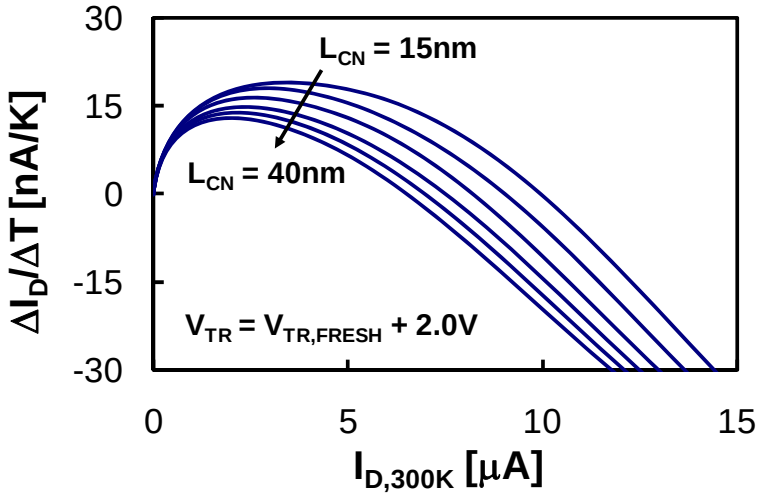

Figure IV.5: *TM curves corresponding to the rectangular charge distributions in Fig. IV.2, for $V_{TR} = V_{TR,FRESH} + 2V$.*

Table IV.1: *Experimental Values of SS_R and TM_{MAX}.*

Program Level	SS_R [mV/decade]	TM_{MAX} [nA/K]
A	137	4.34
B	226	7.25
C	308	12.9
D (V_{GS} = 8V)	365	15.7
D (V_{GS} = 9V)	376	18.0
D (V_{GS} = 10V)	396	21.1
D (CHISEL)	174	5.09

To investigate TM sensitivity to local charge profile, we considered different program biases and mechanisms. As shown in Table IV.1 summarizing experimental TM_{MAX} and SS_R values, cells with the same V_{TR} level show different TM_{MAX} values when programmed under different bias conditions, thus indicating the presence of different nitride charge distributions. Moreover, if CHISEL [83] program mechanism is adopted, the TM peak drops and the whole curve becomes similar to a fresh cell one [85].

To find a correlation between TM_{MAX} and L_{CN}, we simulated the drain current on NROM devices with the nitride charge distributions sketched in Fig. IV.2. We found that both TM peak and zero cross point reduce on increasing L_{CN}, the whole curve becoming very similar to a virgin one, see Fig. IV.5. This is expected, since punchthrough-like phenomena occurring in the channel portion below the storage charge region reduce when the charge distribution spreads over a wider region [80]. In fact, the distance between the channel current density

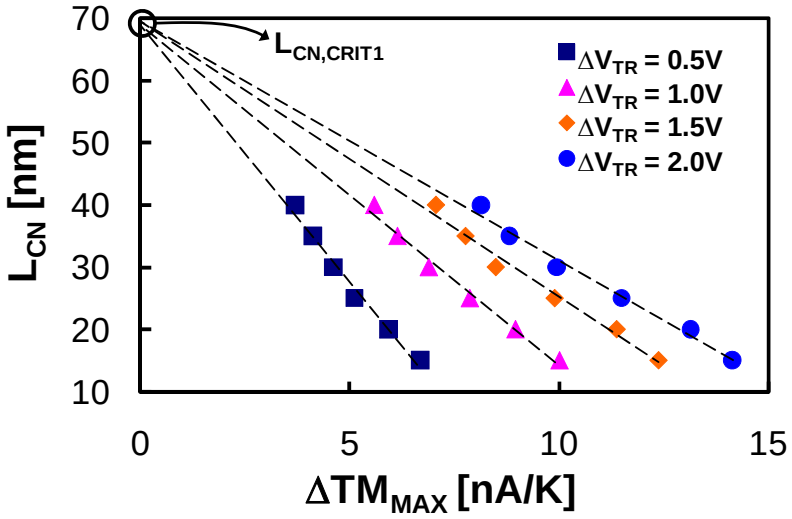

Figure IV.6: *L_{CN} as a function of ΔTM_{MAX} for cells whose V_{TR} shift is 0.5V, 1V, 1.5V, and 2V higher than the threshold voltage of the virgin cell (from the left to the right).*

peak and the silicon/oxide interface reduces gradually on increasing L_{CN}, becoming zero when $L_{CN} \geq L_{CN,CRIT1}$ (see Fig. IV.6), thus bringing the TM plot back to its virgin shape. $L_{CN,CRIT1}$ is the minimum nitride charge region length for short channel effects and bulk punchthrough to be negligible regardless V_{TR}.

Interestingly, the TM_{MAX} reduction shows a quasi-linear dependence on L_{CN} regardless the program level considered. This is clearly shown in Fig. IV.6, where L_{CN} is plotted versus the TM_{MAX} deviation from the virgin cell value (ΔTM_{MAX} = TM_{MAX}-$TM_{MAX,FRESH}$). This finding is very important, since it indicates that, for a given threshold voltage, the length of the charge distribution depends linearly on the TM peak. A general expression of this relationship can be derived, providing a simple tool to estimate the charge region length from experimental TM curves.

$$L_{CN} = k_{TM} \Delta TM_{MAX} + L_{CN,CRIT1} \qquad \text{(IV.3)}$$

k_{TM} depends on ΔV_{TR}, whereas $L_{CN,CRIT1} \approx 70$ nm is constant. k_{TM} and $L_{CN,CRIT1}$ are determined once for a given technology using the simulation procedure discussed in Paragraph IV.2.

IV.4 Subthreshold Slope Monitor

As known from the literature [80]-[82], the subthreshold slope of I_D-V_{GS} curves is sensitive to local charge trapped above the channel, and therefore can be used to develop a tool profiling local charge.

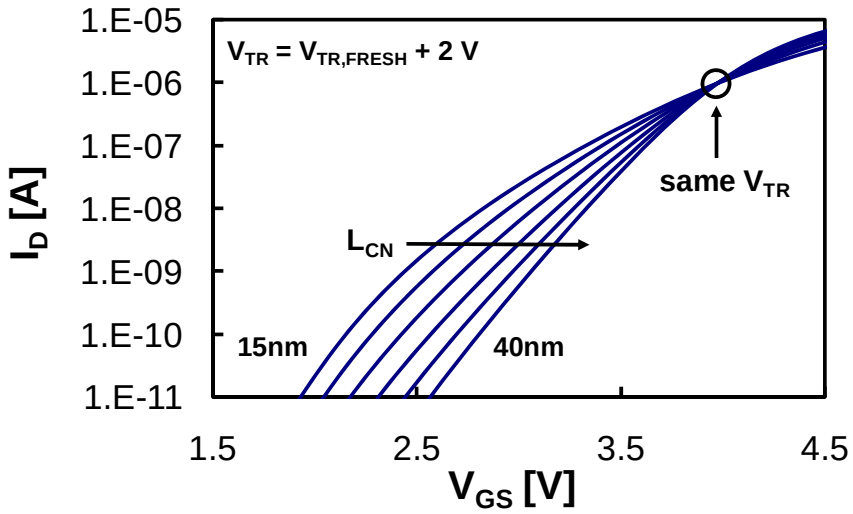

Figure IV.7: *I_D – V_{GS} curves corresponding to the rectangular charge distributions in Fig. IV.2, for $V_{TR} = V_{TR,FRESH} + 2V$.*

SS_R values measured on NROM cells programmed to different V_{TR} levels and under different bias conditions are reported in Table IV.1. Similarly to TM_{MAX}, SS_R varies with V_{TR} (see also Fig. IV.1), while cells with the same V_{TR} show different SS_R values when different program biases are adopted, indicating the presence of different nitride charge profiles.

To gain more insights on the effects of charge distribution features on the subthreshold slope, we employed I_D simulations performed as described in Paragraph IV.2. Similarly to the TM case, we found that the subthreshold slope degradation is reduced as L_{CN} increases, see Fig. IV.7 [87]. Coherently with explanation of the TM_{MAX}-L_{CN} plot, this is related to the reduction of punchthrough-like phenomena occurring when the nitride charge spreads over a wider region.

To relate L_{CN} to SS_R, we plotted L_{CN} versus the SS_R deviation from the virgin cell value ($\Delta SS_R = SS_R - SS_{R,FRESH}$), see Fig. IV.8. Similarly to the TM_{MAX}-L_{CN} plot, the graph obtained shows a linear dependence between L_{CN} and ΔSS_R regardless of the program level considered, thus allowing L_{CN} to be directly estimated from experimental I_D-V_{GS} curves.

$$L_{CN} = k_{SS} \Delta SS_R + L_{CN,CRIT2} \qquad \text{(IV.4)}$$

k_{SS} is a function of ΔV_{TR}, whereas $L_{CN,CRIT2} \approx 60$nm is a constant. k_{SS} and $L_{CN,CRIT2}$ should be determined only once for a given technology using the simulation procedure discussed in Paragraph IV.2.

It is worth noting that $L_{CN,CRIT2}$ is lower than $L_{CN,CRIT1}$, even though they have the same physical meaning. This should be due to the fact that $L_{CN,CRIT1}$ and $L_{CN,CRIT2}$ are calculated through a linear interpolation of different sets of experimental data (SS and TM peak).

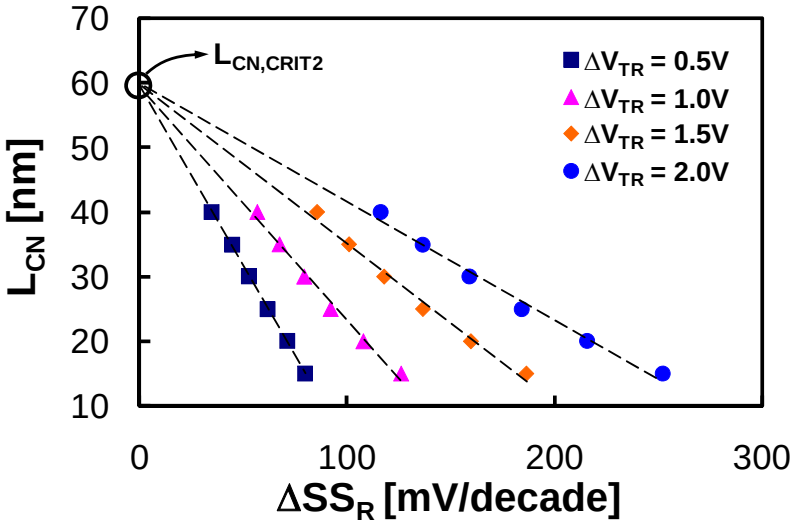

Figure IV.8: *L_{CN} as a function of ΔSS_R.*

IV.5 Charge Profiling Tools Discussion and comparison

In this section, we discuss and compare both charge profiling tools described above. To investigate their accuracy, we compared experimental I_D-V_{GS} curves to device simulations obtained by putting into the nitride the charge profile estimated through TM and SSM formulas. For clarity, we considered separately the case of programmed and erased cells.

IV.5.1 Programmed Cells

SSM and TM formulas provide the same L_{CN} results for NROM cells programmed to V_{TR} levels under different bias conditions, hence the following results are derived by using indistinguishably either SSM or TM. It is worth noting that both (IV.3) and (IV.4) calculate lengths of rectangular-shaped charge regions, whereas more realistic triangular distributions should be considered for programmed NROM devices [80], [85]. Simple formulas allowing calculating length ($L_{CN,T}$) and peak density ($\rho_{CN,TMAX}$) of the triangular charge distribution have been derived in [85], $L_{CNN,R}$ being a constant equal to 10 nm.

$$L_{CN,T} = L_{CN} + 2L_{CNN,R} \tag{IV.5}$$

$$\rho_{CN,T\,MAX} = 2\frac{\rho_{CN} \cdot L_{CN}}{L_{CN,T}} \tag{IV.6}$$

First, we calculated charge distribution profiles for cells programmed under standard

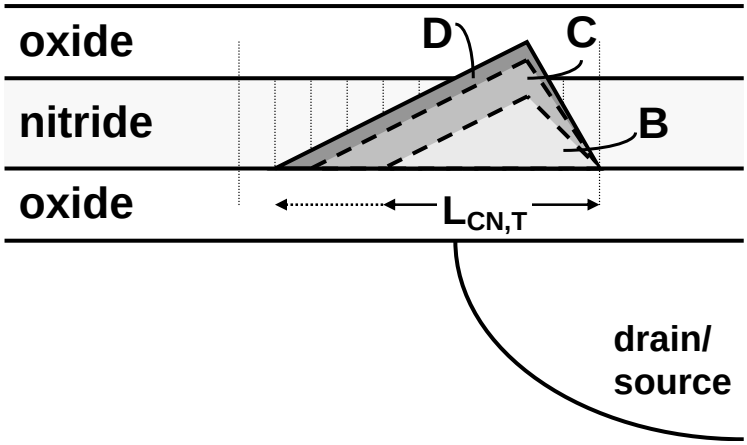

Figure IV.9: *Schematic cross section of the triangular charge distributions calculated for NROM cells programmed at B, C and D levels.*

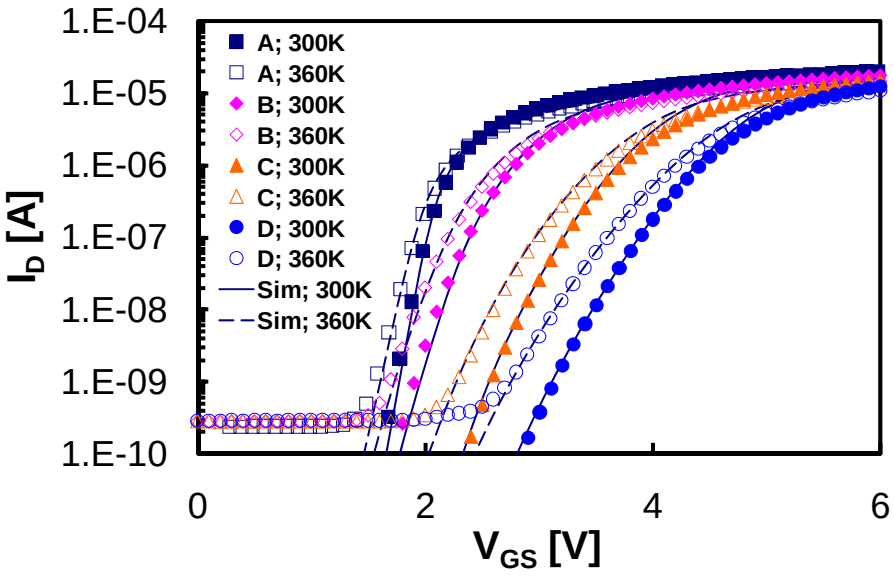

Figure IV.10: *Matching between measured (symbols) and simulated (lines) I_D-V_{GS} curves at both 300K and 360K. For the simulated curves, the triangular charge distributions sketched in Fig. 9 were defined into the nitride layer of the device.*

biases (V_{GS}=9V and V_{DS}=4.5V) from fresh (A) to full program (D) state, see Fig. IV.9. As shown in Fig. IV.10, I_D-V_{GS} simulations performed at different temperatures (300K and 360K) by considering the triangular charge distribution profiles show an excellent agreement with experimental curves, thus confirming the accuracy of above formulas. Further, charge profiles sketched in Fig. IV.9 agree also with simulated CHEI current density shapes, whose peak position from the metallurgical junction and extent above the n-well are ~7.5nm and ~20nm, respectively. Interestingly, Fig. IV.9 provides an intuitive way to monitor the evolution of nitride charge distribution during program: both length and peak density of the nitride charge increase with the program level, differently from the less realistic view reported

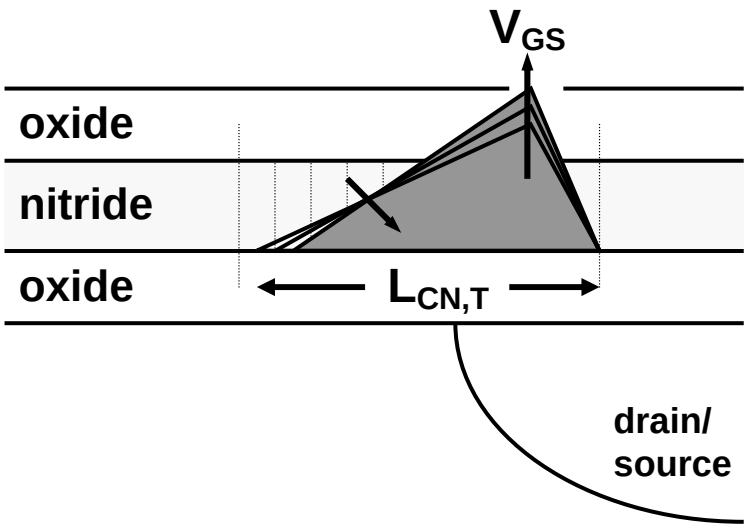

Figure IV.11: *Schematic cross section of the triangular distributions calculated for NROM cells programmed at D level with different V_{GS}. As program V_{GS} increases, L_{CN} is reduced.*

in [81] where L_{CN} was considered constant. This is due to the fact that on increasing V_{TR}, the electron charge buildup reduces the vertical field, which in turn opposes to further electron injection and deflects electrons, finally enlarging nitride charge distribution.

We used SSM and TM tools also to profile charge distribution in NROM devices programmed to D level using different V_{GS}. Figure IV.11 sketches the derived charge profiles. Also in this case (not shown here for brevity), an excellent agreement has been achieved between measured and simulated I_D-V_{GS} curves. Increasing V_{GS} leads to a slight reduction of $L_{CN,T}$ (~2,5nm) and to a slight increase of $\rho_{CN,T}$. As explained in [85], [87] this is due to the fact that the bottom oxide field (F_{OX}) increases with V_{GS}, thus enhancing the maximum charge density that can be stored into the nitride without inverting F_{OX}. Thus, the peak charge density is expected to rise, as found above.

Table IV.2: *Relative uncertainties calculated for L_{CN}.*

Program Level	TM	SSM
B	14.95%	8.25%
C	5.72%	3.34%
D	4.92%	3.79%

Although SSM and TM formulas produce similar results, they rely on different quantities, namely SS_R and TM_{MAX}, both derived from I_D-V_{GS} measurements. In order to compare the two tools, we evaluated the maximum relative uncertainty associated to the calculated L_{CN}. Results are summarized in Table IV.2 for the program levels B, C and D. As reported, SSM results are affected by a smaller uncertainty, due to the different dependence of TM_{MAX} and

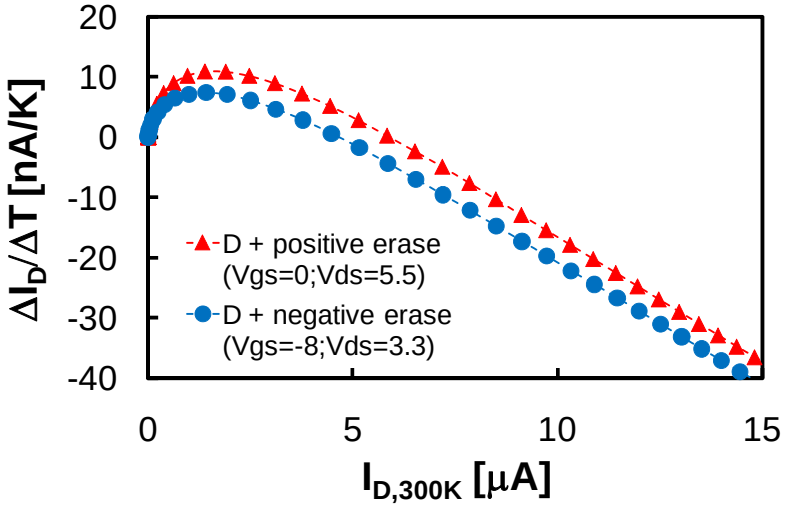

Figure IV.12: *TM curves corresponding to cells initially programmed to level D and erased to level C with positive (empty triangles) and negative (empty circles) erase.*

SS_R formulas on drain current measurements [88]. Thus, since SSM has the smaller uncertainty and it requires only one I_D-V_{GS} measurement performed at room temperature, we believe that it is the best I_D–V_{GS} based tool to derive charge distribution features in programmed NROM devices.

IV.5.2 Erased Cells

NROM cells are erased by using HHI mechanism, and, due to the difficult control of electron and hole injection, the hole distribution is only partially overlapped to the program electron charge [43]. The misalignment between physically separated electron and hole distributions is believed to be one of the major reliability concern for NROM devices [73], [74], [82], hence tools able to infer features of hole distributions are very desirable. In this context, we used both SSM and TM to investigate charge distribution features in erased NROM cells.

We considered programmed (D level) NROM cells erased back to C level by using both positive and negative erase schemes. SS_R values derived from I_D-V_{GS} curves (not shown here for brevity) do not exhibit significant differences, thus suggesting that SSM is not particularly sensitive to electron/hole charge differences in erased cells with the same V_{TR}. On the contrary, TM curves depicted in Fig. IV.12 show different peaks, suggesting different charge distributions in the nitride [85]. To deeply investigate this point, we measured TM curves also on programmed NROM cells erased back to virgin (A) level using both negative and positive erase schemes. TM curves found for both erased and fresh cells are perfectly equal despite the erase scheme adopted, suggesting that TM is sensitive to electron/hole distribution profile only if the cell is not fully erased.

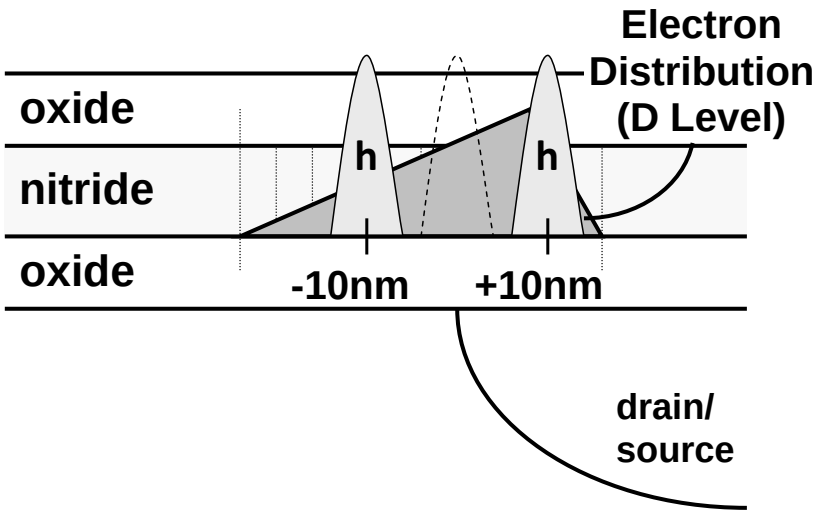

Figure IV.13: *Schematic cross section zoom of hole distributions superimposed on the electron distribution derived for program level D. The hole density was adjusted in order to obtain half erased and fully erased conditions.*

To understand the TM behavior, we used again device simulations. We simulated I_D-V_{GS} curves at both 300K and 360K, assuming some hole charge distributions superimposed to the triangular electron distribution found for NROM cell programmed to D level. Hole distribution centroid is placed at –10nm, 0nm and +10nm from the metallurgic junction, as sketched in Fig. IV.13. The width of hole distribution is adapted so that the bottom oxide field does not invert when erase biases are applied. The hole density is then adjusted to have a Fully Erased cell (FE; the reverse threshold voltage is back to its native value) and Half Erased cell (HE; $V_{TR} = V_{TR,FRESH} + \Delta V_{TR,D}/2$, $\Delta V_{TR,D}$ being the threshold voltage window).

Table IV.3 reports TM_{MAX} values calculated from I_D-V_{GS} simulations when varying hole centroid. In agreement with experimental data, TM_{MAX} values obtained from FE cells are indistinctly close to the fresh cell one, confirming that TM is unable to provide information about hole position when the cell is erased back to its virgin V_{TR}. On the contrary, when cells are half erased, TM_{MAX} values show slight variations with the charge centroid. In agreement with the fact that drain current is not affected by nitride charge above the junction [80], TM sensitivity drops when the hole distribution centroid is moved beyond the junction.

Table IV.3: *SS_R and TM_{MAX} values extracted from simulations for HE and FE cells.*

Hole Centroid Position [nm]	TM_{MAX} [nA/K]		
	Fresh Cell	HE	FE
-10	4.85	15.8	4.88
0		12.6	4.84
+10		12.4	4.79

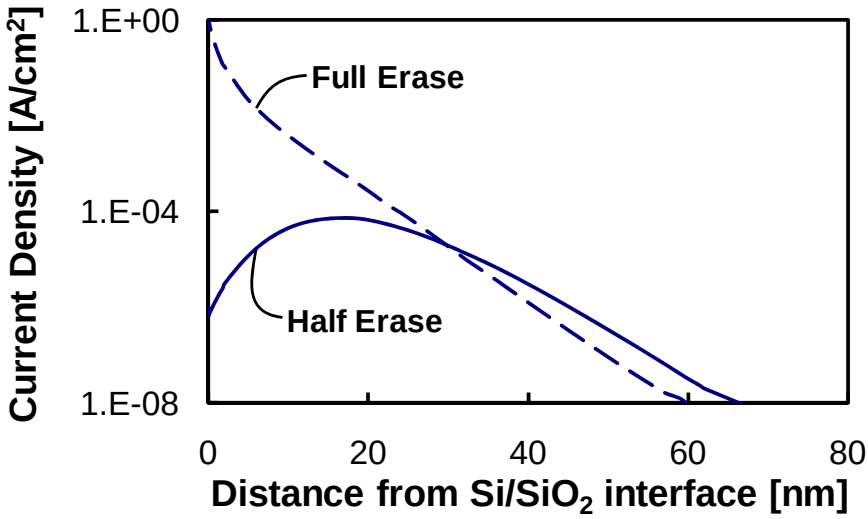

Figure IV.14: *Current density along a vertical cut in the channel at 2nm from the drain junction (i.e. under the nitride charge region). The gate of the device is biased at 1.5 V, corresponding to a device operating in subthreshold regime. Continuous and dashed lines refer to half erase and full erase conditions, respectively.*

Furthermore, despite TM_{MAX} variations to hole centroid, TM formulas allow deriving only length and density of net charge distributions. Although the net charge distributions derived in this way allow reproducing accurately experimental I_D curves [85], no quantitative information are provided on electron and hole distributions. This is due to the fact that TM, and in general all I_D-V_{GS} based charge profiling tools, are sensitive only to the overall vertical field above the channel. For this reason, it is extremely difficult to establish formulas like (IV.2)-(IV.4) to derive hole distribution features in erased cells. In this respect, TM can be considered only a qualitative tool to investigate electron and hole charge distribution characteristics in erased NROM cells.

In order to understand why TM is sensitive to different hole distributions only in half erased NROM cells, we should remember that TM sensitivity is strictly related to punchthrough-like conduction and its temperature sensitivity. When cells are fully erased, the hole distribution completely compensates program electron distribution effects, see Fig. IV.14. From an electrostatic point of view, the classical picture of the maximum I_D density at the silicon/oxide interface is observed: no punchthrough-like conduction occurs, thus explaining the TM insensitivity. On the contrary, in half-erase conditions, holes are not enough to compensate electron distribution, hence punchthrough-like conduction occurs (see Fig. IV.14), demonstrating that TM graph is sensitive to the nitride charge scenario.

IV.5.3 Comparison With Other Charge Profiling Methods

From the above discussion on TM and SSM as tools to profile charge distributions in the nitride, we can derive the following guidelines. If programmed NROM cells are considered, SSM is to be preferred due to its lower uncertainty and application simplicity (only one I_D-V_{GS} measure is needed). On the contrary, if erased cells are considered, both methods are unable to profile physically separated electron and hole distributions, despite slightly higher TM sensitivity. In fact, both SSM and TM are sensitive to the whole vertical electric field, i.e. to the net charge in the nitride portion above the channel.

As known, other techniques, namely CP and GIDL, have been proposed in the literature to profile the charge in the nitride [76]-[78], [82]. The GIDL-based method has been used to qualitatively investigate the hole distribution in erased cells [82]. Physically, this is due to the fact that the maximum sensitivity region of GIDL currents and BTB hole generation is above the n-well [82], where hot holes are injected during erase. For the same reason, GIDL method cannot be used to profile electron distribution in programmed cells. This is demonstrated also by device simulations, showing that GIDL current formulas similar to (IV.3) and (IV.4) are not accurate in profiling electron charge distributions in programmed NROM cells. On the other hand, the charge-pumping technique requires a non-trivial dedicated experimental setup and a complex elaboration of measured data [76]-[78]. Furthermore, its application is strongly limited by several assumptions such as a monotonic V_{TR} profile [76], [77], the absence of trapped charge in fresh devices [77] and a uniform density of surface states, that is also assumed to be unaffected by the program operation [76]-[78]. Further, this technique allows estimating only the charge trapped above the device's channel and, due to the high voltages needed to scan the full channel (applied gate pulses can easily reach more than 8V) the measurement procedure could affect the existing charge distribution [77].

From the above discussion, SSM seems to be the best choice for nitride charge profiling. Its main advantage is to provide a very simple and straightforward procedure to profile the charge distribution. For a given technology, device simulations are needed only once to estimate the parameters of equations (IV.2) and (IV.4). Then, these formulas can be used to directly estimate density and length of the charge distribution from a simple trans-characteristic. Unfortunately, SSM does not allow profiling physically separated electron and hole distribution after erase. In this case, GIDL can provide some insights, although no one of the discussed methods alone allow deriving a quantitative picture of electron and hole distributions in the nitride.

IV.6 Chapter Summary

In this Chapter, two I_D-V_{GS} based tools that are sensitive to charge distribution features in NROM devices are presented and discussed. The first, referred to as Temperature Monitor, exploits temperature effects on I_D-V_{GS} trans-characteristic, whereas the second, referred to as Subthreshold Slope Monitor, exploits effects of local trapped charge on subthreshold slope. Both tools provide simple formulas allowing deriving length and density of net charge distribution. Tool accuracy and sensitivity have been investigated, comparing them also to GILD and CP techniques. Both SSM and TM provide similar results when applied to programmed cells: by assuming the charge features derived from SSM/TM formulas, device simulations accurately reproduce I_D-V_{GS} measurements. Unfortunately, these tools do not let derive quantitative information on electron and hole distributions in erased NROM cells, since they are sensitive only to the net charge above the channel.

CHAPTER V

Hole Distributions in Erased NROM Devices: Profiling Method and Effects on Reliability

In this chapter, a new technique to profile hole distributions in erased NROM devices is presented. Device simulations, compact models and experimental data are combined to derive the hole distribution in erased cells. The accuracy of the derived final charge scenario is verified by comparing experimental data and simulation results. The technique is then used to monitor the evolution of the nitride charge with cycling, and to correlate it to the degradation of memory reliability after cycling.

THE presence of physically separated electron and hole distributions generated by program and erase operations is reported to be one of the main causes of device's retention degradation. Therefore, a deep knowledge of the features and evolution of the nitride storage charge is crucial for reliability, cell optimization, future scalability and multi-level operation.

As discussed in the previous Chapter, several papers published in the literature presented methods to profile the charge distribution in NROM/SONOS devices. Unfortunately, they focus mainly on programmed cells [76], [77], [79], [80], [85], [87], [89], whereas the task of profiling the storage charge in erased memory cells has been only qualitatively addressed [48]. Early works investigated the effects of the program charge features on threshold voltage (V_T) and subthreshold slope (SS) characteristics [79], [80] and, more recently, simple formulas to evaluate length (L_{CN}) and density (ρ_{CN}) of the charge distribution from the I_D-V_{GS}

of a programmed device have been derived [85], [87], [89]. Gate-induced drain leakage (GIDL) current measurements were also employed [48], [77], mainly to confirm results obtained with standard I_D-V_{GS} characteristics and device simulations. However, neither I_D-V_{GS} nor GIDL based techniques can be used to profile hole distributions, being sensitive only to the net charge above the channel [89] or above the junction, respectively. Alternatively, charge-pumping (CP) has been used to characterize the lateral charge distribution in local charge-trapping memories [76]-[78], [90]. Unfortunately, CP can be applied only when the threshold voltage increases monotonically along the channel [76], [77], i.e. on programmed cells, although it has been improperly used also to profile the charge distribution in erased cells [78], [90], where this condition no longer holds. Further, this technique does not allow extracting the full charge profile, since its sensitivity is limited to the channel region.

In this Chapter, a new method that overcomes the abovementioned limitations allowing profiling hole distributions in cycled NROM cells is presented. It exploits and combines simulations from both compact models and device simulations, and GIDL currents measurements. Results obtained with this method are proven by comparing experimental I_{DS}-V_{GS} and $I_{D,GILD}$-V_{GS} characteristics to simulations. This technique is then used to monitor the nitride charge evolution with cycling and to correlate electron and hole distributions to the degradation of memory reliability after cycling.

V.1 Experiments

Samples used in this work are NROM cells manufactured in 0.5-μm technology. Their effective channel length is 0.32μm. Top and bottom oxides are 8 nm thick, whereas the interleaving nitride layer is 6.5 nm thick. The reverse-mode threshold voltage, V_{TR}, is defined as the gate voltage at which the drain current I_D reaches the value of 1μA with V_{DS} = 0.1V. NROM cells were programmed to three different program levels (B, C and D), whose V_{TR} is 0.3 V, 1.1 V, and 2.3 V higher than the threshold voltage of a non-programmed device (A), and then erased back to their native V_{TR} by applying two different erase bias schemes, called "positive erase" (V_{GS} = 0V, V_{DS} = 5.25V), and "negative erase" (V_{GS} = -10V, V_{DS} = 2V), respectively.

Fig. V.1(a) shows experimental I_D-V_{GS} characteristics of programmed (filled symbols) and erased (empty symbols) NROM cells. As shown, the subthreshold slope monotonically increases with the program level and is completely recovered when the cell is erased back to its native state. Fig. V.1(b) shows GIDL current for the same program and erase V_{TR} levels considered in Fig. V.1(a) performed by biasing the drain (V_D=3V) and leaving source and body floating. As shown, the GIDL current at a given V_{GS} increases with the program level

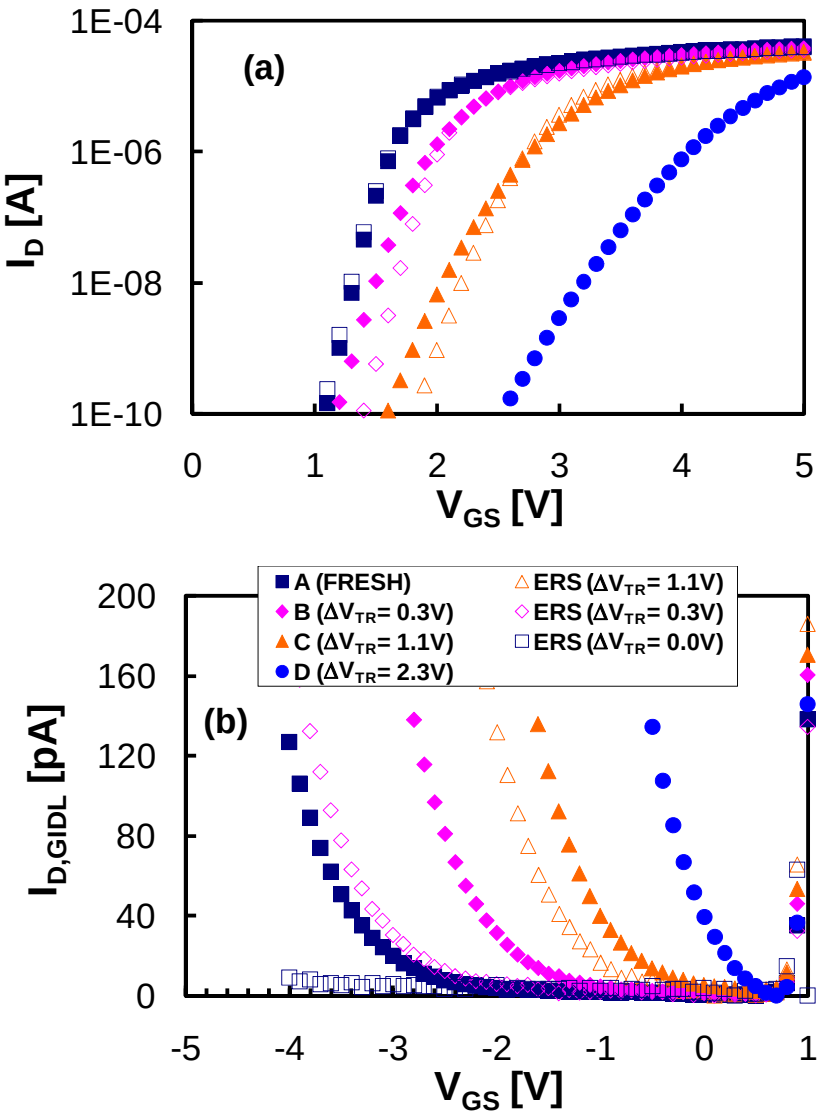

Figure V.1: *Experimental I_D-V_{GS} (a) and $I_{D,GIDL}$-V_{GS} (b) characteristics for NROM cells programmed (filled symbols) and negatively erased (empty symbols) to different levels. DV_{TR} is relative to the threshold voltage of a fresh device.*

(filled symbols), and the whole characteristic is shifted toward less negative gate voltages. An opposite behavior is observed when the cell is erased (empty symbols).

V.2 Hole Charge Profiling

As discussed above, none of the techniques proposed in the literature can be used to correctly profile electron and hole distributions in erased NROM devices. In fact, even if programmed and erased cells with the same V_T have different I_{DS}-V_{GS} subthreshold slopes and GIDL currents, as shown in Fig. V.1(a)-(b), these information can be used only to estimate the net

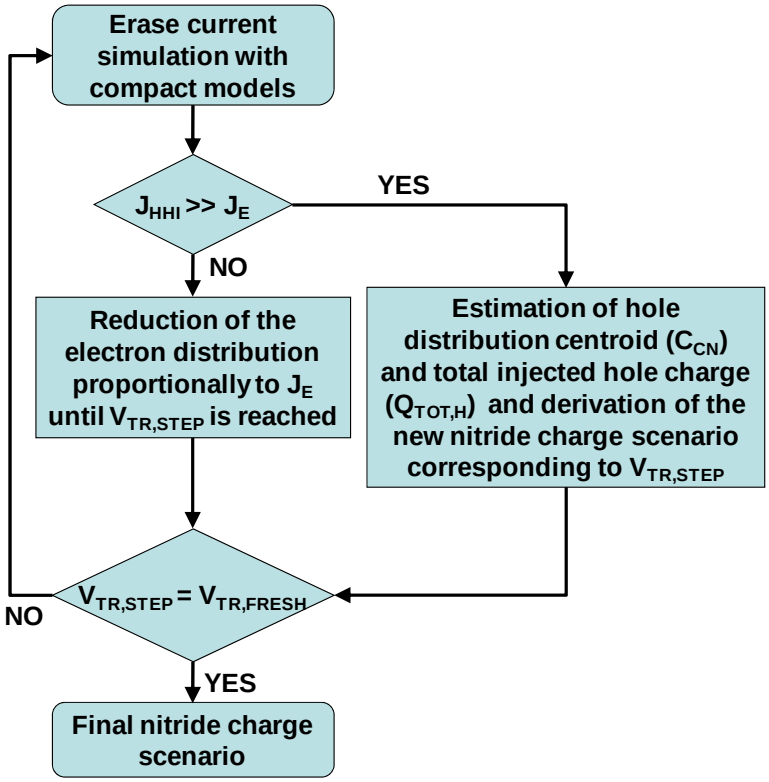

Figure V.2: *Flowchart of the methodology proposed to profile hole charge distribution.*

storage charge, because these physical quantities are sensitive only to the overall electric field [89]. Nevertheless, the $I_{D,GILD}$-V_{GS} curve of a completely erased cell is significantly different from the fresh one, as shown in Fig. V.1(b), suggesting that GIDL currents can be used to derive holes distribution features even when the device is erased back to its initial V_T. Unfortunately, this is not sufficient to profile correctly hole distribution in cycled devices, but can be successfully integrated to this purpose with compact models and commercial drift-diffusion device simulations, resulting in the procedure sketched in Fig. V.2. Starting from the knowledge of the charge distribution at the end of the program operation, obtained as described in [89], this procedure allows deriving the final nitride charge scenario at the end of the erase operation. The dynamic V_{TR} evolution occurring during the erase is divided into a series of static steps. For every step, three basic operations should be performed: i) the physical mechanism dominating the erase operation should be identified, ii) the amount of holes (electrons) that have to be injected (extracted) into (from) the nitride layer of the device should be determined, and iii) the final nitride charge scenario at the end of the step should be computed. For simplicity, we will consider in the following only two steps, that we called Half-Erased (HE) and Full-Erased (FE), whose threshold voltages are given by $V_{TR,FRESH} + \Delta V_{TR,D}/2$ and $V_{TR,FRESH}$, respectively. $V_{TR,FRESH}$ is the V_{TR} of a fresh device, whereas $\Delta V_{TR,D}$ is the threshold voltage window for a NROM cell programmed to the D level.

Table V.1: *Maximum values of simulated J_{HHI} and J_E profiles for devices at the beginning (INITIAL) and at the half (HE) of the erase operation for both positive and negative erase.*

Erase Scheme	$J_{HHI,MAX}$ [A/cm^2]	$J_{E,MAX}$ [A/cm^2]
Positive (INITIAL)	6.85e-9	1.44e-15
Negative (INITIAL)	4.74e-4	7.65e-2
Positive (HE)	2.39e-5	1.07e-10
Negative (HE)	8.76e-4	1.11e-4

The first phase of the procedure is dedicated to identify the physical mechanism playing the dominant role in NROM erase. Even though it is generally known that under *conventional* erase schemes NROM erase is due to hot holes injection [43], electron detrapping may become important when a negative erase scheme is adopted, due to the presence of a high vertical field across the bottom oxide. To evaluate HHI (J_{HHI}) and electron detrapping (J_E) currents, compact models described in [43] were used for simulations. Maxima of J_{HHI} and J_E current simulations are reported Table V.I, for both positive and negative erase schemes. As expected, the electron detrapping contribution is negligible when the positive erase scheme is adopted, whereas it is significant when the negative bias schemes is considered and at the beginning of the erase operation.

If electron detrapping is the dominant erase mechanism (i.e. $J_{HHI} << J_E$), the intermediate threshold voltage $V_{TR,STEP}$ level is reached by simply reducing the electron charge distribution proportionally to the J_E profile, see Fig. V.2. On the contrary, if hot hole injection is the dominant erase mechanism (i.e. $J_{HHI} >> J_E$), holes are injected into the nitride to reach $V_{TR,STEP}$.

The final hole distribution is derived combining GIDL current ($I_{D,GIDL}$) measurements with 2D device simulations. In fact, $I_{D,GIDL}$ is sensitive to the injected hole profile also when the device is fully erased (see Fig. V.1(b)), because Band-To-Band-Tunneling (BTBT) hole generation is very sensitive to the local field above the junction (the maximum sensitivity is ~10 nm beyond the junction above the n-well). To relate hole distribution features to experimental GIDL characteristics, we used the deviation of the experimental GIDL threshold voltage $V_{T,GIDL}$ from the native cell value, $\Delta V_{T,GIDL} = V_{T,GIDL} - V_{T,GIDL(FRESH)}$, as suggested in [48]. $V_{T,GIDL}$ is defined as the gate voltage that must be applied to the device to obtain a 40pA GIDL current with $V_D = 3V$, body and source floating. Then, we employed 2D drift-diffusion device simulations to analyze the effects of centroid C_{CN} and total charge $Q_{TOT,H}$ of hole distribution on GIDL current. We calibrated the BTBT models [92] by reproducing the $I_{D,GIDL}$-V_{GS} curve of a virgin NROM cell. Geometry and doping information fed to the device were obtained as output of a 2D process simulation. To evaluate the effects of C_{CN} and $Q_{TOT,H}$ on GIDL currents, we superimposed rectangular hole distributions having different lengths

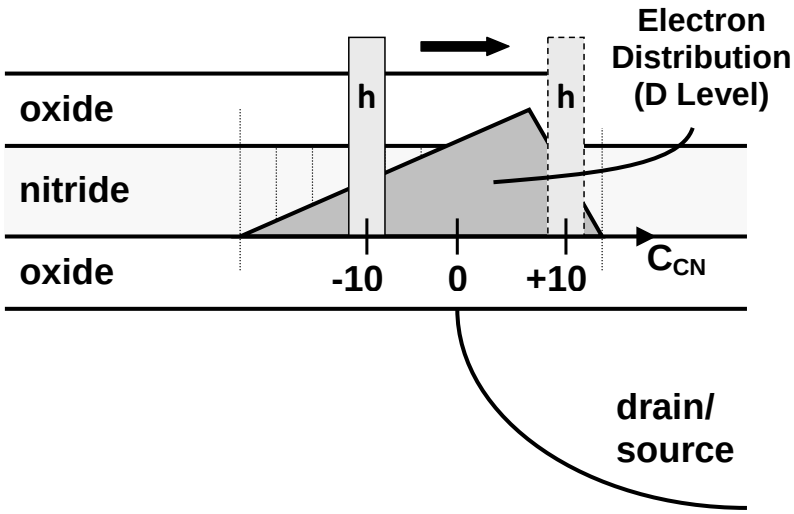

Figure V.3: *Schematic cross section of the rectangular hole distributions superimposed on the program distribution to analyze the impact of C_{CN} and $Q_{TOT,H}$ on $V_{T,GIDL}$.*

($L_{CN,H}$) and positions (C_{CN} = –10nm, -5nm, 0nm, +5nm and +10nm) on the existing charge distribution, as sketched in Fig. V.3 for a programmed cell, while $Q_{TOT,H}$ is adjusted to give the same $V_{TR,STEP}$. Then, we used $\Delta V_{T,GIDL}$ values extracted from GIDL current simulations to build the C_{CN} vs. $\Delta V_{T,GIDL}$ and $Q_{TOT,H}$ vs. $\Delta V_{T,GIDL}$ plots shown in Figs. V.4 and V.5. Using these graphs, it is very easy to derive C_{CN} and Q_{TOT} directly from $V_{T,GIDL}$ measurements. Interestingly, the length of the hole distribution does not affect significantly neither C_{CN} nor $Q_{TOT,H}$ plots, thus easing their estimate. The estimated hole distribution parameters, i.e. C_{CN} and Q_{TOT} are determined unambiguously, as they allow reproducing accurately both I_D-V_{GS} and $I_{D,GIDL}$-V_{GS} curves, that have different sensitivity regions against the charge position along the channel. In this respect, I_D-V_{GS} and $I_{D,GIDL}$-V_{GS} curves are complementary, as I_D-V_{GS} is sensitive the charge above the channel, whereas the GIDL current is affected mainly by the charge above the n-well.

On the other hand, $L_{CN,H}$ cannot be estimated from the above plots, as hole distributions with different $L_{CN,H}$ (and the same C_{CN} and $Q_{TOT,H}$) result in the same $I_{D,GIDL}$-V_{GS} curve. For this reason, the hole distribution length has been derived considering both the profile along the channel of erase current simulations and the $L_{CN,H}$ effect on the threshold voltage.

This hole-profiling method depends on the $V_{TR,STEP}$ considered and on the charge distribution after the program operation. Even though this procedure seems to be slightly complicated, it has several advantages compared other methods, i.e. both those relying on I_{DS}-V_{GS} curves [80], [85], [87], [89] and CP technique [78]-[90]. First, it can be used to profile the nitride charge in the erased state, whereas other techniques cannot be applied. For example, CP requires a monotonically increasing threshold voltage [76], [77], and this condition no longer holds in erased devices. Second, it allows deriving accurately charge distribution

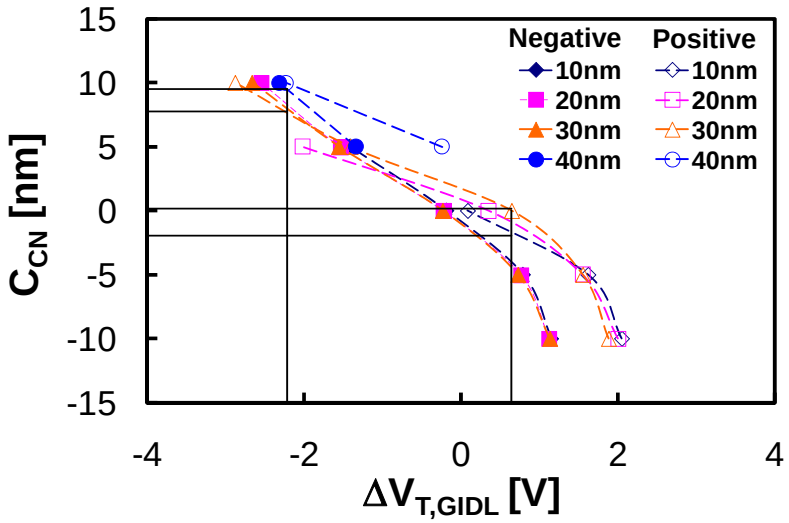

Figure V.4: *C_{CN} vs. $\Delta V_{T,GIDL}$ plot extracted from simulated FE conditions. Empty and filled symbols correspond to negatively and positively erased devices, respectively. Each symbol corresponds to a different width of the considered hole distribution.*

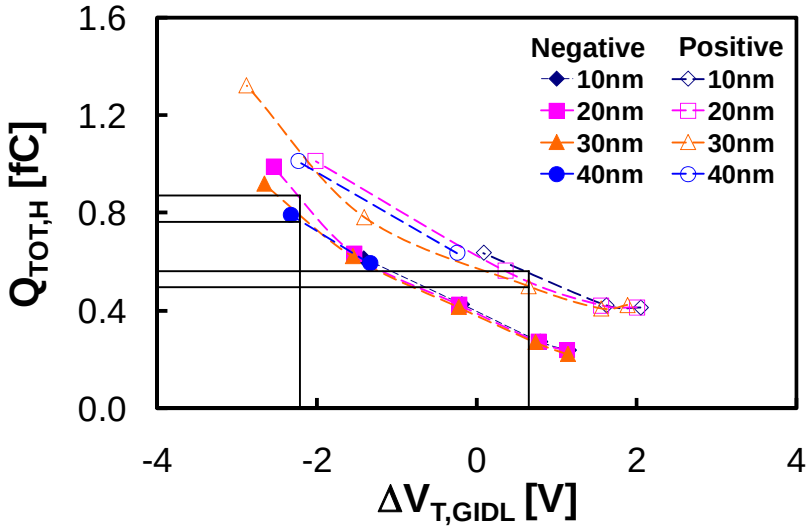

Figure V.5: *$Q_{TOT,H}$ vs. $\Delta V_{T,GIDL}$ plot extracted from simulated FE conditions. Empty and filled symbols correspond to negatively and positively erased devices, respectively. Each symbol corresponds to a different width of the considered hole distribution.*

features along the whole channel, whereas the sensitivity of CP technique is limited to the channel region [77].

V.3 Nitride Charge Evolution With Cycling

The procedure described in the previous Paragraph was adopted to profile hole distributions in NROM cells erased by using positive and negative bias schemes. We determined hole and

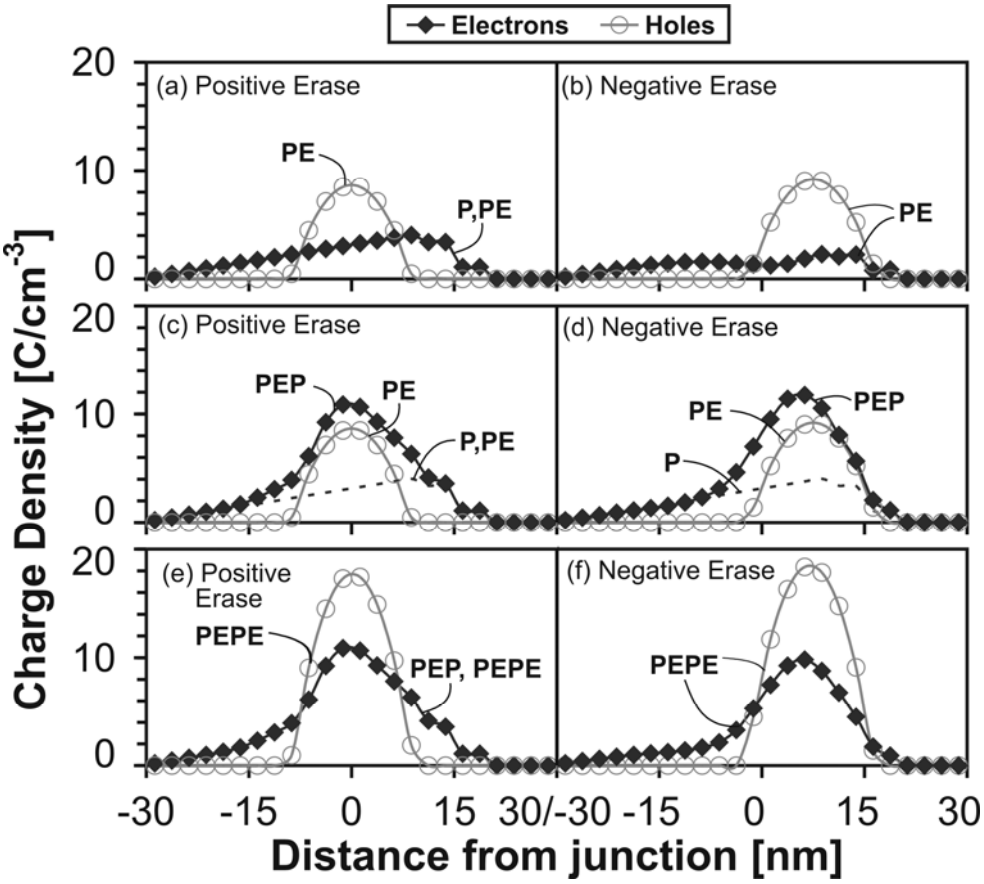

Figure V.6: *Evolution of total electrons and holes charge in the nitride. The labels "P" and "E" refer to program and erase operation, respectively. Positive (negative) values on the x-axis correspond to the junction (channel) region of the device.*

electron distribution after erase following the method described in the previous Section. In the positive erase case, we determined C_{CN} and $Q_{TOT,H}$ using the plots in Figs. V.4 and V.5. For the negative erase, electron detrapping has been included. The amount of electron charge escaping from the nitride has been derived from the J_E profile along the channel. $L_{CN,H}$ was estimated from the profile of the erase current simulations, checking the threshold voltage. Hole distributions derived at the end of the first erase operation are sketched in Fig. V.6(a) and (b) for positive and negative erase schemes, respectively. As expected, hot holes are injected closer to the n-junction when a negative bias scheme is adopted because of the higher vertical field and lower lateral field, whereas the higher lateral field during positive erase allows holes to be injected farther from the drain junction. To test the accuracy of results achieved through this technique, we simulated I_D-V_{GS} and $I_{D,GIDL}$-V_{GS} curves by inserting the estimated charge profile into the device nitride layer. As shown in Fig. V.7(a)-(b), GIDL current simulations reproduce accurately both linear and logarithmic experimental curves, proving the accuracy of this technique. A good agreement is also found when comparing simulated and measured I_D-V_{GS} curves, not shown here for brevity.

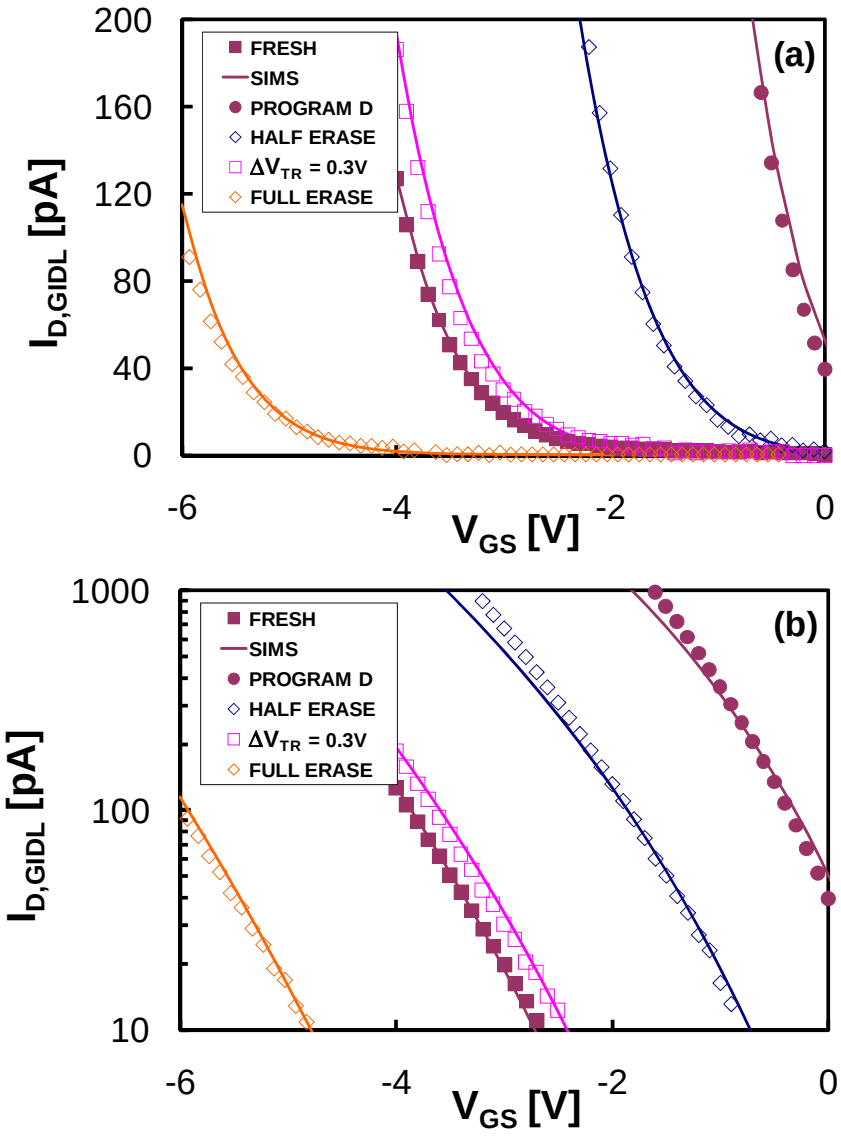

Figure V.7: *Comparison between experimental (symbols) and simulated (lines) GIDL on characteristics for the negative erase case. Simulations were carried out by inserting derived hole distributions into the nitride layer. Both (a) linear and (b) logarithmic scale are shown.*

We used this methodology together with 2D device simulations to monitor the charge evolution in cycled devices, starting from the nitride charge scenario derived at the end of the first cycle, see Fig. V.6(a)-(b). In this case, the procedure described in the previous Section is applied to devices that were re-programmed after erase. The nitride charge scenario at the end of the second program operation has been derived using CHEI simulations [80]. The V_{TR} evolution occurring during program operation is divided into three static steps, that are the B, C and D threshold voltage levels defined in Paragraph V.1. The electron distribution is increased to reach the next intermediate V_{TR} step, assuming that the profile of J_{CHEI} simulations represents the electron injection profile. Repeating this procedure three times leads to the final (level D) program electron distribution, see Fig. V.6(c)-(d).

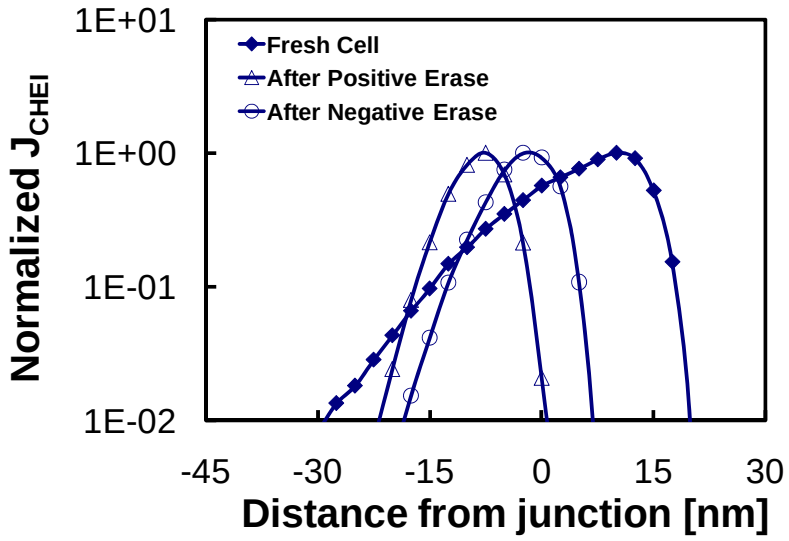

Figure V.8: *Simulated J_{CHEI} profile at the beginning of the second program operation are compared with the J_{CHEI} profile simulated for a fresh device. Positive (negative) values on the x-axis correspond to the junction (channel) region of the device.*

Interestingly, the amount of electron charge above the channel increases with cycling as generally believed [74], and this could affect device reliability. This is related to the fact that the peak of CHEI current density simulated at the beginning of program operation in cycled devices moves into the channel compared to virgin cells, regardless the erase scheme adopted, as clearly shown in Fig. V.8. Physically, this is due to the fact that trapped holes reduce the channel potential drop near the junction, moving the peak of the lateral field into the channel. This explains why the amount of electron charge stored in the nitride portion above the channel increases with cycling as reported in [74]. The most straightforward consequence of this electron increase is that more holes have to be injected to erase the cell, i.e. to compensate the electron effect on the channel current. This is confirmed by hole profiles derived at the end of the second cycle, sketched in Fig. V.6(e)-(f). As shown the amount of holes increases, as well as reliability issues related to the lateral electric field in the nitride.

V.4 Correlation to memory Retention: erase V_{TR} drift

Even though there is not yet an unanimous agreement, it is generally believed that the lateral movement of charge in the nitride plays a role in NROM retention degradation. Among various reliability issues, we focus here on V_{TR} drift measured on erased cells left unbiased at room temperature conditions, see Fig. V.9(a). Since analyses and explanations proposed until now were mainly qualitative [48], we used device simulations and previously derived electron and hole charge distributions to explain the physical mechanism of this phenomenon and also

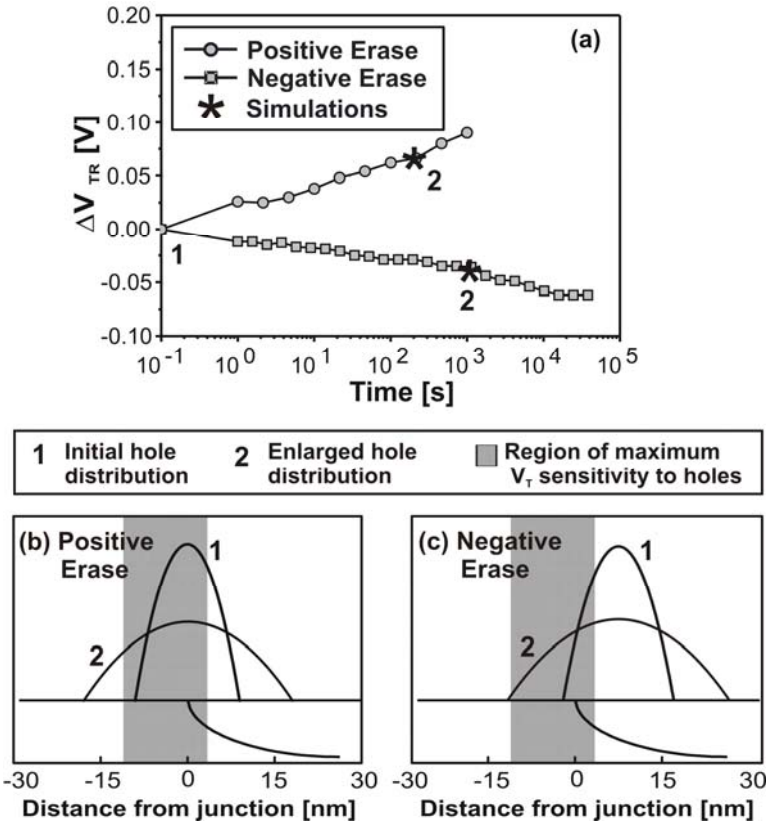

Figure V.9: *Erase state retention losses as a function of the applied erase scheme (a). The retention losses simulated by assuming the hole charge redistributions sketched in (b)-(c) are also shown (stars). The shaded area in (b)-(c) represents the region of maximum sensitivity of V_{TR} against hole charge position.*

its dependence on the erase schemes. To this purpose, we correlated the V_{TR} drift slope to both the lateral field in the nitride (E_{NI}) and the vertical field across the bottom oxide (E_{BOX}), derived from unbiased cell simulations assuming electron and hole charge distributions depicted in Figs. V.6(a)-(b).

If V_{TR} drift would be driven by the vertical field across the bottom oxide, there would be a correlation between E_{BOX} and the V_{TR} drift slope polarity, see Fig. V.9(a). Further, since positively erased cells show a higher V_{TR} drift, this should also indicate a higher E_{BOX} magnitude. On the contrary, simulations show that E_{BOX} polarity does not depend on the erase bias scheme, whereas the bottom oxide magnitude is higher for negatively erased cells. Both these findings seem to exclude that the vertical movement of charge across the oxide is the root cause of V_{TR} drift.

On the contrary, the lateral movement of holes can explain both positive and negative V_{TR} drifts. To obtain a quantitative explanation, we assumed that trapped electrons cannot move after injection [74], [93], whereas the hole displacement driven by E_{NI} [48], [74], [90], [93] is responsible of the lateral redistribution of trapped charge. Independently of the mechanism of

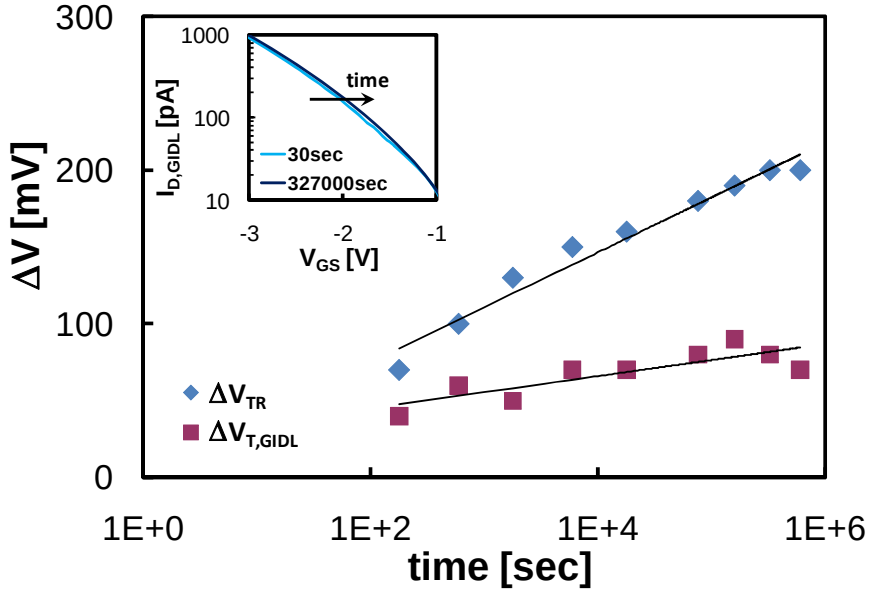

Figure V.10: *Variations of V_{TR} and $V_{T,GIDL}$ as a function of the retention time, for a device half erased under a positive erase scheme. The inset shows logarithmic $I_{D,GIDL}$-V_{GS} curves for two different retention times.*

charge transport in the nitride, probably driven by the internal field E_{NI} [90], [93]-[95], we enlarged the hole distribution (L_{CN} was doubled), keeping the total charge amount constant, as sketched in Figs. V.9(b)-(c). Then, we used device simulations to evaluate the effects on V_{TR} of the widening of hole distributions. Results depicted as stars in Fig. V.9(a) show that the opposite polarity of V_T shifts is well reproduced by the hole distribution enlargement. Physically, this is related to the relative movement of trapped holes with respect to the region where the hole charge effect on V_{TR} is maximum, shown as shaded area in Figs. V.9(b)-(c). This region has been determined by evaluating through device simulations the V_{TR} shift occurring when a narrow rectangular hole packet is moving along the channel of a programmed NROM cell. As expected, when the hole packet is moved from the center of the channel towards the n-junction, its effect on V_{TR} becomes more evident. When a positive erase scheme is adopted, holes are mainly injected inside this region, as shown in Fig. V.9(b). Then, a consistent fraction of them moves outside due to the E_{NI}-driven hole distribution enlargement, thus determining the V_{TR} increase. The opposite happens if a negative erase scheme is adopted. In this case, holes are injected outside the maximum sensitivity region, see Fig. V.9(c), and the E_{NI}-driven hole distribution widening moves them into the maximum sensitivity region, reducing V_{TR}.

According to the above explanation and to the simulation results shown in Fig. V.4 and V.5, the GIDL characteristics before and after the V_{TR} drift should not vary significantly, since $L_{CN,H}$ effects on $I_{D,GIDL}$ are negligible. This is confirmed by the experimental results shown in Fig. V.10 for a device that was programmed to level D and subsequently half-erased

with a positive erase scheme. The evolution of both V_{TR} and $V_{T,GIDL}$ is reported as a function of the retention time. As can be seen, $V_{T,GIDL}$ is approximately constant, whereas ΔV_{TR} increases (~200mV after 10^6 seconds), completely confirming our simulation results.

V.5 Chapter Summary

In this Chapter, a new technique allowing hole distributions profiling in NROM device, which combines compact model and 2D drift-diffusion simulations with experimental GIDL currents, has been presented. Electron discharge effects have been taken into account, and the accuracy of the final charge scenario obtained has been verified by comparing drain and GIDL currents to simulations.

Monitoring charge evolution after program and erase operations allows explaining some general mechanisms related to NROM reliability, usually only qualitatively addressed. First, it has been demonstrated that in cycled devices the amount of electrons in the nitride portion above the channel increases, because holes injected above the junction during erase shift the maximum lateral field into the channel.

Second, it has been proved that V_{TR} drifts occurring in NROM cells left unbiased in the erased state is due to the lateral migration of trapped holes. The model presented allows explaining also the polarity dependence on the erase scheme adopted.

CHAPTER VI

Modeling TANOS Memory Program Transients to Investigate Charge Trapping Dynamics

This chapter describes a novel physics-based drift-diffusion model of TANOS program transients. Shockley-Read-Hall (SRH) theory is used to describe the capture process, whereas Thermal Emission (TE) with field-induced trap energy barrier lowering, TAT and Trap-to-Band (TBT) tunneling mechanisms are considered for the electron emission process. The model also accounts for the electron transport in the nitride conduction band and is used to investigate electron trapping/detrapping dynamics in the nitride. Trapping process is found to be independent from the energy of injected electrons, while detrapping is dominated by trap-to-band tunneling mechanisms. The propoesd model allows monitoring the evolution of trapped charge during TANOS program transients providing useful information for the optimization of the TANOS memory devices.

TANOS-type memory is one of the most promising candidate for scaled NAND Flash technology, allowing improved reliability and scaling perspectives [14], [97]. In order to optimize device reliability and performance it is fundamental to understand the physical mechanisms of charge trapping/detrapping in the silicon nitride trapping layer and derive the features of the trapped charge distribution. In this scenario, accurate models of TANOS operations are strongly demanded.

Despite several recent [98]-[100] as well as earlier [101], [102] efforts to model program operations in TANOS and SONOS memories, a clear assessment of the physical mechanisms involved is still lacking. Majority of the models proposed in the literature agree on the physics

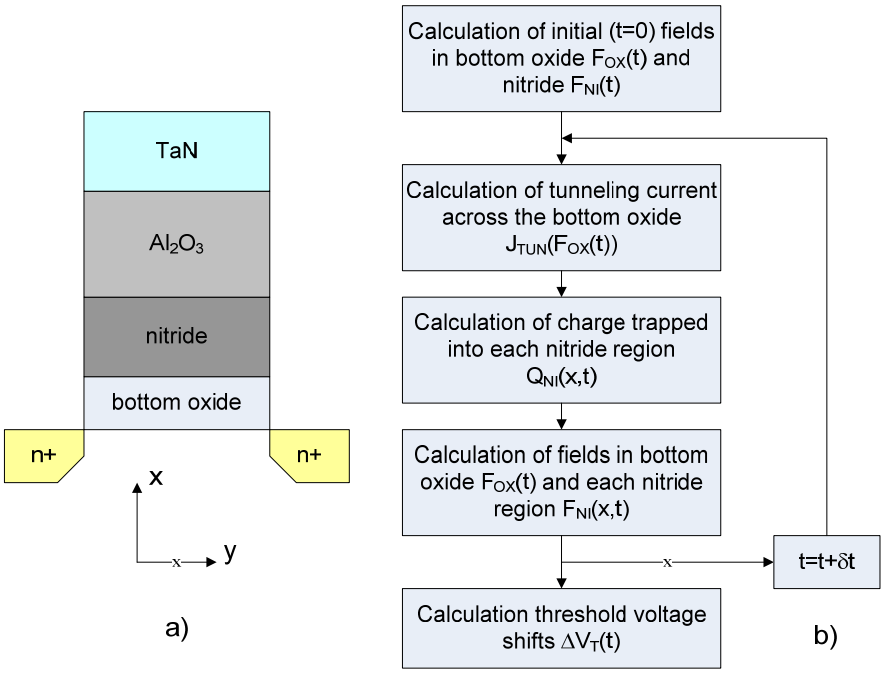

Figure VI.1: *(a) Cross section schematic of the TANOS memory devices used in this work. (b) Schematic flow chart of the model describing V_T shift during program operation.*

of the electron trapping process, which is usually explained according to the Shockley-Read-Hall theory [103], [104]; in addition, thermalization of the injected electrons and their transport in the silicon nitride layer have been also considered recently [100], [105]. On the contrary, the dominant mechanism of the electron detrapping process during programming, either Thermal Emission, or Poole-Frenkel (PF) conduction, or Trap-to-Band Tunneling, has not been yet unambiguously identified.

In this Chapter, a novel physics-based model describing TANOS program operations is presented and used to investigate the electron trapping/detrapping kinetics. By fitting modeling results to the experimental data we derive the trap characteristics (energy and spatial distribution), that are crucial for improving TANOS reliability and performance.

VI.1 Physics-Based Model

A schematic cross section of the TANOS memory device is shown in Fig. VI.1, together with the flow chart of the model we developed. In order to model the program transient of the TANOS device, the nitride layer is discretized in energy and space into a matrix of bins.

At each simulation step, the tunneling current flowing across the bottom oxide, J_{TUN}, is calculated using the model reported in [106], which considers a multi-phonon trap-assisted tunneling conduction mechanism, including random defect generation and charge

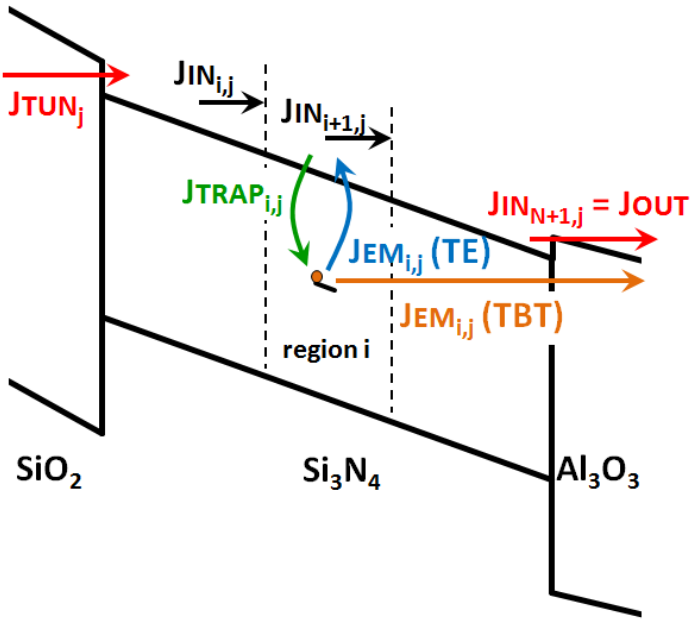

Figure VI.2: *Schematic representation of CB diagram and charge fluxes considered in the model. Index "i" and "j" refer to space and time discretization, respectively. J_{IN} is the current entering the nitride region, J_{TRAP} and J_{EM} are the charge fluxed related respectively to capture and emission processes, J_{TUN} and J_{OUT} are the currents entering and leaving the nitride at the SiO_2/Si_3N_4 and Si_3N_4/Al_2O_3 interfaces, respectively.*

quantization effects at the Si/SiO_2 interface. The electric fields in the bottom oxide, F_{OX}, in the nitride, F_{NI}, and in the alumina layer are computed solving the Poisson equation through the whole stack.

Once the electron current entering the nitride CB at the SiO_2/Si_3N_4 interface and the electric fields are known, the density of electrons moving in the CB of the trapping layer, n_F, and the density of electrons trapped into nitride defects, n_T, are calculated in every nitride region considering the current fluxes reported in Fig. VI.2.

The free electron density is computed by solving the system of linear equations describing current continuity and drift-diffusion (DD) in every nitride regions and at their interfaces

$$\frac{qL_{REG}\left(n_{F_{i,j}} - n_{F_{i,j-1}}\right)}{t_j - t_{j-1}} = J_{IN_{i,j}} - J_{IN_{i+1,j}} - J_{TRAP_{i,j}} + J_{EM_{i,j}} \tag{VI.1}$$

$$J_{IN_{i,j}} = q\mu\left(n_{F_{i,j}}F_{NI_{i,j}} + kT\frac{n_{F_{i,j}} - n_{F_{i+1,j}}}{L_{REG}}\right) \tag{VI.2}$$

q is the electron charge, L_{REG} is the length of each nitride region, $t_j - t_{j-1}$ is the simulation time step, μ is the constant electron mobility in the nitride (10^{-4} $Vm^{-2}s^{-1}$ [107]), k is the Boltzmann's constant, and T is the temperature. Writing Equations (VI.1) and (VI.2) for each

of the *N* nitride regions leads to a system of *2N* equations that can be solved considering the boundary conditions at the SiO_2/Si_3N_4 and Si_3N_4/Al_2O_3 interfaces, i.e. the electron current densities entering into the CB of the first nitride region at the SiO_2/Si_3N_4 interface and leaving the CB of the last nitride region through the alumina oxide at the Si_3N_4/Al_2O_3 interface

$$J_{IN_{1,j}} = J_{TUN} \tag{VI.3}$$

$$J_{IN_{N+1,j}} = J_{OUT} = qn_{F_{N,j}}\mu F_{NI_{N,j}} P_{OUT} \tag{VI.4}$$

P_{OUT} is the tunneling probability of the free electrons at the Si_3N_4/Al_2O_3 interface. J_{TUN} and J_{OUT} are calculated using the model reported in [106].

The charge trapped into each nitride region is calculated through

$$\frac{n_{T_{i,j}} - n_{T_{i,j-1}}}{t_j - t_{j-1}} = J_{TRAP_{i,j}} - J_{EM_{i,j}} \tag{VI.5}$$

$$J_{TRAP_{i,j}} = J_{IN_{i,j}}\sigma_T L_{REG} \int N_{T_{i,j}}(E)\left[1 - f_{T_{i,j}}(E)\right] dE \tag{VI.6}$$

$$J_{EM_{i,j}} = qf_E L_{REG} \int N_{T_{i,j}}(E)\, f_{T_{i,j}}(E) P_{EM_{i,j}}(E) dE \tag{VI.7}$$

σ_T is the capture cross section of nitride traps, N_T and f_T are the density and the occupation probability of traps inside every space and energy (E) nitride bin, f_E is the attempt-to-escape-frequency (10^9 Hz [98], [105]), P_{EM} is the trap emission probability computed including TE, field-induced trap energy barrier lowering, TAT and TBT contributions. Electron capture is governed by the SRH process [103], [104].

The final step is the calculation of the overall threshold voltage shift ΔV_T, which is obtained by summing the contributions of every nitride region

$$\Delta V_{T_j} = \sum_{k=1}^{N} \frac{qn_{T_{k,j}} L_{REG}}{C_{NI_k}} \tag{VI.8}$$

where $C_{NI\,k}$ is the capacitance between the *k*-th nitride region and the TaN gate.

A schematic cross section of the TANOS memory devices we used is shown in Fig. 1(a). The SiO_2 tunnel oxide was thermally grown on the (100) Si substrate, followed by the silicon nitride layer deposition. Different thicknesses of tunnel oxide, t_{OX}, and nitride trapping layer, t_{NI}, are considered, as shown in Table VI.1. Al_2O_3 was used as a blocking oxide. PVD TaN was then deposited to complete the gate stack.

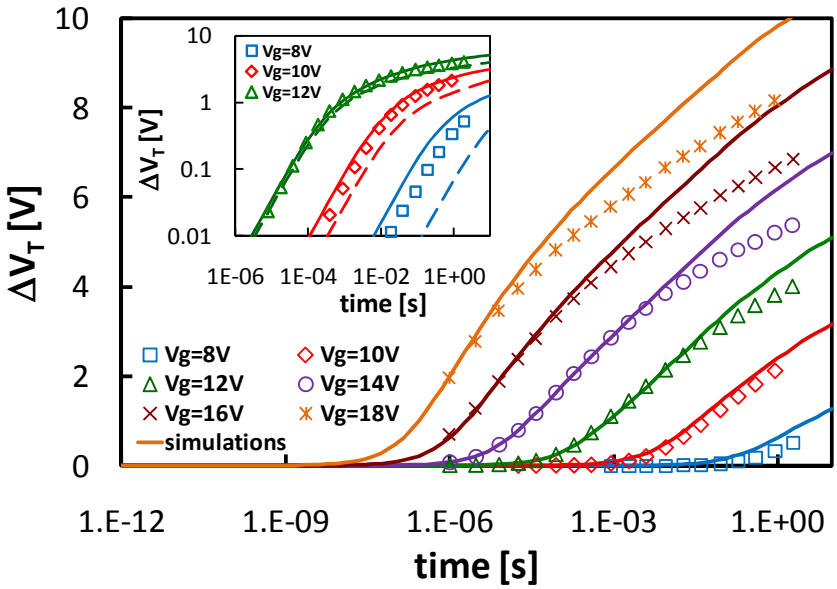

Figure VI.3: *Measurements (symbols) and simulations (lines) of V_T shifts in TANOS memory (sample C) when varying the gate voltage V_G. A uniform trap density across the nitride of $N_T=7.5\cdot10^{19}$ cm^{-3} is considered with $\sigma_T=7\cdot10^{-15}$ cm^2. Dashed lines in the inset depict simulations performed not considering TAT in the calculation of the program current density.*

Table VI.1: *Features of the TANOS memory devices used in this thesis.*

Sample	t_{OX} [nm]	t_{NI} [nm]	t_{AL} [nm]
A	3	5	11.5
B	3.5	5	11.5
C	4	5	11.5
D	4	8.7	11.5

VI.2 Electron Trapping Dynamics

In order to investigate trapping dynamics, we first performed simulations neglecting electron detrapping from occupied nitride traps considering a uniform trap density of $N_T=7.5\cdot10^{19}$ cm^{-3} across the nitride layer. Results for the sample C are shown in Fig. VI.3. The agreement between V_T measurements and simulations is rather accurate for all the program voltages considered, except at longer program times when the detrapping current (which we purposely neglected) comes into play.

Fig. VI.3 allows deriving two important insights in the physics of the electron trapping in the silicon nitride film. First, the electron capture probability is found to be independent from the electron energy and electric field, thus confirming that the trapping is a pure SRH process contrary to the conclusion in [100]. In fact, using a constant trap capture cross section ($\sigma_T=7\cdot10^{-15}$ cm^2) allows reproducing accurately the measured V_T shifts at low ΔV_T values

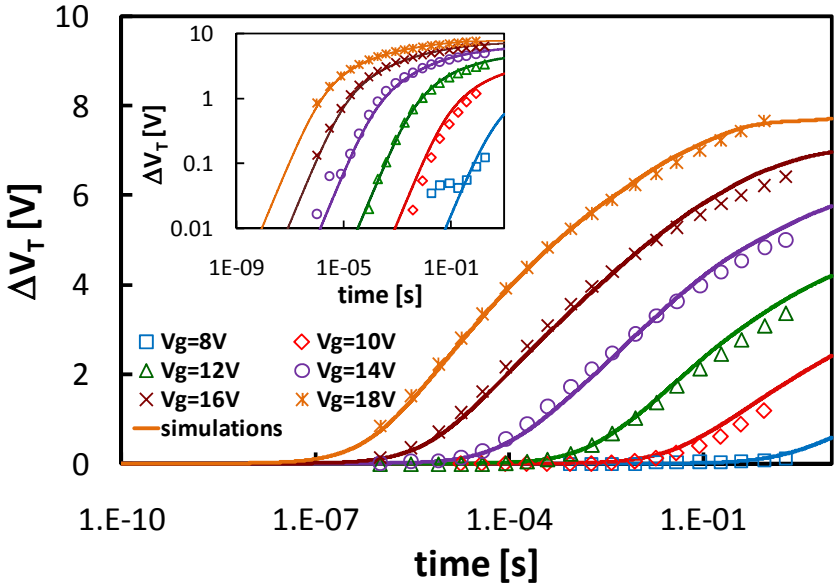

Figure VI.4: *Comparison between V_T shifts measured (symbols) and simulated (solid lines) for sample D. Simulations include both thermal and tunnel-based emission contributions.*

independently from the applied gate voltage V_G, i.e. from the electron energy and initial electric field.

Second, the accurate calculation of the injected current flow across the tunnel oxide is crucial for achieving a high modeling accuracy. In particular, the inset of Fig. VI.3 shows that the TAT contribution across the bottom oxide (oxide defect parameters are E_T=1.6-1.8eV, $N_T=5\cdot10^{16}$ cm^{-3}, $\sigma_T=10^{-14}$ cm^2) must be accounted for to reproduce V_T shifts at low fields, i.e. low V_G and high program times. When TAT contribution is neglected, the program current is underestimated and simulations do not agree with measurements, see dashed lines in the inset of Fig. VI.3.

Then we performed simulations including the processes of the electron emission from the traps to investigate the energy profile of nitride traps and the evolution of the trapped charge distribution during program. We included both thermally activated, such as PF and TE, and tunnel-based, such as TAT and TBT (from traps to nitride and Al_2O_3 conduction bands), physical mechanisms. For the sample D, results are shown in Fig. VI.4. An excellent agreement with the experimental data was obtained for all the program voltages in the total time range proving that the developed model reflects correctly the physics governing the TANOS program operation. Nitride defect parameters used in simulations agree with the values reported in the literature: E_T=1.9-2.7eV, $N_T=7.5\cdot10^{19}$ cm^{-3}, $\sigma_T=7\cdot10^{-15}$ cm^2 [105], [108], [109]. Uniform distributions of traps were considered for both space and energy. Using the same set of parameters we obtained an excellent agreement with the measurements on the samples A, B and C. Results for sample C are shown in Fig. VI.5.

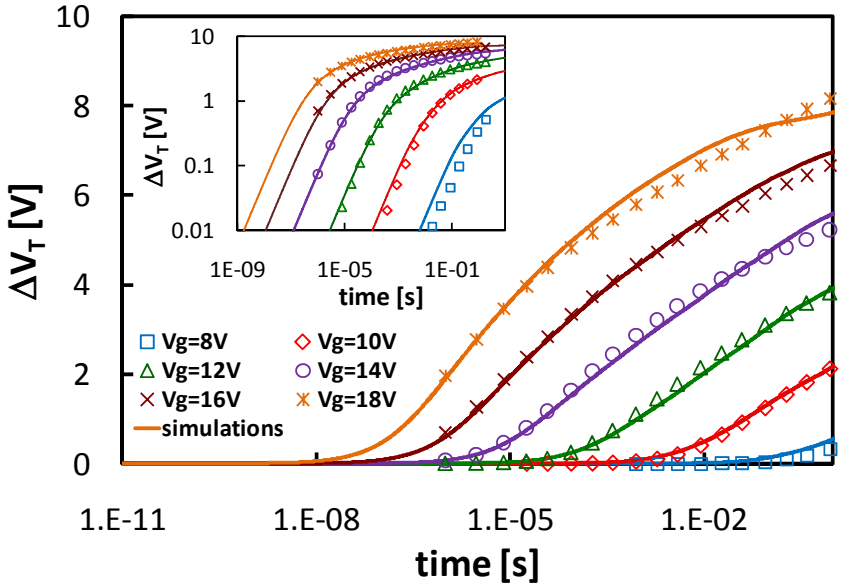

Figure VI.5: *Comparison between V_T shifts measured (symbols) and simulated (solid lines) for sample C. Simulations include both thermal and tunnel-based emission contributions.*

It is important to notice that the derived trap energy profile depends strongly on the considered detrapping mechanism. In fact, by considering only the thermal emission, we extracted E_T=1.2-1.8eV, which differs significantly from the value estimated including also TBT and TAT (E_T=1.9-2.7eV). This suggests that TE or PF mechanisms cannot be responsible for electron detrapping, as the shallower energy levels associated to them would lead to a fast saturation of the program V_T due to the very high TBT/TAT probabilities. This indicates that tunneling processes are the dominant electron detrapping mechanisms during TANOS program.

VI.3 Evolution of the Trapped Charge

Simulation results allow gaining important insights on the evolution of the trapped charge during the program operation. Fig. VI.6 shows the dependence of the charge centroid, C_{CN}, observed at the end of V_T program transient on the stack composition and program conditions.

As shown in the inset of Fig. VI.6, C_{CN} is almost constant during the program transient in agreement with [110]-[111], hence results shown in Fig. VI.6 are valid over the program operation time. In order to have a fair comparison between the data obtained on different stacks, C_{CN} is normalized with respect to the thickness of the nitride layer and it is plotted versus the equivalent field across the oxide. Noticeably, the charge centroid depends on the oxide field, and it is closer to the Si_3N_4/Al_2O_3 interface at low fields (i.e. low program

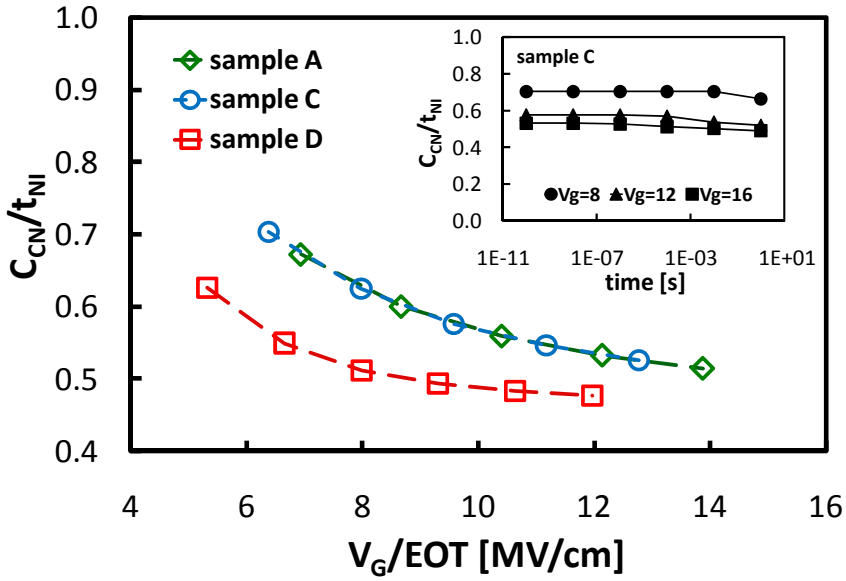

Figure VI.6: *Charge centroid C_{CN} normalized with respect to the thickness of the nitride layer as a function of the equivalent electric field across the tunnel oxide. The inset shows the evolution of the charge centroid with program time at different program voltages (sample C). EOT is the Equivalent Oxide Thickness of the stack, whereas V_G is the voltage applied during program. C_{CN}/t_{NI}=0 corresponds to the SiO_2/Si_3N_4 interface.*

voltages), contrarily to what would occur if the capture efficiency would reduce on increasing the electron energy.

With respect to the stack composition, the results in Fig. VI.6 provide interesting information about nitride charge distribution and dynamics. First, C_{CN} does not depend on the bottom oxide thickness. In fact, charge centroids observed in samples A and C (t_{OX}=3nm and 4nm, respectively) overlap completely. Second, C_{CN} depends on the thickness of the nitride layer. As shown in Fig. VI.6, the normalized centroid position moves to the oxide/nitride interface as the nitride thickness is increased, since the effect of electron accumulation at the Si_3N_4/Al_2O_3 interface reduces on increasing t_{NI} and electrons trap closer to the SiO_2/Si_3N_4 interface.

VI.4 Chapter Summary

In this Chapter a new physics-based drift-diffusion model for TANOS program transient has been presented. Experimental results measured on TANOS devices with different oxide and nitride thicknesses are reproduced with a great accuracy using an unique set of parameters, proving that the model catches correctly TANOS programming physics.

The model has been used to investigate electron trapping/detrapping dynamics. Trapping process is found to be independent from the energy of injected electrons, while electron de-trapping is dominated by tunnel processes.

The evolution of the nitride charge during program has been investigated considering the charge centroid as a figure of merit. It is found that the charge centroid is almost constant during the program transient and depends on the thickness of the nitride layer. These information are vital for the optimization of TANOS memory cells.

CHAPTER VII

Conclusions

THIS thesis presents part of the work carried out during the three years of author's research toward his Ph.D. in the XXI cycle, *Dottorato in Scienze dell'Ingegneria* doctorate course.

Author's research activity focused on the reliability and modeling of innovative nonvolatile memory devices and was motivated by the extensive research efforts made by both industry and academia to find a valid alternative to the mainstream Floating Gate technology, which is facing severe scaling challenges. Among the different solutions that are currently investigated, this work focused on the study of high-k based and charge-trapping memory devices like NROM and TANOS.

In the field of high-k materials, the research activity was devoted to the investigation of the feasibility of high-k based band-gap engineered barriers as tunnel dielectrics for future Flash technologies (Chapter III). To this purpose, a novel physics-based statistical Monte Carlo-like simulator was developed to model the leakage currents flowing through high-k dielectric stacks taking into account both tunneling and defect-assisted contributions. The simulation capabilities of the model were proved by reproducing leakage currents measured on large area capacitors with different SiO_2/high-k dielectric stacks. The excellent agreement between experiments and simulations allowed extracting the characteristics of the defects that are responsible for the undesired conduction through the high-k stack. Finally, the statistical capabilities of the model were used to verify at the array level if the theoretical retention improvements predicted by the introduction of BGE barriers as tunnel dielectric in FG devices do hold also when typical high-k defects are considered. Results indicated that symmetric Hf-based BGE barriers do not allow to satisfy usual retention requirements.

Noticeably, the MC leakage current simulator was also successfully used to investigate the breakdown of SiO_2/HfO_2 gate dielectric stacks in high-k/metal gate logic transistors [112].

It provided fundamental insights on the traps responsible for the overall degradation and breakdown of the metal/high-k gate stacks and on their evolution during dielectric degradation.

In the framework of charge-trapping memories, author's research activity focused on the modeling and reliability of NROM and TANOS devices. For what concerns NROM cells, the activity was devoted to the development of ad-hoc techniques to profile electron and hole charge distributions after both program and erase operations. Two I_D-V_{GS} based tools presented in Chapter IV were developed and used to derive compact formulas allowing calculating length and density of the electron charge distribution in programmed NROM devices. These formulas allowed monitoring the evolution of nitride charge distribution during program, thus providing a useful tool for the optimization of the program operation. As the tools presented in Chapter IV are sensitive only to the net charge above the channel, a new technique to profile hole distributions in erased NROM devices was developed and presented in Chapter V. The technique was used to monitor the charge evolution after program and erase operations, providing insights on device's reliability. First, it was demonstrated that in cycled cells the amount of electrons in the nitride portion above the channel increases, leading to a lower erase efficiency. This was found to be related to the shift of the program field peak into the channel due to holes injected above the junction during erase. Second, it was proven that the threshold voltage shift occurring in NROM cells left unbiased in the erased state is due to the lateral migration of trapped holes. The polarity dependence of the V_{TR} drift on the erase scheme adopted was also explained.

The author worked also on the development of a novel physics-based drift-diffusion model of program transients in TANOS memory devices (Chapter VI). The model was shown to reproduce with great accuracy experimental data measured on TANOS memories with different stacks, allowing investigating the physics of electron trapping/detrapping dynamics. It was found that the trapping process is independent from the energy of injected electrons. Second, tunneling processes were found to dominate electron detrapping during the program operation. The model was also used to gain important insights on the evolution of the nitride charge during program. It was found that the charge centroid is almost constant during the program transient and depends on the thickness of the nitride layer. These information are fundamental for the optimization of TANOS memory cells.

Bibliography

[1] F. Masuoka, M. Asano, H. Iwahashi, T. Komuro, and S. Tanaka, "A new flash E^2PROM cell using triple polysilicon technology," *IEDM Tech. Dig.*, pp. 464-467, 1984.

[2] Roger Barth, "Test Challenges beyond 2010," Global STC Conference (GSC), May 14-16 2007, Napa, CA.

[3] Y. P. Tsividis, *Operation and Modeling of the MOS Transistor*. New York: McGraw-Hill, 1987.

[4] D. Khang and S.M. Sze "A floating gate and its application to memory devices," *Bell Sys. Tech. J.*, vol. 46, p. 1288, 1967.

[5] D. Frohman-Bentchkowsky, "Memory behaviour in a floating gate avalanche injection MOS (FAMOS) structure," *Appl. Phys. Lett.*, vol.18, p.332, 1971.

[6] D. Frohman-Bentchkowsky, "A fully decoded 2048 bit electrically programmable MOS-ROM," *IEEE ISSCC Dig. Tech. Pap.*, p.152, 1980.

[7] H.A.R. Wegener, A.J. Lincoln, H.C. Pao, M.R. O'Connel and R.E. Oleksiak, "The variable threshold transistor, a new electrically alterable, non-destructive read-only storage device," *IEDM Tech. Dig.*, Washington, D.C., 1967.

[8] P. C. Chen, "Threshold-alterable Si-gate MOS devices," *IEEE Trans. Electron Devices*, vol. 24, no. 5, pp. 584-586, 1977.

[9] E. Suzuki, H. Hiraishi, K. Ishi, and Y. Hayashi, "A low voltage alterable EEPROM with metal-oxide-nitride-oxide-semiconductor (MONOS) structure," *IEEE Trans. Electron Devices*, vol. 30, no. 2, pp. 122-127, 1983.

[10] B. Eitan, P. Pavan, I. Bloom, E. Aloni, A. Frommer and D. Finzi, "Can NROM, a 2-bit trapping storage cell, give a real challenge to floating gate cells?," *Proc. SSDM*, Tokyo, Japan, pp. 522-524, Sept. 1999.

[11] M. Specht *et al.*, "Charge trapping memory structures with Al_2O_3 trapping dielectric for high-temperature applications," *Solid State Electronics*, vol. 49, no. 5, pp. 716-720, 2005.

[12] J. Buckley *et al.*, "In-depth investigation of Hf-based high-k dielectrics as storage layer of charge-trap NVMs," *IEDM Tech. Dig.*, 2006.

[13] T.-M. Pan and W.-W. Yeh, "High-Performance High-*k* Y_2O_3 SONOS-Type Flash Memory," *IEEE Trans. Electron Devices*, vol. 55, no. 9, pp. 2354-2360, 2008.

[14] C.H. Lee, K. I. Choi, M. K. Cho, Y. H. Song, K. C. Park and K. K., "A novel SONOS structure of SiO_2/SiN/Al_2O_3 with TaN metal gate for multi-giga bit flash memories," *IEDM Tech. Dig.*, pp. 613–616, 2003.

[15] P. Pavan, R. Bez, P. Olivo, E. Zanoni, "Flash memory cells – An overview", *Proc. of the IEEE*, vol. 85, N. 8, pp. 1248-1271, 1997.

[16] S. T. Wang, "On the I-V characteristics of floating-gate MOS transistors," *IEEE Trans. Electron Devices*, vol. 26, no. 9, pp. 1292-1294, 1979.

[17] C. Hu, "Lucky-electron model of channel hot electron emission," *IEDM Tech. Dig.*, pp.223-226, 1979.

[18] R. H. Fowler and L. Nordheim, "Electron emission in intense electric fields," *Proc. Roy. Soc.* (London), A119, pp. 173–181, 1928

[19] P. E. Cottrell, R. R. Trooutman, and T. H. Ning, "Hot-electron emission in n-channel IGFET's," *IEEE Trans. Electron Dev.*, vol. 26, no. 4, pp. 520-532, 1979.

[20] B. Eitan, and D. Frohman-Bentchkowsky, "Hot-Electron injection into the oxide in n-channel MOS devices," *IEEE Trans. Electron Dev.*, Vol. 28, no. 3, pp. 328-340, 1981.

[21] G. A. Baraff, "Distribution functions and ionization rates for hot-electrons in semiconductors," *Phys. Rev.*, vol. 128, no. 6, pp. 2507-1517, 1962.

[22] L. Esaki, "Long journey into tunneling," *Proc. of IEEE*, vol. 62, pp. 825-831, 1974

[23] J. Moll, *Physics of Semiconductors.* New York: McGraw-Hill, 1964.

[24] M. Lenziger and E.H. Snow, "Fowler-Nordheim tunneling into thermally grown SiO_2," *J. Appl. Phys.,* vol. 40, no. 1, pp. 278-283, 1969.

[25] B. Fulford. (2002, June 24). "Unsung hero". Forbes. [Online]. Available: http://www.forbes.com/global/2002/0624/030.htm

[26] International Technology Roadmap for Semiconductors (ITRS) 2008 Update, "Process Integration, Devices, and Structures"

[27] F. Masuoka, M. Momodomi, Y. Iwata, and R. Shirota, "New ultra high density EPROM and flash EEPROM with NAND structure cell," *IEDM Tech. Dig.*, pp. 552-555, 1987.

[28] R. Moazzami and C. Hu, "Stress induced current in thin silicon dioxide films," *IEDM Tech. Dig.*, pp. 139-142, 1992.

[29] D. Ielmini, A. S. Spinelli, A. L. Lacaita, and A. Modelli, "Statistical modeling of reliability and scaling projections for Flash memories," *IEDM Tech. Dig.*, pp.703-706, 2001.

[30] L. Larcher, and P. Pavan, "Statistical simulations for flash memory reliability analysis and prediction," *IEEE Trans. Electron Dev.*, vol. 51, no. 10, pp. 1636-1643, 2004.

[31] J. Lee *et al.*, "Effects of Floating-Gate Interference on NAND Flash Cell Operation," *IEEE Electron Device Lett.*, vol. 23, no.5, pp. 264-266, 2002.

[32] L. Larcher, A. Padovani, P. Pavan, P. Fantini, A.Calderoni, A. Mauri and A. Benvenuti, "Modeling NAND Flash memories for IC design," *IEEE Electron Device Lett.*, vol. 29, no.10, pp. 1152-1154, 2008.

[33] J. Evans and R. Womack, "An Experimental 512-bit nonvolatile memory with ferroelectric storage cell," *IEEE J. Sol. St. Circ.*, vol. 23, no. 5, p. 1171-1175, 1998.

[34] J.-H. Kim, D. J. Jung, Y. M. Kang, H. H. Kim, W. W. Jung, J. Y. Kang, E. S. Lee, H. Kim, J. Y. Jung, S. K. Kang, Y. K. Hong, S. Y. Kim, H. K. Koh, D. Y. Choi, J. H. Park, S. Y. Lee, H. S. Jeong and K. Kim, "A Highly Reliable FRAM (Ferroelectric Random Access Memory)," *Proc. of IRPS*, pp. 554-557, 2007.

[35] M. Lim *et al.*, "SBT-Based Ferroelectric FET for Nonvolatile Non-Destructive Read Out (NDRO) Memory Applications," *Integrated Ferroelectrics*, vol. 27, pp. 71- 80, 1999.

[36] S. Tehrani, J. Slaughter, M Deherrera, B. Engel, N. Rizzo, J. Salter, M. Durlam, R. Dave, J. Janesky, B. Butcher, K. Smith, and G. Grynkewich, "Magnetoresistive random access memory using magnetic tunnel junctions," *Proc. of the IEEE*, pp. 703-714, 2003.

[37] S. Lai and T. Lowrey, "OUM - A 180nm Nonvolatile Memory Cell Element Technology for Standalone and Embedded Applications," *IEDM Tech. Dig.*, pp. 36.5.1-4, 2001.

[38] F. Capasso, F. Beltram , R. J. Malik, and J. F. Walker, "New floating-gate AlGaAs/GaAs memory devices with graded-gap electron injector and long retention times," *IEEE Electron Device Lett.*, vol. 9, no. 10, pp. 377-379, 1988.

[39] K. K. Likharev, "Layered tunnel barriers for nonvolatile memory devices," *Appl. Phys. Lett.*, vol. 73, no. 15, pp. 2137-2139, October 1998.

[40] B. Govoreanu, P. Blomme, M. Rosmeulen, J. Van Houdt, and K. De Meyer, "VARIOT: A multilayer tunnel barrier concept for low-voltage nonvolatile memory devices," *IEEE Electron Device Lett.*, vol.24, no.2, pp. 99-101, Feb 2003.

[41] J. Robertson, "High dielectric constant gate oxides for metal oxide Si transistors," *Rep. Prog. Phy.*, vol. 69, p. 327-396, 2006.

[42] S. Verma, E. Pop, P. Kapur, K. Parat, and K. C. Saraswat, "Operational voltage reduction of flash memory using high-k composite tunnel barriers," *IEEE Electron Device Lett.*, vol. 29, no. 3, pp. 252-254, 2008.

[43] L. Larcher, P. Pavan, and B. Eitan, "On the physical mechanism of the NROM memory erase," *IEEE Trans. Electron Devices*, vol.51, no. 10, pp. 1593-1599, 2004.

[44] A. Shappir, E. Lusky, G. Choen, I. Bloom, M. Janai, and B. Eitan, "The two-bit NROM reliability," *IEEE Trans. Device Mater. Rel.*, vol.4, no. 4, pp. 397-403, 2004.

[45] B. Eitan, G. Cohen, A. Shappir, E. Lusky, A. Givant, M. Janai, I. Bloom, Y. Polansky, O. Dadashev, A. Lavan, R. Sahar and E. Maayan, "4-bit per cell NROM reliability," *IEDM Tech. Dig.*, pp. 539-542, 2005.

[46] E. Lusky, Y. Shacham-Diamand, I. Bloom, and B. Eitan, "Electron Retention Model for Localized Charge in Oxide-Nitride-Oxide (ONO) Dielectric," *IEEE Electron Device Lett*., vol. 23, no. 9, pp. 556-558, 2002.

[47] M. Janai, "Data retention, endurance, and acceleration factors of NROM devices," *Proc. of IRPS*, pp. 502–505, 2003.

[48] A. Shappir, Y. Shacham-Diamand, E. Lusky, I. Bloom, and B. Eitan, "Lateral charge transport in the nitride layer of the NROM nonvolatile memory device," *Microelectronic. Eng*., vol. 72, no. 1–4, pp. 426–33, 2004.

[49] A. Shappir, D. Levy, Y. Shacham-Diamand, E. Lusky, I. Bloom, and B. Eitan, "Spatial characterization of localized charge trapping and charge redistribution in the NROM device," *Solid-State Electron*., vol. 48, no. 9, pp. 1489–1495, 2004.

[50] C. H. Lee, S. H. Hur, Y. S. Shin, I. H. Choi, D. G. Park, and K. Kim, "A novel structure of SiO_2/SiN/High-k dielectrics, Al_2O_3 for SONOS type flash memory," *Proc. of SSDM*, p.162, 2002.

[51] M. H. White, D. A. Adams, and J. Bu, "On the go with SONOS," *IEEE Circuits Devices Mag.*, vol. 16, no. 4, pp. 22-31, 2000.

[52] M. T. Bohr, R. S. Chau, T. Ghani, and K. Mistry, "The high-k solution," *IEEE Spectrum*, vol. 44, no. 10, pp. 29-35, 2007.

[53] L. Larcher, and P. Pavan, "Statistical simulations to inspect and predict data retention and program disturbs in Flash memories," *IEDM Tech. Dig.*, pp. 165-168, 2003.

[54] A. Padovani, L. Larcher, A. Chimenton, and P. Pavan, "Monte-Carlo simulations of Flash memory array retention" in *Proc. of IEEE VLSI-TSA*, pp. 156-157, 2007.

[55] L. Larcher, P. Pavan, F. Pellizzer, and G. Ghidini, "A new model of gate capacitance as a simple tool to extract model parameters," *IEEE Trans. Electron Dev.*, vol. 48, no. 5, pp. 935-945, 2001.

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

[57] C. D. Young *et al*., "Electron trap generation in high-k gate stacks by constant voltage stress," *IEEE Trans. Device Mater. Rel.*, vol. 6, no. 2, pp. 123-131, 2006.

[58] A. Schenk and H. Hermann, "A new model for long term charge loss in EPROMs," *Proc. of SSDM*, pp. 494-496, 1994.

[59] M. Hermann and A. Schenk, "Field and high temperature dependence on the long term charge loss in erasable programmable read only memories: Measurements and Modeling," *J. Appl. Phys.*, vol. 77, no. 9, pp. 4522-4540, 1995.

[60] L. Larcher, "Statistical simulation of leakage currents in MOS and flash memory devices with a new multiphonon trap-assisted tunneling model," *IEEE Trans. Electron Dev.*, vol. 50, no. 5, pp. 1246-1253, 2003.

[61] Shin-ichi Takagi, Naoki Yasuda, and Akira Toriumi, "A new I-V model for stress-induced leakage current including inelastic tunneling", *IEEE Trans. Electron Dev.,* vol. 46, no. 2, pp. 348-354, 1999.

[62] L. Larcher, A. Paccagnella, and G. Ghidini, "A new model of Stress Induced Leakage Current in gate oxides", *IEEE Trans. Electron Dev.*, vol. 48, no. 2, pp. 285-288, 2001.

[63] D. Ielmini, A. S. Spinelli, M. A. Rigamonti, A. L. Lacaita, "Modeling of SILC based on electron and hole tunneling. II. Steady-state," *IEEE Trans. Electron Dev.*, vol. 47, no. 6, pp. 1266-1272, 2000.

[64] B. Riccò, G. Gozzi, and M. Lanzoni, "Modeling and simulation of Stress-Induced Leakage Current in ultrathin SiO_2 films," *IEEE Trans. Electron Dev.*, vol. 45, no. 7, pp. 1554-1560, 1998.

[65] H.-C. Wen *et al*., "Comparison of effective work function extraction methods using capacitance and current measurement techniques," *IEEE Electron Device Lett.*, vol. 27, no. 7, pp. 598-601, 2006.

[66] Y. Kamimuta, M. Koike, T. Ino, M. Suzuki, M. Koyama, Y. T. and A. Nishiyama, "Determination of Band Alignment of Hafnium Silicon Oxynitride/Silicon (HfSiON/Si) Structures using Electron Spectroscopy," *J. J. Appl. Phys.*, vol. 44, no. 3, pp. 1301-1305, 2005.

[67] M. Koike *et al.*, "Dielectric properties of noncrystalline HfSiON," *Phys. Rev. B*, vol. 73(125123), 2006.

[68] A. Kerber *et al.*, "Charge trapping and dielectric reliability of SiO2-Al2O3 gate stacks with TiN electrodes," *IEEE Trans. Electron Dev.*, vol. 50, no. 5, pp. 1261-1268, 2003.

[69] S. Meng, C. Basceri, B. W. Busch, G. Derderian, and G. Sandhu, "Leakage mechanisms and dielectric properties of Al2O3/TiN-based metal-insulator-metal capacitors," *Appl. Phys. Lett.*, vol. 83, no. 21, pp. 4429-4431, 2003.

[70] D. Heh *et al*., "Spatial distributions of trapping centers in SiO_2/HfO_2 gate stack," *IEEE Trans. Electron Devices*, vol. 54, no. 6, pp. 1338-1345, 2007.

[71] J.-P. Han *et al.*, "Energy distribution of interface traps in High-k gated MOSFETs" *VLSI Symp. Tech. Dig*., pp. 161-162, 2003.

[72] International Technology Roadmap for Semiconductors (ITRS) 2005, "Emerging Research Devices,"

[73] M. Janai and B. Eitan "Data retention, endurance and acceleration factors of NROM devices," in *Proc. of IRPS*, pp. 502-505, 2003.

[74] M. Janai, B. Eitan, A. Shappir, e. Lusky, I. Bloom, and G. Choen, "Data retention reliability model of NROM nonvolatile memory products," *IEEE Trans. Device Mater. Rel.*, vol.4, pp. 404-415, Sept. 2004.

[75] E. Lusky, Y. Shacham-Diamand, I. Bloom, and B. Eitan, "Electrons retention model for localized charge in Oxide-Nitride-Oxide (ONO) dielectric," *IEEE Electron Device Lett.*, vol.23, no. 9, pp. 556-558, 2002.

[76] M. Rosmeulen, L. Breuil, M. Lorenzini, L. Haspeslagh, J. Van Houdt, and D. De Meyer, "Characterization of the spatial charge distribution in local charge-trapping memory devices using the charge-pumping technique," *Solid-State Electron.*, Vol. 48, pp. 1525-1530, 2004.

[77] P. B. Kumar, P. R. Nair, R. Sharma, S. Kamohara, and S, Mahapatra, "Lateral profiling of trapped charge in SONOS Flash EEPROMs programmed using CHE injection," *IEEE Trans. Electron Devices*, vol. 53, no. 4, pp. 698- 705, 2006.

[78] A. Furnémont, M. Rosmeulen, J. Van Houdt, H. Maes, K. De Meyer, "Cycling behaviour of nitride charge profile in NROM-type memory cells," in *Proc. of the 21st Non-Volatile Semiconductor Memory Workshop*, pp. 66-67, Feb. 2006.

[79] E. Lusky, Y. Shacham-Diamand, I. Bloom, and B. Eitan, "Characterization of channel hot electron injection by the subthreshold slope of NROM™ device," *IEEE Electron Device Lett.*, vol. 22, no. 11, pp. 556-558, 2001.

[80] L. Larcher, G. Verzellesi, P. Pavan, E. Lusky, I. Bloom, and B. Eitan, "Impact of programming charge distribution on threshold voltage and subthreshold slope of NROM memory cells," *IEEE Trans. Electron Devices*, vol. 49, no. 11, pp. 1939-1946, 2002.

[81] A. Shappir, Y. Shacham-Diamand, E. Lusky, I. Bloom, and B. Eitan, "Subthreshold slope degradation model for localized-charge-trapping based on non-volatile memory devices," *Solid State Electron.*, vol. 47, pp. 937-941, 2003.

[82] A. Shappir, D. Levi, Y. Shacham-Diamand, E. Lusky, I. Bloom, and B. Eitan, "Spatial characterization of localized charge trapping and charge redistribution in the NROM device," *Solid State Electron.*, vol. 48, pp. 1489-1495, 2004.

[83] S. Mahapatra, S. Shukuri, and J. Bude, "CHISEL Flash EEPROM-Part I: Performance and Scaling," *IEEE Trans. Electron Devices*, vol.49, no. 7, pp. 1296-1301, 2002.

[84] C. Lombardi, S. Manzini, A. Saporito, and M. Vanzini, "A physically based mobility model for numerical simulation of nonplanar devices," *IEEE Trans. On Computer-Aided Design*, vol. 7, no. 11, pp. 1164-1171, 1988.

[85] L. Avital, A. Padovani, L. Larcher, I. Bloom, R. Arie, P. Pavan, and B. Eitan, "Temperature Monitor: a New Tool to Profile Charge Distribution in NROM™ Memory Devices," *Proc. of IRPS*, pp. 534-540, 2006.

[86] M.-Y. Liu, Y.-W. Chang, N.-K. Zous, I. Yang, T.-C. Lu, T. Wang, W. Ting, J. Cu, and C.-Y. Lu "Temperature Effect on Read Current in a Two-Bit Nitride-Based Trapping Storage Flash EEPROM Cell," *IEEE Electron Device Lett.*, vol. 25, no. 7, pp.495-497, 2004.

[87] A. Padovani, L. Larcher, and P. Pavan "Profiling charge distributions in NROM devices," *Proc. PRIME 2006*, pp. 69-72, 2006.

[88] Grabe, M., *Measurement Uncertainties in Science and Technology*, Springer, April 2005.

[89] A. Padovani, L. Larcher, P. Pavan, L. Avital, I. Bloon, and B. Eitan, "I_D-V_{GS} Based Tools to Profile Charge Distributions on NROM™ Memory Devices," accepted for publication on *IEEE Trans. Device Mater. Rel.* vol. 7, no. 1, 2007.

[90] A. Furnémont, M. Rosmeulen, K. van der Zanden, J. Van Houdt, K. De Meyer, H. Maes, "Physical modelling of retention in localized trapping nitride memory devices," in *IEDM Tech. Digest*, 2006, 2006.

[91] A. Padovani, L. Larcher, P. Pavan, "Hole Distributions in NROM Devices: Profiling Technique and Correlation to Memory Retention," *Proc. of IRPS*, pp. 654-655, 2007.

[92] J. J. Liou, "Modeling the Tunneling Current in Reverse-Biased p/n Junctions", *Solid State Electron.*, Vol. 33, No. 7, pp. 971-972, 1990.

[93] S. Manzini, F. Volonté, "Charge transport and trapping in silicon nitride-silicon dioxide dielectric double layers," *J. Appl. Phys.*, Vol. 58, No. 11, pp. 4300-4306, 1985.

[94] K. A. Nasyrov, V. A. Gritsenko, M. K. Kim, H. S. Chae, S. D. Chae, W. I. Ryu, J. H. Sok, J.-W. Lee, and B. M. Kim, "Charge Transport Mechanism in Metal–Nitride–Oxide–Silicon Structures," *IEEE Electron Device Lett.*, Vol. 23, No. 6, pp. 336-338, 2002.

[95] K. A. Nasyrov, V. A. Gritsenko, Yu. N. Novikov, E.-H. Lee, S. Y. Yoon, C. W. Kim, "Two-bands charge transport in silicon nitride due to phonon-assisted trap ionization," *J. Appl. Phys.*, Vol. 96, No. 8, pp. 4293-4296, 2004.

[96] W. J. Tsai, N. K. Zous, C. J. Liu, C. C. Liu, C. H. Chen, T. Wang, S. Pan, and C.-Y. Lu, "Data Retention Behaviour of a SONOS Type Two-Bit Storage Flash Memory Cell," in *IEDM Tech. Digest*, 2001, pp. 719-722.

[97] Y. Park *et al.*, "Highly manufacturable 32Gb multi-level NAND Flash memory with 0.0098 μm^2 cell size using TANOS (Si - Oxide - Al_2O_3 - TaN) cell technology," *IEDM Tech. Dig.*, pp. 29-32, 2006.

[98] A. Paul, Ch. Sridhar, S. Gedam and S. Mahapatra, "Comprehensive simulation of program, erase and retention in charge trapping Flash memories," *IEDM Tech. Dig.*, pp.439-442, 2006.

[99] C. H. Lee *et al.* "Numerical simulation of programming transient behavior in charge trapping storage memory," *NVSMW Tech. Dig.*, pp. 109-110, 2008.

[100] A. Mauri *et al*. "A new physics-based model for TANOS memories program/erase," *IEDM Tech. Dig.*, pp. 555-558, 2008.

[101] F. R. Libsch and M. White, "Charge transport and storage of low programming voltage SONOS/MONOS memory devices," *Solid-State Electron.*, vol. 33, no. 1, pp. 105-126, 1990.

[102] P. J. McWhorter, S. L. Miller and T. A. Dellin, *J. Appl. Phys.*, vol. 68, no. 4, 1902-1909, 1990.

[103] W. Shockley and W. T. Read, "Statistics of the recombination of holes and electrons," *Phys. Rev.*, vol. 87, no. 5, pp. 835-842, 1952.

[104] R. N. Hall, "Electron-hole recombination in Germanium," *Physical Rev.*, vol. 87, p. 387, 1952.

[105] E. Vianello *et al.*, "Impact of the charge transport in the conduction band of the retention of Si-Nitride based memories," *ESSDERC Tech. Dig.*, 107-110, 2008.

[106] A. Padovani *et al.*, "Statistical modeling of leakage currents through SiO2/High-k dielectrics stacks for non-volatile Memory applications," *IRPS Tech. Dig.*, pp. 616-620, 2008.

[107] I. Ay and H. Tolunay, "Steady-state and transient photoconductivity in hydrogenated amorphous silicon nitride films," *Solar Energy Mat. & Solar Cells*, vol. 80, pp. 209-216, 2003.

[108] J. Robertson and M. J. Powell, "Gap states in silicon nitride," *Appl. Phys. Lett.*, no. 44, vol. 4, pp. 415-417, 1984.

[109] E. Lusky, Y. Shacham-Diamand, A. Shapir, I. Bloom, and B. Eitan, "Traps spectroscopy of the Si_3Ni_4 layer using localized charge-trapping nonvolatile memory device," *Appl. Phys. Lett.*, vol. 85, no. 4, pp. 669-671, 2004.

[110] A. Arreghini *et al*., "Experimental characterization of the vertical position of the trapped charge in Si Nitride-based nonvolatile memory cells," *IEEE Trans. Electron Devices*, vol. 55, no. 5, pp. 1211-1218, 2008.

[111] H.-T. Lue, P.-Y. Du, S.-Y. Wang, K.-Y. Hsieh, R. Liu, and C.-Y. Lu, "A study of Gate-Sensing and Channel-Sensing (GSCS) transient analysis method—part I: fundamental theory and applications to study of the trapped charge vertical location and capture efficiency of SONOS-type devices," *IEEE Trans. Electron Devices*, vol. 55, no. 8, pp. 2218-2227, 2008.

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

Author's Publications

Here follows the index of publications by the author.

Journals

[J1] A. Padovani, L. Larcher, P. Pavan, L. Avital, I. Bloom, and B. Eitan, "Id-Vgs Based Tools to Profile Charge Distributions on NROM Memory Devices," *IEEE Transactions on Device and Materials Reliability*, vol.7, no.1, pp.97-104, March 2007.

[J2] (Invited) A. Padovani, L. Larcher, A. Chimenton, P. Pavan, P. Olivo, "Dielectric Reliability for Future Logic and Non-Volatile Memory Applications: a Statistical Simulation Analysis Approach," *ECS Transactions - ULSI vs. TFT Conference*, vol.8, no.1, pp.237-242, July 2007.

[J3] A. Padovani, L. Larcher, P. Pavan, "Hole Distributions in Erased NROM Devices: Profiling Method and Effects on Reliability," *IEEE Transactions on Electron Devices*, vol.55, no.1, pp.343-348, January 2008.

[J4] L. Larcher, A. Padovani, P. Pavan, P. Fantini, A.Calderoni, A. Mauri and A. Benvenuti, "Modeling NAND Flash memories for IC design," *IEEE Electron Device Letters*, vol.29, no.10, pp.1152-1154, October 2008.

[J5] L. Larcher, P. Pavan, A. Padovani and G. Ghidini, "A technique to extract high-k IPD stack layer thicknesses from C-V measurements," submitted for publication on *IEEE Electron Device Letters*.

[J6] A. Padovani, L. Larcher, D. Heh, and G. Bersuker, " Modeling TANOS Memory Program Transients to Investigate Charge Trapping Dynamics," submitted for publication on *IEEE Electron Device Letters*.

Conference Proceedings

[C1] L. Avital, A. Padovani, L. Larcher, I. Bloom, R. Arie, P. Pavan, and B. Eitan, "Temperature Monitor: a New Tool to Profile Charge Distribution in NROM Memory Devices," *IEEE International Reliability Physics Symposium*, San Jose, California, 2006, pp. 534-540.

[C2] A. Padovani, L. Larcher, and P. Pavan, “Profiling charge distributions in NROM devices,” *Ph.D. Research in Microelectronics and Electronics*, Otranto, Jun. 2006, pp. 69-72.

[C3] A. Padovani, L. Larcher, P. Pavan, “Hole Distributions in NROM Devices: Profiling Technique and Correlation to Memory Retention,” *IEEE International Reliability Physics Symposium*, Phoenix, Arizona, 2007, pp. 654-655.

[C4] A. Padovani, L. Larcher, A. Chimenton, P. Pavan, “Monte-Carlo Simulations of Flash Memory Array Retention,” *IEEE International Symposium on VLSI Technology, Systems and Applications (VLSI-TSA)*, Hsinchu, Taiwan, 2007, pp. 156-157.

[C5] A. Padovani, A. Chimenton, P. Olivo, P. Fantini, L. Vendrame, and S. Mennillo, “Statistical Methodologies for Integrated Circuits Design,” *Ph.D. Research in Microelectronics and Electronics*, in Bordeaux, France, 2-5 July 2007, pp. 277-280.

[C6] L. Larcher, A. Padovani, I. Rimmaudo, P. Pavan, A. Calderoni, G. Molteni, F. Gattel, and P. Fantini, “Modeling NAND Flash memories for circuit simulations,” *International Conference on Simulations of Semiconductor Processes and Devices*, Austria, 25-27 September, 2007, pp. 293-296.

[C7] A. Padovani, L. Larcher, S. Verma, P. Pavan, P. Majhi, P. Kapur, K. Parat, G. Bersuker, and K. Saraswat, "Feasibility of SiO2/Al2O3 Tunnel Dielectric for Future Flash Memories Generations," *International Conference on ULtimate Integration on Silicon (ULIS)*, Italy, 12-14 March, 2008, pp. 111-114.

[C8] A. Padovani, L. Larcher, S. Verma, P. Pavan, P. Majhi, P. Kapur, K. Parat, G. Bersuker, and K. Saraswat, "Statistical modeling of leakage currents through SiO2/high-k dielectric stacks for non-volatile memory applications," *IEEE International Reliability Physics Symposium*, Phoenix, Arizona, 2008, pp. 616-620.

[C9] G. Puzzilli, F. Irrera, P. Pavan, L. Larcher, A. Arya, V. Della Marca, A. Padovani, and A. Pirovano, "On the RESET-SET Transition in Phase Change Memories," *European Solid-State Device Research Conference*, Edinburgh, Scotland, 15-19 September 2008, pp.158-161.

[C10] G. Bersuker, D. Heh, C. Young, H. Park, P. Khanal, L. Larcher, A. Padovani, P. Lenahan, J. Ryan, B. H. Lee, H. Tseng, and R. Jammy, "Breakdown in the metal/high-k gate stack: identifying “weak link” in the multilayer dielectric," *IEEE International Electron Devices Meeting*, San Francisco, California, 15-17 December, 2008, pp.791-794.

[C11] S. Verma, G. Bersuker, D. C. Gilmer, A. Padovani, H. Park, A. Nainani, D. Heh, J. Huang, J. Jiang, K. Parat, P. D. Kirsch, L. Larcher, H.-H. Tseng, K. C. Saraswat and R. Jammy, "A Novel Fluorine Incorporated Band Engineered (BE) Tunnel (SiO2/HfSiO/SiO2) TANOS with excellent Program/Erase & Endurance to 10E5 cycles," accepted at the *IEEE International Memory Workshop*, to be held in Monterey, California, 10-14 May, 2009.