Characterization and modeling of advanced charge trapping non volatile memories

V. Della Marca

Université d'Aix-Marseille

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE D'AIX-MARSEILLE

Spécialité MICRO ET NANOELECTRONIQUE

dans le cadre de **l'Ecole Doctorale :** Sciences pour l'Ingénieur : Mécanique, Physique, Micro et Nanoélectronique

Préparée au sein de la société **STMicroelectronics à Rousset** en collaboration avec l'équipe

Mémoires de l'**Institut Matériaux Microélectronique Nanosciences de Provence (IM2NP)**

et le **CEA-Leti à Grenoble**

CHARACTERIZATION AND MODELING OF ADVANCED CHARGE TRAPPING NON VOLATILE MEMORIES

par

Vincenzo Della Marca

Directeur de thèse : **Frédéric LALANDE**

A soutenir publiquement le 24/06/2013 devant le jury composé de:

Frédéric LALANDE	Professeur, Université d'Aix-Marseille	Directeur
Paolo PAVAN	Professeur, Università di Modena e Reggio Emilia	Rapporteur
Laurent BREUIL	Docteur, IMEC - Leuven	Rapporteur
Gabriel MOLAS	Docteur, Cea-Leti (Minatec) Grenoble	Examinateur
Jérémy POSTEL-PELLERIN	Maitre de conférences, Université d'Aix-Marseille	Examinateur
Gilles REIMBOLD	HDR, CEA-Leti, Université de Grenoble	Examinateur
Pascal MASSON	Professeur, Université de Nice Sophia Antipolis	Examinateur
Jean-Luc OGIER	Docteur, ST Microelectronics, Rousset	Examinateur

To myself

Acknowledgments

First and foremost I want to thank my industrial chef in STMicroelectronics (Rousset) Jean-Luc Ogier. He has taught me (like a second mother). I appreciate all his contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating. Not less I want to thank my academic advisor Frédéric Lalande and his collaborator Jérémy Postel-Pellerin. It has been an honor to be their Ph.D. student. They introduced me at the university of Marseille and they gave me the possibility to use innovative equipments for my researches. Equally Gabriel Molas that supervised me at CEA-Leti (Grenoble) and introduced me to Lia Masoero, together we reached important results concerning our researches in a funny atmosphere.

I am especially grateful for the support of Silicon Nanocrystal team: Philippe Boivin, Antonello Scanni, Nando Basile, Olivier Pizzuto and in particular Julien Amouroux. He was Ph.D. student as me and expert in beer tasting. I would also like to acknowledge Laurent Lopez and Arnaud Regnier for the fruitful discussion and paper collaborations. Moreover, all the Electrical Characterization Group of ST: Guillaume Just, Lorin Martin, Olivier Paulet, Lionel Bertorello, Yohan Joly, Luc Baron, Marco Mantelli, Patrick Poire, Marion Carmona, Jean-Sebastian Culoma and Benjamin Rebouffaut.

Ellen Blanchet helped me for the thesis preparation in a satisfactory English :-).

My parents Anna and Michele, finally they learned to use Skype in order to encourage me always, in spite of the distance.

I want to thanks all my Italian friends and the people that I met in France very useful to develop my French language and not only this. Furthermore, my time in Aix-en-Provence was made enjoyable in large part due to Dario, Diego and Stefania that became a part of my life and in particular, for the parties and special nights.

Elisabetta and Dave I cannot explain their importance in my life and during the permanence in France, since the first day….they know!

Finally the last person that I met here, Elisabeth. She is *tunneling* my life as the more energetic electron present in nature….BABOOOM!

Contents

General introduction

Walking down the street, inside an airport or a university, it is impossible not to notice some people speaking or sending messages with their smartphones, others are painting a picture on their tablets, and all this is happening while we are transferring the data of our research from a smart card to a laptop. The wish to communicate and to keep all the information in our pocket, has lead to the development of embedded and portable device technology. Suddenly, with the coming of social networks, we need to exchange comments, articles, pictures, movies and all other types of data with the rest of the word, regardless of our position. In a "touch" we can access the information that always needs to be stored in larger quantity; not one single bit of what belongs to us must be lost and the devices must be extremely reliable and efficient. In this scenario the microelectronics industry is continuously evolving and never ceases to astonish. As a consequence, over the last decade, the market of semiconductor integrated circuits (IC) for embedded applications has exploded too. The request of customers commands the market of low energy consumption portable devices. Particular attention is paid to Flash memories that actually represent the most important media to store each type of data. Depending on application characteristics, different architectures and devices have been developed over the last few years in order to satisfy all the needs of customers. Size scaling, faster access time and lower energy consumption have been the three pillars of scientific research in micro and nano electronic devices over the last few years.

Starting from these philosophical considerations we performed an experimental study on silicon nanocrystal memory that represents one of most attractive solutions to replace the standard Flash floating gate device. The aim of this thesis is to understand the physical mechanisms that govern the silicon nanocrystal cell behavior, to optimize the device architecture and to compare the results found with the standard Flash to verify performance improvement.

In the first chapter, we will present the economic context, the evolution and the working of EEPROM-Flash memories. Then, a detailed description of the technology, the functioning and their scaling limits will be provided. Finally we will expose the possible solutions to overcome these problems and the thesis framework.

The second chapter will present the experimental setup and the methods of characterization used to measure the performances of silicon nanocrystal memory cell. Moreover the impact of relevant technological parameters such as: the nature of nanocrystals, silicon nitride presence, channel doping dose and tunnel oxide thickness, will be analyzed. A memory cell stack optimization is also proposed to match the Flash floating gate memory performance.

In the third chapter the impact of main technological parameters on silicon memory cell reliability (endurance and data retention) is studied. The performance of silicon nanocrystal memories for applications functioning within a wide range of temperatures [-40°C; 150°C] is also evaluated reaching for the first time a 1Mcycles endurance with a 4V programming window. Finally the proposed optimized cell is compared to the standard Flash floating gate.

Chapter four describes a new dynamic technique of measurement for the drain current consumption during the hot carrier injection. This enables the cell energy consumption to be evaluated when a programming operation occurs. This method is applied for the first time to the floating gate and silicon nanocrystals memory devices. A study on programming scheme and the impact of technological parameter is presented in this chapter. In addition the silicon nanocrystal and floating gate cells are compared. Finally we demonstrate that is possible to reach a sub-nanojoule energy consumption saving a 4V programming window.

Finally in the chapter five the conclusion of this work will be analyzed in order to highlight the main experimental results. Moreover the basis for a future work will be presented.

Chapter 1 – Flash memories: an overview

1.1 Introduction

The aim of this first chapter is to present the economic context, the role and the evolution of non-volatile memories. In this context we will present the Flash floating gate device and the physical mechanisms used to transfer electric charge from and into the floating gate. Then the limits of this device and the existing solutions to overcome them will be introduced. In particular, we will focus on the silicon nanocrystal memory that represents the object of this thesis.

1.2 The industry of semiconductor memories

1.2.1 The market of non-volatile memories

Over the last decade, the market of non volatile memories has been boosted, driven by the increasing number of portable devices (figure 1.1). All the applications require higher and higher performance such as: high density, low power consumption, short access time, low costs, and so on [Changhyun '06]. This is why the business of Flash memories gained market segments at the expense of other types of memory (figure 1.2). Although the market is growing continuously, the price of memory device is decreasing (figure 1.3).

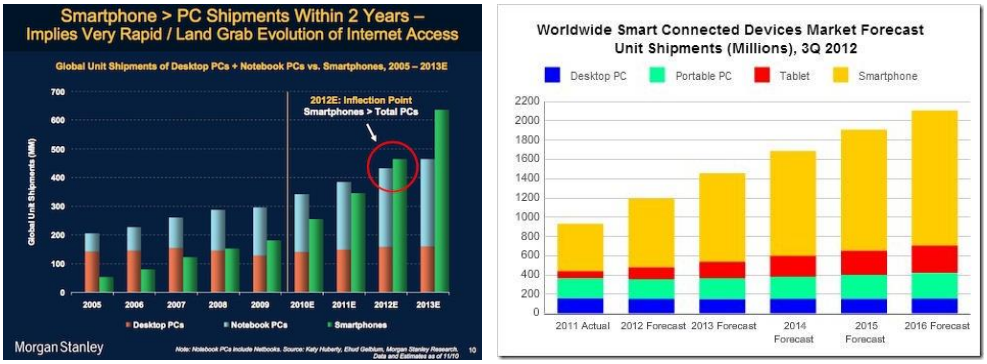
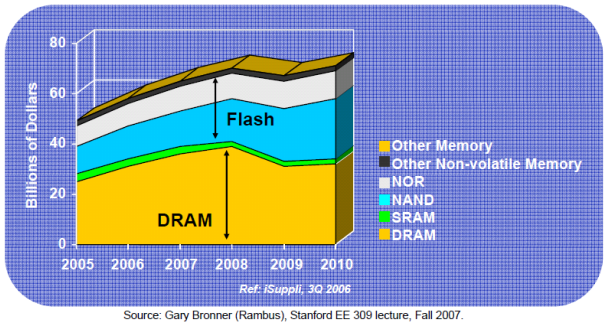


Figure 1. 1. Evolution and forecast of portable devices market (source: muniwireless.com and trak.in)



Source: Gary Bronner (Rambus), Stanford EE 309 lecture, Fall 2007.

Figure 1. 2Memory market: Flash and DRAM dominate[Philip Wong '08].

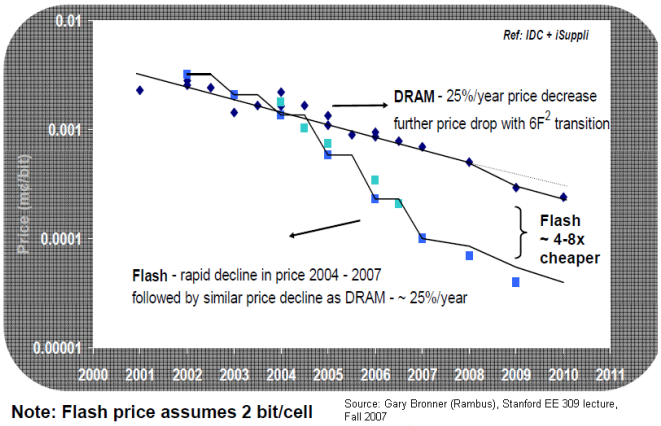


Figure 1. 3. DRAM and Flash price outlook [Philip Wong '08].

As the memory market enters the Gigabit and GHz range with consumers demanding ever better performance and more diversified applications, new types of devices are being developed in order to keep up with the scaling requirements for cost reduction. In this scenario, memories play an important role. The "ideal" memory should be a memory that retains the stored information even when it is not powered (non volatile) with a high integration density, that can be infinitely written/re-written (infinite endurance), with ultra high program/erase/read operations, and a zero energy consumption. Because the "ideal" device does not exist, different types of memories have been studied in order to develop one or more of these properties according to their final application [Masoero '12a] (figure 1.4). In the next section, the most important semiconductor memories will be summarized.

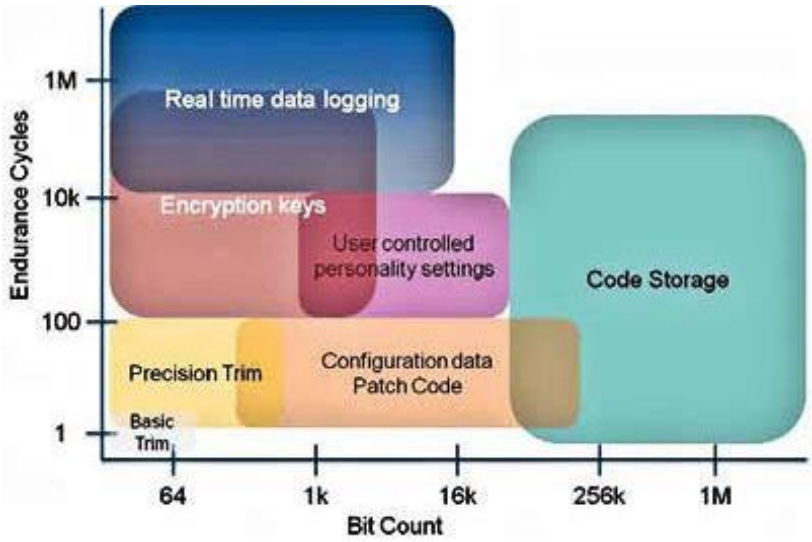


Figure 1. 4. Mapping of typical applications into NVM space [Zajac '10]. "Bit count" is the amount of data that can be stored in a given block.

1.2.2 Memory classification

There are various possibilities to classify semiconductor memories; one is to consider their electrical characteristics (figure 1.5).

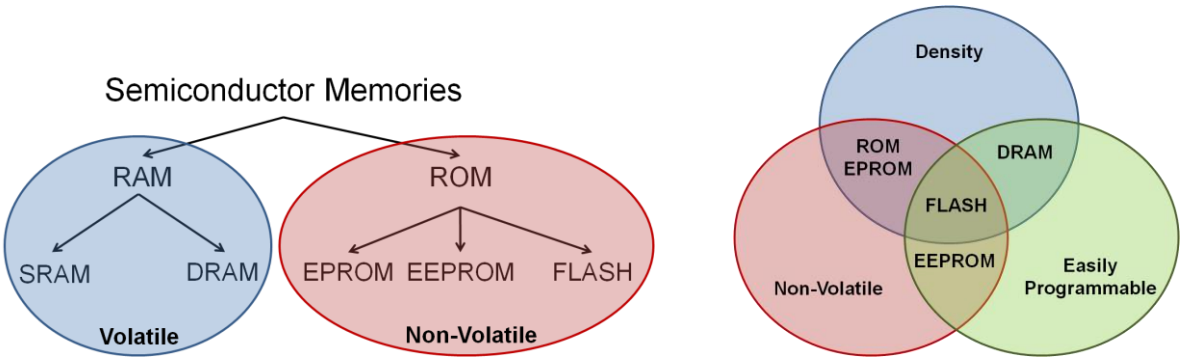


Figure 1. 5. Left: Overview of the non volatile semiconductor memories; Right: Semiconductor memory classification by different performance criteria.

Volatile Memories: are fast memories that are used for temporary storage data since they lose the information when the power is turned off. We can divide them into two types:

Static Random Access Memory (SRAM). The information is maintained as long as they are powered. They are made up of flip-flop circuitry (six transistors in a particular configuration). Because of its large number of components SRAM is large in size and cannot compete with the density typical of other kinds of memories.

Dynamic Random Access Memory (DRAM). These memories lose the information in a short time. They are made up of a transistor and a capacity where the charge is stored. They are widely used in processors for the temporary storage of information. As the capacitor loses the charge, a refresh or recharge operation is needed to maintain the right state.

Non-Volatile Memories: they retain the information even when the power is down. They have been conceived in order to store the information without any power consumption for a long time. This thesis concerns the study of charge storage non volatile memories that are a subgroup of the semiconductor memories. However it is important to remember that there are other devices were the information can be stocked. A very common storage device is the magnetic disk; its main drawback being the long access time and the sensitivity to magnetic fields. Another example of non-volatile memory is the CD technology developed in the late 1970s which uses an optical media that can be read fast, but necessitating a pre-recorded content. Here we will only describe the memory based on semiconductor technology:

Read Only Memory (ROM). This is the first non-volatile semiconductor memory. It consists in a simple metal/oxide/semiconductor (MOS) transistor. Thus its cell size is potentially the smallest of any type of memory device. The memory is programmed by channel implant during the fabrication process and can never be modified. It is mainly used to distribute programs containing microcode that do not need frequent update (firmware).

Programmable Read Only Memory (PROM). It is similar to the ROM memory mentioned above, but the programming phase can be done by the user. It was invented in 1956 and can constitute a cheaper alternative to the ROM memory because it does not need a new mask for new programming.

Erasable Programmable Read Only Memory (EPROM). This memory could be erased and programmed by the user, but the erase has to be done by extracting the circuit and putting it under ultraviolet (UV) radiations. The particularity of this device is the presence of a "floating gate" between the control (top) and tunnel (bottom) oxides. In 1967 D. Khang and S. M. Sze proposed a MOS-based non-volatile memory based on a floating gate in a metal-insulator-metal-insulator-semiconductor structure [Kahng '67]. At the time, however, it was almost impossible to deposit a thin oxide layer (<5nm) without introducing fatal defects. As a consequence a fairly thick oxide layer was adopted and this type of device was developed for the first time at Intel by [Frohman-Bentchkowsky '71].

Electrically Erasable Programmable Read Only Memory (EEPROM). In this memory both the write and erase operations can be electrically accomplished, without removing the chip from the motherboard. The EEPROM cell features a select transistor in series to each floating gate cell. The select transistor increases the size of the memories and the complexity of array organization, but the memory array can be erased bit per bit.

Flash memory is a synthesis between the density of EPROM and the enhanced functionality of EEPROM. It looks like EEPROM memory but without the select transistor. Historically, the name comes from its fast erasing mechanism. Because of these properties and the new applications (figure 1.6) the flash memory market is growing at a higher average annual rate than DRAM and SRAM, which makes it today the most produced memory (figure 1.2). Depending on their applications, flash memories can used in two different architectures that we introduce here and we will describe in next section. **NOR** flash memory provides random memory access and fast reads useful for pulling data from the memory. The **NAND**, on the other hand, reads data slowly but has fast write speeds and high density.

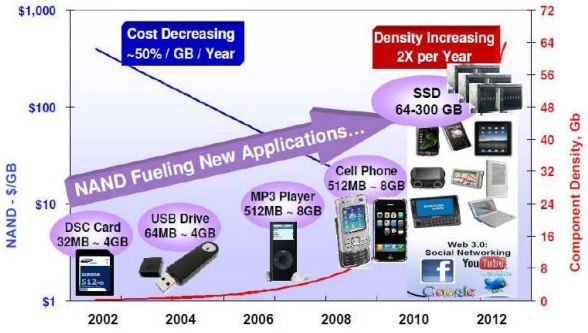


Figure 1. 6. Market trend of NAND Flash memories in portable applications [Bez '11].

1.2.3 Flash memory architectures

Flash memories are organized in arrays of rows (word lines or WL) and columns (bit lines or BL). The type of connection determines the array architecture (figure 1.7).

NOR: The NOR architecture was introduced for the first time by Intel in 1988. The cells are connected in parallel and in particular, the gates are connected together through the wordline, while the drain is shared along the bitline. The fact that the drain of each cell can be selectively selected enables a random access of any cell in the array. Programming is generally done by channel hot electron (CHE) and erasing by Fowler-Nordheim (FN). NOR architectures provide fast reading and relatively slow programming mechanisms. The presence of a drain contact for each cell limits the scaling to $6F^2$, where F is the smallest lithographic feature. Fast read, good reliability and relatively fast write mechanism make NOR architecture the most suitable technology for the embedded applications requiring the storage of codes and parameters and more generally for execution-in-place. The memory cells studied in this thesis will be integrated in a NOR architecture for embedded applications.

NAND: Toshiba presented the NAND architecture development in 1987 in order to realize ultra high density EPROM and Flash EEPROM [Masuoka '87]. This architecture was introduced in 1989 and presented all the cells in series where the gates were connected by a wordline while the drain and the source terminals were not contacted. The absence of contacts means that the cell cannot be selectively addressed and the programming can be done only by Fowler-Nordheim. On the other hand, it is possible to reach an optimal cell size of $4F^2$, thus a 30% higher density than in NOR cells. In NAND architecture programming is relatively fast but the reading process is quite slow as the reading of one cell is done by forcing the cell in the same bit line to the ON state. The high density and the slow reading but fast writing speeds make NAND architecture suitable for USB keys, storing digital photos, MP3 audio, GPS and many other multimedia applications.

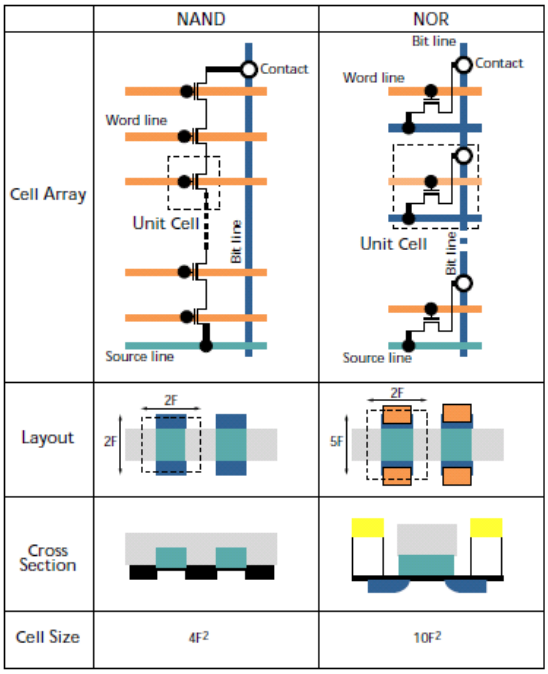


Figure 1. 7. Architectures of NAND (left) and NOR (right) memory array (source: micron.com).

1.3 Floating gate cell

The floating gate cell is the basis of the charge trap memory. The understanding of the basic concepts and functionalities of this device are fundamental and studied in this thesis. In this part we will describe flash memory operations. The operation principle is the following (figure. 1.8a): when the cell is erased there are no charges in the floating gate and the threshold voltage (Vt) is low (Vte). On the contrary when the memory is programmed (or written) the injected charge is stored in the floating gate layer and the threshold voltage value is high (Vtp). To know the state of the memory (e.g. the amount of trapped charge) it is just necessary to bias the gate with moderate read voltage (Vg) that is between (Vte) and (Vtp) and then determine if the current flows through the channel (ON state) or not (OFF state).

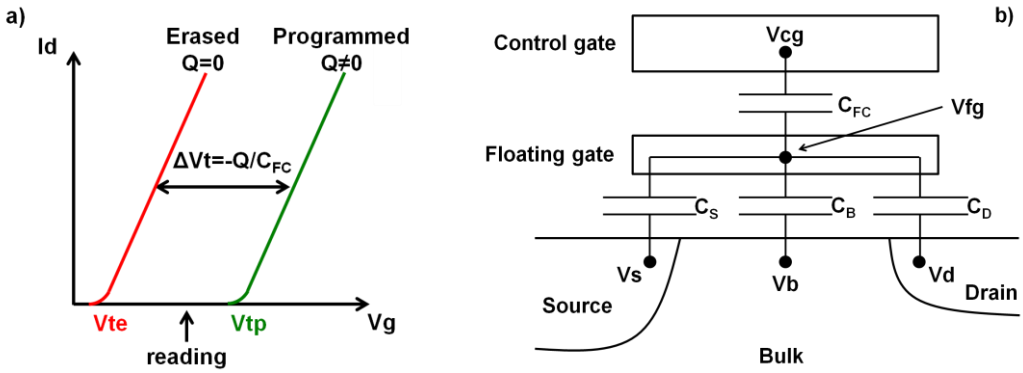


Figure 1. 8. a) I-V trans-characteristics of a floating gate device for two different values of charge stored within the floating gate (Q=0 and Q≠0). b) Schematic cross section of a floating gate transistor. The model using the capacitance between the floating gate and the other electrodes is described [Cappelletti '99].

The schematic cross section of a generic FG device is shown in figure 1.8b; the upper gate is the control gate (CG) and the lower gate, completely isolated within the gate dielectric, is the floating gate (FG). The simple model shown in figure 1.8b helps to understand the electrical behavior of a FG device. C_{FC}, C_S, C_B, and C_D are the capacitances between the FG and control gate, source, drain, and substrate regions, respectively. The potentials are described as follows: V_{FG} is the potential on the FG, V_{CG} is the potential on the control gate, and V_S, V_D and V_B are potentials on source, drain, and bulk, respectively [Pavan '97].

1.3.1 Basic structure: capacitive model

The basic concepts and the functionality of a FG device are easily understood if it is possible to determine the FG potential. Consider the case when no charge is stored in the FG, i.e., Q=0.

$$Q = 0 = C_{FC}(V_{FG} - V_{CG}) + C_S(V_{FG} - V_S) + C_B(V_{FG} - V_B) + C_D(V_{FG} - V_D) \qquad (1)$$

Where V_{FG} is the potential on the FG, V_{CG} is the potential on the control gate, and V_S, V_D and V_B are potentials on source, drain, and bulk, respectively. We name:

$$C_T = C_{FC} + C_S + C_D + C_B \qquad (2)$$

The total capacitance of the FG, and

$$\alpha_J = \frac{C_J}{C_T} \qquad (3)$$

The coupling factor relative to the electrode J, where J can be one of G, D, S and B, the potential on the FG due to capacitive coupling is given by

$$V_{FG} = \alpha_G V_{GS} + \alpha_D V_{DS} + \alpha_S V_S + \alpha_B V_B \qquad (4)$$

It should be pointed out that (4) shows that the FG potential does not depend only on the control gate voltage but also on the source, drain, and bulk potentials [Pavan '97]. When the device is biased into conduction and the source is grounded, V_{FG} can be written approximately as [Wu '92]:

$$V_{FG} = Vt_{FG} + \alpha_G(V_G - Vt) + \alpha_D(V_D - V_{Dt}) \qquad (6)$$

Where α_G and α_D are the coupling factors, Vt_{FG} is the FG threshold voltage (i.e., the V_{FG} value at which the device turns on), while V_{Dt} is the drain voltage used for reading measurement. The control gate threshold voltage (Vt) is obviously dependent on the charge (Q) possibly stored in the FG and is typically given in the form:

$$Vt = \frac{Vt_{FG}}{\alpha_G} - \frac{Q}{C_{FG}} - \frac{\alpha_D}{\alpha_G} V_{Dt} \qquad (7)$$

When (7) is substituted into (6), the following well-known expression for V_{FG} is obtained:

$$V_{FG} = \alpha_G V_G + \alpha_D V_D + \frac{Q}{C_T} \qquad (8)$$

In particular the Vt shift (ΔVt) due to the programming operation is derived approximately as:

$$\Delta Vt = Vt - Vt_0 = -\frac{Q}{\alpha_G C_T} = -\frac{Q}{C_{FC}} \qquad (9)$$

Where Vt_0 is the threshold voltage when Q=0. Equations (8) and (9) reveal the importance of the gate coupling factor (α_G): (8) shows that high α_G induces a floating gate potential close to the applied control gate bias and consequently, the gate coupling ratio needs to be high for provide a good programming and erasing efficiency. On the other hand (9) indicates that high α_G reduces the impact of the storage charge to the programming window (ΔVt). The international roadmap for semiconductor [ITRS '12] indicates that the best trade-off is achieved with a α_G between 0.6 et 0.7.

1.3.2 Programming mechanisms

We describe in this section the two main methods to program a Flash memory cell: Fowler-Nordheim (FN) [Fowler '28] and the channel hot electron (CHE) [Takeda '83].

Fowler-Nordheim programming

The Fowler-Nordheim programming operation is performed by applying a positive high voltage on the control gate terminal (about 20V) and keeping source drain and bulk grounded (figure 1.9a). The high electric field generated through the tunnel oxide creates a gate current due to the FN tunneling of charge from the channel to the floating gate [Chang '83].

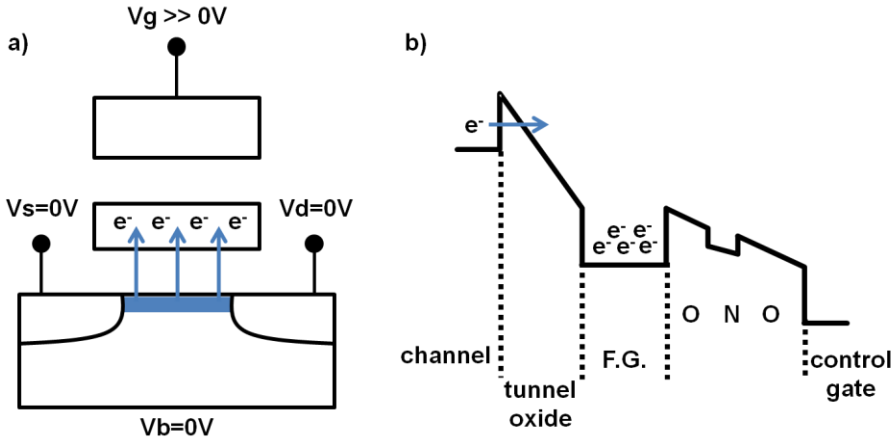


Figure 1. 9. a) FN programming mechanism representation. b) Band diagram of a floating gate memory during FN programming operation.

In this tunnel effect, the electrons flow from the conduction band of the silicon into the floating gate through the triangular energy barrier of the tunnel oxide (figure 1.9b). During the FN programming, the number of trapped electrons in the floating gate increases. As a consequence, the floating gate potential decreases and hence, the electric field through the tunnel oxide decreases. The charge injection will continue until the cancellation of electric field in the tunnel oxide. This is due to the maximum drop potential through the interpoly dielectric layer (ONO). This operation is relatively slow (order of milliseconds), but the energy consumption can be considered negligible because no current flows in the channel.

Channel Hot Electron programming

This operation is done keeping bulk and source grounded and applying a positive high voltage on gate (order of 10V) and drain (order of 5V) terminals (figure 1.10). The electrons are first strongly accelerated in the pinchoff region by the high lateral electric field induced by the drain/source bias. Then the electrons that have reached a sufficiently high kinetic energy are injected into the floating gate thanks to the vertical electric field induced by the positive voltage applied on the gate electrode [Ning '78] [Takeda '85] [Chenming '85]. Programming by channel hot electron is faster than FN (few microseconds). Furthermore the CHE is efficiency poor (only a few electrons are injected over the total amount of electrons that flow from source to drain [Simon '84]), and consequently high power consumption is reached. We remember that this programming mechanism is the main one used in this work to characterize the memory cells.

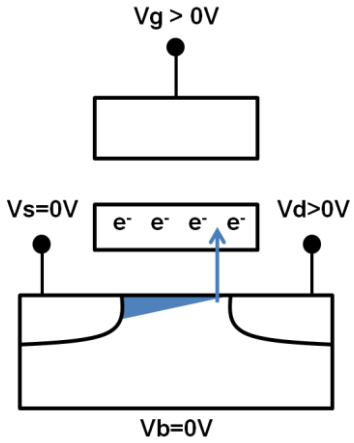


Figure 1. 10. Channel Hot Electron (CHE) programming mechanism representation.

1.3.3 Erase mechanisms

There are mainly four ways to erase the Flash floating gate cell; the schematic representations are shown in figure 1.11.

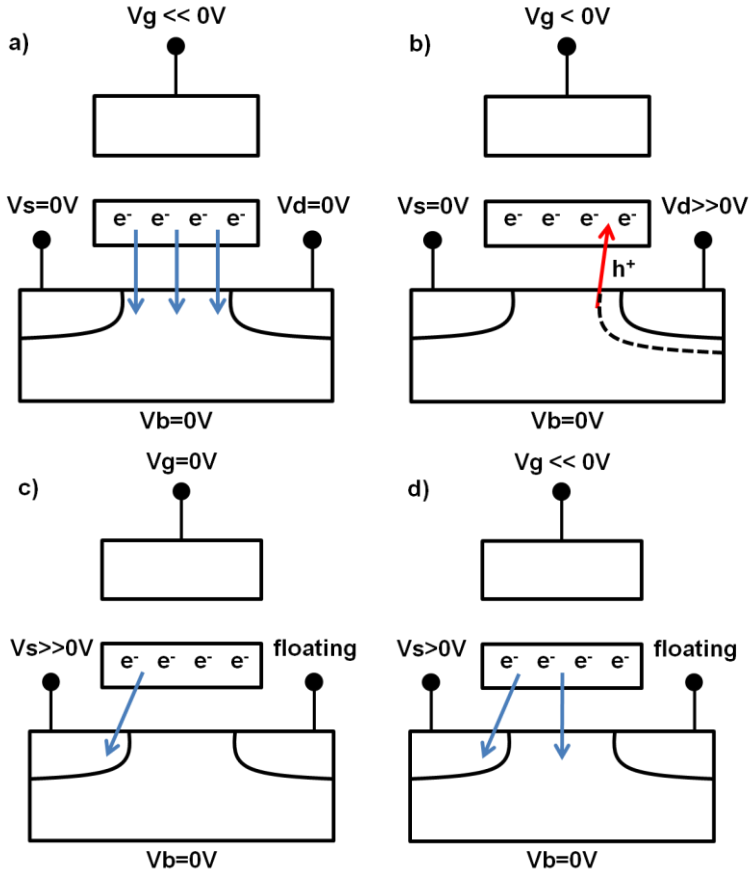


Figure 1. 11. Flash floating gate schematics of erase mechanisms: a) Fowler-Nordheim, b) Hot Hole Injection (HHI), c) source erasing, d) mix source-gate erasing.

Fowler-Nordheim erase

As for the programming operation, the source, drain and bulk are generally kept grounded while a strong negative voltage (order of -15V) is applied to the gate terminal. In this case, electrons are forced to flow from the floating gate to the semiconductor bulk. This method is slow, but the erasing is uniform on the channel surface (figure 1.11a). This is the preferred mechanism to erase the memory cells in NOR architecture. In the next chapter we will discuss about its effect on studied samples.

Hot Hole Injection (HHI) erase

This mechanism consists in accelerating the holes produced by reverse biasing of drain/bulk junction and by injecting them into the floating gate thanks to the vertical electric field [Takeda '83]. Figure 1.11b shows that this is done by keeping the bulk and source grounded and biasing positively the drain (order of 5V) and negatively the gate (about -10V). HHI erasing method is fast, localized near the drain and could induce the SILC (Stress Induced Leakage Current) phenomenon more easily than the methods listed above.

Source erasing

This forces electrons to flow from the floating gate into the source junction by FN tunneling. This erasing is done by applying a positive voltage of about 15V on the source and keeping bulk and gate grounded (figure 1.11c). In order to prevent current through the channel, the drain is kept floating. There are three main drawbacks to this method the erasing is localized near the source, it needs a strong source/gate overlap and it requires the application of a high voltage on the source terminal.

Mix source-gate erase

This is a mix between the source and the FN erasing. Electrons are erased both through the source and the channel. The principle is to share the high voltage needed in the source erasing between the gate and the source electrodes. As a result a negative bias of about -10V is applied on the gate and a positive bias of about 5V on the source. Again, the drain is kept floating in order to prevent source to drain current (figure 1.11d).

1.3.4 Evolution and limits of Flash memories

We explained at the beginning of the chapter that the new applications have commanded the semiconductor market and the research development. Since the invention of flash memory cell, the progress on device architectures and materials has been huge. The “ideal” memory should have:

• high density solution

• low power consumption

• non-volatility

• fast read/write/erase

• random read/write access

• endurance against write/erase cycles

• scalability with low cost

• compatibility with logic circuits and integration

This is the final objective of semiconductor research. As the “ideal” device does not exist different types of memories have been invented in order to push some specific properties. For these reasons, as shown in figure 1.12, memory technology development did not pursue a single technology solution, but rather it has oriented in many different direction over time [Hidaka '11] [Baker '12].

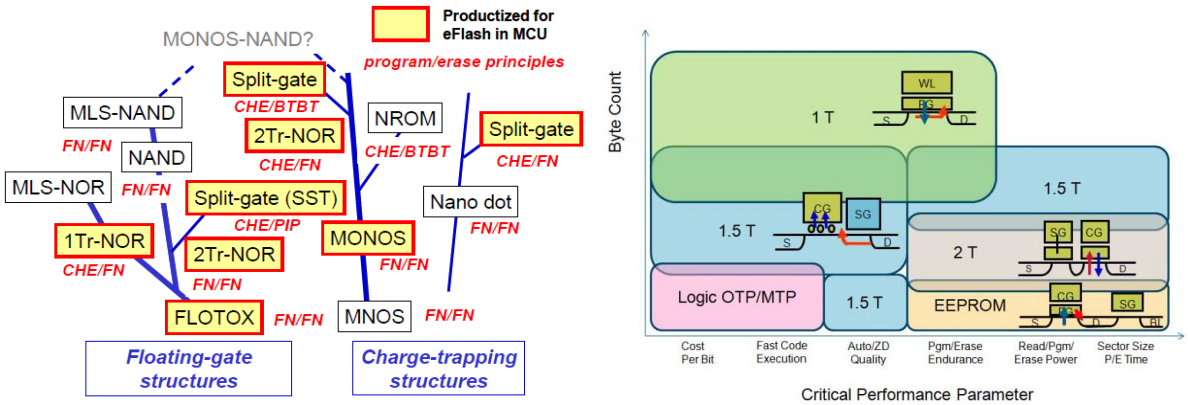


Figure 1. 12. Evolution of Flash and embedded-Flash memory technology (left) [Hidaka '11]. Mapping of common eNVM architectures to the NVM byte count and critical characteristics (right) [Baker '12].

Here, a category for non volatile storage with absolute minimum cost/bit is also shown. In this section we will first introduce the device scaling and the related challenges and then we will present flash cell developments. It is worth noting that the solutions found for the flash

memory cell can be used in embedded memory. In fact, even if in embedded memories there are minor constraints on the cell dimensions, research has always provided smaller non-volatile memory for embedded applications that have to face-off with the same flash scaling limits.

1.3.4.1 Device scaling

During the last 30 years Flash cell size has shrunk from 1.5um to 25nm doubling the memory capacity every year. In table 1.1 we report the international technology roadmap for semiconductor 2009 that forecasts the future trends of semiconductor technology.

ITRS 2012 - Process Integration, Devices, and Structures					
	2013	2016	2019	2022	2026
Nor flash technology node – F (nm)	45	38	32	28	22 ?
Cell size – area factor in multiplies of F2	12	12	12-14	14-16	14-16
Physical gate legth (nm)	110	100	90	85	85
Interpoli dielectric thickness (nm)	13-15	13-15	11-13	11-13	11-13

Table 1. 1. Table 1.1: Summary of the technological requirements for Flash NOR memories as stated in ITRS 2012 roadmap [ITRS '12]. White cell color: manufacturable solutions exist and are being optimized. Yellow cell color: manufacturable solutions are known. Red cell color: unknown manufacturable solutions.

We can see that even if the trend is maintained and the cell scaled down in the years to come, some technological solutions are still not known. Moreover, scaling beyond the 28nm will be very difficult if no revolutionary technology is adopted. The main issues that limit device miniaturizing are:

Stress Induced Leakage Current (SILC). During each erase/write cycle the stress degrades the tunnel oxide and the cell slowly loses its capacity to store electric charges (figure 1.13).

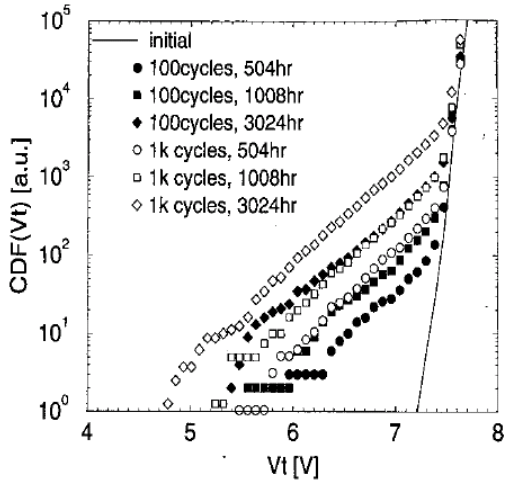


Figure 1. 13. Experimental cumulative distribution functions of bits vs. threshold voltage, measured at different times after different P/E cycling conditions[Hoefler '02].

This phenomenon, increases as the tunnel oxide is thinned. This is due to the defects induced in the oxide by the electrons that passing through it during program/erase operations [Pavan '97] [Ling-Chang '06] [Hoefler '02] [Belgal '02] [Kato '94] [Chimenton '01]. Consequently, the retention depends on the number of cycles and on the tunnel oxide thickness, but the physical scaling of this latter is limited to 6-7 nm.

Short Channel Effects (SCE). SCE appear when the gate length dimensions are so short that the gate control of the channel is lowered due to the influence of the source and drain potentials (figure 1.14). This parasitic effect produces the Drain Induced Barrier Lowering (DIBL) phenomenon [Yau '75], which results in threshold voltage decrease and the degradation of the subthreshold slope. Because of DIBL, the "OFF" current (I_{OFF}) increases and the power consumption reaches values incompatible with the advanced technology node requirements [Brews '80] [Fichtner '80] [Yau '74] [Fukuma '77]. Moreover, elevated I_{OFF} currents result in some disturb of the memory cell, especially in the erased state.

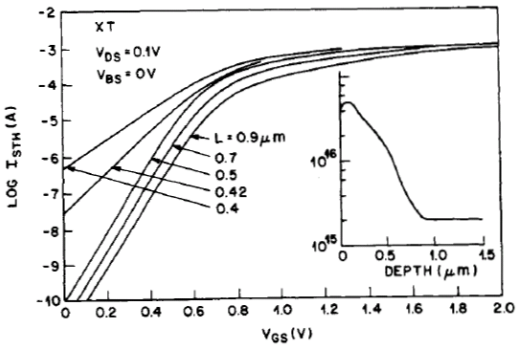


Figure 1. 14. Subthreshold current of MOS transistor as a function of gate voltage with the channel length as parameter. The insert is the calculated boron profile below the silicon surface in the channel [Fichtner '80].

Disturb. We consider here the main disturb effects due to the programming and reading operations done on unselected cell of a NOR memory array. It is to be remembered that in this thesis work, the electrical characterizations are based on the principle that the memory cells will be integrated in a NOR architecture for embedded applications.

- ***Programming disturb***. This impacts the unselected cells in the same bitline and wordline of selected cell. In the first case a drain stress is produced and the programmed cells can lose part of their charge due to FN drain stress on tunnel oxide causing hot hole injection (see the cell A in figure 1.15). The second case, represented in figure 1.15, cell B concerns a gate stress that can be induced on programmed cells (charge lost due to the stress through the ONO) or on erased cells (charge trapping due to the stress through the tunnel oxide).

- ***Read disturb***. In this case the selected cell can suffer from parasitic programming at low gate voltage; furthermore the unselected cells are gate stressed too (figure 1.15 cell C).

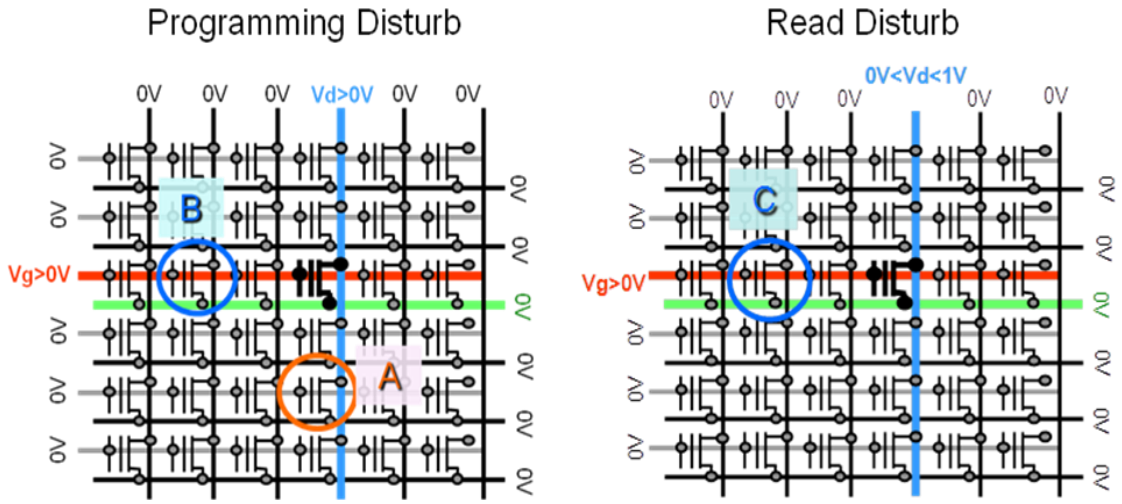


Figure 1. 15.Programming disturb (left) and read disturb (right) condition in NOR Flash memory array.

Coupling factor. As the cell dimensions are scaled down, the tunnel and interpoly dielectric should be scaled accordingly. To maintain the gate coupling factor (α_G) requirement, most Flash structures have the word line (control gate) wrapped around the side-walls of the floating gate to increase the ONO capacitance (figure 1.16). However because the retention of the cell should remain higher than 10 years, the dielectrics have to scale at a slower rate. The cell scaling reduces the distances between the neighboring cells and the contacts. This means that parasitic capacitances have to be taken into account for the coupling factor calculation, and we will explain our model in chapter 4 section 2.2.

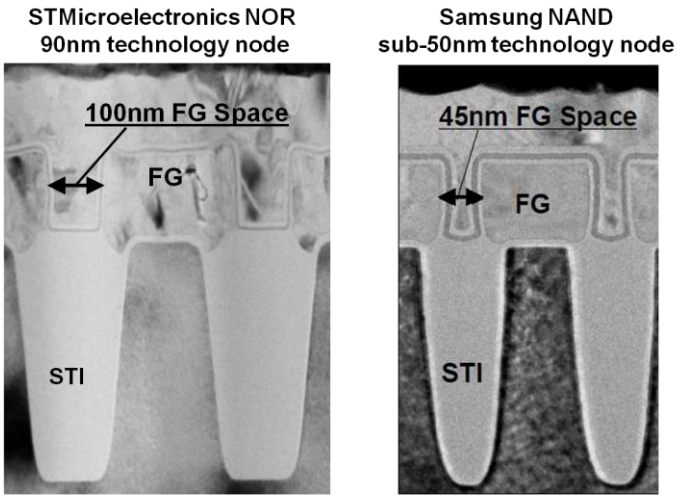
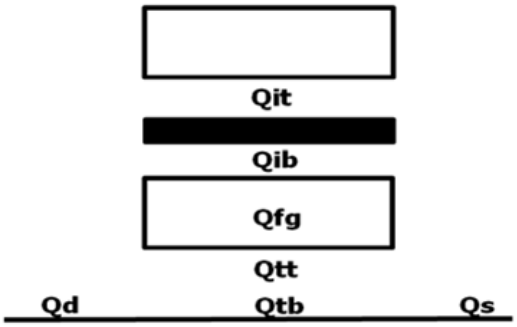


Figure 1. 16. TEM pictures of STMicroelectronics 90nm NOR Flash (left) and Samsung sub-50nm NAND Flash right) [Kim '05].

Parasitic Charge Trapping. In scaled memories the reduction in the number of stored electrons leads to a higher influence of the parasitic charge trapping on threshold voltage shift [Prall '10]. Figure 1.17 shows various locations within a NAND cell, thus programmed and erased by FN, where the parasitic charge can be trapped. The results of a TCAD simulation show that with the memory scaling, the number of electrons located outside the floating gate starts to dominate the cell threshold voltage shift. We will see in following chapters, how this parameter impacts the memory cell behavior.



	50nm	35nm	25nm
Qtb	4	2	1
Qtt	9	7	4
Qib	22	17	9
Qit	149	103	100
Qs	33	9	5
Qd	61	16	10
Qfg	18	12	10

Figure 1. 17. Locations of parasitic charge in a NAND cell (left). Number of electrons required in each location to shift the cell VTH by 100mV (right) [Prall '10].

Random Doping Fluctuation. The threshold voltage shift due to random variations in the quantity and position of doping atoms is an increasingly problem as device dimensions shrink. In figure 1.18 the mean and 3σ for the number of doping atoms are shown as a function of feature size. As device size scales down, the total number of doping atoms in the channel decreases, resulting in a larger variation of doping numbers, and significantly impacting threshold voltage. It has been documented [Frank '99] that at 25nm node, the Vt can be expected to vary of about 30% purely due to the random doping fluctuation.

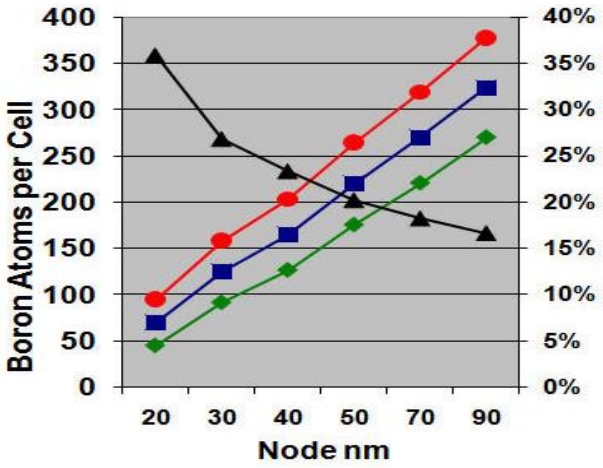


Figure 1. 18. Number of Boron atoms per cell (mean: square, -3σ: diamond, +3σ: circle vs. feature size). The triangle shows the ±3σ percentage divided by the mean [Prall '10].

1.3.5 Alternative solutions

In this section we will describe some of the envisaged modification to the classical flash memory cell in order to overcome the scaling limits presented in the previous section.

1.3.5.1 Tunnel dielectric

In a flash memory the tunnel dielectric has the double role of tunneling media during programming operations and electrostatic barrier in order to preserve the stocked charge. Moreover we must avoid the creation of defects during the programming operations that can induce the SILC and degrade the retention and cycling performance. This technological challenge can be solved by engineering the tunnel barrier. As shown in figure 1.19 crested barriers can provide both sufficient programming and retention. Several crested barriers have been tested: the most common one consists in an ONO layer [Hang-Ting '05], but other combinations have also been experimented ($SiO_2/Al_2O_3/SiO_2$ [Blomme '09], SiO_2/AlN [Molas '10]).

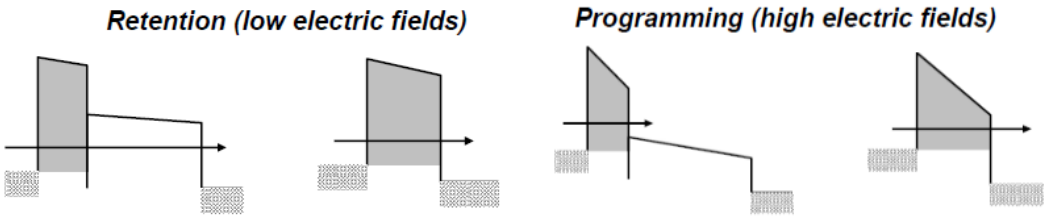


Figure 1. 19. Principle of operation of crested barrier [Buckley '06].

1.3.5.2 Interpoly material

Maintaining a constant coupling ratio at a value of 0.6-0.7 is a great scaling challenge. The use of high-k dielectric in the interpoly dielectric is envisaged to reduce the total EOT while maintaining or even increasing the gate coupling ratio. The choice of the high-k must take into account that for most of them the high dielectric constant comes at the expense of a narrower band gap (figure 1.20). This narrowed band gap can cause leakage current during retention operation [Casperson '02]. In particular Alumina dielectric is employed in the TANOS (TaN/Al2O3/Si3N4/SiO2/Si) memory, proposed for the first time by Samsung in 2005 [Yoocheol '05], Despite the envisaged advantages, high-k materials are not as well known as the silicon oxide and they need further development before they can be integrated in the memory market. One of the main problems is that they inevitably introduce defects that can induce trap assisted conduction and degrade the memory operations [Leroux '04] [Wilk '01] [Ping-Hung '08].

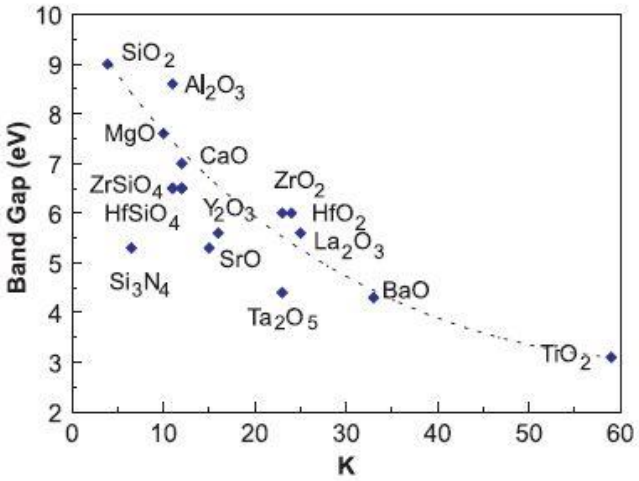


Figure 1. 20. Relationship between the dielectric constant and band gap [Robertson '05].

1.3.5.3 Control Gate

During erasing operation, flash memories employing SiN as a charge trapping layer show a Vt saturation phenomenon (figure 1.21). This has been explained by the back-tunnelling effect. The back-tunnelling effect occurs during FN erasing operations when, due to the high negative bias applied on the gate electrode, the electrons flow from the poly-silicon gate to the charge trapping layer preventing the memory from completely erasing. To overcome this problem and in order to suppress the depletion capacitance of poly-silicon floating gate, different metallic materials with a high work function have been successfully tried improving erasing dynamic [Chang Hyun '03] .

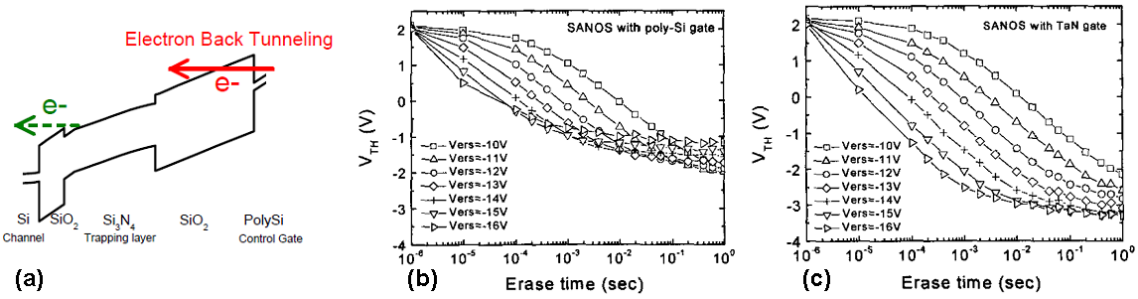


Figure 1. 21. (a) Schematic explaining electron back tunneling phenomena. (b) Erase characteristics of SANOS device with n+ poly-Si gate and (c) TaN/n+ poly-Si gate.

1.3.5.4 Trapping layer

In figure 1.22a we show the schematics of continuous floating gate cell and discrete charge trapping layer. In the first case the trapped charge is free to move along the polysilicon floating gate. This makes the device very sensitive to SILC. A discrete charge trapping layer is the solution envisaged to avoid the charge loss if an electric path is generated in tunnel

oxide. Replacing of polysilicon layer by a discrete charge trapping layer (figure 1.20b) enables the localization of the stored charge that is not free to move, which reduces the impact of the SILC on data retention. In case of defect creation in tunnel oxide, only a part of the global stored charge can be lost. The most often used material for charge trapping layer is the silicon nitride (Si_3N_4). However other materials like HfSiON, AlN and Si, have also been studied [Lai '05] [Wang '05] [Xuguang '04] [Monzio Compagnoni '04]. The subject of this thesis work is to characterize the silicon nanocrystal cell where the polysilicon charge trapping layer is replaced by silicon nanocrystals with and without the SiN capping layer. The results concerning the programming window, the reliability and the energy consumption will be shown in next chapters and compared to the standard floating gate.

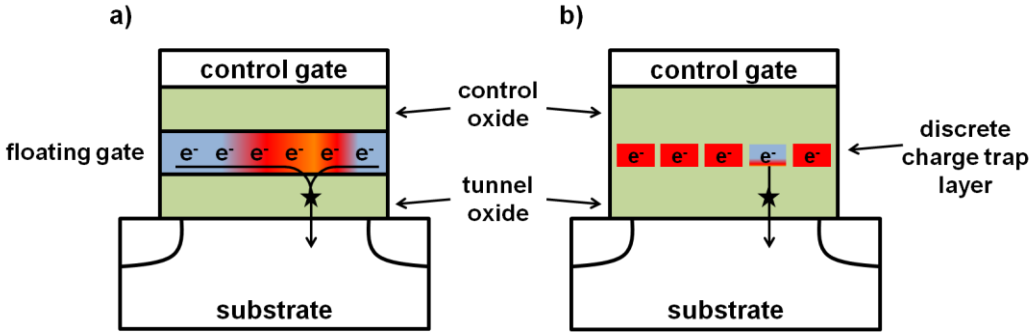


Figure 1. 22. Schematic diagrams representing (SILC) phenomena for (a) continuous floating gate cell (b) discrete charge trapping layer.

1.4 Silicon nanocrystal memory: state of the art

The market of nonvolatile Flash memories, for portable systems, requires lower and lower energy and higher reliability solutions. The silicon nanocrystal Flash memory cell appears as one promising candidate for embedded applications. The functioning principle of discrete charge trapping silicon nanocrystal memories (Si-nc) is similar to floating gate devices. In this thesis we consider the integration of Si-nc memories in NOR architecture for embedded applications programmed by channel hot electron and erased by Fowler-Nordheim mechanisms.

There are many are the advantages to using this technology:

- Robustness against SILC and RILC (Radiation Induced Leakage Current), this enables to scale the tunnel oxide thickness to be scaled down to 5nm, while the ten year data retention constraint is guaranteed. Moreover the operation voltages can be decreased too [Compagnoni '03] [Monzio Compagnoni '04]. Further

improvements can be achieved using cells with a high number of nanocrystals [De Salvo '03].

- Full compatibility with standard CMOS fabrication process encouraging industrial manufacturability, reducing the number of masks with respect to the fabrication of floating gate device [Muralidhar '03] [Baron '04] and ease of integration [Jacob '08].

- Decrease in cell disturb, due to the discrete nature of nanocrystals and their smaller size than a floating gate, the coupling factor between the gate and drain is reduced as well as the disturbs between neighboring cells.

- Multi level applications, the threshold voltage of a silicon nanocrystal transistor depends on the position of stored charge along the channel [Crupi '03] [De Salvo '03].

Despite these peculiarities two main drawbacks characterize the Si-nc memories:

- The weak coupling factor between the control gate and nanocrystals. This implies finding a method to keep the program/erase voltages small and to take advantage of the decrease in tunnel oxide thickness [De Salvo '01].

- The spread in the surface fraction covered with Si-nc limiting this type of cell for high integration density applications [Gerardi '04].

IBM presented the first Si-nc memory at IEDM [Tiwari '95] in order to improve the DRAM (Dynamic Random Access Memory) performance using a device with characteristics similar to EEPROM. The polysilicon floating gate is replaced by silicon nanocrystals grown on tunnel oxide by Low Pressure Chemical Vapor Deposition (LPCVD) two step process. This type of fabrication enables the size and density of nanocrystals to be controlled separately [Mazen '03] [Mazen '04].

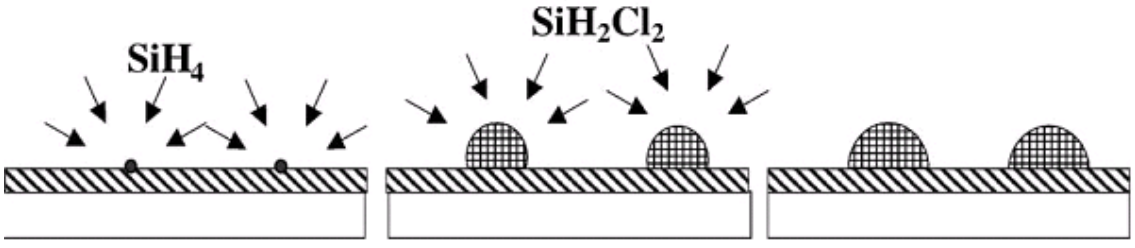


Figure 1. 23. Schematic representation of the nucleation and growth two step process[Mazen '03].

Other techniques of fabrication have also been developed: ionic implantation [Hanafi '96], annealing of SRO (Silicon Rich Oxide) layers deposition [Rosmeulen '02] and aereosol deposition [De Blauwe '00]. Thanks to these research works, Motorola demonstrated the interest in using this device for non-volatile applications by developing a 4Mb memory array [Muralidhar '03]. In addition STMicroelectronics in collaboration with CEA-Leti presented their 1Mb memory array [De Salvo '03]. The three main actors in the industry of silicon nanocrystal memories are STMicroelectronics, Atmel and Freescale.

STMicroelectronics, in collaboration with CEA-Leti, presented in 2003 a 1Mb nanocrystal CAST (Cell Array Structure Test, figure 1.24a) where the Si-nc were fabricated with a two step LPCVD process [De Salvo '03] [Gerardi '04]. This structure is programmed and erased by Fowler-Nordheim tunneling, and the write/erase characteristics are reported in figure 1.24b.

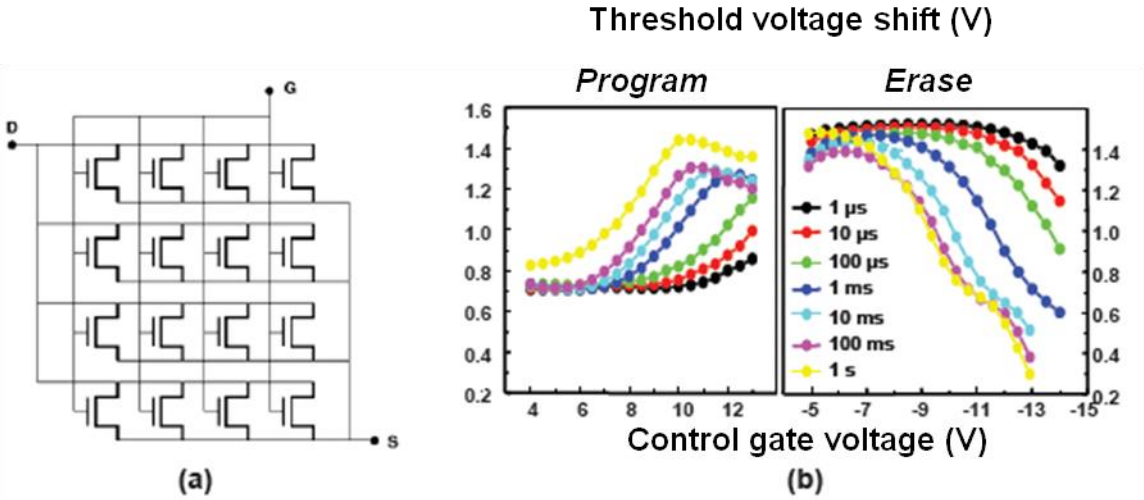


Figure 1. 24. a) Schematic of a CAST structure. b) Program/erase characteristics in fully Fowler-Nordheim regime [De Salvo '03].

As a result STMicroelectronics presented in 2007 a 16Mb Flash NOR memory array divided into 32 sectors of 512kb [Gerardi '07a] [Gerardi '07b]. The silicon nanocrystals were grown on a tunnel oxide, 5nm thick, with a diameter between 3nm and 6nm and a density of $5 \cdot 10^{11} nc/cm^2$. To complete the stack an ONO layer was used as a control oxide (EOT=12nm). In figure 1.25a the program/erase threshold voltage distributions of 16Mb memory array are plotted. In this case the cells have been programmed by channel hot electron and erased by Fowler-Nordheim reaching a 3V programming window in case of the average of distributions and 800mV for the worst case. Moreover Gerardi highlighted the problem of parasitic charge trapping in ONO layer during the cycling (figure 1.25b).

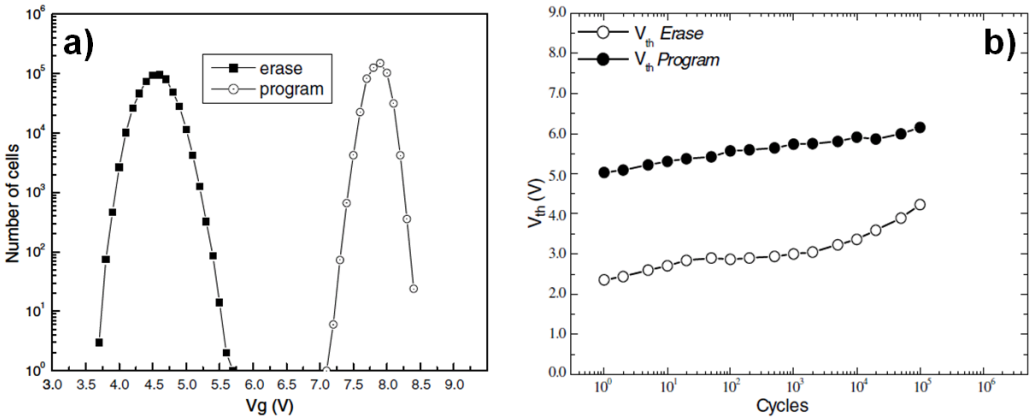


Figure 1. 25. a) Program and erase threshold voltage distributions for one sector of 512 kb of nanocrystal memory cells. b) Evolution of the program/erase threshold voltages of a Si-nc memory cell showing that the program/erase levels are shifted due to electron trapping in the ONO [Gerardi '07b].

Finally they processed the silicon nanocrystal memory cell in order to assume a cylindrical shape, which greatly benefits improve the coupling ratio (figure 1.26a). In addition, they used an optimized ONO control dielectric, enabling the reduction of the parasitic charge trapping during cycling (figure 1.26b); this type of cell was integrated in a 4Mb NOR array [Gerardi '08].

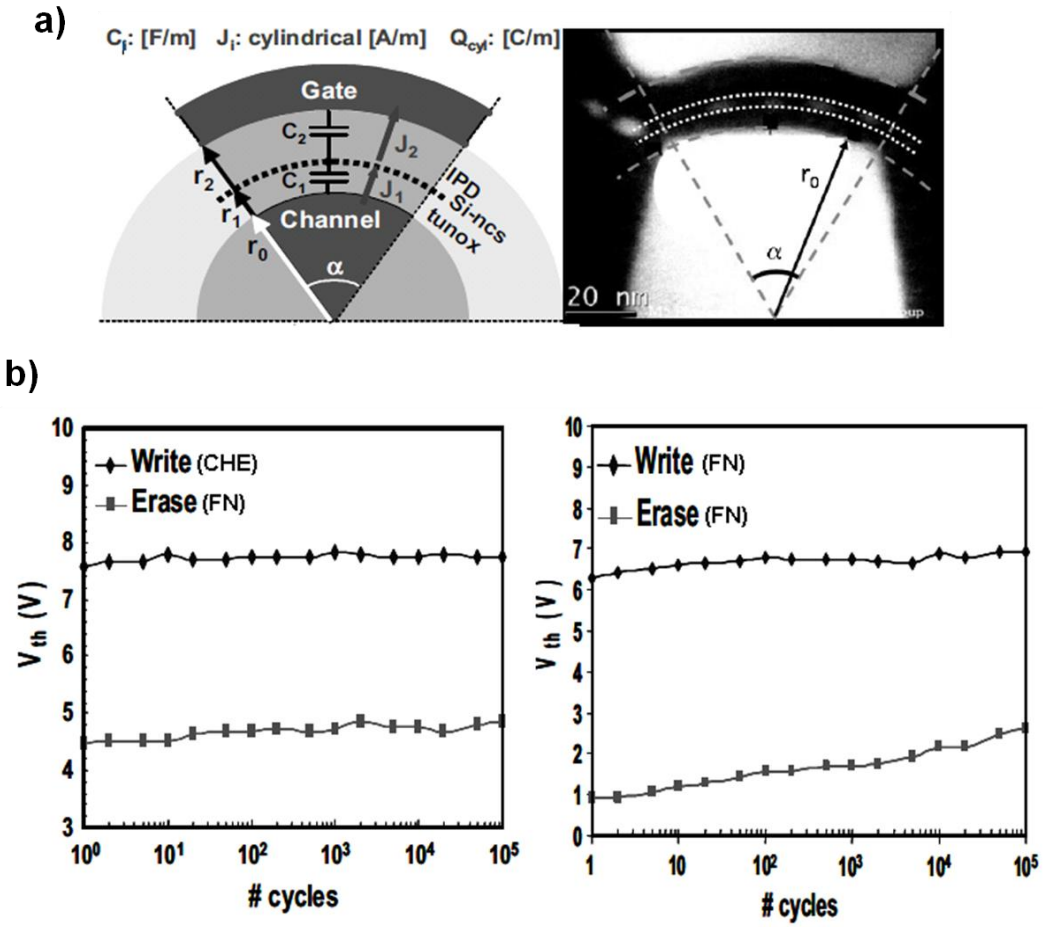


Figure 1. 26. a) Cross-section of the cylindrical-shaped structure and corresponding TEM image on the right. b) Endurance characteristic of a Si-nc cell by using CHE/FN and FN/FN program/erase operations.

ATMEL presented an extensive set of experiments, resulting from a joint venture with CEA-Leti. In this context, they demonstrated the validity of the discrete storage node concept on a 32Mb Si-NC NOR Flash memory product, fabricated in a 130 nm technology platform. The presented results gave a quantitative estimation of the maturity of two step LPCVD Si-NC technology [Jacob '07]. Their study focused on the impact of silicon nanocrystal size on memory cell performance integrating a HTO (High Temperature Oxide) as control oxide in order to avoid the parasitic charge trapping in ONO layer shown by STMicroelectronics results. In figure 1.27b the written and erased threshold voltages, measured on a set of 40 memory cells, are shown (corresponding to different dot sizes, see figure 1.27a) [Jacob '08].

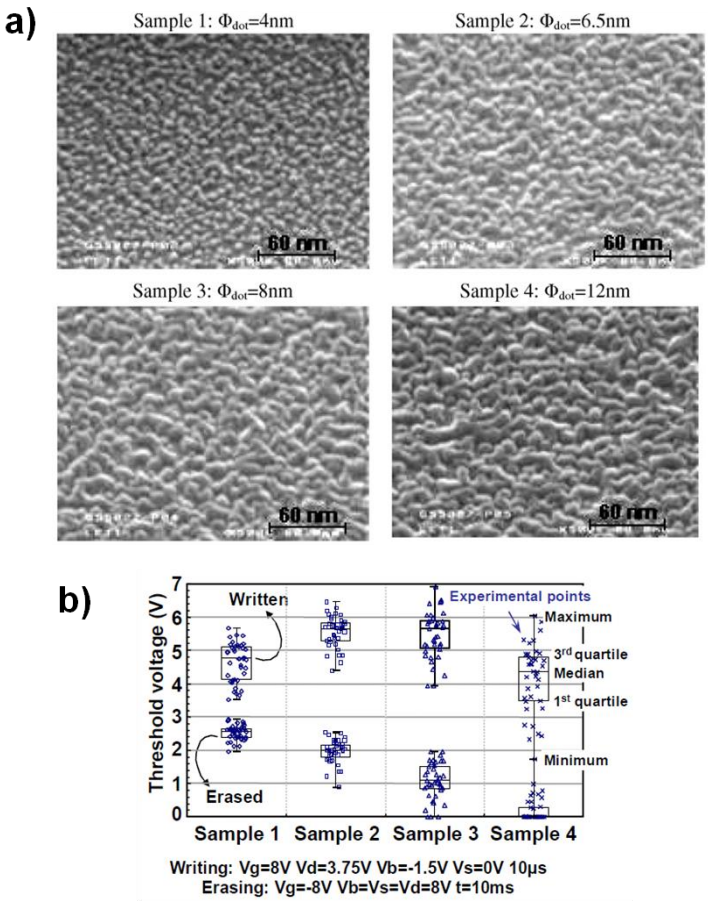


Figure 1. 27. a) SEM images of Si-NCs with same dot nucleation step and different dot growing times. b) written and erased Vth of bitcells with different Si-NCs [Jacob '08].

It clearly appears that increasing the Si-nc size, the programming window is increased too. Indeed, this result well agrees with the theoretical model [De Salvo '01] which states that the programming window linearly increases with the floating gate surface portion covered by the Si-NCs. In fact it was demonstrated for the Si-nc cell that the dynamic charging/discharging Si-dot memory corresponds better to a FG memory device operation rather than to a pure capture/emission trap-like behavior [De Salvo '01]. Starting from the capacitive model of

Flash floating gate (section 1.3.1), and by considering the discrete nature of nanocrystals, the coefficient α_D can be neglected and the equation (8) can be rearranged as:

$$V_{FG} = \alpha_G V_G + \frac{Q}{C_T} \qquad (10)$$

In this FG-like approach, we define a parameter that takes into account the surface portion covered by the nanocrystals (R_{nc}). It corresponds to a weighting factor for the trapped charges in the MOSFET threshold voltage; the Vt shift in this case takes into account this parameter and can be written as:

$$\Delta Vt = Vt - Vt_0 = -\frac{Q \cdot R_{nc}}{C_{FC}} \qquad (11)$$

This approach will be considered as fundamental in the next chapters in order to improve the Si-nc memory cell coupling factor and thus the programming window.

We reported in figure 1.28 the results shown in [Jacob '08] concerning cell reliability using the HTO control oxide and keeping the silicon nanocrystal size constant.

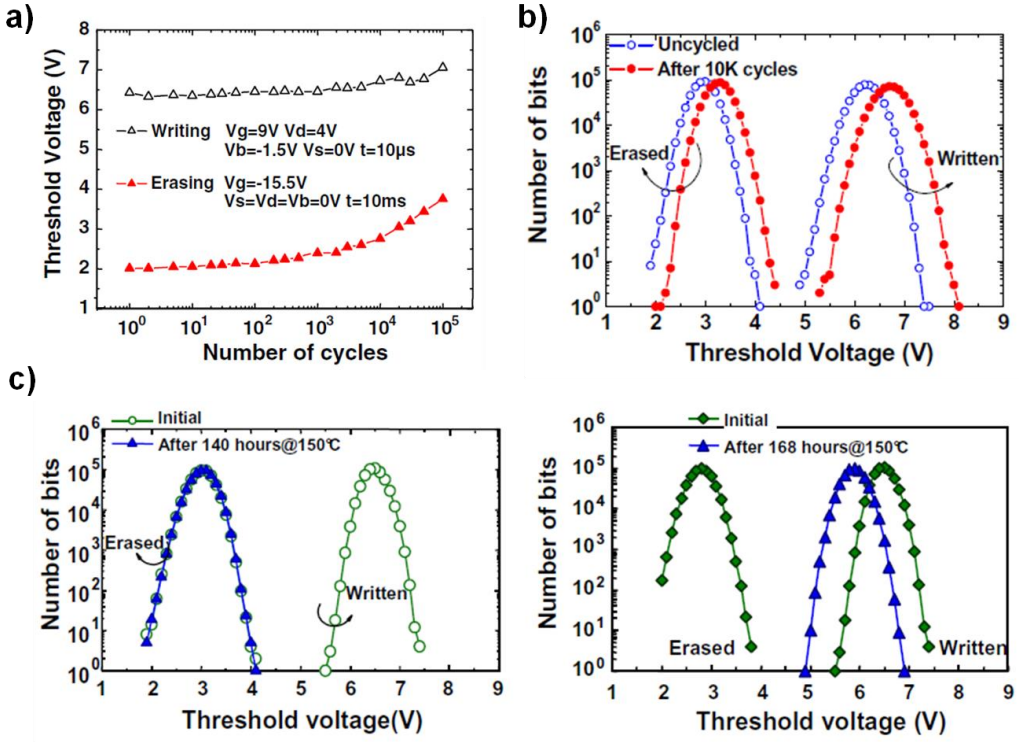


Figure 1. 28. a) Endurance data for a memory bitcell b) Threshold voltage distributions of erased and written states of two different sectors, measured before and after 10k write/erase cycles. c) Data retention at 150°C on two different uncycled 512Kb sectors [Jacob '08].

Figure 1.28a shows the program/erase endurance data of a memory bitcell corresponding to a sample with 9nm Si-nc diameter and 10^{12}nc/cm² Si-nc density. The cell has been programmed by CHE adding a substrate biasing that enables trapping charge in the nanocrystals positioned toward the center of channel; this improves the programming window but degrades the analog circuit area and the cell consumption. The initial programming window (larger than 4V) decreases to 3V after 100kcycles, showing that even if the ONO layer is substituted by the HTO a parasitic charge trapping occurs. This appears also in figure 1.28b where the threshold voltage distributions of erased and programmed states of two different sectors are shown, measured before and after 10k write/erase cycles. In this case the programming window of tail bits is 600mV. To conclude, the data retention is reported in figure 1.28c. At 150°C the programming window loses 500mV after 168h.

Freescale was created by Motorola in 2004 when the studies on silicon nanocrystal memory cell had already been started [Muralidhar '03]. Freescale did a comparative study on the importance of control dielectric, using HTO and ONO samples because the latter, with its silicon nitride layer, represents a barrier against the parasitic oxidation of silicon nanocrystals and decreases the leakage current in the memory stack. As a drawback the parasitic charge trapping is present during the programming operations. In figure 1.29 we report the endurance characteristics of HTO and ONO samples fabricated by Freescale. For the HTO sample the threshold voltages remain stable up to 1kcycles, and afterwards their increase is explained by the parasitic charge trapping in the oxide. In the case of ONO sample, the electrons trapping in the silicon nitride layer starts immediately with the first program/erase cycles figure 1.29.

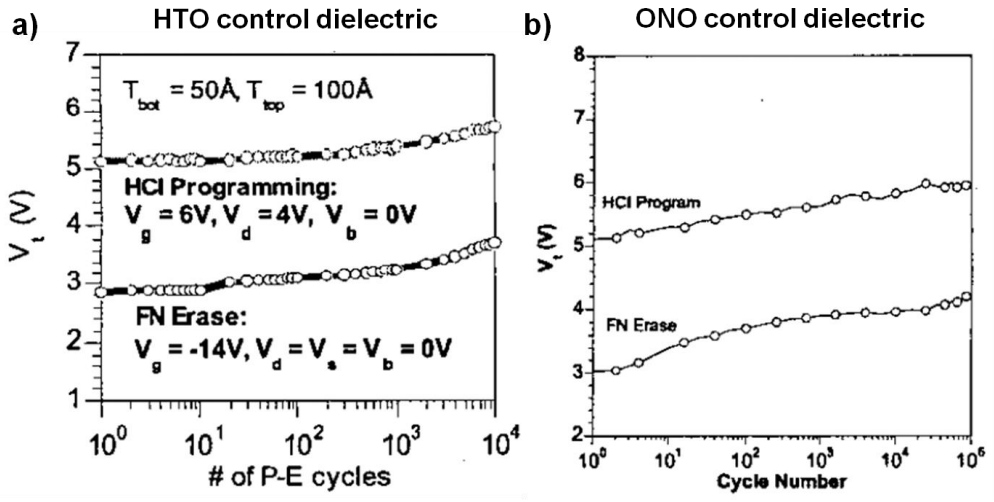


Figure 1. 29. Endurance characteristics of silicon nanocrystal cells integrating a) HTO [Steimle '04] and b) ONO [Muralidhar '04] control dielectric.

This parasitic charge trapping impacts also the data retention (figure 1.30). It is thus important to minimize it to reach the 20 year target. They demonstrated the advantage of discrete nature of silicon nanocrystals on data retention and read disturb; it enables the tunnel oxide thickness to be decreased and hence the program/erase voltages.

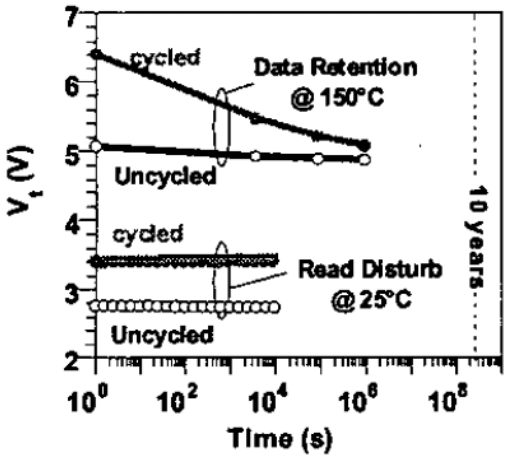


Figure 1. 30. Program state data retention and erased slate READ disturb characteristics for a nanocrystal NVM bitcell with a 5mn tunnel oxide. Exhibited charge loss in cycled case is attributed to detrapping of parasitic charge [Muralidhar '04].

Further studies have been performed concerning the impact of silicon nanocrystals size and density [Rao '05] [Gasquet '06]. Figure 1.31a shows that the covered area impacts the program/erase speed and the saturation level of the programming window.

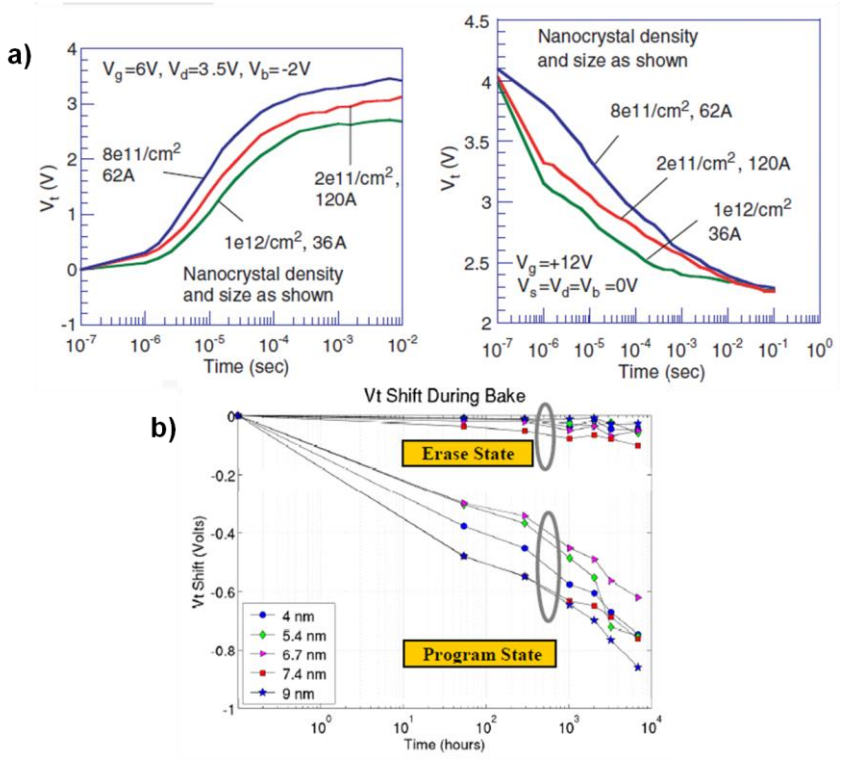


Figure 1. 31. a) HCI program and FN erase speed (with positive gate voltage) for devices with 4.5nm bottom oxide and 12nm top oxide, with different nanocrystal depositions [Rao '05]. b) 200ºC bake Vt shift for samples with different nanocrystals size [Gasquet '06].

The hot carrier injection speed increases with the covered area, while the Fowler-Nordheim erase operation is more efficient with smaller nanocrystals. This is due to the presence of HTO and the Coulomb blockade effect. Data retention measurements have been also carried out on a 4Mb memory array. The samples had a 5nm tunnel oxide and 10nm HTO (figure 1.31b). Here the data retention loss is shown during a 200ºC bake. The erased state is very close to neutral charge so the Vt shift is small while most of the variation in program state response originates in the first 54 hours of bake and appear uncorrelated to nanocrystal size. Moreover, Freescale decided to integrate silicon nanocrystals in high scalable Split Gate memories (figure 1.32) [Sung-Taeg '08] [Yater '09], where it is possible to control the current consumption during the hot carrier injection for low energy embedded applications [Masoero '11] [Masoero '12b].

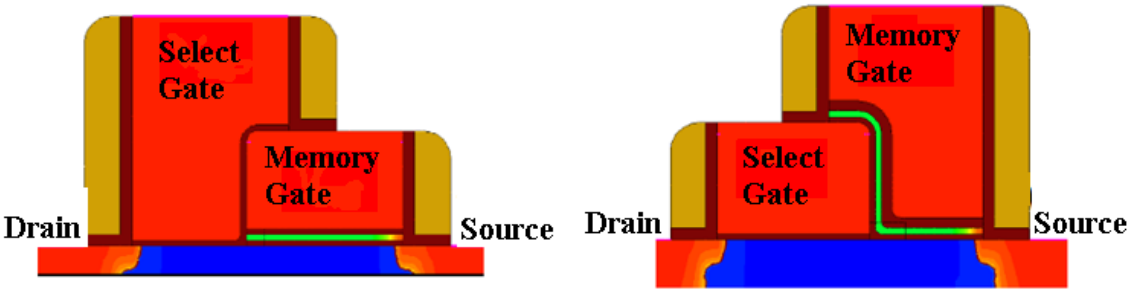


Figure 1. 32. Schematic of Split Gate with memory first Left) or Access first (Right) configuration [Masoero '12a].

Recent results of endurance and data retention are reported in figure 1.33.

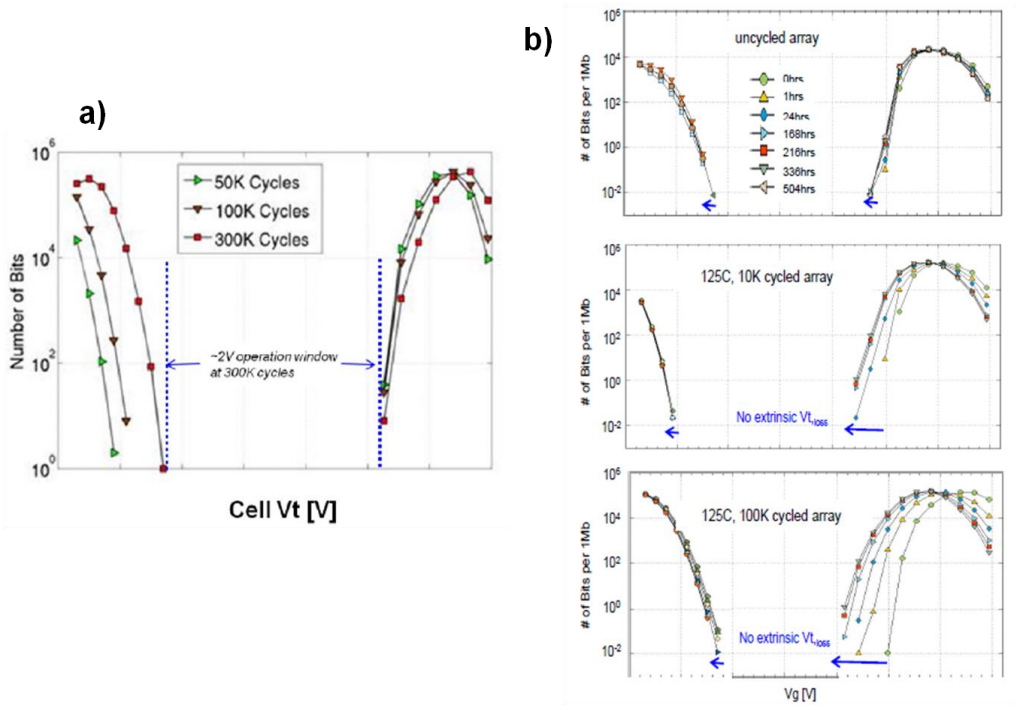


Figure 1. 33. a) Erase and program Vt distributions of cycles up to 300K at 25°C. b) Bake retention characteristics at 150C with fresh, 10K and 100K cycled parts of 125°C cycling temperature [Sung-Taeg '12].

The cycling experiments (figure 1.33a) show program and erase Vt distribution width that remain approximately constant throughout extended cycling and a substantial operating window is maintained even after 300Kcycles. Concerning the data retention, due to the inherent benefits of NC-based memories, no extrinsic charge loss was observed on fresh and cycled parts (figure 1.33b). The average loss for 504hrs for uncycled arrays is about 70mV and for 10K and 100K cycled arrays it is 250mV and 400mV, respectively [Yater '11] [Sung-Taeg '12]. Finally all these studies underline the importance of achieving a good coupling factor to improve the programming window and thus cell endurance, paying attention to the tunnel oxide thickness that plays an important role for the data retention and disturbs.

1.5 Flash technology for embedded applications

The 1T silicon nanocrystal technology is not the only solution to replace the Flash floating gate. In particular for the market of embedded applications the Flash memory array is integrated in the microcontroller products with SRAM, ROM and logic circuits achieving System on a Chip solution (SoC). This type of integrated circuit enables the fabrication costs reduction due to the compatibility with the CMOS process, by improving the system performance because the code can be executed directly from the embedded Flash. The most important applications for embedded products are the smart card and automotive, where low energy consumption, fast access time and high reliability are required (figure 1.34). In this scenario each one of main industrial actors searches the best compromise between cell area, performance and cost. In figure 1.35 we show the mainstream Flash concepts proposed by the top players of SoC manufacturers [Strenz '11].

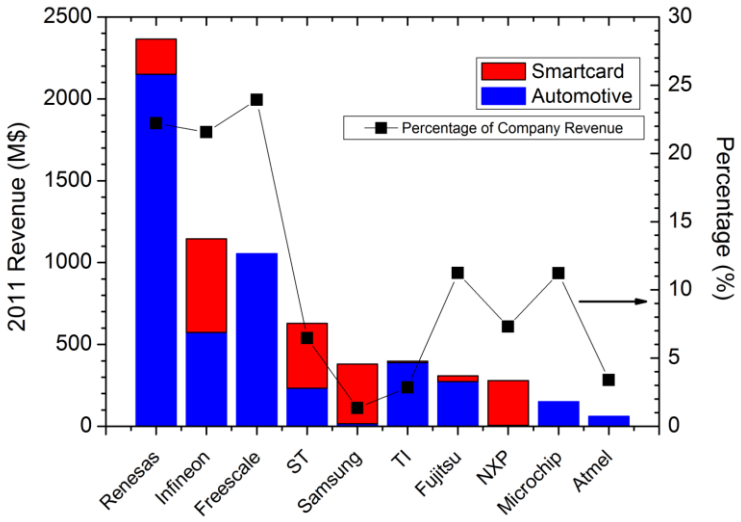


Figure 1. 34. Automotive microcontroller and smartcard market (source: iSupply Q1 2012 update).

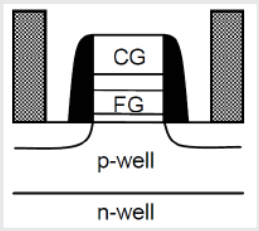
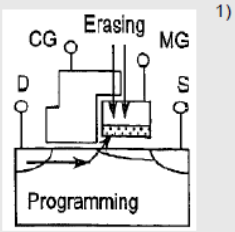
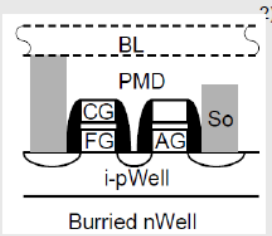
	1T-NOR	Splitgate	2T-NOR
			
P/E	(1) CHE / FN (2) FN / FN	(1) SSI / FN (2) SSI / HHI	FN/FN
Storage	Floating Gate	(1) Floating Gate (2) ONO	Floating Gate
Area	10 - 30 F^2	20 - 30 F^2	25 - 100 F^2
Main Players	(1) FSL, ST, TI (2) Infineon	(1) SST-licensees (2) Renesas	NXP, Samsung

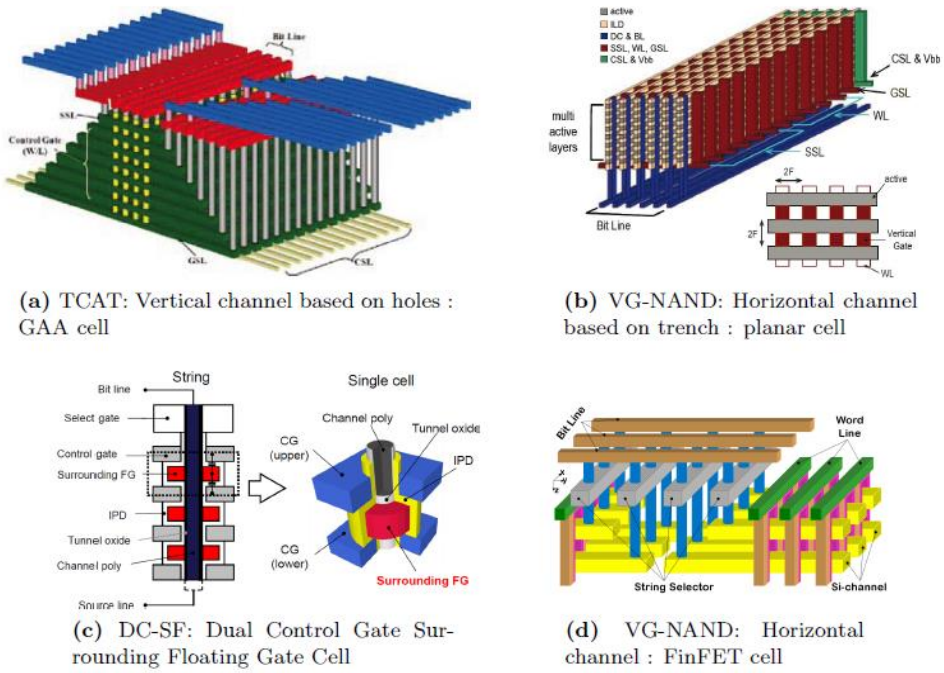
Figure 1. 35. Mainstream Flash integration concept [Strenz '11].

Although a large variety of different cell concepts can be found in sell, only three main concepts in terms of bitcell structure dominate the market – all of them using NOR array configuration: 1T stacked gate concepts, splitgate concepts as well as 2-transistor NOR concepts. Due to highly diverging product requirements there is a variety of concepts tailored to specific applications. Looking into development of new nodes a clear slowdown of area shrink potential can be observed for classical bitcell concepts while reliability requirements are tightened rather than relaxed. This increases the pressure for new, emerging cell concepts with better shrink potential. We used this brief analysis (Robert Strenz, Infineon - Workshop on Innovative Memory Technologies, Grenoble 2012), to highlight the concept that the industry push its technology to overcome the problem of scaling cost.

1.6 Innovative solutions for non volatile memory

Since the ultimate scaling limitation for charge storage devices is too few electrons, devices that provide memory states without electric charges are promising to scale further. Several non-charge-storage memories have been extensively studied and some commercialized, and each has its own merits and unique challenges. Some of these are uniquely suited for special applications and may follow a scaling path independent of NOR and NAND flash. Some may eventually replace NOR or NAND flash. Logic states that do not depend on charge storage eventually also run into fundamental physical limits. For example, small storage volume may be vulnerable to random thermal noise, such as the case of superparamagnetism limitation for MRAM. One disadvantage of this category of devices is that the storage element itself cannot

also serve as the memory selection (access) device because they are mostly two-terminal devices. Even if the on/off ratio is high, two terminal devices still lack a separate control (e.g. gate) that can turn the device off in normal state. Therefore, these devices use 1T-1C (FeRAM), 1T-1R (MRAM and PCRAM) or 1D-1R (PCRAM) structures. It is thus challenging to achieve small ($4F^2$) cell size without an innovative access device. In addition, because of the more complex cell structure that must include a separate access (selection) device, it is more difficult to design 3-D arrays that can be fabricated using just a few additional masks like those proposed for 3-D NAND [ITRS '12] [Jiyoung '09] [Tae-Su '09] [SungJin '10] [Eun-Seok '11] (figure 1.36).



(a) TCAT: Vertical channel based on holes : GAA cell

(b) VG-NAND: Horizontal channel based on trench : planar cell

(c) DC-SF: Dual Control Gate Surrounding Floating Gate Cell

(d) VG-NAND: Horizontal channel : FinFET cell

Figure 1. 36. An overview on 3D array integration of charge trapped Flash NAND: a) [Jiyoung '09], b)[Tae-Su '09], c) [SungJin '10], d) [Eun-Seok '11].

1.6.1 Ferroelectric Random Access Memory (FeRAM)

FeRAM devices achieve non-volatility by switching and sensing the polarization state of a ferroelectric capacitor. To read the memory state the hysteresis loop of the ferroelectric capacitor must be traced and the data must be written back after reading. Because of this "destructive read," it is a challenge to find ferroelectric and electrode materials that provide both adequate change in polarization and the necessary stability over extended operating cycles. The ferroelectric materials are foreign to the normal complement of CMOS fabrication materials and can be degraded by conventional CMOS processing conditions. Thus, the ferroelectric materials, buffer materials and process conditions are still being refined. So far, the most advanced FeRAM [Hong '07] is substantially less dense than NOR

and NAND flash. It is fabricated at least one technology generation behind NOR and NAND flash, and not capable of MLC. Thus, the hope for near term replacement of NOR or NAND flash has faded. However, FeRAM is fast, low power and low voltage, which makes it suitable for RFID, smart card, ID card and other embedded applications. In order to achieve density goals with further scaling, the basic geometry of the cell must be modified while maintaining the desired isolation. Recent progress in electrode materials show promise to thin down the ferroelectric capacitor [ITRS '12] and extend the viability of 2-D stacked capacitor through most of the near-term years. Beyond this the need for 3-D capacitors still remains a formidable challenge.

1.6.2 Magnetic Random Access Memory (MRAM)

MRAM devices employ a magnetic tunnel junction (MTJ) as the memory element. An MTJ cell consists of two ferromagnetic materials separated by a thin insulating layer that acts as a tunnel barrier. When the magnetic moment of one layer is switched to align with the other layer (or to oppose the direction of the other layer) the effective resistance to current flow through the MTJ changes. The magnitude of the tunneling current can be read to indicate whether a ONE or a ZERO is stored. Field switching MRAM probably is the closest to an ideal “universal memory”, since it is non-volatile and fast and can be cycled indefinitely, thus may be used as NVM as well as SRAM and DRAM. However, producing magnetic field in an IC circuit is both difficult and inefficient. Nevertheless, field switching MTJ MRAM has successfully been done in products. In the near term, the challenge will be the achievement of adequate magnetic intensity fields to accomplish switching in scaled cells, where electromigration limits the current density that can be used. Therefore, it is expected that field switch MTJ MRAM is unlikely to scale beyond 65 nm node. Recent advances in “spin-torque transfer (STT)” approach, where a spin-polarized current transfers its angular momentum to the free magnetic layer and thus reverses its polarity without resorting to an external magnetic field, offer a new potential solution [Miura '07]. During the spin transfer process, substantial current passes through the MTJ tunnel layer and this stress may reduce the writing endurance. Upon further scaling the stability of the storage element is subject to thermal noise, thus perpendicular magnetization materials are projected to be needed at 32 nm and below [ITRS '12].

1.6.3 Resistive Random Access Memory (RRAM)

RRAM is also a promising candidate next-generation universal memory because of its shorter write time, large R-ratio, multilevel capability, and relatively low write power consumption. However, the switching mechanism of RRAM remains unclear. RRAM are based on binary metal oxides has been attracting increasing interest, owing to its easy fabrication, feasibility of 3-D (stacked) arrays, and promising performances. In particular, NiO and HfO based RRAM have shown low voltage and relatively fast programming operations [Russo '09] [Vandelli '11]. RRAM functionality is based on the capability to switch the device resistance by the application of electrical pulses or voltage sweeps. In the case of metal–oxide-based RRAM devices, the switching mechanism has been recognized to be a highly localized phenomenon, where a conductive filament is alternatively formed and destroyed (at least partially) within the dielectric layer. Several physical interpretations for the switching processes have been proposed, including trap charging in the dielectric, space-charge-limited conduction processes, ion conduction and electrodeposition, Mott transition, and Joule heating. Such a large variety of the proposed physical mechanisms can be explained in part by the different dielectric and electrode materials and by the different procedures used in the experiments (unipolar or bipolar experiments). This aspect represents a limit today for the cell behavior understanding and a comprehensive physical picture of the programming behavior in RRAM device is still to be developed. This device results to be highly scalable, but limited by the size of select transistor in cell architecture. Another drawback is due to the high voltage necessary to create the conductive filament the first time to switch from the pristine state in a conductive state. This “first programming” operation has to be performed during the manufacturing process, thus increasing the fabrication complexity.

1.6.4 Phase Change Random Access Memory (PCRAM)

PCRAM devices use the resistivity difference between the amorphous and the crystalline states of chalcogenide glass (the most commonly used compound is $Ge_2Sb_2Te_5$, or GST) to store the logic ONE and logic ZERO levels. The device consists of a top electrode, the chalcogenide phase change layer, and a bottom electrode. The leakage path is cut off by an access (selection) transistor (or diode) in series with the phase change element. The phase change write operation consists of: (1) RESET, for which the chalcogenide glass is momentarily melted by a short electric pulse and then quickly quenched into an amorphous solid with high resistivity, and (2) SET, for which a lower amplitude but longer pulse (usually

>100 ns) anneals the amorphous phase into a low resistance crystalline state. The 1T-1R (or 1D-1R) cell is larger or smaller than NOR flash, depending on whether MOSFET or BJT (or diode) is used and the device may be programmed to any final state without erasing the previous state, which provides substantially faster programming throughput. The simple resistor structure and the low voltage operation also make PCRAM attractive for embedded NVM applications [ITRS '12]. The major challenges for PCRAM are the high current required to reset the phase change element and the relatively long set time. Interaction of phase change material with electrodes may pose long-term reliability issues and limit the cycling endurance. This is a major challenge for DRAM-like applications. Because PCRAM does not need to operate in page mode (no need to erase), it is a true random access, bit alterable memory like DRAM. The scalability of the PCRAM device to < 5 nm has been recently demonstrated using carbon nanotubes as electrodes [Feng '10] [Jiale '11] and the reset current followed the extrapolation line from larger devices. In at least one case, cycling endurance of 10^{11} was demonstrated [Kim '10].

1.7 Conclusion

In this chapter, we have presented the framework of this thesis. In the first part the economic context, the classification and the architectures of semiconductors memory were presented. Then, the Flash floating gate memory cell was described as well as its capacitive model that characterizes this device. Furthermore, the main program/erase mechanisms implemented in memory arrays are explained highlighting the importance of channel hot electron programming operation and the Fowler-Nordheim erasing for this thesis work. We thus presented the flash scaling limits and the proposed solutions; we explained the advantages of using a charge trapping layer instead of the continuous floating gate and a high-k control dielectric instead of the classical silicon oxide. Finally, we introduced the silicon nanocrystal memory cell that is the central point of this thesis. In particular we reported the state of the art of charge trap silicon nanocrystal memory, listing the various trials performed in the past. We introduced the impact on cell performances and reliability of some technical parameters: silicon nanocrystal size and density, control oxide, and the cell active shape. The object of this thesis will be to find the best tradeoff between some technological parameters, in order to optimize the programming window, the reliability and the energy consumption of our silicon nanocrystal cell.

Bibliography of chapter 1

[Baker '12] K. Baker, *"Embedded Nonvolatile Memories: A Key Enabler for Distributed Intelligence"*, in Memory Workshop (IMW), 2012 4th IEEE International, 20-23 May 2012, pp. 1-4, http://dx.doi.org/10.1109/imw.2012.6213637

[Baron '04] T. Baron, F. Mazen, J. M. Hartmann, P. Mur, R. A. Puglisi, S. Lombardo, G. Ammendola, and C. Gerardi, *"Growth and characterization of LPCVD Si quantum dots on insulators"*, Solid-State Electronics, vol. 48, 2004, pp. 1503-1509, http://dx.doi.org/10.1016/j.sse.2004.03.015

[Belgal '02] H. P. Belgal, N. Righos, I. Kalastirsky, J. J. Peterson, R. Shiner, and N. Mielke, *"A new reliability model for post-cycling charge retention of flash memories"*, in Reliability Physics Symposium Proceedings, 2002. 40th Annual, 2002, pp. 7-20, http://dx.doi.org/10.1109/relphy.2002.996604

[Bez '11] R. Bez, "Non-Volatile Memory Technology trends", presented at the Micro and Nano-Electronics 2DAYs, 2011,

[Blomme '09] P. Blomme and J. Van Houdt, *"Scalability of Fully Planar NAND Flash Memory Arrays Below 45nm"*, in Memory Workshop, 2009. IMW '09. IEEE International, 10-14 May 2009, pp. 1-2, http://dx.doi.org/10.1109/imw.2009.5090607

[Brews '80] J. R. Brews, W. Fichtner, E. H. Nicollian, and S. M. Sze, *"Generalized guide for MOSFET miniaturization"*, Electron Device Letters, IEEE, vol. 1, 1980, pp. 2-4, http://dx.doi.org/10.1109/EDL.1980.25205

[Buckley '06] J. Buckley, Ph.D. Thesis, *"Etude de memoires Flash integrant des dielectriaues high-k en tant qu'oxyde tunnel ou couche de stockage"*, Ecole Doctorale Electronique, Electrotechnique, Automatique et Traitement du Signal, Universitè de Grenoble

[Cappelletti '99] P. Cappelletti, C. Golla, P. Olivo, and E. Zanoni, "*Flash Memories*" Kluwer Academic Publishers.

[Casperson '02] J. D. Casperson, L. D. Bell, and H. A. Atwater, *"Materials issues for layered tunnel barrier structures"*, Journal of Applied Physics, vol. 92, 2002, pp. 261-267, http://dx.doi.org/10.1063/1.1479747

[Chang '83] C. Chang, M. Liang, x, S, C. Hu, and R. W. Brodersen, *"Carrier tunneling related phenomena in thin oxide MOSFET's"*, in Electron Devices Meeting, 1983 International, 1983, pp. 194-197, http://dx.doi.org/10.1109/iedm.1983.190474

[Chang Hyun '03] L. Chang Hyun, C. Kyung In, M. K. Cho, S. Yun Heub, P. Kyu-Charn, and K. Kinam, *"A novel SONOS structure of SiO2/SiN/Al2O3 with TaN metal gate for multi-giga bit flash memories "*, in Electron Devices Meeting, 2003. IEDM '03 Technical Digest. IEEE International, 8-10 Dec. 2003, pp. 26.5.1-26.5.4, http://dx.doi.org/10.1109/iedm.2003.1269356

[Changhyun '06] K. Changhyun, *"Future Memory Technology Trends and Challenges"*, in Quality Electronic Design, 2006. ISQED '06. 7th International Symposium on, 27-29 March 2006, pp. 513-513, http://dx.doi.org/10.1109/isqed.2006.69

[Chenming '85] H. Chenming, C. T. Simon, H. Fu-Chieh, K. Ping-Keung, C. Tung-Yi, and K. W. Terrill, *"Hot-electron-induced MOSFET degradation - Model, monitor, and improvement"*, Electron Devices, IEEE Transactions on, vol. 32, 1985, pp. 375-385, http://dx.doi.org/10.1109/T-ED.1985.21952

[Chimenton '01] A. Chimenton, P. Pellati, and P. Olivo, *"Analysis of erratic bits in flash memories"*, Device and Materials Reliability, IEEE Transactions on, vol. 1, 2001, pp. 179-184, http://dx.doi.org/10.1109/7298.995831

[Compagnoni '03] C. M. Compagnoni, D. Ielmini, A. S. Spinelli, A. L. Lacaita, C. Previtali, and C. Gerardi, *"Study of data retention for nanocrystal Flash memories"*, in Reliability Physics Symposium Proceedings, 2003. 41st Annual. 2003 IEEE International, 30 March-4 April 2003, pp. 506-512, http://dx.doi.org/10.1109/relphy.2003.1197800

[Crupi '03] I. Crupi, D. Corso, G. Ammendola, S. Lombardo, C. Gerardi, B. DeSalvo, G. Ghibaudo, E. Rimini, and M. Melanotte, *"Peculiar aspects of nanocrystal memory cells: data and extrapolations"*, Nanotechnology, IEEE Transactions on, vol. 2, 2003, pp. 319-323, http://dx.doi.org/10.1109/TNANO.2003.820515

[De Blauwe '00] J. De Blauwe, M. Ostraat, M. L. Green, G. Weber, T. Sorsch, A. Kerber, F. Klemens, R. Cirelli, E. Ferry, J. L. Grazul, F. Baumann, Y. Kim, W. Mansfield, J. Bude, J. T. C. Lee, S. J. Hillenius, R. C. Flagan, and H. A. Atwater, *"A novel, aerosol-nanocrystal floating-gate device for non-volatile memory applications"*, in Electron Devices Meeting, 2000. IEDM '00. Technical Digest. International, 2000, pp. 683-686, http://dx.doi.org/10.1109/iedm.2000.904411

[De Salvo '01] B. De Salvo, G. Ghibaudo, G. Pananakakis, P. Masson, T. Baron, N. Buffet, A. Fernandes, and B. Guillaumot, *"Experimental and theoretical investigation of nano-crystal and nitride-trap memory devices"*, Electron Devices, IEEE Transactions on, vol. 48, 2001, pp. 1789-1799, http://dx.doi.org/10.1109/16.936709

[De Salvo '03] B. De Salvo, C. Gerardi, S. Lombardo, T. Baron, L. Perniola, D. Mariolle, P. Mur, A. Toffoli, M. Gely, M. N. Semeria, S. Deleonibus, G. Ammendola, V. Ancarani, M. Melanotte, R. Bez, L. Baldi, D. Corso, I. Crupi, R. A. Puglisi, G. Nicotra, E. Rimini, F. Mazen, G. Ghibaudo, G. Pananakakis, C. M. Compagnoni, D. Ielmini, A. Lacaita, A. Spinelli, Y. M. Wan, and K. van der Jeugd, *"How far will silicon nanocrystals push the scaling limits of NVMs technologies?"*, in Electron Devices Meeting, 2003. IEDM '03 Technical Digest. IEEE International, 8-10 Dec. 2003, pp. 26.1.1-26.1.4, http://dx.doi.org/10.1109/iedm.2003.1269352

[Eun-Seok '11] C. Eun-Seok, Y. Hyun-Seung, J. Han-Soo, C. Gyu-Seog, P. Sung-Kye, and L. Seok-Kiu, *"A Novel 3D Cell Array Architecture for Terra-Bit NAND Flash Memory"*, in Memory Workshop (IMW), 2011 3rd IEEE International, 22-25 May 2011, pp. 1-4, http://dx.doi.org/10.1109/imw.2011.5873207

[Feng '10] X. Feng, A. Liao, D. Estrada, and E. Pop, *"Ultra-low power phase change memory with carbon nanotube interconnects"*, in Device Research Conference (DRC), 2010, 21-23 June 2010, pp. 253-254, http://dx.doi.org/10.1109/drc.2010.5551972

[Fichtner '80] W. Fichtner, E. N. Fuls, R. L. Johnston, T. T. Sheng, and R. K. Watts, *"Experimental and theoretical characterization of submicron MOSFETs"*, in Electron Devices Meeting, 1980 International, 1980, pp. 24-27, http://dx.doi.org/10.1109/iedm.1980.189743

[Fowler '28] R. H. Fowler and L. Nordheim, *"Electron Emission in Intense Electric Fields"*, Proceedings of the Royal Society of London. Series A, vol. 119, May 1, 19281928, pp. 173-181, http://dx.doi.org/10.1098/rspa.1928.0091

[Frank '99] D. J. Frank, T. Yuan, I. Meikei, and H. S. P. Wong, *"Monte Carlo modeling of threshold variation due to dopant fluctuations"*, in VLSI Technology, 1999. Digest of Technical Papers. 1999 Symposium on, 1999, pp. 169-170, http://dx.doi.org/10.1109/vlsit.1999.799397

[Frohman-Bentchkowsky '71]D. Frohman-Bentchkowsky, *"A fully-decoded 2048-bit electrically-programmable MOS ROM"*, in Solid-State Circuits Conference. Digest of Technical Papers. 1971 IEEE International, Feb 1971, pp. 80-81, http://dx.doi.org/10.1109/isscc.1971.1154978

[Fukuma '77] M. Fukuma and M. Matsumura, *"A simple model for short channel MOSFET's"*, Proceedings of the IEEE, vol. 65, 1977, pp. 1212-1213, http://dx.doi.org/10.1109/PROC.1977.10676

[Gasquet '06] H. P. Gasquet, R. F. Steimle, R. A. Rao, and R. Muralidhar, *"Impact of Nanocrystal Size on NVM Array Reliability Performance"*, in Non-Volatile Semiconductor Memory Workshop, 2006. IEEE NVSMW 2006. 21st, 2006, pp. 64-65, http://dx.doi.org/10.1109/.2006.1629496

[Gerardi '04] C. Gerardi, B. DeSalvo, S. Lombardo, and T. Baron, *"Performances of Si nanocrystal memories obtained by by CVD and their potentialities to further scaling of non-volatile memories"*, in Integrated Circuit Design and Technology, 2004. ICICDT '04. International Conference on, 2004, pp. 37-43, http://dx.doi.org/10.1109/icicdt.2004.1309902

[Gerardi '07a] C. Gerardi, V. Ancarani, R. Portoghese, S. Giuffrida, M. Bileci, G. Bimbo, O. Brafa, D. Mello, G. Ammendola, E. Tripiciano, R. Puglisi, and S. A. Lombardo, *"Nanocrystal Memory Cell Integration in a Stand-Alone 16-Mb NOR Flash Device"*, Electron Devices, IEEE Transactions on, vol. 54, 2007a, pp. 1376-1383, http://dx.doi.org/10.1109/TED.2007.895868

[Gerardi '07b] C. Gerardi, S. Lombardo, G. Ammendola, G. Costa, V. Ancarani, D. Mello, S. Giuffrida, and M. C. Plantamura, *"Study of nanocrystal memory integration in a Flash-like NOR device"*, Microelectronics Reliability, vol. 47, 2007b, pp. 593-597, http://dx.doi.org/10.1016/j.microrel.2007.01.024

[Gerardi '08] C. Gerardi, G. Molas, G. Albini, E. Tripiciano, M. Gely, A. Emmi, O. Fiore, E. Nowak, D. Mello, M. Vecchio, L. Masarotto, R. Portoghese, B. De Salvo, S. Deleonibus, and A. Maurelli, *"Performance and reliability of a 4Mb Si nanocrystal NOR Flash memory with optimized 1T memory cells"*, in Electron Devices Meeting, 2008. IEDM 2008. IEEE International, 15-17 Dec. 2008, pp. 1-4, http://dx.doi.org/10.1109/iedm.2008.4796823

[Hanafi '96] H. I. Hanafi, S. Tiwari, and I. Khan, *"Fast and long retention-time nano-crystal memory"*, Electron Devices, IEEE Transactions on, vol. 43, 1996, pp. 1553-1558, http://dx.doi.org/10.1109/16.535349

[Hang-Ting '05] L. Hang-Ting, W. Szu-Yu, L. Erh-Kun, S. Yen-Hao, L. Sheng-Chih, Y. Ling-Wu, C. Kuang-Chao, J. Ku, H. Kuang-Yeu, L. Rich, and L. Chih-Yuan, *"BE-SONOS: A bandgap engineered SONOS with excellent performance and reliability"*, in Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International, 5-5 Dec. 2005, pp. 547-550, http://dx.doi.org/10.1109/iedm.2005.1609404

[Hidaka '11] H. Hidaka, *"Evolution of embedded flash memory technology for MCU"*, in IC Design & Technology (ICICDT), 2011 IEEE International Conference on, 2-4 May 2011, pp. 1-4, http://dx.doi.org/10.1109/icicdt.2011.5783209

[Hoefler '02] A. Hoefler, J. M. Higman, T. Harp, and P. J. Kuhn, *"Statistical modeling of the program/erase cycling acceleration of low temperature data retention in floating gate nonvolatile memories"*, in Reliability Physics Symposium Proceedings, 2002. 40th Annual, 2002, pp. 21-25, http://dx.doi.org/10.1109/relphy.2002.996605

[Hong '07] Y. K. Hong, D. J. Jung, Y. M. Kang, H. S. Kim, W. W. Jung, H. K. Koh, J. H. Park, D. Y. Choi, S. E. Kim, W. S. Ann, H. H. Kim, J. H. Kim, W. U. Jung, E. S. Lee, S. Y. Lee, H. S. Jeong, and K. Kim, *"130 nm-technology, 0.25 μm2, 1T1C FRAM cell for SoC (system-on-a-chip)-friendly applications"*, in VLSI Technology, 2007 IEEE Symposium on, 12-14 June 2007, pp. 230-231, http://dx.doi.org/10.1109/vlsit.2007.4339704

[ITRS '12] ITRS, "INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS", Process integration, devices, and structures. Technical report 2012, http://www.itrs.net/Links/2012ITRS/Home2012.htm 2012.

[Jacob '07] S. Jacob, G. Festes, S. Bodnar, R. Coppard, J. F. Thiery, T. Pate-Cazal, T. Pedron, B. De Salvo, L. Perniola, E. Jalaguier, F. Boulanger, and S. Deleonibus, *"Integration of CVD silicon nanocrystals in a 32Mb NOR flash memory"*, in Solid State Device Research Conference, 2007. ESSDERC 2007. 37th European, 11-13 Sept. 2007, pp. 410-413, http://dx.doi.org/10.1109/essderc.2007.4430965

[Jacob '08] S. Jacob, B. De Salvo, L. Perniola, G. Festes, S. Bodnar, R. Coppard, J. F. Thiery, T. Pate-Cazal, C. Bongiorno, S. Lombardo, J. Dufourcq, E. Jalaguier, T. Pedron, F. Boulanger, and S. Deleonibus, *"Integration of CVD silicon nanocrystals in a 32 Mb NOR flash memory"*, Solid-State Electronics, vol. 52, 2008, pp. 1452-1459, http://dx.doi.org/10.1016/j.sse.2008.04.032

[Jiale '11] L. Jiale, R. G. D. Jeyasingh, C. Hong-Yu, and H. S. P. Wong, *"A 1.4uA Reset Current Phase Change Memory Cell with Integrated Carbon Nanotube Electrodes for Cross-Point Memory Application"*, in VLSI Technology (VLSIT), 2011 Symposium on, 14-16 June 2011, pp. 100-101,

[Jiyoung '09] K. Jiyoung, A. J. Hong, S. Min Kim, E. B. Song, P. Jeung Hun, H. Jeonghee, C. Siyoung, J. Deahyun, M. Joo -Tae, and K. L. Wang, *"Novel Vertical-Stacked-Array-Transistor (VSAT) for ultra-high-density and cost-effective NAND Flash memory devices and SSD (Solid State Drive)"*, in VLSI Technology, 2009 Symposium on, 16-18 June 2009, pp. 186-187,

[Kahng '67] K. Kahng and S. M. Sze, *"A floating gate and its application to memory devices"*, Bell Syst. Tech. J., vol. 46, 1967, p. 1288, http://dx.doi.org/10.1109/T-ED.1967.16028

[Kato '94] M. Kato, N. Miyamoto, H. Kume, A. Satoh, T. Adachi, M. Ushiyama, and K. Kimura, *"Read-disturb degradation mechanism due to electron trapping in the tunnel oxide for low-voltage flash memories"*, in Electron Devices Meeting, 1994. IEDM '94. Technical Digest., International, 11-14 Dec 1994, pp. 45-48, http://dx.doi.org/10.1109/iedm.1994.383470

[Kim '10] I. S. Kim, S. L. Cho, D. H. Im, E. H. Cho, D. H. Kim, G. H. Oh, D. H. Ahn, S. O. Park, S. W. Nam, J. T. Moon, and C. H. Chung, *"High performance PRAM cell scalable to sub-20nm technology with below 4F2 cell size, extendable to DRAM applications"*, in VLSI Technology (VLSIT), 2010 Symposium on, 15-17 June 2010, pp. 203-204, http://dx.doi.org/10.1109/vlsit.2010.5556228

[Kim '05] K. Kim, *"Technology for sub-50nm DRAM and NAND flash manufacturing"*, in Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International, 5-5 Dec. 2005, pp. 323-326, http://dx.doi.org/10.1109/iedm.2005.1609340

[Lai '05] C. H. Lai, A. Chin, K. C. Chiang, W. J. Yoo, C. F. Cheng, S. P. McAlister, C. C. Chi, and P. Wua, *"Novel SiO2/AlN/HfAIO/IrO2 memory with fast erase, large ΔVth and good retention "*, in VLSI Technology, 2005. Digest of Technical Papers. 2005 Symposium on, 14-16 June 2005, pp. 210-211, http://dx.doi.org/10.1109/.2005.1469271

[Leroux '04] C. Leroux, J. Mitard, G. Ghibaudo, X. Garros, G. Reimbold, B. Guillaumor, and F. Martin, *"Characterization and modeling of hysteresis phenomena in high K dielectrics"*, in Electron Devices Meeting, 2004. IEDM Technical Digest. IEEE International, 13-15 Dec. 2004, pp. 737-740, http://dx.doi.org/10.1109/iedm.2004.1419276

[Ling-Chang '06] H. Ling-Chang, K. An-Chi, J. R. Shih, L. Yao-Feng, K. Wu, and K. Ya-Chin, *"Statistical modeling for postcycling data retention of split-gate flash*

memories", Device and Materials Reliability, IEEE Transactions on, vol. 6, 2006, pp. 60-66, http://dx.doi.org/10.1109/TDMR.2006.870354

[Masoero '11] L. Masoero, G. Molas, F. Brun, M. Gely, J. P. Colonna, V. Della Marca, O. Cueto, E. Nowak, A. De Luca, P. Brianceau, C. Charpin, R. Kies, A. Toffoli, D. Lafond, V. Delaye, F. Aussenac, C. Carabasse, S. Pauliac, C. Comboroure, G. Ghibaudo, S. Deleonibus, and B. De Salvo, *"Scalability of split-gate charge trap memories down to 20nm for low-power embedded memories"*, in Electron Devices Meeting (IEDM), 2011 IEEE International, 5-7 Dec. 2011, pp. 9.5.1-9.5.4, http://dx.doi.org/10.1109/iedm.2011.6131522

[Masoero '12a] L. Masoero, Ph.D. Thesis, *"Etude d'architectures et d'empilements innovants de mémoires Split-Gate (grille séparée) à couche de piégeage discret"*, Université de Grenoble,

[Masoero '12b] L. Masoero, G. Molas, V. Della Marca, M. Gely, O. Cueto, J. P. Colonna, A. De Luca, P. Brianceau, C. Charpin, D. Lafond, V. Delaye, F. Aussenac, C. Carabasse, S. Pauliac, C. Comboroure, P. Boivin, G. Ghibaudo, S. Deleonibus, and B. De Salvo, *"Physical Understanding of Program Injection and Consumption in Ultra-Scaled SiN Split-Gate Memories"*, in Memory Workshop (IMW), 2012 4th IEEE International, 20-23 May 2012, pp. 1-4, http://dx.doi.org/10.1109/imw.2012.6213686

[Masuoka '87] F. Masuoka, M. Momodomi, Y. Iwata, and R. Shirota, *"New ultra high density EPROM and flash EEPROM with NAND structure cell"*, in Electron Devices Meeting, 1987 International, 1987, pp. 552-555, http://dx.doi.org/10.1109/iedm.1987.191485

[Mazen '03] F. Mazen, T. Baron, A. M. Papon, R. Truche, and J. M. Hartmann, *"A two steps CVD process for the growth of silicon nano-crystals"*, Applied Surface Science, vol. 214, 2003, pp. 359-363, http://dx.doi.org/10.1016/S0169-4332(03)00521-X

[Mazen '04] F. Mazen, L. Mollard, T. Baron, S. Decossas, and J. M. Hartmann, *"Preferential nucleation of silicon nano-crystals on electron beam exposed SiO2 surfaces"*, Microelectronic Engineering, vol. 73–74, 2004, pp. 632-638, http://dx.doi.org/10.1016/j.mee.2004.03.005

[Miura '07] K. Miura, T. Kawahara, R. Takemura, J. Hayakawa, S. Ikeda, R. Sasaki, H. Takahashi, H. Matsuoka, and H. Ohno, *"A novel SPRAM (SPin-transfer torque RAM) with a synthetic ferrimagnetic free layer for higher immunity to read disturbance and reducing write-current dispersion"*, in VLSI Technology, 2007 IEEE Symposium on, 12-14 June 2007, pp. 234-235, http://dx.doi.org/10.1109/vlsit.2007.4339706

[Molas '10] G. Molas, J. P. Colonna, R. Kies, D. Belhachemi, M. Bocquet, M. Gely, V. Vidal, P. Brianceau, L. Vandroux, G. Ghibaudo, and B. De Salvo, *"Investigation of charge-trap memories with AlN based band engineered storage layers"*, in Memory Workshop (IMW), 2010 IEEE International, 16-19 May 2010, pp. 1-4, http://dx.doi.org/10.1109/imw.2010.5488309

[Monzio Compagnoni '04] C. Monzio Compagnoni, D. Ielmini, A. S. Spinelli, A. L. Lacaita, and C. Gerardi, *"Study of nanocrystal memory reliability by CAST structures"*, Solid-State Electronics, vol. 48, 2004, pp. 1497-1502, http://dx.doi.org/10.1016/j.sse.2004.03.014

[Muralidhar '03] R. Muralidhar, R. F. Steimle, M. Sadd, R. Rao, C. T. Swift, E. J. Prinz, J. Yater, L. Grieve, K. Harber, B. Hradsky, S. Straub, B. Acred, W. Paulson, W. Chen, L. Parker, S. G. H. Anderson, M. Rossow, T. Merchant, M. Paransky, T. Huynh, D. Hadad, C. Ko-Min, and B. E. White, Jr., *"A 6 V embedded 90 nm silicon nanocrystal nonvolatile memory"*, in Electron Devices Meeting, 2003. IEDM '03

Technical Digest. IEEE International, 8-10 Dec. 2003, pp. 26.2.1-26.2.4, http://dx.doi.org/10.1109/iedm.2003.1269353

[Muralidhar '04] R. Muralidhar, R. F. Steimle, M. Sadd, R. Rao, C. T. Swift, E. J. Prinz, J. Yater, L. Grieve, K. Harber, B. Hradsky, S. Straub, B. Acred, W. Paulson, W. Chen, L. Parker, S. G. H. Anderson, M. Rossow, T. Merchant, M. Paransky, T. Huynh, D. Hadad, C. Ko-Min, and B. E. White, Jr., *"An embedded silicon nanocrystal nonvolatile memory for the 90nm technology node operating at 6V"*, in Integrated Circuit Design and Technology, 2004. ICICDT '04. International Conference on, 2004, pp. 31-35, http://dx.doi.org/10.1109/icicdt.2004.1309900

[Ning '78] T. H. Ning, *"Hot-electron emission from silicon into silicon dioxide"*, Solid-State Electronics, vol. 21, 1978, pp. 273-282, http://dx.doi.org/10.1016/0038-1101(78)90148-X

[Pavan '97] P. Pavan, R. Bez, P. Olivo, and E. Zanoni, *"Flash memory cells-an overview"*, Proceedings of the IEEE, vol. 85, 1997, pp. 1248-1271, http://dx.doi.org/10.1109/5.622505

[Philip Wong '08] H. S. Philip Wong, "Emerging memories", ed: Stanford University, Stanford, California, USA, 2008, http://asia.stanford.edu/events/spring08/slides402S/0403-Wong.pdf

[Ping-Hung '08] T. Ping-Hung, C.-L. Kuei-Shu, L. Chu-Yung, W. Tien-Ko, P. J. Tzeng, C. H. Lin, L. S. Lee, and M. J. Tsai, *"Novel SONOS-Type Nonvolatile Memory Device With Optimal Al Doping in HfAlO Charge-Trapping Layer"*, Electron Device Letters, IEEE, vol. 29, 2008, pp. 265-268, http://dx.doi.org/10.1109/LED.2007.915380

[Prall '10] K. Prall and K. Parat, *"25nm 64Gb MLC NAND technology and scaling challenges invited paper"*, in Electron Devices Meeting (IEDM), 2010 IEEE International, 6-8 Dec. 2010, pp. 5.2.1-5.2.4, http://dx.doi.org/10.1109/iedm.2010.5703300

[Rao '05] R. A. Rao, H. P. Gasquet, R. F. Steimle, G. Rinkenberger, S. Straub, R. Muralidhar, S. G. H. Anderson, J. A. Yater, J. C. Ledezma, J. Hamilton, B. Acred, C. T. Swift, B. Hradsky, J. Peschke, M. Sadd, E. J. Prinz, K. M. Chang, and B. E. White Jr, *"Influence of silicon nanocrystal size and density on the performance of non-volatile memory arrays"*, Solid-State Electronics, vol. 49, 2005, pp. 1722-1727, http://dx.doi.org/10.1016/j.sse.2005.10.022

[Robertson '05] J. Robertson, *"Interfaces and defects of high-K oxides on silicon"*, Solid-State Electronics, vol. 49, 2005, pp. 283-293, http://dx.doi.org/10.1016/j.sse.2004.11.011

[Rosmeulen '02] M. Rosmeulen, E. Sleeckx, and K. De Meyer, *"Silicon-rich-oxides as an alternative charge-trapping medium in Fowler-Nordheim and hot carrier type non-volatile-memory cells"*, in Electron Devices Meeting, 2002. IEDM '02. International, 8-11 Dec. 2002, pp. 189-192, http://dx.doi.org/10.1109/iedm.2002.1175810

[Russo '09] U. Russo, D. Ielmini, C. Cagli, and A. L. Lacaita, *"Filament Conduction and Reset Mechanism in NiO-Based Resistive-Switching Memory (RRAM) Devices"*, Electron Devices, IEEE Transactions on, vol. 56, 2009, pp. 186-192, http://dx.doi.org/10.1109/TED.2008.2010583

[Simon '84] T. Simon, K. Ping-Keung, and H. Chenming, *"Lucky-electron model of channel hot-electron injection in MOSFET'S"*, Electron Devices, IEEE Transactions on, vol. 31, 1984, pp. 1116-1125, http://dx.doi.org/10.1109/T-ED.1984.21674

[Steimle '04] R. F. Steimle, R. Rao, M. Sadd, C. Swift, B. Hradsky, S. Straub, T. Merchant, M. Stoker, C. Parikh, S. Anderson, M. Rossow, J. Yater, B. Acred, K. Harber, E. Prinz, B. E. White, Jr., and R. Muralidhar, *"Silicon nanocrystals: from Coulomb*

blockade to memory arrays", in Nanotechnology, 2004. 4th IEEE Conference on, 16-19 Aug. 2004, pp. 290-292, http://dx.doi.org/10.1109/nano.2004.1392328

[Strenz '11] R. Strenz, *"Embedded Flash Technologies and their Applications: Status & Outlook"*, in Electron Devices Meeting, 2011. IEDM 2011. IEEE International, Dec. 2011, pp. 1-4,

[Sung-Taeg '08] K. Sung-Taeg, J. Yater, H. CheongMin, J. Shen, N. Ellis, M. Herrick, H. Gasquet, W. Malloch, and G. Chindalore, *"Si Nanocrystal Split Gate Technology Optimization for High Performance and Reliable Embedded Microcontroller Applications"*, in Non-Volatile Semiconductor Memory Workshop, 2008 and 2008 International Conference on Memory Technology and Design. NVSMW/ICMTD 2008. Joint, 18-22 May 2008, pp. 59-60, http://dx.doi.org/10.1109/nvsmw.2008.23

[Sung-Taeg '12] K. Sung-Taeg, B. Winstead, J. Yater, M. Suhail, G. Zhang, C. M. Hong, H. Gasquet, D. Kolar, J. Shen, B. Min, K. Loiko, A. Hardell, E. LePore, R. Parks, R. Syzdek, S. Williams, W. Malloch, G. Chindalore, Y. Chen, Y. Shao, L. Huajun, L. Louis, and S. Chwa, *"High Performance Nanocrystal Based Embedded Flash Microcontrollers with Exceptional Endurance and Nanocrystal Scaling Capability"*, in Memory Workshop (IMW), 2012 4th IEEE International, 20-23 May 2012, pp. 1-4, http://dx.doi.org/10.1109/imw.2012.6213668

[SungJin '10] W. SungJin, L. KiHong, S. DaeGyu, K. BeomYong, K. MinSoo, B. JinHo, H. JiHye, K. SungJun, L. BoMi, J. YoungKyun, C. SungYoon, S. ChangHee, Y. HyunSeung, C. SangMoo, H. Kwon, S. Aritome, P. SungKi, and H. Sungjoo, *"Novel 3-dimensional Dual Control-gate with Surrounding Floating-gate (DC-SF) NAND flash cell for 1Tb file storage application"*, in Electron Devices Meeting (IEDM), 2010 IEEE International, 6-8 Dec. 2010, pp. 29.7.1-29.7.4, http://dx.doi.org/10.1109/iedm.2010.5703447

[Tae-Su '09] J. Tae-Su, K. Joong-Sik, H. Sang-Min, O. Young-Hoon, R. Kwang-Myung, C. Seoung-Ju, C. Su-Ock, J.-G. Oh, S. Bhardwaj, K. Jungtae, K. Yong-Taik, M. Nagoga, C. Seon-Yong, S.-C. Moon, S.-W. Chung, S.-J. Hong, and P. Sung-Wook, *"Highly scalable Z-RAM with remarkably long data retention for DRAM application"*, in VLSI Technology, 2009 Symposium on, 16-18 June 2009, pp. 234-235,

[Takeda '83] E. Takeda, Y. Nakagome, H. Kume, and S. Asai, *"New hot-carrier injection and device degradation in submicron MOSFETs"*, Solid-State and Electron Devices, IEE Proceedings I, vol. 130, 1983, pp. 144-150, http://dx.doi.org/10.1049/ip-i-1.1983.0026

[Takeda '85] E. Takeda, Y. Ohji, and H. Kume, *"High field effects in MOSFETS"*, in Electron Devices Meeting, 1985 International, 1985, pp. 60-63, http://dx.doi.org/10.1109/iedm.1985.190891

[Tiwari '95] S. Tiwari, F. Rana, K. Chan, H. Hanafi, C. Wei, and D. Buchanan, *"Volatile and non-volatile memories in silicon with nano-crystal storage"*, in Electron Devices Meeting, 1995. IEDM '95., International, 10-13 Dec 1995, pp. 521-524, http://dx.doi.org/10.1109/iedm.1995.499252

[Vandelli '11] L. Vandelli, A. Padovani, L. Larcher, G. Broglia, G. Ori, M. Montorsi, G. Bersuker, and P. Pavan, *"Comprehensive physical modeling of forming and switching operations in HfO2 RRAM devices"*, in Electron Devices Meeting (IEDM), 2011 IEEE International, 5-7 Dec. 2011, pp. 17.5.1-17.5.4, http://dx.doi.org/10.1109/iedm.2011.6131574

[Wang '05] Y. Q. Wang, P. K. Singh, W. J. Yoo, Y. C. Yeo, G. Samudra, A. Chin, W. S. Hwang, J. H. Chen, S. J. Wang, and D. L. Kwong, *"Long retention and low voltage operation using IrO/sub 2/HfAlO/HfSiO/HfalO gate stack for memory application"*, in

Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International, 5-5 Dec. 2005, pp. 162-165, http://dx.doi.org/10.1109/iedm.2005.1609295

[Wilk '01] G. D. Wilk, R. M. Wallace, and J. M. Anthony, *"High-κ gate dielectrics: Current status and materials properties considerations "*, Journal of Applied Physics, vol. 89, 2001, pp. 5243-5275, http://dx.doi.org/10.1063/1.1361065

[Wu '92] C.-Y. Wu and C.-F. Chen, *"Physical model for characterizing and simulating a FLOTOX EEPROM device"*, Solid-State Electronics, vol. 35, 1992, pp. 705-716, http://dx.doi.org/10.1016/0038-1101(92)90041-a

[Xuguang '04] W. Xuguang, L. Jun, B. Weiping, and K. Dim-Lee, *"A novel MONOS-type nonvolatile memory using high-κ dielectrics for improved data retention and programming speed "*, Electron Devices, IEEE Transactions on, vol. 51, 2004, pp. 597-602, http://dx.doi.org/10.1109/TED.2004.824684

[Yater '09] J. Yater, M. Suhail, S. T. Kang, J. Shen, C. Hong, T. Merchant, R. R. H. Gasquet, K. Loiko, B. Winstead, S. Williams, M. Rossow, W. Malloch, R. Syzdek, and G. Chindalore, *"16Mb Split Gate Flash Memory with Improved Process Window"*, in Memory Workshop, 2009. IMW '09. IEEE International, 10-14 May 2009, pp. 1-2, http://dx.doi.org/10.1109/imw.2009.5090570

[Yater '11] J. Yater, C. Hong, S. T. Kang, D. Kolar, B. Min, J. Shen, G. Chindalore, K. Loiko, B. Winstead, S. Williams, H. Gasquet, M. Suhail, K. Broeker, E. LePore, A. Hardell, W. Malloch, R. Syzdek, Y. Chen, Y. Ju, S. Kumarasamy, H. Liu, L. Lei, and B. Indajang, *"Highly Optimized Nanocrystal-Based Split Gate Flash for High Performance and Low Power Microcontroller Applications"*, in Memory Workshop (IMW), 2011 3rd IEEE International, 22-25 May 2011, pp. 1-4, http://dx.doi.org/10.1109/imw.2011.5873213

[Yau '74] L. D. Yau, *"A simple theory to predict the threshold voltage of short-channel IGFET's"*, Solid-State Electronics, vol. 17, 1974, pp. 1059-1063, http://dx.doi.org/10.1016/0038-1101(74)90145-2

[Yau '75] L. D. Yau, *"Simple I/V model for short-channel i.g.f.e.t.s in the triode region"*, Electronics Letters, vol. 11, 1975, pp. 44-45, http://dx.doi.org/10.1049/el:19750034

[Yoocheol '05]S. Yoocheol, C. Jungdal, K. Changseok, L. Changhyun, P. Ki-Tae, L. Jang-Sik, S. Jongsun, K. Viena, C. Byeongin, S. Jaesung, K. Dongchan, C. Hag Ju, and K. Kinam, *"A novel NAND-type MONOS memory using 63nm process technology for multi-gigabit flash EEPROMs"*, in Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International, 5-5 Dec. 2005, pp. 327-330, http://dx.doi.org/10.1109/iedm.2005.1609341

[Zajac '10] C. Zajac, *"Choose the right non-volatile memory IP"*, EE Times-Asia, 2010, http://www.eetindia.co.in/STATIC/PDF/201006/EEIOL_2010JUN01_STOR_TA_01.pdf

Chapter 2 - Electrical characterization of silicon nanocrystal memories

2.1 Introduction

In this section the results concerning the programming window electrical characterization are presented. The programming window of the silicon nanocrystal memory cell was measured using a defined experimental protocol developed in the STMicroelectronics-Rousset electrical characterization laboratory. With this procedure we evaluated the impact of main technological parameters on the programming window: silicon nanocrystal size and density (covered area), presence of silicon nitride capping layer, channel doping dose and tunnel oxide thickness. The results were compared to the state of the art in order to understand how to improve the cell performance using a CMOS process fully compatible with the existing STMicroelectronics method. The chapter will conclude with the benchmarking of silicon nanocrystal cell versus the standard Flash floating gate.

2.2 Experimental details

One of the main limits of silicon nanocrystal memories is the narrow programming window [Gerardi '08] [Monzio Compagnoni '03] [De Salvo '03]. In order to evaluate how to improve the memory cell performance, it is important to develop manual and automatic tests.

2.2.1 Experimental setup

The electrical characterization of the silicon nanocrystal cell was performed using manual and automatic probe stations. The first was used to measure the programming/erase kinetic characteristics, while the second was used to obtain statistical information concerning the dispersion on wafer. In figure 2.1 a picture of the manual test bench is shown.

- Manual prober
  - Thermo chuck equipped
- Switch matrix (optional)
  - To connect the instruments to the sample
- Tester
  - Electrical parameter analyser (HP4156)
  - LCR meter (HP4284)
  - Pulse Generator (HP8110)
- Computer (LABView system) to drive the prober and the tester

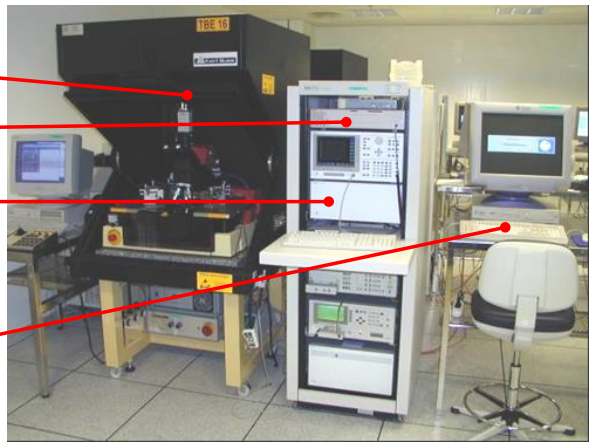


Figure 2. 1. Manual test bench.

The manual probe station is driven by the LABView system in order to command the instruments of the bench with *homemade* software. In particular the test bench is equipped with a HP4156 electrical parameter analyzer, a HP4284 LCR precision meter and a HP8110 pulse generator. The switch matrix enables the instruments to be connected to the 200mm wafer (sample). Using the LABView program we are able to measure the program/erase cell characteristics using different biasing conditions. The presence of thermo chuck enables measurement in the range of temperature from -40°C up to 250°C. Figure 2.2 shows the automatic probe bench system. The test program is developed in HP BASIC language and controlled by SIAM (Automatic Identification System of Model Parameters). It is a dedicated software for parametric tests. It was possible to load in the automatic prober up to 25 wafers to obtain statistical results. With an accurate prober calibration the station is able to probe the wafers by applying the same test program. The bench was equipped with a HP4142 electrical parameter analyzer, a HP4284 LCR precision meter and a HP8110 pulse generator.

- Automatic Prober
  - Wafer handling
  - Wafer alignment
- Matrix test head, probe card
  - Connect the instruments to the samples
- Tester
  - LCR meter (HP4284)
  - Pulse Generator (HP8110)
  - Parameter analyzer (HP4142)
- Computer (SIAM system) To drive the prober and the tester

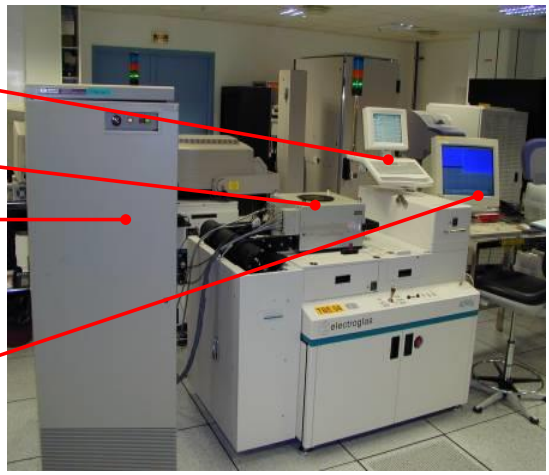


Figure 2. 2. Automatic test bench.

2.2.2 Methods of characterization

In order to characterize the programming window of the silicon nanocrystal cell and compare it with the characterizations obtained on standard Flash floating gate in NOR architecture, we used an appropriate method to program the cell by channel hot electron and to erase it by Fowler-Nordheim mechanism. The measurement protocol was divided into two parts and it was kept unchanged for all samples. The first part was performed using the automatic bench; by applying only one program/erase cycle we were able to evaluate the programming window dispersion on the whole wafer. The evaluation of programming window dispersion was performed using a fixed 5µs programming pulse, a gate voltage (Vg) of 9V, a drain voltage (Vd) of 4.2V, source and bulk voltages (Vs and Vb respectively) of 0V. Concerning the erase phase, a pulse of 90ms was applied on the gate terminal using Vg=-18V, while the drain, source and bulk terminals were grounded (Vd=Vs=Vb=0V). The second part of the experiments was performed using the manual prober station. Its purpose was to apply program/erase pulses with different durations and amplitudes to get the kinetic evolution of the threshold voltage in channel hot electron and Fowler-Nordheim regime. The two methods are described below. Figures 2.3(a) and 2.3(b) showthe signals used during the channel hot electron programming kinetic, and the Fowler-Nordheim erase kinetic, in that order.

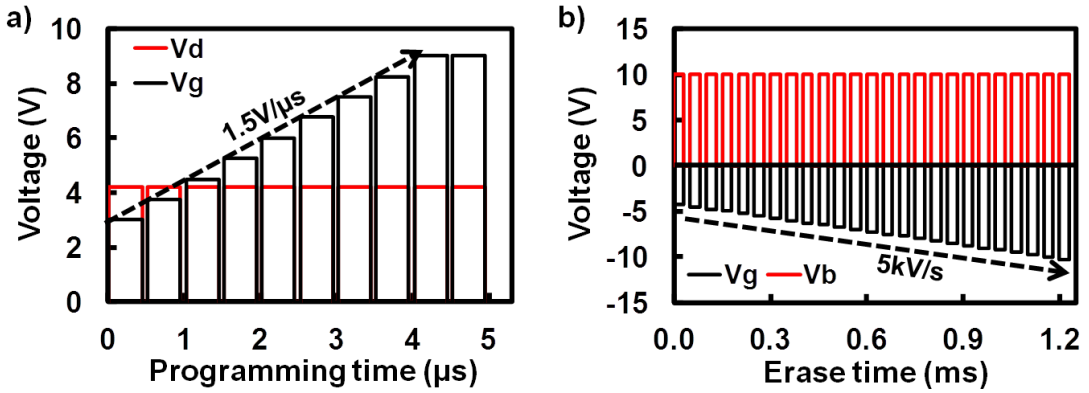


Figure 2. 3. Signals applied during a) the channel hot electron programming operation and b) Fowler-Nordheim erase operation.

The staircases have been used to emulate the ramps generated in STMicroelectronics products. The programming kinetic was performed by applying 4.2V pulses on drain terminal and a staircase from 3V to 9V with a step of 0.75V, followed by an additional pulse of 1µs on gate terminal. The duration of each pulse was 0.5µs in order to obtain a 1.5V/µs ramp. For the erase kinetic, 10V pulses were applied on drain, source and bulk terminals, while a staircase from -4V to -10V was applied on gate terminal. In this way it was possible to reach the gate-bulk voltage of 20V. This represents the maximum voltage value available in STMicroelectronics products. The step amplitude was 0.25V, while the duration was 50µs in order to emulate a 5kV/s ramp. After each pulse the cell state was read by measuring the gate voltage needed to drive a fixed drain current of 1µA. The programming window (ΔVt) is calculated as the difference between the programmed threshold voltage (Vtp) and the erased threshold voltage (Vte):

$$\Delta Vt = Vtp - Vte \qquad (1)$$

2.3 Impact of technological parameters

After the technical details described above, we are going to present the results of electrical characterization in terms of programming window. The aim is to understand the impact of main technological parameters on the programming window and how to merge them to obtain the best results with respect to the Flash floating gate.

2.3.1 Effect of silicon nanocrystal size

The impact of silicon nanocrystals size on the programming window has already been analyzed in depth [De Salvo '01] [Rao '05]. Nevertheless it is important to evaluate its effect in STMicroelectronics products. In [Jacob '08] it appears that the bigger the Si-NCs are, the

larger the programming window is. This is due to the increasing surface portion covered by the nanocrystals. In this work, the studied samples have a channel width of W=90nm and a channel length of L=180nm. The cell stack is described in figure 2.4.

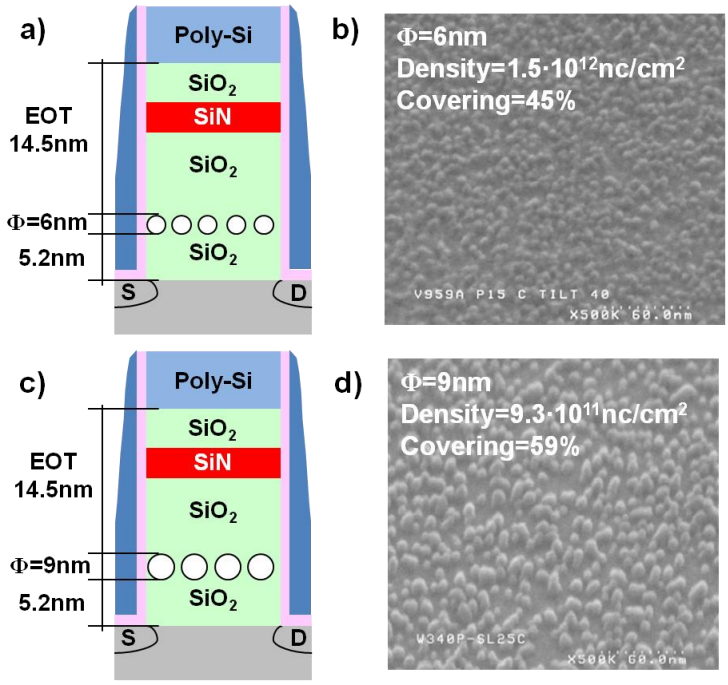


Figure 2. 4. Silicon nanocrystal cells used to evaluate the impact of nanocrystal size. (a) and (c) are the schematics of cells that integrated average 6nm and 9nm nanocrystal diameter. (b) and (d) are the corresponding 40° tilted CDSEM pictures used to measure the nanocrystal size, density.

On a p-type substrate a 5.2nm thick tunnel oxide was grown. The average silicon nanocrystal size and density have been measured in-line using a Critical Dimension Scanning Electron Microscopy (CDSEM) technique [Amouroux '12]. In this case we compare the samples with two different diameters (Φ) of 6nm and 9nm. Then, to complete the stack, the ONO (Oxide-Nitride-Oxide) Inter-Poly Dielectric (IPD) layer was deposited to reach 14.5nm of Equivalent Oxide Thickness (EOT). On a silicon nanocrystal cell it is not possible to measure capacitance to calculate the EOT of ONO thickness, because of the discrete nature of nanocrystals. The ONO thickness is thus considered to be unchanged with respect to the standard Flash floating gate cell, because the fabrication process and the recipe are unchanged for the two devices. As for the F.G. it was measured at the end of process with capacitance-voltage characterizations and Transmission Electron Microscopy analysis. Figure 2.5a shows the average values and the dispersion of program/erase threshold voltages obtained with the statistical measurements (30 samples). The minimum program/erase states to target the Flash floating gate are highlighted. With our extrapolation it is necessary to increase the nanocrystal diameter up to 14nm, maintaining the cell stack unchanged, in order

to achieve good programming window. Using the measured size and density (figure 2.4).we plot the correlation between the programming window and the percentage of covered area in figure 2.5b; increasing the covered area, thus the coupling factor, the programming window increases because of the higher number of trapped charges. Using this cell structure to achieve the minimum programming window of 4V, the 95% of covered area is needed. That is not coherent with the discrete nature of Si-nc cell, as detailed in section 1.4. Moreover, the programming window increases with the silicon nanocrystal size as well as the dispersion on wafer. Increasing the diameter by 3nm enables increasing the programming window of 0.5V that is not sufficient for our application.

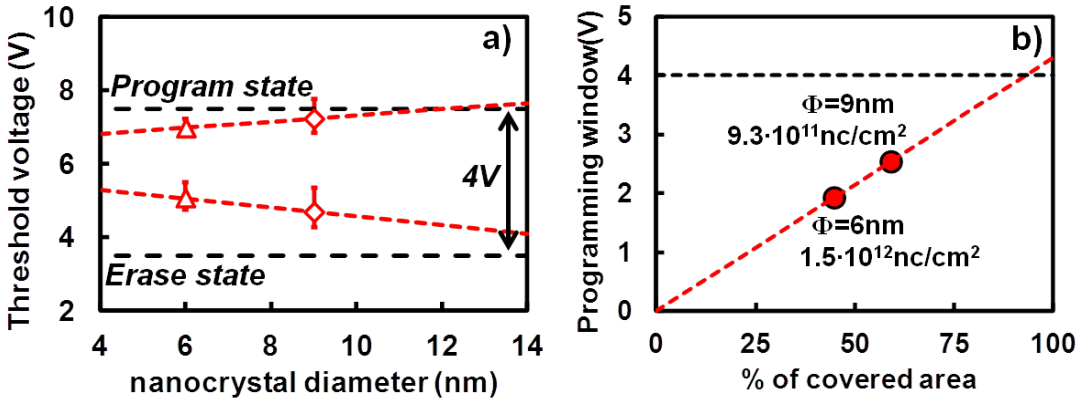


Figure 2. 5. a) Statistical results of program/erase threshold voltage measured for samples with different silicon nanocrystal sizes (Φ=6nm and Φ=9nm). b) Programming window as a function of covered area.

After these preliminary evaluations we used the staircases, described in paragraph 2.2, to measure the program/erase kinetic characteristics. In figure 2.6a the results concerning the considered samples are compared. As expected, we notice that by increasing the nanocrystal size, the programming window is improved because the covered area increases as well as the cell coupling factor.

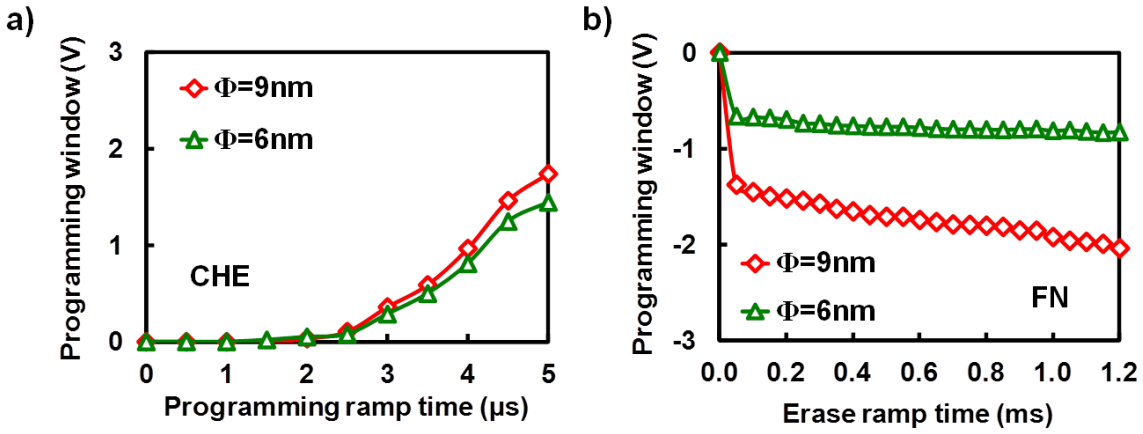


Figure 2. 6. a) Channel hot electron (CHE) programming kinetic and b) Fowler-Nordheim (FN) erase kinetic measured on samples with Φ=6nm and Φ=9nm. The program/erase pulses are described in section 2.2.

In literature it is shown that the FN erase speed can increase when the nanocrystal size decreases [Rao '05]. This is true for specific cell architectures and in particular for small nanocrystal diameters. However, when increasing the nanocrystal size and the covered area, the coulomb blockade effect and quantum mechanisms can be neglected [De Salvo '01], thus the coupling factor dependence becomes predominant. It is important to notice that the programming window of figures 2.5 and 2.6 cannot be compared because two different program/erase schemes are used. With the ramped gate voltage the program/erase efficiency decreases with respect to the box pulses used for the statistical measurements shown in figure 2.5. To conclude we can confirm that increasing the silicon nanocrystal size and thus, the covered area, the programming window is improved. In particular this effect is more evident on FN erase operation.

2.3.2 Effect of silicon nitride capping layer

In literature it is shown how to improve the programming window using high-k material interpoly dielectrics [Molas '07] instead of the ONO layer. One alternative solution is represented by the hybrid Si-nc cell. It is demonstrated that the capping layer on Si-ncs increases the number of trapping sites. Moreover it protects the nanocrystals from oxidation during the interpoly dielectric deposition [Steimle '03] [Colonna '08] [Chen '09]. In this paragraph we analyze the impact of silicon nitride (Si_3N_4) capping layer on the memory cell programming window. The studied samples are shown in figure 2.7.

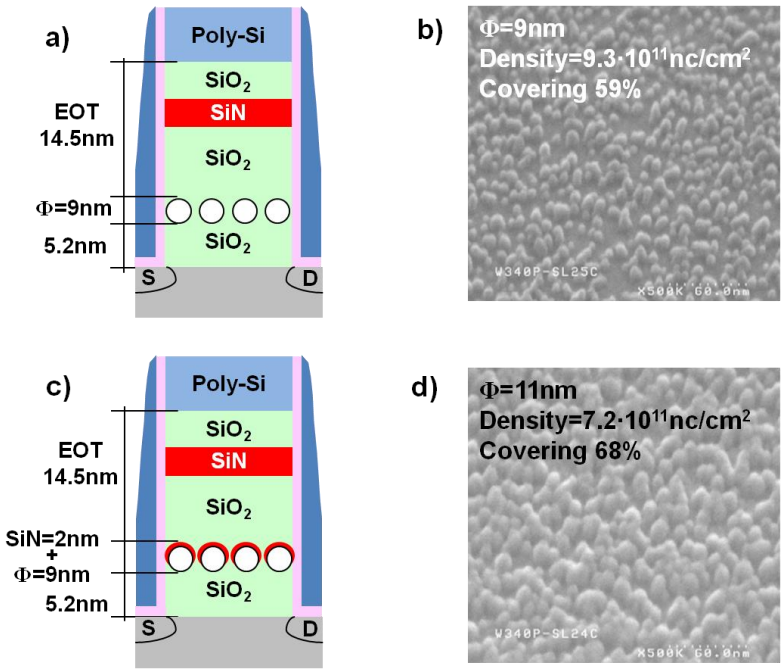


Figure 2. 7. Silicon nanocrystal cells used to evaluate the impact of nanocrystal size. a) and c) are the schematics of cells that integrated average 9nm nanocrystals diameter with and without Si_3N_4 capping layer. b) and d) are the corresponding 40° tilted CDSEM pictures used to measure the nanocrystal size and density.

The knowledge of silicon nitride growth process gained from experimental trials in STMicroelectronics and CEA-Leti laboratories led to the estimation that the Si_3N_4 capping layer was 2nm thick [Amouroux '13]. The average size measured in-line using CDSEM image processing was thus 11nm. Using the techniques described in paragraph 2.2 we characterized the programming window dispersion on wafer (figure 2.8) and the program/erase kinetics (figure 2.9).

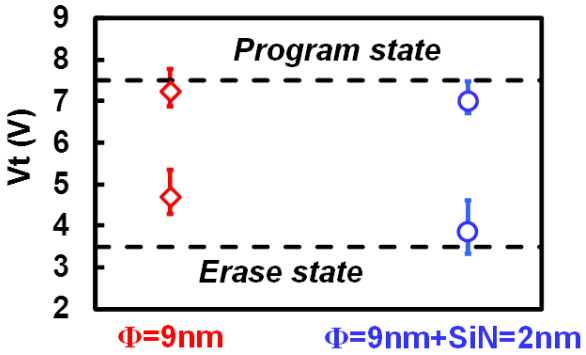


Figure 2. 8. Statistical results of program/erase threshold voltage measured for samples with and without Si_3N_4 capping layer on silicon nanocrystals.

The results plotted in figure 2.8 show the programming window is increased by 1V when the nanocrystals are capped by Si_3N_4 layer, maintaining the threshold voltage dispersion unchanged, thanks to the increased number of trapping sites [Chen '09]. Nevertheless, the expected limits of program/erase levels are not still reached. The program/erase kinetics, in figure 2.9, are achieved using the ramps previously described.

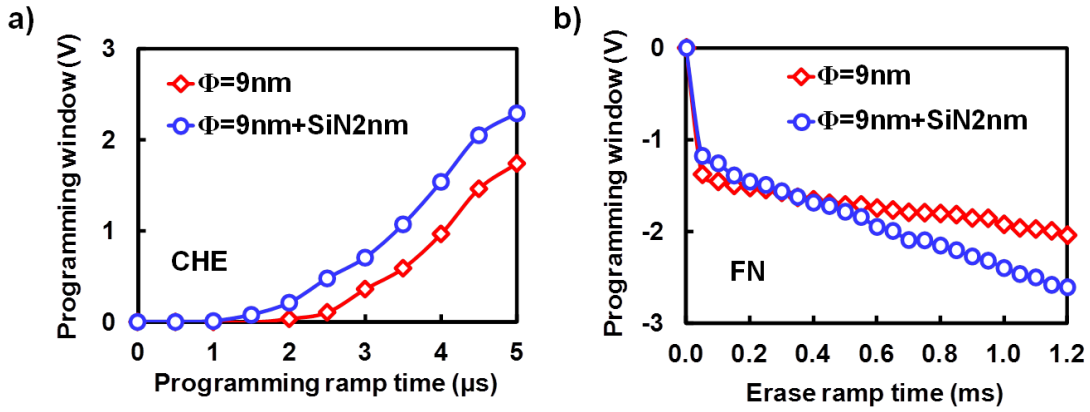


Figure 2. 9. a) Channel hot electron (CHE) programming kinetic and b) Fowler-Nordheim (FN) erase kinetic measured on samples with and without Si_3N_4 capping layer.
The program/erase pulses are described in section 2.2.

The final programming window is increased with the Si_3N_4 capping layer thanks to the increase in charge trapping probability and the improvement in coupling factor (figure 2.9). In particular the former improves the channel hot electron programming efficiency, while the covering ratio increase improves in particular the erase operation. It is worth noting that the

programming windows of figure 2.8 and 2.9 are not directly comparable because the mechanisms of program erase are different due to the different scheme box or ramp pulses. Finally, these results suggest the silicon nitride capping layer is a solution to increase the programming window maintaining the nanocrystal physical parameters (size and density) constant. In section 3 we will analyze the impact of Si_3N_4 capping layer on Si-nc cell reliability.

2.3.3 Effect of channel doping dose

The impact of channel doping dose (CDD) on the threshold voltage of a MOS transistor is well known and well analyzed in literature [Brews '78] [Brews '79a].

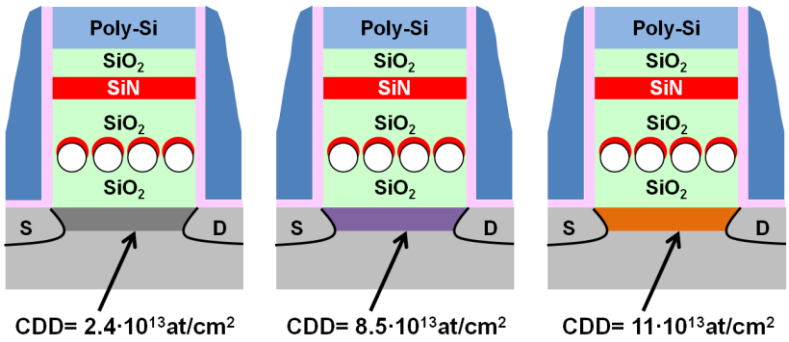


Figure 2. 10. Schematics of silicon nanocrystal cells with different channel doping doses (CDD).

In this section we show the experimental results concerning the programming window achieved using the silicon nanocrystal cell with 9nm silicon nanocrystals capped by 2nm Si_3N_4 trapping layer, where the channel doping dose is varied. More precisely three CDD are used: $2.4 \cdot 10^{13}$at/cm^2, $8.5 \cdot 10^{13}$at/cm^2, and $11 \cdot 10^{13}$at/cm^2 (figure 2.10). The aim of this trial is to increase the injection probability due to the higher number of carriers in the channel surface. In figure 2.11a the programming window dispersion on wafer is shown.

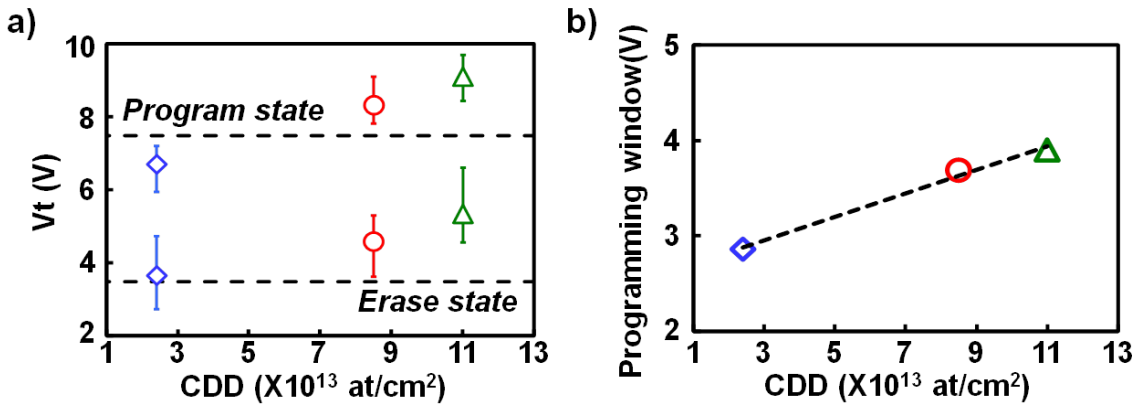


Figure 2. 11. a) Statistical results of program/erase threshold voltage measured for samples with different channel doping doses (CDD).b) Linear dependence of programming window as a function of CDD.

It is important to notice that in spite of the sample dispersion, the trends related to the threshold voltage shift versus the channel doping dose are coherent; tp and Vte shift toward higher levels, due to the threshold voltage dependence on CDD [Brews '79a] [Brews '79b] [Booth '87]. As expected, the average programming window increases with the channel doping dose (figure 2.11b). The programming window tends to saturate for the higher CDD; this can be due to the high density of introduced ions and the consequent narrowing of depleted zone in the channel. The minimum width of depleted zone depends on the centroid position of the doping dose portion which lies in the depletion region, as explained in [Brews '79a]. The program/erase kinetic characteristics, using ramped gate voltage, are plotted in figure 2.12.

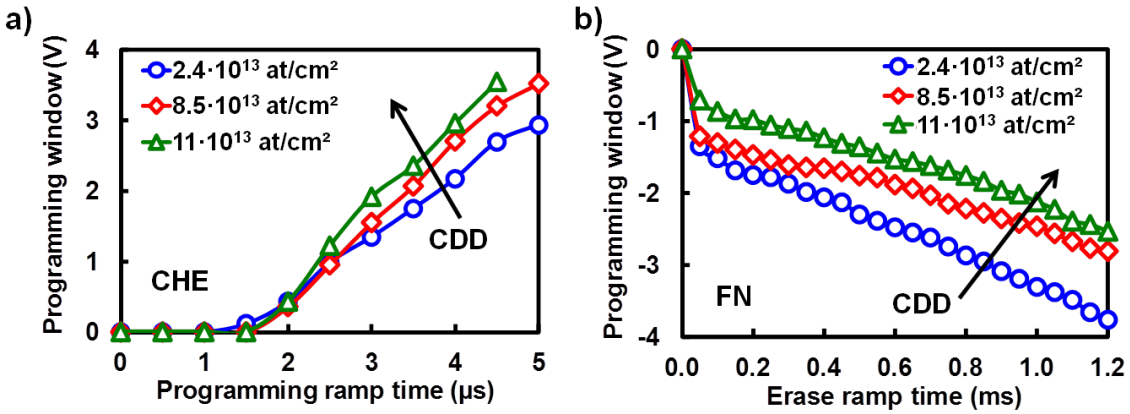


Figure 2. 12. a) Channel hot electron (CHE) programming kinetic and b) Fowler-Nordheim (FN) erase kinetic measured on samples with different channel doping doses (CDD). The program/erase ramps are described in section 2.2.

We showed in figure 2.11a the threshold voltage dependence on CDD; hence we performed kinetic measurements forcing the programmed and erased threshold voltages to the levels of Vtp=8.2V and Vte=3.5V. These levels represent the targets used in STMicroelectronics products. The Vt adjusting operation enabled the different devices to be compared in order to evaluate the impact of channel doping dose on programming window only. We notice the improvement in programming efficiency when CDD=$11\cdot10^{13}$at/cm^2 (figure 2.12a) and at the same time, the erase efficiency is decreased (figure 2.12b). The programming window trend is the same as statistical measurements made with box pulses (increasing the CDD the programming window increases), but the absolute values are not comparable because, in this case, the Vt adjusting is performed and also the program/erase scheme is different (box versus ramp). In this section we showed that by increasing the channel doping dose, the programming window increases, but the threshold voltage adjusting is needed to reach good levels of programmed and erased states. In section 3.1 we demonstrated that the erase operation can be improved with the increase of nanocrystal size and density. We can affirm

therefore it is important to find the best tradeoff between the channel doping dose and the channel covered area in order to optimize the programming window. Finally we decided to use the higher channel doping dose (CDD=$11 \cdot 10^{13}$at/cm^2) for cell optimization, as reported at the end of the chapter.

2.3.4 Effect of tunnel oxide thickness variation

The last technological parameter we varied in order to improve the Si-nc cell performances was the tunnel oxide thickness (tunox), due to the well known dependence of Fowler-Nordheim tunneling on tunox [Fowler '28] [Punchaipetch '06]. In literature alternative solutions to the SiO_2 tunnel oxide integrating high-k materials are proposed [Fernandes '01] [Maikap '08]. The studied samples are described in figure 2.13a; three values of SiO_2 tunnel oxide thickness were considered: tunox=3.7nm, tunox=4.2nm, tunox=5.2nm. It is not possible to measure the tunnel oxide thickness of the silicon nanocrystal cell because the nature of nanocrystals does enable them to be contacted directly. We decided thus to measure the capacitance of memory stack integrated in large structures applying -7V on gate terminal (30 samples). We verified that the measurement did not introduce a parasitic charge trapping in nanocrystals. Hence we extrapolated the EOT of memory stack and we compared it with theoretical results (figure 2.13b). The difference between the calculated and measured values is due to the fact that the calculated EOT does not take in to account the substrate and gate doping in capacitance stack. However one can notice that the relative variation of EOT stack thickness corresponds to the variation of tunnel oxide thickness of the three samples. Furthermore, Transmission Electron Microscopy (TEM) photos were taken to measure the physical tunnel oxide thickness at the end of fabrication process (figure 2.13c). Using a specific image processing based on the light contrast of the TEM picture we measured the physical thicknesses of our samples that correspond to the expected ones.

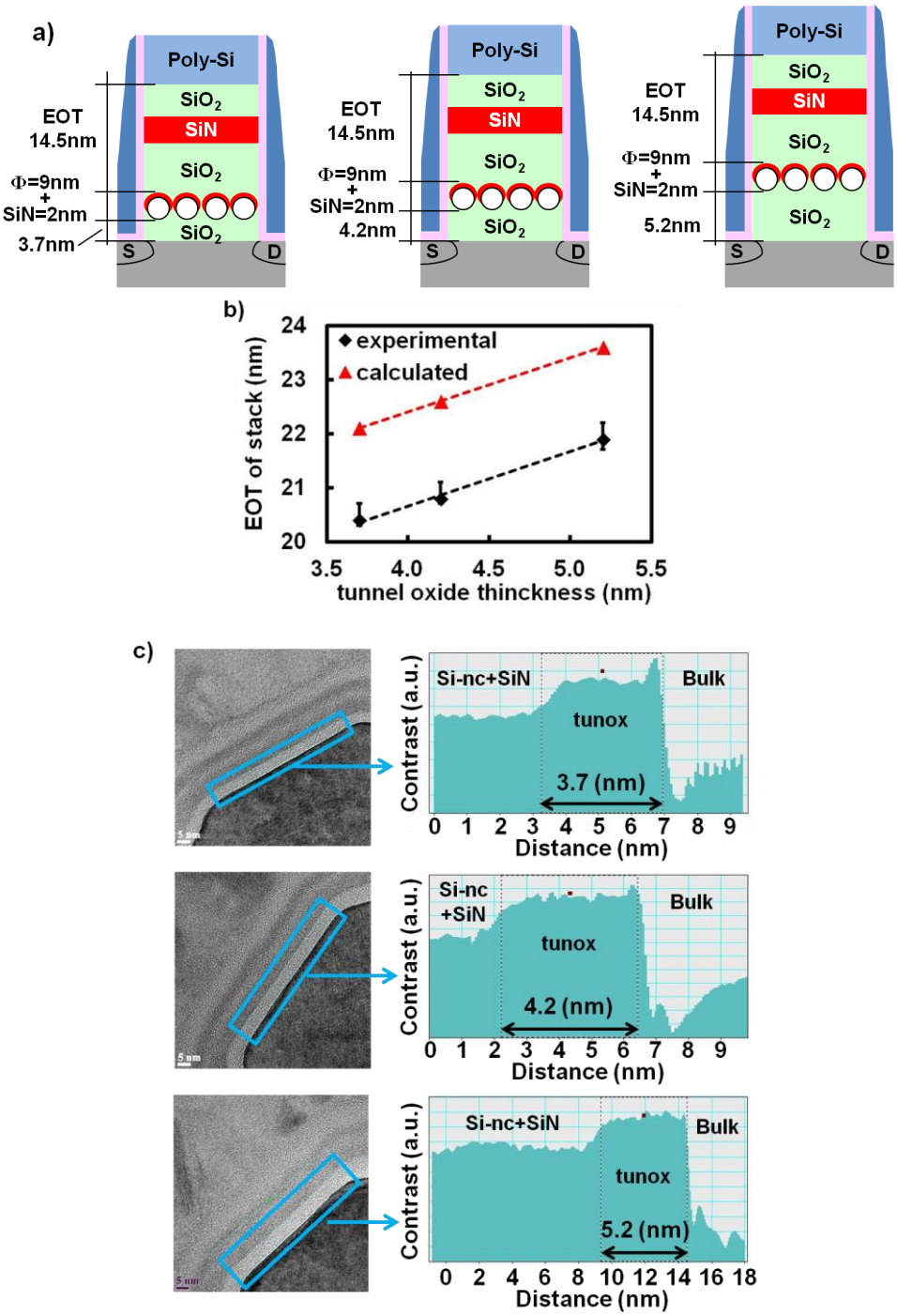


Figure 2. 13. a) Cell schematics of Si-nc+SiN devices with: tunox=3.7nm, tunox=4.2nm, tunox=5.2nm. b) Measured and calculated EOT of memory stack. C) TEM analysis used to measure the tunnel oxide thicknesses.

Also for this case we performed experiments to evaluate the programming window dispersion on wafer (30 samples are tested); the results are reported in figure 2.14. The dispersion is greater than 1V, due to the process variation. This limits the data interpretation, but the trend is clear by increasing the tunnel oxide thickness, the programming window decreases. Here the cell reaches a satisfactory erase level, using the tunox=4.2nm and tunox=3.7nm, but as in the case of channel doping dose variation, the Vt adjusting is needed due to the impact of tunnel oxide thickness on cell threshold voltage [Yu-Pin '82] [Koh '01].

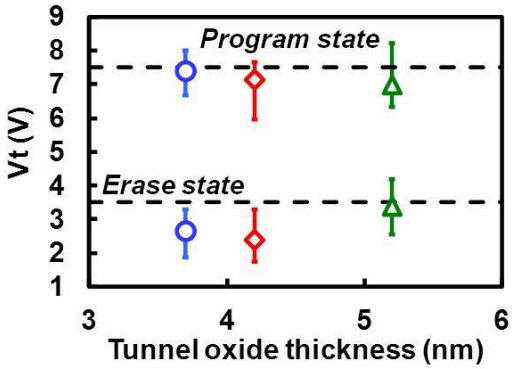


Figure 2. 14. Statistical results of program/erase threshold voltage measured for samples with different tunnel oxide thicknesses: 3.7nm, 4.2nm and 5.2nm.

Figure 2.15 shows the results of CHE program and FN erase kinetic characteristics, using the pulses described in section 2.2.

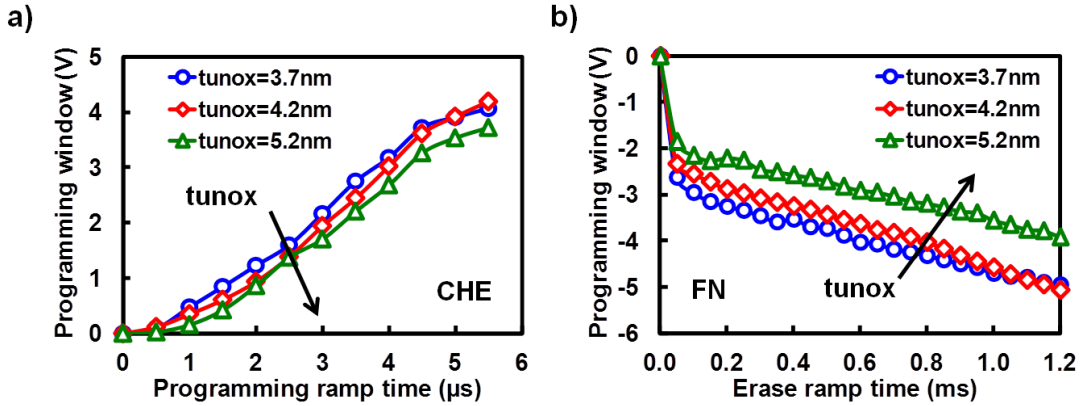


Figure 2. 15. a) Channel hot electron (CHE) programming kinetic and b) Fowler-Nordheim (FN) erase kinetic measured on samples with different tunnel oxide thickness (tunox). The program/erase pulses are described in section 2.2.

Also in this case, before the kinetic characterizations, the program/erase threshold voltages were fixed: Vte=3V and Vtp=8.2V, in order to evaluate the impact of tunnel oxide thickness on programming window only. One can notice that the tunnel oxide thickness has a limited influence on channel hot electron programming operation. This is due to the dominant role of horizontal electric field on hot carrier injection probability [Eitan '81]. In table 2.1 we reported the vertical electric field (ξ_{vert}) in tunnel oxide calculated using a 9V gate voltage and we show that, for the considered tunnel oxide thicknesses, the tunox variation slightly impacts the vertical electric field [Tam '84].

tunox (nm)	Vg (V)	EOT (nm)	ξvert (MV/cm)	Relative variation
3.7	9	22.1	4.07	-
4.2	9	22.6	3.98	2%
5.2	9	23.6	3.81	6%

Table 2. 1. Vertical electric fields calculated for samples with different tunnel oxide thicknesses: 3.7nm, 4.2nm and 5.2nm, applied during the programming operation.

Instead, during the Fowler-Nordheim erase operation, the gate voltage reaches -20V and only the vertical electric field is present. The F-N erase only depends on the applied gate voltage and tunnel oxide thickness [Fowler '28], keeping the temperature constant. We can thus affirm that the impact of tunnel oxide thickness is more relevant on erase operation; in particular a thickness of 4.2nm is sufficient to achieve the expected 4V programming window in less than 1ms. As a consequence of these considerations we performed program/erase Fowler-Nordheim characterizations using box pulses of variable duration and Vg=±18V (figure 2.16), in order to complete the evaluation of programming window dependence on tunnel oxide thickness. It is worth noting that 100ms are necessary to obtain a 4V programming window by writing with a gate voltage of 18V for the sample with 3.7nm tunnel oxide thickness. The impact of tunnel oxide thickness on erase operation is more important, because the ΔVt=6V is reached in 100ms for the same sample.

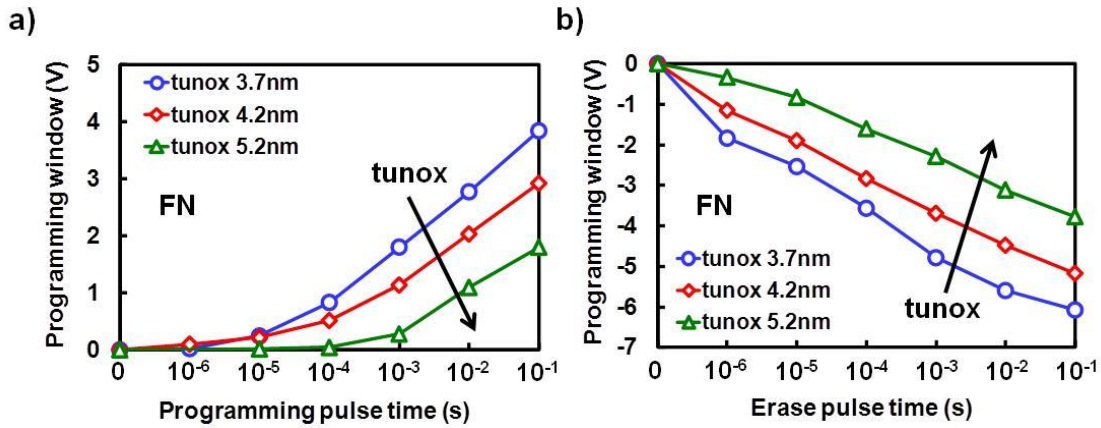


Figure 2. 16. a) Program and b) erase Fofler-Nordheim characteristics of Si-nc cell using three different tunnel oxide thicknesses (tunox): 3.7nm, 4.2nm and 5.2nm.

In figure 2.17 the ΔVt, using 100ms of program/erase time, is plotted as a function of tunnel oxide thickness. We show the dependence of Fowler-Nordheim program/erase on tunnel oxide thickness. The two operations are not symmetrical because by applying a positive or negative voltage on gate terminal, the channel zone is depleted, or not, which in turns varies the bulk surface potential. Moreover this technological parameter impacts the reliability characteristic which will be described in the next chapter. Finally it is important to find a satisfactory tradeoff between tunnel oxide thickness and channel doping dose in order to adjust the program/erase threshold voltages. Moreover in order to reach a 4V programming window the maximum tunnel oxide thickness is 4.2nm for this cell architecture.

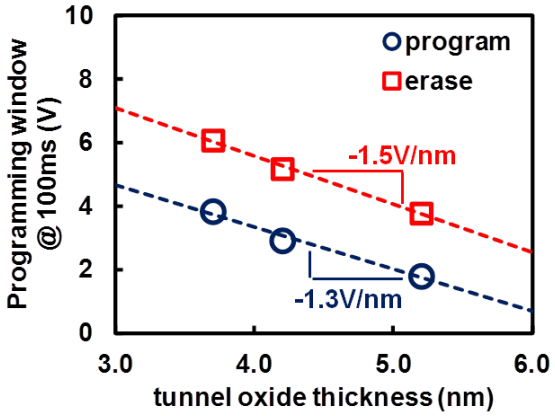


Figure 2. 17. Dependence of programming window, measured using Vg=±18V after 100ms, versus tunnel oxide thickness for the programmed and erased states.

2.4 Programming window cell optimization

In the last paragraph we evaluated the impact of the main technological parameters on the programming window of the STMicroelectronics silicon nanocrystal memory cell. The aim of this analysis was to define the best way to improve the programming window using the standard program/erase pulses used for the Flash floating gate memory cell, maintaining the cell dimensions unchanged. Below we summarize the main conclusion of previous studies that we have taken into account in order to optimize the silicon nanocrystal cell:

- When increasing the silicon nanocrystal size, thus the covered cell area, the programming window increases and in particular the Fowler-Nordheim erase operation is improved. We noticed that using the standard memory cell stack a covering of 95% is required to reach the level of 4V programming window that is not coherent with the silicon nanocrystal principle of functioning. In order to improve the programming window and optimize the stack of Si-nc cell, we considered as a fundamental point to increase the coupling factor as explained for the Flash floating gate by [Wang '79] [Wu '92] [Pavan '97] [Esseni '99] and in the chapter 1. Two different recipes have been developed to achieve 9nm and 12nm silicon nanocrystals reaching respectively 46% and 76% covered area. Furthermore with the coupling factor optimization it was possible to decrease the ONO layer thickness down to10.5nm to increase the vertical electric field during the erase operation. This thickness value is chosen in accordance with the recipes available in the STMicroelectronics production line.

- The presence of Si_3N_4 capping layer on silicon nanocrystals increases the charge trapping probability and the covered channel area. The coupling factor is increased

and then the programming window increases. Observing the CDSEM pictures we noticed that the Si_3N_4 capping layer grows around the silicon nanocrystals. Hence if their size is big enough it is possible to reach the contact and the coalescence of hybrid nanocrystals. In this case it is not possible to confirm if the programming window improvement is due to the Si_3N_4 capping layer presence or if it is due to the covered area increasing. In figure 2.7 we have shown the isolation of nanocrystals in a tilted CDSEM picture. In figure 2.18 we plot on the same graph the results of figure 2.5 and figure 2.8 showing that the programming window due to the Si_3N_4 presence can be extrapolated by the trend obtained when varying the covered area. In this case we can consider that the improvement in programming window depends mainly on covered area and slightly on charge trapping probability increasing. Even if the presence of nanocrystal Si_3N_4 capping layer is helpful to improve the programming window, we decided to avoid this process step in order to minimize the effects of parasitic charge trapping [Steimle '04] [Gerardi '07] [Bharath Kumar '06]. This choice will be explained and described in the next chapter on cell reliability.

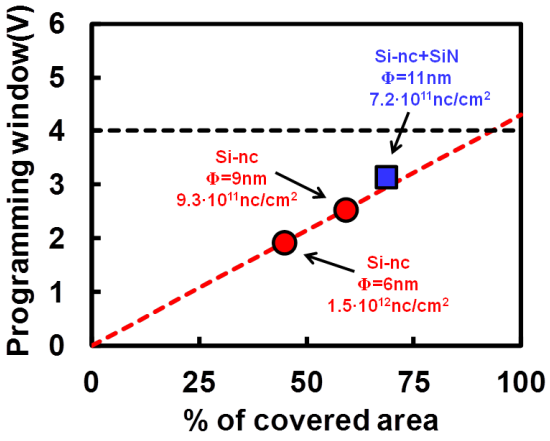


Figure 2. 18. Programming window as a function of covered area, silicon nanocrystal (Si-nc) and hybrid silicon nanocrystal (Si-nc+SiN) are compared.

- We have shown that it is possible to improve the programming window by increasing the channel doping dose, but paying attention to the shift of threshold voltages. By increasing the channel doping dose up to 10^{14}at/cm^2, a 20% programming window gain is achieved. In this case the adjusting of program/erase threshold voltages is needed and to do this it is important to find the best tradeoff with the variation of other parameters nanocrystal size and tunnel oxide thickness. After these considerations we decided to use 10^{14}at/cm^2 as the CDD for the optimized silicon nanocrystal cell in order to reach a higher programming window; the details will be given below.

- Finally we studied the impact of tunnel oxide thickness on program and erase operations. In particular we demonstrated during the channel hot electron programming the tunnel oxide thickness slightly impacts the programming window because of the dominant dependence on lateral electric field. On the contrary, this technological parameter strongly impacts the Fowler-Nordheim operations. In particular we showed the effect on erase operation using both the ramped gate voltage and box pulses. In this latter case an improvement of 1.5V/nm can be achieved. In order to reach the 4V programming window a tunnel oxide of 4.2nm or thinner is needed. As for channel doping dose, the program/erase threshold voltages are shifted by the tunnel oxide variation and a Vt adjusting operation is needed.

The layers stacked in the optimized nanocrystal cell are shown in figure 2.19. The SiO_2 4.2nm thick tunnel oxide was grown on a p-type substrate doped on a surface with a dose of 10^{14}at/cm^2. Two different recipes were developed to grow 9nm and 12nm silicon nanocrystals. The cell stack is completed with an ONO layer with an equivalent oxide thickness of 10.5nm. We showed in chapter 1 that by decreasing the ONO thickness, the capacitance between the control gate and the floating gate is increased and the programming window is thus decreased; to compensate this effect we increased the silicon nanocrystal size up to 12nm. Furthermore the Fowler-Nordheim erase operation can be improved by decreasing the ONO thickness and thus increasing the vertical electric field on tunnel oxide. Using these two nanocrystal fabrication recipes we obtained the following samples:

- Sample 1: Φ=9nm; density=$7.3\cdot10^{11}$nc/cm^2; covering=46.4%
- Sample2: Φ=12nm; density=$6.7\cdot10^{11}$nc/cm^2; covering=75.7%

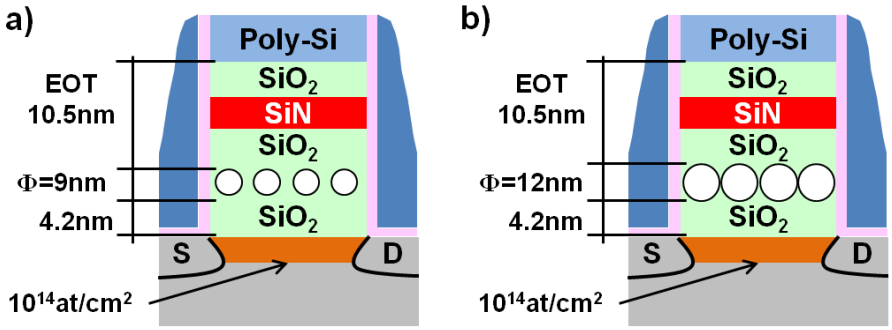


Figure 2. 19. Schematic of optimized silicon nanocrystal cell; nanocrystals with two different sizes are implemented: a) Φ=9nm, b) Φ=12nm.

In figure 2.20 the program/erase kinetic characteristics are plotted for the optimized stacks; the dispersion on wafer is also highlighted (30 samples tested). The first point we notice is the limited dispersion with respect to the nanocrystal diameter which is comparable for the two samples. Once again we demonstrated that for the silicon nanocrystal cell the covered area slightly impacts the channel hot electron programming, while the Fowler-Nordheim erase operation is strongly improved. The cell with the higher covered area can be erased in 0.4ms reaching a programming window of 4.7V which is thus greater than the 4V minimum programming window target; the program/erase threshold voltages were fixed: Vte=3V and Vtp=8.2V. In this case the quantum and/or the Coulomb blockade effects are negligible because of the large size of nanocrystals and the thick tunnel oxide. [De Salvo '01]. In conclusion with the optimized cell architecture, it is possible to reach the 4V programming window using the standard program/erase ramps described in section 2.2. In the following chapters we will continue the study of the silicon nanocrystal cell by analyzing its reliability and energy consumption.

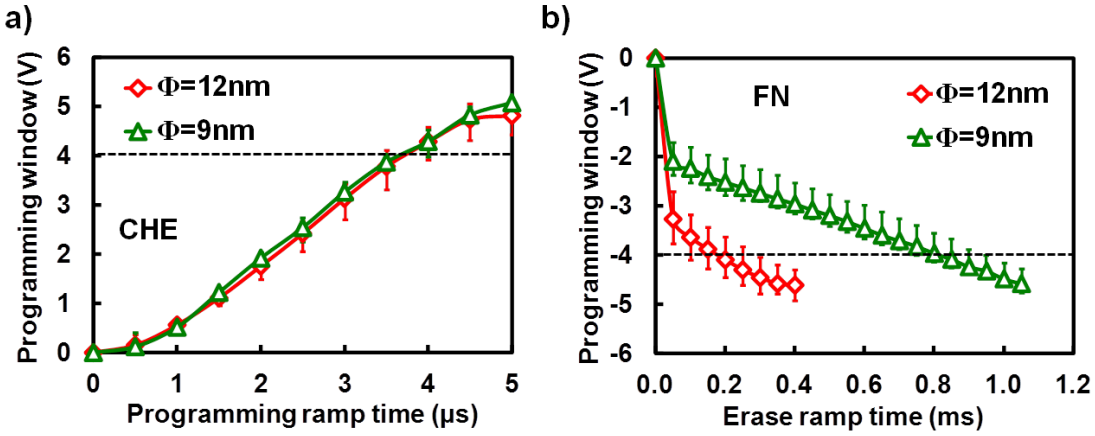


Figure 2. 20. a) Channel hot electron (CHE) programming kinetic and b) Fowler-Nordheim (FN) erase kinetic measured on samples with optimized memory stack and two different nanocrystals size: Φ=9nm and Φ=12nm. The program/erase pulses are described in section 2.2.

2.5 Benchmarking with Flash floating gate

To conclude this paragraph we compare the results obtained on optimized silicon nanocrystal memory cell (Φ=12nm; density=$6.7 \cdot 10^{11}$nc/cm^2; covering=75.7%) with the standard Flash floating gate keeping the cell size constant. To compare these devices the program/erase levels were fixed to Vte=3V and Vtp=8V. In figure 2.21 the program kinetic characteristic is shown for each device. For the optimized Si-nc cell the performances are the same of floating gate cell; the 4V minimum programming window is reached using a channel hot electron operation in 3.5μs using the ramped gate voltage.

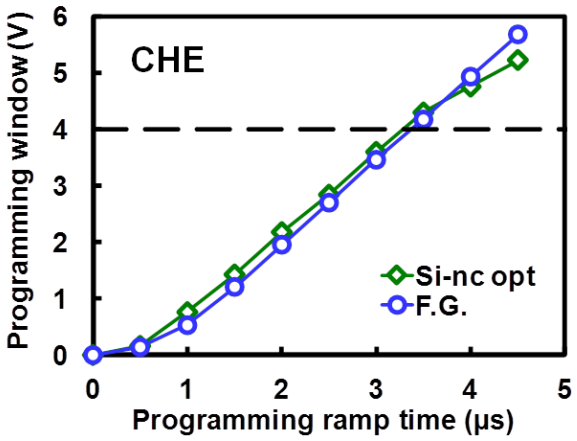


Figure 2. 21. Channel hot electron programming kinetic characteristics measured for the optimized silicon nanocrystal memory cell and Flash floating gate.

Moreover in figure 2.22 the erase kinetic characteristics are plotted, using the ramped gate voltage. The erase performances are improved with respect to the floating gate memory cell thanks to the thinner tunnel oxide thickness and the coupling factor increase.

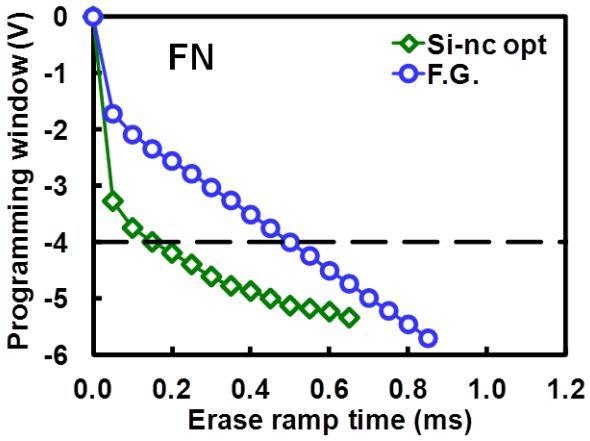


Figure 2. 22. Fowler-Nordheim erase kinetic characteristics measured for the optimized silicon nanocrystal memory cell and Flash floating gate.

The erasing time to reach the minimum 4V programming window is 0.2ms for the optimized Si-nc cell achieving a gain of 60% with respect to the Flash floating gate. Considering these last results the programming window width can be increased up to 5V by adjusting the program/erase ramp time. This is important with regard to the programming window degradation after cycling, as explained in the next chapter. In conclusion all the trials varying different technological parameters (nanocrystals size and density, Si_3N_4 capping, channel doping dose and tunnel oxide thickness) have enabled us to optimize the silicon nanocrystal cell programming window in order to make it comparable with the Flash floating gate memory cell. The aim is to substitute the floating gate and thus decrease the wafer costs. In the next chapter we will compare the optimized silicon nanocrystal cell with the Flash floating gate memory according to the reliability results (endurance and data retention).

Bibliography of chapter 2

[Amouroux '13] A. Amouroux, Ph.D. Thesis, *"Procédé de croissance et caractérisation de nanocristuax de silicium pour l'intégration dans les mémoires non volatiles"*, Université Aix-Marseille,

[Amouroux '12] J. Amouroux, E. Faivre, P. Boivin, C. Muller, D. Deleruyelle, L. Fares, P. Maillot, M. Putero, and E. Jalaguier, *"Extraction of physical parameters on silicon nanocrystals devoted to non-volatile memories"*, in Semiconductor Conference Dresden-Grenoble (ISCDG), 2012 International, 24-26 Sept. 2012, pp. 169-172, http://dx.doi.org/10.1109/iscdg.2012.6360029

[Bharath Kumar '06] P. Bharath Kumar, E. Murakami, S. Kamohara, and S. Mahapatra, *"Endurance and Retention Characteristics of SONOS EEPROMs Operated using BTBT Induced Hot Hole Erase"*, in Reliability Physics Symposium Proceedings, 2006. 44th Annual., IEEE International, 26-30 March 2006, pp. 699-700, http://dx.doi.org/10.1109/relphy.2006.251331

[Booth '87] R. V. Booth, M. H. White, W. Hon-Sum, and T. J. Krutsick, *"The effect of channel implants on MOS transistor characterization"*, Electron Devices, IEEE Transactions on, vol. 34, 1987, pp. 2501-2509, http://dx.doi.org/10.1109/T-ED.1987.23341

[Brews '78] J. R. Brews, *"A charge-sheet model of the MOSFET"*, Solid-State Electronics, vol. 21, 1978, pp. 345-355, http://dx.doi.org/10.1016/0038-1101(78)90264-2

[Brews '79a] J. R. Brews, *"Threshold shifts due to nonuniform doping profiles in surface channel MOSFET's"*, Electron Devices, IEEE Transactions on, vol. 26, 1979a, pp. 1696-1710, http://dx.doi.org/10.1109/T-ED.1979.19674

[Brews '79b] J. R. Brews, *"Subthreshold behavior of uniformly and nonuniformly doped long-channel MOSFET"*, Electron Devices, IEEE Transactions on, vol. 26, 1979b, pp. 1282-1291, http://dx.doi.org/10.1109/T-ED.1979.19594

[Chen '09] L.-L. Chen, C.-H. Chang, Y.-S. Lin, and Y.-H. Wu, *"Hybrid TiN nanocrystals/Si3N4 nonvolatile memory featuring low voltage operation by spinodal phase segregation"*, in Device Research Conference, 2009. DRC 2009, 22-24 June 2009, pp. 143-144, http://dx.doi.org/10.1109/drc.2009.5354957

[Colonna '08] J. Colonna, G. Molas, M. W. Gely, M. Bocquet, E. Jalaguier, B. De Slavo, H. Grampeix, P. Brianceau, K. Yckache, A.-M. Papon, G. Auvert, C. Bongiorno, and S. Lombardo, *"Realization of Hybrid Silicon core/silicon Nitride Shell Nanodots by LPCVD for NVM Application"*, in Materials Research Society Proceedings, http://dx.doi.org/10.1557/PROC-1071-F02-02

[De Salvo '01] B. De Salvo, G. Ghibaudo, G. Pananakakis, P. Masson, T. Baron, N. Buffet, A. Fernandes, and B. Guillaumot, *"Experimental and theoretical investigation of nano-crystal and nitride-trap memory devices"*, Electron Devices, IEEE Transactions on, vol. 48, 2001, pp. 1789-1799, http://dx.doi.org/10.1109/16.936709

[De Salvo '03] B. De Salvo, C. Gerardi, S. Lombardo, T. Baron, L. Perniola, D. Mariolle, P. Mur, A. Toffoli, M. Gely, M. N. Semeria, S. Deleonibus, G. Ammendola, V. Ancarani, M. Melanotte, R. Bez, L. Baldi, D. Corso, I. Crupi, R. A. Puglisi, G. Nicotra, E. Rimini, F. Mazen, G. Ghibaudo, G. Pananakakis, C. M. Compagnoni, D. Ielmini, A. Lacaita, A. Spinelli, Y. M. Wan, and K. van der Jeugd, *"How far will silicon nanocrystals push the scaling limits of NVMs technologies?"*, in Electron Devices Meeting, 2003. IEDM '03 Technical Digest. IEEE International, 8-10 Dec. 2003, pp. 26.1.1-26.1.4, http://dx.doi.org/10.1109/iedm.2003.1269352

[Eitan '81] B. Eitan and D. Frohman-Bentchkowsky, *"Hot-electron injection into the oxide in n-channel MOS devices"*, Electron Devices, IEEE Transactions on, vol. 28, 1981, pp. 328-340, http://dx.doi.org/10.1109/T-ED.1981.20336

[Esseni '99] D. Esseni, A. Della Strada, P. Cappelletti, and B. Ricco, *"A new and flexible scheme for hot-electron programming of nonvolatile memory cells"*, Electron Devices, IEEE Transactions on, vol. 46, 1999, pp. 125-133, http://dx.doi.org/10.1109/16.737450

[Fernandes '01] A. Fernandes, B. DeSalvo, T. Baron, J. F. Damlencourt, A. M. Papon, D. Lafond, D. Mariolle, B. Guillaumot, P. Besson, P. Masson, G. Ghibaudo, G. Panankakis, F. Martin, and S. Haukka, *"Memory characteristics of Si quantum dot devices with SiO_2 /ALD Al_2O_3 tunneling dielectrics "*, in Electron Devices Meeting, 2001. IEDM '01. Technical Digest. International, 2001, pp. 7.4.1-7.4.4, http://dx.doi.org/10.1109/iedm.2001.979455

[Fowler '28] R. H. Fowler and L. Nordheim, *"Electron Emission in Intense Electric Fields"*, Proceedings of the Royal Society of London. Series A, vol. 119, May 1, 19281928, pp. 173-181, http://dx.doi.org/10.1098/rspa.1928.0091

[Gerardi '07] C. Gerardi, V. Ancarani, R. Portoghese, S. Giuffrida, M. Bileci, G. Bimbo, O. Brafa, D. Mello, G. Ammendola, E. Tripiciano, R. Puglisi, and S. A. Lombardo, *"Nanocrystal Memory Cell Integration in a Stand-Alone 16-Mb NOR Flash Device"*, Electron Devices, IEEE Transactions on, vol. 54, 2007, pp. 1376-1383, http://dx.doi.org/10.1109/TED.2007.895868

[Gerardi '08] C. Gerardi, G. Molas, G. Albini, E. Tripiciano, M. Gely, A. Emmi, O. Fiore, E. Nowak, D. Mello, M. Vecchio, L. Masarotto, R. Portoghese, B. De Salvo, S. Deleonibus, and A. Maurelli, *"Performance and reliability of a 4Mb Si nanocrystal NOR Flash memory with optimized 1T memory cells"*, in Electron Devices Meeting, 2008. IEDM 2008. IEEE International, 15-17 Dec. 2008, pp. 1-4, http://dx.doi.org/10.1109/iedm.2008.4796823

[Jacob '08] S. Jacob, B. De Salvo, L. Perniola, G. Festes, S. Bodnar, R. Coppard, J. F. Thiery, T. Pate-Cazal, C. Bongiorno, S. Lombardo, J. Dufourcq, E. Jalaguier, T. Pedron, F. Boulanger, and S. Deleonibus, *"Integration of CVD silicon nanocrystals in a 32 Mb NOR flash memory"*, Solid-State Electronics, vol. 52, 2008, pp. 1452-1459, http://dx.doi.org/10.1016/j.sse.2008.04.032

[Koh '01] M. Koh, W. Mizubayashi, K. Iwamoto, H. Murakami, T. Ono, M. Tsuno, T. Mihara, K. Shibahara, S. Miyazaki, and M. Hirose, *"Limit of gate oxide thickness scaling in MOSFETs due to apparent threshold voltage fluctuation induced by tunnel leakage current"*, Electron Devices, IEEE Transactions on, vol. 48, 2001, pp. 259-264, http://dx.doi.org/10.1109/16.902724

[Maikap '08] S. Maikap, W. Banerjee, P. J. Tzeng, T. Y. Wang, C. H. Lin, T. C. Tien, L. S. Lee, J. R. Yang, M. J. Kao, and M. J. Tsai, *"Highly Thermally Stable and Reproducible of ALD RuO_2 Nanocrystal Floating Gate Memory Devices with Large Memory Window and Good Retention"*, in VLSI Technology, Systems and Applications, 2008. VLSI-TSA 2008. International Symposium on, 21-23 April 2008, pp. 50-51, http://dx.doi.org/10.1109/vtsa.2008.4530793

[Molas '07] G. Molas, M. Bocquet, J. Buckley, J. P. Colonna, L. Masarotto, H. Grampeix, F. Martin, V. Vidal, A. Toffoli, P. Brianceau, L. Vermande, P. Scheiblin, M. Gely, A. M. Papon, G. Auvert, L. Perniola, C. Licitra, T. Veyron, N. Rochat, C. Bongiorno, S. Lombardo, B. De Salvo, and S. Deleonibus, *"Thorough investigation of Si-nanocrystal memories with high-k interpoly dielectrics for sub-45nm node Flash NAND applications"*, in Electron Devices Meeting, 2007. IEDM 2007. IEEE

International, 10-12 Dec. 2007, pp. 453-456, http://dx.doi.org/10.1109/iedm.2007.4418971

[Monzio Compagnoni '03] C. Monzio Compagnoni, D. Ielmini, A. S. Spinelli, A. L. Lacaita, C. Gerardi, L. Perniola, B. De Salvo, and S. Lombardo, *"Program/erase dynamics and channel conduction in nanocrystal memories"*, in Electron Devices Meeting, 2003. IEDM '03 Technical Digest. IEEE International, 8-10 Dec. 2003, pp. 22.4.1-22.4.4, http://dx.doi.org/10.1109/iedm.2003.1269342

[Pavan '97] P. Pavan, R. Bez, P. Olivo, and E. Zanoni, *"Flash memory cells-an overview"*, Proceedings of the IEEE, vol. 85, 1997, pp. 1248-1271, http://dx.doi.org/10.1109/5.622505

[Punchaipetch '06] P. Punchaipetch, K. Ichikawa, Y. Uraoka, T. Fuyuki, A. Tomyo, E. Takahashi, and T. Hayashi, *"Experimental investigation of tunnel oxide thickness on charge transport through Si nanocrystal dot floating gate memories"*, Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures, vol. 24, 2006, pp. 1271-1277, http://dx.doi.org/10.1116/1.2198852

[Rao '05] R. A. Rao, H. P. Gasquet, R. F. Steimle, G. Rinkenberger, S. Straub, R. Muralidhar, S. G. H. Anderson, J. A. Yater, J. C. Ledezma, J. Hamilton, B. Acred, C. T. Swift, B. Hradsky, J. Peschke, M. Sadd, E. J. Prinz, K. M. Chang, and B. E. White Jr, *"Influence of silicon nanocrystal size and density on the performance of non-volatile memory arrays"*, Solid-State Electronics, vol. 49, 2005, pp. 1722-1727, http://dx.doi.org/10.1016/j.sse.2005.10.022

[Steimle '03] R. F. Steimle, M. Sadd, R. Muralidhar, R. Rajesh, B. Hradsky, S. Straub, and B. E. White, Jr., *"Hybrid silicon nanocrystal silicon nitride dynamic random access memory"*, Nanotechnology, IEEE Transactions on, vol. 2, 2003, pp. 335-340, http://dx.doi.org/10.1109/TNANO.2003.820817

[Steimle '04] R. F. Steimle, R. Rao, M. Sadd, C. Swift, B. Hradsky, S. Straub, T. Merchant, M. Stoker, C. Parikh, S. Anderson, M. Rossow, J. Yater, B. Acred, K. Harber, E. Prinz, B. E. White, Jr., and R. Muralidhar, *"Silicon nanocrystals: from Coulomb blockade to memory arrays"*, in Nanotechnology, 2004. 4th IEEE Conference on, 16-19 Aug. 2004, pp. 290-292, http://dx.doi.org/10.1109/nano.2004.1392328

[Tam '84] S. Tam and H. Chenming, *"Hot-electron-induced photon and photocarrier generation in Silicon MOSFET's"*, Electron Devices, IEEE Transactions on, vol. 31, 1984, pp. 1264-1273, http://dx.doi.org/10.1109/T-ED.1984.21698

[Wang '79] S. T. Wang, *"On the I-V characteristics of floating-gate MOS transistors"*, Electron Devices, IEEE Transactions on, vol. 26, 1979, pp. 1292-1294, http://dx.doi.org/10.1109/T-ED.1979.19595

[Wu '92] C.-Y. Wu and C.-F. Chen, *"Physical model for characterizing and simulating a FLOTOX EEPROM device"*, Solid-State Electronics, vol. 35, 1992, pp. 705-716, http://dx.doi.org/10.1016/0038-1101(92)90041-a

[Yu-Pin '82] H. Yu-Pin, J. Mize, T. Mozdzen, T. O"Keefe, J. Pinto, and R. Worley, *"Ultra-thin gate-oxide characteristics and MOS/VLSI scaling implications"*, in Electron Devices Meeting, 1982 International, 1982, pp. 98-102, http://dx.doi.org/10.1109/iedm.1982.190223

Chapter 3 – Reliability of silicon nanocrystal memory cell

3.1 Introduction

In this section we present the results concerning our study on the reliability of the silicon nanocrystal memory cell. The data retention, thus the charge loss, is evaluated for different temperatures starting from a fixed programmed state [Gerardi '02]. The aim of endurance experiments is to evaluate the cell functionality after a large number of program/erase cycles, typically 100k. In particular, the results will show the impact of some technological parameters on cell reliability such as: silicon nanocrystal size and density, silicon nitride capping layer, channel doping dose and tunnel oxide. The comprehension of experimental results will be useful to improve the cell performances similarly to the previous chapter devoted to programming window characterization. At the end we will present the results obtained for the optimized STMicroelectronics nanocrystal cell and we will compare them to the standard Flash floating gate.

3.2 Data retention: impact of technological parameters

The data retention experiments have been performed by programming the silicon nanocrystal cell with manual and automatic test bench described in chapter 2. Let us evaluate the effect of main technological parameters on data retention and chose the best cell architecture configuration to optimize the performance.

3.2.1 Effect of silicon nitride capping layer

We have seen in chapter 2 that hybrid silicon nanocrystal memory is an attractive solution to improve the cell programming window [Steimle '03] [Colonna '08] [Chen '09] [Tsung-Yu '10]. Moreover in literature many papers present the integration of high-k materials as a good option to achieve better cell performances [Lee '03] [Molas '07]. At STMicroelectronics we were not able to integrate high-k materials in the process flow so we decided to develop a hybrid solution by capping the silicon nanocrystals using the silicon nitride (Si_3N_4) to compare it with standard Si-nc cell. The presence of this layer improves the programming window by increasing the covered area and generating a higher number of trapping sites with respect to the simple Si-nc utilization as explained in chapter 2. In literature the effect of temperature on charge loss is also explained when a silicon nitride layer is used to store the charges. This is the case for SONOS and TANOS memories [Tsai '01] [Chung '07] [Padovani '10] [Molas '10] [Min-Feng '12]. The hybrid silicon nnocrystal cell has thus demonstrated higher operation speed than a plain SONOS memory, while maintaining better retention characteristic than a pure Si nanocrystal memory [Steimle '03] [Chen '09] [Hung-Bin '12]. We reported in figure 3.1 our results of data retention at 150°C and 250°C of silicon nanocrystal cell with and without the 2nm Si_3N_4 capping layer; the tunnel oxide thickness is 5.2nm (see the device description in figure 2.7).

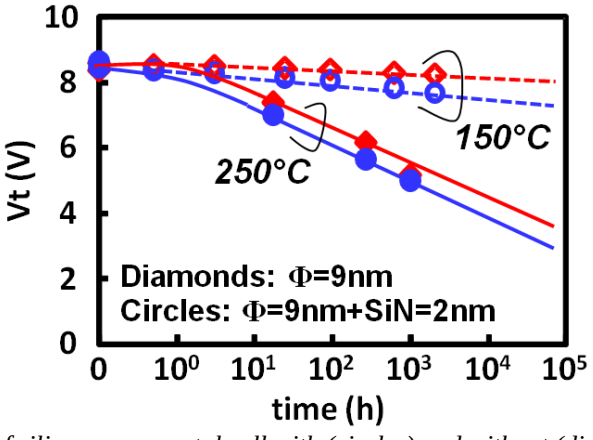


Figure 3. 1. Data retention of silicon nanocrystal cell with (circles) and without (diamonds) Si_3N_4 capping layer at 150°C and 250°C.

The charge loss is strongly activated by temperature. At 250°C the threshold voltage shift is around 2.5V after 100h, while it is only 300mV at 150°C. Furthermore, the presence of Si_3N_4 capping layer increases the charge loss with respect to the samples where the nanocrystals are not capped. Hence, in this case, the data retention characteristic of hybrid silicon nanocrystal cell is not improved. The examples described in literature of hybrid Si-nc cell show nanocrystals embedded and completely surrounded by silicon nitride layer (figure 3.2a), while in our case the nanocrystals are grown on SiO_2 tunnel oxide and only capped by Si_3N_4 layer (figure 3.2b). The band diagram of figure 3.2a shows that if the nanocrystals are surrounded by silicon nitride layer, the barrier to consider for data retention is 7.2nm thick (tunox=5.2nm and SiN=2nm). In our case (figure 3.2b), the barrier between Si-ncs and the substrate is thinner and represented by the tunnel oxide only.

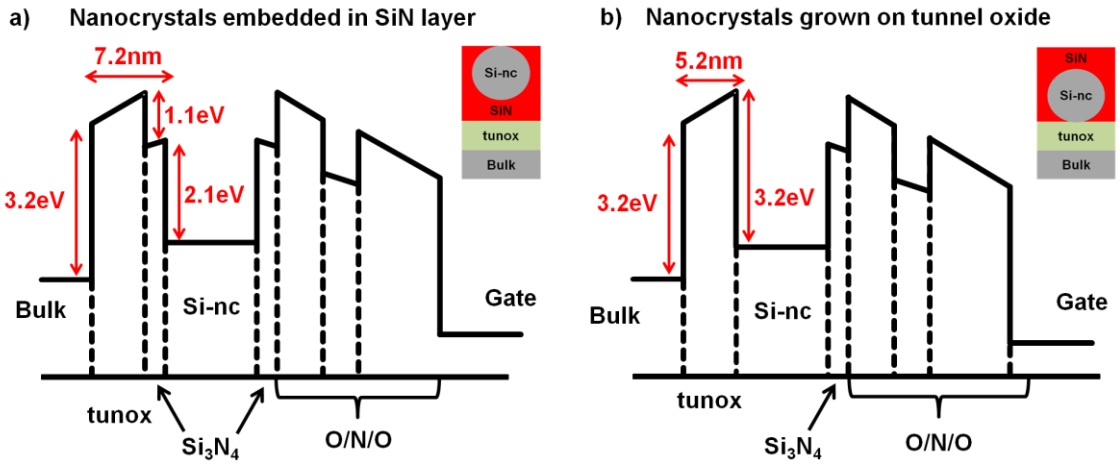


Figure 3. 2. Band diagrams of hybrid silicon nanocrystal cell: a) the silicon nanocrystals are embedded in Si_3N_4 trapping layer, b) the silicon nanocrystals are grown on SiO2 tunnel oxide and capped by Si_3N_4.

This explains why our hybrid Si-nc+SiN memory loses more charges with respect to the simple Si-nc cell. The difference can be due to the portions of trapped charge at the tunox/SiN interface around the Si-nc. These trapped charges are more energetic and when the temperature increases can be easily lost. In this case the presence of Si_3N_4 layer does not improve the cell data retention because is not sandwiched between the nanocristal and the bulk to increase the barrier thickness. Rather it caps the nanocrystals and contacs directly the tunnel oxide, creating a parasitic charge trapping at the tunox/SiN interface.

3.2.2 Effect of channel doping dose

In figure 3.3 we plot the data retention results for the samples where the channel doping dose (CDD) is changed. In this case the tunnel oxide thickness is 5.2nm and the nanocrystals are capped with the Si_3N_4 layer (Φ=9nm+SiN=2nm). The charge loss at 150°C is not impacted by the CDD. The slight difference with the data presented in figure 3.1 can be due to the parasitic charge trapping at the tunox/SiN interface caused by the irregular Si_3N_4 layer deposition on the wafer; nevertheless, the difference is not relevant for cell behavior understanding.

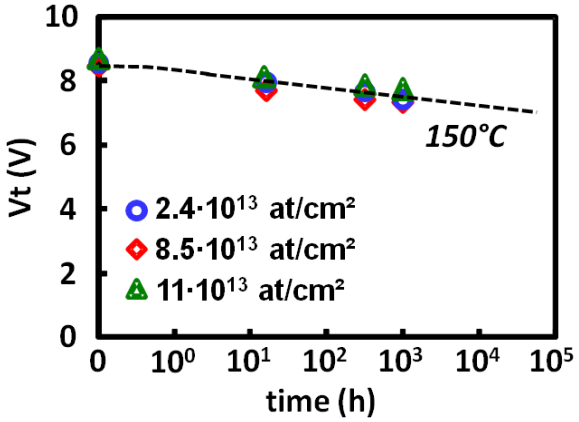


Figure 3. 3. Data retention of silicon nanocrystal cell with different channel doping doses at 150°C. The tunnel oxide thickness is 5.2nm and the Si-nc are capped by Si_3N_4 layer (Φ=9nm+SiN=2nm).

3.2.3 Effect of tunnel oxide thickness

The most important parameter to evaluate the charge loss during the data retention is the tunnel oxide thickness (tunox), because as shown in figure 3.2b, it defines the thickness of barrier between the nanocrystals and the substrate. In literature the direct dependence on tunox of charge loss in terms of applied electric field is explained [Weihua '07] [Ghosh '10]. The mechanisms of charge loss can be activated since it depends on the tunnel oxide thickness and the type of traps generated by the Si-nc fabrication process. We considered samples with three different tunnel oxide thicknesses: 3.7nm, 4.2nm and 5.2nm (description in figure 2.13). Figure 3.4 shows the data retention results at 27°C, 150°C and 250°C. In these samples the silicon nanocrystals are capped by the Si_3N_4 trapping layer. Ideally one memory cell must keep the charge stored for 10 years at 27°C. Considering the charge loss, the cell program level has to be greater than the program verify level; in our case this level is fixed at 5.7V. The data retention specification is reached only using a tunnel oxide thickness of 5.2nm, whereas if a tunox=4.2nm is used, the charge loss constraint is very close to the suitable limit. Moreover, the charge loss mechanism is strongly accelerated if the temperature

is increased. This is due to the parasitic charge trapping at the tunox/SiN interface [Chung '07]. As previously explained, the charge loss mechanism in this case is similar to the SONOS and TANOS memories [Hung-Bin '12]. In table 3.1 we summarized the percentage of lost charge after 186 hours for all temperatures considered and tunnel oxide thicknesses.

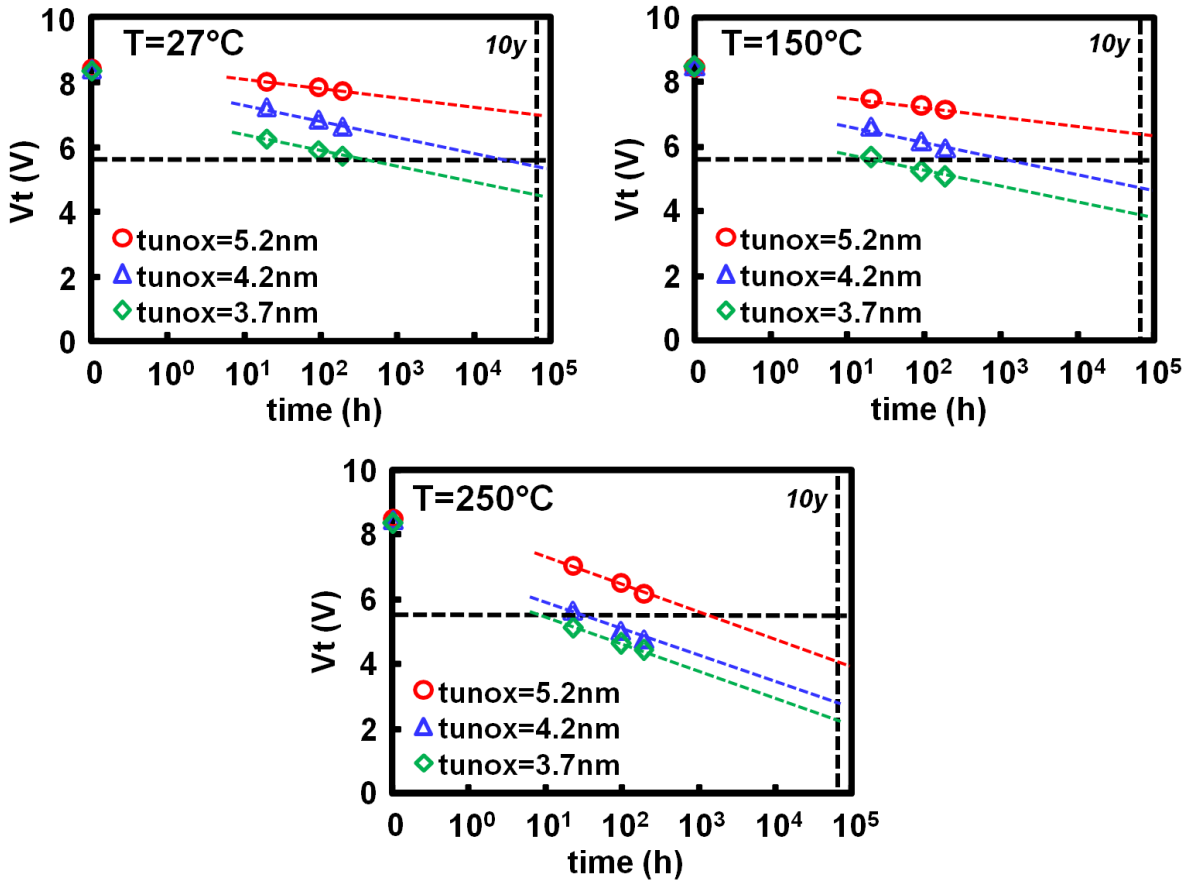


Figure 3. 4. Data retention at 27°C, 150°C, 250°C for hybrid silicon nanocrystal cell (Si-nc+SiN) varying the tunnel oxide thickness: 3.7nm, 4.2nm, 5.2nm.

Percentage of charge loss after 186h

	Temperature (°C)		
tunox (nm)	**27**	**150**	**250**
5.2	8%	15%	27%
4.2	21%	30%	43%
3.7	31%	40%	47%

Table 3. 1. Percentage of charge loss after 186h

In figure 3.5 we show the Arrhenius plot of retention time, defined as the time necessary to reach a Vt=6V that is very near the program verify level. The slopes of extrapolated trends enable achieving the charge loss activation energy (Ea) to be attained for each tunnel oxide thickness considered embedded in our memory stack. As expected, when increasing the tunnel oxide thickness, a higher energy is necessary to activate the charge loss mechanism [Weihua '07] [Ghosh '10]. The extrapolated quantities are comparable to the amounts found

in literature [Gerardi '08] [Lee '09] [Gay '12]. However the differences are due to the cell architectures.More particularly these activation energies can be impacted by the parasitic charge trapping in Si_3N_4 capping layer. We can conclude that is possible to achieve the data retention specification using a tunnel oxide 5.2nm thick in the range of temperature [27°C; 150°C], but as we demonstrated in chapter 2, in accordance with the literature, this parameter impacts the coupling factor and the Fowler-Nordheim erase efficiency. To improve the performance, the cell architecture was changed, avoiding the Si_3N_4 capping layer and thus minimizing the parasitic charge trapping (see section 3.4)

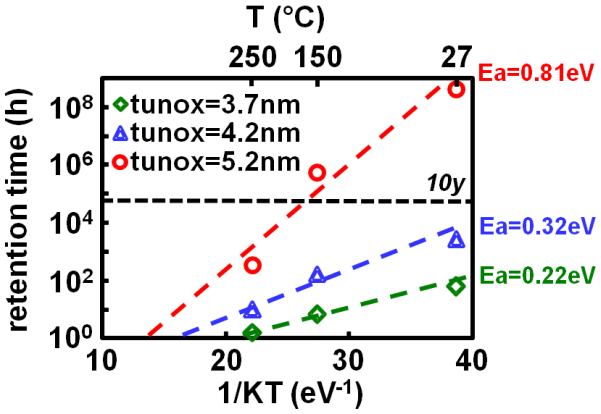


Figure 3. 5. Arrhenius plot of retention time (defined as the time corresponding to to reach Vt=6V) for temperatures of 27°C, 150°C and 250°C.

3.3 Endurance: impact of technological parameters

Another important point to evaluate in the key criterion of reliability is the endurance of silicon nanocrystal memory cell. We present in this chapter the memory device degradation in terms of programming window closure when the number of program/ease cycles is increased. We investigate the effect of different technological parameters such as: silicon nanocrystal size and density (covering area), silicon nitride capping layer (Si_3N_4), channel doping dose and tunnel oxide thickness. The aim is to understand the Si-nc cell behavior and chose the best architecture configuration to optimize the memory performance.

3.3.1 Impact of silicon nanocrystal size

We described in chapter 2 the increasing of nanocrystal size, thus the covering of channel area, corresponds to an increase of the programming window. Using the same samples we reported in figure 3.7, the results of 100kcycles endurance experiments. The cells were programmed by channel hot electron using Vg=9V, Vd=4.2V and tp=5µs; the Fowler-Nordheim erase was performed with Vg=-18V, using a ramp of 5kV/s followed by a plateau of 90ms (te). A schematic of the cycling signals is reported in figure 3.6.

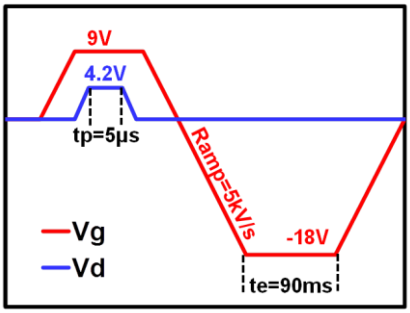


Figure 3. 6. Schematic of signals used for endurance experiments.

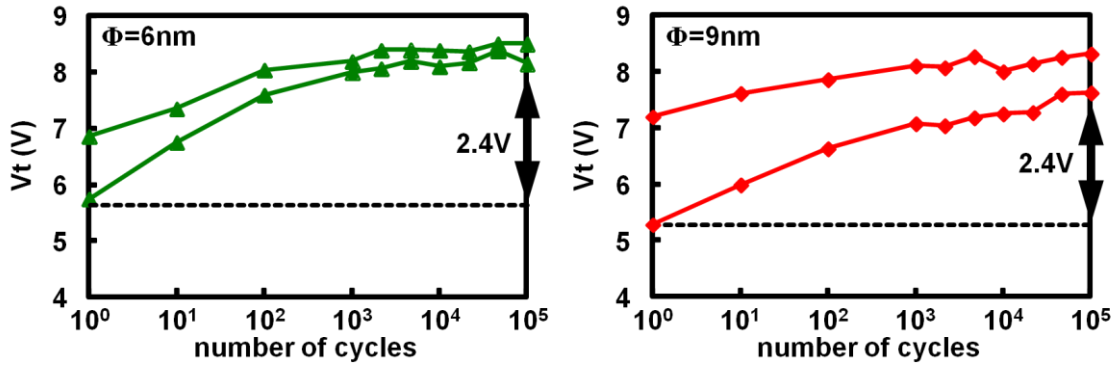


Figure 3. 7. Endurance characteristics of silicon nanocrystal cells comparing the samples with different nanocrystal size: Φ=6nm and Φ=9nm.The cells are programmed by CHE (Vg=9V, Vd=4.2V, tp=5µs) and erased by FN (Vg=-18V, ramp=5kV/s+te=90ms).

The programming window at the beginning is 1.9V for the bigger nanocrystals and 1.1V using the 6nm nanocrystals. This result is coherent with the data reported in chapter 2. The erase operation is improved by increasing the channel covered area (9nm Si-nc), because the coupling factor is increased. However, in both cases we notice an important shift of threshold voltages, in particular for the erase state (2.4V). This can be due to the parasitic charge trapping in ONO interpoly dielectric layer [Jacob '07] [Gerardi '07] [Hung-Bin '12]. The electric field applied during the FN erase is not sufficient to evacuate the trapped charges during the programming operation in the SiN layer of ONO stack. The same phenomenon is present in SONOS and TANOS memories. Despite, the improvement due to the increasing of silicon nanocrystal size (or covered area), cell functioning does not reach the specification of 100kcycles. The cell with the 6nm silicon nanocrystals, after only 10cycles, has the erase threshold voltage at the same level of program threshold voltage; hence it is impossible to discriminate the two states for the reading circuits. For the sample with 9nm embedded Si-nc the endurance limit is 10kcycles. Finally, to improve the programming window and the erase efficiency, thus the cell endurance, one way is to increase the silicon nanocrystal covered area. Further development of cell architecture is needed to reach a good programming window, such as Si_3N_4 capping layer, channel doping dose, tunnel oxide thickness, but we will study these aspects in the next sections.

3.3.2 Impact of silicon nitride capping layer

We repeated the endurance experiments on the samples with the silicon nanocrystals capped by the silicon nitride (Si_3N_4) layer, by keeping the program/erase conditions unchanged (figure 3.6). In figure 3.8 we reported the results of cycling experiments, comparing the Si-nc cell (Φ=9nm) with and without the silicon nitride capping layer.

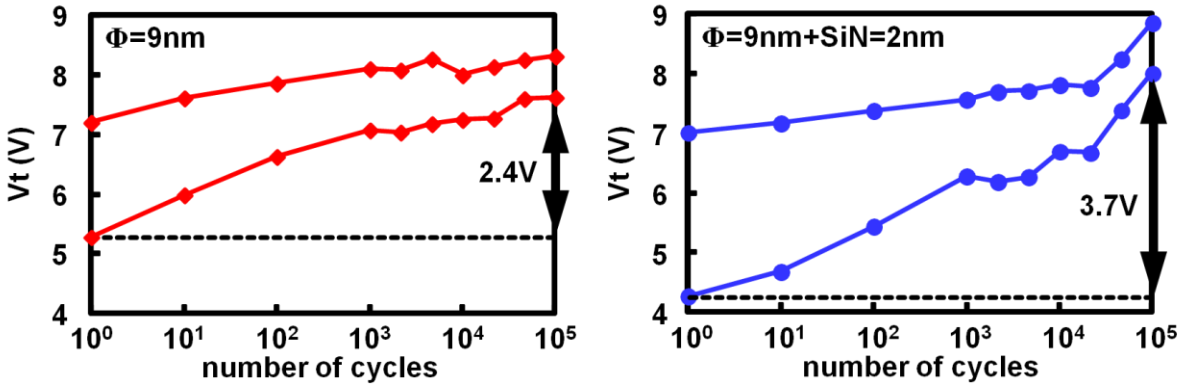


Figure 3. 8. Endurance characteristics of silicon nanocrystal cell comparing the samples with and without the Si_3N_4 capping layer: Φ=9nm and Φ=9nm+SiN=2nm.The cells are programmed by CHE (Vg=9V, Vd=4.2V, tp=5µs) and erased by FN (Vg=-18V, ramp=5kV/s+te=90ms).

The programming windows at the beginning of cycling experiments are coherent with the data reported in chapter 2. The Si_3N_4 capping layer enables improvement on the programming window, but the shift of program/erase threshold voltage is increased too. This effect is due to the parasitic charge trapping in silicon nitride and at the tunox/SiN interface that generates a 3.7V shift of Vte. The shift of threshold voltages and the low erase efficiency cause the programming window closure after 30kcycles. In order to summarize the results, we reported in table 3.2 the values of programming window before and after the cell cycling and the program/erase threshold voltage (Vtp and Vte) shifts. To conclude this section we can affirm that the Si_3N_4 capping layer enables, in this case, a 30% programming window gain to be achieved which is maintained after 100k cycles, but the large quantity of parasitic charge cumulated during the experiments determines a great shift of program/erase threshold voltages limiting the cell functioning at 30kcycles.

Si-nc	Programming window @1cycle	Programming window @100kcycles	Vtp shift	Vte shift	Endurance limit
Φ=6nm	1.1V	0.4V	1.6V	2.4V	10cycles
Φ=9nm	1.4V	0.7V	1.1V	2.4V	10kcycles
Φ=9nm+SiN=2nm	2.7V	0.8V	1.8V	3.7V	30kcycles

Table 3. 2. Programming window before and after 100k program/erase cycles, and program/erase threshold voltage shift; the studied samples have a different silicon nanocrystals diameter: Φ=6nm, Φ=9nm and Si_3N_4 capping layer (Φ=9nm+SiN=2nm).

3.3.3 Impact of channel doping dose

We have previously seen in chapter 2 that an increasing of the channel doping dose enables the programming window to be increased and it generates a shift of program/erase threshold voltages. The results concerning the endurance experiments are shown and discussed below, using the same samples described in figure 2.7c (and figure 2.10) with three different channel doping doses: $2.4 \cdot 10^{13}$at/cm^2, $8.5 \cdot 10^{13}$at/cm^2 and $11 \cdot 10^{13}$at/cm^2. The cells were programmed by channel hot electron and erased by Fowler-Nordheim as shown in figure 3.6, but using te=10ms. The results of experiments are plotted in figure 3.9.

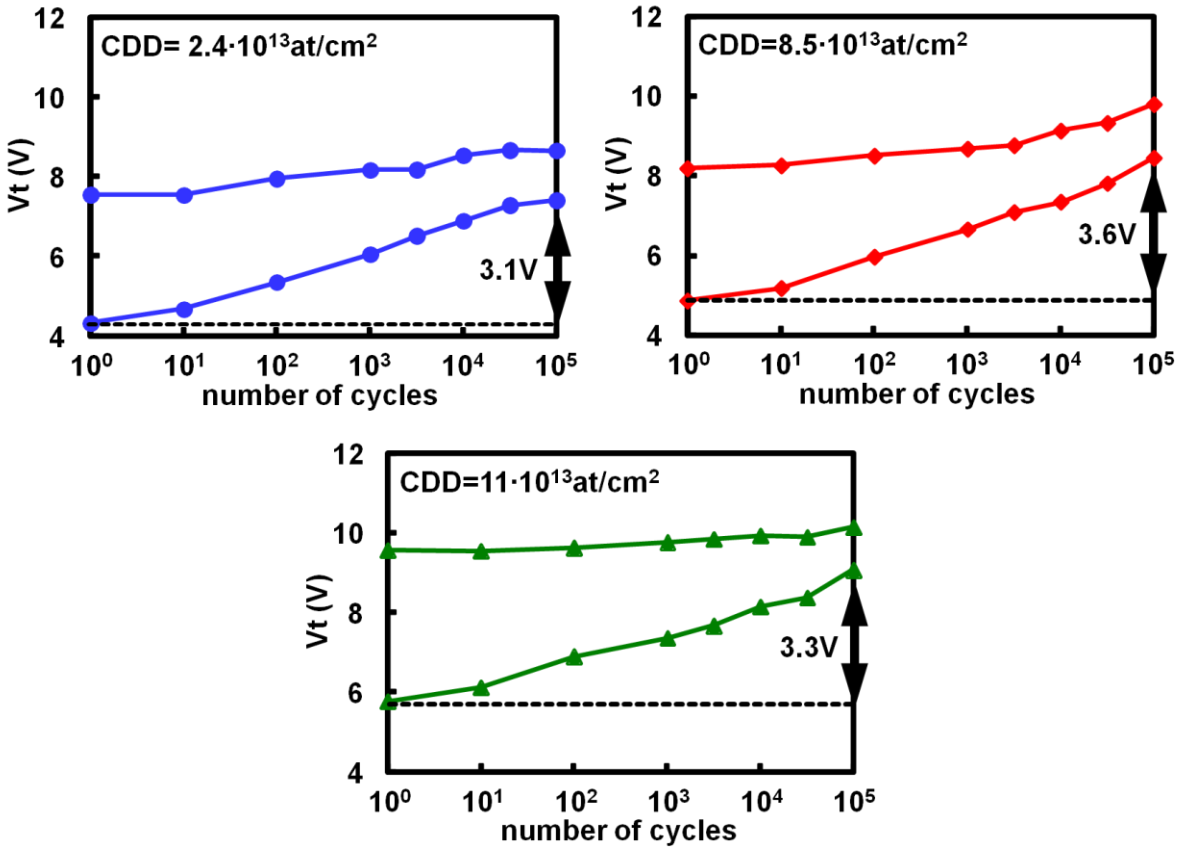


Figure 3. 9. Endurance characteristics of silicon nanocrystal cell comparing the samples with different channel doping doses (CDD): $2.4 \cdot 10^{13}$at/cm^2, $8.5 \cdot 10^{13}$at/cm^2 and $11 \cdot 10^{13}$at/cm^2. The cells are programmed by CHE (Vg=9V, Vd=4.2V, tp=5µs) and erased by FN (Vg=-18V, ramp=5kV/s+te=10ms).

One can notice the shift of programming window toward higher voltages at the cycling beginning due to the increase of CDD as described in chapter 2. Moreover in this case the parasitic charge trapping in Si_3N_4 capping layer is present. This is demonstrated by the shift of threshold voltage during the cycling. The closure of programming window is evident for the three samples, but the highest CDD presents the most stable programming threshold voltage and only the erase state shifts. In order to summarize the results, we reported in table 3.3 the values of programming windows before and after the cell cycling and the

program/erase threshold voltages shifts. For the first time we achieved 100kcycles endurance characteristic by keeping the programmed and erased states separated using the CDD=1.1·10^{13}at/cm^2 and the erase time te=10ms, but we can note that the programming window is only 1.1V. That is not enough to achieve a good cell functioning, hence other improvements on the memory cell stack are necessary. In the next section the impact of tunnel oxide thickness will be studied.

CDD (at/cm^2)	Programming window @1cycle	Programming window @100k cycles	Vtp shift	Vte shift	Endurance limit
2.4·10^{13}	3.2V	1.3V	1.1V	3.1V	90kcycles
8.5·10^{13}	3.3V	1.3V	1.6V	3.6V	70kcycles
11·10^{13}	3.8V	1.1V	0.6V	3.3V	>100kcycles

Table 3. 3. Programming window before and after 100k program/erase cycles, and program/erase threshold voltage shift; the studied samples have a different channel doping doses: 2.4·10^{13}at/cm^2, 8.5·10^{13}at/cm^2 and 11·10^{13}at/cm^2.

3.3.4 Impact of tunnel oxide thickness

The last technological parameter we studied is the tunnel oxide thickness. We varied it between 3.7nm and 5.2nm, using the silicon nanocrystals capped by silicon nitride layer (Φ=9nm+SiN=2nm). The program/erase conditions are the same as the last experiments where the CDD is varied (figure 3.6 with te=10ms). In figure 3.10 we reported the results of endurance experiments.

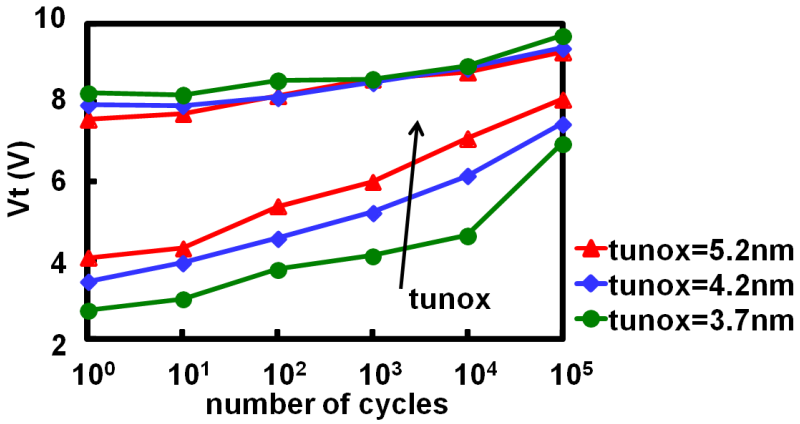


Figure 3. 10. Endurance characteristics of silicon nanocrystal cell comparing the samples with different tunnel oxide thickness (tunox): 3.7nm, 4.2nm and 5.2nm. The cells are programmed by CHE (Vg=9V, Vd=4.2V, tp=5µs) and erased by FN (Vg=-18V, ramp=5kV/s+te=10ms).

One can notice the limited influence of tunnel oxide thickness on CHE programming operation, while the FN erase is strongly impacted by this technological parameter. The endurance experiments confirm the erase efficiency improvement when the tunnel oxide thickness is decreased. This is due to the higher vertical electric field applied, as explained in

chapter 2. The results of programming windows before and after the cell cycling, and the program/erase threshold voltages shifts, are summarized in table 3.4. Once again the program/erase threshold voltage shift demonstrates the parasitic charge trapping in the silicon nitride layer. This effect slightly depends on tunnel oxide thickness because the charges are trapped during the CHE injection (independent on tunox). Concerning the erase operation we notice the improvement gained using thinner tunnel oxide (3.7nm) which enables the programmed and the erased state to be kept separate, while the working condition are very closed to the limit of good functioning for the case of the sample with tunox=4.2nm. In conclusion, to improve the programming window and the cell endurance, the experiments suggest minimizing the tunnel oxide thickness, but this contradicts the considerations on data retention, where the cell performance is increased by using a thicker tunnel oxide. It is important, thus, to find the best tradeoff between all the technological parameters to satisfy the data retention and endurance specifications.

tunox (nm)	Programming window @1cycle	Programming window @100k cycles	Vtp shift	Vte shift	Endurance limit
3.7	5.5V	2.7V	1.5V	4.2V	>100kcycles
4.2	4.5V	1.9V	1.4V	4.0V	>100kcycles
5.2	3.5V	1.2V	1.7V	4.0V	60kcycles

Table 3. 4. Programming window before and after 100k program/erase cycles, and program/erase threshold voltage shift; the studied samples have different tunnel oxide thicknesses: 3.7nm, 4.2nm and 5.2nm.

3.4 Silicon nanocrystal cell optimization

In this chapter we evaluated the impact of the main technological parameters of STMicroelectronics silicon nanocrystal memory cell. The results reported here concerning data retention and endurance experiments have been confirmed in literature, in particular the charge loss activation energy and the threshold voltage shift during the cycling due to the parasitic charge trapping. The aim of this analysis was to define the best way to achieve better results in terms of reliability of Si-nc memory cell.

In order to optimize the silicon nanocrystal cell we have taken into account the following points:

- In literature it is shown that the data retention is not strongly impacted by the nanocrystal size [Crupi '03] [Weihua '07] [Gasquet '06]. However this parameter impacts directly the programming window and the erase efficiency as shown in

chapter 2. Moreover, the increasing of covered area leads to a good cell functioning after 100k program/erase cycles with better erase efficiency during the cycling. In any case we noticed that the increasing of covered area, thus the cell coupling factor, using the studied cell architecture, is not sufficient to erase the parasitic charge trapping in the ONO control dielectric [Gerardi '07] [Jacob '07]. Hence it is important to increase the nanocrystal size and density, but further improvements on cell architecture are needed.

- The presence of silicon nitride capping layer on nanocrystals increases the charge trapping probability and the cell covered area. We described in section 3.2.1 the differences between the samples with nanocrystals entirely surrounded by Si_3N_4 layer and the samples where Si-nc are grown on the SiO_2 tunnel oxide and afterwards capped by Si_3N_4. Our cell corresponds to the second case. Thus, there are no benefits concerning the data retention due to the Si_3N_4 presence because the barrier to consider for the data retention corresponds only to the tunnel oxide thickness. Moreover, the presence of parasitic charge trapping at the tunox/Si_3N_4 interface facilitates the charge loss at high temperature. Concerning the cell endurance using the silicon nitrite capping layer, the coupling factor is increased and thus the programming window increases too, but the parasitic charge trapping in the Si_3N_4 does not enable good cell functionality after 100k program/erase cycles with the described memory stack. To avoid the parasitic charge trapping that worsens the Si-nc cell reliability for the optimization, we decided to skip this process step by adjusting the other technical parameters to achieve better results.

- The data retention is unchanged when varying the channel doping dose. It is thus possible achieve a programming window gain by increasing the channel doping dose and adjusting the program/erase levels if needed (chapter 2). Moreover, the increasing of CDD decreases the shift of programmed threshold voltage because the programmed threshold voltage is close to the saturation level; the CDD=10^{14}at/cm^2 is chosen for the optimized silicon nanocrystal memory cell.

- Finally we showed the charge loss dependence on tunnel oxide thickness and we extrapolated the activation energies related to the different samples. As in the case of erase operation, the charge loss increases as well as the tunnel oxide thickness. We noticed that a 5.2nm tunnel oxide is needed to achieve the data retention specification

for temperatures up to 150°C. On the other hand, the tunnel oxide thickness strongly impacts the erase Fowler-Nordheim operation. Hence, to achieve good cell functioning after 100k program/erase cycles, a tunnel oxide thickness of 3.7nm has to be used. Fo our study it is important now, to evaluate the cell behavior using a 4.2nm tunnel oxide, embedded in a different architecture without silicon nitride capping layer and an optimized ONO stack, in order to increase the vertical electric field and improve the program/erase efficiency and cell reliability.

3.4.1 Data retention optimization

It is clear now that if on the one hand the Si_3N_4 capping layer increases the programming window, on the other hand the Si-nc cell reliability is strongly degraded. We show below the results obtained using the silicon nanocrystal cell with the optimized stack. The latter is described in chapter 2 (figure 2.19): tunox=4.2nm, Φ=9nm or 12nm and the EOT of ONO layer is 10.5nm. In figure 3.11 we plotted the data retention results at 27°C, 150°C and 250°C of Si-nc cell with smaller nanocrystals (Φ=9nm). The charge loss is accelerated by the temperature, but the specification at 10 years is reached up to 150°C, despite the thinner tunnel oxide thickness with respect to the samples with the Si_3N_4 capping layer of figure 3.4.

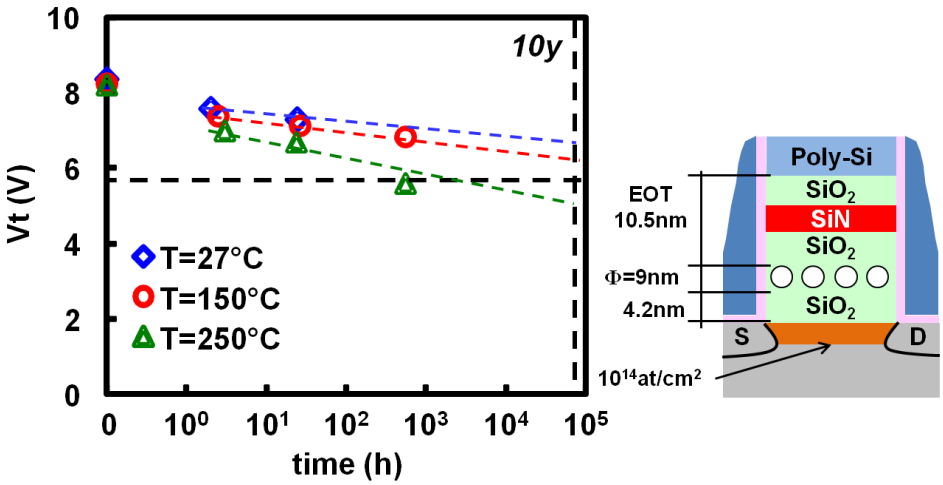


Figure 3. 11. Data retention at 27°C, 150°C, 250°C for the optimized silicon nanocrystal cell (tunox= 4.2nm).

In order to show the gain reached by avoiding the Si_3N_4 capping layer in figure 3.12, we plot the data retention characteristics at 150°C of optimized Si-nc cell (Φ=9nm) compared with the data of hybrid nanocrystals (Si-nc+SiN) cell; the two samples have the same tunnel oxide thickness (tunox=4.2nm). In this case the optimized Si-nc cell is able to retain the stored charge up to 10 years at 150°C. This result demonstrates that the data retention strongly improves by avoiding the silicon nitride capping layer.

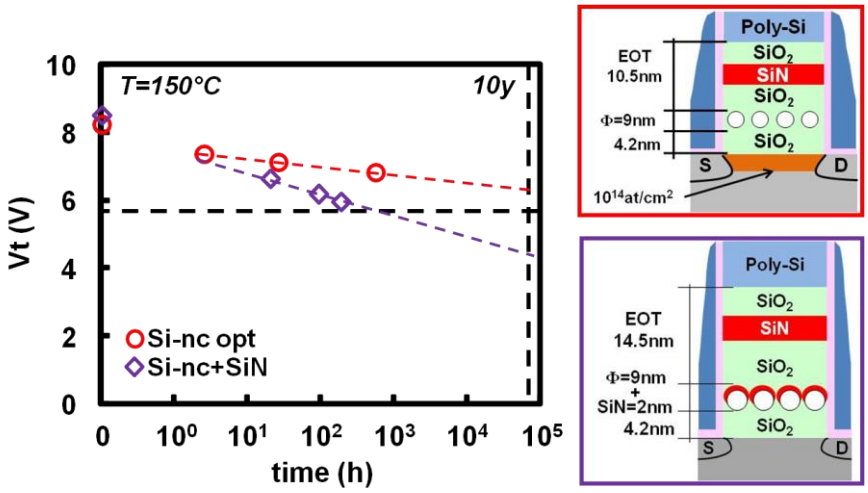


Figure 3. 12. Data retention characteristics of hybrid silicon nanocrystal cell and optimized Si-nc cell, integrating the same tunnel oxide thickness (4.2nm).

In the optimized memory stack silicon nanocrystals have been embedded with different size and density: Φ=9nm density=$7.3 \cdot 10^{11}$nc/cm^2 and Φ=12nm density=$6.7 \cdot 10^{11}$nc/cm^2 (figure 2.18). In figure 3.13 we compare the shift of programmed and erased threshold voltages versus time at 150°C of the optimized cell integrating the different silicon nanocrystal types. The impact of silicon nanocrystal size on data retention is limited, and the programming window closure due to the charge loss is unchanged. To complete the study on data retention of the optimized Si-nc memory cell we plotted in figure 3.14 the results before and after cycling, for the first time up to 1Mcycles, concerning the cell with bigger silicon nanocrystals (Φ=12nm). As published in [Monzio Compagnoni '04], quite unexpectedly the stressed cell displays the same Vt drift (i.e. stronger data retention) with respect to the virgin sample. The smaller leakage current is due to the negative charge trapping in the tunnel oxide or electron trapping at deep-trap states at the nanocrystal surface [Monzio Compagnoni '03].

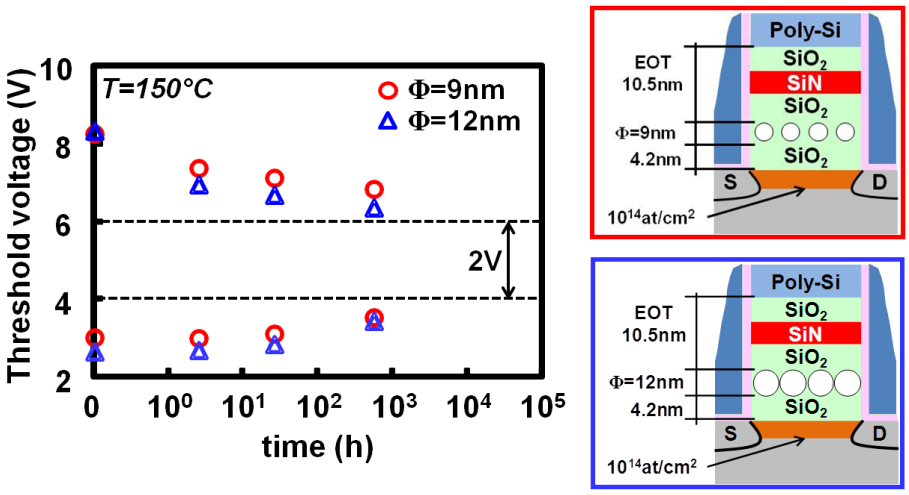


Figure 3. 13. Data retention characteristics of programmed and erased states at 150°C. Silicon nanocrystals with different sizes (Φ=9nm and Φ=12nm) are integrated in the optimized memory cell stack.

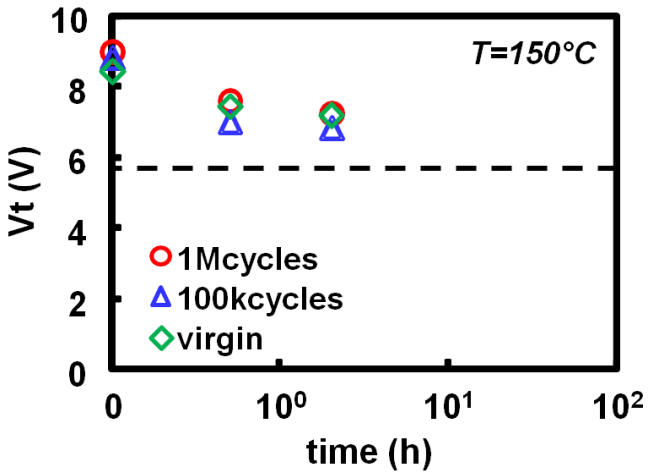


Figure 3. 14. Data retention characteristics of programmed Si-nc cells at 150°C. Stressed and virgin samples are compared.

3.4.2 Endurance optimization

We complete the optimized cell reliability characterization showing the data concerning the endurance cell degradation. The impact of Si_3N_4 on endurance is evaluated above. In figure 3.15 we plot the results of hybrid silicon nanocrystal cell compared with the optimized cell; the cell schematics are also shown. In order to achieve the same programming window (4V) at the beginning of cycling, we used different program/erase conditions:

- Hybrid Si-nc cell (Φ=9nm+SiN=2nm)
  - CHE programming: Vg=9V; Vd=4.2V; tp=5µs.
  - FN erase: Vg=-18V; te=ramp=5kV/s + 10ms.
- Optimized Si-nc cell (Φ=9nm)
  - CHE programming: Vg=9V; Vd=4.2V; tp=1µs.
  - FN erase: Vg=-18V; te=ramp=5kV/s + 1ms.

Using the optimized memory stack it is possible to decrease the programming and erase time thanks to the higher covered area and associated coupling factor. Moreover, by avoiding the Si_3N_4 capping layer, the erase efficiency and the endurance are greatly improved. A very slight shift of program/erase threshold voltages for the optimized sample results in 3.6V programming window after 100kcycles.

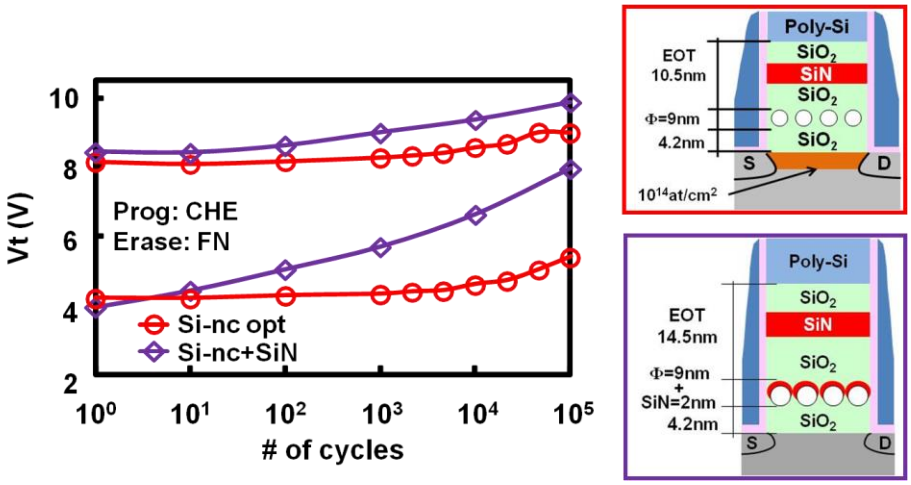


Figure 3. 15. Endurance characteristics of silicon nanocrystal memory comparing the hybrid Si-nc cell and the optimized Si-nc cell (Φ=9nm). The cells schematics are also shown.

This result was reached without a pre-cycling cell treatment. In fact using a positive or negative stress with high voltage, before cycling, helps accelerate the degradation process and improves the endurance performance with less memory window decrease [Yong '10]. The shifts of threshold voltages measured for the optimized Si-nc cell are not so marked, thus the pre-cycling treatment is not needed.

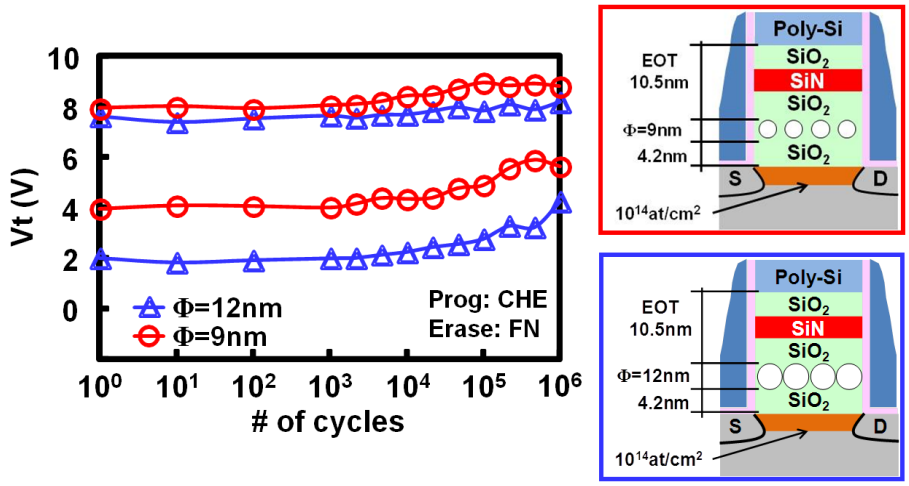


Figure 3. 16. Endurance characteristics of optimized silicon nanocrystal memory comparing two different nanocrystal sizes: Φ=9nm and Φ=12nm. The cells are programmed by CHE (Vg=9V, Vd=4.2V, tp=1μs) and erased by FN (Vg=-18V, ramp=5kV/s+te=1ms).The cells schematics are also shown.

After reaching this good results by using the optimized Si-nc cell we show in figure 3.16, for the first time to our knowledge, the 1Mcycles endurance characteristics of two optimized samples with different nanocrystal sizes (Φ=9nm and Φ=12nm) by achieving a large programming window. The cell with 12nm Si-nc is able to maintain a 4V programming window after 1Mcycles, improving the results published in [Ng '06]. In table 3.5 the values of the programming window before and after the cycling as well as the threshold voltage shifts are reported. Hence to improve the endurance performances up to 1Mcycles it is important to avoid the Si_3N_4 Si-nc capping, to increase the covered area and to use a thinner ONO layer.

Si-nc diameter	Programming window @1cycle	Programming window @1M cycles	Vtp shift	Vte shift	Endurance limit
Φ=9nm	4.0V	3.2V	0.9V	1.7V	>1Mcycles
Φ=12nm	5.6V	4.0V	0.6V	2.2V	>1Mcycles

Table 3. 5. Programming window before and after 1M program/erase cycles, and program/erase threshold voltage shifts; the studied samples have different nanocrystal sizes: Φ=9nm and Φ=12nm .

Using the program/erase conditions of experiments reported in figure 3.16, we repeated the cycling varying the temperature (T=-40°C up to T=150°C). In figure 3.17 the results are shown for the Si-nc cell with the higher covered area (sample with Φ=12nm and density=$6.7 \cdot 10^{11}$nc/cm^2). The programming window after 1Mcyles remains bigger than 4V and its value does not depend on the temperature.

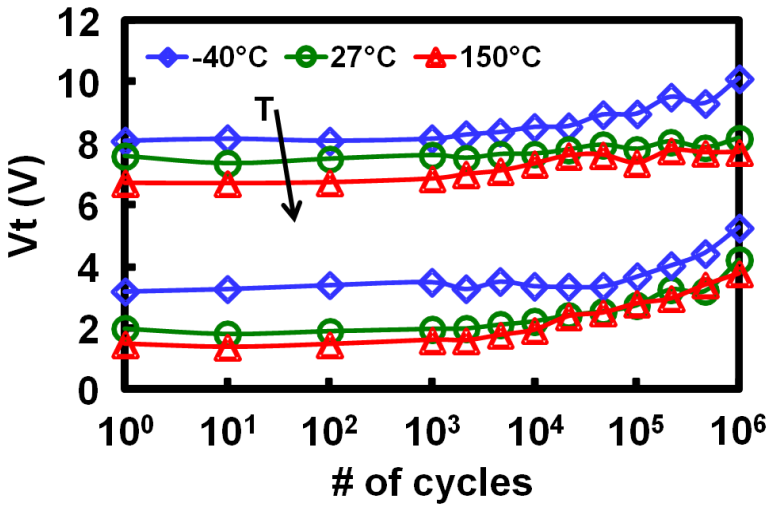


Figure 3. 17. Endurance characteristics of optimized silicon nanocrystal memory (Φ=12nm) at different temperatures: T=-40°C, T=27°C and T=150°C. The cells are programmed by CHE (Vg=9V, Vd=4.2V, tp=1µs) and erased by FN (Vg=-18V, ramp=5kV/s+te=1ms).

One can notice that by increasing the temperature the characteristic shifts toward the lower voltages. Both the programming and the erase operations are impacted by the temperature [Della Marca '13]. The programming efficiency decreases by increasing the temperature because the current in the channel decreases as well as the injection probability [Eitan '81] [Emrani '93]. Moreover, at low temperature, an increase in mobility is observed for Si-nc transistors generating a quasi-linear increase of the threshold voltage [Souifi '03]. In the case of FN erase, the efficiency increases with the temperature. This is justified assuming that the dominant conduction mechanism is assisted by traps [Zhou '09]. Therefore the programming window is bigger than 4V at the first cycle for all the temperatures and the shift of threshold voltages is due to the program/erase conditions that are kept unchanged. In table 3.6 the programming window before and after the cycling as well as the threshold voltage shifts are reported for different temperatures: T=-40°C, T=27°C and T=150°C.

Temperature	Programming window @1cycle	Programming window @1M cycles	Vtp shift	Vte shift	Endurance limit
-40°C	4.9V	4.9V	2.0V	2.0V	>1Mcycles
27°C	5.6V	4.0V	0.6V	2.2V	>1Mcycles
150°C	5.2V	4.0V	1.0V	2.3V	>1Mcycles

Table 3. 6. Programming window before and after 1M program/erase cycles, and program/erase threshold voltage shifts of Si-nc cell (Φ=9nm and Φ=12nm), at different temperature: T=-40°C, T=27°C and T=150°C.

3.5 Benchmarking with Flash floating gate

To conclude this chapter we compare the results concerning the optimized silicon nanocrystal memory cell with the standard Flash floating gate, keeping constant the cell size. In figure 3.18 the data retention at 250°C is shown for each device. We have seen previously (figure 3.11) that the optimized cell can maintain the programmed memory state 10 years up to 150°C. To satisfy our fixed data retention specification and to achieve the Flash floating gate results the cell must keep a program threshold voltage greater than 5.75V at 250°C up to 168h. One can notice that the Si-nc cell is just at the limit of this target, but further efforts concerning the tunnel oxide optimization are required to reach the standard floating gate device. The main constraint is the fast initial charge loss due to the charge trapping in the tunnel oxide, ONO layer and at the interfaces: substrate/tunox, Si-nc/oxide.

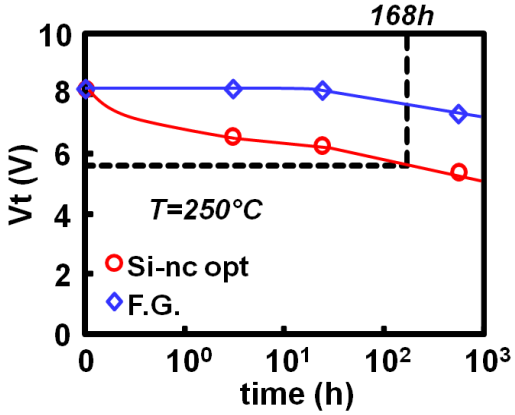


Figure 3. 18. Data retention of optimized silicon nanocrystal (Si-nc opt) and floating gate (F.G.) cells at 250°C.

To improve data retention of the optimized Si-nc cell one way is to increase the tunnel oxide thickness taking in to account the tradeoff with the programming window. Moreover special recipes of tunnel oxide growth can be developed playing on: time and temperature of process, oxide nitridation and preparation of the active surface to silicon nanocrystal nucleation. However these options will be taken in to account in future work. The endurance results are also compared keeping unchanged the program/erase conditions (CHE programming:

Vg=9V, Vd=4.2V, tp=1µs and FN erase: Vg=-18V, ramp=5kV/s+te=1ms). We considered the optimized Si-nc cell with a larger programming window (Φ=12nm); the data are plotted in figure 3.19. As expected, using the same program/erase conditions the Flash floating gate presents a larger programming window at the beginning of cycling (ΔVt=7V), thanks to its superior coupling factor and higher programming efficiency. The most significant threshold voltage degradation determines a major closure of programming window after 1Mcycles (ΔVt=2.8V), while the endurance characteristic remains more stable for the Si-nc cell. This is why for the floating gate device it is important to achieve a larger programming window. To understand the results qualitatively, we report in table 3.7 the programming window before and after the cycling as well as the threshold voltage shifts.

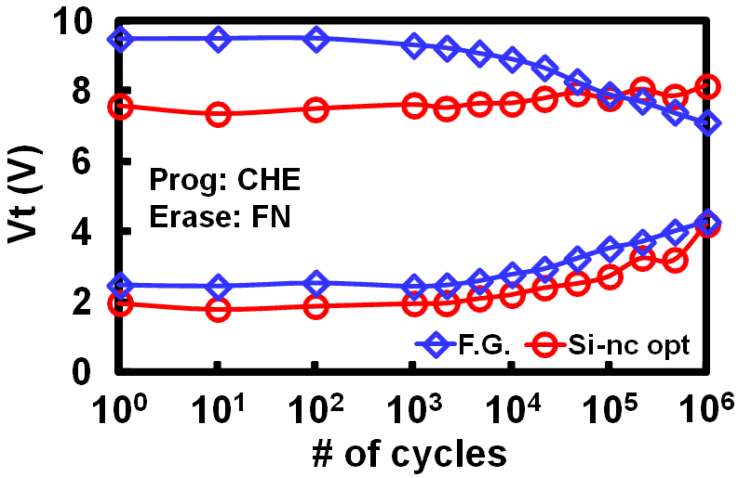


Figure 3. 19. Endurance characteristics of optimized silicon nanocrystal memory (Si-nc, Φ=9nm) compared with the Flash floating gate (F.G.). The cells are programmed by CHE (Vg=9V, Vd=4.2V, tp=1µs) and erased by FN (Vg=-18V, ramp=5kV/s+te=1ms).

	Programming window @1cycle	Programming window @1M cycles	Vtp shift	Vte shift	Endurance limit
Si-nc (Φ=12nm)	5.6V	4.0V	0.6V	2.2V	>1Mcycles
F.G.	7.0V	2.8V	-2.4V	1.8V	>1Mcycles

Table 3. 7. Programming window before and after 1M program/erase cycles, and program/erase threshold voltage shifts; the studied samples are the optimized silicon nanocrystal memory cell (Si-nc Φ=12nm), and the Flash floating gate (F.G.). Cycling conditions: CHE programming (Vg=9V, Vd=4.2V, tp=1µs) and FN erase (Vg=-18V, ramp=5kV/s+te=1ms).

We repeated the endurance experiments on Flash floating gate varying the program/erase conditions in order to achieve the same initial threshold voltage levels (CHE programming: Vg=8V, Vd=3.7V, tp=1µs and FN erase: Vg=-19.5V, ramp=5kV/s+te=1ms). The comparison with the optimized Si-nc cell is shown in figure 3.20. Quite unexpectedly the programming window of Flash floating gate starts to degrade after 100cycles, even if a lower drain and gate

voltages are used. This is not due to the tunnel oxide degradation, but it can be due to the lower programming efficiency caused by lower vertical and horizontal fields during the channel hot electron operation.

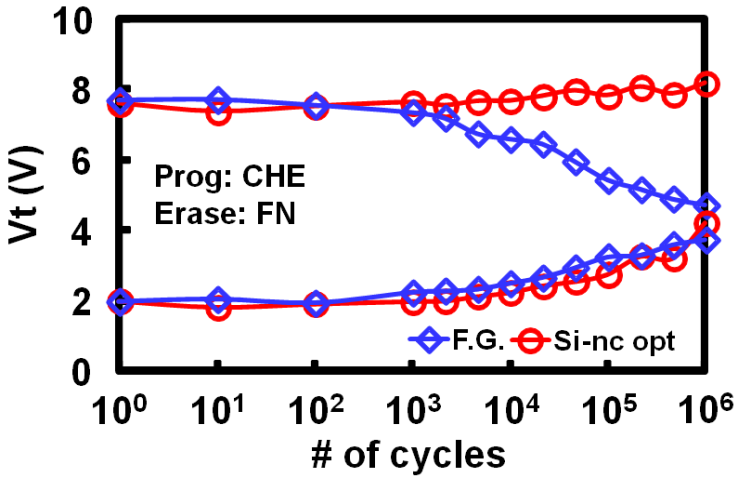


Figure 3. 20. Endurance characteristics of optimized silicon nanocrystal memory (Si-nc, Φ=12nm) programmed by CHE (Vg=9V, Vd=4.2V, tp=1µs) and erased by FN (Vg=-18V, ramp=5kV/s+te=1ms), compared with the Flash floating gate (F.G.) programmed by CHE (Vg=8V, Vd=3.7V, tp=1µs) and erased by FN (Vg=-19.5V, ramp=5kV/s+te=1ms).

In fact after the cycling experiments using Vg=9V and Vd=4.2V, we verified that it is possible to reach the program threshold voltage previously achieved (figure 3.19). Also in this case the Si-nc cell presents a more stable endurance characteristic; the data concerning the programming window before and after the cycling and threshold voltage shifts are reported in table 3.8. Finally we demonstrated for the first time to our knowledge the functioning of a silicon nanocrystal cell up to 1M program/erase cycles. A 4V programming window is preserved in a wide range of temperature [-40°C; 150°C]. As a drawback the silicon nanocrystal cell present a higher charge loss than the floating gate at 250°C.

	Programming window @1cycle	Programming window @1M cycles	Vtp shift	Vte shift	Endurance limit
Si-nc (Φ=12nm)	5.6V	4.0V	0.6V	2.2V	>1Mcycles
F.G.	5.7	1	-3	1.8	1Mcycles

Table 3. 8. Programming window before and after 1M program/erase cycles, and program/erase threshold voltage shifts. The studied samples are: optimized silicon nanocrystal memory cell (Si-nc Φ=12nm) programmed by CHE (Vg=9V, Vd=4.2V, tp=1µs) and erased by FN (Vg=-18V, ramp=5kV/s+te=1ms), and the Flash floating gate (F.G.) programmed by CHE (Vg=8V, Vd=3.7V, tp=1µs) and erased by FN (Vg=-19.5V, ramp=5kV/s+te=1ms).

Bibliography of chapter 3

[Chen '09] L.-L. Chen, C.-H. Chang, Y.-S. Lin, and Y.-H. Wu, *"Hybrid TiN nanocrystals/Si3N4 nonvolatile memory featuring low voltage operation by spinodal phase segregation"*, in Device Research Conference, 2009. DRC 2009, 22-24 June 2009, pp. 143-144, http://dx.doi.org/10.1109/drc.2009.5354957

[Chung '07] S. S. Chung, Y. H. Tseng, C. S. Lai, Y. Y. Hsu, E. Ho, C. Chen, L. C. Peng, and C. H. Chu, *"Novel Ultra-Low Voltage and High-Speed Programming/Erasing Schemes for SONOS Flash Memory with Excellent Data Retention"*, in Electron Devices Meeting, 2007. IEDM 2007. IEEE International, 10-12 Dec. 2007, pp. 457-460, http://dx.doi.org/10.1109/iedm.2007.4418972

[Colonna '08] J. Colonna, G. Molas, M. W. Gely, M. Bocquet, E. Jalaguier, B. De Slavo, H. Grampeix, P. Brianceau, K. Yckache, A.-M. Papon, G. Auvert, C. Bongiorno, and S. Lombardo, *"Realization of Hybrid Silicon core/silicon Nitride Shell Nanodots by LPCVD for NVM Application"*, in Materials Research Society Proceedings, http://dx.doi.org/10.1557/PROC-1071-F02-02

[Crupi '03] I. Crupi, D. Corso, G. Ammendola, S. Lombardo, C. Gerardi, B. DeSalvo, G. Ghibaudo, E. Rimini, and M. Melanotte, *"Peculiar aspects of nanocrystal memory cells: data and extrapolations"*, Nanotechnology, IEEE Transactions on, vol. 2, 2003, pp. 319-323, http://dx.doi.org/10.1109/TNANO.2003.820515

[Della Marca '13] V. Della Marca, J. Amouroux, G. Molas, J. Postel-Pellerin, F. Lalande, P. Boivin, and J.-L. Ogier, *"Improved Performance of Silicon Nanocrystal Memories for Application Working Over a Wide Range of Temperature"*, in 223rd Electrochemical Society ECS Meeting, Toronto, Ontario, Canada. , Accepted and to be published,

[Eitan '81] B. Eitan and D. Frohman-Bentchkowsky, *"Hot-electron injection into the oxide in n-channel MOS devices"*, Electron Devices, IEEE Transactions on, vol. 28, 1981, pp. 328-340, http://dx.doi.org/10.1109/T-ED.1981.20336

[Emrani '93] A. Emrani, F. Balestra, and G. Ghibaudo, *"Generalized mobility law for drain current modeling in Si MOS transistors from liquid helium to room temperatures"*, Electron Devices, IEEE Transactions on, vol. 40, 1993, pp. 564-569, http://dx.doi.org/10.1109/16.199361

[Gasquet '06] H. P. Gasquet, R. F. Steimle, R. A. Rao, and R. Muralidhar, *"Impact of Nanocrystal Size on NVM Array Reliability Performance"*, in Non-Volatile Semiconductor Memory Workshop, 2006. IEEE NVSMW 2006. 21st, 2006, pp. 64-65, http://dx.doi.org/10.1109/.2006.1629496

[Gay '12] G. Gay, G. Molas, M. Bocquet, E. Jalaguier, M. Gely, L. Masarotto, J. P. Colonna, H. Grampeix, F. Martin, P. Brianceau, V. Vidal, R. Kies, T. Baron, G. Ghibaudo, and B. De Salvo, *"Performance and Modeling of Si-Nanocrystal Double-Layer Memory Devices With High- k Control Dielectrics"*, Electron Devices, IEEE Transactions on, vol. 59, 2012, pp. 933-940, http://dx.doi.org/10.1109/TED.2012.2182769

[Gerardi '02] C. Gerardi, G. Ammendola, M. Melanotte, S. Lombardo, and I. Crupi, *"Reliability and Retention Study of Nanocrystal Cell Array"*, in Solid-State Device Research Conference, 2002. Proceeding of the 32nd European, 24-26 September 2002, pp. 475-478, http://dx.doi.org/10.1109/essderc.2002.194971

[Gerardi '07] C. Gerardi, V. Ancarani, R. Portoghese, S. Giuffrida, M. Bileci, G. Bimbo, O. Brafa, D. Mello, G. Ammendola, E. Tripiciano, R. Puglisi, and S. A. Lombardo, *"Nanocrystal Memory Cell Integration in a Stand-Alone 16-Mb NOR Flash Device"*,

Electron Devices, IEEE Transactions on, vol. 54, 2007, pp. 1376-1383, http://dx.doi.org/10.1109/TED.2007.895868

[Gerardi '08] C. Gerardi, G. Molas, G. Albini, E. Tripiciano, M. Gely, A. Emmi, O. Fiore, E. Nowak, D. Mello, M. Vecchio, L. Masarotto, R. Portoghese, B. De Salvo, S. Deleonibus, and A. Maurelli, *"Performance and reliability of a 4Mb Si nanocrystal NOR Flash memory with optimized 1T memory cells"*, in Electron Devices Meeting, 2008. IEDM 2008. IEEE International, 15-17 Dec. 2008, pp. 1-4, http://dx.doi.org/10.1109/iedm.2008.4796823

[Ghosh '10] B. Ghosh, L. Hai, B. Winstead, M. C. Foisy, and S. K. Banerjee, *"Retention modeling of nanocrystalline flash memories: A Monte Carlo approach"*, in Semiconductor Electronics (ICSE), 2010 IEEE International Conference on, 28-30 June 2010, pp. 203-205, http://dx.doi.org/10.1109/smelec.2010.5549551

[Hung-Bin '12] C. Hung-Bin, W. Yung-Chun, C. Lun-Chun, C. Ji-Hong, Y. Chao-Kan, and C. Chun-Yen, *"High-Reliability Trigate Poly-Si Channel Flash Memory Cell With Si-Nanocrystal Embedded Charge-Trapping Layer"*, Electron Device Letters, IEEE, vol. 33, 2012, pp. 537-539, http://dx.doi.org/10.1109/LED.2012.2184519

[Jacob '07] S. Jacob, L. Perniola, G. Festes, S. Bodnar, R. Coppard, J. F. Thiery, T. Pedron, E. Jalaguier, F. Boulanger, B. De Salvo, and S. Deleonibus, *"Investigation of Reliability Characteristics of Si Nanocrystal NOR Memory Arrays"*, in Non-Volatile Semiconductor Memory Workshop, 2007 22nd IEEE, 26-30 Aug. 2007, pp. 71-72, http://dx.doi.org/10.1109/nvsmw.2007.4290585

[Lee '09] D. U. Lee, T. H. Lee, E. K. Kim, J.-W. Shin, and W.-J. Cho, *"Analysis of charge loss in nonvolatile memory with multi-layered SiC nanocrystals"*, Applied Physics Letters, vol. 95, 2009, pp. 063501-3, http://dx.doi.org/10.1063/1.3205112

[Lee '03] J. J. Lee, X. Wang, W. Bai, N. Lu, J. Liu, and D. L. Kwong, *"Theoretical and experimental investigation of Si nanocrystal memory device with HfO2 high-k tunneling dielectric"*, in VLSI Technology, 2003. Digest of Technical Papers. 2003 Symposium on, 10-12 June 2003, pp. 33-34, http://dx.doi.org/10.1109/vlsit.2003.1221072

[Min-Feng '12] H. Min-Feng, W. Yung-Chun, T. Shun-Cheng, and C. Jiang-Hung, *"Polycrystalline-Si TFT TANOS Flash Memory Cell With Si Nanocrystals for High Program/Erase Speed and Good Retention"*, Electron Device Letters, IEEE, vol. 33, 2012, pp. 649-651, http://dx.doi.org/10.1109/LED.2012.2188826

[Molas '07] G. Molas, M. Bocquet, J. Buckley, J. P. Colonna, L. Masarotto, H. Grampeix, F. Martin, V. Vidal, A. Toffoli, P. Brianceau, L. Vermande, P. Scheiblin, M. Gely, A. M. Papon, G. Auvert, L. Perniola, C. Licitra, T. Veyron, N. Rochat, C. Bongiorno, S. Lombardo, B. De Salvo, and S. Deleonibus, *"Thorough investigation of Si-nanocrystal memories with high-k interpoly dielectrics for sub-45nm node Flash NAND applications"*, in Electron Devices Meeting, 2007. IEDM 2007. IEEE International, 10-12 Dec. 2007, pp. 453-456, http://dx.doi.org/10.1109/iedm.2007.4418971

[Molas '10] G. Molas, L. Masoero, P. Blaise, A. Padovani, J. P. Colonna, E. Vianello, M. Bocquet, E. Nowak, M. Gasulla, O. Cueto, H. Grampeix, F. Martin, R. Kies, P. Brianceau, M. Gely, A. M. Papon, D. Lafond, J. P. Barnes, C. Licitra, G. Ghibaudo, L. Larcher, S. Deleonibus, and B. De Salvo, *"Investigation of the role of H-related defects in Al2O3 blocking layer on charge-trap memory retention by atomistic simulations and device physical modelling "*, in Electron Devices Meeting (IEDM), 2010 IEEE International, 6-8 Dec. 2010, pp. 22.5.1-22.5.4, http://dx.doi.org/10.1109/iedm.2010.5703414

[Monzio Compagnoni '03] C. Monzio Compagnoni, D. Ielmini, A. S. Spinelli, A. L. Lacaita, C. Gerardi, L. Perniola, B. De Salvo, and S. Lombardo, *"Program/erase dynamics and channel conduction in nanocrystal memories"*, in Electron Devices Meeting, 2003. IEDM '03 Technical Digest. IEEE International, 8-10 Dec. 2003, pp. 22.4.1-22.4.4, http://dx.doi.org/10.1109/iedm.2003.1269342

[Monzio Compagnoni '04] C. Monzio Compagnoni, D. Ielmini, A. S. Spinelli, A. L. Lacaita, and C. Gerardi, *"Study of nanocrystal memory reliability by CAST structures"*, Solid-State Electronics, vol. 48, 2004, pp. 1497-1502, http://dx.doi.org/10.1016/j.sse.2004.03.014

[Ng '06] C. Y. Ng, T. P. Chen, M. Yang, J. B. Yang, L. Ding, C. M. Li, A. Du, and A. Trigg, *"Impact of programming mechanisms on the performance and reliability of nonvolatile memory devices based on Si nanocrystals"*, Electron Devices, IEEE Transactions on, vol. 53, 2006, pp. 663-667, http://dx.doi.org/10.1109/TED.2006.870281

[Padovani '10] A. Padovani, L. Larcher, D. Heh, G. Bersuker, V. Della Marca, and P. Pavan, *"Temperature effects on metal-alumina-nitride-oxide-silicon memory operations"*, Applied Physics Letters, vol. 96, 2010, pp. 223505-223505-3, http://dx.doi.org/10.1063/1.3446835

[Souifi '03] A. Souifi, P. Brounkov, S. Bernardini, C. Busseret, L. Militaru, G. Guillot, and T. Baron, *"Study of trap centres in silicon nanocrystal memories"*, Materials Science and Engineering: B, vol. 102, 2003, pp. 99-107, http://dx.doi.org/10.1016/S0921-5107(03)00013-8

[Steimle '03] R. F. Steimle, M. Sadd, R. Muralidhar, R. Rajesh, B. Hradsky, S. Straub, and B. E. White, Jr., *"Hybrid silicon nanocrystal silicon nitride dynamic random access memory"*, Nanotechnology, IEEE Transactions on, vol. 2, 2003, pp. 335-340, http://dx.doi.org/10.1109/TNANO.2003.820817

[Tsai '01] W. J. Tsai, N. K. Zous, C. J. Liu, C. C. Liu, C. H. Chen, W. Tahui, S. Pan, L. Chih-Yuan, and S. H. Gu, *"Data retention behavior of a SONOS type two-bit storage flash memory cell"*, in Electron Devices Meeting, 2001. IEDM '01. Technical Digest. International, 2001, pp. 32.6.1-32.6.4, http://dx.doi.org/10.1109/iedm.2001.979614

[Tsung-Yu '10] C. Tsung-Yu, W. Yi-Hong, W. C. Y. Ma, K. Po-Yi, W. Kuan-Ti, L. Chia-Chun, Y. Chi-Ruei, Y. Wen-Luh, and C. Tien-Sheng, *"Characteristics of SONOS-Type Flash Memory With In Situ Embedded Silicon Nanocrystals "*, Electron Devices, IEEE Transactions on, vol. 57, 2010, pp. 1895-1902, http://dx.doi.org/10.1109/TED.2010.2051489

[Weihua '07] G. Weihua, L. Shibing, J. Rui, L. Qi, H. Yuan, W. Qin, and L. Ming, *"Analysis of Charge Retention Characteristics for Metal and Semiconductor Nanocrystal Non-volatile Memories"*, in Electron Devices and Solid-State Circuits, 2007. EDSSC 2007. IEEE Conference on, 20-22 Dec. 2007, pp. 141-144, http://dx.doi.org/10.1109/edssc.2007.4450082

[Yong '10] W. Yong, Y. Xiaonan, W. Qin, H. Zongliang, Z. Manhong, Z. Bo, and L. Ming, *"Pre-cycling with higher voltages for endurance improvement of silicon nanocrystal memory device"*, in Solid-State and Integrated Circuit Technology (ICSICT), 2010 10th IEEE International Conference on, 1-4 Nov. 2010, pp. 1265-1267, http://dx.doi.org/10.1109/icsict.2010.5667629

[Zhou '09] X. Zhou, K. Uchida, H. Mizuta, and S. Oda, *"Carrier transport by field enhanced thermal detrapping in Si nanocrystals thin films"*, Journal of Applied Physics, vol. 105, 2009, pp. 124518-5, http://dx.doi.org/10.1063/1.3151688

Chapter 4 – Cell consumption during the channel hot electron programming operation

4.1 Introduction

In this section we present the results concerning the current and energy consumption of floating gate and silicon nanocrystal memory cells during the channel hot electron programming operation. The current consumption evaluation of a Flash floating gate memory cell is measured using current/voltage converter or indirect technique. In this way it is not possible to understand the dynamic cell behavior and to measure the cell performances in a NOR architecture for a programming pulse of several microseconds. Moreover, the indirect method, which will be explained in the chapter, is not functional for silicon nanocrystal memories. In this context we developed a new experimental setup in order to measure dynamically the current consumption during a channel hot electron programming operation. This method helps to understand the dynamic behavior of two devices. The energy consumption is also evaluated using different bias and doping conditions. The aim was to characterize the impact of different parameters about the floating gate cell consumption and find the best tradeoff for the Si-nc cell. Furthermore the consumption due to the leakage of unselected cells in the memory array is measured in order to complete this study. In conclusion the consumption is optimized and compared for both devices, giving new solutions for low power applications.

4.2 Methods of Flash floating gate current consumption measurement

Today one of the most important challenges for Flash floating gate memory cells in view of low power applications is to minimize the current consumption during the Channel Hot Electron (CHE) programming operation. Specific consumption characterization technique is presented in the literature, but requires a complex measurement setup, limited by circuit time constants [Esseni '99] [Esseni '00b] [Maure '09]. As an alternative it is possible to calculate the cell consumption using the static drain current measurement on an equivalent transistor.

4.2.1 Standard current consumption measurement

In literature it is shown how to measure the drain current of a Flash floating gate using a dedicated experimental setup. In [Esseni '99] and [Esseni '00b] an I/V converter (shunt resistance + amplifier) is used to measure the drain current during the programming; two different configurations are shown in figure 4.1.

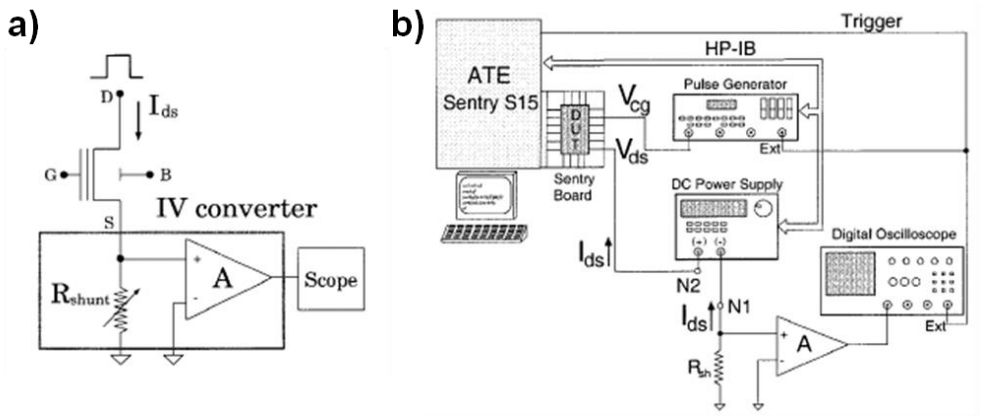


Figure 4. 1. a) Basic experimental setup to measure the programming drain current [Esseni '99]. b) Complex experimental setup used for ramped-gate programming experiments to measure the drain current absorption [Esseni '00b].

The development of the system shown in figure 4.1a introduces some errors on measured results due to the variable shunt resistance, the IV conversion and the coupling between IV converter and the scope. In order to measure the current absorption using a ramped gate programming the setup was improved (figure 4.1b) but the complexity of the system limits the current measurement sensibility and the writing pulse duration. Today the memories for embedded NOR architectures consume a current magnitude of the order of 50 microamperes for shorter time periods (several microseconds). Moreover A. Maure in his PhD thesis evaluated the error due to the IV conversion (figure 4.2), showing its relevance when the programming time decreases [Maure '09].

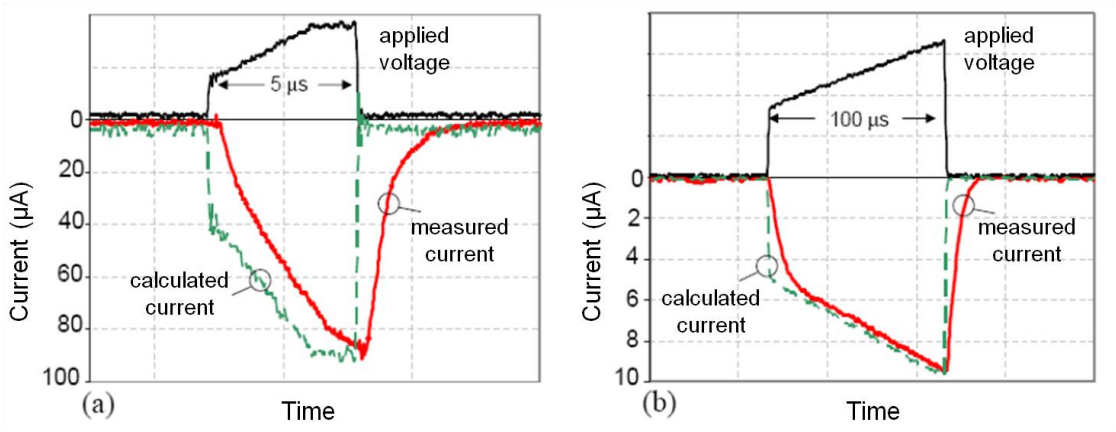


Figure 4. 2. Error evaluation during a fast current measurement performed applying a ramped voltage on a resistance. a)Very fast ramp induces a high current, the parasitic capacitive effect is important. b) The slow ramp shows induces a smaller current decreasing the measurement error.

4.2.2 Indirect current consumption measurement

As an alternative to the direct measurement on floating gate cell, it is possible to calculate the static current consumption during the programming operation starting from the drain current measurement on the equivalent transistor (called dummy cell), where the control gate and floating gate are shorted and the geometric dimensions (channel length and channel width) are kept unchanged. This technique enables the consumption to be calculated, regardless of the programming time, by using a commercial electrical parameter analyzer, while the current is considered to be constant during the programming. In order to explain this method we consider the following formula:

$$V_{FG} = \alpha_G V_G + \alpha_D V_D + \alpha_G \left(\frac{Vt_{treq}}{\alpha_G} - Vt\right) \qquad (4.1)$$

Where:

α_G and α_D : are the coupling factor relative to the gate and drain terminals;

V_G and V_D : control gate and drain voltages during the programming operation;

Vt_{treq} : threshold voltage measured for the equivalent transistor (dummy cell);

Vt : threshold voltage of floating gate cell during the programming operation.

Defining the overdrive voltage as:

$$V_{OV} = V_G - Vt \qquad (4.2)$$

We obtain:

$$V_{FG} = \alpha_G V_{OV} + \alpha_D V_D + Vt_{treq} \qquad (4.3)$$

The coupling factors have been calculated using the capacitance model shown in figure 4.3 and the cell dimensions. In this simple model the parasitic capacitances are also considered:

C_{D1} and C_{S1}: parasitic capacitances due to the coupling between floating gate and drain/source contacts.

C_X : parasitic capacitance due to the coupling between the selected floating gate and the neighbor cells in the same word line.

C_Y : parasitic capacitance due to the coupling between the selected floating gate and the neighbor cells in the same bit line.

We calculated for the standard Flash floating gate: α_G=0.67 and α_D=0.07.

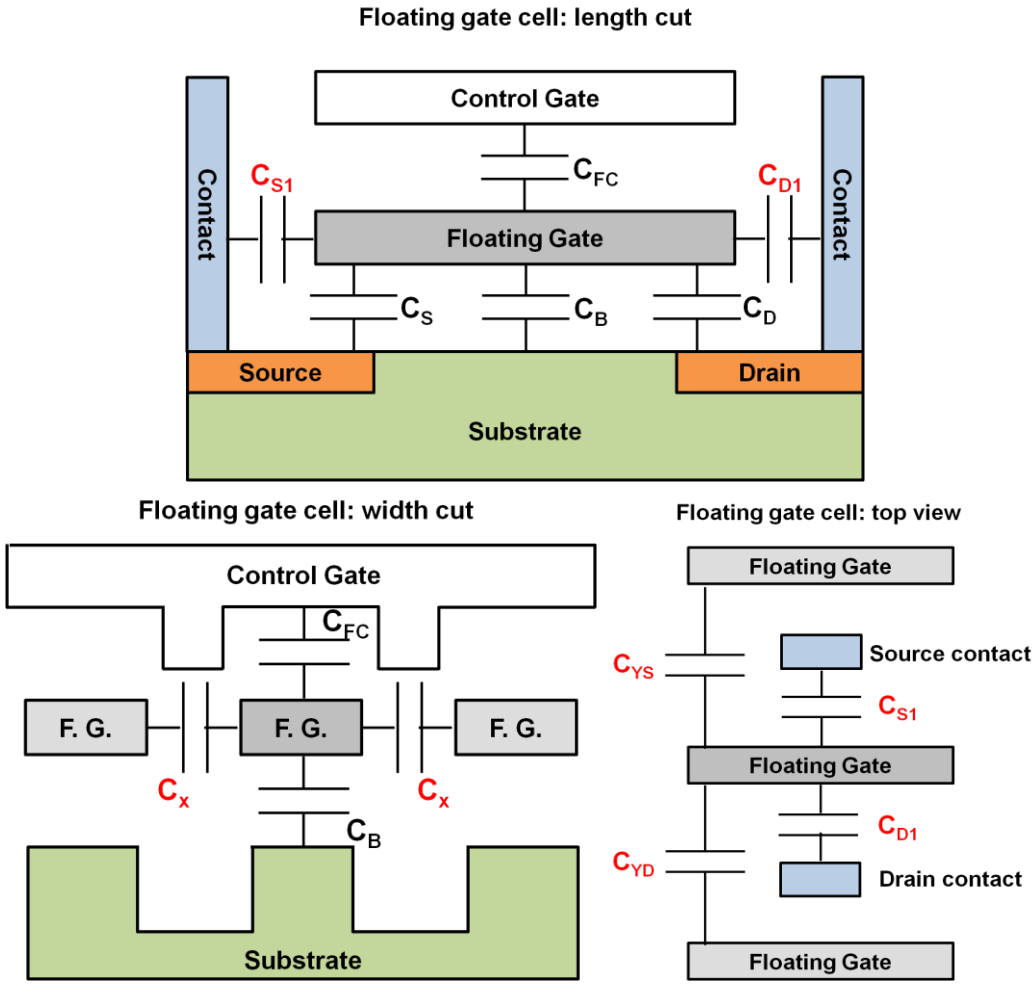


Figure 4. 3. Capacitive model used in order to calculate the Flash floating gate cell coupling factor.

In the case of α_G it is possible to compare the theoretical result with the measured values using different experimental techniques [Wong '92] [Choi '94]. This requires the static measurement of electrical parameters in both a floating gate cell and dummy cell. In figure 4.4 we reported the box plot of α_G calculated as the ratio between the subthreshold slope of dummy cell and subthreshold slope of floating gate cell. The cell dimensions are: W=90nm and L=180nm. One can notice the effect of dispersion on wafer related to the process variations (tunnel oxide integrity, source and drain implantation, channel doping dose, geometrical effect, etc..), and coupling factor calculation (36 samples are tested).

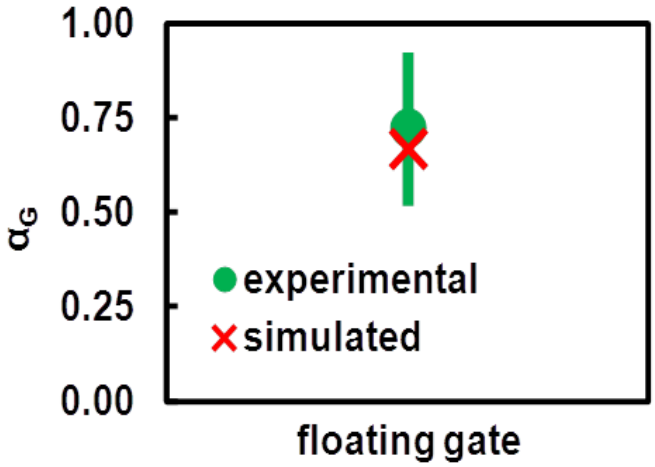


Figure 4. 4. Floating gate coupling factor (α_G) measured with the method of subthreshold slope with the error bar (36 samples) and simulated with the capacitive model.

Another important parameter to measure in order to calculate the floating gate potential evolution during the programming operation is the dummy cell threshold voltage. The results are shown in figure 4.5. The average value of Vt_{treq} is 2.96V with a dispersion of 0.6V on wafer.

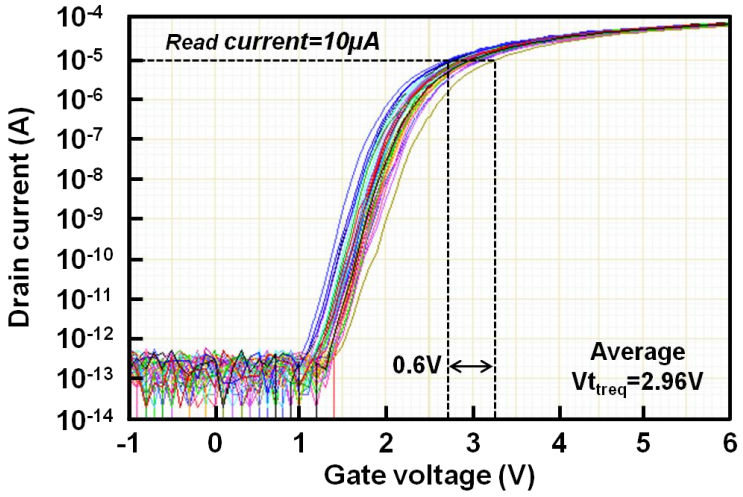


Figure 4. 5. Id-Vg characteristics of 36 dummy cells on wafer. The average value of Vt_{treq} is 2.96V using the read current of 10μA.

Finally the overdrive voltage is measured monitoring the Vt evolution during a programming operation. When the cell is programmed by a ramp, the V_{ov} remains constant as well as the CHE injection [Esseni '99].

In figure 4.6 the measured Vt, obtained by applying the 1.5V/µs ramp on the control gate with a drain voltage of 4.2V, are plotted. The ramped control gate voltage is emulated using a series of pulses as explained in chapter 2. In order to reduce the error due to the V_{ov} calculation, the overdrive is calculated as follow:

$$V_{OV} = V_G - \left(\frac{Vt_n + Vt_{n+1}}{2} \right) \qquad (4.4)$$

Where Vt_n and Vt_{n+1} are the threshold voltages of floating gate cell after the n^{th} and the $(n+1)^{th}$ pulse respectively. The calculated overdrive voltage is reported in figure 4.7.

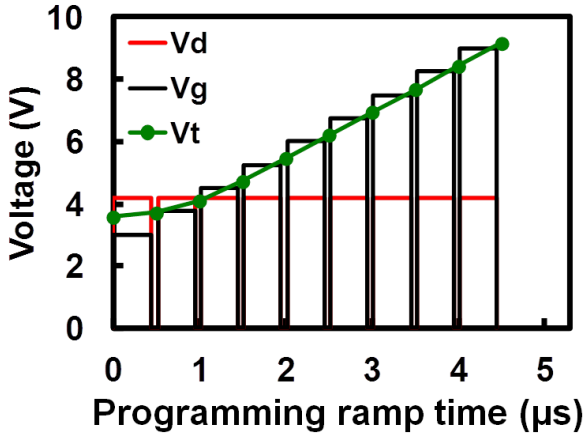


Figure 4. 6. Channel hot electron programming kinetic of floating gate cell. The control gate voltage (Vg) is a 1.5v/µs ramp, Vd=4.2V. Threshold voltage (Vt) read conditions: Id=10µA, Vd=0.7V.

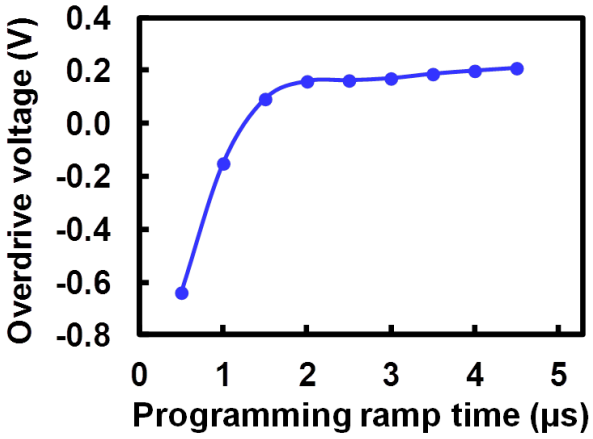


Figure 4. 7. Overdrive voltage during the channel hot electron programming operation calculated using the formula (4.4).

At this point it is possible to calculate the floating gate potential reached by the cell after the channel hot electron programming operation; its value is: V_{FG}=3.3V. In order to measure the static cell consumption, this potential is applied on the gate of the dummy cell maintaining the Vd=4.2V. In figure 4.8 we reported the drain current absorption (Id) under programming conditions measured for the equivalent transistor. During the experimental trials we noticed the degradation of Id level for consecutive measurements on the same device. This is due to the high voltage applied between gate and drain terminals by degrading the tunnel oxide. Thus the measurement has been performed using a sampling time as fast as possible (65µs) depending on the parameter analyzer speed.

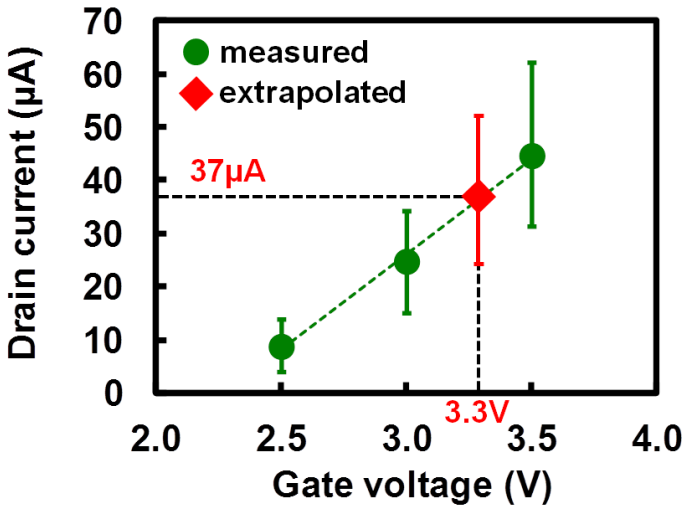


Figure 4. 8. Floating gate current consumption extrapolation, obtained measuring the dummy cell drain current.

Using this indirect method it is possible to evaluate the floating gate cell consumption by using the static measurements on an equivalent transistor (dummy cell). This procedure introduces a significant error due to the spread of dummy cell parameters, by assuming that the drain current absorption remains constant during the programming phase. This means the energy consumption is overestimated with respect to the real conditions. On the other hand the direct measurement on floating gate cell, by using the IV converter described above (figure 1.2), shows relevant limits when the programming pulse is short (several microseconds) due to the presence of parasitic capacitance in the measurement setup. That is why we have been motivated to develop a new measurement technique

4.2.3 New method of current consumption measurement

We discussed in paragraph 2.2 the complex measurement setup for dynamic drain current measurements. This method is affected by high time constant and does not enable measurement in a very short time period. The alternative indirect method of absorbed Id extrapolation that we previously described, is equally inaccurate and does not enable energy consumption calculation. We propose a new technique of measurement in order to measure the drain current during the CHE programming operation by using pulses of several microseconds. Our setup is shown in figure 4.9, where we use the Agilent B1500 equipped with two WGFMU (Waveform Generator and Fast Measurement Unit, Agilent B1530A) [Della Marca '11b] [Della Marca '11c]. In this way it is possible to set the sampling time at 10 ns by measuring the current dynamically. Moreover a power supply source can be connected using a low resistance switch matrix to the FG to complete the device biasing.

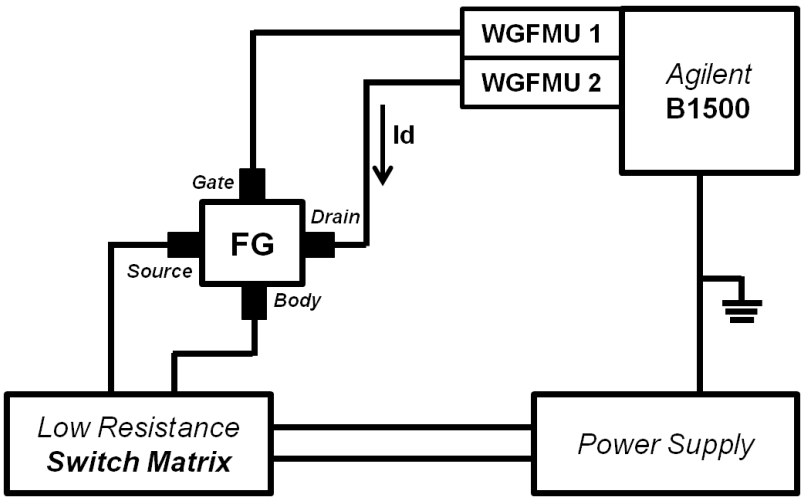


Figure 4. 9. Experimental setup used to perform dynamic drain current (Id) measurements during a channel hot electron programming operation. The device under test (DUT) is a floating gate cell addressed in a memory array of 256 kbit.

In figure 4.10 an example of dynamic drain current of floating gate cell is reported, that is measured using the ramp=1.5V/μs on gate terminal and the drain voltage at 4.2V. The static measured value corresponds to the value extrapolated using the indirect technique. With the developed setup we are able to generate arbitrary pulses on gate and drain terminals, in order to obtain specific ramps or boxes. The applied signals are measured like the dynamic current. One can notice that the drain current is not constant during the programming operation using a ramped gate pulse. The Id becomes constant when the equilibrium condition is reached [Esseni '99] and its quasi-static value decreases when the gate voltage remains constant. Thus the importance of this characterization technique is related to the cell behavior comprehension and the energy consumption calculation.

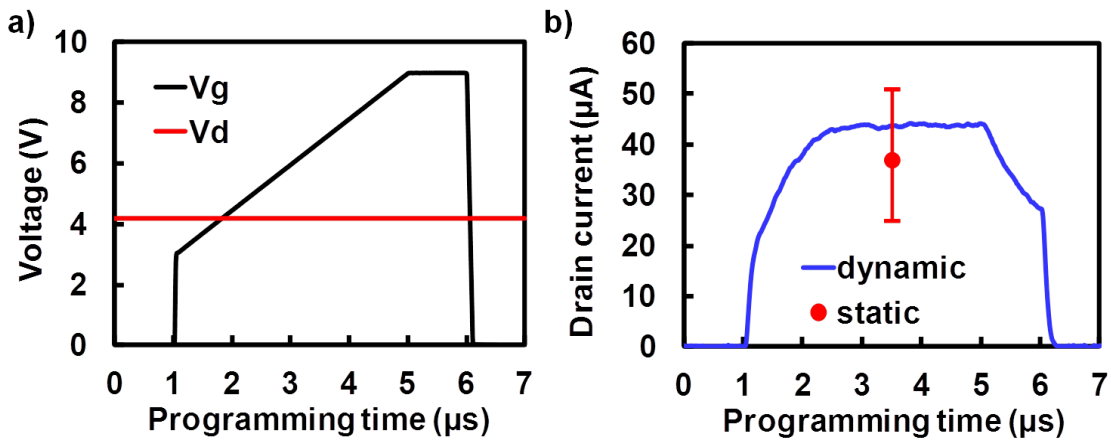


Figure 4. 10. a) Gate and drain voltages (Vg, Vd) generated with the setup of figure 4.9. b) Dynamic drain current measured during the channel hot electron programming operation. The static measurement is the same reported in figure 4.8.

4.3 Floating gate consumption characterization

In order to understand the cell's behavior during the channel hot electron operation and to optimize its performances, we decided to use the dynamic technique of measurement to evaluate the impact of the programming pulse shape; the impact of drain and bulk biases and the impact of technology (channel doping dose and lightly doped drain). The study on current and energy consumption during the programming operation is not limited to the single cell current absorption, but extended to the bitline leakage current as well. In the memory array, the unselected cells connected to the same bitline of selected cell, contribute to the global consumption with their drain/bulk junction leakage [Della Marca '13]. The principle of bitline leakage measurement will be explained in paragraph 4.3.2.

4.3.1 Cell consumption

4.3.1.1 Impact of programming pulse shape

We have seen before that in literature the floating gate behavior is described when the pulse gate is represented by box or by ramp. We applied our dynamic method to measure the drain current and consequently, the energy consumption and the programming window. The aim was to find the best tradeoff to improve the cell performances. In figure 4.11a the boxes applied on the control gate are shown, the ramp speed is 45V/µs and the drain voltage is constant at 4.2V. For all box pulses, the measured current peak is constant (figure 4.11b). When the gate voltage remains constant, the Id current quickly decreases following an exponential law. We have to plot the consumption data in arbitrary units (a. u.) to respect the STMicroelectronics data confidentiality. This means that it is possible to reach low energy

consumption levels and programming in a very short time. After each programming operation, the threshold voltage is measured and the cell is erased at the same start level. In this way, we calculate the programming window (PW) as the difference between programmed and erased threshold voltages. Then, the energy consumption (Ec) is calculated using the following formula:

$$Ec = \int_{t=0}^{t=tp} (Id \cdot Vd)dt \qquad (4.5)$$

Where tp is the programming time.

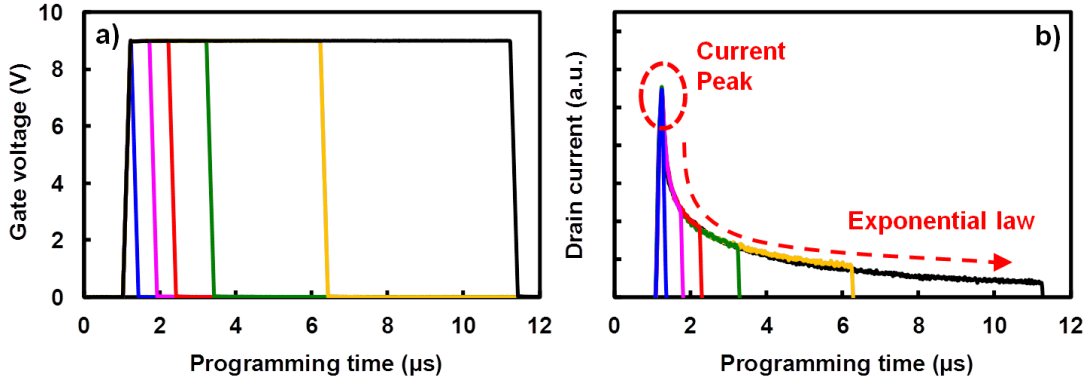


Figure 4. 11. (a) Gate box pulses applied during the channel hot electron operation; (b) drain current measured with dynamic method. Drain voltage is constant (Vd=4.2 V).

In figure 4.12 we report the trends of cell performances for different box pulse durations. The PW and the Ec increase with the box duration; in region I a low energy consumption level can be reached maintaining a satisfactory programming window. The presence of Id peak can be a problem for the logic circuits around the memory array. Furthermore, the designed charge pump layout areas, used to supply the drain terminal, depend on the value of Id current [Munteanu '02].

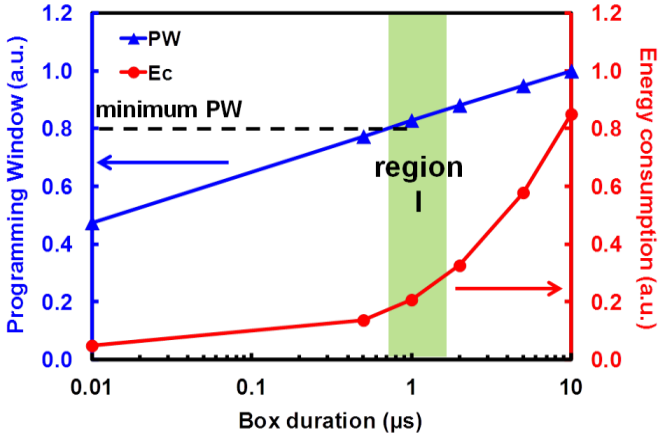


Figure 4. 12. Programming window and energy consumption versus the box duration. In region I a low energy level is reached maintaining the programming window.
The Y scale is normalized with the same factor as figure 4.14.

With the same method, we measured the drain current for different ramps applied on control gate (figure 4.13). It is worth noting that by increasing the ramp speed the Id current peak increases. On the contrary, when the ramp is slower, the Id current is smoothed (no peak), but the programming time increases. As explained before, we calculated the programming window and the energy consumption also in this case; the results are plotted in figure 4.14. The programming window and the energy consumption both decrease with the ramp speed increase. It is possible to reach a very low Ec, maintaining a good PW level, but inferior to the minimum specification (figure 4.14). This specification is due to the sense amplifier sensibility.

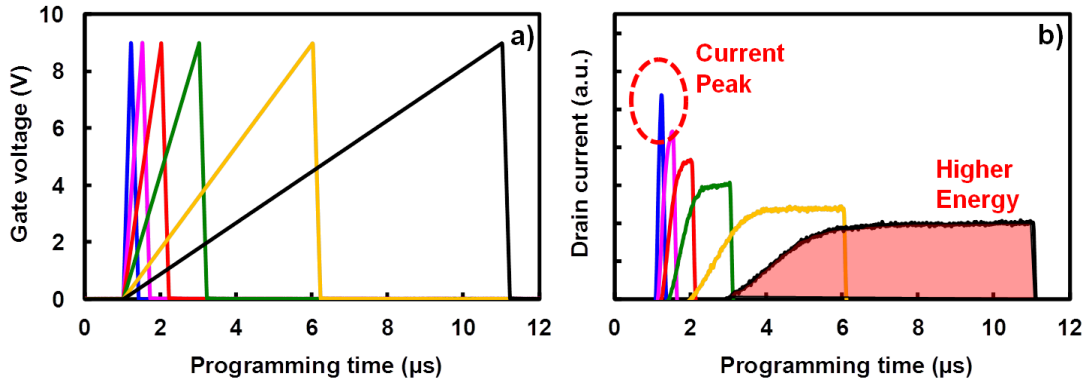


Figure 4. 13. (a) Gate ramp pulses applied during the channel hot electron programming operation; (b) drain current measured with dynamic method. Drain voltage is constant (Vd=4.2 V).

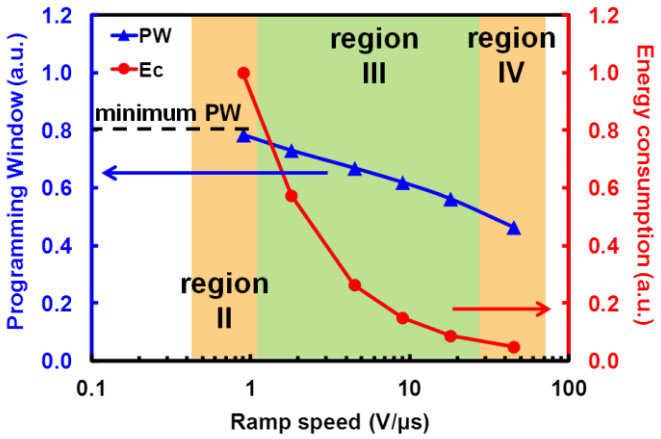


Figure 4. 14. Programming window and energy consumption trends versus the ramp speed. Three regions are highlighted: region II higher energy; region III best tradeoff; region IV higher current peak. The Y scale is normalized with the same factor as figure 4.12.

Three particular regions are highlighted in figure 4.14. Region II is where the Id is lowest with the highest energy consumption. On the contrary in the region IV the energy reaches the lowest value, but higher drain current peak is present. The best performances in terms of consumption are obtained in region III, but the programming window does not reach the minimum level. In order to resolve this conundrum, we decided to merge the two pulse types,

generating different combinations of ramp + plateau (figure 4.15). The measured drain currents are shown in figure 4.15b. The current peak becomes smooth decreasing the plateau duration and therefore the ramp speed. But when the Vg is constant, the current exponentially decreases. This is due to the floating gate potential variation that modifies the channel surface potential, thus the drain current. Using the gate pulses of figure 4.15a, the programming window, the energy consumption and the drain current peak are measured and plotted in figure 4.16. One can notice that the programming window and the energy consumption slightly depends on the gate pulse shape if the programming time is kept constant (tp=5µs).

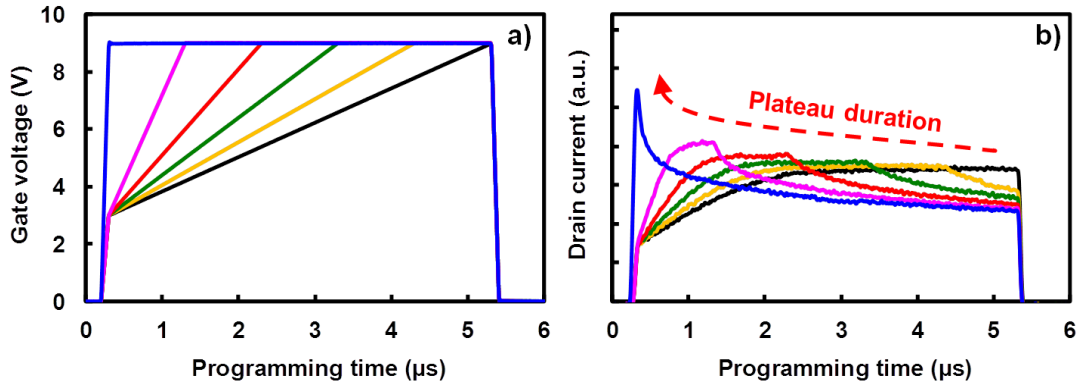


Figure 4. 15. (a) Gate pulses with different plateau durations, applied during the channel hot electron operation; (b) drain current measured with dynamic method; the current peak is smoothed decreasing the plateau duration, thus the ramp speed.

This study enables the gate pulse shape to be chosen with respect to the best compromise between the cell performances (PW, Ec, Id peak). The final amplitude of Vg and the programming time duration are kept constant at 9V and 5µs respectively (charge pump design constraints). In order to optimize the cell performances, we decided to merge a 1.5V/µs ramp with a 1µs plateau in order to avoid the problem of Id peak, maintaining a satisfactory programming window. The results are summarized in Table 1, where the gain/loss percentages are normalized with respect to the case of single box programming.

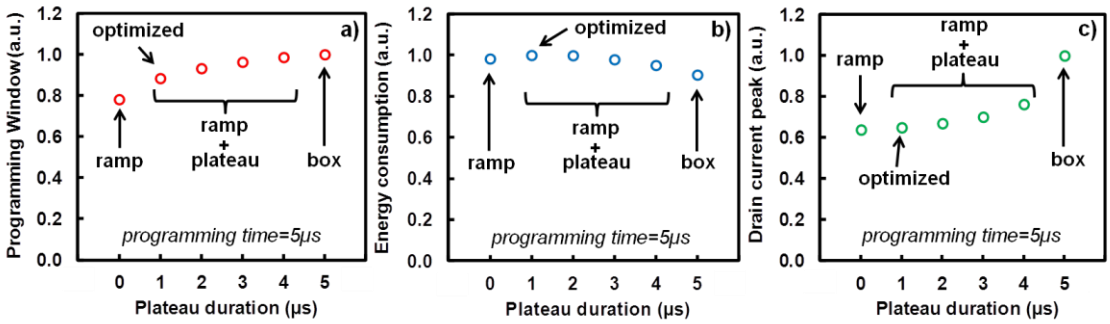


Figure 4. 16. (a) Programming window, (b) energy consumption and (c) drain current peak of floating gate cel,l measured using the pulses shown in figure 4.15a.

	5µs box pulse reference case		
	PW (%)	Ec (%)	Id peak (%)
Ramp	-22	+9	-36
Optimized	-11	+10	-35

Table 4. 1. Summary of results obtained using a single ramp or ramp + plateau (optimized) programming pulses, with respect to the box (reference case).

Using the new dynamic method of measurement, we characterized the device with different ramp and box programming signals. This procedure enables the best programming pulse shape to be chosen with respect to the final embedded low power product application. We have shown one possible optimization, with respect to the standard box pulse. The best trade-off reduces the current consumption by 35%. However this decreases the programming window by 11% and increases the energy consumption by 10%, if the programming duration is kept constant (5µs). Another improvement, concerning floating gate cell performances, can be obtained using the appropriate drain and bulk biases. This study will be presented in the next paragraph.

4.3.1.2 Impact of drain and bulk biases

Using the optimized pulse (ramp + plateau), we decided to study the dynamic cell behavior for several drain (Vd) and bulk (Vb) voltages.

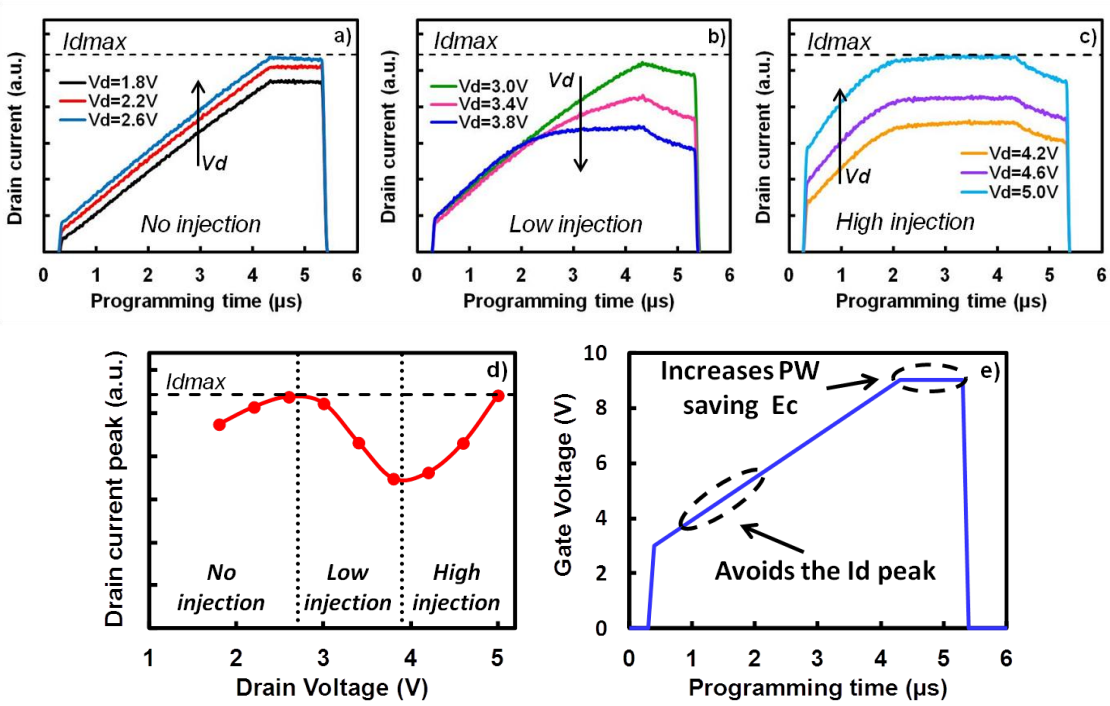


Figure 4. 17. Id currents measured with the dynamic method: (a) No injection zone, (b) Low injection zone, (c) High injection zone. (d) Maximum of drain current versus Vd. The Y scales are normalized with the same factor. (e) Optimized gate ramp pulse (1.5V/µs) + plateau (1µs).

For static characterizations, it is of common knowledge that Id increases with Vd. Static measurements do not take into account the charge injection in the floating gate and the variation of surface potential. Our dynamic measurements show that three different regions of work exist depending on drain biasing. In figure 4.17a-d, Id is measured during the channel hot electron operation, using the optimized gate pulse shown in figure 4.17e. We classified the curves in order to show different trends. In the *No injection zone* the drain current follows the Vg potential, demonstrating a transistor-like behavior that means the injection of electrons into the floating gate does not occur. By Increasing the Vd potential and repeating the same measurements, the drain current decreases due to the start of the injection; this is the *Low injection zone*. In the case of high amounts of Vd (*High injection zone*) the drain current by increasing again. The current variation in this region is attributed to the effects of the high fields applied between gate and substrate, as well as drain and source, and the channel modulation. These effects induce the hot carrier generation, thus increasing the drain current. In this case the channel is completely formed and pinched closely to the drain; the position of the pinch-off point is modulated by the drain voltage [Benfdila '04] [Moon '91] [Wang '79]. In figure 4.17d, it is worth noting that we found an optimal value of drain current between the low and high injection zones (Vd=3.8V). After each programming pulse, we measured the threshold voltage in order to evaluate the cell performances as shown in figure 4.18, where the PW and the Ec are plotted for different Vd. Clearly, by lowering the drain voltage to 3.8V with respect to Vd=5V, the energy consumption is minimized, with gain around 55% versus only 15% loss on the programming window.

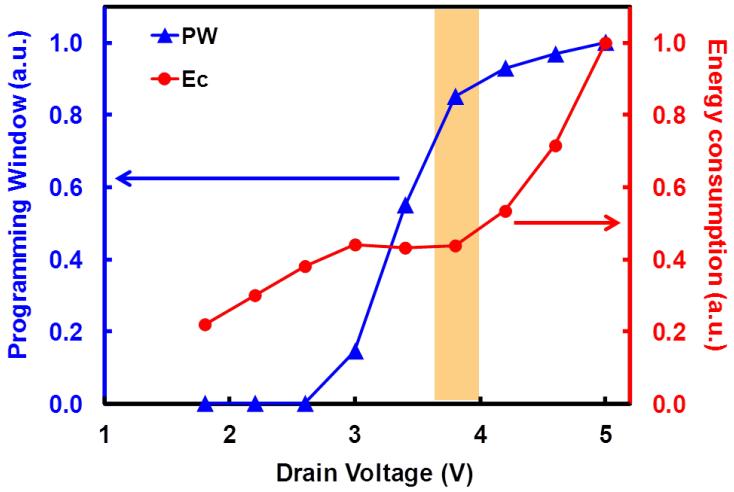


Figure 4. 18. Programming window and energy consumption versus drain voltage; the point of minimum energy is highlighted.

We repeated the experiment with the reverse body bias to benefit from the CHISEL effect [Takeda '83] [Driussi '04] [Esseni '00a]. The results are reported in figure 4.19. By increasing the amplitude of bulk bias the injection efficiency is increased, reaching a bigger programming window. In the meantime, the energy consumption of drain charge pump decreases due to the current reduction, allowing relaxed design constraints. Here we only considered this contribution, but by adding the substrate biasing, a bulk current is present. This current impacts the size of bulk charge pumps.

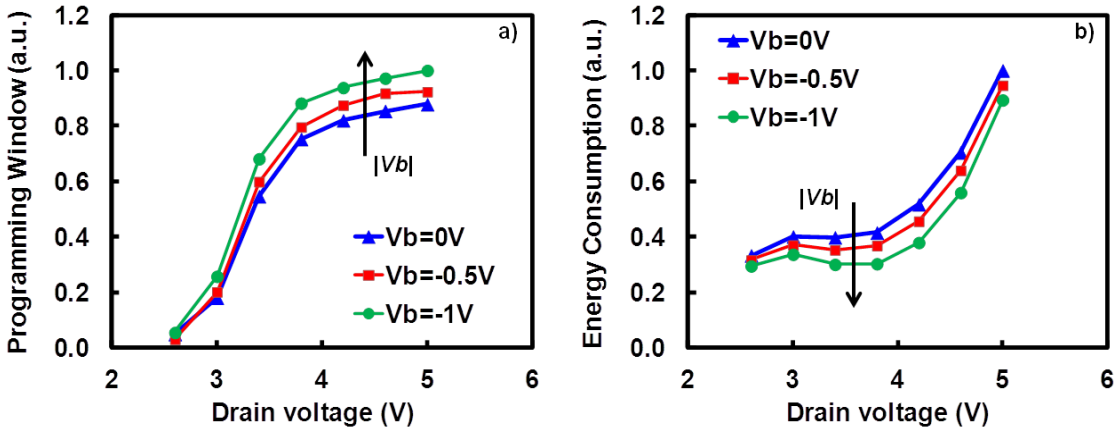


Figure 4. 19. Programming window and energy consumption versus drain voltage; the point of minimum energy is highlighted.

The limit for the maximum value of Vb depends on the breakdown voltage of drain/bulk junction. Finally several drain and bulk biases have been analyzed, optimizing the programming signals in order to reduce the current consumption, still keeping satisfactory performances. By decreasing the drain voltage from 4.2V to 3.8V, a diminution of 15% in terms of energy consumption can be achieved. Moreover, using the CHISEL effect with the reverse body bias, further drain current reduction is possible.

4.3.1.3 Impact of channel doping dose

After this study on programming pulses, where we understood the significant role of surface channel potential, we decided to modify an important technological parameter which can impact the cell consumption: the Channel Doping Dose (CDD). We tested the device described above using the optimized gate pulse (ramp + plateau) presented in figure 4.17e, with Vd=3.8V and Vb=-0.5V. The drain-bulk bias is chosen in order to minimize the stress of junction, and to decrease the bitline leakage of memory array (see next section). For each measurement we started from the same erased threshold voltage (Vte) in order to evaluate and compare the effect of CDD. Three different values of doping dose are used; the results are shown in figure 4.20. We can notice that an increase of CDD leads to the improvement of

the programming window and the energy consumption. When the channel doping dose is increased, the drain/bulk junction doping profile becomes more abrupt, and the cell threshold voltage (Vth) increases. Thus the electrons are turned into a more energetic state and the probability of injecting electrons in the floating gate increases, leading to better programming efficiency. Moreover, the drain current absorption decreases. Since the Vte increases as well as the CDD, it is necessary to adjust the erase pulse settings by increasing the erase time. After this study on single cell we will analyze in the next section the current consumption due to the unselected memory cells connected to the same bit line of programmed cell. Finally the global consumption of an entire bitline of 512 cells will be evaluated.

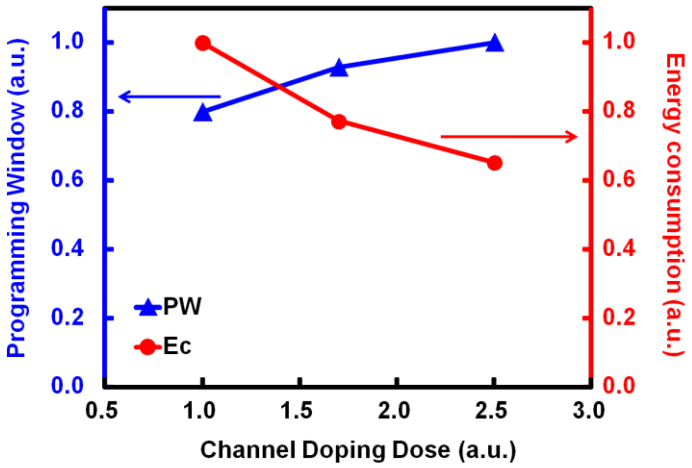


Figure 4. 20. Trends of programming window and energy consumption versus channel doping dose variation.

4.3.2 Bitline leakage

Another interesting point to evaluate in a NOR-architecture memory array is the BitLine Leakage (BLL), due to the Gate Induced Drain Leakage (GIDL) current in Band-to-Band Tunneling (BBT) regime during the CHE programming operation [Rideau '04]. As highlighted in previous studies [Mii '92] [Orlowski '89] [Touhami '01], several technological parameters, such as cell LDD doping (dose, tilt, and energy), drain–gate overlap, or STI shape have an impact on electric fields in the drain–bulk junction, responsible for GIDL. In this section, the impact of arsenic LDD implantation energy on bitline leakage measurements is presented through electrical characterizations, performed on a dummy cell structure.

4.3.2.1 Impact of lightly doped drain implantation energy

It is traditionally known that LDD process is used in classic MOSFETs to reduce lateral electric field by forming a gradual drain junction doping near the channel and as a consequence, to decrease the hot electron effect at the drain side [Nair '04] [Yimao '11]. In our case, we decided to use the LDD implants for the floating gate memory cell, in order to decrease the BLL caused by the drain bias of the unselected cells and to find a tradeoff between programming efficiency degradation and BLL improvement. As illustrated in figure 4.21, BLL measurement corresponds to the sum of the leakage current of all unselected cells on the same bitline (511 cells in this case), while a cell is being programmed. In order to measure it, we varied the gate potential of dummy cell between -3V (to emulate the floating gate potential of the programmed cell) and 1V (erased cell). The biasing conditions are chosen with regard to the results of section 3.1.2: Vd=3.8V, Vb=-0.5V and Vs=0V.

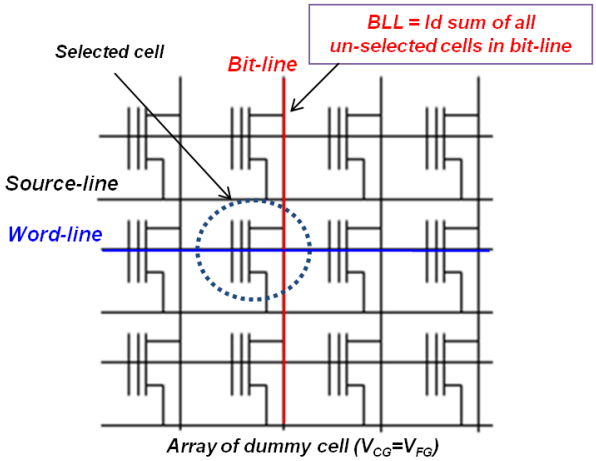


Figure 4. 21. Bit-line leakage measurement principle.

In the array we define the global consumption (I_{global}) as the sum of programming drain current of selected cell (I_{sel}) and leakage current of 511 unselected cells in the bitline (I_{unsel}). Moreover I_{unsel} is composed of two main current contributions: punchthrough leakage current (I_{pt}) and drain/bulk junction leakage current (I_{jc}). In figure 4.22a we show that the Ipt contribution is negligible with respect to I_{jc}, demonstrating that the bitline leakage current is mainly due to the drain/bulk junction contribution. One can notice that by increasing the LDD implantation energy the punchthrough current decreases. Even if the effective channel length slightly decreases, the doping variation in the LDD zones increases the barrier height of junctions. The global leakage of the unselected cell decreases with the LDD implantation energy; moreover the worst case is represented by the programmed cells in the bitline (figure 4.22b). In order to summarize the results of current consumption we plotted in figure 4.22c the percentages of global bitline current consumption (selected cell + 511unselected cells).

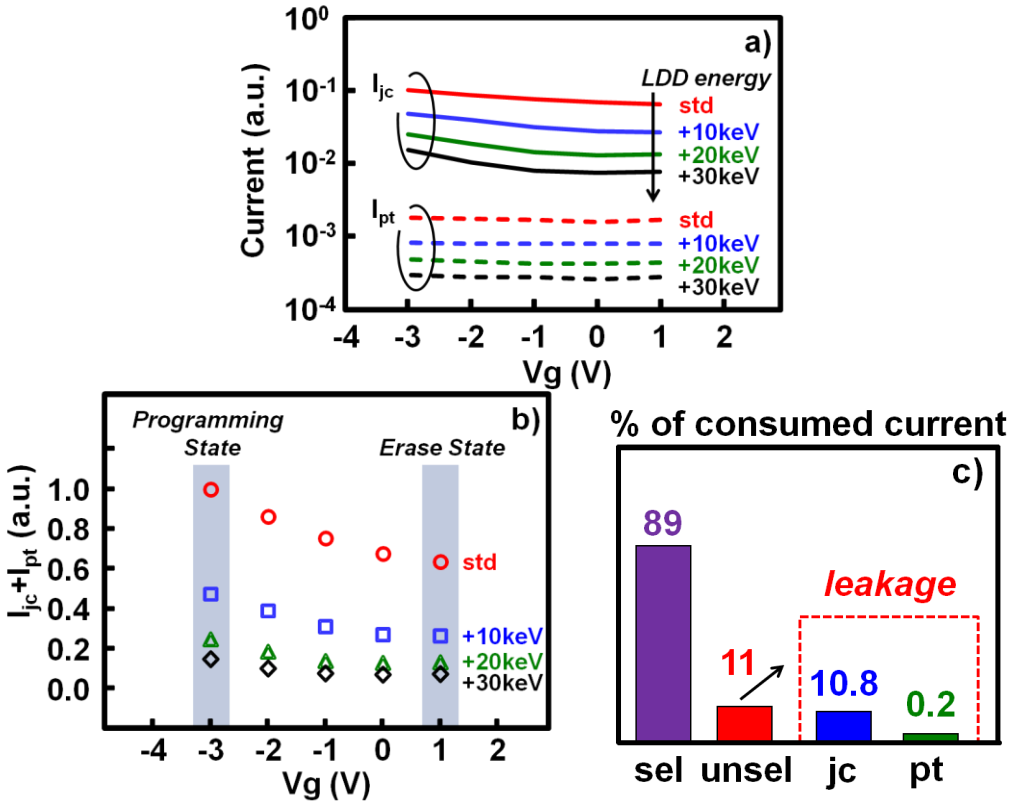


Figure 4. 22. (a) Effect of LDD implantation energy on the main leakage contributions of unselected cell: junction current (Ijc) and punchthrough current (Ipt) for different gate voltages. (b) Sum of Ijc and Ipt versus Vg for different LDD implantation energies. (c) Percentages of global bitline current consumption (selected cell + 511unselected cells).

4.3.2.2 TCAD simulations of LDD implantation

To understand the effect of LDD concerning the BLL mechanism, we performed TCAD investigation using Synopsys commercial tool for both process and electrical simulations. The process simulator parameters, i.e. doping diffusion and segregation coefficients, have been fine-tuned in order to obtain electrical results in accordance with experimental data. For electrical simulations, the hydrodynamic transport model has been adopted. Figure 4.23a shows the 2D cell net active doping profiles of LDD and drain–bulk junction, for several implantation energies. In figure 4.23b, the net doping profile along a horizontal cut below the Si/SiO2 interface is reported and five regions are identified, from left to right: source/LDD/channel/LDD and drain. We observe that the net doping level at the channel-LDD region (also corresponding to the gate–drain overlap region) decreases with implantation energy increase. It has been shown that doping and surface potential gradients have an impact on GIDL through the lateral electric field [Parke '92] [Rideau '10]. In the present case, a less abrupt net doping profile in the channel-LDD region for the highest implantation energy (figure 4.23b) leads to a lower lateral electric field and a smaller leakage current.

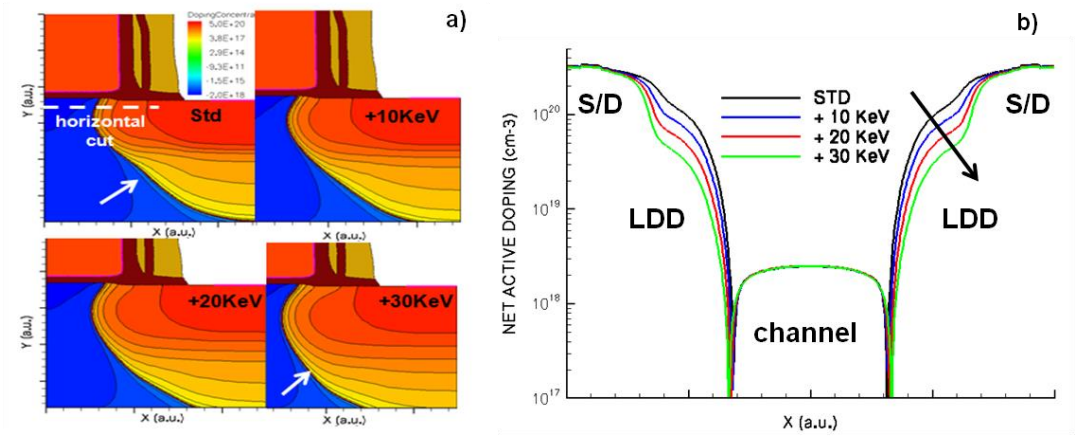


Figure 4. 23. TCAD simulations: (a) LDD cartography and (b) doping profiles at Si/SiO2 interface, for different implantation energies.

Figure 4.24a shows the distribution of the absolute value of the electric field (E) obtained by TCAD simulations. The electric field E is the highest at the gate edge, but its value at this point becomes smaller as LDD implantation energy increases. The lateral (Ex) and vertical (Ey) components of the electric field at the Si/SiO2 interface are shown in figures 4.24b and c respectively. It can be noticed that if, on the one hand, no significant variation is seen on the vertical electric field peak (figure 4.24c) and on the other hand, the lateral electric field peak decreases as LDD implantation energy increases (figure 4.24b). As previously mentioned, the reduction of the lateral electric field, and thus of the global electric field, decreases the leakage current of the drain–bulk junction, due to Band-to-Band Tunneling. Although the cell LDD implantation energy increase could help decrease the bitline leakage, we also have to take into account its impact on cell performances during the programming operation. In what follows, we will focus on the impact of implant energy on the write efficiency. Programming is performed on a standard cell using CHE injection. We bias the control gate and the drain with 9V and 3.8V box pulses respectively, and the bulk with -0.5V. In figure 4.25a the programming window and the bitline leakage are plotted versus the LDD implantation energy. This graph highlights the fact that the programming window is impacted by LDD energy and decreases as the energy increases, due to the reduction of the lateral electric field contribution. A satisfactory trade-off can be found reaching a gain of 49% in terms of BLL reduction, losing only 6% of PW and increasing by +10keV the standard LDD implantation energy. Further improvements can be made, with a +20keV increase, gaining 70% of BLL reduction against less than 10% loss on PW. This study has been performed by keeping the channel doping dose unchanged. In order to find the best tradeoff, it is important to take into account that when the CDD is increased, the BLL is increased too, because the lateral electric field is enhanced (figure 4.25b).

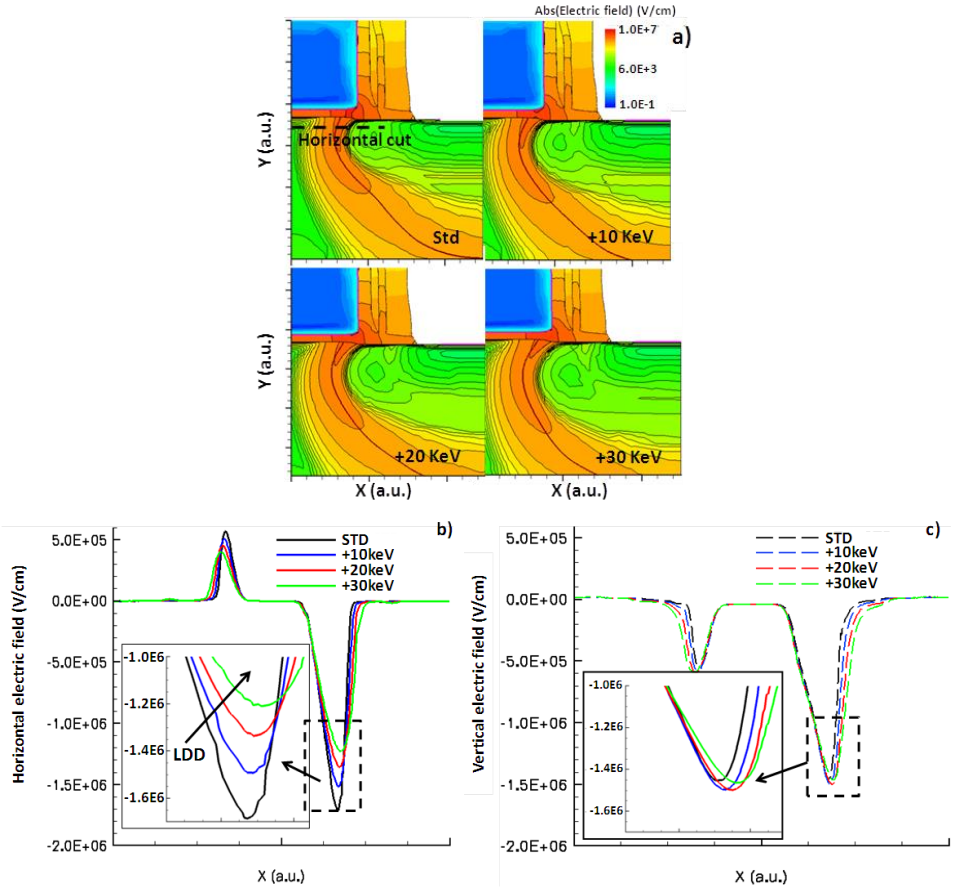


Figure 4. 24. TCAD simulations: (a) global electric field distribution; (b) lateral field Ex and (c) vertical field Ey along the channel for different LDD implantation energies.

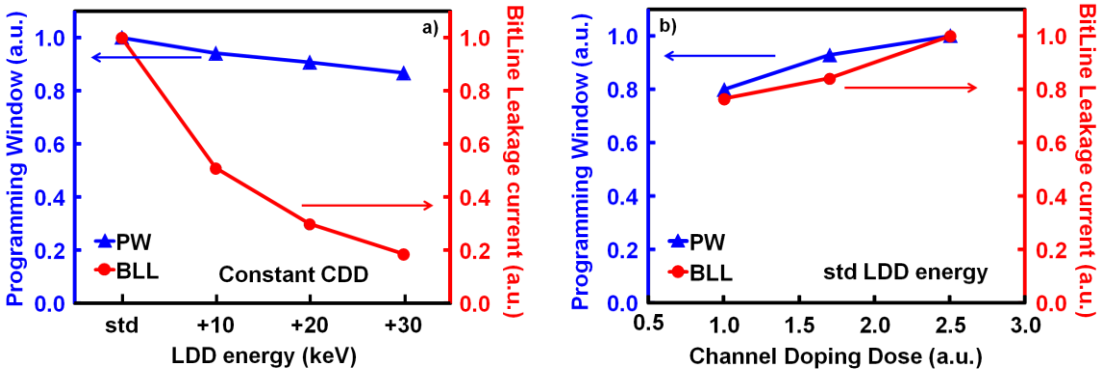


Figure 4. 25. Programming window and bitline leakage versus (a) LDD implantation energy (constant channel doping dose), and (b) channel doping dose (standard LDD energy implantation).

In conclusion we found a programming scheme optimization for the floating gate cell using the new dynamic method measurement for the drain current consumption. The study enables the best tradeoff to be found depending on cell application in terms of dynamic consumption and programming window. In addition we considered the impact of drain and bulk biases highlighting the optimum point of work for our technology using Vd=3.8V and Vb=-0.5V. Finally the impact of channel doping dose and lightly doped drain implantation energy have been studied to improve the consumption due to the unselected cells of bitline.

4.4 Silicon nanocrystal cell consumption characterization

The study of floating gate cell consumption helped understand the main parameters that can impact this aspect during the channel hot electron programming operation. We have seen the importance of new setup development for dynamic current consumption. Moreover, this method becomes compulsory because the discrete nature of nanocrystals does not enable a dummy cell to be designed that is useful for static measurement and thus the drain current consumption extrapolation. Here we present the impact of programming scheme and tunnel oxide thickness on s hybrid silicon nanocrystal cell (Si-nc+SiN). Finally we will show the consumption results on optimized Si-nc cell compared with those on the standard floating gate.

4.4.1 Impact of programming pulse shape

After the study on the floating gate cell, we used the same setup for dynamic current measurements to evaluate also the hybrid Si-nc cell with a 5.2nm tunnel oxide thickness. This device has been chosen for its higher programming window due to the SiN presence that increases the charge trap probability (see chapters 2 and 3). The programming window and consumption are evaluated by using box and ramp pulses and also considering the optimization used in the case of floating gate (ramp+plateau). In figure 4.26 we show the applied gate voltage pulses while the drain voltage is constant (Vd=4.2V); the source and bulk terminals are grounded (Vs=Vb=0V). The programming pulse varies from a box to a ramp of 1.2V/µs; between these two conditions each ramp is followed by a plateau of different duration to improve the programming window, while the programming time is kept constant (tp=5µs).

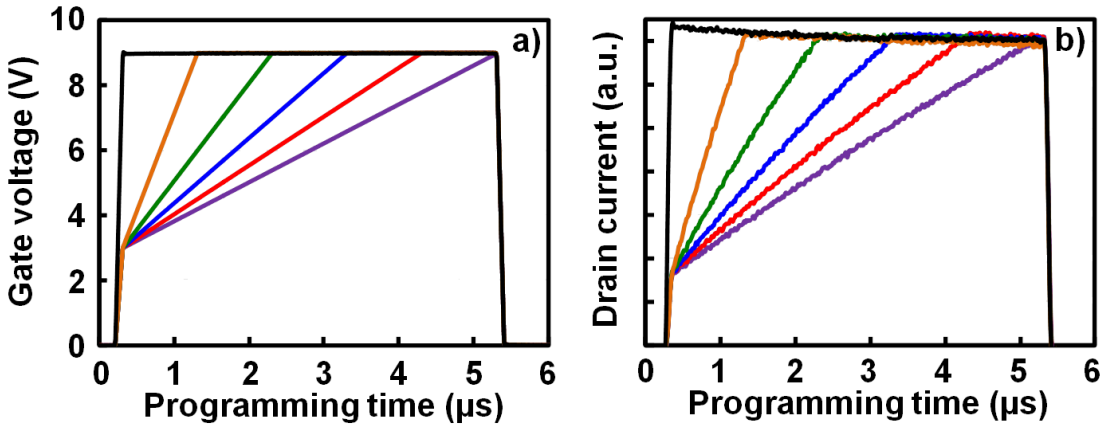


Figure 4. 26. (a) Gate pulses with different plateau durations, applied during the CHE operation; (b) drain current of Si-nc+SiN cell, measured with dynamic method; the current follows the gate potential.

We program the cell always by starting from the same threshold voltage, in order to measure the drain current and the programming window after each pulse. Using the dynamic measurement setup and considering the Id behavior, we can observe that the current follows the gate potential variation over the time. The cell behavior is very different with respect to the floating gate cell results observed above. These results, obtained for the Si-nc cell, suggest a transistor-like behavior during the programming operation. In particular we can notice that the peak is not present when the box pulse is applied, but the current remains constant during the programming phase. The design of memory array control circuits have to take into account this aspect, but no overshoot is possible. In figure 4.27 we reported a simple schematic of silicon nanocrystal cell in order to explain its behavior. During the CHE injection the charges are trapped only in the silicon nanocrystals and SiN capping layer close to the drain area. In this zone the horizontal electric field is stronger and the electrons are accelerated so as to be injected in trapping layers [Tam '84] [Takeda '85] [Ning '78] [Chenming '85]. During this operation a high potential is applied on drain terminal (Vd=4.2V). Then the Space Charge Region (SCR) on drain-body junction increases its area, reducing the channel length. Thus the stored charges remain trapped in Si-ncs and SiN over the drain and SCR zone. While the gate voltage changes, the hybrid Si-ncs on channel zone are not charged and so their potential remains constant. Consequently, the substrate surface potential depends on gate voltage only, which dynamically drives the drain current during the channel hot electron operation [Della Marca '11a].

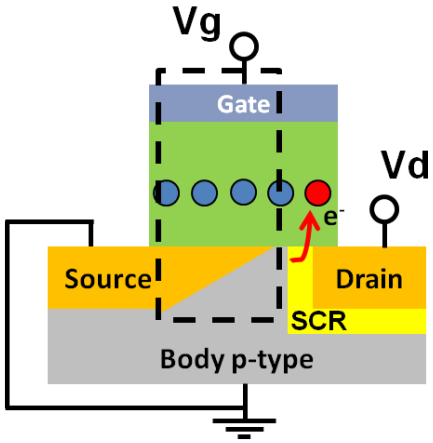


Figure 4. 27. Scheme of a Silicon Nanocrystal cell behavior during the channel hot electron programming operation.

With these measurements and using the formula (4.5) we can compare the results obtained for Si-nc cell and the results reported in figure 4.16 for the Flash floating gate. The aim is to evaluate the relation between the shape of the gate pulse, the energy and the programming window of the two devices (figure 4.28).

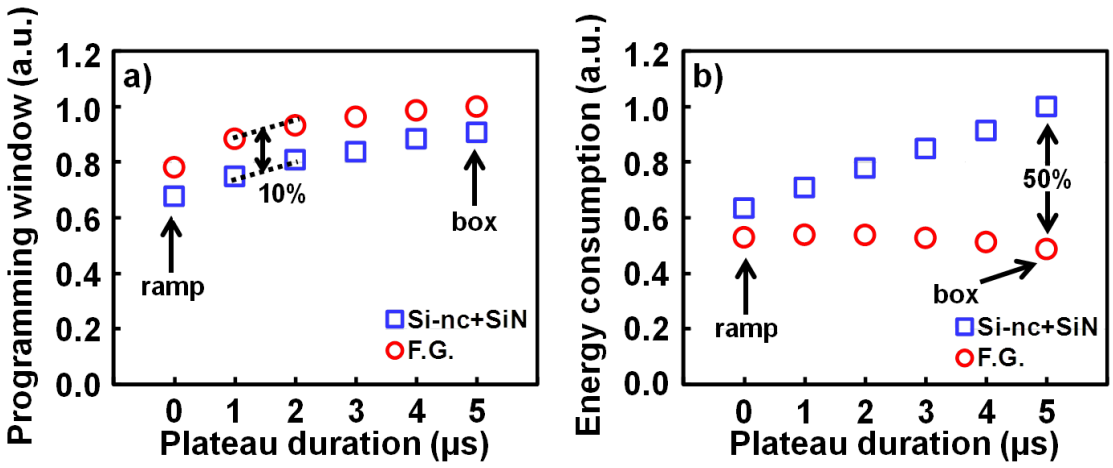


Figure 4. 28. (a) Programming window and (b) energy consumption comparison of Si-nc+SIN and floating gate cells, measured using the pulses shown in figure 4.26a. The F.G. data are also shown in figure 4.16.

Concerning the programming window, the hybrid Si-nc and the floating gate cells have the same behavior. One can notice that by increasing the plateau duration (or the ramp speed) the programming window increases too, and the difference between the two devices remains constant (10%). This difference is due to the higher coupling factor of Flash floating gate. Considering the Id consumption of Si-nc+SiN cell (figure 4.26b), it is easy to understand the linear dependence of energy consumption on plateau duration, while Ec slightly decreases for the Flash floating gate. For each pulse it is evident that the hybrid Si-nc cell reaches a smaller programming window consuming more energy; in particular to achieve a good programming window (greater than 80%), the Si-nc+SiN cell can use up to 50% more energy. It is necessary to use box pulses to increase the programming window, but also to decrease the programming time in order to reduce the energy consumption. To do this we repeated the experiment on another Si+SiN cell using box pulses to evaluate the dependence on programming window and energy on programming time. The results are shown in figure 4.29, the maximum value of Id is slightly different with respect to figure 4.26 and this is due to the wafer dispersion; in the two figures the same y scale is used. To program the cell we used Vg=9V and Vd=4.2V. The drain current follows the gate voltage during the programming. Thus the calculated energy is a linear function of pulse duration. After each pulse the programming window was also measured and the results are plotted in figure 4.30. The aim is to find the best programming pulse shape to improve the cell performances, as we did for the flash floating gate cell.

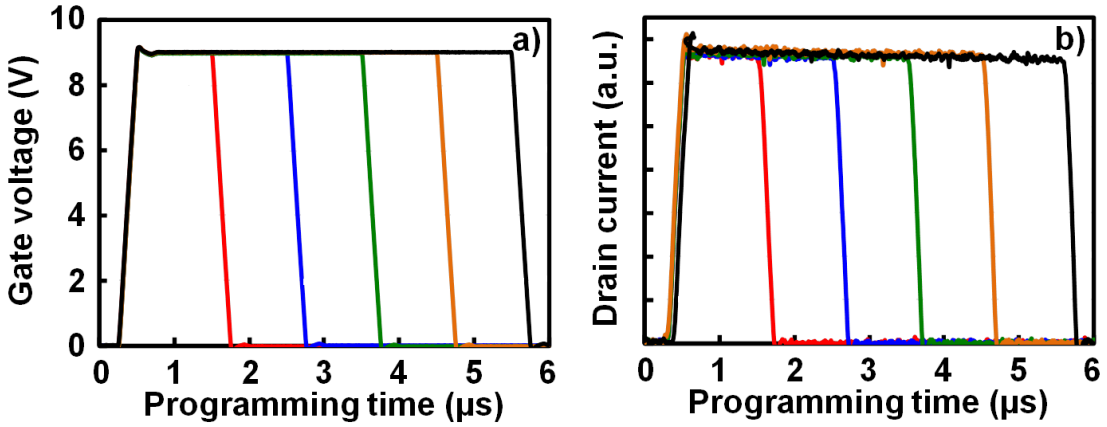


Figure 4. 29. (a) Gate box pulses with different durations, applied during the channel hot electron operation; (b) drain current of Si-nc+SiN cell, measured with dynamic method; the current follows the gate potential.

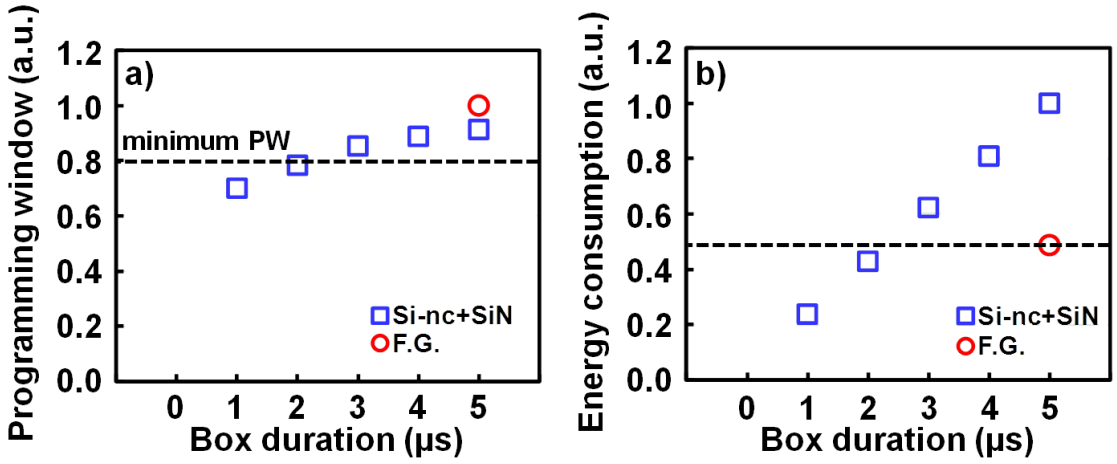


Figure 4. 30. (a) Programming window and (b) energy consumption comparison of Si-nc+SIN and floating gate cells, measured using the box pulses shown in figure 4.29a. The Y scale is the same as in figure 4.28.

In figure 4.30 we notice that the minimum programming window level is reached using 2µs box pulses. This enables the energy under the floating gate consumption to be decreased by using a 5µs box pulses. We can now establish that for the Si-nc+SiN cell, the best performances are obtained using box pulses of short duration. Unlike the floating gate cell, where an optimized pulse was defined by merging a ramp followed by a plateau in order to avoid current peak during the programming, for the hybrid Si-nc cell, the current level is constant. This leads to less design constraints and less disturb in logic and analog circuits around the memory array. To complete the study we report the experimental data of dynamic drain current absorption, of the programming window and the energy consumption in figures 4.31 and 4.32. We demonstrated that the drain current follows the gate potential, (see also the figures 4.26 and 4.29), confirming the linear dependence of energy consumption on programming time.

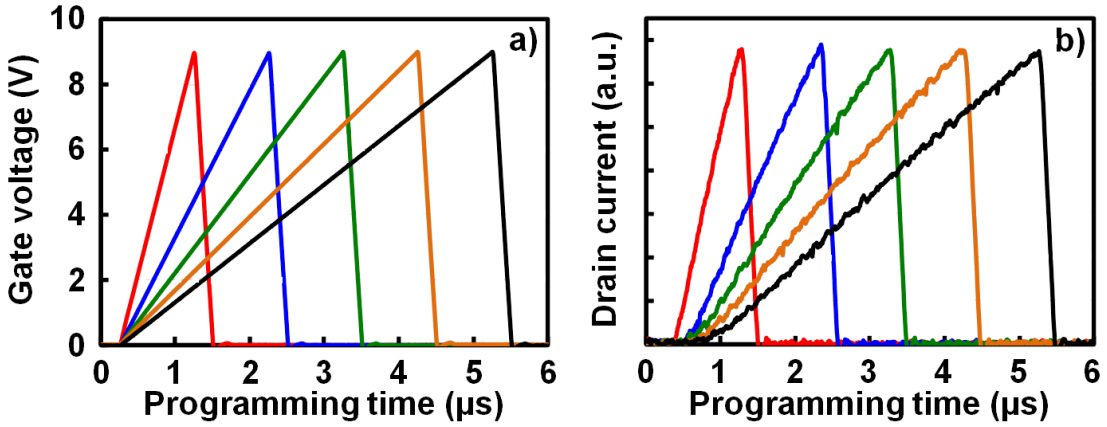


Figure 4. 31. (a) Gate ramp pulses with different durations, applied during the channel hot electron operation; (b) drain current of Si-nc+SiN cell, measured with dynamic method; the current follows the gate potential.

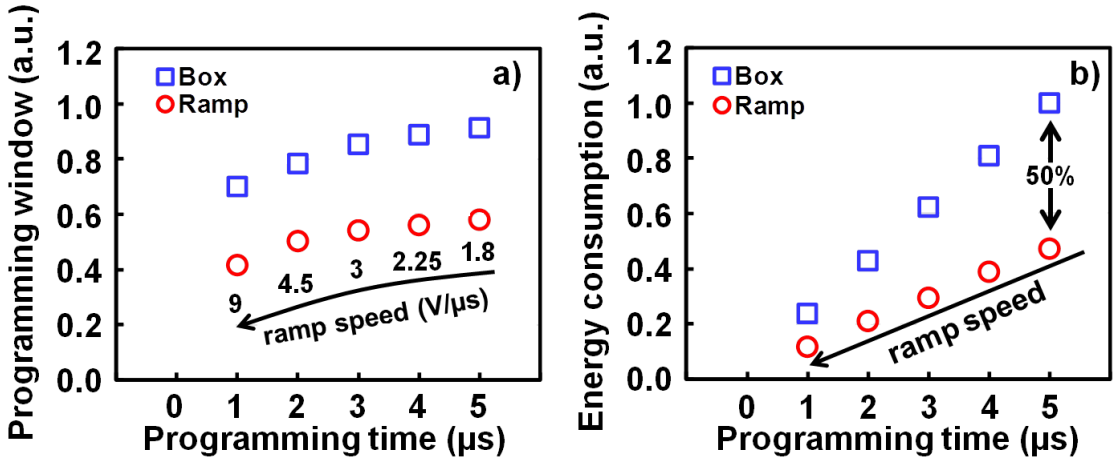


Figure 4. 32. (a) Programming window and (b) energy consumption of Si-nc+SIN, measured using box and ramped pulses respectively shown in figure 4.29a and 4.31a. The Y scale is the same as in figure 4.30.

In figure 4.32 we compared the experimental results obtained for the hybrid silicon nanocrystal cell when it is programmed by box or by ramp. When the programming time increases, the consumed energy as well as the programming window are increased. In this case the ramp speed is varied as a function of programming time (tp); so the programming window decreases when the ramp speed is increased because tp decreases too. The difference between ramp and box pulses is constant and independent of programming time: it is 40% for programming window and 50% for the energy consumption. The energy increasing is linear but the programming window tends to saturate when increasing the programming time. This means that in order to reach a satisfactory programming window level, a long programming time is necessary. Today, for electronic low power applications, speed is a fundamental parameter in order to be competitive on market. This is why we consider the box programming scheme to be the best solution for the silicon nanocrystal cell.

4.4.1.1 TCAD simulations of current consumption

In order to confirm the experimental results on silicon nanocrystal and floating gate cells, the behavior of these two devices is simulated using a commercial TCAD simulator. We chose a two-dimensional (2D) approach in order to evaluate the process impact, improve the program stability and reduce the simulation time. We produced process simulations using the parameter set provided by *Advanced Calibration* and we perform the electrical simulations with Sentaurus Device 2010.3. In figure 4.33 the simulated structures are shown. In the case of the floating gate cell, the 2D approach implies considering the device width dimension as a scaling parameter for the currents and the wings area as an additional coupling capacitance between control gate and floating gate. The programming operation increases the electrostatic potential on the whole floating gate area (figure 4.33a). The Si-nc+SiN cell simulations only need to scale the currents and the nanocrystal size and density. These two parameters are based on process results. After the programming operation the charged nanocrystals are very close to drain and SCR (figure 4.33b).

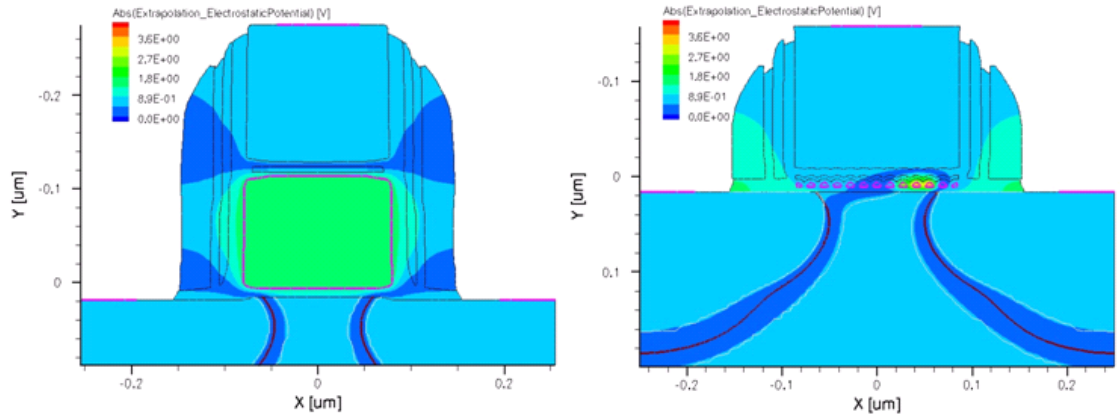


Figure 4. 33. Designed structures used in TCAD simulation after programming operation: a) floating gate cell, and b) Si-nc+SiN cell.

We used the same calibrated hydrodynamic set model for both devices, except for the Channel Hot Electron injection model. The Spherical Harmonic Expansion (SHE) model presented in [Seonghoon '09] was chosen for FG cell programming simulations, while the Fiegna model [Fiegna '93] is used for Si-nc+SiN cell. During the simulations we noticed that the SHE model can reproduce dynamic current simulations of each device, but not the programming window, in the case of Si-nc+SiN cell, even if the adjustment parameters are considered. Hence we decided to use in this case the Fiegna model that offers the best compromise to simulate the Si-nc+SiN channel hot electron programming operation. Figure 4.34 shows concordance between TCAD simulations and dynamic drain current measurements.

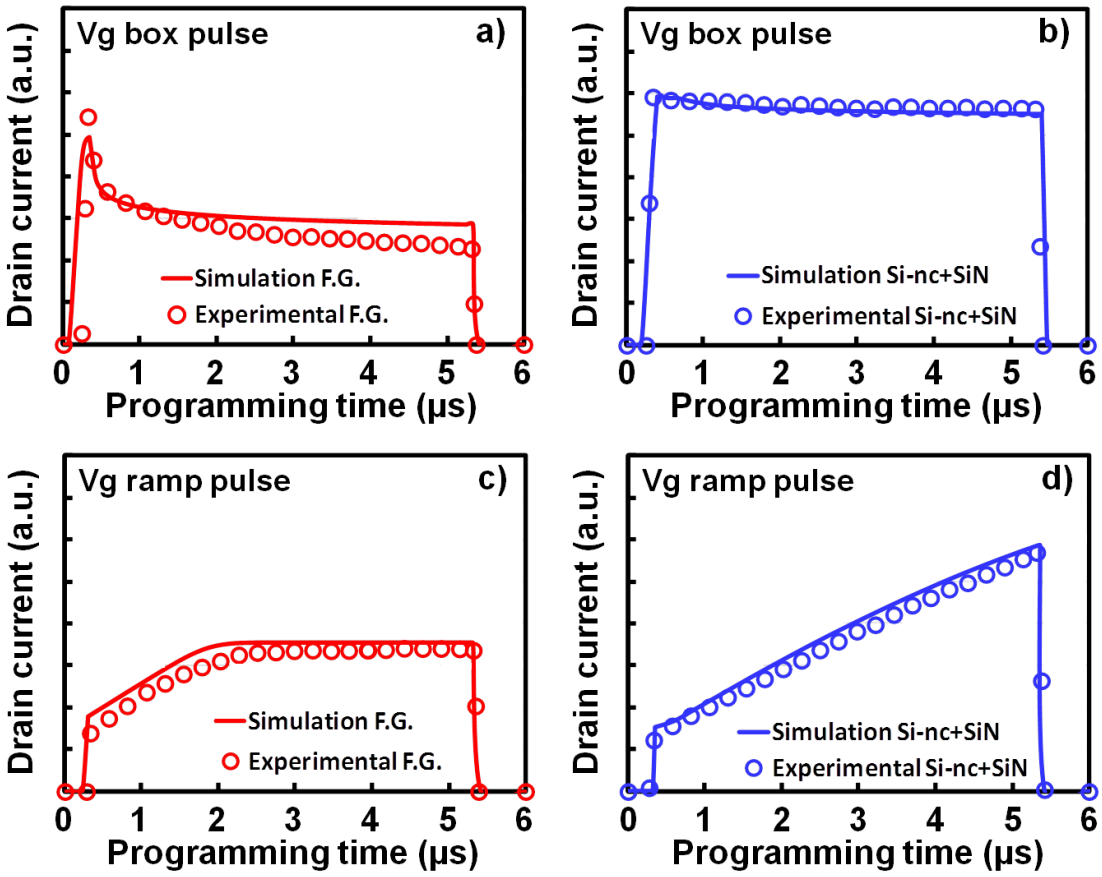


Figure 4. 34. Drain current in floating gate (F.G.) and hybrid silicon nanocrystal (Si-nc+SiN) cells using: a)and b) gate voltage box pulse or c) and d) gate voltage ramp pulse on gate; Vg=9V, Vd=4.2V, Vb=Vs=0V. Experimental data and simulations are compared; all the graphs have the same Y axe scale.

In figure 4.34 the results of dynamic current measurements, obtained for the floating gate and the Si-nc+SiN cells, are shown in the case of programming box and ramp pulses. These are also compared with the simulations of our TCAD model previously presented. For each case there is a satisfactory quantitative concordance. As described above we were able to simulate the programming window level after each pulse and to calculate the consumed energy. In figure 4.35 is reported the case of the Si-nc+SiN cell, by using the box pulses with different durations. The experimental data used to fit the simulations are the same as those plotted in figure 4.30. One can notice that the concordance between data and simulation predicts the cell behavior by varying the voltage bias and pulse shape. In this case we used the simulations to confirm our explanation concerning the cell functioning. Thus the charge localization maintains the absorbed drain current constant. This leads to an increase of the cell consumption and to suppress the current peak during the channel hot electron operation.

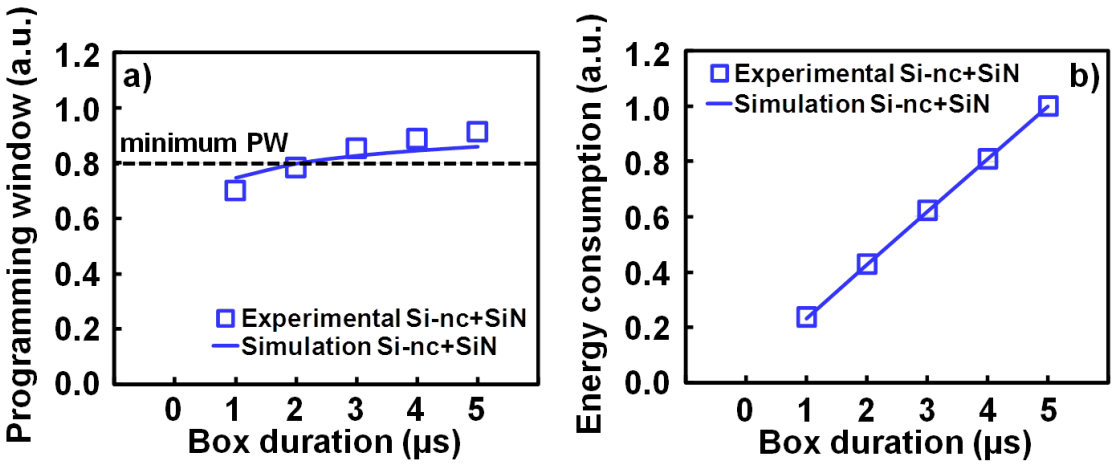


Figure 4. 35. Dependence of programming window and energy consumption on box pulse duration for the hybrid silicon nanocrystal cell. Experimental data and simulations are compared.

4.4.1.2 Hybrid silicon nanocrystal cell programming scheme optimization

Previously, we described the floating gate and hybrid silicon nanocrystal cell behavior using the experimental data obtained with the dynamic current measurements and TCAD simulations. The Si-nc+SiN cell does not present a drain current peak during the channel hot electron programming, but the consumption is higher than standard Flash floating gate. In order to reduce the consumed energy we decided to use short programming pulses by maintaining a satisfactory programming window level. In order to summarize all results, we plot in figure 4.36 the value of energy consumption calculated for the F.G. and Si-nc+SiN cells, while the programming window is kept constant (PW=4V). To reduce the design constraints, it is possible to optimize the programming operation of Si-nc+SiN cell using a box pulse. To reach 4V of programming window the following conditions are used:

- Floating gate: Vg_{ramp}=1.2V/µs; Vg_{max}=9V; Vd=4.2V; Vs=Vb=0V; tp=5µs
- Si-nc+SiN: Vg_{ramp}=2V/µs+plateau=2µs; Vg_{max}=9V; Vd=4.2V; Vs=Vb=0V; tp=5µs
- Si-nc+SiN optimization: Vg_{box}=9V; Vd=4.2V; Vs=Vb=0V; tp=2µs

This optimization enables a gain of 25% in terms of energy consumption and of 60% concerning the programming time with respect to the Flash floating gate performance if it is programmed by ramp to avoid the current peak.

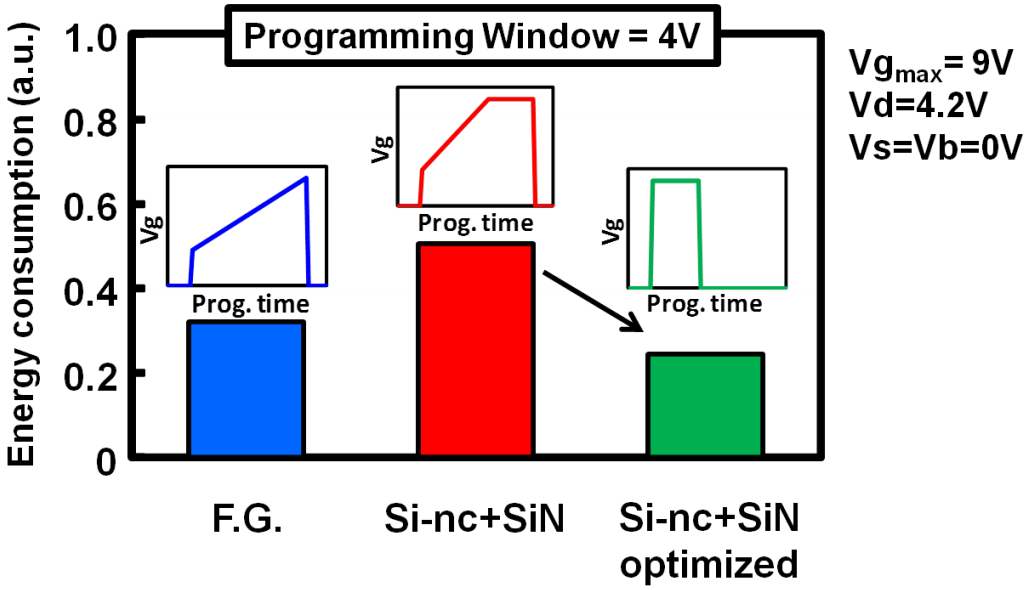


Figure 4. 36. Results of programming pulse optimization in terms of energy consumption.

4.4.2 Impact of gate and drain biases

In the preceding paragraph we studied the effect of programming pulse shape of Si-nc+SiN cell, keeping the biasing conditions constant. In order to optimize the cell consumption we further investigate the effect of gate and drain biases. In figure 4.37 the programming window and the energy are plotted keeping the programming time constant (tp=1µs); both are calculated as explained above. It is confirmed that the box pulse is more efficient in terms of programming window than the ramp, at the expense of higher energy consumption. Moreover the impact of drain and gate voltages is shown. The increase in drain voltage by 0.4V leads to the same gain in terms of memory window as when increasing Vg by 1.2V. The horizontal electric field becomes predominant in channel hot electron operation, a small variation of Vd thus implies a relevant increase of the programming window. Comparing the ramp and box pulses, when gate voltage is varied (figure 4.37 a and b), we notice that the energy consumption is twice as big as when the box pulses are used independently of the programming time. In addition the difference of programming windows, calculated for both cases, increases with gate voltage which means that the box pulse achieves the best results regardless of Vg. This increases the vertical electric field during the CHE operation, hence the charge injection probability is increased. In figure 4.37 c) and d) the impact of drain voltage on programming window and energy consumption is shown. We confirm the best results in terms of PW are reached using the box pulse and we notice the exponential increase of energy for the higher amounts of Vd.

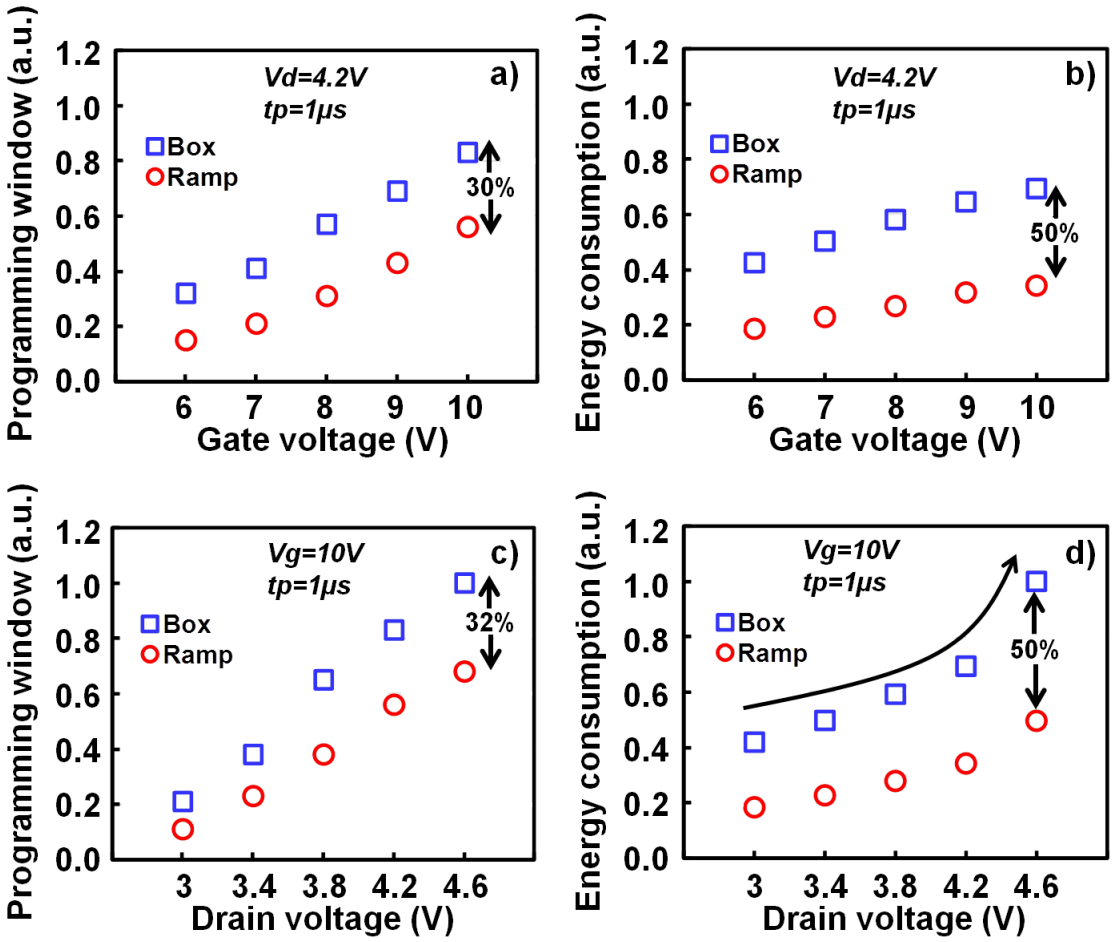


Figure 4. 37. Programming window and energy consumption as a function of gate (a-b) and drain (c-d) voltages. Ramp and box gate pulses are used.

To better understand the cell performances as a function of programming time and biases, the programming energy is plotted as a function of the programming window (figure 4.38). The goal is to increase the programming window by keeping constant the energy consumption using the box pulse and optimizing the biasing conditions. The abrupt variation of gate voltage (high ramp speed) during the programming operation starts the hot carrier generation at the beginning of the programming phase. The hot electron injection starts when Vg≈Vd [Takeda '83] [Takeda '85]. Thus, by using a ramp, the programming efficiency is lower and a longer time is needed to program the cell correctly. In figure 4.38 we notice that the programming window tends to saturate when the programming time is increased, leading to higher consumption. This is due to the quantity of injected charges that decreases the vertical electric field during the programming operation. We demonstrated once again that the box pulse increases the programming window by keeping the programming time and the biasing condition constant.

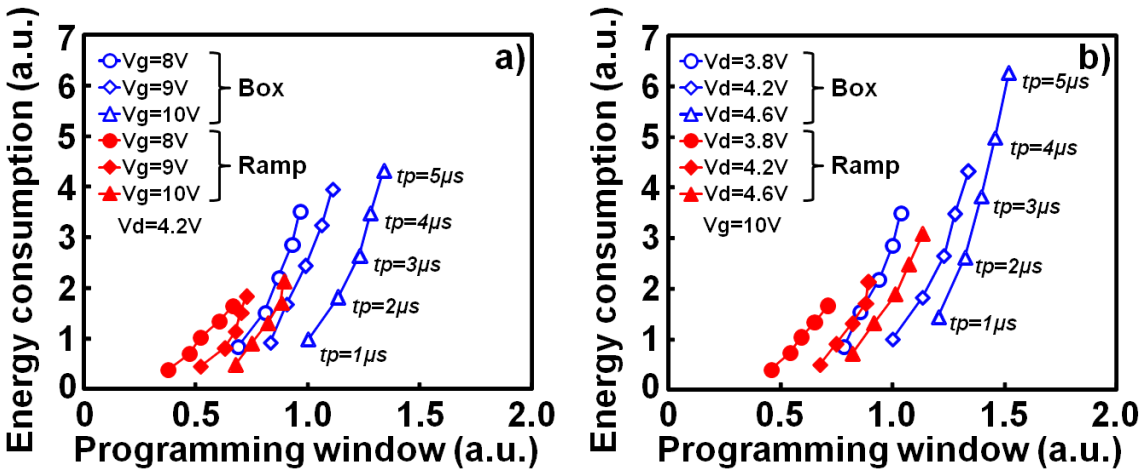


Figure 4. 38. Energy consumption as a function of programming window for different gate (a) and drain (b) voltages. The Y scale is comparable to that in figure 4.37.

The points of figure 4.38, using the programming time of 1μs, are the same as shown in figure 4.37. In this way it is possible to compare the two figures with the arbitrary units. Considering these results we defined the programming efficiency (PE) as the ratio between the programming window (PW) and the energy consumption (Ec).

$$PE = \frac{PW}{Ec} \qquad (4.6)$$

In figure 4.39 we report the programming efficiency in the case of box pulse, that has shown the best programming window results, for different biasing conditions.

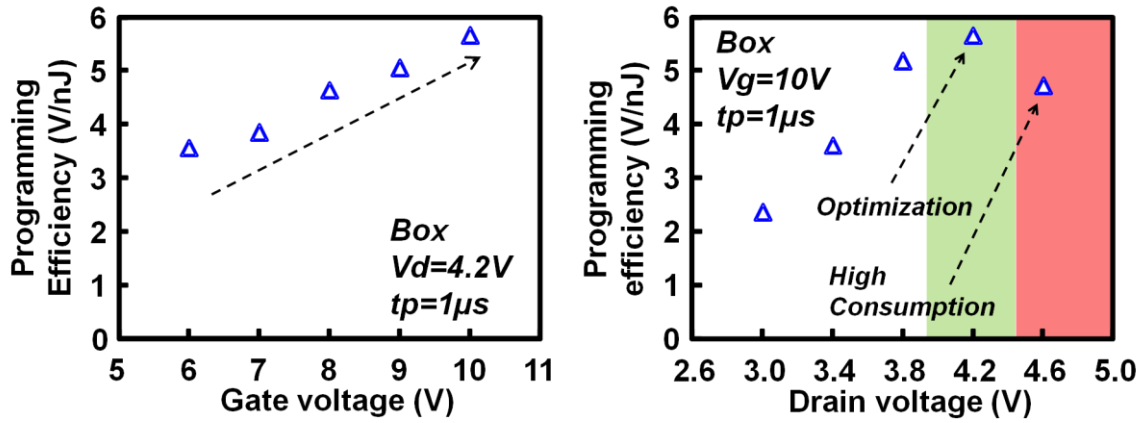


Figure 4. 39. Efficiency (PW/Ec) vs gate (left) and drain (right) voltage, using tp=1μs.

One can notice the linear dependence on the gate voltage (or vertical electric field). On the other hand, an optimized efficiency is measured for Vd=4.2V. This is due to the exponential behavior of energy consumption versus drain voltage. When the drain voltage is higher than 4.2V, the programming injection tends to saturate while the drain current increases which increases the consumption and reduces the programming efficiency [Della Marca '12]

[Masoero '12]. We can now affirm that to optimize the programming operation, it is necessary to use the box programming pulse with the higher gate voltage and Vd=4.2V which, in this case, represents the point of higher programming efficiency. Using the optimized programming scheme (Box pulse, Vg=10V and Vd=4.2V), we plotted programming efficiency in figure 4.40 as a function of the programming time. The greatest efficiency is reached using shorter programming time; as previously shown the shorter box pulse represents the best compromise for a satisfactory programming window and energy consumption. The graph shows that silicon nanocrystal memories are suitable for fast programming operations, representing a satisfactory trade-off between programming window and energy consumption.

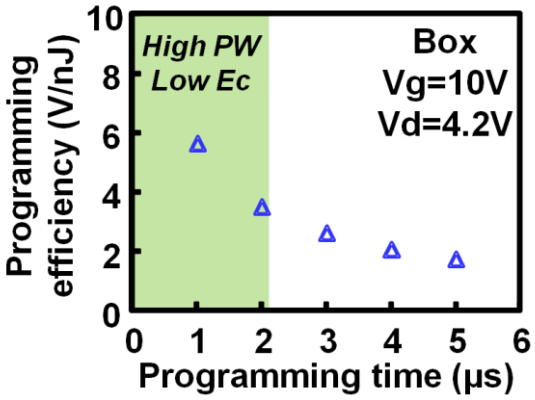


Figure 4. 40. Efficiency as function of programming time for Si-nc cell, using box gate pulses (Vg=10V, Vd=4.2V).

4.4.3 Impact of tunnel oxide thickness

After the study of programming scheme to improve the consumption of the silicon nanocrystal cell, in this section we show the impact of tunnel oxide thickness (tunox) on the programming operation (programming window and energy consumption) and data retention. First we performed the programming kinetic experiments using cumulative box pulses (duration of 0.5µs) for two tunnel oxide thicknesses (figure 4.41). In one case the amplitude of gate voltage remains constant (Vg=9V) during the kinetic, but in the other case Vg changes between 3V and 9V in order to emulate the programming ramp of 1.5V/µs; the drain voltage is kept constant at 4.2V. A very small impact of the tunnel oxide thickness on the threshold voltage shift is observed when a box pulse is applied, in accordance with Hot Carrier Injection theory. The Vt kinetics are differ significantly when box and ramp are compared. This is due to the different programming speed. The hot electron injection starts when Vg≈Vd, thus in the case of box it starts after the first 0.5µs pulse and after 1µs the 80% of global charge is stored. Using the ramp programming, the Vt characteristic is smoother

than the case of box pulse, because of the gradual hot carrier injection. In table 4.2 we reported the values of vertical electric field calculated during the first 0.5µs, considering zero charges stored in silicon nanocrystals for two different tunnel oxide thicknesses. We can notice that the difference of 0.2 MV/cm is small compared to the maximum E_{vert}=4.2MV/cm, thus the 1nm variation of tunox becomes negligible for the channel hot electron operation because the horizontal electric field is dominant.

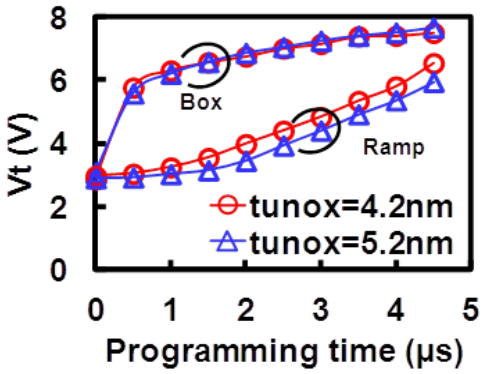


Figure 4. 41. Channel hot electron threshold voltage (Vt) kinetic characteristics using box and ramp pulses, for tunnel oxide thicknesses of 4.2nm and 5.2nm.

	Gate voltage=9V Programming time=0.5µs	
tunox (nm)	4.2	5.2
ΔVt (V)	2.9	2.5
E_{vert} (MV/cm)	4.2	4

Table 4. 2. Vertical electric fields (E_{vert}) and threshold voltage shifts (ΔVt) calculated for different tunnel oxide thicknesses (tunox) using a box pulse (tp=0.5µs; Vg=9V).

In figure 4.42 the energy consumption is plotted as a function of the programming window using the optimized programming conditions identified before (box pulse, Vg=10V and Vd=4.2V). The XY scale axes are the same as for figure 4.38. Moreover we highlighted the levels of minimum programming window, acceptable for good cell functionality, and the level of sub-nanojoule energy consumption. In figure 4.42, once again the tunnel oxide thickness has a limited influence on the consumed energy and the programming window during the CHE programming. In particular to produce the minimum programming window, a programming pulse of 1µs is sufficient for the two tunnel oxide thicknesses with a consumption energy lower than 1nJ. We show in figure 4.43 the programming efficiency calculated for two different tunnel oxides varying the box pulse duration, in order to conclude the study on tunnel oxide thickness and to demonstrate that this technological parameter has a limited impact on programming efficiency. The efficiency is higher using short pulses and is independent of the tunnel oxide thickness; which is consistent with the results previously presented.

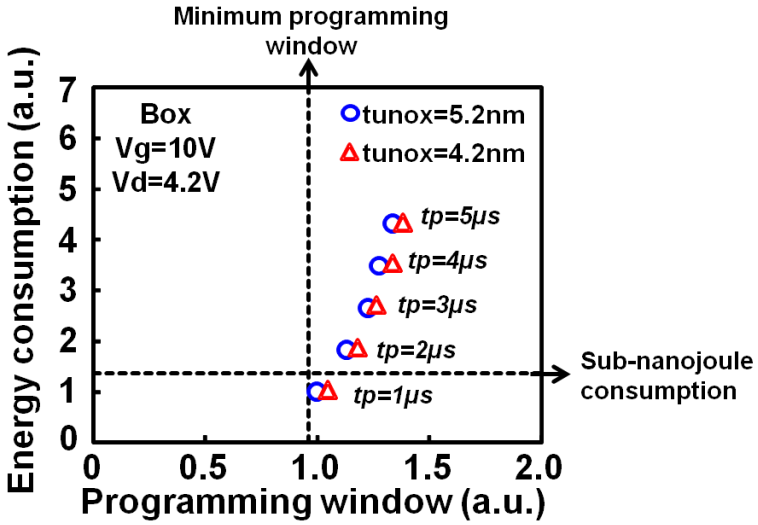


Figure 4. 42. Energy consumption as a function of the programming window of cells with 4.2nm and 5.2nm tunnel oxide thicknesses using box pulses (Vg=10V, Vd=4.2V).

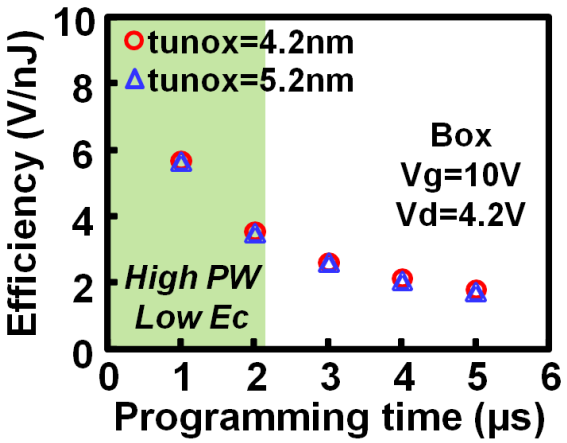


Figure 4. 43. Efficiency as function of programming time for Si-nc cell, using box gate pulses, calculated for tunox=4.2nm and tunox=5.2nm (Vg=10V, Vd=4.2V).

In this section we studied the impact of programming scheme on energy consumption for silicon nanocrystal memory cell. We propose an optimization concerning the programming pulse shape demonstrating the greatest efficiency of box pulse versus the ramp. The linear dependence on gate voltage is shown, while an optimum point of work is found for Vd=4.2V. The consumption has been reduced down to 1nJ generating a satisfactory programming window. Moreover the best trade-off to improve the cell efficiency is found by using very fast pulses regardless of tunnel oxide thickness.

4.5 Optimized cell consumption

In previous sections we showed the characterization results of floating gate and hybrid silicon nanocrystal cell consumption; we related it with the programming window in order to find an optimized programming scheme. The best compromise for floating gate cell was to use a programming ramp followed by a plateau in order to reduce the drain current peak and to maintain a satisfactory programming window. In the case of the Si-nc+SiN cell we found the best compromise using very short box pulse on gate terminal. Here we report the results found using the optimized silicon nanocrystal cell with 4.2nm tunnel oxide, 12nm silicon nanocrystal size and the thinner ONO (EOT=10.5nm). Using box pulses of different durations, we characterized the drain current absorption; in figure 4.44 we reported the results using Vg=10V and Vd=4.2V, the Y scale axe is the same as in figure 4.29 so as to compare the optimized cell with Si-nc+SiN cell. We notice that the drain current does not follow the gate potential, but in this case it decreases during the programming time, showing a similar behavior to that of the floating gate cell, where a current peak is present. We explained previously that the different behavior between Si-nc+SiN and floating gate cell is justified by the localization or not of trapped charges. In the case of Si-nc cell we can affirm that the localization effect due to the SiN layer is not present. Moreover the bigger size of nanocrystals produces a charge distribution toward the center of device channel, modifying the potential of substrate surface. In figure 4.45 we reported the cell schematics and the relative drain current measurements in order to compare the behavior of Si-nc+SiN, optimized Si-nc and floating gate cells.

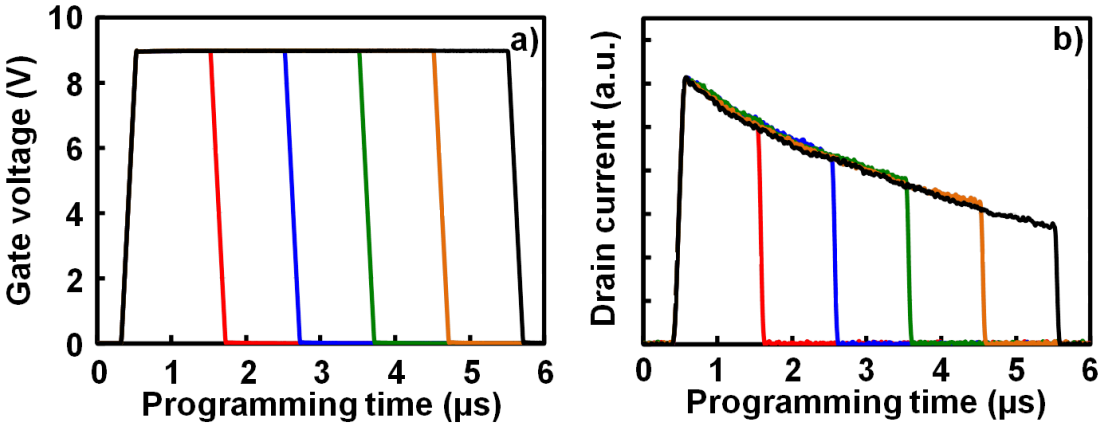


Figure 4. 44. (a) Gate box pulses with different durations, applied during the channel hot electron operation; (b) drain current of optimized Si-n cell, measured with dynamic method. The Y scale is the same as in figure 4.29.

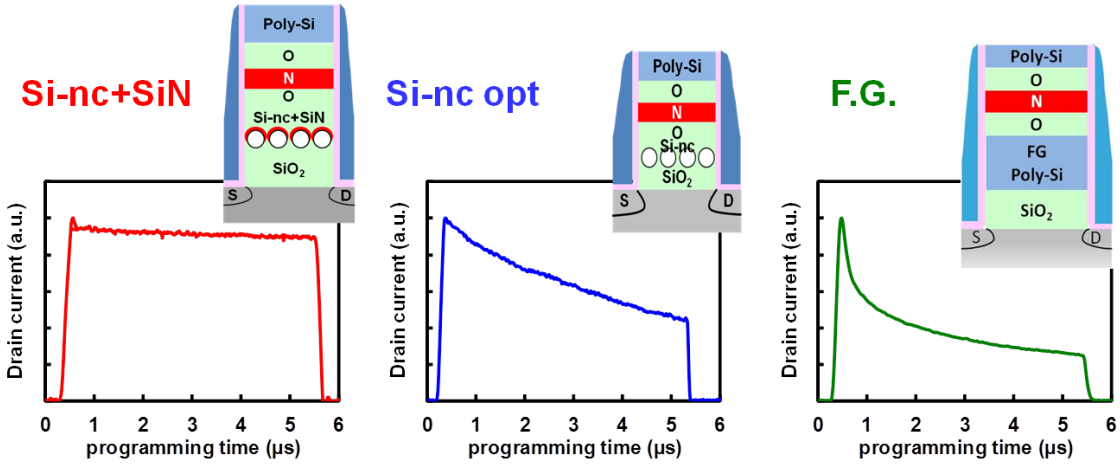


Figure 4. 45. Dynamic drain current measured for: hybrid silicon nanocrystal cell (Si-nc+SiN), optimized Si-nc cell and floating gate (F.G.) cell, using a box programming pulse on control gate.

Figure 4.45 shows that the Si-nc cell, where large nanocrystals (Φ=12nm) are embedded with a high density, has a response half-way between Si-nc+SiN and floating gate cells. This means it is possible to control the dynamic current modifying the nanocrystal size and density, thus the covering area. Previously we explained the *transistor-like* behavior of hybrid Si-nc+SiN cell where the drain current follows the gate potential thanks to the charge localization close to the drain. Instead, in the case of the floating gate device, a peak of current is detected when a box pulse is applied on gate terminal. During the hot carrier injection the charges flow throw the silicon nanocrystals toward the source side modifying the substrate surface potential and thus the vertical and horizontal electric fields. The charge diffusion can be due to the single electron interactions between neighbor silicon nanocrystals with different electric potentials [Matsuoka '92]. Moreover the nanocrystal coalescence during the growth process can reduce the distances up to the contact creating percolative paths from the drain side to the source side. Taking into account the dynamic behavior of the Si-nc cell and the fact that the box is more efficient with respect to the ramp pulse, in particular for short pulse duration, we repeated the measurements by varying drain and gate biases and we compared these results with the hybrid Si-nc cell. In figure 4.46 we reported the programming window and the energy consumption as a function of Vg (a-b) and Vd (c-d) using 1µs box programming pulses, of Si-nc and Si-nc+SiN cells. We notice that the optimized silicon nanocrystal cell presents a better programming window due to the higher covering area (coupling factor) and lower energy consumption due to the lower drain current. The dependence on gate voltage (vertical electric field) is linear as in the case of the Si-nc+SiN cell. With high drain voltage the programming window starts to saturate for Vd=3.8V, while the consumed energy exponentially increases. These results directly impact the cell efficiency.

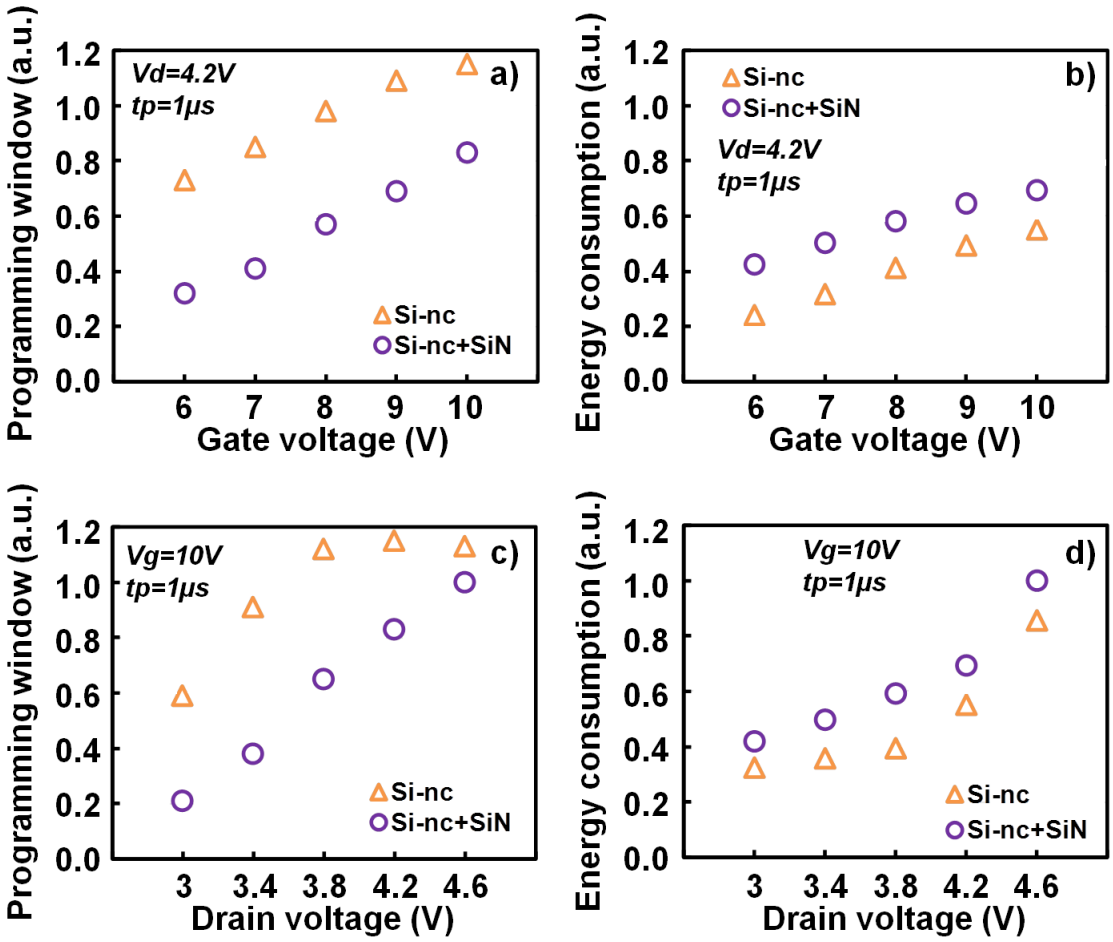


Figure 4. 46. Programming window and energy consumption as a function of gate (a-b) and drain (c-d) voltages of optimized Si-nc and Si-nc+SiN cells The Y scale axe is the same as in figure 4.37.

In figure 4.47a we can notice that for the optimized Si-nc cell the programming efficiency decreases in spite of the Si-nc+SiN cell trend.

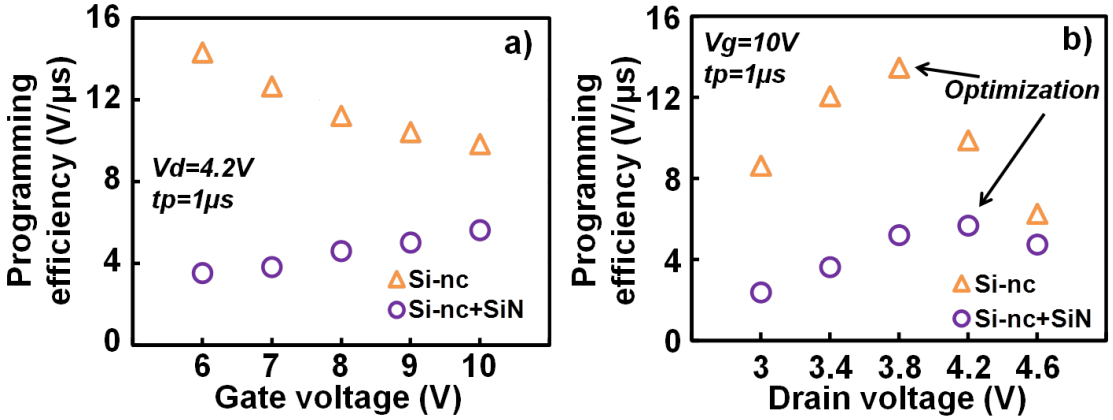


Figure 4. 47. Programming efficiency (PW/Ec) vs gate (a) and drain (b) voltage, using tp=1µs, of optimized Si-nc and Si-nc+SiN cells.

This is because a low Vg is sufficient to produce high programming window close to saturation while the drain current dependence on Vg increases the consumed energy. On the other hand by varying the drain voltage (figure 4.47b) we note that the programming

efficiency is greatly increased with respect to the Si-nc+SiN cell; the optimization of Si-nc cell is reached at Vd=3.8V when the programming window starts to saturate. To decreasing the drain voltage optimization from 4.2V to 3.8V is important not only for the gain concerning the consumption, but also for reduction in the size of charge pump circuits that can deliver less current to the memory array. The area reduction of these circuits directly impacts the final chip cost hence the importance of this improvement. Finally we plotted in figure 4.48 the consumed energy as a function of the programming window, comparing the two nanocrystal devices. The optimized cell has 1.5V of gain on programming window keeping the energy level constant. Depending on the final target, it is possible to decrease the programming time in order to reduce the energy consumption for very fast and ultra low energy applications.

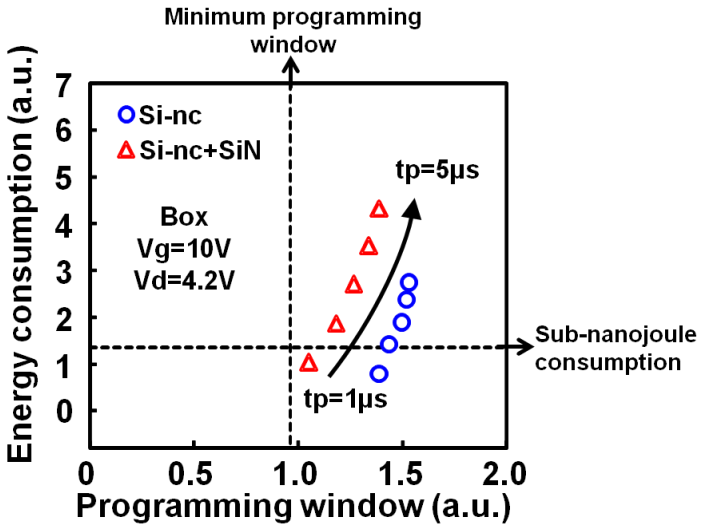


Figure 4. 48. Energy consumption as a function of the programming window of Si-nc+SiN and optimized Si-nc cells with 4.2nm tunnel oxide thickness using box pulses (Vg=10V, Vd=4.2V).The XY scale axes are the same as in figure 4.42.

4.6 Benchmarking with Flash floating gate

To conclude the report on cell energy consumption, in this paragraph we compare the main results based on the previous study of the silicon nanocrystal cell with Flash floating gate. We have shown before the cell behavior under different biasing conditions and the impact of some technological parameters (tunnel oxide thickness, channel doping dose, ONO thickness). In order to compare these two different devices, we plotted in figure 4.49 the programming window and the consumed energy using box pulses of different duration. The devices were tested using the proper drain voltage optimization shown previously: Vd=4.2V for the Si-nc+SiN; Vd=3.8V for F.G. and Si-nc cells; the gate voltage is kept unchanged (Vg=9V). The optimized Si-nc cell is able to reach a programmed level comparable with that

of the floating gate cell, but the consumed energy remains higher because of the higher drain current; the worst case is represented by the Si-nc+SiN cell. Using the experimental data we extrapolated the power laws describing cell behavior and we show that by using very fast pulses the Si-nc cell consumption can be improved but it does not go down to the floating gate level. The results of ramped gate voltage (figure 4.50) show that the Si-nc optimized cell is completely comparable with the floating gate in terms of programming window and energy consumption. Even if the maximum drain current is grater, in the case of Si-nc, the different dynamic behavior enables to reach very similar performance.

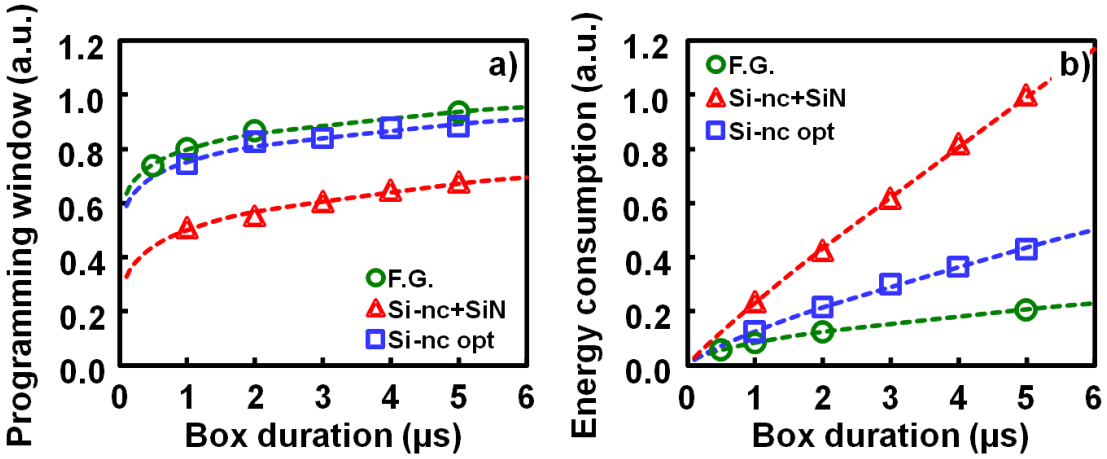


Figure 4. 49. (a) Programming window and (b) energy consumption comparison of optimized silicon nanocrystal (Si-nc opt), hybrid silicon nanocrystal (Si-nc+SiN) and floating gate (F.G.) cells, measured using box pulses.

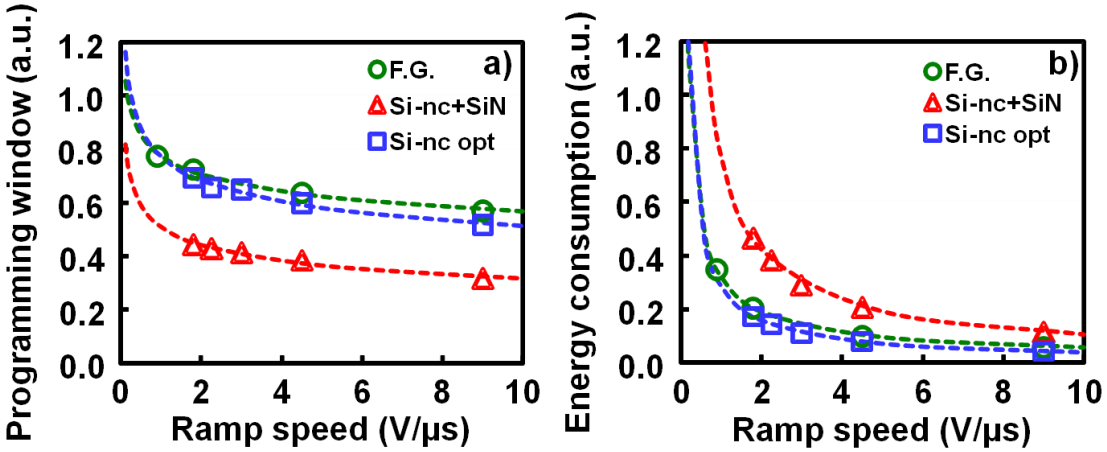


Figure 4. 50. (a) Programming window and (b) energy consumption comparison of optimized silicon nanocrystal (Si-nc opt), hybrid silicon nanocrystal (Si-nc+SiN) and floating gate (F.G.) cells, measured using ramp pulses.

The dynamic measurements verify the presence of current peak in the case of the floating gate and we explained above that it can cause disturbs in analog and digital circuits around the memory array. In the case of the optimized Si-nc cell, the current decrease during programming operation is not abrupt and it can be tolerated depending on design constraints.

To conclude the chapter we show in figure 4.51 the performance of Si-nc, hybrid Si-nc and floating gate cells found using the optimized drain voltages and keeping the gate voltage unchanged (Vg=9V). The programming time is fixed at 2µs in order to propose benchmarking in the case of low energy and fast application. We notice that when the box pulse is applied the floating gate consumption is 50% less than for the Si-nc optimized cell, on the contrary the programming window can be considered as equivalent. Spite when using programming ramp, the silicon nanocrystal cell shows the best consumption level. The optimized silicon nanocrystal cell can be considered as a good alternative to the flash floating gate in terms of programming speed and energy consumption maintaining a satisfactory level of programming window, but further efforts are necessary to overcome the Flash floating gate memory cell.

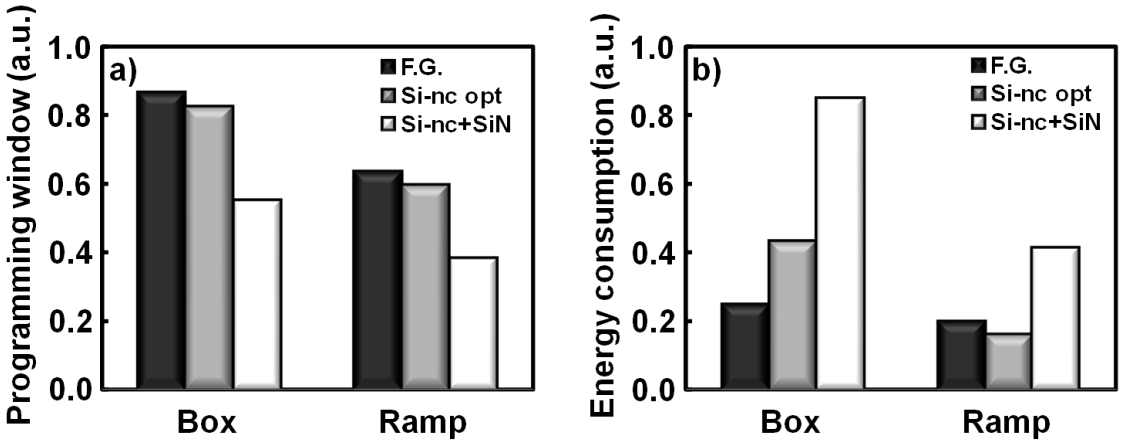


Figure 4. 51. (a) Programming window and (b) energy consumption comparison of optimized silicon nanocrystal (Si-nc opt), hybrid silicon nanocrystal (Si-nc+SiN) and floating gate (F.G.) cells using box and ramp pulses when the programming time is fixed to 2µs.

Bibliography of chapter 4

[Benfdila '04] A. Benfdila and M. T. Bensidhoum, "On the current saturation in low dimensional MOSFETs", in Microelectronics, 2004. 24th International Conference on, 16-19 May 2004, pp. 299-302 vol.1, http://dx.doi.org/10.1109/icmel.2004.1314622

[Chenming '85] H. Chenming, C. T. Simon, H. Fu-Chieh, K. Ping-Keung, C. Tung-Yi, and K. W. Terrill, "Hot-electron-induced MOSFET degradation - Model, monitor, and improvement", Electron Devices, IEEE Transactions on, vol. 32, 1985, pp. 375-385, http://dx.doi.org/,

[Choi '94] W. L. Choi and D. M. Kim, "A new technique for measuring coupling coefficients and 3-D capacitance characterization of floating-gate devices", Electron Devices, IEEE Transactions on, vol. 41, 1994, pp. 2337-2342, http://dx.doi.org/10.1109/16.337446

[Della Marca '11a] V. Della Marca, J. Amouroux, J. Delalleau, L. Lopez, J. Ogier, J. Postel-Pellerin, F. Lalande, and G. Molas, "Energy consumption optimization in nonvolatile silicon nanocrystal memories", in Semiconductor Conference (CAS), 2011 International, 17-19 Oct. 2011, pp. 339-342, http://dx.doi.org/10.1109/smicnd.2011.6095810

[Della Marca '11b] V. Della Marca, J.-L. Ogier, J. Postel-Pellerin, F. Lalande, and G. Molas, "New Direct Measurement Technique of Programming Current for Flash Memory Cell Energy Consumption Optimization", presented at the IEEE SISC, 2011b,

[Della Marca '11c] V. Della Marca, A. Regnier, J. Ogier, R. Simola, S. Niel, J. Postel-Pellerin, F. Lalande, and G. Molas, "Experimental study to push the Flash floating gate memories toward low energy applications", in Semiconductor Device Research Symposium (ISDRS), 2011 International, 7-9 Dec. 2011, pp. 1-2, http://dx.doi.org/10.1109/isdrs.2011.6135271

[Della Marca '12] V. Della Marca, L. Masoero, G. Molas, J. Amouroux, E. Petit-Faivre, J. Postel-Pellerin, F. Lalande, E. Jalaguier, S. Deleonibus, B. De Salvo, P. Boivin, and J. Ogier, "Optimization of programming consumption of silicon nanocrystal memories for low power applications", in Semiconductor Conference Dresden-Grenoble (ISCDG), 2012 International, 24-26 Sept. 2012, pp. 65-68, http://dx.doi.org/10.1109/iscdg.2012.6359988

[Della Marca '13] V. Della Marca, G. Just, A. Regnier, J. L. Ogier, R. Simola, S. Niel, J. Postel-Pellerin, F. Lalande, L. Masoero, and G. Molas, "Push the flash floating gate memories toward the future low energy application", Solid-State Electronics, vol. 79, 2013, pp. 210-217, http://dx.doi.org/, http://www.sciencedirect.com/science/article/pii/S0038110112003048

[Driussi '04] F. Driussi, D. Esseni, and L. Selmi, "Performance, degradation monitors, and reliability of the CHISEL injection regime", Device and Materials Reliability, IEEE Transactions on, vol. 4, 2004, pp. 327-334, http://dx.doi.org/10.1109/TDMR.2004.837208

[Esseni '99] D. Esseni, A. Della Strada, P. Cappelletti, and B. Ricco, "A new and flexible scheme for hot-electron programming of nonvolatile memory cells", Electron Devices, IEEE Transactions on, vol. 46, 1999, pp. 125-133, http://dx.doi.org/10.1109/16.737450

[Esseni '00a] D. Esseni, L. Selmi, A. Ghetti, and E. Sangiorgi, "Injection efficiency of CHISEL gate currents in short MOS devices: physical mechanisms, device

implications, and sensitivity to technological parameters", Electron Devices, IEEE Transactions on, vol. 47, 2000a, pp. 2194-2200, http://dx.doi.org/10.1109/16.877183

[Esseni '00b] D. Esseni, C. Villa, S. Tassan, and B. Ricco, "Trading-off programming speed and current absorption in flash memories with the ramped-gate programming technique", Electron Devices, IEEE Transactions on, vol. 47, 2000b, pp. 828-834, http://dx.doi.org/10.1109/16.831000

[Fiegna '93] C. Fiegna, E. Sangiorgi, and L. Selmi, "Oxide-field dependence of electron injection from silicon into silicon dioxide", Electron Devices, IEEE Transactions on, vol. 40, 1993, pp. 2018-2022, http://dx.doi.org/10.1109/16.239743

[Masoero '12] L. Masoero, G. Molas, V. Della Marca, M. Gely, O. Cueto, J. P. Colonna, A. De Luca, P. Brianceau, C. Charpin, D. Lafond, V. Delaye, F. Aussenac, C. Carabasse, S. Pauliac, C. Comboroure, P. Boivin, G. Ghibaudo, S. Deleonibus, and B. De Salvo, "Physical Understanding of Program Injection and Consumption in Ultra-Scaled SiN Split-Gate Memories", in Memory Workshop (IMW), 2012 4th IEEE International, 20-23 May 2012, pp. 1-4, http://dx.doi.org/10.1109/imw.2012.6213686

[Matsuoka '92] H. Matsuoka, T. Ichiguchi, T. Yoshimura, and E. Takeda, "Single electron charging effect in coupled quantum dots of a Si-MOSFET", in Electron Devices Meeting, 1992. IEDM '92. Technical Digest., International, 13-16 Dec 1992, pp. 781-784, http://dx.doi.org/10.1109/iedm.1992.307474

[Maure '09] A. Maure, Ph.D. Thesis, "Modelisation Compacte de la Cellule Memoire Flash et Optimisation des Phases d'Ecriture", Université de Provence (Aix-Marseille I),

[Mii '92] T. Mii, R. Kleinhenz, W. Noble, J. Johnson, A. Bryant, and M. Jaffe, "First experimental verification of collection length limited gate induced drain leakage", in Electron Devices Meeting, 1992. IEDM '92. Technical Digest., International, 13-16 Dec 1992, pp. 155-158, http://dx.doi.org/10.1109/iedm.1992.307331

[Moon '91] B. J. Moon, C. K. Park, K. Lee, and M. Shur, "New short-channel n-MOSFET current-voltage model in strong inversion and unified parameter extraction method", Electron Devices, IEEE Transactions on, vol. 38, 1991, pp. 592-602, http://dx.doi.org/10.1109/16.75171

[Munteanu '02] D. Munteanu and A. M. Ionescu, "Modeling of drain current overshoot and recombination lifetime extraction in floating-body submicron SOI MOSFETs", Electron Devices, IEEE Transactions on, vol. 49, 2002, pp. 1198-1205, http://dx.doi.org/10.1109/TED.2002.1013276

[Nair '04] D. R. Nair, S. Mahapatra, S. Shukuri, and J. D. Bude, "Drain disturb during CHISEL programming of NOR flash EEPROMs-physical mechanisms and impact of technological parameters", Electron Devices, IEEE Transactions on, vol. 51, 2004, pp. 701-707, http://dx.doi.org/10.1109/TED.2004.825821

[Ning '78] T. H. Ning, "Hot-electron emission from silicon into silicon dioxide", Solid-State Electronics, vol. 21, 1978, pp. 273-282, http://dx.doi.org/, http://www.sciencedirect.com/science/article/pii/003811017890148X

[Orlowski '89] M. K. Orlowski and C. Werner, "Model for the electric fields in LDD MOSFETs. II. Field distribution on the drain side", Electron Devices, IEEE Transactions on, vol. 36, 1989, pp. 382-391, http://dx.doi.org/10.1109/16.19940

[Parke '92] S. A. Parke, J. E. Moon, H. C. Wann, P. K. Ko, and C. Hu, "Design for suppression of gate-induced drain leakage in LDD MOSFETs using a quasi-two-dimensional analytical model", Electron Devices, IEEE Transactions on, vol. 39, 1992, pp. 1694-1703, http://dx.doi.org/10.1109/16.141236

[Rideau '04] D. Rideau, A. Dray, F. Gilibert, F. Agut, L. Giguerre, G. Gouget, M. Minondo, and A. Juge, "Characterization & modeling of low electric field gate-induced-drain-

leakage [MOSFET]", in Microelectronic Test Structures, 2004. Proceedings. ICMTS '04. The International Conference on, 22-25 March 2004, pp. 149-154, http://dx.doi.org/10.1109/icmts.2004.1309469

[Rideau '10] D. Rideau, V. Quenette, D. Garetto, E. Dornel, M. Weybright, J. P. Manceau, O. Saxod, C. Tavernier, and H. Jaouen, "Characterization & modeling of gate-induced-drain-leakage with complete overlap and fringing model", in Microelectronic Test Structures (ICMTS), 2010 IEEE International Conference on, 22-25 March 2010, pp. 210-213, http://dx.doi.org/10.1109/icmts.2010.5466816

[Seonghoon '09] J. Seonghoon, A. Wettstein, C. Woosung, F. M. Bufler, and E. Lyumkis, "Gate Current Calculations Using Spherical Harmonic Expansion of Boltzmann Equation", in Simulation of Semiconductor Processes and Devices, 2009. SISPAD '09. International Conference on, 9-11 Sept. 2009, pp. 1-4, http://dx.doi.org/10.1109/sispad.2009.5290216

[Takeda '83] E. Takeda, Y. Nakagome, H. Kume, and S. Asai, "New hot-carrier injection and device degradation in submicron MOSFETs", Solid-State and Electron Devices, IEE Proceedings I, vol. 130, 1983, pp. 144-150, http://dx.doi.org/10.1049/ip-i-1.1983.0026

[Takeda '85] E. Takeda, Y. Ohji, and H. Kume, "High field effects in MOSFETS", in Electron Devices Meeting, 1985 International, 1985, pp. 60-63, http://dx.doi.org/10.1109/iedm.1985.190891

[Tam '84] S. Tam and H. Chenming, "Hot-electron-induced photon and photocarrier generation in Silicon MOSFET's", Electron Devices, IEEE Transactions on, vol. 31, 1984, pp. 1264-1273, http://dx.doi.org/10.1109/T-ED.1984.21698

[Touhami '01] A. Touhami and A. Bouhdada, "Modeling of gate-induced drain leakage current in n-type metal--oxide--semiconductor field effect transistor", Journal of Applied Physics, vol. 89, 2001, pp. 1880-1884, http://dx.doi.org/10.1063/1.1327607

[Wang '79] S. T. Wang, "On the I-V characteristics of floating-gate MOS transistors", Electron Devices, IEEE Transactions on, vol. 26, 1979, pp. 1292-1294, http://dx.doi.org/10.1109/T-ED.1979.19595

[Wong '92] M. Wong, D. K. Y. Liu, and S. S. W. Huang, "Analysis of the subthreshold slope and the linear transconductance techniques for the extraction of the capacitance coupling coefficients of floating-gate devices", Electron Device Letters, IEEE, vol. 13, 1992, pp. 566-568, http://dx.doi.org/10.1109/55.192842

[Yimao '11] C. Yimao, Z. Xing, and H. Ru, "Counter-lightly-doped-drain (C-LDD) structure for Multi-level cell (MLC) NOR flash memory free of drain disturb", in Solid-State Device Research Conference (ESSDERC), 2011 Proceedings of the European, 12-16 Sept. 2011, pp. 207-210, http://dx.doi.org/10.1109/essderc.2011.6044198

Chapter 5 - Conclusion

In this thesis work we discussed silicon nanocrystal memory cell characterization and modeling. After a detailed study of recent silicon nanocrystal implication in memory devices we optimized the memory stack. We characterized the programming window by varying different technological parameters such as nanocrystal size and density, presence of Si_3N_4 capping layer, channel doping dose and tunnel oxide thickness. The aim of the experiments was to understand the cell behavior in order to improve the coupling factor and to minimize the parasitic charge trapping. Furthermore the results concerning the cell reliability achieved with our proposed optimization have been shown. Satisfactory data retention at 150°C has been shown, and for the first time a 1Mcycles endurance, within a range of temperature between -40°C and 150°C with a 4V programming window, has been presented. Finally we developed a new technique to measure the cell current consumption during the channel hot electron programming operation. This technique has been applied for the first time to the Flash floating gate and the silicon nanocrystal memory cell. We described the dynamic behavior and how to improve the cell energy consumption to reach a sub-nanojoule consumption by maintaining a 4V programming window. The silicon nanocrystal cell and the floating gate performances are compared at the section end in table 5.1.

In the first chapter the economic context, the evolution and the classification of semiconductor memories was presented. Then the Flash memory operations needed to understand this thesis were reviewed. We thus presented the Flash memory scaling limits and the proposed solutions. We explained the advantages of using a discrete charge trapping layer instead of the continuous floating gate and the importance of control dielectric instead of the classical silicon oxide. Finally, we introduced the silicon nanocrystal memory solution. In particular we reported the state of the art of charge trap silicon nanocrystal cell, which is the object of this thesis. This option reduces the mask number in process fabrication and scales the memory stack thickness, hence decreasing the operating voltages. Moreover the silicon nanocrystal cell can produce satisfactory data retention results using a thin tunnel oxide.

The second chapter describes the experimental setup used to characterize the programming window of silicon nanocrystal memory with automatic and manual tests. The impact of different technological parameters on program/erase speed is studied. The results have shown that to improve the programming window it is important to increase the silicon nanocrystal size and density (covered area), further improvement can be made by capping the silicon

nanocrystal with the SiN layer. These results are coherent with the literature. Moreover, the channel doping dose can be increased to improve the programming efficiency, but in this case, an adjusting of program/erase threshold voltage is needed. The last technological parameter studied in this chapter was the tunnel oxide thickness. We demonstrated that its variation strongly impacts the Fowler-Nordheim operation, while the channel hot electron programming is slightly dependent. After these considerations we optimized the silicon nanocrystal memory stack comparing this device with the Flash floating gate. Concerning the channel hot electron programming operation, the Si-nc cell reached the same performance as floating gate, which means a 4V programming window in 3.5µs with a ramped gate voltage (1.5V/µs). For the Fowler-Nordheim erase operation an improvement is seen with respect the floating gate device due to a thinner tunnel oxide. The optimized Si-nc cell is erased with a ramped gate (5kV/s) in 200µs in despite of the floating gate that reaches a 4V programming window in 500µs.

In chapter three we have shown the results of reliability experiments performed on silicon nanocrystal memory. The variation of technological parameters has been also evaluated. In particular the presence of silicon nitride capping layer on nanocrystals increases the charge trapping probability and the cell covered area. In our case the nanocrystals were not entirely surrounded by Si_3N_4, layer but they were grown on the SiO_2 tunnel oxide and afterwards capped by Si_3N_4. We demonstrated that there are no benefits in this configuration concerning the data retention, due to the Si_3N_4 presence, because the physical tunnel barrier that separates the nanocrystals from the substrate corresponds to the tunnel oxide thickness only. Furthermore the Si_3N_4 capping layer enables the parasitic charge trapping at the tunox/SiN interface by increasing the charge loss with the temperature. Moreover, the presence of parasitic charge trapping at the tunox/Si_3N_4 interface facilitates the charge loss at high temperature. Concerning the cell endurance using the silicon nitrite capping layer, the coupling factor is increased, thus the programming window increases too, but the parasitic charge trapping in the Si_3N_4 does not lead to produce satisfactory cell functionality after 100k program/erase cycles. Another important point is the charge loss dependence on tunnel oxide thickness. We extrapolated the activation energies related to the different samples. Obviously the charge loss increases when the tunnel oxide thickness decreases. We noticed that a 5.2nm tunnel oxide is needed to achieve the data retention specification for temperatures up to 150°C. On the other hand, the tunnel oxide thickness strongly impacts the erase Fowler-Nordheim operation, hence to obtain satisfactory cell functioning after 100k program/erase

cycles, a tunnel oxide thickness of 3.7nm has to be used. Finally we evaluated the cell behavior using a 4.2nm tunnel oxide, embedded in a different architecture without the silicon nitride capping layer and with an optimized ONO stack (EOT=10.5nm) in order to increase the vertical electric field which improves the program/erase efficiency and the cell reliability. We demonstrated for the first time the silicon nanocrystal cell functioning up to 1M program/erase cycles by maintaining a 4V programming window in a wide range of temperatures from -40°C to 150°C. Moreover, by avoiding the SiN capping, the data retention is also improved for cycled samples. The silicon nanocrystal covered area and the channel doping dose increase improve the programming window, hence the endurance performance. Furthermore we have shown that these two technological parameters do not impact the data retention results. Finally the Si-nc cell was compared with the floating gate. The endurance experiments have shown better behavior for the Si-nc cell up to 1M program/erase cycles, while the charge loss is higher at 250°C due to the thinner tunnel oxide.

We presented a new dynamic technique of drain current measurement **in chapter four**. This innovative method is presented for the first time in literature and can be used for different cell architectures (floating gate, silicon nanocrystal and split gate). We characterized the consumption of floating gate and silicon nanocrystal cell under various bias conditions and programming schemes. We compared in particular the ramp and box pulses on gate terminal during a channel hot electron programming operation. In this way we optimized the programming pulses for the two devices in order to minimize the energy consumption and the drain current peak. For the Flash floating gate we propose to use a ramped gate followed by a plateau, while a box pulse can be used in the case of the silicon nanocrystal memory cell. Using TCAD simulation we explained the *transistor-like* behavior of the silicon nanocrystal cell when the SiN capping layer is used. This is due to the discrete nature of the charge trapping layer and thus the localization of trapped charge. The optimized silicon nanocrystal cell was also characterized showing an intermediate behavior between the floating gate and the hybrid Si-nc cell. This decreases the cell consumption by increasing the programming window and thus the cell programming efficiency. Using the optimized gate and drain biasing we demonstrated that it is possible to obtain a 4V programming window with sub-nanojoule energy consumption. Finally we compared the optimized silicon nanocrystal cell with the Flash floating gate. The programming time has been fixed to 2µs in order to propose benchmarking in the case of low energy and fast application. We notice that when the box

pulse is applied the floating gate consumption is 50% less than with the Si-nc optimized cell, except for short pulses when the performances of the two devices become more similar. In spite of this, using the ramp programming, the silicon nanocrystal cell has the best consumption level. The optimized silicon nanocrystal cell can be considered as a good alternative to the flash floating gate in terms of programming speed and energy consumption while keeping a satisfactory level of programming window.

	Silicon Nanocrystal	Floating gate
Programming (Vg_{ramp}=3V to 9V; 1.5V/µs; Vd=4.2V)	**3.5µs** to reach a 4V programming window	**3.5µs** to reach a 4V programming window
Erase (Vg_{ramp}=0V to -18V; 5kV/s)	**200µs** to reach a 4V programming window	**500µs** to reach a 4V programming window
Data retention @ 250°C	**2.3V** lost in 168h	**0.5V** lost in 168h
Endurance 1Mcycles	Programming window of 4V	Programming window of 2.8V
Energy consumption (1µs box pulse)	**0.36nJ** to reach a 5V programming window	**0.24nJ** to reach a 5.4V programming window

Table 5. 1.Performance comparison achieved for optimized silicon nanocrystal and floating gate memory cell.

5.1 Perspectives

In this thesis we focused in particular on the silicon nanocrystal cell current consumption during the channel hot electron programming operation. Hereafter we propose two interesting points to study in a future work.

Dependence of current consumption on silicon nanocrystals size and density

We explained the cell behavior when the SiN capping layer is used. Using large nanocrystals it is possible to activate a mechanism of charge diffusion in the charge trapping layer. This explains behavior similar to the floating gate device, where the charge is distributed on the channel area. Preliminary experimental results lead us to say that is possible to drive the drain current consumption, controlling the charge diffusion mechanism. We conclude that the charges in silicon nanocrystals positioned on the drain side (HCI zone) can control the programming window, while the stored charges in Si-ncs, close to the source, control the

consumed current. The charge diffusion changes the cell behavior emulating a double transistor functioning where the transistor on the source side acts as an access transistor with a threshold voltage that depends on the quantity of the trapped charges. To increase the ability to control the channel current, one way is to increase the charge diffusion toward the source side. In figure 5.1 we show a schematic of the silicon nanocrystal cell where the mechanism of charge diffusion is enabled.

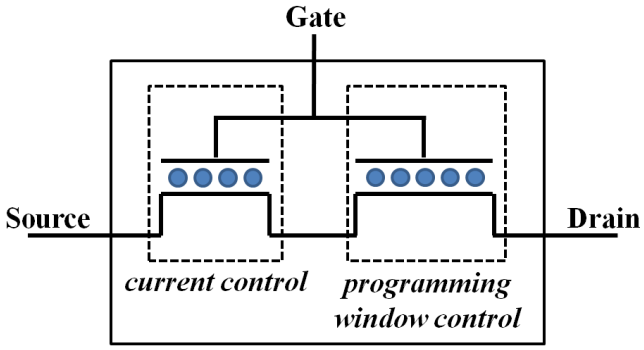


Figure 5. 1. Schematic of silicon nanocrystal cell when the charge diffusion mechanism is enabled.

New Cell architecture (ATW-Flash)

As an alternative to the 1T silicon nanocrystal memory and 1.5T Split Gate, we proposed a new cell architecture called ATW-Flash (Asymmetrical Tunnel Window). The schematic of this cell is shown in figure 5.2.

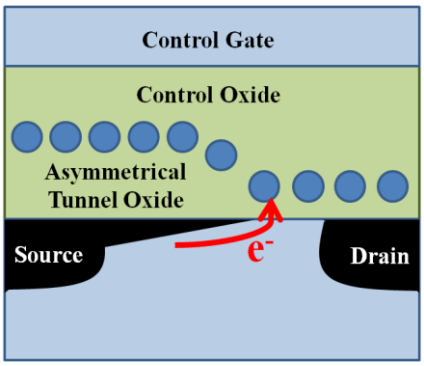


Figure 5. 2. Schematic of Asymmetrical Tunnel Window Flash (ATW-Flash)

Using the asymmetrical tunnel oxide thickness it is possible to improve the hot carrier injection by increasing the vertical electric field when the channel is pinched-off. Moreover the consumed current is controlled by the tunnel oxide thickness in the source region. In this way the programming efficiency can be improved. This cell presents advantages concerning to current consumption of a split gate cell in a small area. The real current consumption can be evaluated with our dynamic method of measurement. As a drawback, the high electric field generated in the tunnel oxide zone, where the thickness is varied, can stress the device, thus limiting the endurance performance.

Chapter 6 - Résumé du travail de thèse en français

6.1 Présentation de la thèse

Ce travail de thèse concerne une étude expérimentale et de modélisation sur les mémoires à nanocristaux de silicium qui représentent une des solutions les plus intéressantes pour remplacer le dispositif Flash à grille flottante. L'objectif de ce travail de thèse est de comprendre les mécanismes physiques qui gouvernent le comportement de la cellule à nanocristaux de silicium afin d'optimiser l'architecture du dispositif et comparer les résultats avec la cellule Flash standard.

Dans le premier chapitre nous présenterons le contexte économique, l'évolution et le fonctionnement des mémoires Flash-EEPROM. Ensuite une description détaillée de la technologie, du comportement et des limitations de la miniaturisation sera fournie. En conclusion nous exposerons les solutions possibles pour résoudre ces problèmes.

Le deuxième chapitre présentera le setup expérimental ainsi que les méthodes de caractérisation utilisées dans le but de mesurer les performances de la cellule mémoire à nanocristaux de silicium. De plus, l'impact des principaux paramètres technologiques comme par exemple : la nature des nanocristaux, la présence d'une couche de nitrure de silicium, la dose de dopage du canal et l'épaisseur de l'oxyde de tunnel, sera analysé. Une optimisation de l'empilement de la cellule mémoire sera aussi proposée pour pouvoir comparer les performances avec celles de la cellule Flash à grille flottante.

Dans le troisième chapitre, l'impact des principaux paramètres technologiques sur la fiabilité (endurance et rétention des données) sera étudié. Les performances de la mémoire à nanocristaux de silicium pour des applications fonctionnelles dans un intervalle étendu de température [-40°C; 150°C] seront aussi évaluées en montrant pour la première fois une endurance de la cellule jusqu'à 1 million de cycles avec une fenêtre de programmation finale de 4V. Pour conclure la cellule optimisée proposée sera comparée à la cellule Flash à grille flottante.

Le chapitre quatre décrit une nouvelle technique de mesure dynamique pour le courant de drain consommé pendant l'injection d'électrons chauds. Cette procédure permet l'évaluation de la consommation d'énergie lorsqu'une opération de programmation est conclue. Cette méthode est appliquée pour la première fois aux cellules mémoires à grille flottante et à nanocristaux de silicium. Une étude concernant la typologie des pulses durant la programmation et l'impact des paramètres technologiques sera présentée dans ce chapitre.

Enfin, nous démonterons qu'il est possible de rejoindre une consommation énergétique inférieure à 1nJ en sauvegardant une fenêtre de programmation de 4V.

Le manuscrit terminera avec une conclusion générale qui résume les différents résultats obtenus dans ce travail de thèse, avant de proposer quelques perspectives.

6.2 Le marché des mémoires à semi-conducteur

Pendant les dix dernières années, le marché des mémoires à semi-conducteur a subi une forte augmentation, grâce à l'énorme quantité de produits tels que les smartphones et autres tablettes vendus de part le monde (figure 6.1). Toutes ces applications demandent de plus en plus de hautes performances telles que: faible consommation d'énergie, temps d'accès courts, bas coûts, etc. C'est pourquoi le commerce des mémoires Flash gagne des parts de marché par rapport aux autres types de mémoires. Malgré le fait que le marché est en train de croître en continu, le prix des dispositifs mémoires diminue.

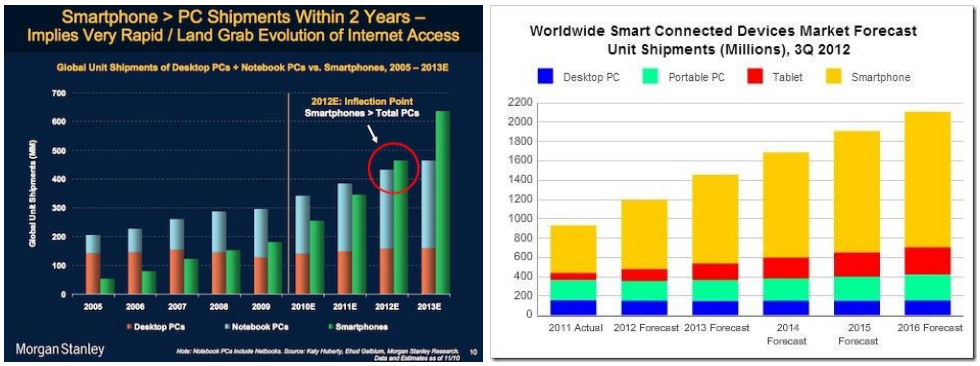


Figure 6. 1. Evolution et prévisions du marché des appareils portables (source: muniwireless.com et trak.in)

6.3 Les mémoires à nanocristaux de silicium

Le marché des mémoires flash non volatiles appliqué aux systèmes portables demande toujours une plus faible consommation d'énergie et des solutions à haute fiabilité. Les cellules mémoires Flash à nanocristaux de silicium paraissent être une des candidates les plus prometteuses pour des applications intégrées. Le principe de fonctionnement des mémoires à nanocristaux de silicium (Si-nc) à piégeage de charges discret est similaire à celui des dispositifs à grille flottante. Dans ce travail de thèse nous considérons l'intégration des mémoires Si-nc dans une architecture NOR pour des applications dédiées, programmées et effacées respectivement par les mécanismes: électrons chauds et Fowler-Nordheim.

Les avantages de l'utilisation de cette technologie sont:

- Résistance contre le SILC et RILC: ceci permet de diminuer l'épaisseur de l'oxyde de tunnel en dessous de 5nm, en conservant toujours une contrainte de dix ans pour la rétention de données. Par ailleurs, les tensions des opérations d'écritures et d'effacement peuvent être également diminuées.

- Compatibilité avec le procédé de fabrication CMOS standard afin d'encourager la production industrielle, en réduisant le nombre de masques utilisés par rapport à la fabrication du dispositif à grille flottante.

- Diminution des effets de perturbations de la cellule mémoire: Grâce à la nature discrète des nanocristaux ainsi qu'à leur petite taille, le facteur de couplage entre la grille et le drain est réduit autant que les interférences entre cellules voisines.

- Application multi-niveaux: La tension de seuil d'un transistor à nanocristaux de silicium dépend de la position de la charge stockée tout le long du canal.

En dépit de ces particularités, deux inconvénients importants caractérisent la mémoire Si-nc:

- Le faible facteur de couplage entre la grille de contrôle et les nanocristaux.

- La dispersion sur la surface recouverte avec les nanocristaux qui limite ce type de cellule pour des applications à haute densité d'intégration.

Des études importantes ont été amenées par les sociétés STMicroelectronics, Atmel et Freescale, qui ont démontré la possibilité de pouvoir obtenir des résultats intéressants en termes de fenêtre de programmation satisfaisante, haute fiabilité et intégration au sein d'architectures mémoires spécifiques (Split Gate).

6.4 Caractérisation électrique de la cellule mémoire à nanocristaux

Dans ce chapitre, nous évaluerons l'impact des principaux paramètres technologiques sur la fenêtre de programmation de la cellule mémoire à nanocristaux produite au sein de STMicroelectronics. Le but de cette analyse a été de définir la meilleure façon d'améliorer la fenêtre de programmation en utilisant toujours les pulses de programmation et d'effacement standards utilisées pour la cellule à mémoire Flash à grille flottante. Ensuite nous résumerons les conclusions principales obtenues par les études de caractérisations électriques de la cellule.

- Lorsque la taille des nanocristaux est augmentée, donc par conséquent la surface recouverte de la cellule, la fenêtre de programmation augmente et en particulier l'opération d'effacement par Fowler-Nordheim est améliorée. Nous avons constaté que l'utilisation de l'empilement mémoire standard nécessitait un recouvrement de 95% pour obtenir une fenêtre de programmation de 4V, pourcentage non cohérent avec le principe de fonctionnement de la cellule à nanocristaux de silicium. Afin d'améliorer la fenêtre de programmation et d'optimiser l'empilement de la cellule Si-nc nous avons considéré comme point clé l'augmentation du facteur de couplage comme expliqué dans la littérature pour les mémoires Flash à grille flottante. Deux différentes recettes ont été développées afin d'obtenir des nanocristaux de silicium avec une taille moyenne de 9 nm et 12nm qui permettent d'arriver respectivement à une surface recouverte de 46% et 76%. De plus, avec l'optimisation du facteur de couplage il a été possible de diminuer l'épaisseur de la couche d'ONO jusqu'à 10,5nm d'épaisseur équivalente, ceci a permis d'augmenter le champ électrique vertical pendant l'opération d'effacement. Cette valeur d'épaisseur a été choisie en accord avec les recettes disponibles dans la ligne de production de STMicroelectronics.

- La présence de la couche Si_3N_4 recouvrant les nanocristaux de silicium augmente la probabilité de piégeage de charge et la surface de canal recouverte. Le facteur de couplage est augmenté et donc la fenêtre de programmation augmente aussi. Dans les observations CDSEM nous avons noté que la couche Si_3N_4 pousse autour des nanocristaux de silicium. Dans ce cas il n'est possible de confirmer si les améliorations obtenues concernant la fenêtre de programmation dont dues à la présence de la couche Si_3N_4 ou à l'augmentation de la surface recouverte. Dans la figure 6.2 nous montrons les résultats concernant la fenêtre de programmation obtenus en utilisant des échantillons avec différentes dimensions de nanocristaux de silicium et avec la couche Si_3N_4. Dans ce cas nous pouvons considérer que l'amélioration obtenue dépend principalement de la surface recouverte et faiblement de l'augmentation de probabilité de piégeage de charge. Même si la présence de la couche Si_3N_4 est utile pour améliorer la fenêtre de programmation, nous avons décidé d'éviter cette étape process afin de minimiser les effets de piégeage de charges parasites.

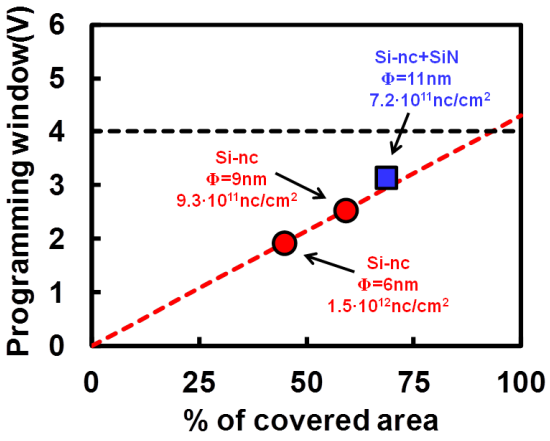


Figure 6. 2. Fenêtre de programmation en fonction de la surface recouverte. Les cellules à nanocristaux de silicium (Si-nc) et hybride (Si-nc+SiN) sont comparées.

- Nous avons montré qu'il est possible d'augmenter la fenêtre de programmation en augmentant la dose de dopage dans le canal, en considérant toujours le décalage des tensions de seuil. En augmentant la dose de dopage de canal jusqu'à 10^{14}at/cm^2, une augmentation de 20% de la fenêtre de programmation a été obtenue. Dans ce cas la régulation des tensions de programmation/effacement est nécessaire, donc afin de ce faire il est important de trouver le meilleur compromis lorsque d'autres paramètres sont modifiés: taille des nanocristaux et épaisseur de l'oxyde de tunnel. Après ces considérations nous avons décidé d'utiliser une dose de dopage de canal de 10^{14}at/cm^2 pour l'optimisation de la cellule avec l'objectif d'améliorer la fenêtre de programmation.

- Enfin nous avons étudié l'impact de l'épaisseur de l'oxyde de tunnel sur les opérations de programmation et effacement. Nous avons démontré que pendant la programmation par injection d'électrons chauds l'épaisseur de l'oxyde de tunnel impacte faiblement la fenêtre de programmation de part la dépendance prédominante du champ électrique horizontal. A contrario, ce paramètre technologique impacte fortement les opérations par Fowler-Nordheim. En particulier nous avons montré l'effet sur l'opération d'effacement. Afin d'obtenir une fenêtre de programmation de 4V, un oxyde de tunnel épais de 4.2nm maximum est nécessaire.

L'empilement optimisé de la cellule à nanocristaux apparaît dans la figure 6.3. Un oxyde de tunnel (SiO2) d'épaisseur 4.2nm a été crû sur un substrat de silicium dopé type-p ayant en surface une dose de 10^{14}at/cm^2. Deux recettes distinctes ont été développées pour pouvoir faire croître des nanocristaux de silicium avec une dimension moyenne de 9nm et 12nm.

L'empilement de la cellule a été complété avec une couche ONO d'épaisseur équivalente de 10.5nm.

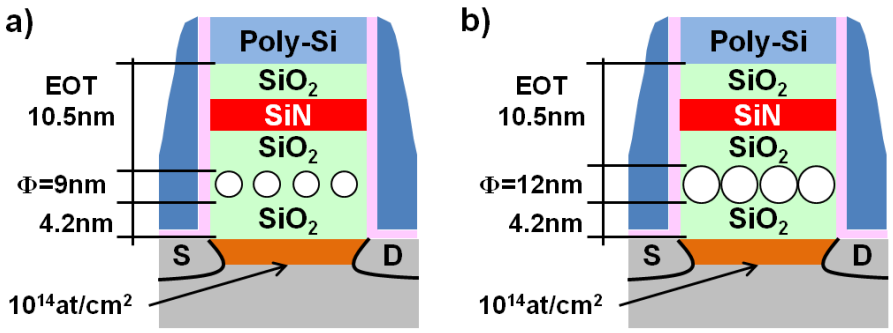


Figure 6. 3. Schéma de la cellule à nanocristaux de silicium optimisé; des nanocristaux avec deux tailles différentes sont mises en œuvre: a) Φ = 9 nm, b) Φ = 12nm.

Pour conclure cette partie nous avons comparé les résultats obtenus pour la cellule optimisée Si-nc avec la Flash standard à grille flottante. Dans la figure 6.4 nous montrons les caractéristiques de la cinétique de programmation des deux dispositifs. Pour la cellule optimisée Si-nc les performances sont les mêmes que la cellule à grille flottante; la fenêtre de programmation minimale de 4V est obtenue en utilisant une programmation par électrons chauds d'une durée de 3.5µs.

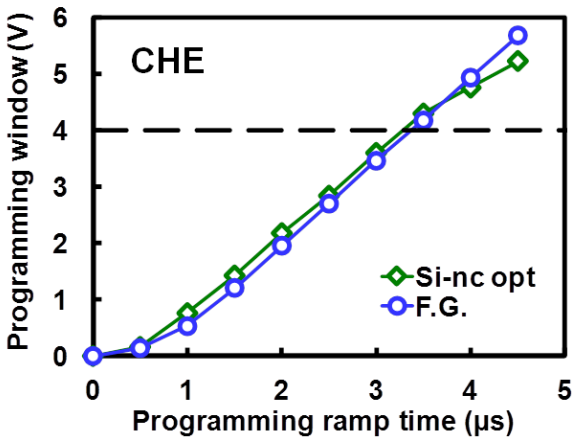


Figure 6. 4. Caractéristiques de la cinétique de programmation par porteurs chauds (Vg_ramp=1.5V/µs, Vg=[3V; 9V], Vd=4.2V).

Les caractéristiques de la cinétique d'effacement apparaissent quant à elles dans la figure 6.5. Les performances d'effacement ont été améliorées par rapport à la cellule mémoire à grille flottante grâce à un oxyde de tunnel d'épaisseur moindre ainsi qu'à un facteur de couplage plus important.

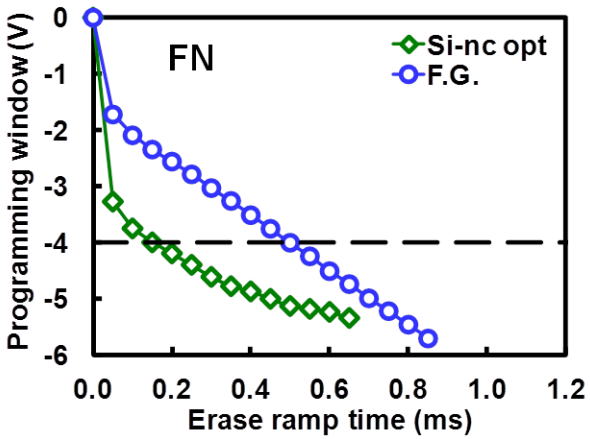


Figure 6. 5. Caractéristiques de la cinétique d'effacement par Fowler-Nordheim (Vg_ramp=5kV/s; Vg=[-14V; -18V]).

Le temps d'effacement pour obtenir la fenêtre de programmation minimale de 4V est de 0.2ms pour la cellule optimisée Si-nc permettant un gain de 60% par rapport à la cellule Flash à grille flottante. Pour conclure, tous les essais réalisés en faisant varier les différents paramètres technologiques ont permis d'optimiser la fenêtre de programmation de la cellule Si-nc avec l'objectif de pouvoir substituer la cellule à grille flottante et donc de diminuer les coûts de production. Dans le prochain paragraphe seront comparés les résultats concernant la fiabilité de la mémoire à nanocristaux de silicium et à grille flottante.

6.5 Fiabilité de la cellule à nanocristaux de silicium

Dans ce chapitre nous allons évaluer l'impact des principaux paramètres technologiques sur la fiabilité de la cellule mémoire à nanocristaux de silicium. Les résultats expérimentaux reportés ici concernant la rétention des données et l'endurance ont été comparés avec la littérature, et plus particulièrement la perte de charge et la dégradation de la tension de seuil pendant le cyclage dû au piégeage parasite des charges dans la couche de nitrure ont été analysés. L'objectif de cette analyse a été de définir la meilleure façon d'obtenir des résultats satisfaisants en termes de fiabilité. Afin d'optimiser la cellule à nanocristaux nous avons pris en considération les points suivants:

- Dans la littérature a été montré que la rétention des données n'est pas fortement impactée par la taille des nanocristaux. Cependant ce paramètre a une influence directe sur la fenêtre de programmation et l'efficacité d'effacement comme écrit précédemment. De plus, l'augmentation de la surface recouverte amène à un meilleur fonctionnement de la cellule après 100k cycles de programmation/effacement. Dans tous les cas nous avons noté que l'augmentation de la surface recouverte, donc du

facteur de couplage, en utilisant l'empilement mémoire standard, n'est pas suffisante pour effacer les charges parasites piégées dans le diélectrique de contrôle ONO. Il est donc important d'augmenter la taille ainsi que la densité des nanocristaux, même si d'autres améliorations sont nécessaires.

- La présence d'une couche de nitrure de silicium recouvrant les nanocristaux augmente la probabilité de piéger des charges et augmente aussi la surface recouverte de la cellule. Dans ce cas il n'y a pas d'amélioration concernant la rétention des données liée à la présence de la couche de Si_3N_4 car la barrière qui doit être considérée correspond seulement à l'épaisseur de l'oxyde de tunnel. On remarque que la présence des charges parasites piégées à l'interface entre l'oxyde de tunnel et le nitrure facilite la perte des électrons à température élevée. Concernant l'endurance de la cellule où la couche de Si_3N_4 sur les nanocristaux a été intégrée, le facteur de couplage est augmenté et donc la fenêtre de programmation est augmentée elle aussi. Par contre ce piégeage parasite ne permet pas d'obtenir un fonctionnement de la cellule satisfaisant après 100k cycles de programmation/effacement en utilisant l'empilement mémoire standard. Afin d'éviter le piégeage parasite qui empire la fiabilité de la cellule à nanocristaux, nous avons décidé de s'affranchir de l'étape de dépôt de la couche Si_3N_4.

- La rétention des données reste inchangée lorsque la dose de dopage du canal (CDD) est modifiée. Il est possible par conséquent d'obtenir un gain sur la fenêtre de programmation en augmentant la dose de dopage de canal mais dans ce cas une régulation des niveaux programmés/effacés doit être réalisée. Le CDD=10^{14}at/cm^2 a été choisi pour la cellule mémoire optimisée à nanocristaux de silicium.

- Enfin, nous avons montré la dépendance de la perte des charges de l'épaisseur de l'oxyde de tunnel et nous avons extrapolé les énergies d'activation pour chacun des échantillons. Comme dans le cas de l'opération d'effacement, la perte des charges augmente lorsque l'épaisseur de l'oxyde de tunnel est diminuée. Nous avons constaté qu'une épaisseur d'oxyde de tunnel de 5.2nm est nécessaire pour parvenir à la spécification de rétention des données pour les températures jusqu'à 150°C. D'autre part, l'épaisseur d'oxyde de tunnel a un impact considérable sur l'opération d'effacement faite par Fowler-Nordheim. Donc pour obtenir une fenêtre de programmation adéquate après 100k cycles un oxyde de tunnel de 3.5nm doit être

utilisé. Pour notre étude il a été important d'évaluer le comportement de la cellule en utilisant un oxyde de tunnel de 4.2nm dans une architecture mémoire où la couche ONO a été optimisée et la couche Si_3N_4 n'est pas déposée.

En conclusion de ce paragraphe nous avons comparé les résultats concernant la cellule à nanocristaux optimisée avec la cellule Flash à grille flottante. Dans la figure 6.6 la rétention des données à 250°C est représentée pour chacun des deux dispositifs. Pour satisfaire la contrainte de rétention des données la cellule doit maintenir un niveau de la tension de seuil supérieur à 5,75V à 250°C pendant 168 heures. On peut observer que la cellule Si-nc est à la limite de cet objectif cependant d'autres efforts seront nécessaires afin de pouvoir encore améliorer les performances et donc atteindre les résultats obtenus en utilisant la cellule à grille flottante. La principale contrainte est représentée par la perte de charges rapide initiale due au piégeage d'électrons dans l'oxyde de tunnel, dans la couche ONO et dans les interfaces relatives.

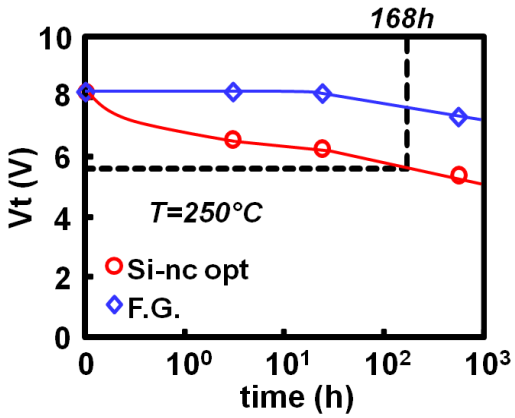


Figure 6. 6. Rétention des données de la cellule à nanocristaux de silicium optimisé (Si-nc opt) et à grille flottante (F.G.) à la température de 250°C.

Les résultats concernant l'endurance ont été aussi comparés et les conditions de programmation/effacement sont restées inchangées (programmation: Vg=9V, Vd=4.2V, tp=1µs et effacement: Vg=-18V, ramp=5kV/s+te=1ms). Les résultats expérimentaux pour la cellule optimisée à nanocristaux de silicium sont montrés dans la figure 6.7. Comme nous pouvions nous y attendre, la cellule Flash à grille flottante présente une fenêtre de programmation plus importante au début du cyclage, ceci grâce à son meilleur facteur de couplage et à sa meilleure efficacité de programmation. La dégradation la plus importante de la tension de seuil détermine une fermeture majeure de la fenêtre de programmation après 1 millions de cycles, En même temps la caractéristique de l'endurance est plus stable pour la cellule Si-nc.

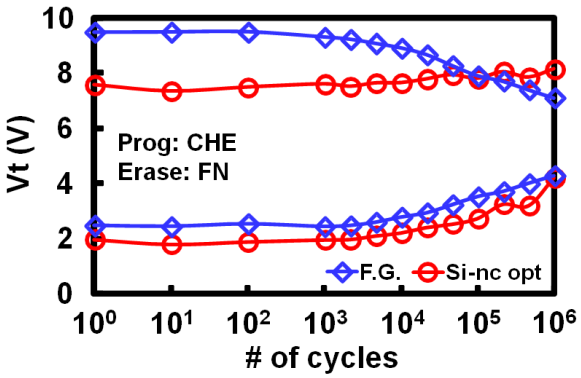


Figure 6. 7. Caractéristiques d'endurance des cellules à nanocristaux de silicium optimisé (Si-nc opt) et à grille flottante (F.G.).

En conclusion nous avons démontré pour la première fois, à notre connaissance, le fonctionnement d'une cellule à nanocristaux de silicium jusqu'à 1 million de cycles de programmation/effacement. Une fenêtre de programmation de 4V est ainsi préservée dans un intervalle de température étendu [-40°C ; 150°C] (figure 6.8). En contrepartie le principal inconvénient pour la cellule à nanocristaux est représenté par la perte des charges importante à 250°C.

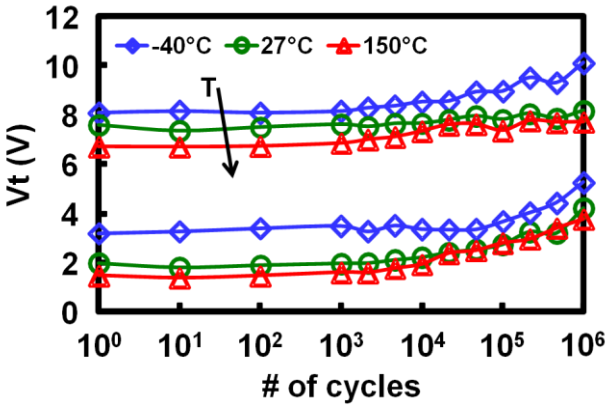


Figure 6. 8. Caractéristiques d'endurance des cellules à nanocristaux de silicium optimisé dans un intervalle de température étendu [-40°C ; 150°C].

6.6 Consommation de la cellule pendant une opération de programmation par injection d'électrons chauds

Dans cette section nous présentons les résultats concernant la consommation de courant et d'énergie de la cellule Flash à grille flottante et à nanocristaux de silicium pendant une opération de programmation faite par injection d'électrons chauds. L'évaluation de la consommation du courant d'une cellule Flash à grille flottante peut être effectuée à travers des manipulations utilisant un convertisseur courant/tension ou par technique indirecte. De cette manière là il n'est pas possible de comprendre le comportement dynamique et de mesurer les performances de la cellule implémentée dans une architecture NOR pour des

pulses de programmation de quelques microsecondes. D'autre part la méthode indirecte de calcul du courant consommé n'est pas fonctionnelle pour les mémoires à nanocristaux de silicium. Dans ce contexte nous avons développé une nouvelle méthode expérimentale qui permet de mesurer dynamiquement la consommation du courant pendant une opération de programmation effectuée par injection d'électrons chauds. Cette méthode a permis de comprendre le comportement dynamique des deux dispositifs. L'énergie consommée a été aussi évaluée en utilisant différentes conditions de polarisation. L'objectif a été de caractériser l'impact de différents paramètres sur la consommation et de trouver le meilleur compromis afin d'améliorer les performances des deux cellules mémoires en question. De plus, la consommation due aux fuites des cellules non sélectionnées dans le plan mémoire a été mesurée pour compléter l'étude. Nous allons décrire les principaux résultats obtenus pour la cellule à nanocristaux de silicium et la cellule Flash à grille flottante.

6.7 Optimisation de la consommation énergétique

Les caractérisations électriques effectuées sur la cellule à grille flottante et sur la cellule à nanocristaux recouverts par la couche Si_3N_4 ont permis d'optimiser les polarisations ainsi que la forme des signaux utilisées pendant l'opération de programmation. Pour la cellule Flash à grille flottante il a été démontré que le meilleur compromis entre l'énergie consommée et la fenêtre de programmation se trouve être l'utilisation d'une rampe de programmation suivie par un plateau. Cette façon de programmer la cellule permet de réduire respectivement le pic du courant de drain et de maintenir une fenêtre de programmation adéquate. Dans le cas de la cellule à nanocristaux de silicium recouverts par la couche Si_3N_4 (Si-nc+SiN), nous avons trouvé que les meilleurs résultats sont acquis lorsque des pulses très courts sont appliqués au terminal de grille. Nous reportons ci-dessous les résultats expérimentaux mesurés pour une cellule à nanocristaux avec un oxyde de tunnel épais de 4.2nm, des nanocristaux avec une taille moyenne de 12nm et une fine couche d'ONO (10.5nm). Dans la figure 6.9 sont montrés les résultats de la consommation de drain lorsque des pulses de formes carrées et de différentes durées sont appliqués à la grille de contrôle. Nous remarquons que le courant de drain ne suit pas le potentiel de la grille de contrôle, mais dans ce cas il diminue pendant le temps de programmation. Ce type de comportement est similaire à celui d'une cellule Flash à grille flottante où un pic de courant est mesuré. Nous avons démontré alors que la différence de comportement entre la cellule à grille flottante et la cellule Si-nc+SiN est justifiée par la localisation de la charge piégée. Dans le cas de la cellule optimisée nous pouvons affirmer

que l'effet de la localisation due à la présence de la couche SiN n'est pas présent. Par ailleurs la grosse taille des nanocristaux produit une distribution de la charge vers le centre du canal et modifie le potentiel de surface de canal. Dans la figure 6.10 nous avons représenté les schémas des divers empilements mémoire et les mesures relatives au courant de drain afin de pouvoir comparer le comportement des trois cellules : Si-nc, Si-nc+SiN et F.G.

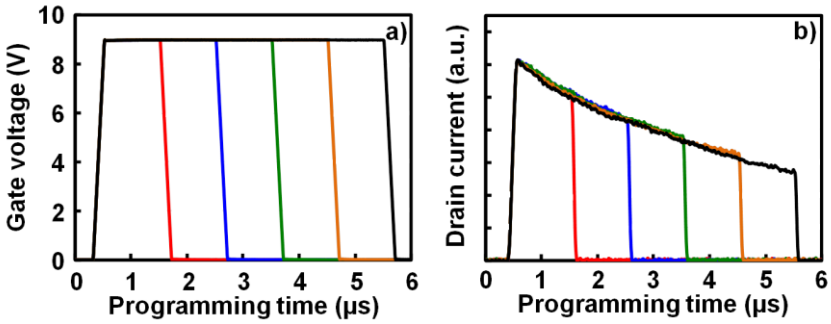


Figure 6. 9. a) Pulses de différentes durées appliquées à la grille de contrôle. b) Résultats de la consommation du courant de drain.

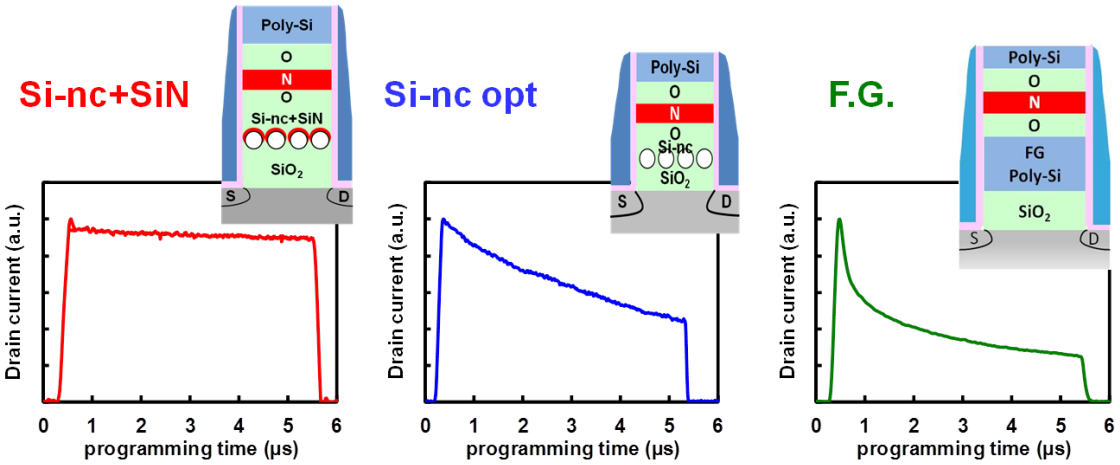


Figure 6. 10. Courant de drain dynamique mesuré pour les cellules à nanocrystaux des silicium avec et sans la couche SiN, et à grille flottante.

La figure 6.10 montre que la cellule Si-nc optimisée a une réponse à mi-chemin entre les cellules Si-nc+SiN et grille flottante. Cela signifie qu'il est possible de contrôler le courant dynamique en faisant varier la taille et la densité des nanocristaux, donc la surface recouverte. Nous avons expliqué que la cellule Si-nc+ SiN a un comportement de type transistor où le courant de drain suit le potentiel de la grille de contrôle de part la localisation de la charge à côté de la zone de drain. Au contraire pour le dispositif à grille flottante un pic de courant est détecté lorsqu' une impulsion de forme carrée est appliquée sur la grille de contrôle. Pendant l'injection de porteurs chauds les charges diffusent à travers les nanocristaux en direction de la source en faisant varier le potentiel de surface du substrat et donc les champs électriques. En prenant compte le comportement dynamique de la cellule à nanocristaux optimisée et le fait que les pulses de formes carrées sont plus efficaces que les rampes de programmation,

nous avons caractérisé la cellule mémoire pour différentes conditions de polarisation des terminaux de grille de contrôle et de drain. Enfin nous avons trouvé des conditions de programmation optimisées qui sont: Vg=10V, Vd=3.8V et un temps de programmation de 1µs. De cette manière il est possible d'atteindre les contraintes pour les applications à faible consommation d'énergie qui sont définies: fenêtre de programmation supérieure à 4V et énergie inférieure à 1nJ.

6.8 Comparaison des performances avec la cellule Flash à grille flottante

Pour conclure le rapport sur la consommation d'énergie, nous allons comparer maintenant les principaux résultats mesurés sur les cellules à nanocristaux de silicium optimisées et la cellule Flash à grille flottante. Après avoir comparé les différents comportements dynamiques des cellules on montre dans la figure 6.11 l'énergie consommée par rapport à la fenêtre de programmation en utilisant des pulses carrées de différentes durées. La cellule Si-nc optimisée est capable d'atteindre des niveaux de programmation comparables à celui de la cellule à grille flottante, néanmoins, l'énergie consommée reste plus élevée à cause d'un courant de drain plus important. En utilisant les données expérimentales nous avons extrapolé les lois qui décrivent le comportement de la cellule et nous avons démontré qu'en usant des pulses de programmation plus rapides la consommation de la cellule Si-nc optimisée peut être améliorée et devenir comparable à celles de la cellule standard à grille flottante. La cellule à nanocristaux que nous avons développé peut être donc considérée comme une alternative importante à la cellule Flash à grille flottante en termes de vitesse de programmation et d'énergie consommée en gardant toujours un niveau de fenêtre de programmation satisfaisant.

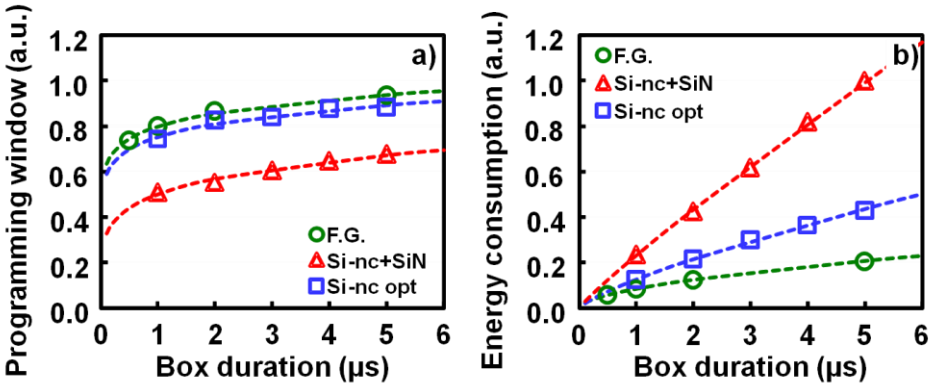


Figure 6. 11. a) Fenêtre de programmation et b) énergie consommée en utilisant des pulses de programmation carré de différentes durées.

6.9 Conclusion générale

Dans ce travail de thèse nous avons caractérisé et modélisé les cellules mémoires à nanocristaux de silicium. Suite à une étude détaillée des récentes implications des nanocristaux dans des dispositifs mémoire, nous avons optimisé l'empilement de la mémoire. Nous avons par conséquent caractérisé la fenêtre de programmation en faisant varier différents paramètres technologiques: taille et densité des nanocristaux, présence de la couche Si_3N_4, dose de dopage du canal et épaisseur de l'oxyde de tunnel. L'objectif des expériences a été de comprendre le comportement de la cellule afin d'améliorer le facteur de couplage et de minimiser le piégeage des charges parasites. Les résultats concernant la fiabilité de la cellule ont montré une rétention des données satisfaisante à 150°C et pour la première fois une endurance allant jusqu'à 1 million de cycles dans l'intervalle de température entre -40°C et 150°C avec une fenêtre de programmation de 4V. Enfin nous avons développé une technique innovante de mesure du courant dynamique pendant une opération de programmation faite par injection d'électrons chauds. Cette technique a été appliquée pour la première fois dans le but d'étudier les cellules Flash à grille flottante et à nanocristaux de silicium. Nous avons décrit le comportement dynamique et comment améliorer la consommation d'énergie pour atteindre une consommation inférieure à 1nJ toujours en gardant une fenêtre de programmation de 4V. Comme perspectives nous proposons de continuer à investiguer le phénomène de diffusion des charges dans la couche de piégeage de nanocristaux durant l'opération de programmation avec l'objectif de diminuer la consommation énergétique. En alternative il est possible d'intégrer les nanocristaux dans des architectures mémoires différentes de la cellule mémoire standard.

Caractérisation et Modélisation des Mémoires Avancées non Volatiles à Piégeage de Charge

Les mémoires à nanocristaux de silicium sont considérées comme l'une des solutions les plus intéressantes pour remplacer les grilles flottantes dans les mémoires Flash pour des applications de mémoires non-volatiles embarquées. Ces nanocristaux sont intéressants pour leur compatibilité avec les technologies de procédé CMOS, et la réduction des coûts de fabrication. De plus, la taille des nanocristaux garantie un faible couplage entre les cellules et la robustesse contre les effets de SILC. L'un des principaux challenges pour les mémoires embarquées dans des applications mobiles et sans contact est l'amélioration de la consommation d'énergie afin de réduire les contraintes de design de cellules. Dans cette étude, nous présentons l'état de l'art des mémoires Flash à grille flottante et à nanocristaux de silicium. Sur ce dernier type de mémoire une optimisation des principaux paramètres technologiques a été effectuée pour permettre l'obtention d'une fenêtre de programmation compatible avec les applications à faible consommation d'énergie. L'étude s'attache à l'optimisation de la fiabilité de la cellule à nanocristaux de silicium. On présente pour la première fois une cellule fonctionnelle après un million de cycles d'écriture et effacement dans une large gamme de températures [-40°C;150°C], et qui est capable de retenir l'information pendant dix ans à 150°C. Enfin, une analyse de la consommation de courant et d'énergie durant la programmation montre l'adaptabilité de la cellule pour des applications à faible consommation. Toutes les données expérimentales ont été comparées avec les résultats d'une cellule standard à grille flottante pour montrer les améliorations apportées.

Mots Clés: Mémoires à nanocristaux de silicium; grille flottante; consommation d'énergie; fenêtre de programmation; fiabilité; température

Characterization and Modeling of Advanced Charge Trapping Non Volatile Memories

The silicon nanocrystal memories are one of the most attractive solutions to replace the Flash floating gate for nonvolatile memory embedded applications, especially for their high compatibility with CMOS process and the lower manufacturing cost. Moreover, the nanocrystal size guarantees a weak device-to-device coupling in an array configuration and, in addition, for this technology it has been shown the robustness against SILC. One of the main challenges for embedded memories in portable and contactless applications is to improve the energy consumption in order to reduce the design constraints. Today the application request is to use the Flash memories with both low voltage biases and fast programming operation. In this study, we present the state of the art of Flash floating gate memory cell and silicon nanocrystal memories. Concerning this latter device, we studied the effect of main technological parameters in order to optimize the cell performance. The aim was to achieve a satisfactory programming window for low energy applications. Furthermore, the silicon nanocrystal cell reliability has been investigated. We present for the first time a silicon nanocrystal memory cell with a good functioning after one million write/erase cycles, working on a wide range of temperature [-40°C; 150°C]. Moreover, ten years data retention at 150°C is extrapolated. Finally, the analysis concerning the current and energy consumption during the programming operation shows the opportunity to use the silicon nanocrystal cell for low power applications. All the experimental data have been compared with the results achieved on Flash floating gate memory, to show the performance improvement.

Key words: Silicon nanocrystal memories; floating gate; energy consumption; programming window; reliability; temperature