University of Crete
Department of Physics
Microelectronics Research Group

CMOS COMPATIBLE NON-VOLATILE MEMORY DEVICES BASED ON III-N QUANTUM DOTS

Master Thesis
Microelectronics & Optoelectronics Graduate Program

Elena Papadomanolaki

Supervising Professor:
Eleftherios Iliopoulos

Crete, 2012

CONTENTS

Acknowledgements

I would like to thank my supervising professor, Mr. Eleftherios Iliopoulos, for his valuable guidance and help, and for appreciating my work. I would also like to express my thanks to Mr. Panagiotis Dimitrakis and Mr. Pascal Normand at IMEL/NCSR "Demokritos" for their indispensable collaboration on this project, as well as Caroline Bonafos at CEMES, CNRS, Toulouse, France for her contribution. I would like to thank Maria Kayambaki, whose help for the experimental part of my thesis was vital. My thanks also go to the members of my thesis committee and to anyone who takes an interest in this work.

What is more, I would like to thank my friends and colleagues at the University of Crete, Savvas Eftyhis, Alexandros Tzimis, and Katerina Papamichail for making me feel welcome and for pretending to listen to my academic problems. My thanks go to all my old friends, Maria, Effie, Giorgos, Iro, and Marina, for their love and for always providing a safe escape when I need it. I owe my gratitude to my parents, Afroditi and Manolis, for making my education possible and for other things, too many to be listed. My deepest love and thanks lastly go to my sister, Maria, and my dogs, Rabito and Roza, for the precious moments of stress relief.

Abstract

To investigate memory properties of GaN QDs, MOS capacitors were realized with GaN nanocrystals embedded in the SiO_2 dielectric. Samples were grown on n-type Si and Al was used as the metal gate. A reference sample with no QDs was also developed. High-frequency capacitance-voltage measurements were performed which revealed a wide hysteresis for all the samples except for the reference which exhibited no hysteresis for a full voltage sweep. C-V characteristics suggest electron injection that takes place from the substrate but limited hole injection. The width of the hysteresis was found to increase with the quantity of GaN embedded in the oxide. Pulse measurements confirm an asymmetry between electron and hole injection and suggest that erasing is more challenging than programming for these devices. Retention measurements at room temperature revealed good retention properties meeting the ten-year threshold for non-volatility. Finally, transient current measurements were performed to further investigate injection mechanisms.

The continuous floating gate transistor faces several challenges, the most significant of which is its scalability problem, and this does not allow NVMs to follow the rate of shrinkage of other integrated circuits. For that reason, several alternatives to the FG-MOSFET have been proposed in literature with the nanoparticle FG-MOSFET being a very prominent candidate. In this direction, GaN QDs could present two key advantages: compared to Si, GaN is considered to have a negative conduction band offset which could result in improved retention characteristics; in addition, it is fully compatible with current CMOS manufacture technologies. The current thesis investigates the potential of GaN QDs in NVMs.

The first chapter introduces the reader in the basic concepts of NVM, as well as the fundamental operation of the FG-MOSFET. The second chapter includes the theoretical background that is necessary for the understanding of this work. Physics of the MOS system are first described with emphasis on the capacitance of the device. The paragraphs that follow look into the properties of both the continuous and the nanoparticle floating gate respectively. Finally, chapter 3 describes the experimental process, the results and the analysis in full detail.

List of Figures

List of Tables

CHAPTER 1

Introduction to Nonvolatile Memories

Semiconductor memory is an essential part of modern information processors. The length of time that the memory can retain the data is a property called *retention* and the unpowered retention time parameter is the measure of *non-volatility*. A volatile memory will typically have a worst-case retention time of less than a second. A nonvolatile memory (NVM) is usually specified as a worst-case unpowered retention time of 10 years, although this parameter can vary from days to years depending on the specific memory technology and application. Integrated circuit nonvolatile memories are frequently further classified in terms of the degree of functional flexibility available for modification of stored contents (table 1.0-1). It should be mentioned here that out of the different categories of nonvolatile memory, Flash memory is of great significance. Flash memory is an EEPROM where the entire chip or a subarray within the chip may be erased at one time and it has been the dominant form of NVM over the years.

The conventional NVM device is the floating-gate MOS Field Effect Transistor (FG-MOSFET). The FG-MOSFET makes use of charge storage on a *floating gate*, a thin layer formed within the gate insulator of a field effect transistor between the gate electrode (the control gate) and the channel, usually made of polysilicon. The basic structure of a FG-MOSFET is presented in figure 1.0-1. The amount of charge on the floating gate determines if the transistor will conduct when a fixed set of 'read' bias conditions are applied. Because the floating gate is surrounded by insulators, it can retain its charge for long periods of time and in this way non-volatility is ensured.

The present chapter reviews the function principles of a simple NVM cell and introduces important characteristics and basic concepts that are useful in the study of nonvolatile memories.

Acronym	Definition	Description
ROM	Read-only memory	Memory contents defined manufacture and not modifiable.
EPROM	Erasable programmable ROM	Memory is erased by exposure to UV light and programmed electrically.
EEPROM	Electrically erasable programmable ROM	Memory can be both erased and programmed electrically. The se of 'EE' implies block erasure rather than byte erasable.
E^2PROM	Electrically erasable programmable ROM	Memory can be both erased and programmed electrically as for EEPROM, but the use of 'E^2' implies byte alterability rather than block erasable.

Table 1.0-1: Nonvolatile Memory Functional Capability Classifications.

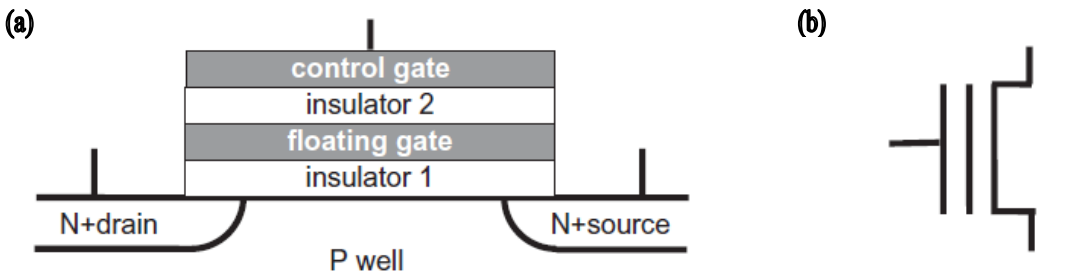


Figure 1.0-1: Floating-gate transistor: (a) elements of the transistor structure and (b) circuit symbol.

1.1 THE FLOATING GATE: STRUCTURE AND BASIC OPERATION

1.1.1 Structure of a FG-MOSFET

The basic structure of a FG-MOSFET is again presented in more detail in figure 1.1-1.

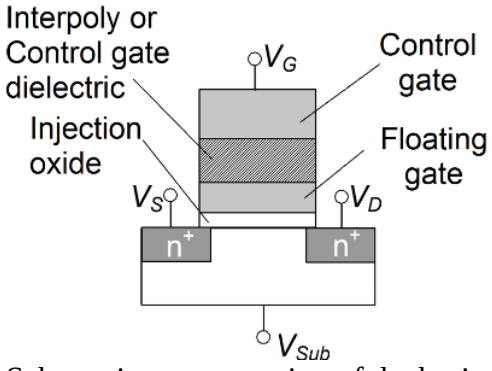


Figure 1.1-1: Schematic representation of the basic FG-MOSFET.

As it is shown, like a plain MOS Field Effect Transistor, it includes a source and a drain, two highly doped regions on a silicon substrate (body). The current will be conducted from source to drain (for a p channel) or from drain to source (for an n channel) when an inversion layer is formed. The inversion layer is formed, again, like a regular MOSFET, by the bias applied to the gate, only for a FG-MOSFET we call it a 'control gate' to distinguish it from the *floating gate*. The floating gate is a thin layer, usually made of polysilicon that lies between the control gate and the substrate and is isolated from both by two layers of dielectric: the injection oxide that separates it from the body and the interpoly dielectric that separates it from the control gate. The injection oxide (also called the tunnel oxide) is usually a thin (≥6nm) layer of silicon oxide and it is so called because it is where the charge of the floating gate is injected through. The interpoly dielectric (or control dielectric or control oxide) comprises of one or more types of dielectric; its purpose is not to allow the charges to escape to the control gate and should, therefore, be somewhat thicker than the injection oxide.

1.1.2 Operation Principles with Basic Equations

For a plain MOSFET device, the threshold voltage is usually defined as the gate voltage where an inversion layer forms at the interface between the oxide and the substrate. The inversion layer's forming allows the flow of carriers along the gate-source junction and, practically, the threshold voltage is the voltage at which there are sufficient minority carriers in the inversion layer to make a low resistance conducting path between the MOSFET source and drain.

The operation principle of FG-NVM is essentially based on the switching of the MOSFET's threshold voltage V_{th} between a low and a high value. This is achieved by storing an amount of charge into the floating gate. Since the floating gate controls at

least part of the underlying transistor channel, the charge on this gate will directly influence the current in the channel. The mechanism is in short, the following:

For simplicity, let us assume the charge that is stored in the floating gate comprises of electrons, even though a similar case can be made if the stored charge is positive (holes). The storage of negative charge reduces the channel potential and hence decreases the current flowing from the source to the drain. A negatively charged floating gate increases the threshold voltage and leads to a parallel shift of the transfer characteristics I_{DS}-V_{GS} with respect to the uncharged transistor (fig. 1.1-2). The shift can be calculated by the following expression:

$$V_{thW} = V_{th0} + Q_W / C_{IPD}$$

where V_{thW} and V_{th0} are the threshold voltage values after and before electron storage respectively, Q_W is the amount of stored electrons (in Cb) and $C_{IPD} = \frac{\varepsilon_{IPD} A}{t_{IPD}}$ is the capacitance (in F) of the interpoly dielectric. A is the gate area, ε_{IPD} is the dielectric constant of the interpoly dielectric and t_{IPD} is the interpoly dielectric thickness. This high V_{th} state is commonly named *Write* or *Program* state. If the stored mount of charge is decreased to

$$V_{thE} = V_{th0} + Q_E / C_{IPD} .$$

This new state is usually named *Erase* state and the difference between V_{thW} and V_{thE} is called *memory window*:

$$\Delta V_{th} = V_{thW} - V_{thE} = Q / C_{IPD} .$$

The operation that corresponds to the transfer of information from the outside world into the memory cell is referred to as the programming or write operation, whereas the operation that establishes the initial condition of the cell is called the erase operation.

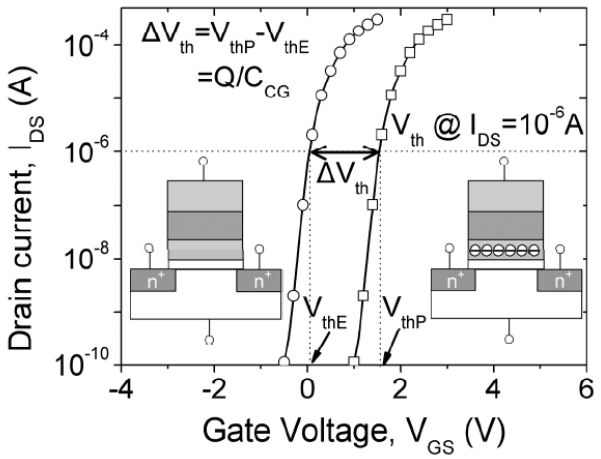


Figure 1.1-2: I-V characteristics of a memory cell in program and erase states and the extraction of the memory window.

Finally, the readout operation, communicates the current state (Write or Erase) of the memory cell to the outside world. The individual state of a cell, corresponding to the logic '0' or '1', is detected by applying a gate voltage in between the V_{thW} and V_{thE} values. Specific circuits sense the current flowing from the source to drain, I_{DS}, and indicate whether the transistor is 'ON' or 'OFF', i.e. whether the cell is in the

erase or the program state respectively. The readout operation uses a readout voltage that lies somewhere between the two distinct threshold voltages (fig. 1.1-3).

We note, that in the case that the stored charge is positive, the threshold voltage accordingly shifts to lower voltages, since the channel potential is enhanced:

$$V_{th} = V_{th0} - Q / C_{IPD}$$

In any case, the low V_{th} state is usually considered as the erase state while the high V_{th} state is the write state. In order to set the memory cell in the program state (logic state '1'), the floating gate should be charged negatively. Erasure of the programme state (logic state '0') requires at the very least the extraction of the stored electrons from the floating gate even though it is possible to have injection of holes; this further widens the memory window.

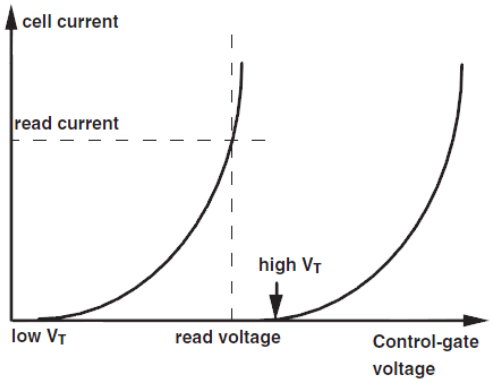


Figure 1.1-3: Reading principle for floating gate devices.

Gate Coupling Ratio

The floating-gate device, as operated from the control gate and the drain, acts as a metal-oxide-semiconductor transistor but with slightly different characteristics. The main reason is that the external control gate of the cell is not directly controlling the transistor channel. The floating gate, however, has a more direct impact on the transistor channel while it is biased only through capacitive coupling from other terminals (mainly the control gate).

A main parameter for characterizing memory cells is, therefore, the gate coupling ratio, which is defined as the ratio of the control-gate-to-floating-gate capacitance and the overall floating-gate capacitance (fig. 1.1-4). Basically, it quantifies the percentage of control gate voltage that is coupled form the control gate to the floating gate in any operation mode. The subthreshold slope of the memory cell as measured from this control gate is smaller than that of the floating-gate transistor by a factor equal to this gate-coupling ratio.

A general expression for the coupling ratio is:

$$\alpha = \frac{C_{PP}}{\sum_j C_j}$$

which becomes, if one considers only the capacitance of the tunnel oxide to be important:

$$\alpha = \frac{C_{PP}}{C_{TO} + C_{PP}}$$

where $C_{TO} = \frac{\varepsilon_{TO} A}{t_{TO}}$ is the capacitance of the injection (tunnel) oxide. In case the tunnel oxide and the interpoly dielectric are made of the same material, the coupling factor becomes $\alpha = \frac{t_{TO}}{t_{TO} + t_{IPD}}$.

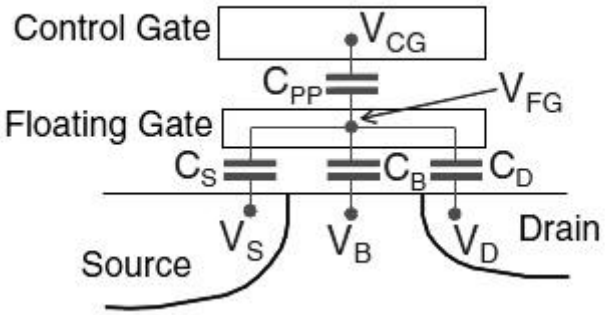


Figure 1.1-4: Schematic cross-section of a floating gate NVM cell showing the intrinsic capacitances ($C_B = C_{TO}$).

1.2 MEMORY CHARACTERISTICS: CONCEPTS AND DEFINITIONS

The floating gate transistor is the most fundamental and most common of all non-volatile memory devices and the study of its function is essential if we are to gain an understanding of the elemental theory of non-volatile memory. Of course, there are many more different kinds of devices for non-volatile memory; however, in order to characterize all of them one needs to familiarize themselves with some of the concepts that are usually used to evaluate their behaviour and their effectiveness. In this paragraph, we attempt to introduce these concepts, give definitions and explain their importance.

1.2.1 Memory Window

As it was mentioned earlier, the memory window of a memory cell is the difference between the higher and the lower values of the threshold voltages of its two distinct states, the program (or write) state and the erase state. The memory window plays a particularly important role in the effectiveness of the readout operation. To read the state of a memory device, a set of bias voltages is applied. A voltage pulse is applied on the control gate and the amplitude of that pulse lies somewhere between the lower and the higher threshold voltages of the device. Subsequently, the current that flows through the device is sensed by a highly sensitive current detecting circuit: if the current is higher (lower) than a predetermined value at these certain bias conditions then the cell is considered to be erased (programmed). If the memory window is too narrow there is a greater chance of an incorrect readout. Especially when dealing with large arrays of memory cells, the extracted threshold voltage may exhibit a statistical distribution around a central value (fig. 1.2-1), and a wide memory window makes it easier to avoid overlapping.

It becomes therefore obvious why the memory window is one of the critical measures for the evaluation of the performance of a non-volatile memory cell.

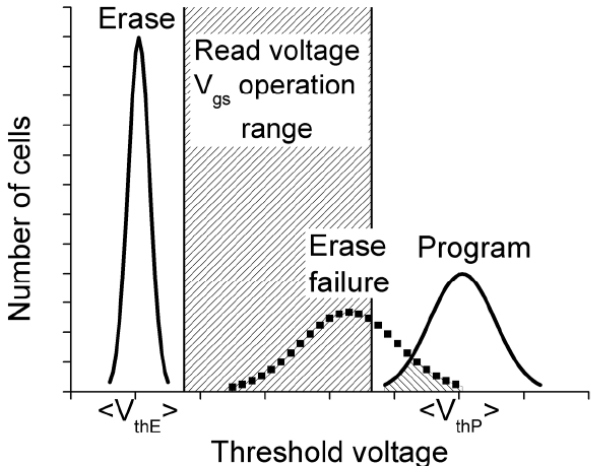


Figure 1.2-1: Representation of threshold voltage distribution for the program and erase states of a Flash memory cell.

The memory window can be measured if one extracts the current-voltage I_{DS}-V_{GS} characteristics of a NVM memory cell after applying a series of positive and negative

voltages on the control gate. As mentioned above, the pulses result in the shifting of the characteristics to either more positive or more negative values (fig. 1.2-2). The threshold voltage can then be calculated for each one of the characteristics using one of the many methods in literature, the simplest and most common of which is the constant current method. The constant current method evaluates V_{th} as the gate voltage corresponding to a given arbitrary constant drain current, a typical value for which is $(W_m/L_m)*10^{-7}$ where W_m and L_m are the mask channel width and length respectively.

Of course, when extracting the memory window as described above, one must keep in mind that the resulting value depends both on the duration and the amplitude of the pulse (fig. 1.2-3), even though after a certain voltage the window stops to vary significantly (fig 1.2-3a).

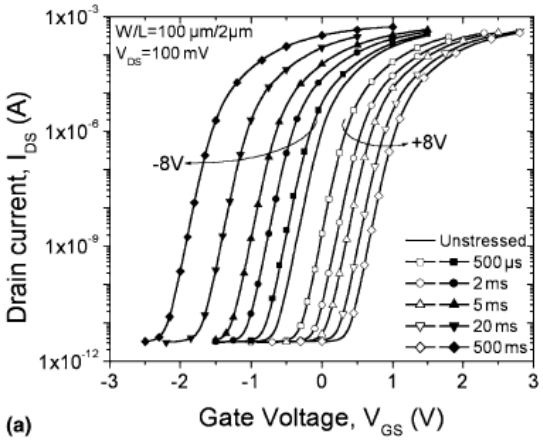


Figure 1.2-2: Transfer characteristics of a memory device after application of ±8V voltage pulsed of various durations.

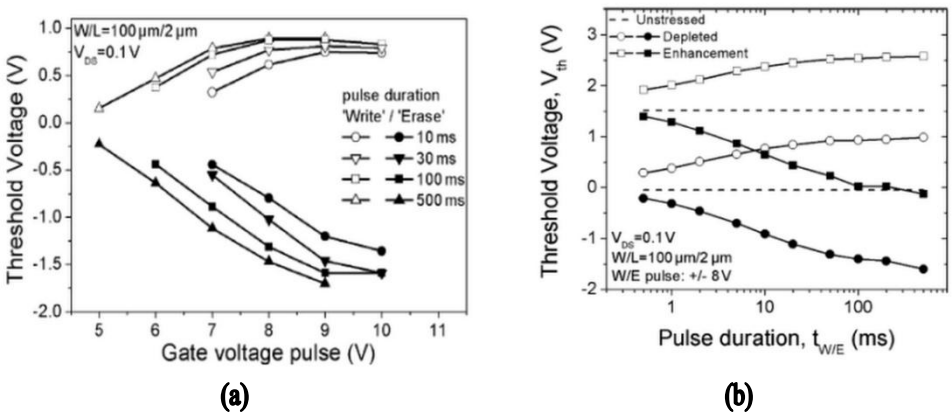


Figure 1.2-3: Memory windows of NVM devices after application of different voltage pulses (a) dependence of the memory window on the amplitude of the pulse and (b) dependence on the pulse duration.

Frequently, however, a different method is used in order to estimate the memory window. The method involves performing capacitance-voltage (C-V) measurements on the NVM device. The charge storage results in a hysteresis in the C-V characteristic as the voltage sweeps from inversion to accumulation and back to

inversion (fig. 1.2-4). The width of the hysteresis in the characteristic is a measure of the memory window, and is, of course, dependent on the range of the voltage sweep. This is more thoroughly discussed in following chapters.

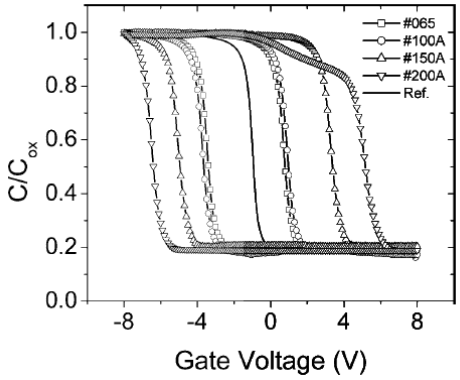


Figure 1.2-4: C-V characteristics of a NVM cell.

The advantages of this method is that it is simple and gives quick results even though sometimes it is possible to overestimate the memory window. Most importantly, it does not require a four terminal device: two terminal MOS-C-type structures suffice for the measurement and such a structure can be much more easily realized.

1.2.2 Retention

The most important requirement for a nonvolatile memory cell is the ability to preserve the stored data for long periods of time without the need for external power supply. That ability is called *charge* or *data retention* and the key aspect of a NVM device is its *retention time*. The retention time can be defined as the time between data storage and the first erroneous readout. For true non-volatility at least ten years of charge retention are required. In any other case the NVM cell is deemed unreliable.

Each non-volatile storage technology employs a particular storage mechanism and properties associated with that mechanism and its implementation format will determine the retention characteristics of the device. For FG-MOSFETs as discussed above, the storage mechanism is to hold charge on its floating gate.

The measurement procedure of the retention time comprises of three basic steps. First, the memory device under test is selected and programmed or erased. Next, the device is stored in a controlled temperature place (e.g. a furnace) with no bias to its electrodes. After a pre-determined time has elapsed, the device is taken out of the storage place and the threshold voltage is measured. This procedure is repeated until the total storage (or waiting, or bake) time exceeds 10^4 s. Semi-logarithmic plots of V_{th} versus the waiting time (t) at different temperatures are presented in figure 1.2-5. V_{th} then shows a dependence on time that can be described by the equation: $V_{th} = K_1 - K_2 \log(t)$, where K_1 and K_2 are constants. Of course, any change in V_{th} is proportional to the charge loss:

$$\Delta V_{th} = KQ(t).$$

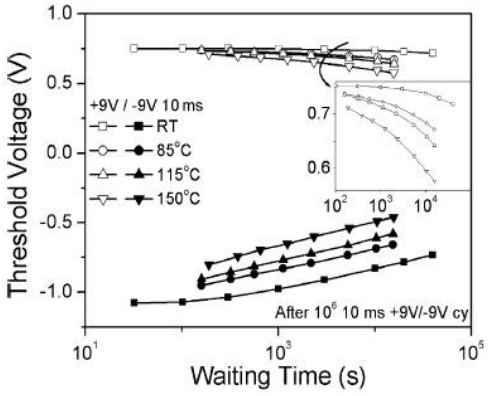


Figure 1.2-5: Data retention characteristics of NVM devices at room temperature (RT), 85°, 115°, and 150° C. Samples have also been stressed with 10^6 P/E cycles (10 ms +9/-9V).

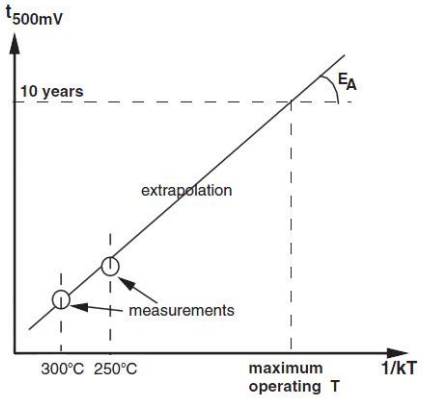


Figure 1.2-6: Arrhenius diagram of a floating gate device.

Data retention is sensitive to many parameters. The most important of them are (a) the device history, (b) the ambient temperature (raise in temperature results in the increase of the K_2 parameter in the relationship above), and (c) the thickness and the quality of the injection and control oxide layers. Dependence on temperature is particularly significant because for higher temperatures the charge loss proceeds with faster rates. That is the reason why in practice, temperature accelerated tests are used (typically 250° to 300° C) that enhance the charge loss mechanism.

The bake time corresponding to a given threshold voltage shift is displayed as a function of the inverse of the absolute temperature. Figure 1.2-6 shows such a log (time) versus 1/T result, which is known as an Arrhenius plot. From practical experience, it is known that this relationship describes the temperature dependence of the charge loss mechanism quite well. This not only allows data taken at high temperatures to be extrapolated to predict what will happen at room or operating temperatures, but also helps extract the activation energy E_A of the dominant charge loss mechanism. And once the activation energy is known, one can then identify the charge loss mechanism itself. Mathematically, if the abscissa is 1/kT (with k Boltzmann's constant), the slope of this curve will be equal to the activation energy:

$$t_{\Delta V_t} \sim \exp(E_A / kT).$$

It should be noted that temperature accelerating testing has its drawbacks and sometimes yields retention times that can be misleading because the operation temperatures of a NVM cell are in reality different. In addition, the memory window of a device is affected by the readout operations and such tests do not take that into account.

1.2.3 Endurance

The processes employed to write and/or read cells will result in stresses that eventually degrade the properties of the memory or disturb the contents of the memory. *Endurance* is the term used to describe the ability of a device to withstand these stresses, and it is quantified as a minimum number of erase-write cycles or write-read cycles that the chip can be expected to survive. For quite a number of years the industry has used 100,000 cycles as the minimum competitive endurance requirement. High performance NVMs must even meet the requirement for one million program/erase cycles. The write/erase cycling endurance (or endurance for short) of a floating-gate device is the major reliability characteristic because it describes the gradual degradation of the cell and, therefore, its lifetime in terms of the number of write/erase cycles that can be applied before failure.

For a single cell, endurance is usually monitored by measuring the threshold voltage window as a function of the number of applied write/erase cycles (fig 1.2-7). The lower V_{th} should not increase above a certain value in order to guarantee sufficient readout current. An important remark is that the cell-level endurance tests always underestimate the lifetime of the cell because a large number of cycles are applied to the cell in a very short period in order to reduce test time. In practice, there will be more time in between cycles, and the degradation will be relaxed because of recovery effects.

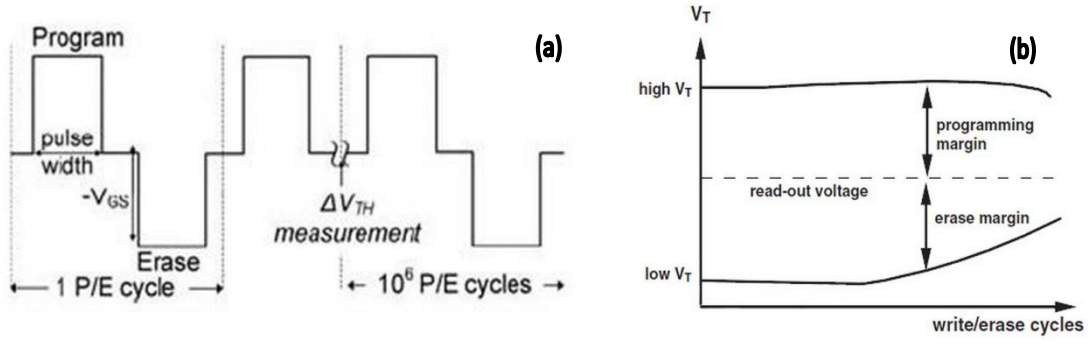


Figure1.2-7: (a) Schematic representation of the voltage pulse sequence which is used in endurance tests. (b) Endurance characteristics of a floating-gate device.

1.2.4 Other Performance Aspects of NVM Cells

As we have mentioned earlier, the memory devices are programmed, erased and read by applying voltage pulses of specific amplitudes and durations. The duration of

those pulses is a measure of the *speed* of the NVM devices and for fast memory cells short duration pulses are desired. It is also desirable that the device operates with as low voltages as possible so that the *power consumption* is kept to minimum values.

Another very important issue with non-volatile memories is the size of the device. Clearly, it is preferable to have memory cells that are as small as possible because the smaller the devices, the more of them can be integrated on a single microchip. Unfortunately, as the dimensions of the device shrink, the FG-MOSFET is faced with several challenges. The main problem is that as the injection oxide gets thinner it becomes easier for the stored charge to tunnel from the floating gate back to the substrate and the result is that retention characteristics deteriorate dramatically. Additionally, devices for logic applications that do not face such problems move a lot faster to shorter technology nodes. The result is a technology mismatch between memory devices and devices for logic applications and that makes difficult the implementation of embedded memory units. To overcome these difficulties alternative ideas for memory cells have been proposed one of which is the nanoparticle non-volatile memory cell which is the subject of this thesis.

REFERENCES

[1] Brewer, J. E. (ed.), and Gill M. (ed.), (2008) *Nonvolatile Memory Technologies with Emphasis on Flash: A Comprehensive Guide to Understanding and Using NVM Devices*. New Jersey: John Wiley and Sons.

[2] Khriachtchev, L. (ed.) (2009) *Silicon Nanophotonics: Basic Principles, Present Status and Perspectives*. Singapore: Pan Stanford Publishing Pte Ltd.

[3] Dimitrakis, P., (2006) *Nanoparticle Electronics*. Ph.D., National Technical University of Athens.

[4] Dimitrakis, P., Kapetanakis, E., Normand, P., Skarlatos, D., Tsoukalas, D., Beltsios, K., Clavarie, A., Benassayag, G., Bonafos, C., Chassaig, C., Carrada, M., and Soncini, V. (2003) *MOS Memory Structures by Very-Low-Energy Implanted Si in Thin SiO_2.* Materials Science and Engineering, B101, pp. 14-18.

[5] Dimitrakis, P., Kapetanakis, E., Tsoukalas, D., Skarlatos, Bonafos, C., Benassayag, G., Clavarie, A., Perego, M., Fanciulli, M., Soncini, V., Sotgiu, R., Agarwal, A., Ameen, M., Sohl, C., and Normand, P. (2004) *Silicon Nanocrystal Memory Devices Obtained by Ultra-Low Energy Ion-Beam Synthesis*. Solid State Electronics, 48, pp. 1511-1517.

[6] Ortiz-Conde, A., Garcia, F. J., Lion J. J., Cerdeira, A., Estrada, M., and Yue, Y. (2002) *A Review of Recent MOSFET Threshold Voltage Extraction Methods. Microelectronics Liability*, 42, pp. 583-596.

[7] Ghetti, A., Bortesi, L., Vendrame, L. (2005) *3D Simulation Study of Gate Coupling and Gate Cross-Interference in Advanced Floating Gate Non-Volatile Memories*. Solid State Electronics, 49, pp. 1805-1812.

CHAPTER 2

Theoretical Background

The floating-gate MOS device, which is, as discussed in the previous chapter, the fundamental cell for non-volatile memory, is based on the operation of the MOS-Field-Effect-Transistor, the most common of all the electronic devices in integrated circuits. So do a lot of other types of non-volatile memories including nanocrystal memory cells. The MOSFET, in turn, takes advantage of the properties of the Metal-Oxide-Semiconductor structure. For that reason, it becomes necessary that we examine some of the operation principles of the MOS junction, and the MOSFET, in this chapter. The focus is, of course, on those properties relevant to the behavior of memory devices and for a full review the reader is prompted to go through the related literature.

So far, only the general concepts of non-volatile memory have been introduced. Yet, the need to optimize memory devices requires a full understanding of their function as well as the challenges they face. To do that, this chapter next takes a closer look at the FG-MOSFET and the physics involved in the mechanisms that make its operation possible. It also looks into their problematic areas and proposes possible solutions to those problems.

To overcome some of the difficulties that come with the FG-MOSFET current research is looking into a number of alternatives, each one of which has both benefits and weaknesses. One of the most prominent candidates to replace the floating-gate transistor in the future is the nanocrystal memory cell. The idea is that the floating-gate layer within the dielectric of the transistor is replaced with nanoparticles of a semiconducting material (such as silicon, germanium or, in our case, gallium nitride). The advantages of such a structure over the simple floating-gate device are many, the most important of which is, maybe, that it is not limited by the thickness of the

interpoly dielectric. Nanoparticle memory cells was first proposed by Tiwari in 1996; since then significant progress has been made and recent tests to evaluate their effectiveness have produced very promising results. Nanocrystals (or quantum dots) are briefly discussed in order to gain a perspective of their properties, after an analysis of the operation of a nanoparticle memory cell, as it is understood, is given. Lastly, we argue about the reasons that might make gallium nitride a suitable material for the nanocrystals in memory applications.

2.1 THE MOS CAPACITOR AND THE MOSFET

2.1.1 Electronics of the Metal-Oxide-Semiconductor System

The Metal-Oxide-Silicon structure takes advantage of the fact that silicon, a semiconductor, turns into a very effective insulator when oxidized. In reality, the MOS system is a simpler, yet still more specific case of the Metal-Insulator-Oxide system, which means that any appropriate insulator can be used in place of the silicon dioxide with similar effects.

The device is presented in figure 2.1-1a. We assume acceptor-doped silicon throughout this paragraph for our analysis, but mention is given to the n-type silicon substrate when it is deemed necessary. On top of the silicon lies a layer of silicon dioxide (SiO_2) of thickness t_{ox}. The metal, usually aluminium, also called the *gate*, provides an electrode at which the voltage can be fixed. It should be noted here, that a lot of the times, especially in recent devices, polysilicon is used instead of the metal without this affecting the operation of the device. The resulting structure is basically a simple capacitor with its top electrode on the metal-oxide interface, its bottom electrode somewhere in the bulk silicon and a dielectric in between. The energy band diagram of this system in equilibrium is given in figure 2.1-1b. X_{Si} and X_{ox} are the electron affinities of the silicon and the oxide accordingly and Φ_M is the metal work function. The metal and the semiconductor form then two plates of a charged capacitor and this capacitor is charged to a voltage that corresponds to the difference between the metal and the silicon work function. However, it should be noted that frequently, especially in practical applications this difference is ideally considered to be zero.

Flat-Band Condition

Now let us consider the case when a bias voltage is applied between the gate and the silicon, and let us also assume that the applied voltage is set to a value that exactly compensates the difference in the work functions of the metal and the semiconductor:

$$V_{FB} = \Phi_M - \Phi_S .$$

This voltage is called *flat-band voltage* and the MOS system is then said to be in *flat-band condition*, which as it has been mentioned is sometimes ideally considered to occur at 0V. The name flat-band means that the band diagram of the system is then 'flattened' (fig. 2.1-2).

Accumulation

Subsequently, if the silicon is held at ground and the voltage applied to the metal is kept negative but increased in magnitude, holes will begin to accumulate on the oxide-semiconductor interface so that the surface will eventually have a greater density of holes than the acceptor density. This condition is called *accumulation* and the region at the surface containing the increased hole population is known as the *accumulation layer*.

Depletion

In the case that a positive voltage is applied between the metal and the silicon negative charge will be stored at the silicon surface and positive on the metal. What this means is that holes will be drawn by the field away from the semiconductor-metal interface leaving positively charged acceptors close to the surface. The area where holes have left their acceptors exposed is called *depletion region* or *depletion layer*

and the system is said to be in conditions of *surface depletion*. The width of the depletion layer is x_d and Q_d is the total charge of the depletion layer:

$$Q_d = -qN_a x_d ,$$

where N_a is the acceptor density of the silicon substrate, and q the electron charge in coulombs. The width of the depletion layer x_d increases as we move to higher voltages.

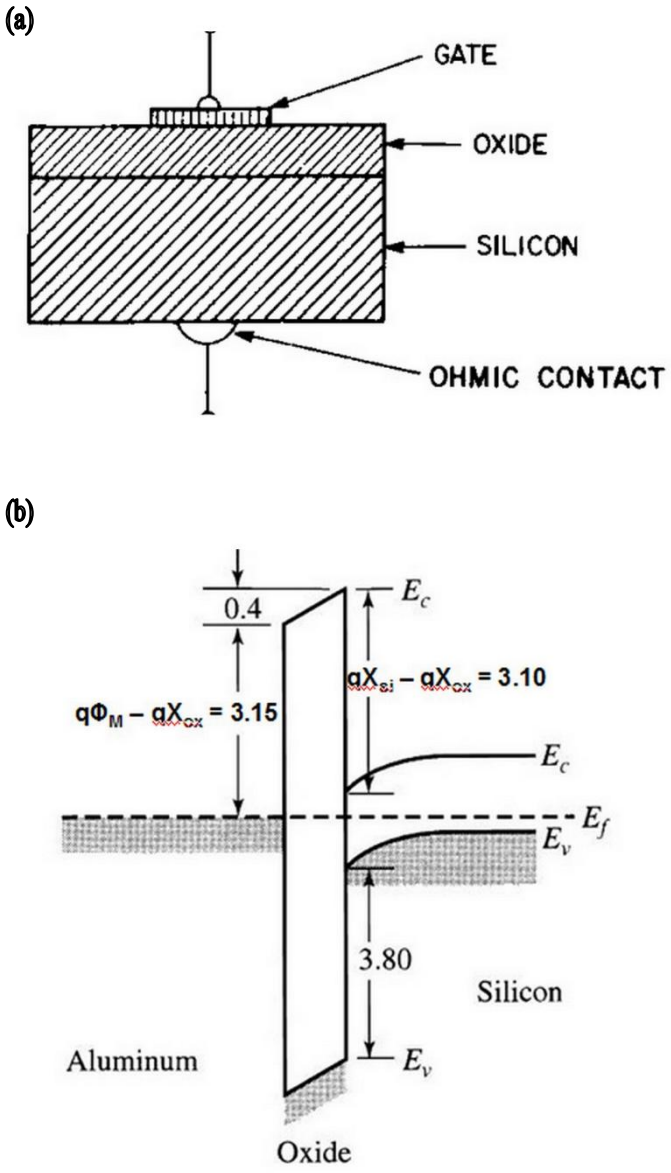


Figure 2.1-1: (a) Structure of the MOS Capacitor. (b) Band diagram of the MOS capacitor in equilibrium.

Inversion

If the voltage applied to the metal is increased further, the field at the surface of the silicon will also increase and generation of carriers will eventually exceed recombination where majority carriers have been depleted. Then the width of the depletion region will have reached a maximum which can be found to be equal to:

$$x_{d\max} = \sqrt{\frac{4\varepsilon_S |\phi_p|}{qN_a}}, \qquad \text{where } \phi_p = \frac{E_i - E_f}{q}.$$

E_f is the Fermi level in the bulk of the semiconductor, E_i is the intrinsic Fermi level, and ε_S is the dielectric constant of the semiconductor.

The electron-hole pairs generated are separated by the field, the holes being swept into the bulk and the electrons moving to the oxide-silicon interface to form the *inversion layer*, so called because the surface contains more electrons than holes even though the silicon is doped with acceptor impurities. The inversion layer is the channel used in MOS Field Effect transistor to conduct current between the source and the drain of the transistor. The voltage at which the inversion layer occurs is sometimes called *threshold voltage* (V_{th} or V_T) and will be discussed in more detail in a following paragraph. As soon as the inversion layer is formed, the system is said to go into *inversion*. The inversion layer comprises then of negative charge (Q_n) which adds to the total charge on the semiconductor surface so that the total charge is the sum of the negative charges of the depletion and the inversion layers: $Q = Q_d + Q_n$.

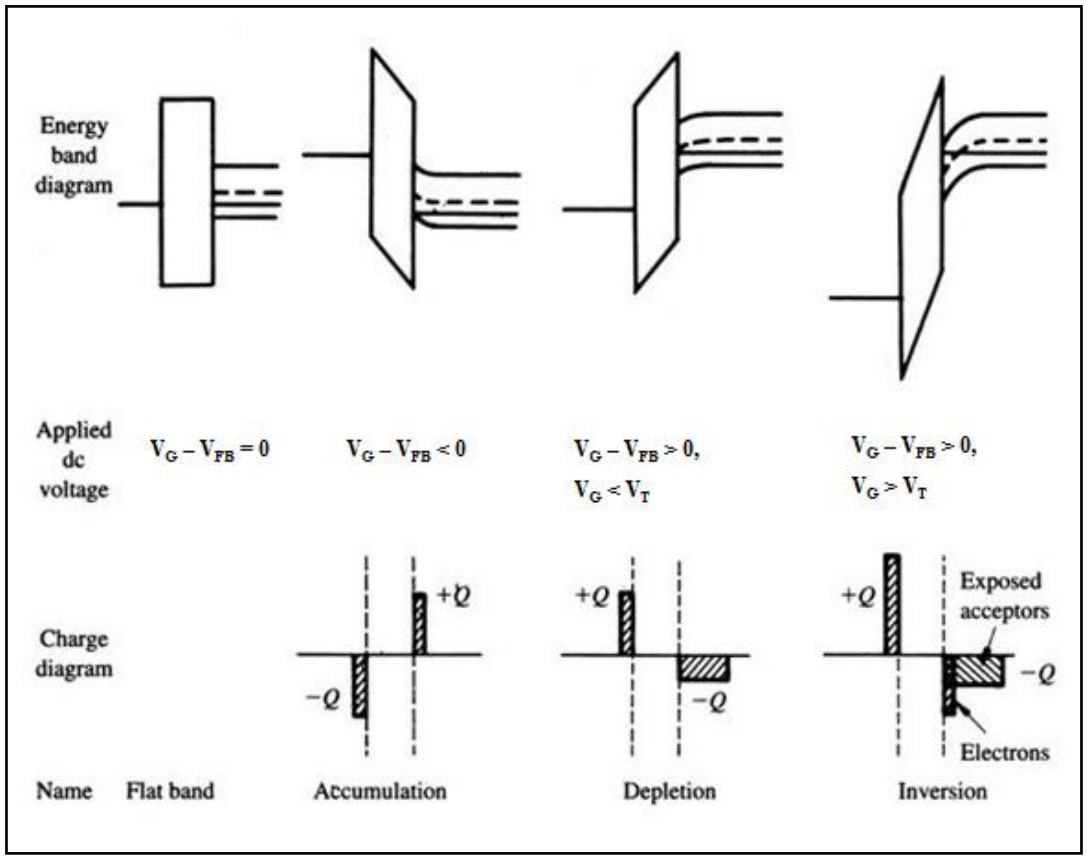


Figure 2.1-2: Band diagrams and charge distributions for the MOS capacitor.

The band diagrams for all of the above cases, as well as the charge densities, are collectively presented in figure 2.1-2. The above discussion can easily be made for a donor-doped silicon substrate, if we keep in mind that in that case, the electrons will be the majority carriers and the holes the minority carriers; then, the charges discussed change their sign. Additionally, a MOS capacitor on n-type silicon, will move from

accumulation to inversion as the gate voltage becomes more and more *negative* instead of more positive, as was the case for p-type silicon.

2.1.2 MOS Capacitance

As it has been already stated, the MOS system is, in fact, a capacitor. What makes it so important is that its capacitance does not remain constant; it rather depends on the voltage bias that is applied at the gate. To study how the bias voltage affects the capacitance of the system, we assume a dc signal applied to the gate, and a small ac voltage superposed on that dc bias, which will help measure the system impedance.

Let us first consider that the MOS system described in the previous paragraph is biased with a steady voltage that causes the silicon surface to be accumulated (for a p-type silicon sample, it has been pointed out this corresponds to a negative applied voltage). The excess holes at the surface are pulled very close to the oxide, and therefore the capacitance measured will be that of the oxide itself:

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}},$$

where ε_{ox} is the oxide dielectric constant, t_{ox} the oxide thickness, and C is capacitance per unit area.

As the surface area becomes less accumulated, and the system approaches flat-band conditions, the capacitance will start to decrease. When the gate bias finally reaches the flat-band voltage the capacitance has be found to be given by the following expression:

$$C_{FB} = \frac{1}{1/C_{ox} + L_D/\varepsilon_s},$$

where $L_D = \sqrt{\frac{\varepsilon_s kT}{q^2 N_a}}$ is the Debye length, T the temperature in K and k the Boltzmann constant.

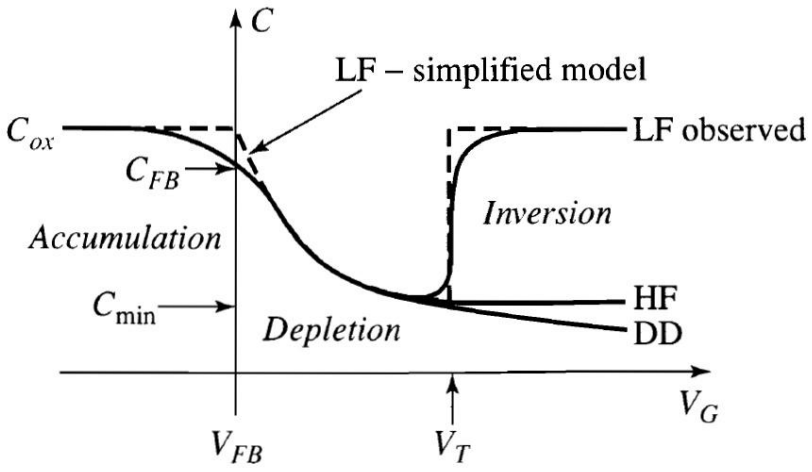


Figure 2.1-3: General behaviour of C-V curves of an ideal MOS system under different dc bias and ac small signal conditions.

When the gate voltage becomes more positive than the flat-band voltage, holes are repelled from the surface of the silicon and the system is in depletion. Under this condition, relatively straightforward electrostatic analysis shows that the overall capacitance C corresponds to the capacitance obtained by a series connection of the oxide capacitance and the capacitance C_s across the surface depletion region:

$$C_{depletion} = \frac{1}{1/C_{ox} + 1/C_s} = \frac{1}{1/C_{ox} + x_d/\varepsilon_s},$$

where x_d is the width of the surface depletion layer, which depends upon gate bias as well as the doping and oxide properties. From the above equation it is obvious that the capacitance of the system decreases as the depletion region widens.

When the gate bias is increased sufficiently to invert the surface, a new feature must be considered to describe the MOS capacitance behaviour. The inversion layer at the MOS surface results from the generation of minority carriers; hence, the population of the inversion layer can change only as fast as carriers can be generated within the depletion region near the surface. This limitation causes the measured capacitance to be a function of the frequency of the ac signal used to measure the small-signal capacitance of the system (fig. 2.1-3).

The simplest case arises when both the dc gate-bias voltage and the small-signal measuring voltage are changed very slowly so that the silicon can always approach equilibrium. In this case, the signal frequency is low enough so that the inversion-layer population can ‘follow’ it. The capacitance of the MOS system is just that associated with charge storage on either side of the oxide; its value is approximately C_{ox}. In this case the C-V characteristic of the system follows the curve that is marked LF (low frequency) in figure 2.1-3: going from C_{ox} in the accumulation region of bias through a decreasing region as the surface traverses the depletion region and moving back up to C_{ox} when the surface becomes inverted.

When the ac measuring signal is changed rapidly while the dc bias voltage is varied slowly, the inversion layer cannot respond to the measuring signal. The capacitance then corresponds to the series combination of the oxide capacitance and the depletion-region capacitance, as was true under depletion bias. Since the depletion region reaches a maximum width x_{dmax} when the system goes into strong inversion, the measured capacitance approaches a value corresponding to the series connection of the oxide capacitance and the capacitance associated with the maximum depletion-region width. It remains constant at this value as the bias voltage is increased further (HF-high frequency curve in figure 2.1-3).

There is a final capacitance behaviour called *deep depletion* (DD on the C-V characteristic of figure 2.1-3). It corresponds to the experimental situation in which both the gate bias voltage V_G and the small-signal measuring voltage vary at a faster rate than can be accommodated by generation in the surface depletion-region. Since the inversion layer cannot form, the depletion region becomes wider than x_{dmax}. The capacitance in this mode is given by the series combination of the oxide capacitance and the capacitance of the depletion region, only in this case $x_d > x_{dmax}$ and C does not reach a minimum.

The equivalent circuits for all of the above cases are presented in figure 2.1-4.

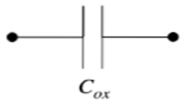
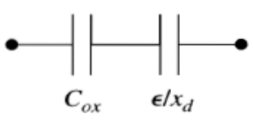
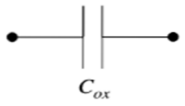
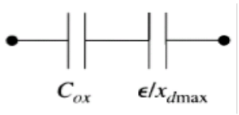
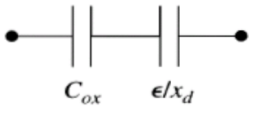
Accumulation	
Depletion	 $x_d < x_{dmax}$
Inversion – HF (high frequency bias conditions)	
Inversion – LF (low frequency bias conditions)	
Deep depletion	 $x_d > x_{dmax}$

Figure 2.1-4: Equivalent circuits for the overall capacitance of the MOS System under different bias conditions.

The Effect of the oxide and interface charge on the Flat-Band Voltage: So far we have tried to describe the behaviour of the MOS system which essentially depends on how the charge is distributed in the silicon and metal interface and how that distribution is affected by applying a voltage bias on the metal. We have considered the ideal case when no charge can exist within the oxide. However, very frequently and despite any efforts to eliminate it, several kinds of charges sometimes present themselves in the bulk of the dielectric or the surface and one has to consider their effect on the system behaviour, especially when one is interested in the MOS system as a platform for charge storage, as is the case for non-volatile memories.

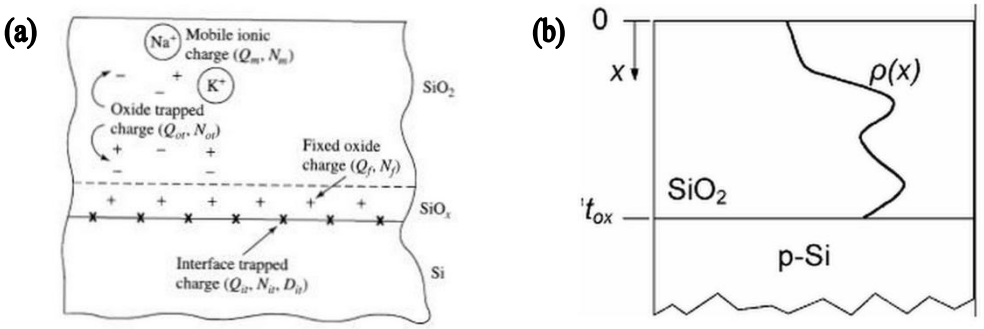


Figure 2.1-5: (a) Four categories of oxide charge in the MOS system (b) Charge density distribution in an oxide layer.

Oxide charge is the result of defects in the dielectric, it generally depends on how effective the process of the forming of the layer is performed and it usually is categorized in four groups (fig. 2.1-5a): Q_f the fixed interface charge density, Q_{ot} the oxide trapped-charge density, Q_{it} the interface trapped-charge density, and Q_m the mobile charge density.

The oxide charge causes the C-V curves of the MOS system to distort. The exact analysis is fairly complicated and different for each kind of charge. For example, fixed oxide charge causes the C-V characteristic to shift along the voltage sweep. In this paragraph, we only make a brief mention of how the oxide charge affects the flat-band voltage because it will prove useful for the analysis that follows in the next chapters; should one wish so, one can look into the related literature.

We have mentioned above that the flat-band voltage where there is no oxide charge is given by the expression:

$$V_{FB} = \Phi_M - \Phi_S = \Phi_{MS}.$$

When we take into consideration the effect of oxide charges, the above equation can be found to take the form:

$$V_{FB} = \Phi_{MS} - \frac{Q_f}{C_{ox}} - \gamma\frac{Q_m}{C_{ox}} - \gamma\frac{Q_{ot}}{C_{ox}} - \frac{Q_{it}}{C_{ox}}, \text{ where:}$$

$$\gamma = \frac{\int_0^{t_{ox}} \frac{x}{t_{ox}}\rho(x)dx}{\int_0^{t_{ox}} \rho(x)dx}.$$

$\rho(x)$ is the charge density distribution through the oxide layer (fig. 2.1-5b), and $Q = \frac{1}{t_{ox}}\int_0^{t_{ox}} x\rho(x)dx$.

If now we consider only the trapped charge in the oxide, with a random distribution ρ, then, we get:

$$V_{FB} = \Phi_{MS} - \gamma\frac{Q_{ot}}{C_{ox}},$$

and this results in a shift in the C-V characteristic, as well as in the flat-band voltage. This will be again discussed in the following paragraphs.

2.1.3 The MOS Field Effect Transistor and The Threshold Voltage

This paragraph briefly describes the operation of the MOS Field Effect Transistor. It does not attempt a lengthy analysis, nor does it contain a full mathematical description as such a thing would stray from the purpose of the present study; it rather focuses on a few characteristics of the device that will be useful for future reference.

The basic structure of an n-channel MOSFET is shown in figure 2.1-6a. It consists of the MOS structure of the previous paragraphs, with a surface inversion layer or *channel* extending between two diffused junctions. These two junctions are electrically disconnected unless there in an n-type inversion layer at the surface to provide a conduction channel between them. When the surface is inverted and a voltage is applied between the junctions, electrons can enter the channel at one junction, called the *source*, and leave at the other, called the *drain*. The density of free

charge in the channel is controlled by a field that extends from the gate electrode to the silicon through the insulating layer.

The characteristics of the device are presented in figure 2.1-6b. When a positive voltage V_G is applied to the gate relative to the substrate, mobile negative charges (electrons) gets attracted to Si-oxide interface. These induced electrons form the channel. For a given value of V_G, the current I_D increases with V_D, and finally saturates.

The reason it is called an 'n-channel' MOSFET is because the inversion layer (or the channel) consists of electrons. The same principles apply for a p-MOSFET, developed on n-type silicon, with a channel that consists of holes.

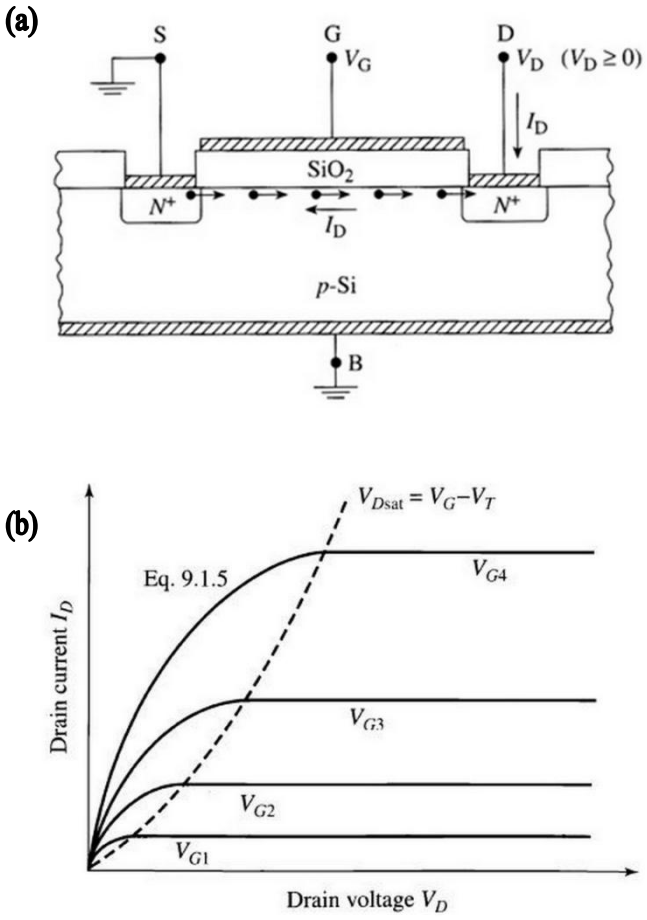


Figure 2.1-6: (a) Schematic representation of the n-channel MOSFET. (b) Typical I-V characteristics of a MOSFET device. $V_{G4} > V_{G3} > V_{G2} > V_{G1}$.

In order for the channel to form, however, and for current to be able to flow, the gate voltage must be higher than the *threshold voltage* V_T (or V_{th}) of the device, as we have mentioned before. A full electronic analysis yields the threshold voltage to equal:

$$V_T = V_{FB} + V_C + 2|\phi_p| + \frac{1}{C_{ox}}\sqrt{2\varepsilon_S q N_a \left(2|\phi_p| + V_C + V_B\right)}.$$

V_{FB} is the flat-band voltage of the MOS system, V_C is the voltage applied to the transistor source, and V_B the voltage applied to the body (substrate).

2.2 PHYSICS OF THE FG-MOSFET

2.2.1 Charge Storage

The best place to begin to understand the physics of the FG-MOSFET is to consider the energy levels involved. Figure 2.2-1b shows the band structure for a simple floating gate device (fig. 2.2-1a) where the silicon substrate is shown on the left. This band diagram might depict a two-terminal single memory cell (like to the MOS capacitor) or can be part of a more complex device with three or more terminals. The n-type control gate is shown instead of a metal on the right, and an n-type polysilicon floating gate is in the middle, sandwiched between two silicon dioxide layers. The floating gate, embedded within insulators, is isolated from the exit or entry of charge by the high-energy barrier between the conduction band in the polysilicon and the conduction bands in the top and bottom SiO_2 layers. These barriers, much greater than the thermal energy, provide non-volatile retention of the charge. In order to change the amount of charge stored on the floating gate it is necessary to change the potential of the floating gate relative to the potential on the opposite side of either SiO_2 layer until some conduction mechanism is invoked that can overcome or tunnel through the barrier.

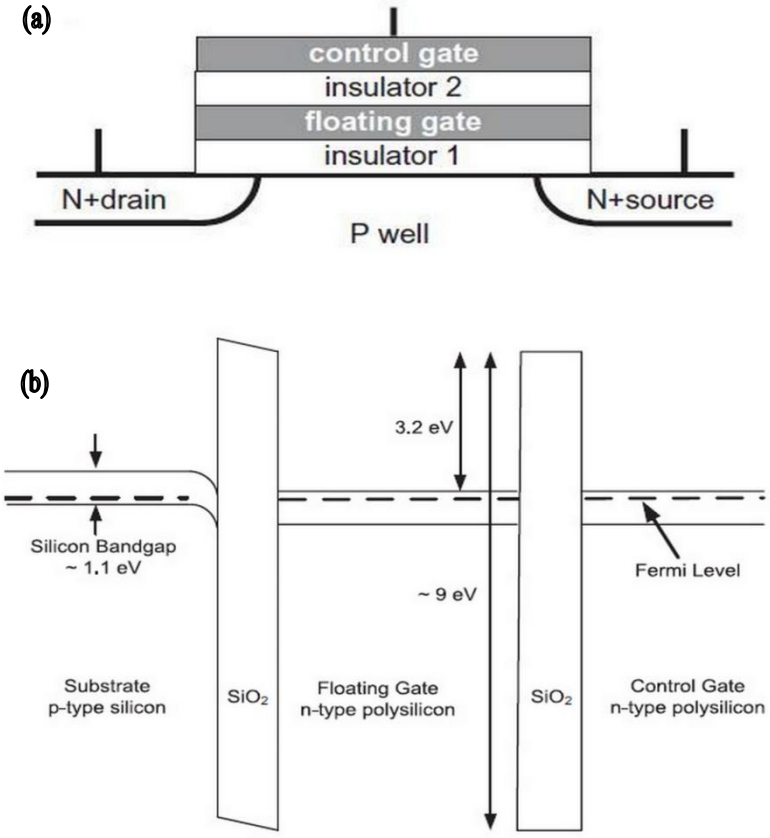


Figure 2.2-1: (a) The FG-MOSFET. (b) Energy band diagram for a typical floating-gate structure.

Different strategies may be selected to overcome the energy barriers. Two conduction mechanisms in common use are channel hot-electron injection (CHEI) and

Fowler-Nordheim (FN) tunnelling. Hot-carrier injection may be used to add electrons to a floating gate (i.e. programming). FN tunnelling may be used to remove or add electrons to a floating gate (i.e. erase or program). Other mechanisms (enhanced tunnelling, substrate hot electron injection, etc.) can be used but, since they are not as common, we will not be concerned with such procedures.

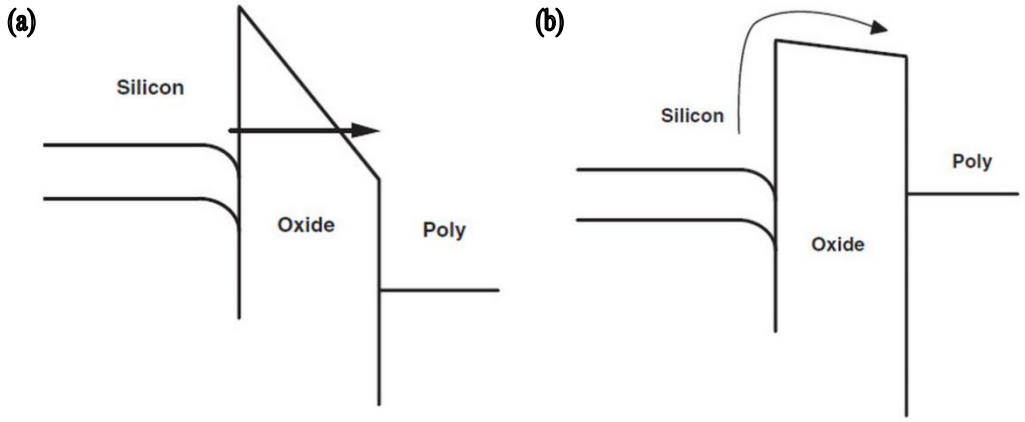


Figure 2.2-2: (a) Energy band representation of Fowler-Nordheim tunnelling through thin oxides; electrons in the silicon conduction band tunnel through the triangular energy barrier. (b) Energy band representation of channel hot-electron injection (CHEI) in an n-channel MOS transistor.

Fowler-Nordheim Tunnelling: FN tunnelling is in fact a field-assisted electron tunnelling mechanism. When a large voltage is applied across a polysilicon-SiO_2-silicon structure, its band structure has a shape as indicated in figure 2.2-2a. Due to the high electric field, the electrons in the silicon conduction band see a triangular energy barrier of which the width is dependent on the applied field. The height of the barrier is determined by the electrode material and the band structure of SiO_2.

At sufficiently high fields, the width of the barrier becomes sufficiently small for electron to tunnel through the barrier from the silicon conduction band into the oxide into the oxide conduction band. A simple expression for the FN current is given:

$$J = \alpha E_{inj}^2 \exp\left(\frac{-E_c}{E_{inj}}\right)$$

with:

$$\alpha = \frac{q^3}{8\pi h \phi_b}\frac{m}{m^*}$$

$$E_c = \frac{4\sqrt{2m^*}\phi_b^{3/2}}{3\eta q}$$

where h is Planck's constant, ϕ_b is the energy barrier at the injecting interface (3.2eV for Si-SiO_2), E_{inj} is the electric field at the injecting interface, q the charge of a single electron, m the free electron mass, m* the effective ass of an electron on the bandgap of SiO_2 and $\eta = h/2\pi$.

More accurate expressions for the FN current density have been expressed and are available in literature, but in general, the current shows an exponential dependence on the applied field.

FN tunnelling is very common in non-volatile memories for programming and erasing. It provides low currents and, hence, low power operation. It has also shown good endurance and current controllability for the devices that employ it. It has, however, some disadvantages such as the large voltages it requires for tunnelling and the demand for good quality of the tunnel oxide.

Channel Hot-Electron Injection: At sufficient drain bias (fig. 2.2-3), the minority carriers that flow in the channel of a MOS transistor are heated by the large electric fields seen at the drain side of the channel and their energy distribution is shifted higher. This phenomenon gives rise to impact ionization at the drain, by which both minority and majority carriers are generated. The highly energetic majority carriers are normally collected at the substrate contact and for the so-called substrate current. The minority carriers, on the other hand, are collected at the drain. A second consequence of carrier heating occurs when some of the minority carriers gain enough energy to allow them to surmount the SiO_2 energy barrier. If the oxide field favors injection, these carriers will be injected over the barrier into the gate insulator and give rise to the so-called channel hot-carrier injection gate current. The corresponding energy band diagram is schematically represented for the case of an n-channel transistor in figure 2.2-2b. The electrons gain enough energy from the lateral field at the drain side of the channel to surmount the energy barrier between the silicon and the oxide.

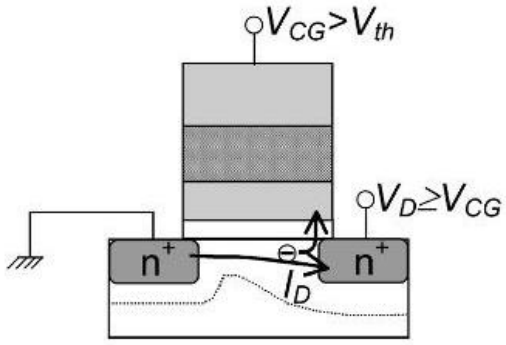


Figure 2.2-3: Bias conditions for Channel Hot-Electron Injection mechanism for the storing of charge on the floating gate.

An important difference with respect to the FN tunnelling mechanism is that CHEI can only bring electrons onto the floating gate; it cannot be used to remove them again. What is more, in contrast to the FN tunnelling case, no closed-form analytical expression exists for the channel hot-electron injection current. There are several models (e.g. the lucky electron models, the effective electron models, etc.) but all of those models are merely qualitative. It can be shown, however, that the gate current increases with thinner gate oxides, shallower junctions, smaller effective channel lengths, and higher substrate doping levels.

From the point of view of the memory cell operation, it can be concluded that the CHEI mechanism provides a fast programming method, however, requiring a

considerable channel current resulting in high power consumption. On the other hand, because of the small oxide field, cells relying on CHEI for programming are less susceptible to oxide breakdown and stress-induced leakage currents.

2.2.2 FG-MOS Capacitance

As it has been discussed in paragraph 2.1.2, charges that find themselves trapped in the oxide of a MOS junction with a density distribution ρ(x), cause the flat-band voltage of the system to shift, according to the expression:

$$V_{FB} = \Phi_{MS} - \gamma \frac{Q_{ot}}{C_{ox}}, \text{ with}$$

$$\gamma = \frac{\int_0^{t_{ox}} \frac{x}{t_{ox}} \rho(x) dx}{\int_0^{t_{ox}} \rho(x) dx}, \text{ and}$$

$$Q_{ot} = \int_0^{t_{ox}} \rho(x) dx .$$

If no charges exist within the oxide then the flat-band voltage equals Φ_{MS}, and so it becomes obvious that the shift has an absolute value of

$$\left|\Delta V_{FB}\right| = \left|\gamma \frac{Q_{ot}}{C_{ox}}\right|,$$

and also, that **the flat-band voltage of the system shifts to more negative values (to the left) if the charge in question is positive and to more positive values (to the right) if the charge is negative**.

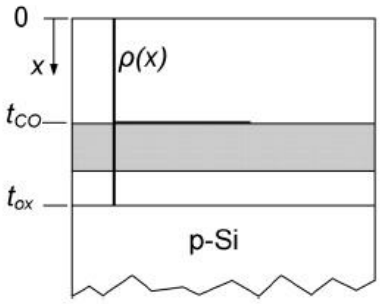


Figure 2.2-4: Schematic representation of the charge density distribution of the stored charge in a floating gate memory cell.

Let us now assume a floating gate, in the middle of the oxide of a MOS system, whose function is to store charge, and let us also ideally assume that no other charges exist in the oxide, except those held in the floating gate. When the floating gate is not charged, no shift manifests in the flat-band voltage of the system: it equals Φ_{MS}, as it is argued above. However, when the floating gate is in a state when it holds stored charge, we can mathematically describe that charge as having a distribution of (fig. 2.2-4):

$$\rho(x) = \begin{cases} \rho_0, & x = t_{co} \\ 0, & x \neq t_{co} \end{cases}.$$

This yields $\gamma = \frac{t_{co}}{t_{ox}}$ and finally:

$$|\Delta V_{FB}| = \frac{t_{co}}{t_{ox}} \frac{|Q_{ot}|}{C_{ox}} = \frac{t_{co}|Q_{ot}|}{\varepsilon_{co}}.$$

Following the discussion above, if the charge that is held in the floating gate comprises of electrons then the shift of the flat-band voltage will be to higher values, whereas if the holes are stored in the floating gate, the shift of the flat-band voltage will be to the left. The shift in the flat-band voltage results in a shift in the C-V characteristic of the device and what this means is that when one attempts to perform a capacitance versus voltage measurement, a hysteresis will appear in the curve, as the sample is swept from depletion to accumulation and back to depletion.

Figure 2.2-5a shows such a measurement on a floating gate device, where the voltage is swept from a positive bias V_{bias1} to a negative bias V_{bias2} and then back to V_{bias1}. The dashed curve represents the state of the cell with no stored charge and the flat-band voltage that corresponds to the C_{FB} is in this case $V_{FB}{}^0=\Phi_{MS}$. The continuous curve shows storage of electrons in the floating gate and gives V_{FB} as described in the equations above. The shift is, therefore:

$$|\Delta V_{FB}| = |V_{FB} - V_{FB}^0|.$$

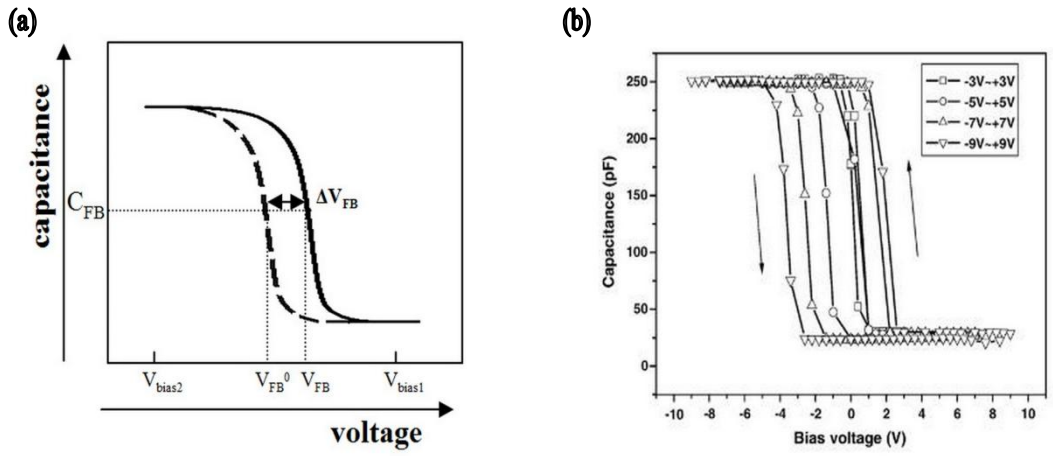


Figure 2.2-5: (a) Hysteresis in the C-V characteristic of a floating gate memory device. The dashed curve corresponds to the state without any stored charge on the floating gate. (b) Read C-V tests performed on a memory device.

When the positive voltage V_{bias1} is applied on the gate of the FG-MOS system, electrons are injected into the floating gate from the substrate. The result is that the flat-band voltage, and with it the entire C-V curve, will shift to the right. When then the negative V_{bias2} is applied, the electrons that were trapped in the floating gate will tunnel back to the substrate, so that when the voltage is swept back to positive values, the C-V curve is left where it should be without any shift.

A similar case can be made when the injected charge comprises of holes, only this time the shift will be to the left. It should be noted, however, that both types of charge can be stored in a floating gate. For example, in the case described above, when the negative bias is applied, it could have the effect of both electrons escaping into the substrate and holes being injected in the floating gate in the same time. The result would be that the C-V curve will shift to even lower values. All these cases are shown in figure 2.2-6.

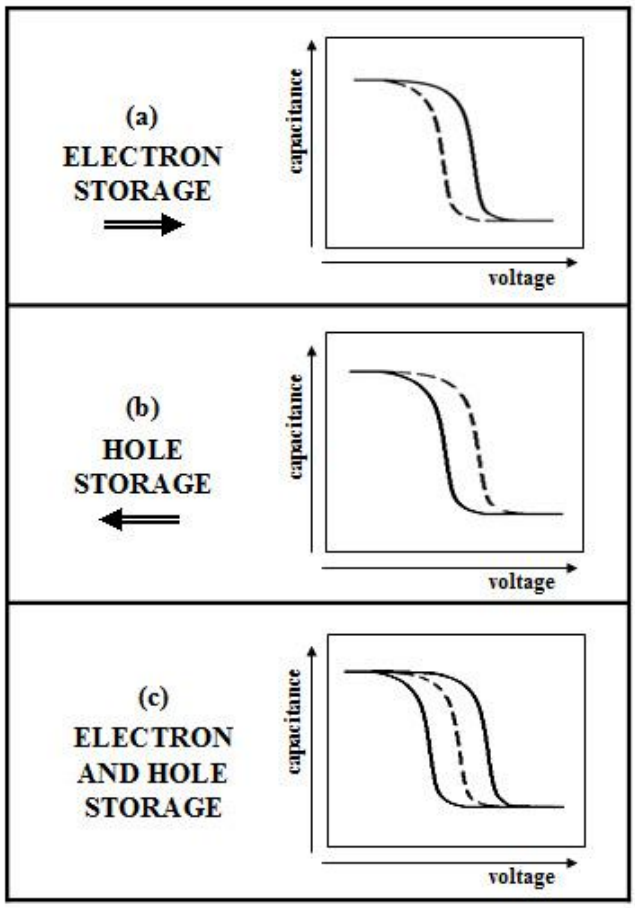


Figure 2.2-6: Shift of the C-V curves for different types of charge stored in the floating gate. The dashed curve corresponds to the state without any stored charge on the floating gate.

We remind here that the operation of a memory cell is based on the switching of the threshold voltage of a MOSFET between a high and a low value, each one of which corresponds to either the write state (for the higher V_T) or the erase state (for the lower V_T); the difference between these voltages is called the memory window and is a critical characteristic of a non-volatile memory device (see chapter. 1). The reason the hysteresis of the C-V curve is so important is that the threshold voltage of a MOSFET is given, as was mentioned in paragraph 2.1.3, by the expression:

$$V_T = V_{FB} + V_C + 2|\phi_p| + \frac{1}{C_{ox}}\sqrt{2\varepsilon_S q N_a \left(2|\phi_p| + V_C + V_B\right)}.$$

Because the flat-band voltage is the only variable in this expression that is affected by the state of a memory cell, it can be concluded for the memory window of such a device that:

$$MemoryWindow = \Delta V_T = \Delta V_{FB}$$

and this allows for the memory window to be extracted by performing a capacitance-voltage measurement on an FG-MOS structure, which is an easier method than extracting the threshold voltage from the current-voltage characteristics of the device.

It should be pointed out, however, that in real experimental measurements, one cannot always be certain of what is the curve that corresponds to no charge on the floating gate, and that makes it difficult to determine the sign of the shift and consequently the type of carriers stored in the floating gate, and one may have to make assumptions or additional measurements (fig. 2.2-5b).

It should also be taken into account that, in such tests, the amplitude of the shifts depends on the limits of the voltage sweep because they affect the total charge that will be stored. What is more, in practical applications where a voltage pulse is used to program a non-volatile memory cell, the total charge also depends on the duration of the pulse and additional tests should be performed should that dependence need determining.

2.2.3 Limits of the FG-MOSFET

An NVM cell would ideally possess all of the desirable features that would make its function efficient. These include fast speeds for the write, erase and read operations, infinite endurance, low power consumption and non-volatility (with a minimum ten-year retention).

The FG-MOSFET, which has been the most common among NVM devices, despite its virtues, faces several limits concerning these features. Most FG-MOSFETs require high programming voltages and this also limits the further scaling of the cells. Typical speeds are of the order of 10μs to 2ms which is rather slow, especially compared to the speeds of volatile memory devices. Their endurance is limited to 10^5 to 10^6 write/erase cycles by damage to the tunnel oxides due to repeated charge injection through them.

Many of these weaknesses of the FG-MOSFET are due to the fact that the tunnel oxide cannot be thinner than a certain limit. When the tunnel dielectric is too thin, charge tunnels or leaks off the floating gate through the tunnel dielectric and 10-year data retention cannot be met. Despite years of research and development, the minimum tunnel dielectric thickness for floating - gate memory cells using SiO_2 as the tunnel dielectric continues to be limited to approximately 6-7 nm. This limitation is due to Fowler-Nordheim tunnelling current or due to oxide defects created after many write/erase cycles.

In order to overcome these performance boundaries of NVMs it becomes essential for the tunnel oxide to drop to smaller thicknesses. However, this cannot be done at the expense of the retention characteristics of a memory cell and the 10-year limit is not negotiable. Another reason why it is quite difficult to reduce the size of NVM cells is the fact that as NVM cells are organized in large matrix of cells closely packed. As the FG devices are packed more closely to each other, each floating gate becomes coupled not only to the control gate, or other parts of the device itself but there also appears coupling with other nearby cells (fig. 2.2-7). This coupling

introduces a dependence of the FG potential of a cell on the FG potential of the nearby cells. Clearly, this ‘interference’ depends on the electrical status of nearby cells which means that the program or erase operation on a cell affects the memory state of the neighbouring cells.

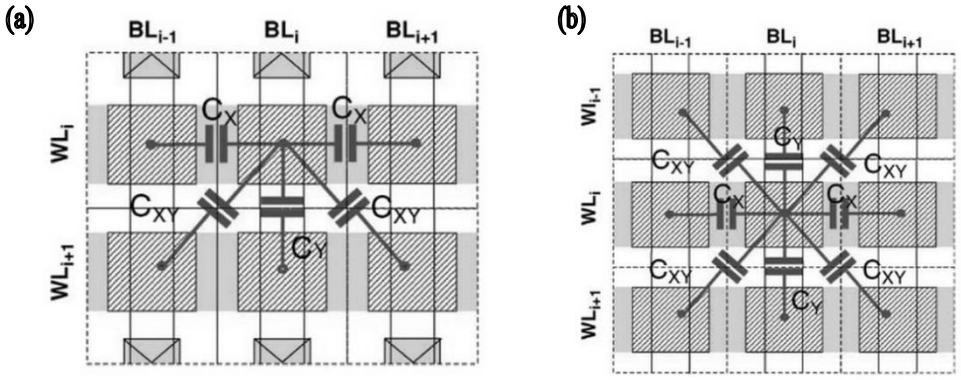


Figure 2.2-7: Schematic drawing of parasitic capacitance among nearby cells in (a) NOR Flash array and (b) a NAND Flash array.

These scaling limitations have led to a gap between the technologies of logic devices and FG-NVM cells. This is unfortunate, not only because NVM finds it difficult to continue to follow Moore’s law, which states that the number of transistors that can be placed on an integrated circuit must double approximately every two years, but also because there is demand for NVMs integrated on the same chip as a logical circuit and obviously a design of an integrated circuit with embedded both logic and memory devices cannot be realized following the design rules of the smallest node technology.

2.3 NANOPARTICLE MEMORY DEVICES

2.3.1 Structure and Operation

Many of the issues concerning the limits of floating-gate memory devices can be faced by replacing the conventional continuous and conductive polysilicon floating gate with a layer where the charge is stored in discrete sites which are mutually electrically isolated. One way of creating such a floating gate layer is to use nanogranular materials made of a high density of conducting islands, commonly referred to as *nanoparticles* (NPs), *nanocrystals* (NCs) or *quantum dots* (QDs). These islands should form a two dimensional (2-D) array with an appropriate interdot distance to prohibit the direct tunnelling of charges between adjacent dots (mutual electrical isolation). Such an alternative memory cell is shown in figure 2.3-1 and is usually called *nanofloating gate memory* or *nanoparticle FG memory* (NP-NVM), or *nanocrystal memory* (NC-NVM). Tiwari et al. were the first to propose a memory device that utilized charge storage in silicon nanocrystals in 1996. Since then, several variations of the NP-NVM have been suggested. Those include memories with nanoparticles made of a number of different materials (metal or semiconducting), as well as differentiations in the structure, such as stacked arrays of QDs.

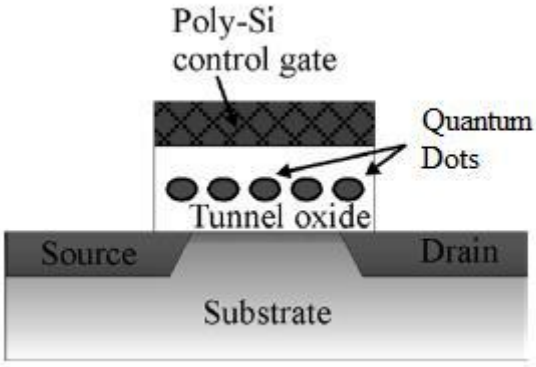


Figure 2.3-1: Schematic representation of a nanocrystal memory cell.

The energy band diagram of a nanoparticle memory cell is shown in figure 2.3-2 at flat-band conditions. This band diagram, in 1-D describes the basic operation principle of any nanoparticle based memory cell. For the sake of simplicity let us assume that only electrons are taking part in the memory operation. Nevertheless, hole action can be described by the same mechanisms like electrons assuming a band diagram horizontally symmetric to the one in figure 2.3-2. According to the figure, the NP layer is separated from the cathode (electron injection side) and anode electrodes by two layers made of different materials with physical thicknesses t_i and t_s respectively. The layer nearby the cathode is the injection layer, while the other one is the blocking layer. Their energy barrier heights with respect to the NPs are Φ_{BI} and Φ_{BB} respectively. Finally, Φ_{BA} and Φ_{BC} denotes the anode and cathode barrier heights respectively.

Very much like with the conventional floating-gate memory devices, several physical mechanisms are available to accomplish this charge transfer, but the most commonly used ones are either channel hot electron injection (CHEI) or Fowler-

Nordheim (FN) tunnelling for the write operation, and FN tunnelling for the erase operation.

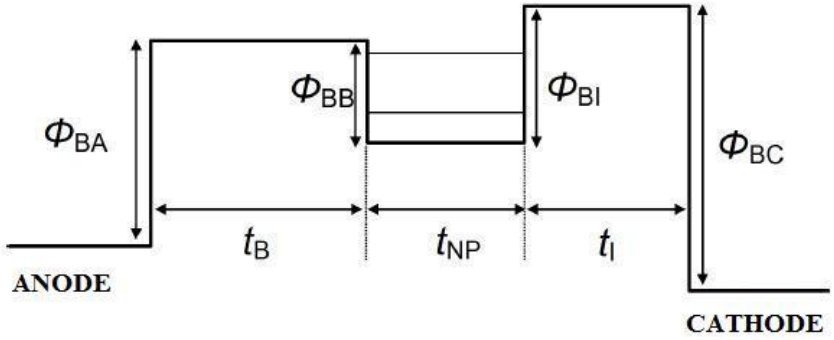


Figure 2.3-2: One dimensional energy band diagram under flat-band conditions describing the basic structure for NP-based memory cells.

2.3.2 Memory Window and Capacitance

To extract the memory window for a nanoparticle memory cell, first we have to consider the density distribution of the charge that gets stored in the nanocrystals. Figure 2.3-3 presents the approximation used to describe the charge storage in the nanocrystals: we will consider that in each quantum dot, ν electrons are stored, and those electrons are located in the center of the nanocrystal. Additionally, we consider the surface density of the nanoparticles to be n_{nc}; t_{nv} is the diameter of the ncs, t_{co} the thickness of the control oxide, and t_{ox} the total oxide thickness (quantum dots included).

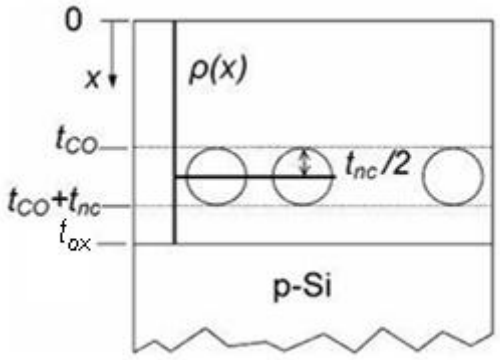


Figure 2.3-3: Schematic representation of the charge density distribution of the stored charge in NP-based memory.

Then the charge density distribution can be written as

$$\rho(x) = \begin{cases} \rho_0, & x = x_0 \\ 0, & x \neq x_0 \end{cases}, \quad \text{with } \rho_0 = n_{nc} \nu q, \quad \text{and } x_0 = t_{co} + \frac{t_{nc}}{2}.$$

If we consider equivalent oxide thicknesses for the different material that the quantum dots are made of, we have:

$$x_0^{eq} = t_{co} + \frac{\varepsilon_{ox} t_{nc}}{2\varepsilon_{si}}, \text{ and}$$

$$t_{ox}^{eq} = t_{co} + \frac{\varepsilon_{ox} t_{nc}}{\varepsilon_{si}} + t_{TO}.$$

where ε_{nc} is dielectric constant of the nanocrystals, and ε_{co} the dielectric constant of the oxide. $t_{TO} = t_{ox} - t_{CO} - t_{nc}$.

Using these equivalent oxide thicknesses, we get:

$$\gamma = \frac{t_{co} + \frac{\varepsilon_{ox} t_{nc}}{2\varepsilon_{si}}}{t_{co} + \frac{\varepsilon_{ox} t_{nc}}{\varepsilon_{si}} + t_{TO}}$$

$$\left|\Delta V_{FB}\right| = \left|\gamma \frac{Q_{ot}}{C_{ox}}\right| \Rightarrow \left|\Delta V_{FB}\right| = \left|\Delta V_{th}\right| = \frac{q \nu n_{nc}}{\varepsilon_{CO}} \left(t_{CO} + \frac{1}{2} \frac{\varepsilon_{CO}}{\varepsilon_{nc}} t_{nc} \right),$$

which can be rewritten as:

$$\boxed{MemoryWindow = \frac{q \nu n_{nc}}{\varepsilon_{CO}} \left(t_{CO} + \frac{1}{2} \frac{\varepsilon_{CO}}{\varepsilon_{nc}} t_{nc} \right)}.$$

The above expression for the Memory Window is used extensively in the study of nanoparticle memories. Assuming that each quantum dot can hold one carrier (ν=1), one can extract the surface density of the nanocrystals n_{nc}. Furthermore, if that density is somehow known (e.g. from AFM or TEM tests), the number of electrons that can be stored in each nanoparticle can be calculated.

Another expression for the memory window, takes into account not only the density of the nanoparticles but also the Ratio R, which is proportional to the density n_{nc} and the diameter t_{nc} of the ncs:

$$R = n_{nc} A_{nc} = n_{nc} \pi t_{nc}^2 / 4,$$

where A_{nc} is the area occupied by a single nanocrystal in the plane. The memory window is then given by

$$\Delta V_{th} = R \frac{Q_{nc}}{C_{ox}'}, \text{ with}$$

$$C_{ox}' = \varepsilon_{ox} S (1 - R) / (t_{CO} / t_{TO})$$

where S is the gate area. However, this is not an expression for the memory used very often.

As with conventional floating-gate cells, the nanocrystal floating gate can hold either electrons, or holes, or both, with the respective shift in the flat-band voltage and the C-V characteristic of the device. The same discussion that was made in paragraph 2.2.2 holds here, with the conclusions being summarized in figure 2.2-6. It is reminded that when negative charges are stored in the floating gate (or, in this case, the quantum dots), the flat-band voltage shifts a higher value, whereas storage of positive charges results in the shifting of the flat-band voltage to the left.

Here, however, we consider one more very important aspect of the C-V curve of non-volatile memories: the direction of the hysteresis. To extract the capacitance-voltage characteristics of an MOS-like memory structure, the voltage is swept from the inversion bias voltages to accumulation and back. If the device has memory behaviour, then this is exhibited by form of a hysteresis in the C-V curve. This hysteresis can have either a clockwise or a counter-clockwise direction, and the direction provides information regarding the part of the device where carriers are injected from.

Let us consider a MOS capacitor with nanocrystals inside of the oxide and let us also assume that only electrons are stored in the quantum dots. To perform the C-V test, the device is biased like it is shown in figure 2.3-4. When positive voltage is applied on the control gate, this will result in the electrons moving up. If injection takes place from the substrate, that means charging of the nanoparticles. There is a case, however, that electrons are injected from the control gate electrode. Then, electrons moving up will mean the loss of charge from the nanocrystals (erase operation). Respectively, when the negative bias is applied, electrons will be moving down. This means, erasing of the memory in the first case, and programming in the latter. **This obviously translates in a *clockwise* hysteresis, when electrons are injected from the control gate electrode and a *counter-clockwise* one when electrons are injected from the substrate** (fig. 2.3-5).

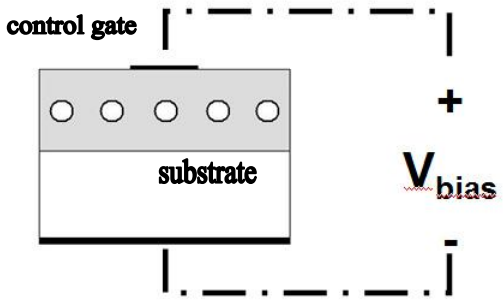


Figure 2.3-4: Applying bias to the nanoparticle memory MOS capacitor.

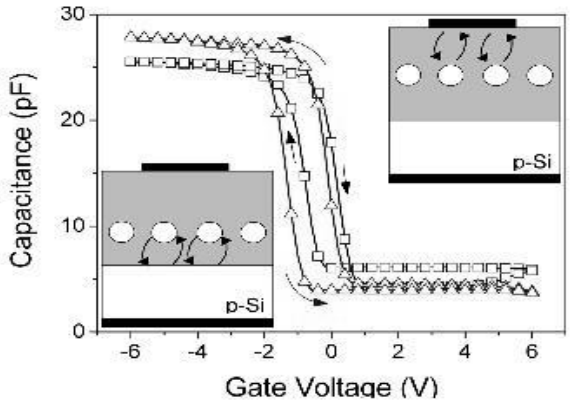


Figure 2.3-5: Experimental high-frequency (1MHz) C-V for two NP-MOS capacitor memory cells where the 2-D array of nanoparticles is sited at different distances from the Si substrate, leading to different exchange mechanisms.

Injection of carriers from the control gate could mean faulty fabrication of the memory cells, as it is desirable the charging of the device to happen with carriers coming from the substrate, and it could mean a control oxide with very small thickness. The case described above is shown in figure 2.3-6 along with three more cases of different types of hysteresis. N-type and p-type substrates are separately looked into, but for the sake of simplicity, all the cases assume storage of a single type of carrier. Nonetheless, one can form conclusions for more complicated cases based on these results.

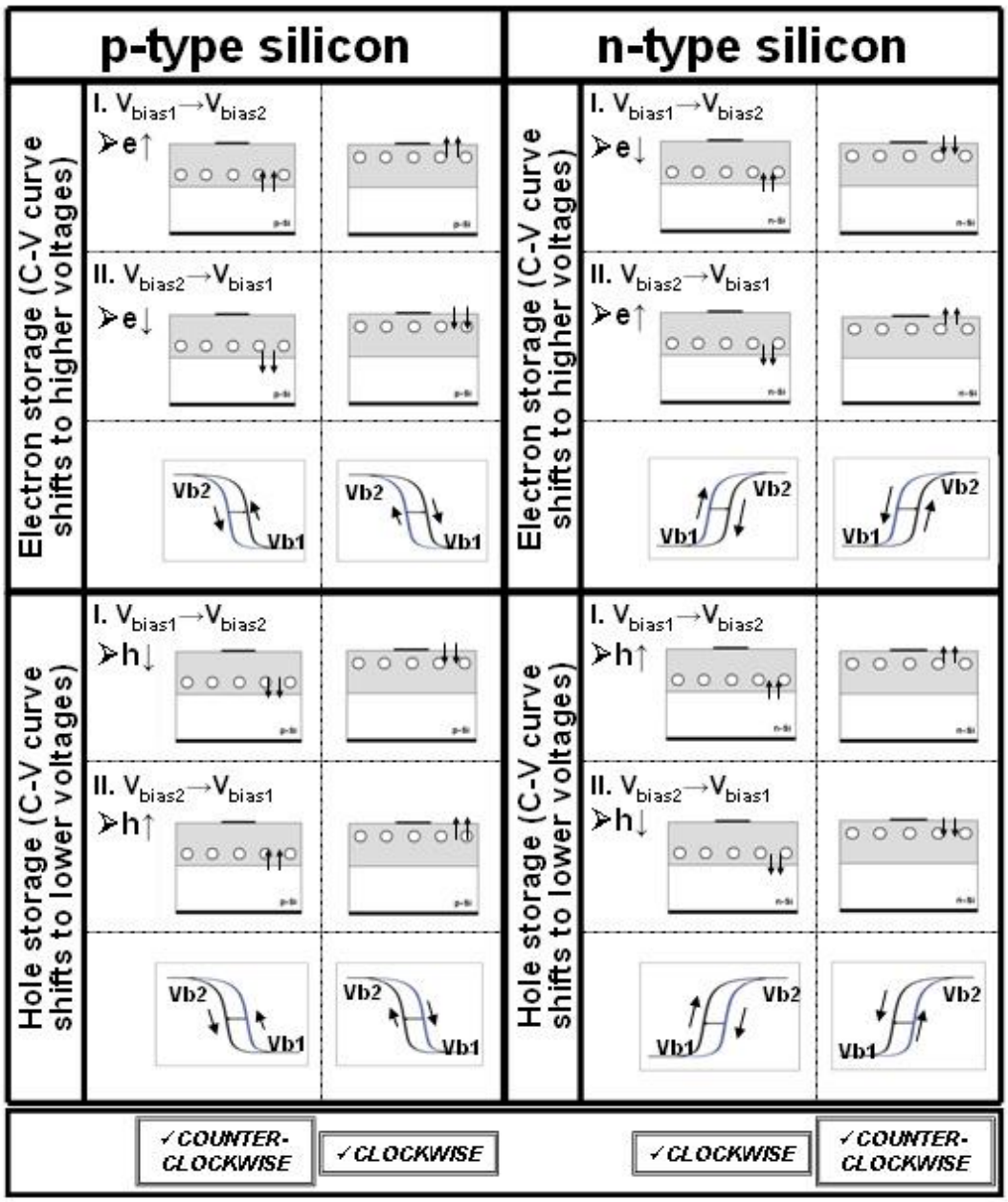


Figure 2.3-6: Direction of the hysteresis loop in relation to the place of carrier injection.

2.3.3 Nanoparticle Memory Cells versus Conventional Floating Gate Devices

As compared to conventional stacked-gate NVM devices, nanocrystal charge-storage offers several advantages, the main one being the potential to use thinner tunnel oxides without sacrificing non-volatility. This is a quite attractive proposition, since reducing the tunnel oxide thickness is key to lowering operating voltages and/or increasing operating speeds.

One main disadvantage of a uniform floating gate is that charge storage becomes very sensitive to oxide defects. Storing charge on a single node -the FG node- makes the conventional memory structure particularly prone to failure of the FG isolation. One weak spot in the tunnel oxide is sufficient to create a fatal discharge path, compromising long-term non-volatility. In a nanocrystal NVM device charge is not stored on a continuous FG poly-Si layer, but instead, on the layer of the discrete, mutually isolated quantum dots. As a result, when a defect presents itself in the oxide, it will affect charge storage only on the neighbouring quantum dot, whereas the rest of the nanocrystals will continue to store charge as effectively as before (fig. 2.3-7)

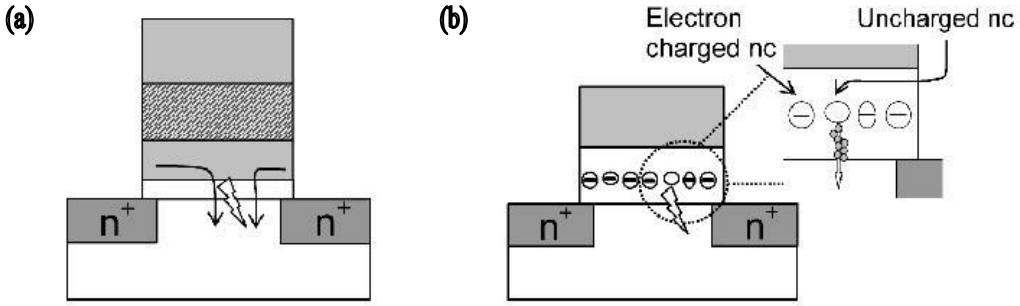


Figure 2.3-7: The stored charge in a conventional FG NVM (a) is leaked entirely to the channel and the stored information is lost, while in a NP memory cell (b) only the quantum dot above the defect is discharged and hence the stored information remains.

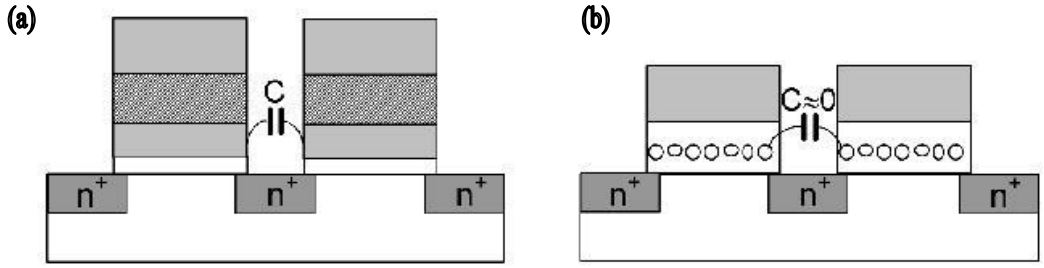


Figure 2.3-8: Schematic representation of capacitive coupling between adjacent NVM cells in the case of (a) a continuous FG and (b) nanoparticle FG.

From an integration point of view, the floating gate coupling among neighbouring cells that was discussed in paragraph 2.2.2, is eliminated in nanocrystal memory cells due to the nanometre size of the coupling elements (fig. 2.3-8) and so device density can be increased with the minimum crosstalk effects.

There are other important advantages though. First, nanocrystal memories use a simplified fabrication process as compared to conventional stacked-gate FG NVM's by avoiding the fabrication complications and costs of a dual-poly process. Further,

due to the absence of drain to FG coupling, nanocrystal memories suffer less from drain-induced-barrier-lowering (DIBL) and therefore have intrinsically better punchthrough characteristics. One way to exploit this advantage is to use a higher drain bias during the read operation, thus improving memory access time. Alternatively, it allows the use of shorter channel lengths and therefore smaller cell area (i.e., lower cost).

Finally, nanocrystal memories are characterized by excellent immunity to stress induced leakage current (SILC) and oxide defects due to the distributed nature of the charge storage in the nanocrystal layer.

There are, however, some intrinsic weaknesses as well. Of particular importance is the low capacitive coupling between the external control gate and the nanocrystal charge storage layer. This does not only result in higher operating voltages, thus offsetting the benefits of the thinner tunnel oxide, it also removes an important design parameter (namely the coupling ratio) typically used to optimize the performance/reliability trade-off. More importantly, FG-based Flash devices are a proven, mature, and trusted concept, with a substantial track record of demonstrated reliability. New technologies face time-compression diseconomies, and may find it difficult to match such history within a reasonable time frame.

2.4 QUANTUM DOTS

2.4.1 Confinement: Properties of Quantum Dots

Nanoscale science, as the name implies, is concerned with the study of materials with a dimension of roughly 1 to 100 nm. Such structures belong to a group known as low-dimensional structures (2D-0D). The dimensionality refers to the number of directions in which the carriers of the material act as free carriers. When a material is spatially confined in all three directions, a 0D structure is formed. This is called a *quantum dot*. Nanoparticles (or nanocrystals) used in non-volatile memories are treated like quantum dots, that is, it is presumed that their properties are like to the electronic properties of quantum dots. For this reason, it is helpful that we look into some of those properties in this chapter.

Size and Shape of QDs

There are many different techniques used for the fabrication of quantum dots. Some of them are the aerosol technique, Low Pressure Chemical Vapour Deposition (LPCVD), sputtering techniques, and Molecular Beam Epitaxy (MBE). Different techniques result in different shapes of quantum dots, and those include spherical, semi spherical, or ellipsoidal.

The size of the nanoparticles plays a very important role in affecting its properties. The main electronic property that is affected is the density of states as a function of the system energy, which is transformed to delta functions at specific energy levels (fig. 2.4-1).

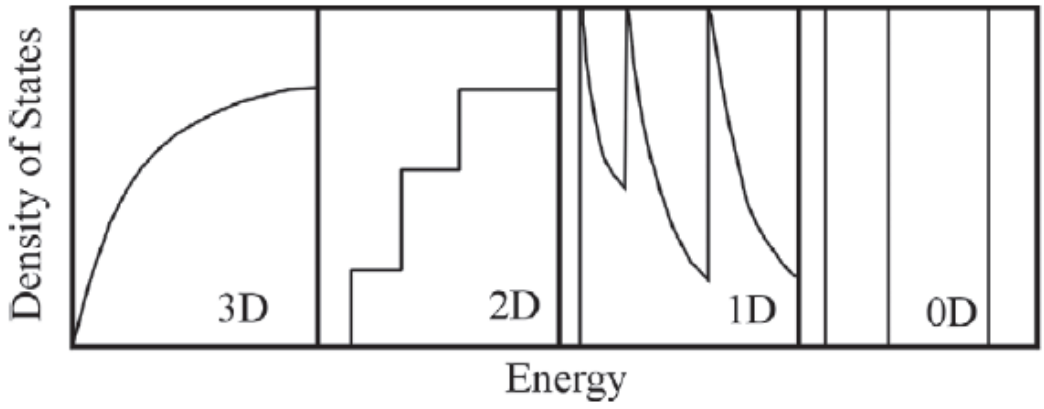


Figure 2.4-1: Density of states vs. energy for a bulk material (3D), quantum well (2D), quantum wire (1D), and quantum dot (0D).

The energy band-gap also becomes different to that of the bulk material and is dependent on the quantum dot size. Generally, the smaller the size of the nanoparticle, the larger the band gap, the greater the difference in energy between the highest valence band and the lowest conduction band becomes, therefore more energy is needed to excite the dot and concurrently, more energy is released when the crystal returns to its resting state. A better understanding of the energy band gap variation, as well as the discretization of the electronic density of states can be obtained by writing and solving the Schrodinger equation for an electron in a sphere of radius α and

considering the potential is zero inside the sphere and infinite everywhere outside the sphere. The solution for this problem of quantum mechanics yields allowed energy levels that are described by the equation:

$$E_{n0} = \frac{n^2 \pi^2 \hbar^2}{2m^* \alpha^2}, \text{ where n = 1, 2, 3, …}$$

In the above expression, ħ is Plack's constant, and m* the effective electron mass, and n the allowed energy level. It is obvious that the energy levels are discrete and that the energy is inversely proportional to the square of the confinement dimension α, which means the smaller the quantum dot, the higher the energy level.

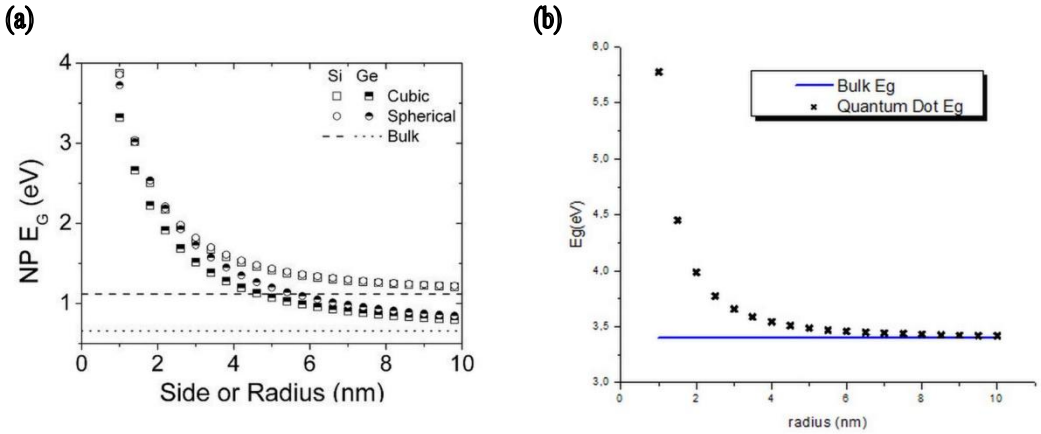


Figure 2.4-2: (a) Dependence of the energy band-gap on the effective dimension of spherical and cubic Si and Ge nanoparticles. (b) Dependence of the energy band-gap on the radius of a spherical GaN QD.

Precise calculations for the prediction of the electronic properties of nanoparticles require knowledge of the crystal lattice properties and the shape of the QDs. Different theories (k•p, tight-binding) have been used to calculate the allowed electron energy levels, the density of states and the energy band-gap variation as a function of the QD size. Figure 2.4-2a shows the dependence of the energy band-gap E_g on the effective dimension of spherical and cubic Si and Ge nanoparticles. Significant departure from the bulk material is predicted and observed when the QD dimension is lower than 6.5nm. Usually, the difference in band-gap energy between the QD and the bulk material is called *confinement energy* ΔE_G.

Assuming an electron-hole system confined in a spherical nanocrystal, the following expression has been derived for ΔE_G:

$$\Delta E_G = \frac{\hbar^2 \pi^2}{2\alpha^2}\left(\frac{1}{m_e^*} + \frac{1}{m_h^*}\right) - \frac{1.8q^2}{\varepsilon\alpha} - 0.24E_{Ry}^*, \text{ with}$$

$$E_{Ry}^* = 13.6058\frac{1}{\varepsilon^2}\left(\frac{m_0}{m_e^*} + \frac{m_0}{m_h^*}\right)^{-1}.$$

where ε is the optical dielectric constant of the QD material, α the radius of the QD and m_e^*, and m_h^* the effective electron and hole masses respectively. E_{Ry}^* denoted the effective Rydberg energy in eV, and m_0 is the free electron mass.

Assuming spherical quantum dots and using the above expression, figure 2.4-2b shows the dependence of the band-gap on the radius α, for gallium nitride.

Coulomb Blockade

The transformation of continuous energy bands of a bulk material to discrete energy levels at the nanoscale and the subsequent quantum-size effects motivate the use of nanoparticles as discrete storage nodes in memory devices. The retention of a stored charge is determined not only by the thicknesses of the oxide but also by the barrier height of the well, which in turn depends of the energy band-gaps of the material. However, we must consider one more critical quantum mechanical parameter especially in regard with the carrier injection process.

Let us consider a thin insulating barrier between two conducting electrodes (fig. 2.4-3a). The insulating material can be said to have a capacitance per area C=ε/d. According to the laws of quantum mechanics, there is a larger than zero possibility for an electron on one side of the barrier to reach the other side. When a bias voltage is applied, this means that there will be a current, neglecting additional effects, the tunnelling current will be proportional to the bias voltage.

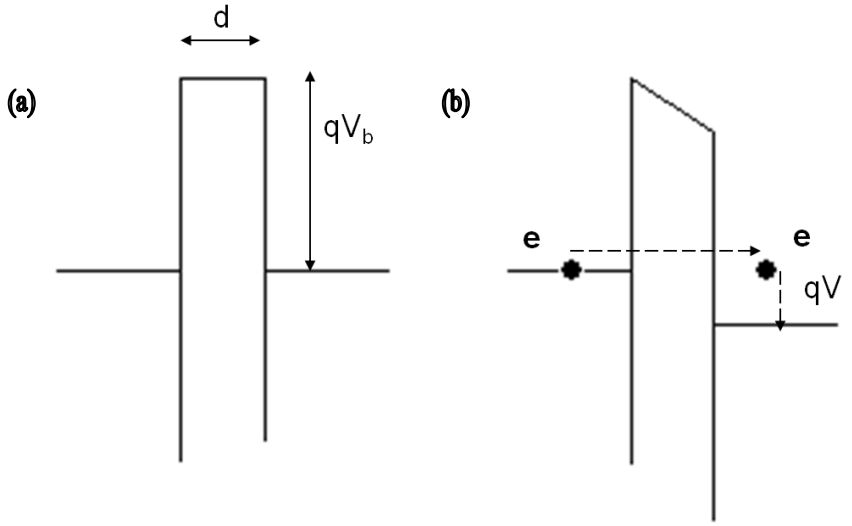


Figure 2.4-3: Schematic representation of an electron tunnelling through a barrier.

Due to the discreteness of electrical charge, current through a tunnel junction is a series of events in which exactly one electron tunnels through the tunnel barrier (fig. 2.4-3b). The tunnel junction capacitor is charged with one elementary charge by the tunnelling electron, causing an energy build-up of $\Delta E = q^2/2C$. If the capacitance is very small, the voltage build-up can be large enough to prevent another electron from tunnelling. The electrical current is then suppressed at low bias voltages and the resistance of the device is no longer constant. The increase of the differential resistance around zero bias is called the *Coulomb Blockade*.

A similar phenomenon occurs when a carrier is stored in a quantum dot. Once an electron is stored in a nanoparticle, its potential energy increases by

$$\Delta E = \left[(\nu+1)^2 - \nu^2\right] q^2 / 2C_\Sigma$$

where ν is the total number of stored electrons in the QD, C_Σ is the QD self capacitance and q the electronic charge. This energy prohibits the next electron to be stored in the QD and thereby, the carrier injection process is suppressed and the corresponding nanoparticle device exhibits a self-limited injection mechanism.

2.4.2 Materials for Quantum Dots

As it has already been stated above, the ideal goal in optimizing non-volatile memory devices is to achieve fast write/erase speeds without any cost for the long retention time. For this purpose, we need to create an asymmetry in charge transport through the gate dielectric to maximize the $I_{G/Write/Erase}/I_{G/Retention}$ ratio. In order to accomplish this with nanocrystal memories, several techniques have been suggested, such as tunnel barrier engineering and double-stacked arrays of quantum dots.

Another very appealing approach is to engineer the depth of the potential well at the storage node (i.e. the nanocrystal) thus creating an asymmetrical barrier between the substrate and the storage nodes, which translates into a small barrier for writing and a large barrier for retention (fig. 2.4-4a). The idea is that a large $I_{G/Write/Erase}/I_{G/Retention}$ ratio can be achieved even for very thin oxides by aligning the nanocrystal energy level to be within the Si band gap under retention and above the conduction band edge under erase. Because writing is performed by tunnelling electrons from the Si substrate into the nanocrystals (thus can always find available states to tunnel into) and can have current similar to $I_{G/Erase}$, fast write/erase and long retention time can be achieved simultaneously.

Towards this direction, memory devices that utilize metal nanoparticles have been looked into, as metals can provide a large variety of work functions to choose from for the well potential, thus making the engineering much easier. Apart from the range of work functions, metal nanocrystals have a number of other advantages as well: higher density of states around the Fermi level, stronger coupling with the conduction band channel, and smaller energy perturbation due to confinement issues.

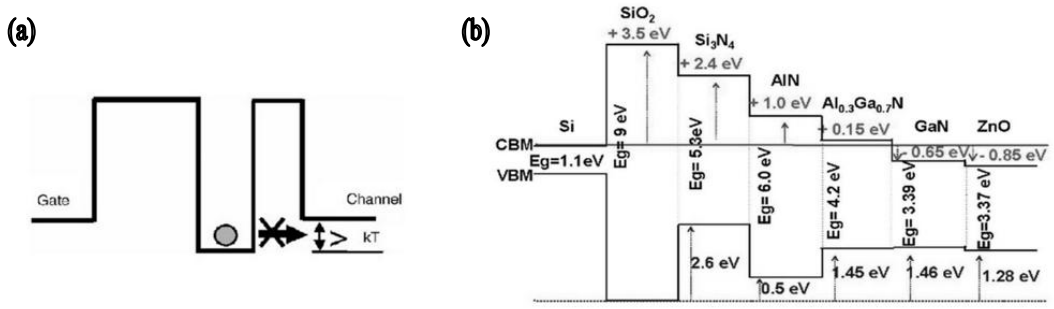


Figure 2.4-4: (a) Energy barriers of a nanocrystal memory in the case of metallic nanoparticles with higher work function than silicon (channel area). (b) CBO (Conduction Band Offset) values of various materials compared to that of silicon {reference [17]}.

Despite all their virtues, metal nanoparticles have a major disadvantage which greatly limits their potential commercial applicability: they are not compatible with current CMOS technology and their fabrication requires special techniques. This means that even if they possess excellent qualities, they might have a very increased manufacture cost and industry could take some time before it can put such devices to practice.

This is the reason why nanocrystal memory research has been focused on semiconducting materials, most prominently silicon (Si) and germanium (Ge). Neither Si nor Ge QDs, however, exhibit the desirable negative offset compared to the

conduction band minimum of the silicon substrate. Fortunately, we do not have to abandon this idea of negative offset altogether, as there exist semiconducting materials that can have their electron energy level within the band gap of Si (fig. 2.4-4b, and reference [17]). In particular, a semiconductor with very appealing band alignment compared to both silicon and silicon dioxide seems to be gallium nitride, although there is some debate concerning the band offsets of these materials and this has yet to be fully confirmed.

Gallium Nitride: Gallium nitride (GaN) is a direct band-bap semiconductor commonly used in optoelectronic, high-power and high-frequency devices. It is a very hard material with wurzite crystal structure and a wide band gap of approximately 3.4eV. It is mechanically stable and has high heat capacity and thermal conductivity.

A one-dimensional energy band diagram under flat-band conditions of a non-volatile MOS type structure that include GaN QDs is presented in figure 2.4-5. It should be noted though, that this diagram does not take into account the quantum confinement effects that should not be ignored when one wishes to extract an exact model behaviour. It shows that electrons inside the nanoparticles exhibit a negative conduction band offset with respect to the silicon substrate conduction band. In this case, as we have explained, the injected electrons can be trapped into the nanocrystals at low voltage, and the retention time can be significantly improved due to the higher energy barrier the carriers have to overcome in order to tunnel back to the Si substrate.

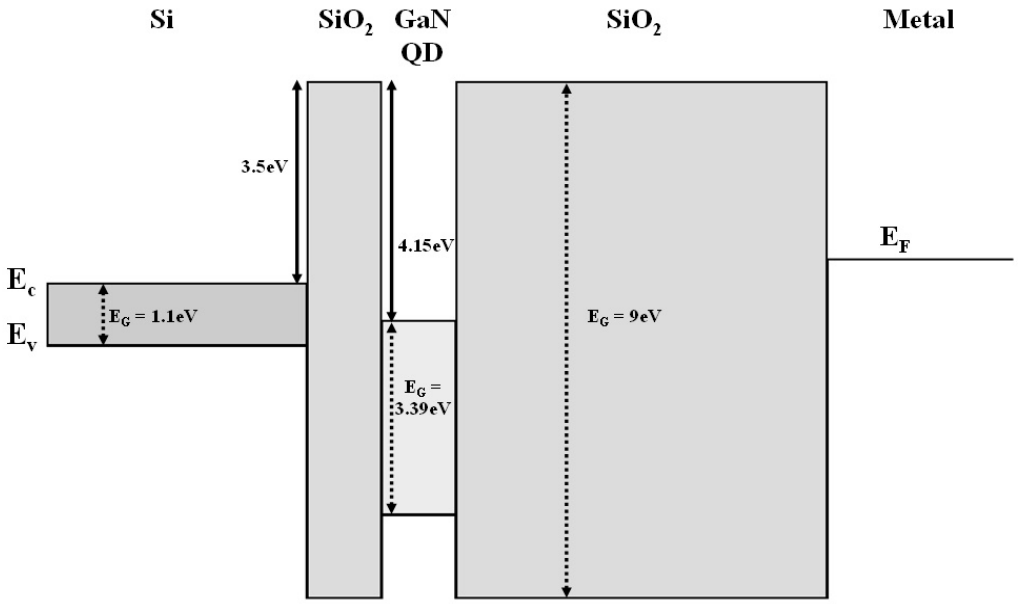


Figure 2.4-5: One-dimensional energy band diagram of a non-volatile cell of the MOS type with GaN QDs.

There are other reasons, however, why GaN would make a good candidate material for nanoparticles. The wider band gap should lead to even longer retention times. This may seem trivial since Si and Ge QDs have proven to have excellent retention

characteristics. Considering though that storage loss is a thermally activated process with storage time decreasing exponentially as temperature rises, GaN could perform much better in higher temperatures if such a thing is necessary.

REFERENCES

[1] Muller, R. S., Kamins, T.I., (2003) *Device Electronics for Integrated Circuits*. ^^Ed.^^ New Jersey: John Wiley and Sons.

[2] Nicollian, E. H., Brews J. R., (1982) *MOS (Metal Oxide Semiconductor) Physics and Technology*. New Jersey: John Wiley and Sons.

[3] Schroder, D. K, (2006) *Semiconductor Material and Device Characterization*. 3rd Ed. New Jersey: John Wiley and Sons.

[4] Brewer, J. E. (ed.), and Gill M. (ed.), (2008) *Nonvolatile Memory Technologies with Emphasis on Flash: A Comprehensive Guide to Understanding and Using NVM Devices*. New Jersey: John Wiley and Sons.

[5] Khriachtchev, L. (ed.) (2009) *Silicon Nanophotonics: Basic Principles, Present Status and Perspectives*. Singapore: Pan Stanford Publishing Pte Ltd.

[6] Dimitrakis, P., (2006) *Nanoparticle Electronics*. Ph.D., National Technical University of Athens.

[7] Lee, H-R., Samjong, C., Kyoungah, C., and Sangsig K. (2007) *Capacitance-Voltage Characteristics of MOS Capacitors with Ge Nanocrystals Embedded in ZrO_2 Gate Material*. Thin Solid Films, 516, pp. 412-416.

[8] Ghetti, A., Bortesi, L., and Vendrame, L. (2005) *3D Simulation Study of Gate Coupling and Gate Cross-Interference in Advanced Floating Gate Non-Volatile Memories*. Solid State Electronics, 49, pp. 1805-1812.

[9] Gargin, P.A. (2004) *Silicon Nanoelectronics and Beyond*. Journal of Nanoparticle Research, 6, pp. 11-26.

[10] Tiwari, S., Farhan R., Hussein H., Hartstein, A., Crabble, E. F., and Chan, K. (1996) *A Silicon Nanocrystals Based Memory*. Applied Physics Letters, 68/10, pp. 1377-1379.

[11] Chen, W-R., Chang T-C., Liu, P-T., Lin, P-S., Tu, C-H., and Chang, C-Y. (2007) *Formation of stacked Ni Silicide Nanocrystals for Non-volatile Memory application.* Applied Physics Letters, 90 112108.

[12] Dimitrakis, P., Schamm-Chardon S., Bonafos, C., and Normand, P. (2012) Nanoparticle-Based Memories: Concept and Operation Principles. in: Chaughule R.S., and Watawe S.C. (eds.) *Applications of Nanomaterials*. American Scientific Publlishers.

[13] De Blauwe, J. (2002) *Nanocrystal Nonvolatile Memory Devices*. IEEE Transactions on Nanotechnology, vol. 1, no 1, pp. 72-77.

[14] Bukowski, T.J., Simmons, J.H., (2002) *Quantum Dot Research: Current State and Future Prospects*. Critical Reviews is Solid State and Materials Sciences, 27. pp. 119-142.

[15] Kayanuma, Y., (1988) *Quantum Size Effects of Interacting Electrons and Holes in Semiconductor Nanocrystals with Spherical Shape*. Physical Reviews B, vol. 38, no. 14, pp. 9797-9805.

[16] Liu, Z., Lee, C., Narayanan, V., Pei, G., and Kan E. C., (2002) *Metal Nanocrystal Memories - Part I: Device Design and Fabrication*. IEEE Transactions on Electron Devices, vol. 49, pp. 1606-1613.

[17] Tsoukalas, D., Dimitrakis, P., Kolliopoulou, S., and Normand, P., (2005) *Recent Advances in Nanoparticle Memories*. Materials Science and Engineering B, 124-125, pp. 93-101.

[17] Joo, K.-H., Moon C.-R., Lee S.-N., Wang X., Yang J.-K., Yeo, I.-S., Lee, D., Nam, O., Chung, U-I., Moon J. T., and Ryu B.-I., (2006) *Novel Charge Trap Devices with NCBO Trap Layers for NVM or Image Sensor*. IEDM Technical Digest, 10.1109/IEDM.2006.34950.

[18] Dimitrakis, P., Iliopoulos, E., and Normand, P., (2010) *Nanocrystal Memory Device Utilizing GaN Quantum Dots by RF-MBD*. Material Research Society Symposium, 1250-G06-03.

[19] Pankove, J.I. (ed.), and Moustakas, T.D (ed.), (1998) *Gallium Nitride (GaN) I*, vol. 50 in *Semiconductors and Semimetals*, Academic Press.

CHAPTER 3

Device Development and Characterization

At the end of the previous chapter, we explained the reasons why gallium nitride (GaN) would make a suitable choice of a material for nanocrystal NVM. Such a case, however, needs to be investigated more thoroughly for its validity to be proven and research in this particular field is still quite young. This thesis aims to examine if there is in reality any promise in such devices to meet non-volatile memory specifications and make acceptable candidates for the replacement of conventional floating-gate transistors.

For this purpose, a series of two terminal devices of the MOS-type were developed on n-type silicon, with gallium nitride quantum dots embedded inside the silicon dioxide dielectric. Different deposition conditions for the GaN were applied to each sample so as to have devices with quantum dots that vary in size and concentration, in order to study the effect that the nanocrystals have on the memory characteristics of the MOS capacitor. In this chapter, the development of these devices is described as well as the process that was followed to attempt to characterize their electrical properties.

The first step in this study was to perform high-frequency C-V sweeps to observe if the distinctive hysteresis that is characteristic of memory can be observed, and to obtain a measure of how wide the memory window can be. Secondly, a series of pulses was applied to try to assess how the device can be charged and discharged (or programmed and erased).

Retention measurements for these devices show an unusual behavior implying more than one discharging mechanisms. To investigate this, the last part of this study includes transient current measurements at different bias voltages.

3.1 GROWTH OF GAN QDs MOS DEVICES

3.1.1 Fabrication of the Devices

The structure of the devices that were developed is presented in figure 3.1-1. SiO_2 was formed by thermal oxidation on a (100) n-type silicon substrate, which resulted in a thin layer of high quality dielectric of 3.5 nm. Following this, GaN-QDs were formed on the oxide surface by Radio Frequency plasma assisted Molecular Beam Deposition (RF-MBD). Several samples were realized, using different deposition conditions for the GaN in order to achieve a range of quantum dot size and density distributions and investigate their effect on the memory qualities of the devices. The different deposition conditions are presented in table 3.1-1, where the dose is described in monolayers (MLs) of GaN equivalence; one ML is the amount of deposited GaN equivalent to one two-dimensional monolayer (thickness of 0.2593nm). The growth of the QDs will be more thoroughly discussed later on. Next, as a capping dielectric, a ~20nm-layer of SiO_2 was deposited with Low Pressure Chemical Vapour Deposition (LPCVD) by exposing the sample to tetraethyl-orthosilicate (TEOS). A review of these techniques is given in the following paragraph.

For comparison, a control sample without QDS, but treated in similar conditions and exposed to RF N-plasma at the QDs growth step was also grown. All the MOS capacitors use an aluminium gate.

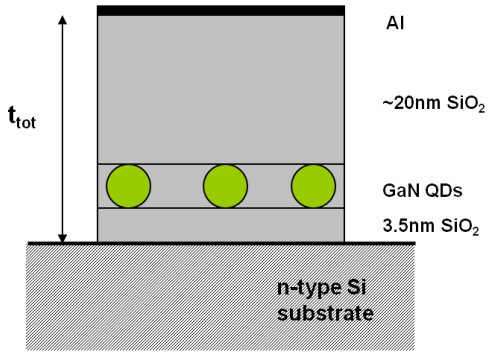


Figure 3.1-1: Structure of the GaN-QDs MOS Device.

Sample	Description of GaN Deposition Conditions
C2011	8 MLs dose GaN QDs
C2012	10 MLs dose GaN QDs
C2013	14 MLs dose GaN QDs
C2014	5 MLs dose GaN QDs
C2015	18 MLs dose GaN QDs
C2016	10 MLs dose GaN QDs [lower III/V]
CRef	No deposition

Table 3.1-1: Description of the fabricated samples.

3.1.2 Description of Fabrication Processes

In order to better understand the process of the fabrication of the samples that were studied, a brief description of the techniques that were used will be given in this paragraph.

Thermal Oxidation: Thermal oxidation involves exposing the silicon substrate to an oxidizing agent at high temperatures (between 800 and 1200 ^{o}C). The result is that oxygen (O_2) diffuses into the wafer and reacts with it, thus forming silicon dioxide (SiO_2). The oxidizing agent can be either water vapour (H_2O - the process is then called wet oxidation) or molecular oxygen (O_2 – dry oxidation). The oxidizing ambient may also contain several percent of hydrochloric acid (HCl) which helps to remove metal ions that may occur in the oxide.

Thermal oxide incorporates silicon consumed from the substrate and oxygen supplied from the ambient. Thus, it grows both down into the wafer and up out of it. For every unit thickness of silicon consumed, 2.27 unit thickness of oxide will appear. Consequently, if a bare silicon surface is oxidized, 44% of the oxide thickness will lie below the original surface, and 56% above it. The rate of growth is often predicted by the Deal-Grove model:

$$t = \frac{X_0^2}{B} + \frac{X_0}{B/A},$$

where t is the time needed to grow oxide of thickness X_0. A, and B are constants that depend on the growth conditions.

It should be noted that dry oxidation produces much better oxide quality than wet oxidation. Similarly, a better oxide results when the process takes place on a (100) Si wafer rather than a (111) wafer.

Low Pressure Chemical Vapour Deposition (LPCVD): The process of Chemical Vapour Deposition (CVD) involves exposing the wafer to one or more volatile precursors which react and/or decompose on the substrate surface to produce the desired deposit. By-products can also be produced, and those are removed by gas flow through the reaction chamber.

The term ‘low pressure’ is used to describe CVD processes that take place at sub-atmospheric pressures. Reduced pressures tend to reduce unwanted gas-phase reactions and improve film uniformity across the wafer. Most modern CVD are either LPCVD or UHVCVD (Ultra High Vacuum Chemical Vapour Deposition).

For the growth of silicon dioxide, common source gases include silane and oxygen, dichlorosilant ($SiCl_2H_2$) and nitrous oxide (N_2O) or tetraethyl orthosilicate (TEOS: $Si(OC_2H_5)_4$). TEOS is frequently used, as it produces relatively pure oxides.

RF Plasma-Assisted Molecular Beam Deposition: Molecular Beam Epitaxy (MBE) or Molecular Beam Deposition (MBD) is a process that takes place in ultra-high vacuum and is characterized by the large mean free path of the reactant species evaporated.

In solid-state MBE, elements, such as gallium, are heated in separate quasi-knudsen effusion cells until they begin to slowly sublimate. The gaseous elements

then condense to the wafer, where they may react with each other. The term ‘beam’ means that evaporated atoms do not interact with each other or with vacuum chamber gases until they reach the wafer, due to the long mean free path of the atoms. MBE allows precise control of thickness for each layer, down to a single monolayer. For this reason it is used to form very delicate structures such as quantum wells and quantum dots. A simple schematic showing the basic parts and operation of a typical MBE system is shown in figure 3.1-2.

In the case of nitrides, such as GaN, there appears an issue with solid-state MBE. Molecular nitrogen is inert at the growth temperatures. In consequence, a more active nitrogen species must be used. One of those species is nitrogen plasma, which contains atomic or excited molecular nitrogen to react in order for the nitride to be formed. The term RF results from the fact that a radio frequency (RF) source is use to generate the plasma. This technique is called RF plasma-assisted Molecular Beam Deposition.

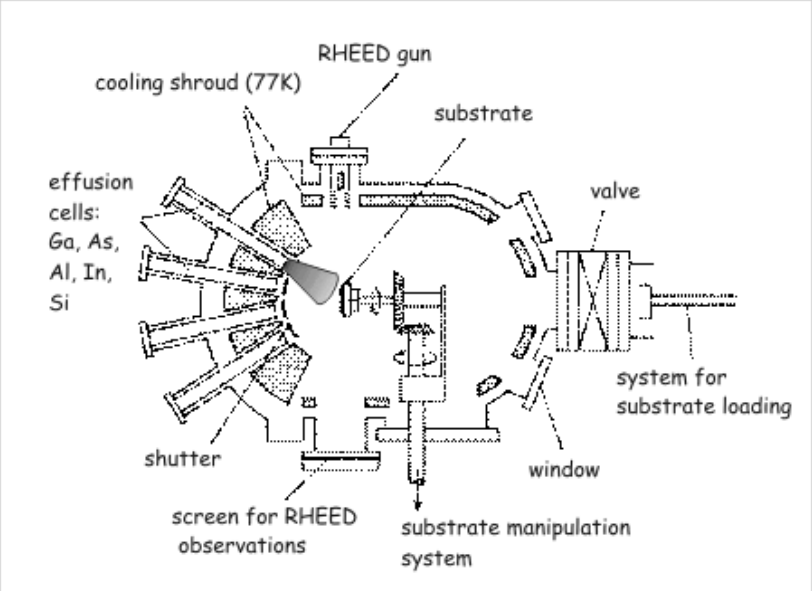


Figure 3.1-2: The basic parts and operation of an MBE system.

3.1.3 Quantum Dot Properties and Formation

Like it was mentioned earlier in this paragraph, Molecular Beam Deposition was used to grow the GaN quantum dots on amorphous SiO_2. Because of the amorphous substrate, the development of the QDs becomes a rather unpredictable process since the rules of epitaxial growth do not apply in this case. Therefore, it is vital that their presence is confirmed and, for that reason, a series of images taken by Transmission Electron Microscopy are presented in this paragraph.

Such an example is presented in figure 3.1-2 for an early sample. In figure 3.1-3a we can clearly see darker areas of material inside of the SiO_2 which should correspond to the QDs. With greater magnification (fig. 3.1-3b) it becomes apparent that those regions have a crystalline structure, even though they present a variety of crystallographic orientations.

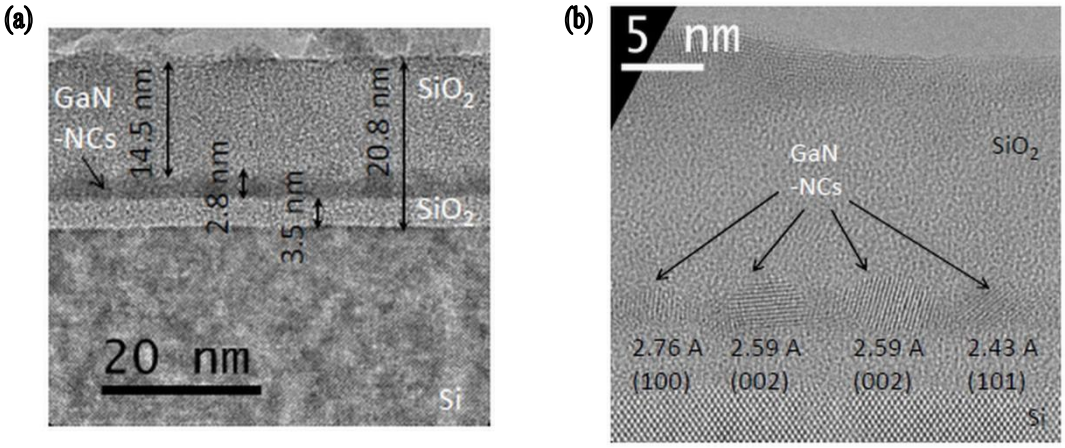


Figure 3.1-3: TEM images for an early sample that show QDs in the oxide.

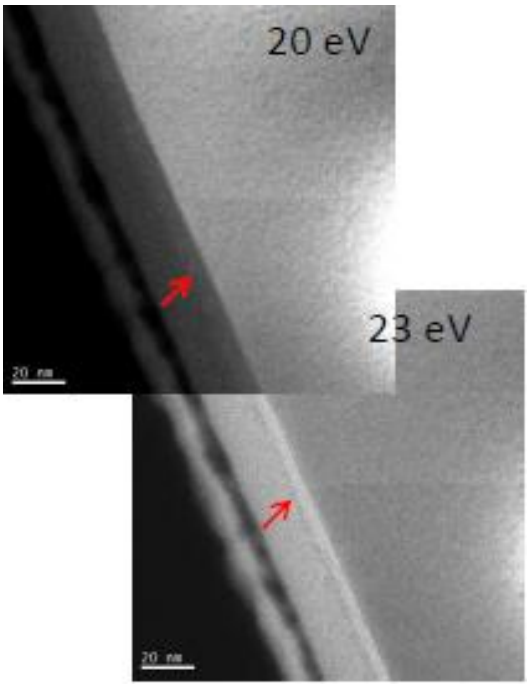


Figure 3.1-4: EFTEM images of the same sample.

The hypothesis that the QDs we see in figure 3.1-3 consist of GaN is supported not only by the fact that their crystalline structure is recognized as matching the one of GaN but also by the Energy Filtered TEM (EFTEM) images. The Electron Energy Loss Spectrum (EELS) of GaN peaks at ~20eV whereas the one of SiO_2 presents a peak at ~23eV. In figure 3.1-4, it is obvious that if the energy used for acquiring an image is at 23eV, which is the energy absorbed by SiO_2, then a dark band appears where the QDs are. When the energy changes to 20eV, which can be absorbed by in turn by GaN, but not so much by SiO_2, then a white band becomes visible in the same region.

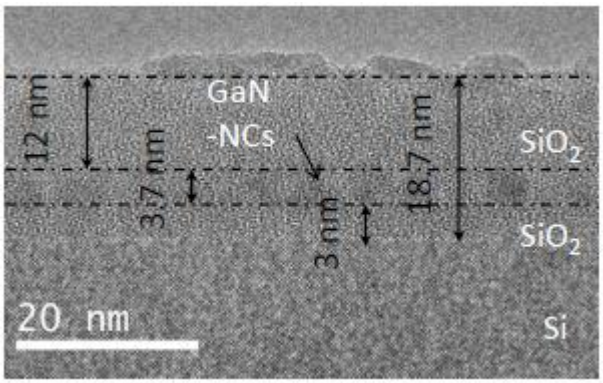


Figure 3.1-5: TEM image for a second sample.

In figure 3.1-5, one can see a second TEM image for a different sample, grown with different deposition conditions.

Unfortunately, there isn't TEM data available for all of the samples studied here. We have images for samples C2014 and C2015 and those are presented in figures 3.1-6 and 3.1-7 respectively. As it can be found in table 3.1, sample C2014 contains a dose of 5MLs of GaN, while C2015 a dose of 18MLs.

The first thing one can notice is that the QDs in C2014, with a diameter of ~3.5nm are much smaller than the ones in C2015 (diameter of ~4.7nm). In fact, for sample C2015 the QDs are so large and close together that unless we look at greater magnification (fig. 3.1-7b), they appear as a single band.

The conclusions we can draw from the information we have here is the following: Firstly, and most importantly, it is confirmed that GaN QDs are formed using MBE as the method of growth, on amorphous SiO_2, nonetheless. The second conclusion we can make, with fair confidence, is that the size of the QDs is primarily affected by the dose of GaN. **The more GaN deposited, the larger the QDs that result from the growth process**. The last part of this study is how the density of the QDs is varied in different deposition conditions. What seems to be the case is that QD-density is affected in a more direct way by the rate of the deposition rather than the quantity of GaN. Namely, **as the deposition becomes slower, the density of the QDs seems to increase**.

These conclusions will be used to try to explain the behaviour of the devices later on in the study. However, it must be pointed out that what the generalizations we have made above, although sound and reasonable, cannot be completely confirmed unless more data becomes available and for that reason, the rest of the samples should be looked into.

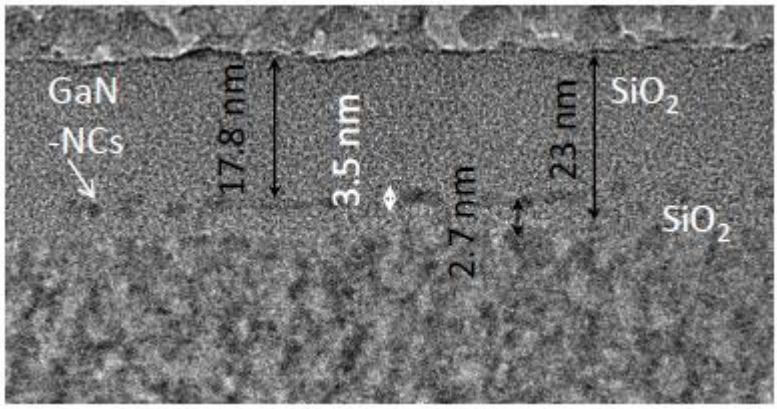


Figure 3.1-6: TEM images for sample C2014.

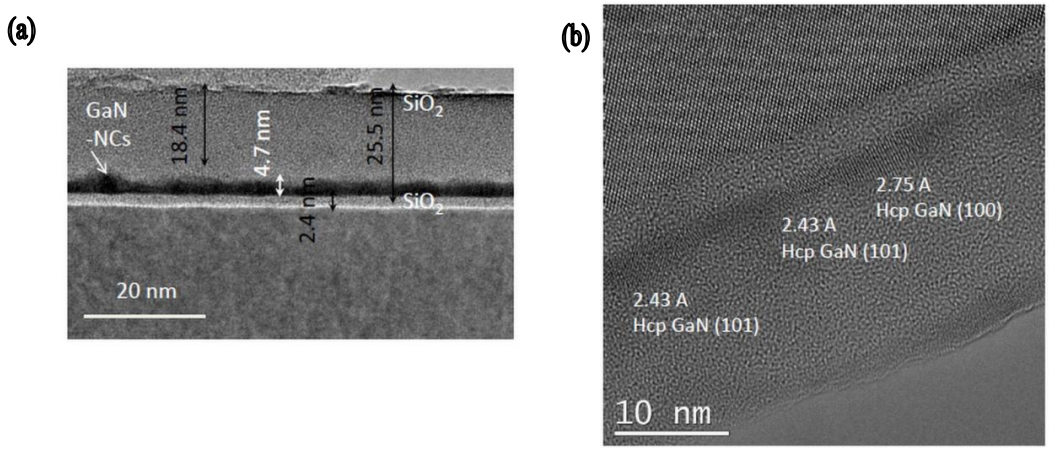


Figure 3.1-7: TEM images for sample C2015.

3.2 CAPACITANCE-VOLTAGE MEASUREMENTS

3.2.1 Experiment and Results

As it has already been explained in paragraph 2.2.2, two-terminal devices like the ones we described above can be used for a first evaluation of non-volatile memory characteristics. More specifically, the hysteresis of the C-V curve can give a quite reliable sense of the memory window that the floating gate (continuous or –in this case- nanocrystal) is responsible for.

For this reason, it follows naturally that the first step in characterizing the memory features of these devices is performing a series of C-V sweeps and observing the hysteresis. An Agilent 4284A Precision LCR Meter was used programmed with LabVIEW code. All the measurements were OPEN/SHORT corrected to eliminate as much of the systematic error as possible. In Appendix A, one can find more details on how this experimental setup works as well as a study on different impedance models that helped determine which models should preferably be used in our case.

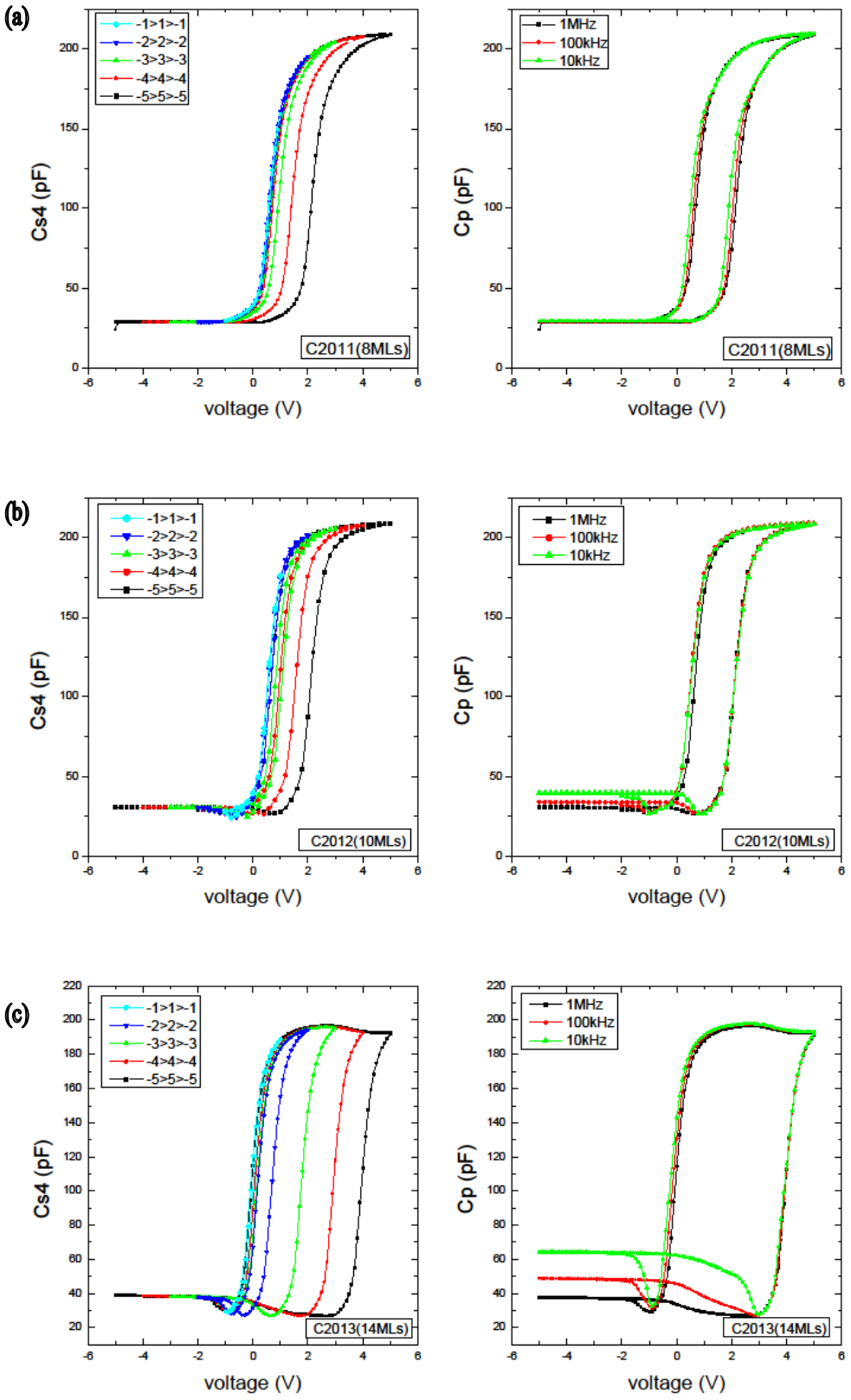
Each of the samples of table 3.1-1 was swept from inversion to accumulation and back to inversion at high frequency (~1MHz) with a step of 200mV:

- -1V→1V→-1V,
- -2V→2V→-2V,
- -3V→3V→-3V,
- -4V→4V→-4V, and
- -5V→5V→-5V.

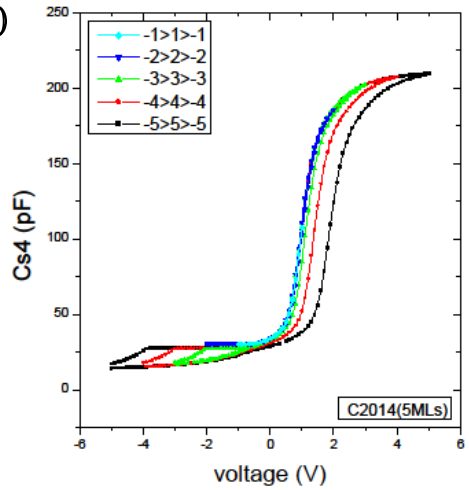
Sweeps were performed at different frequencies as well, 100kHz and 10kHz, to see how the samples behave at lower frequencies.

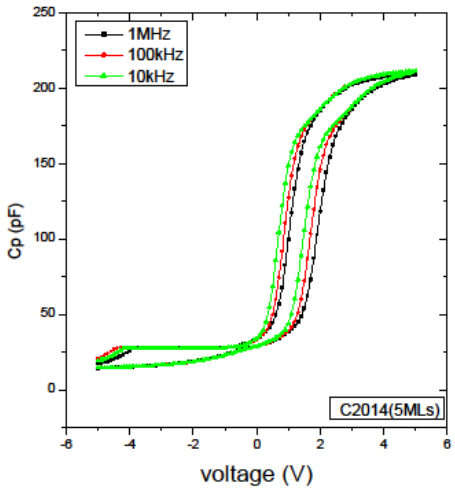
The resulting characteristics are given in figure 3.2-1 for all the samples, including the reference sample. The symbols used for the capacitance (C_p and C_{s4}) correspond to the respective circuit element of the impedance model that was used each time for the extraction of the capacitance. As it has been mentioned above, one can find a detailed description of those models and the manner in which they were used in the appendix.

The first thing of importance here is that it seems that significant hysteresis appears on the C-V characteristics of all the samples except on the one of the reference, the sample without nanocrystals. This implies that any memory attribute of the devices results from the presence of the GaN quantum dots. The reason this is such an important observation is because to create the GaN quantum dots on top of the SiO_2 used for the tunnel oxide, MBD uses a gallium source, as well as a nitrogen source. The presence of N could accidentally result it the formation of Si_3N_4. Silicon nitride can contain a large number of charge trapping sites, and, even though it is non-conducting in itself, it could act as a place for charge storage. This would mean that the devices would appear to possess memory properties but those would be due to the Si_3N_4 instead of the GaN quantum dots. The reference sample is exposed to the same deposition conditions as the rest of the samples; only the Ga source is blocked with the N source active. Therefore, if Si_3N_4 was present in any of the other samples, so it would be in the reference sample and we would see a hysteresis there as well as everywhere else. The fact that we do not means that **we can exclude Si_3N_4 as the origin of the charge storage**.

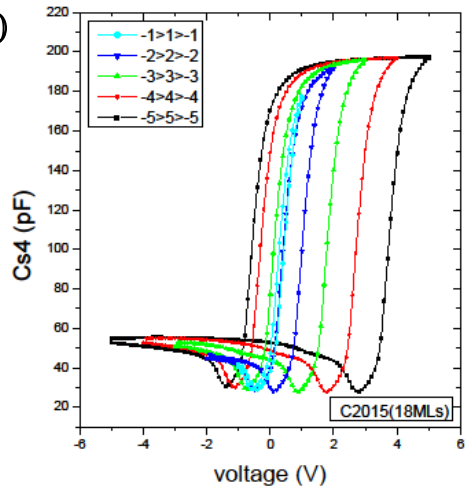

(a)
-1>1>-1
-2>2>-2
-3>3>-3
-4>4>-4
-5>5>-5
Cs4 (pF)
voltage (V)
C2011(8MLs)
1MHz
100kHz
10kHz
Cp (pF)
voltage (V)
C2011(8MLs)
(b)
-1>1>-1
-2>2>-2
-3>3>-3
-4>4>-4
-5>5>-5
Cs4 (pF)
voltage (V)
C2012(10MLs)
1MHz
100kHz
10kHz
Cp (pF)
voltage (V)
C2012(10MLs)
(c)
-1>1>-1
-2>2>-2
-3>3>-3
-4>4>-4
-5>5>-5
Cs4 (pF)
voltage (V)
C2013(14MLs)
1MHz
100kHz
10kHz
Cp (pF)
voltage (V)
C2013(14MLs)

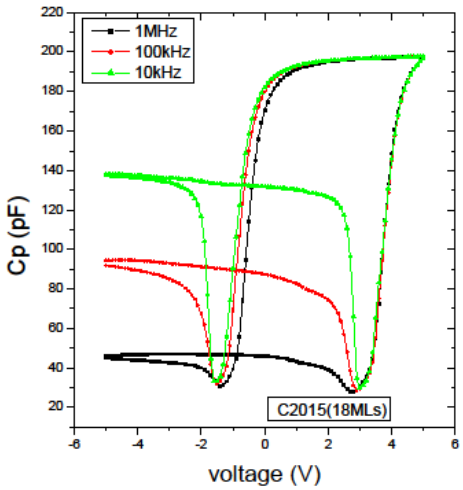
(d)


-1>1>-1
-2>2>-2
-3>3>-3
-4>4>-4
-5>5>-5
Cs4 (pF)
voltage (V)
C2014(5MLs)

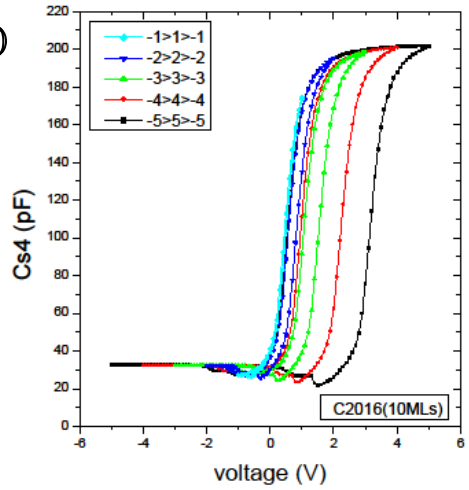

1MHz
100kHz
10kHz
Cp (pF)
voltage (V)
C2014(5MLs)

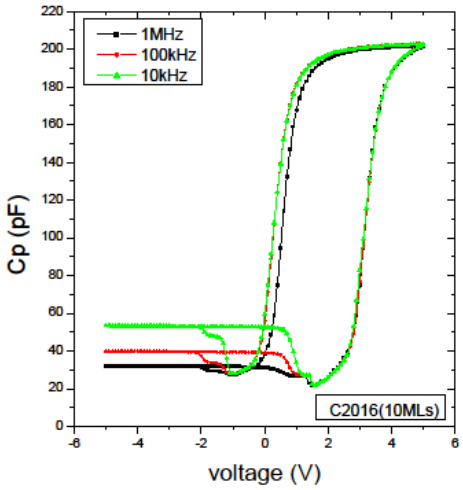
(e)


-1>1>-1
-2>2>-2
-3>3>-3
-4>4>-4
-5>5>-5
Cs4 (pF)
voltage (V)
C2015(18MLs)


1MHz
100kHz
10kHz
Cp (pF)
voltage (V)
C2015(18MLs)

(f)


-1>1>-1
-2>2>-2
-3>3>-3
-4>4>-4
-5>5>-5
Cs4 (pF)
voltage (V)
C2016(10MLs)


1MHz
100kHz
10kHz
Cp (pF)
voltage (V)
C2016(10MLs)

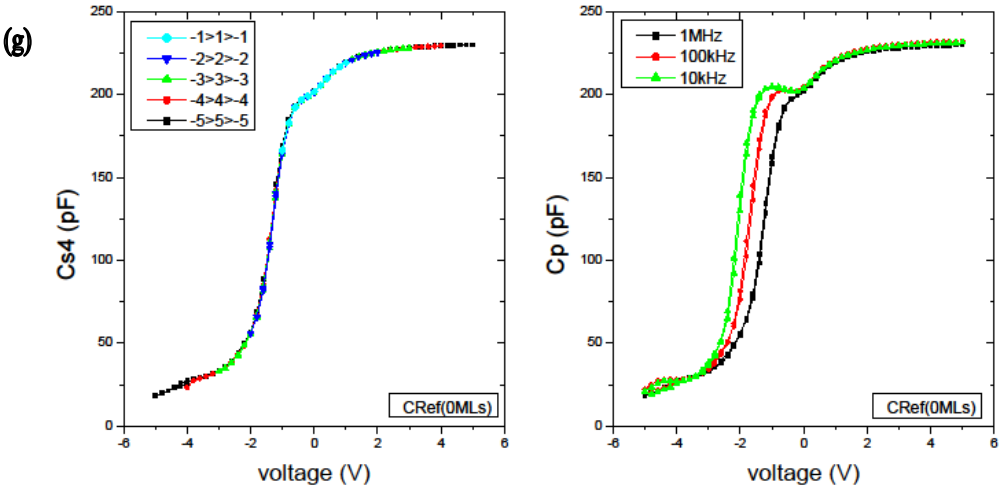


Figure 3.2-1: C-V characteristics for different sweeps (on the left) and for different frequencies (on the right) for samples C2011 (a), C2012 (b), C2013 (c), C2014 (d), C2015 (e), C2016 (f), and CRef (g).

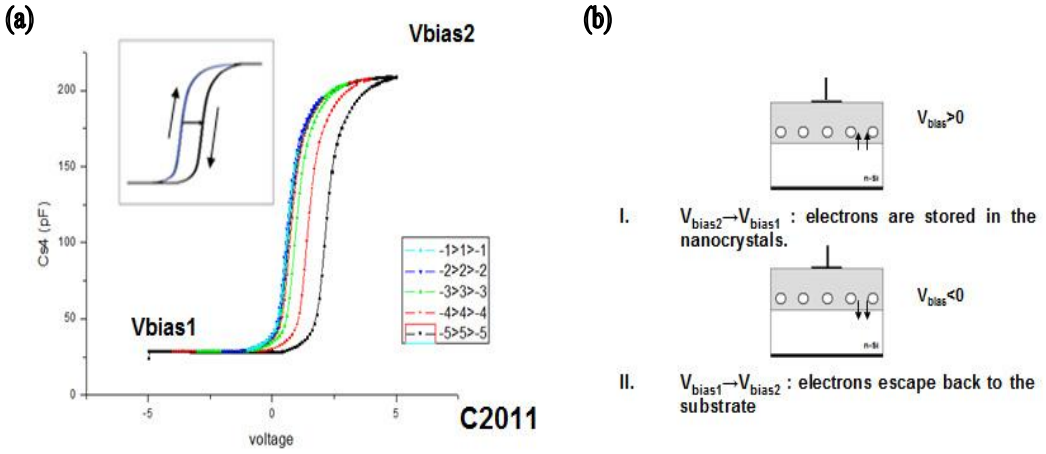


Figure 3.2-2: (a) The appearing hysteresis is clockwise for all the C-V curves. (b) Injection of electrons takes place from the substrate.

Secondly, we notice that as the voltage sweeps are getting wider and wider, the right branch of the characteristics moves to higher and higher voltages while the left branch stays relatively unchanged. As it has been explained in paragraph 2.2.2, when the flat-band voltage shifts to the right, it indicates the storage of negative charges. Since the left branch does not move considerably, we can assume that the storage of positive charges in the nanocrystals is inconsequential. Therefore, presumably **the carriers that are trapped in the QDs and that cause the hysteresis are primarily electrons**.

Another thing worth mentioning here is that the hysteresis on all of the characteristics is clockwise (fig 3.2-2a). According to what has been explained in paragraph 2.3.2 and figure 2.3-6, we can come to the conclusion that **carrier**

injection takes place from the substrate and not the gate. As figure 3.2-2b represents, assuming electrons as the injecting carriers, when the bias is positive, electrons move towards the gate, and they are injected from the substrate into the nanocrystals. Accordingly, when negative bias is applied, electrons will tunnel from the QDs back into the substrate, leaving the nanocrystal floating gate without charges.

3.2.2 Analysis and Discussion

A lot of information about a MOS capacitor can be extracted from its high frequency C-V characteristic. First of all, as it is described in detail in paragraph 2.1.2 the maximum capacitance in a C-V curve, which corresponds to the device being accumulated, is the capacitance it the oxide C_{ox}. Similarly, the minimum capacitance equals C_{inv}, which is the capacitance of the device during inversion. Of course, care must be taken in assigning values for C_{ox} and C_{inv}. Two examples are given in figure 3.2-3 for samples CRef (3.2-3a) and C2015 (3.2-3b), where it can be seen how C_{ox} and C_{inv} are chosen after the two-way sweep characteristics are split into their two separate branches.

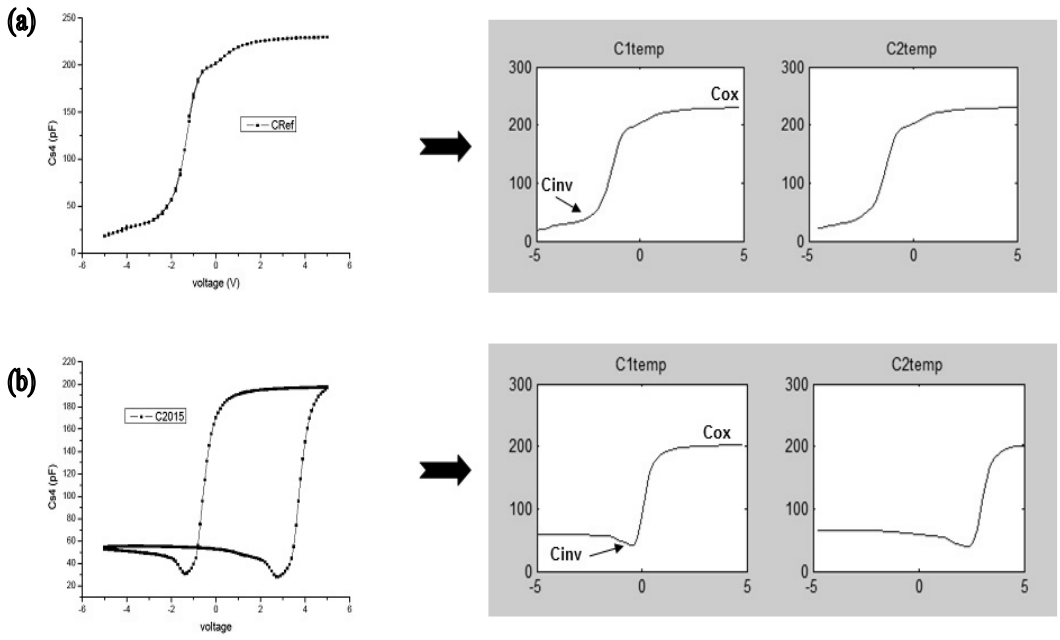


Figure 3.2-3: Assigning C_{ox} and C_{inv} for CRef (a) and C2015 (b).

The high frequency C-V curve can provide information on the doping density of an MOS capacitor with the maximum-minimum capacitance method. An empirical form of this method for silicon is used here:

$$\log(N_D) = 30.38759 + 1.68278 \cdot \log(C_1) - 0.03177 \cdot [\log(C_1)]^2 \text{, with:}$$

$$C_1 = \frac{RC_{ox}}{A(1-R)} \text{, and } R = \frac{Cinv}{Cox}.$$

N_D is the doping density of the substrate in units of cm^{-3} and A is the area of the capacitor in cm^2. Capacitances are given in units of F.

If the doping density is known, the Debye Length can be calculated:

$$L_D = \sqrt{\frac{\varepsilon_{Si} kT}{q^2 N_D}},$$

where T is the temperature in K, k Boltzmann's constant, q the electron charge and ε_{Si} the dielectric constant of silicon.

What is more, if the oxide capacitance is known then the oxide thickness t_i is easily extracted from the well known relationship:

$$C_{ox} = \frac{\varepsilon_{ox} A}{t_i} \Rightarrow t_i = \frac{\varepsilon_{ox} A}{C_{ox}},$$

where ε_{eff} is the effective dielectric constant of the oxide with the nanocrystals inside of it. The calculation of ε_{eff} can be rather complicated. In this study the effective dielectric constant was calculated according to the results presented in reference [4] of this chapter. In this calculation, we approximated the diameter D of the QDs with $D = \alpha x$, where x the number of GaN monolayers deposited for each sample and α the thickness of one monolayer of GaN: α = 2.9Å or α = 0.29nm. Additionally, the volume fraction of the nanocrystals in the oxide is approximated with $f = DA/(t_{tot} - D)A = D/(t_{tot} - D)$, where t_{tot} the total thickness of the oxide as shown in figure 3.1-1.

The flat-band capacitance can be calculated with two different methods. The first one is through the calculation of L_D and is explained in paragraph 2.1.2:

$$C_{fb1} = \frac{C_{ox} C_{fbs}}{C_{ox} + C_{fbs}}, \text{ with } C_{fbs} = \frac{\varepsilon_{Si} A}{L_D}.$$

The second one is an empirical equation for C_{fb}:

$$C_{fb2} = C_{ox} \left(\frac{1}{1 + \frac{136}{t_i} \sqrt{\frac{T/300}{N_D}}} \right),$$

where N_D in cm^{-3} and t_i in cm.

Both of those equations where used with remarkably similar results for the flat-band capacitance.

After the C_{fb} has been found, it should be straightforward to extract the flat-band voltage from the left and the right branch of the C-V curve and calculate the memory window as their difference (see also figure 2.2-5):

$$MemoryWindow = |\Delta V_{FB}| = |V_{FB}^1 - V_{FB}^2|.$$

For all of the above calculations, matlab code was written that can be found in Appendix B (extracts 3 and 4).

The results from the analysis that was described above for all the samples are given in table 3.2-1 and the memory windows plotted versus the range of the sweeps can be seen in figure 3.2-4.

	C2011	C2012	C2013	C2014	C2015	C2016	CRef
Cox(pF)	209	207	204	211	202	203	230
Cinv(pF)	31	35	39	28	41	28	32
Nd(cm^{-3})	$4.18*10^{15}$	$5.63*10^{15}$	$7.86*10^{15}$	$3.20*10^{15}$	$8.90*10^{15}$	$3.15*10^{15}$	$4.25*10^{15}$
Ld(m)	$6.38*10^{-8}$	$2.83*10^{-8}$	$4.65*10^{-8}$	$7.28*10^{-8}$	$4.36*10^{-8}$	$7.35*10^{-8}$	$6.32*10^{-8}$
ti(nm)	27.9	28.3	29.0	27.5	29.7	28.8	25
Cfb1(pF)	119	126	133	113	135	110	126
Cfb2(pF)	119	126	133	112	135	110	126
ε	3.9216	3.9386	3.9839	3.9049	4.0366	3.9386	3.9

Table 3.2-1: Calculated values for all of the samples.

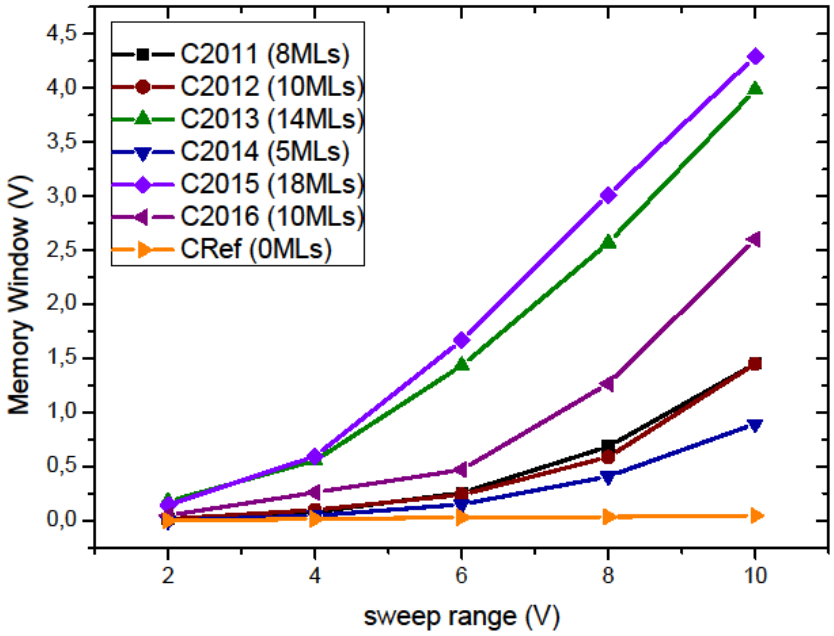


Figure 3.2-4: Memory Window vs the range of the C-V sweep for all the devices.

At this point, it should be noted that there is a second way to extract the memory window of such a device. To perform the C-V measurement, voltage is applied on the gate of the MOS capacitor. As the voltage sweeps from negative to positive values and back, the change of the impedance is monitored by the LCR meter. This sweep, not only affects the imaginary, but also the real part of the impedance. Depending on the model that is used for that impedance, this translates into change for every part of the model accordingly (also see Appendix A). If one assumes a conductance that is connected in parallel with what would be the theoretical capacitance of the device, then as the voltage changes so does the value of this parallel conductance. In particular, it should exhibit peaks close to the flat-band voltage. It follows, that if V_{FB} is different during the voltage sweep forth and the one back, so the peaks of the

conductance will appear accordingly. Therefore, to estimate the memory window, one has but simply to 'measure the distance' of the two peaks in volts. A couple of examples are given in figures 2.3-5a and 2.3-5b for sample C2015.

Although this method was used in a few cases to confirm its validity and even though the results were very similar to the ones in figure 2.3-4 (see fig. 2.3-5c) it is not preferred here because its accuracy depends a lot more on the step with which the voltage sweep is performed and as a result it does not render results as accurate.

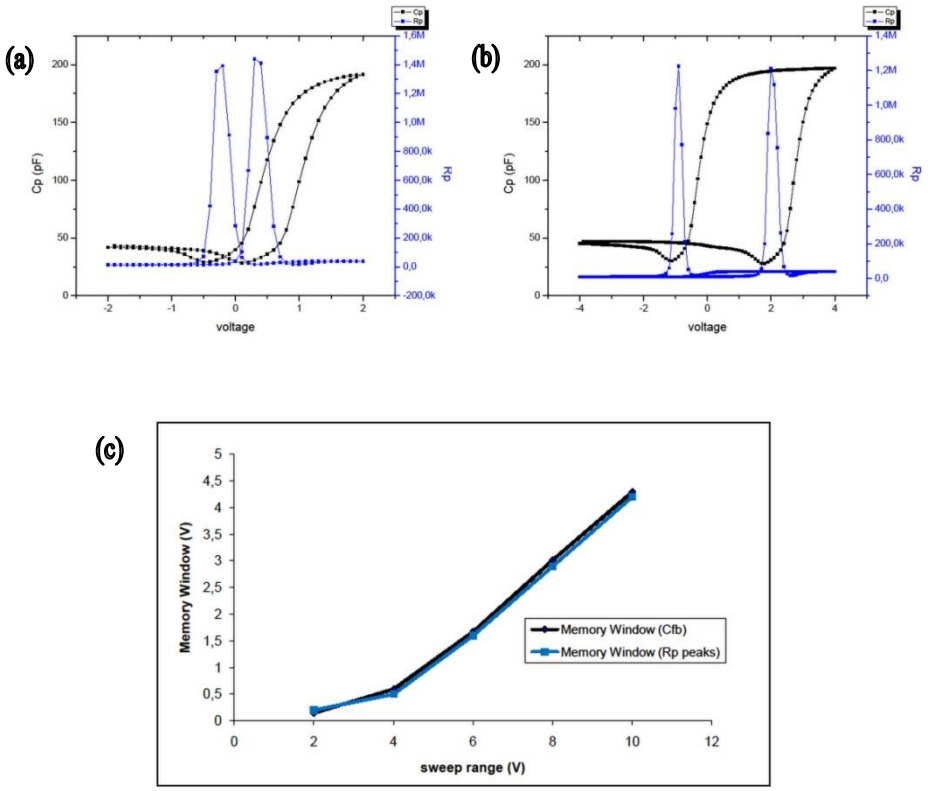


Figure 2.3-5: Extraction of the memory window from the peaks in the parallel conductance: (a), (b) Examples of conductance-voltage characteristics. (c) Comparing the results with the ones extracted from the C-V curves.

Lastly, we can use the memory window extracted from the C-V curves to calculate the surface density of the charge that is stored in each device, which should give us an idea of which sample can hold more charge. For this calculation, based on the discussion in the previous chapter, the following approximation was used:

$$\Delta V_{FB} = \frac{Q}{C_{CO}} = \frac{Q t_{CO}}{\varepsilon_{ox}\varepsilon_0 A} \Rightarrow \frac{Q}{A} = \frac{\varepsilon_{ox}\varepsilon_0 \Delta V_{FB}}{t_{CO}}.$$

These calculations are presented in table 3.2-2, and figure 3.2-6, where one can find the charge density, as well as the total number of electrons that each sample holds in the nanocrystals. It is noted that the measured capacitance corresponds to an area of $410\times410\mu m^2$.

From all of the above, it is obvious that the general trend that is followed is, the higher the quantity of GaN deposited during device fabrication the largest the memory window it exhibits on the C-V characteristics. The only slight deviation from this

observation is between samples C2011 with 10MLs GaN and C2012 with 8 MLs of GaN. Those two have memory windows very close to each other but it should be noted that the quantities of GaN are also quite similar. It also becomes apparent if we examine table 3.2-2 that sample C2015, which contains the greatest amount of GaN can hold the most charge on its nanocrystals, while the reference sample, with no quantum dots, holds practically no electrons. Again, the only sample that deviates from this observation is C2012.

Assuming, from what is known (paragraph 3.1.3) that the dose of GaN affects the size of the quantum dots (more GaN results in larger QDs), whereas the QD density depends on the deposition conditions, a possible explanation for these results could be the following:

A larger QD can hold more carriers than a smaller one, which spatially makes sense, and that is the reason why samples C2013 and C2015 with the largest doses of GaN can clearly hold the most charge. As the QD size gets smaller, however, the Coulomb Blockade phenomenon becomes much more prominent: its effect is inversely proportional to the diameter of the QD. Thus, it becomes important for smaller QDs, while it can be negligible for larger ones. Therefore, for the samples with the lower GaN and, subsequently, with the smaller similarly sized QDs, density could play a much more important role than size. This could explain why sample C2016 exhibits so much more storage ability than C2012. They might both have QDs of similar sizes that can hold the same number of electrons each, but there are much more QDs on C2016 with was realized with a lower rate of deposition.

However, to come to safe conclusions, more information on the size and density of the QDs is essential. For this reason, TEM measurements are due for all of the samples and this will either confirm of discard this hypothesis.

GaN Deposited	**Maximum ΔV_{FB} (-5→5→-5)**	**Charge/unit area (nC/cm^2)**	**Electrons/unit area (e/cm^2)**	**Total number of electrons on a device**
5MLs (2C2014)	0.8935	176.303	$1.100*10^{12}$	$1.85*10^{9}$
8MLs (C2011)	1.4508	286.269	$1.789*10^{12}$	$3.01*10^{9}$
10MLs (C2012)	1.4487	285.854	$1.784*10^{12}$	$2.99*10^{9}$
14MLs (C2013)	3.9832	785.956	$4.906*10^{12}$	$8.25*10^{9}$
18MLs (C2015)	4.2937	847.223	$5.289*10^{12}$	$8.89*10^{9}$
10MLs (C2016)	2.6016	513.342	$3.204*10^{12}$	$5.38*10^{9}$
0MLs (CRef)	0.0448	8.842	$0.005*10^{12}$	$0.09*10^{9}$

Table 3.2-2: Calculated charge density for the devices.

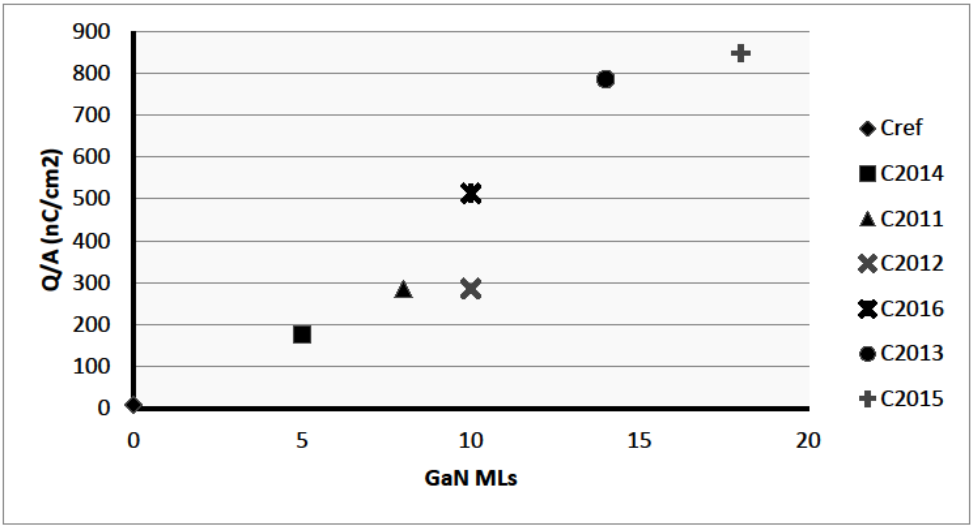


Figure 2.3-6: Charge Density in the devices, as it is affected by the quantity of the deposited GaN.

3.3 PULSE MEASUREMENTS

3.3.1 Experiment and Results

So far, C-V measurements have helped in providing with a first sense of how much charge can be held in the nanocrystals of these devices. However, the charging and discharging (trapping/detrapping) rely on tunnelling mechanisms that are time-dependent and, thus, they are seriously affected by the voltage sweep conditions. Additionally, in memory applications the program, erase and read operations are performed by applying square voltage pulses on the gate of the device. It follows, therefore, that we cannot rely on C-V measurements for a precise picture of how the charging and the discharging works for these devices.

For the reasons described above, a series of pulses was applied on the devices to study their behaviour. The procedure is as follows:

First, a positive voltage pulse is applied on the device. This should result in some electrons being stored in the nanocrystals, and this should in turn mean that the flat-band voltage of the MOS capacitor is higher than before. Consequently, if a high frequency C-V sweep is performed it should appear shifted compared to a sweep on an uncharged device. Again, if a second pulse is applied, this should move the C-V curve even higher. If the range of the sweep is short enough, one can make the hypothesis that the C-V itself does not affect the characteristic. The shift of V_{FB} depends on both the width and the height of the applied pulse.

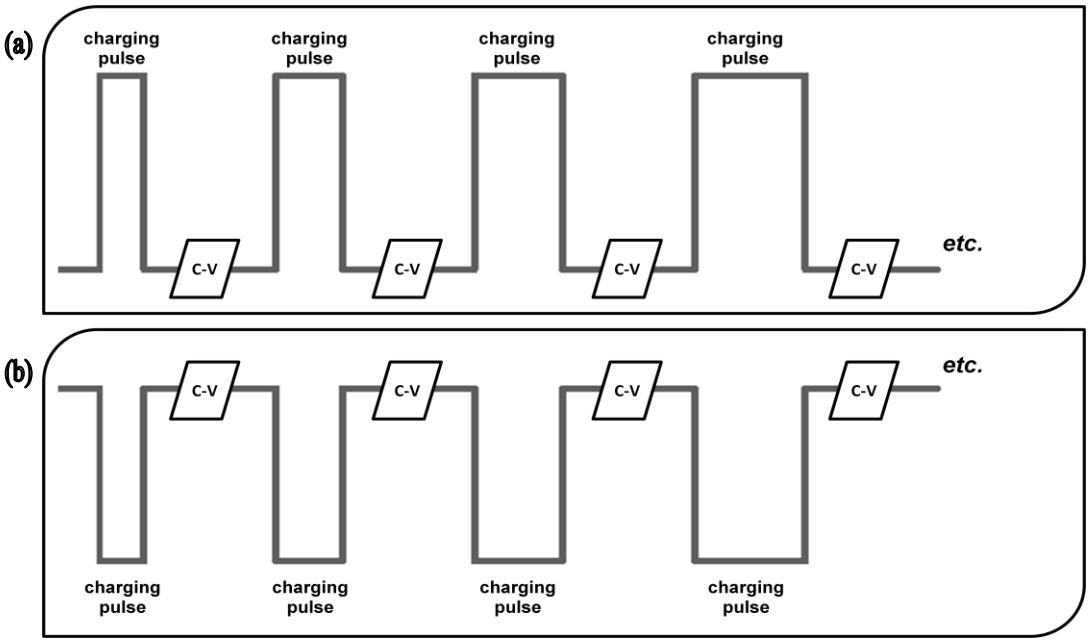


Figure 3.3-1: The voltage pulse sequence, including the high-frequency C-V measurements, that was used that was used for (a) positive pulses (charging of the capacitors) and (b) negative pulses (discharging).

Following this course of thought, a series of positive voltage pulses was applied on each of the sample, starting with an uncharged device. These pulses had fixed height but varied in width: 200μs, 1ms, 10ms, 100ms, 500ms, 1s. Each pulse was followed by a narrow high-frequency C-V sweep to help extract the V_{FB} from. Next, the amplitude of the pulse was increased and another set of pulses of different widths, as

described above, was applied -each with the increased height- with C-V sweeps as narrow as possible following each pulse, and so on (fig. 3.3-1a).

This process causes the C-V characteristic to shift to higher and higher voltages, which means V_{FB} increases and a larger memory window occurs. A couple of examples are given in figure 3.3-2 to show the shifting of the C-V curve.

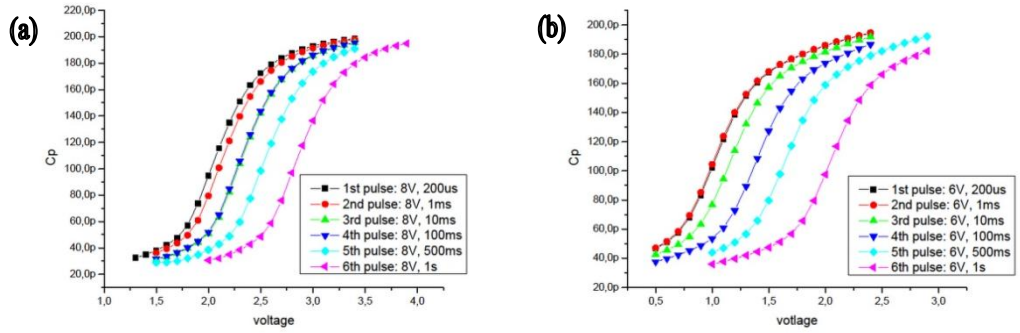


Figure 3.3-2: Examples of how the C-V curve shifts during pulse measurements.

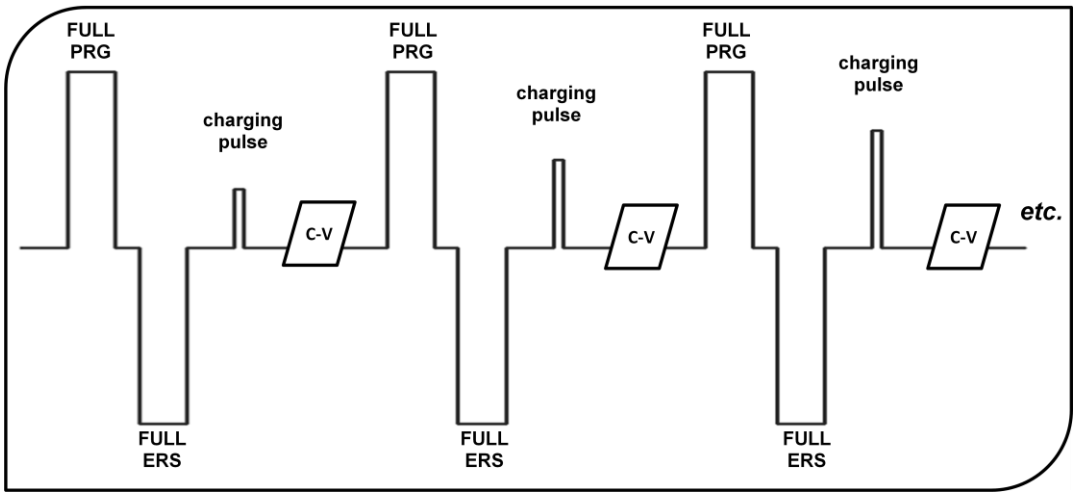


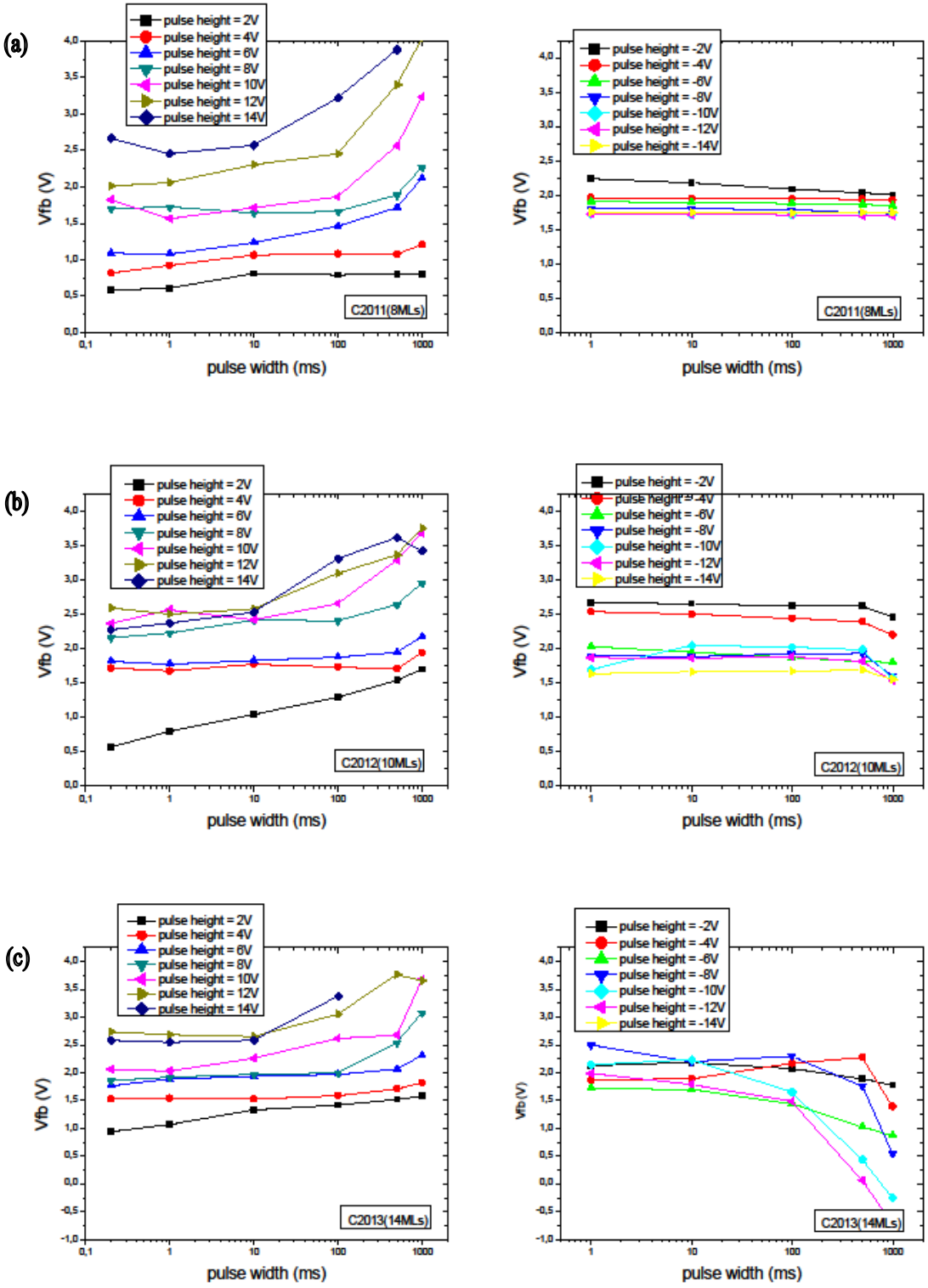
Figure 3.3-3: The voltage pulse sequence usually used to help bring the device to its initial state before applying the charging pulse.

It is noted that the flat-band capacitance for each device was extracted from a wider C-V sweep as described in the previous paragraph in more detail, and this value was used on all C-V curves for the extraction of V_{FB}.

In total, on each sample 7 series of pulses were performed: The heights of each series were 2V, 4V, 6V, 8V, 10V, 12V, and 14V, and each series comprised of 6 pulses with widths that varied from 200μs to 1s as mentioned above. Pulses were applied with a Keithley 2602A System Sourcemeter and again C-V sweeps were performed by an Agilent 4284A Precision LCR Meter.

A similar procedure was followed to observe how the discharging takes place. This time, the device was first charged with a long-lasting voltage pulse of 10s. 10V were used on all of the samples, except on C2016 which did not survive voltage as high, in which case an 8V pulse was preferred. After each device was charged, as with the

charging pulses, we applied a series of voltage pulses of varying widths (1ms, 10ms, 100ms, 500ms, and 1s) but of negative amplitudes: -2V, -4V, -6V, -8V, -10V, -12V, and -14V (fig. 3.3-1b). Again, high-frequency narrow C-V followed each pulse and V_{FB} was extracted.



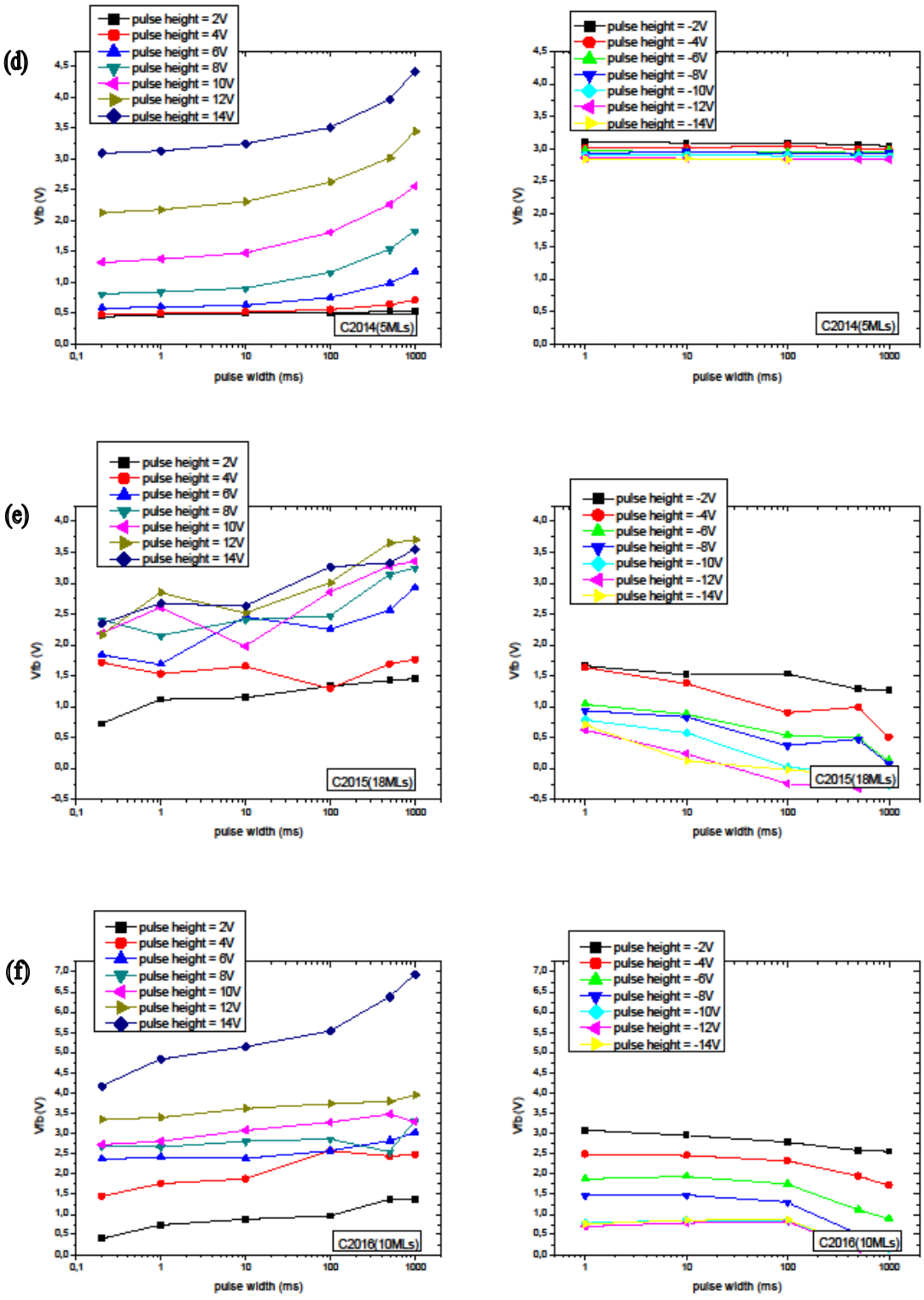


Figure 3.3-4: Pulse Measurements: V_{FB} vs pulse width for (a) C2011, (b) C2012, (c) C2013, (d) C2014, (e) C2015, and (f) C2016.

All C-V sweeps during this experiment were performed at 1MHz, with a step of 100mV. The Matlab code that was written for the extraction of V_{FB} is given in Appendix B (extract 5).

It should be pointed out here that the aim of these measurements is to discover how pulses of different heights and widths affect the charge that is stored in the floating gate. Therefore, it would be more useful if the device returned to the uncharged (erase) state before each program pulse is applied. For this reason, the most common way to perform such tests is to apply a set of pulses that will ensure the device returns to its original state, then apply the charging pulse and, lastly, do a C-V sweep to see how it was affected (fig. 3.3-3). However, as it will also become apparent in the following discussion, these devices present us with a difficulty in erasing them. This makes it difficult to be certain that after the FULL PROGRAM/FULL ERASE pair of pulses that are shown in figure 3.3-3 are applied the device will each time return to the initial state. Thus, the simpler setup of figure 3.3-1 was preferred but one should keep in mind that in our case, each time the state of the device depends not only on the charging pulse, but also on its previous state.

The resulting V_{FB} – pulse-width plots are given in figure 3.3-4 for all of the devices. Applying positive pulses yielded the plots on the left and, for negative pulses, plots on the right were obtained.

We should comment at this point, that the problem with such measurements, apart from the fact that each measurement depends on all the previous ones, is that the measured flat-band voltage, is inevitably affected by the voltages applied during the C-V sweep. In particular, the human factor plays an imperative role here and is very difficult to eliminate, since the range of each sweep is decided by the person performing the experiment. The results presented above can provide us with important conclusions; however, despite the effort that was made to minimize any dependence on subjective decisions, one can see there are still some inconsistencies.

3.3.2 Discussion

It was concluded from the capacitance-voltage measurements that the amount of GaN deposited on the sample during fabrication, plays a critical part in affecting the memory qualities of the devices; in general, the more GaN, the wider the memory window, and the greater the amount of charge that each device can hold in the nanocrystals. It was presumed that this is due to the fact that more monolayers of GaN mean larger quantum dots that can hold more electrons each, than smaller ones.

This can also be seen from results from the pulse measurements. Figure 3.3-5 compares how each sample responds to the same set of pulses. In 3.3-5a, positive pulses, 6V each, of different durations are applied whereas in 3.3-5b the same comparison for a set of negative pulses (-6V) of different widths is presented. It becomes apparent that, with the same set of pulses, it is much easier to charge those devices with more MLs of GaN deposited, such as C2015 and C2013. Samples with less MLs of GaN, such as C2014 and C2011 are more difficult to charge.

It has been mentioned before, and it now becomes clearer with these results, that discharging these devices is a lot more difficult than charging them. This is true especially for those samples with low GaN deposited. If we take, for example, C2014 with 5MLs of GaN, we can see in figure 3.3-5 that as wider and wider pulses are applied, charge is stored in the QDs and V_{FB} is shifted to higher values, the difference reaching approximately ~0.5V from the narrowest to the wider pulse. If, however, we look at how negative pulses affect a charged device of the same sample, V_{FB} stays practically unchanged for the set of applied pulses ($\Delta V_{FB} \approx 0V$).

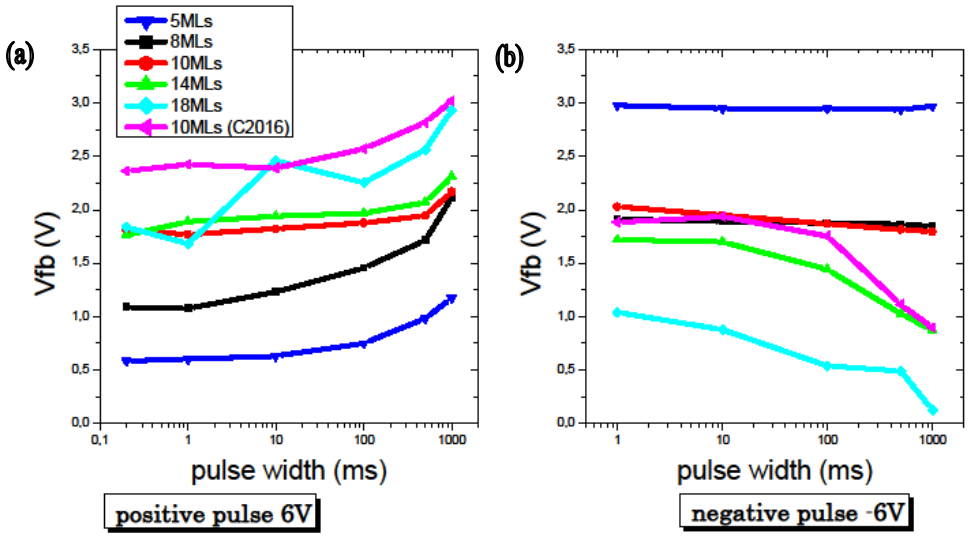


Figure 3.3-5: (a) The effect of a series of positive pulses of amplitude 6V on the samples, (b) The effect of a series of negative pulses of amplitude -6V on the samples.

We can imagine the reason why this happens if we take another look at the band diagram of the device (fig. 3.3-6). Again, this does not take into account quantum confinement effects on the energy gap. In any case, the differences would be minuscule and it would not prove any more illustrative than this.

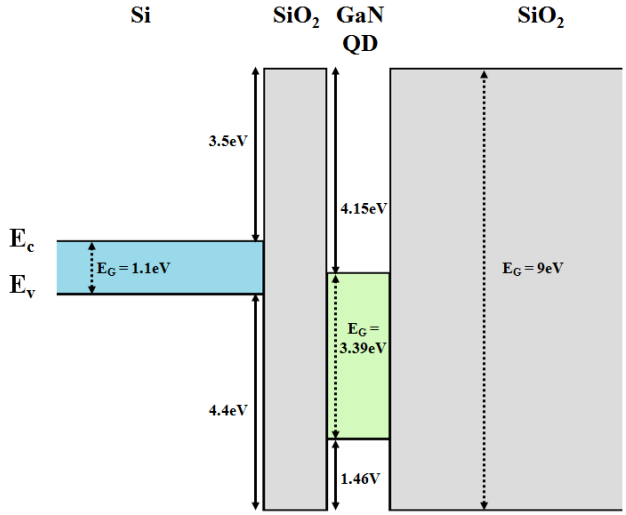


Figure 3.3-6: Band Diagram for the GaN QD memory devices, based on the data found in reference [4].

It is plain form figure 3.3-6 that electrons have to overcome an energy barrier of 3.5eV to be stored in the QDs. Once they enter the nanocrystals, however, a higher barrier of ~4.1eV has to be overcome for them to return to the substrate, making discharging harder. Another way for V_{FB} to return back to erase would be if holes

were injected form the silicon into the nanocrystals, thus reducing the total amount of charge. For that to happen, however, holes have to tunnel through an even higher barrier of ~4.4eV which makes this latter case unlikely as well, leaving erasing the device as a challenge.

Samples with higher amount of GaN seem to be easier to discharge. There is, however, one more factor that one needs to take into account. Samples with more GaN seem to be a lot more unpredictable in their behaviour with the pulse measurements. Devices with lower doses appear to yield a lot more consistent results and are a lot more stable, not being easily affected by the C-V sweep that follows each pulse. This should not be a very surprising fact and it could be explained by the quantum dot formation process.

For a fixed MBD deposition rate of Ga and N, the larger the desired dose of GaN, the more time is necessary. However, a quantum dot can start nucleating any time until the deposition is over, and, the sooner the nucleation starts, the larger the quantum dot will be. This means that the longer the deposition takes, the more dispersed the quantum dots will be in terms of their size. It is understood then that with a low dose of GaN the size of the quantum dots will have a distribution less dispersed than what would result from a longer deposition (figure 3.3-7). It is obvious that if the quantum dots are not very uniformly shaped, they will start behaving more unpredictably.

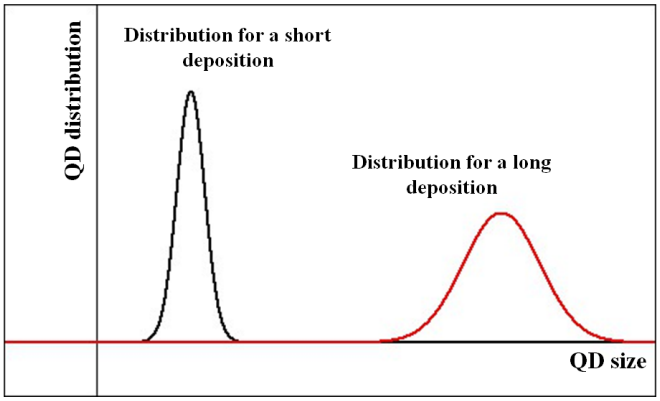


Figure 3.3-7: The duration of the deposition affects the quantum dot size.

Lastly, plots of the flat-band voltage versus the pulse height are presented in figures 3.3-8 and 3.3-9 for overview.

Of course, it has to be clear that experiments on such devices are still quite young and it is early in research to know what happens concerning charge storage and escape mechanisms. As it has been discussed, we can safely say that electrons are the charge carriers and that injection takes place from the substrate. To fully understand the mechanisms, however, further study is needed. For the moment, to investigate hole injection better, a set of samples will be soon developed on an acceptor-doped silicon substrate. Despite all the uncertainties, there is promise that if this part of the operation is understood, we will be able to develop fast devices that operate on relatively low voltages.

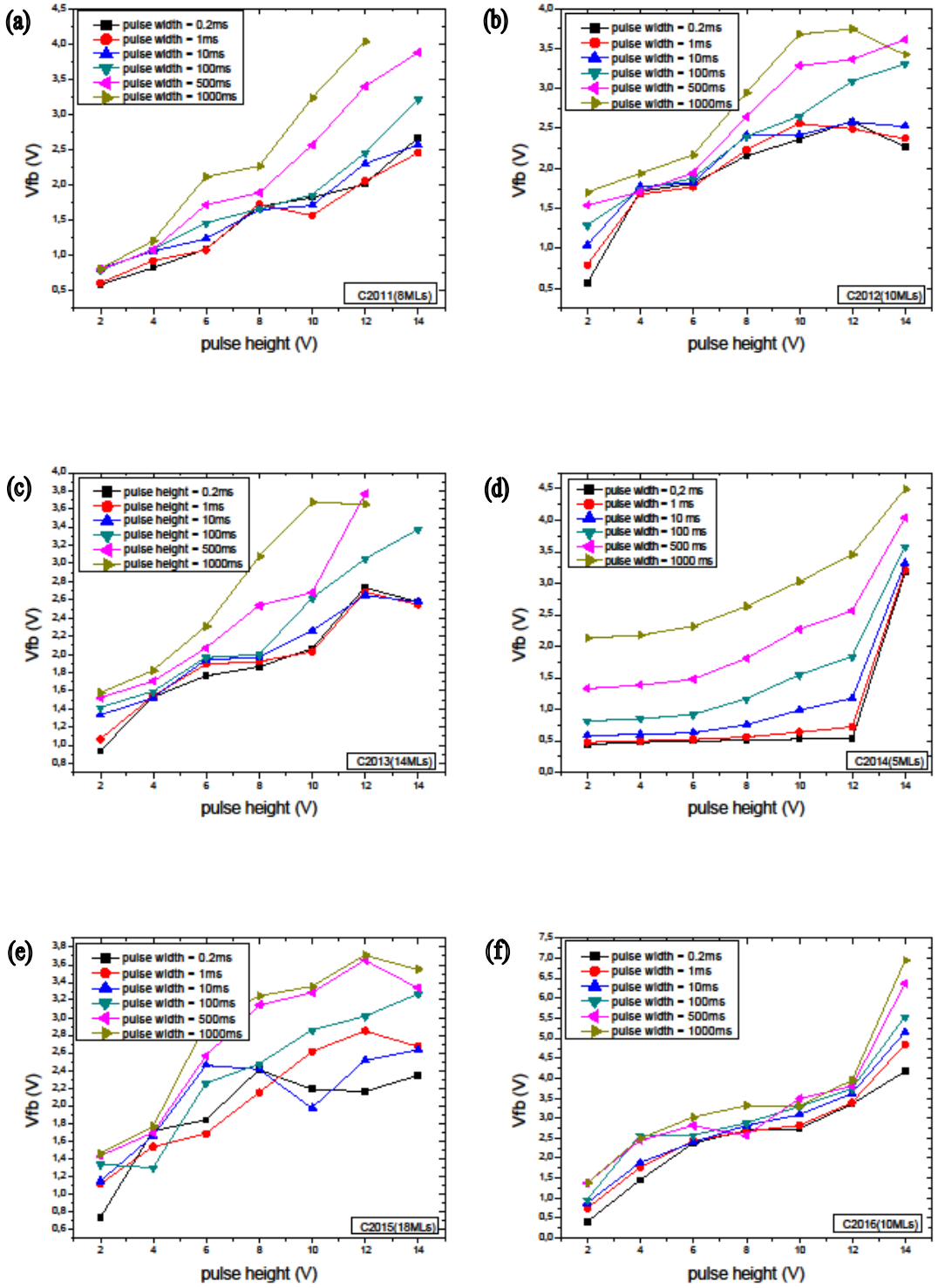


Figure 3.3-8: V_{FB}-pulse height plots for the charging of uncharged devices: (a) C2011, (b) C2012, (c) C2013, (d) C2014, (e) C2015, and (f) C2016.

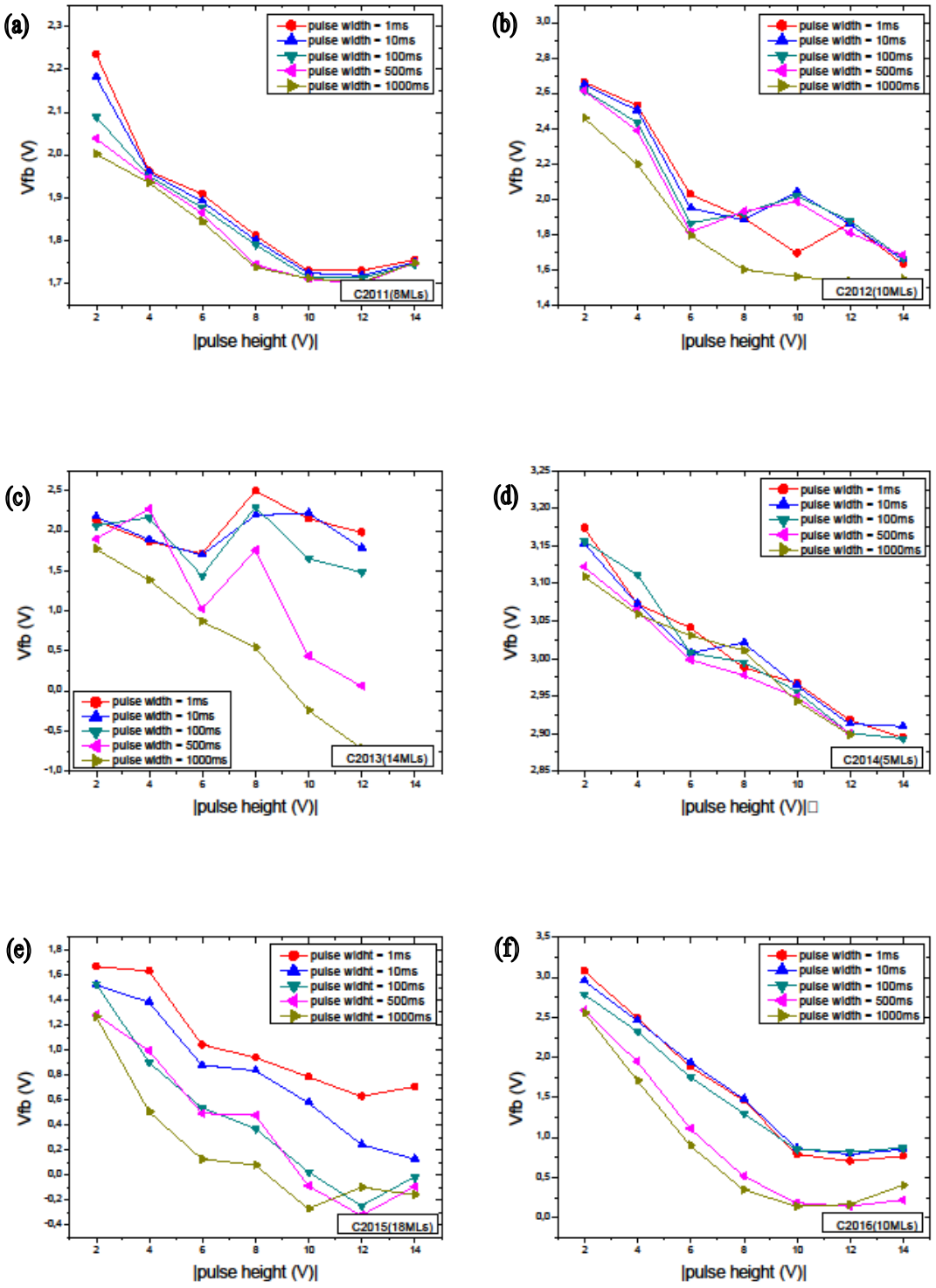


Figure 3.3-9: V_{FB}-pulse height plots for the discharging of charged devices: (a) C2011, (b) C2012, (c) C2013, (d) C2014, (e) C2015, and (f) C2016.

3.4 CURRENT TRANSIENT MEASUREMENTS

3.4.1 Experiment and Results

The samples were subjected to charge retention measurements at room temperature by P. Dimitrakis and P. Normand at the Institute of Microelectronics at the National Centre for Scientific Research 'Demokritos'. The results of these measurements are presented in figure 3.4-1.

The reader might recall from paragraph 1.2.2 that V_{th} (and, subsequently, the memory window ΔV_{FB}) of memory devices is expected to have a dependence on time that is described by the following expression:

$$V_{th} = K_1 - K_2 \log(t).$$

From figure 3.4-1, it becomes apparent that the retention characteristics of the devices seem to be more complicated than this. This indicates that there might be more than a single discharging mechanism involved. If we assume that each process which causes charge to be lost affects the charge retention in a different way, this might explain why the patterns in figure 3.4-1 are observed.

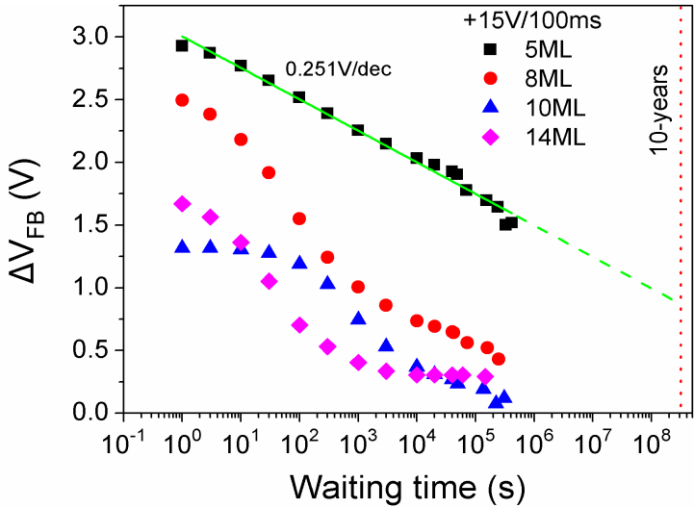


Figure 3.4-1: Retention measurements at room temperature as were obtained by Dimitrakis, P., and Normand, P. at IMEL/NCSR Demokritos.

In order to look further into the possibility of multiple mechanisms, a series of transient measurements were performed. The current that flows through the device at a certain bias was monitored over a period of time. Transient measurements were executed with a Keithley 4200 Semiconductor Characterization System at several bias voltages for each sample, both positive and negative. The dc bias voltage was applied at the gate of the device (anode) and the current was measured at the cathode of the setup (see figure 3.4-2) while it remained grounded.

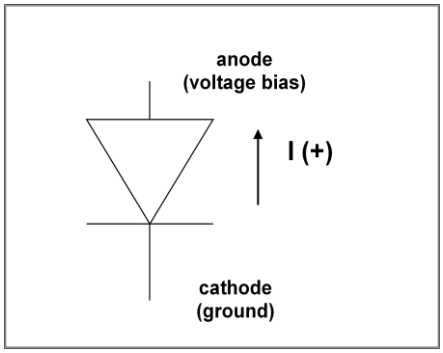


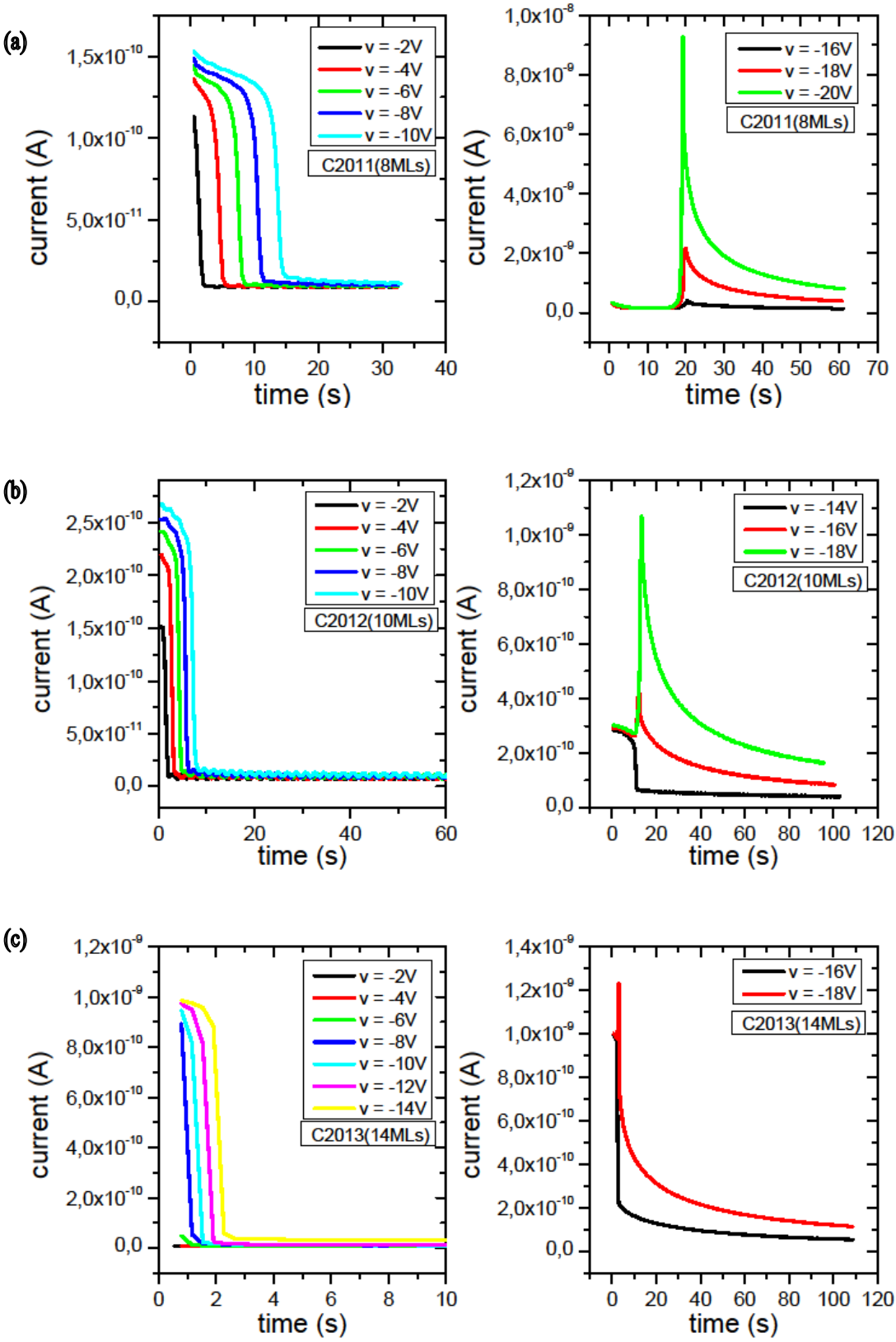
Figure 3.4-2: DUT bias conditions and current flow for the transient measurements.

Results from these measurements are presented in figures 3.4-3 and 3.4-4. It is noted that there appears a slight rippling of the current value due to errors introduced by the experimental setup. On the same note, there also is background noise of the order of 10^{-11}A which corresponds to a practical value of 0A, and this is the value that all of the transient characteristics seem to approach asymptotically for steady-state.

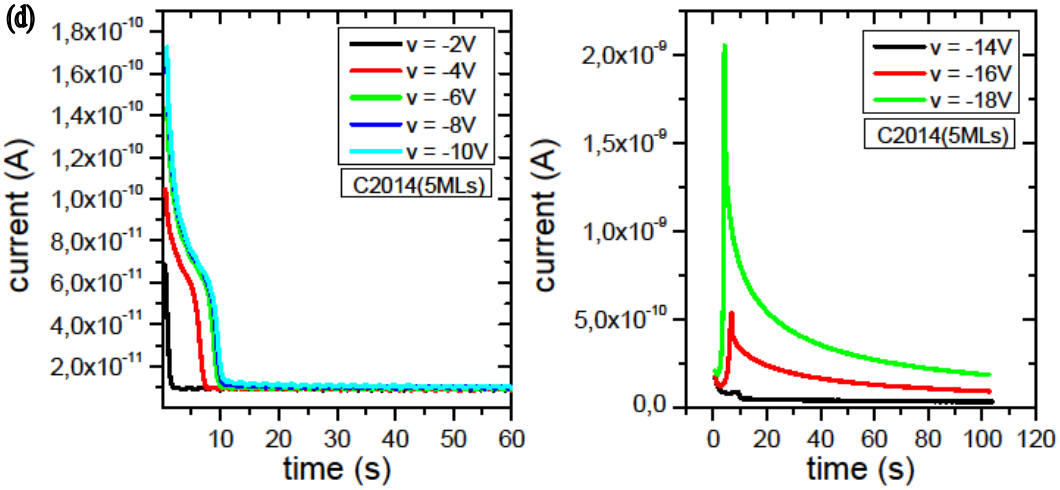
3.4.2 Discussion

Let us begin by taking a closer look at figure 3.4-1. The first thing one notices, as has been described in the previous paragraph, is that the retention characteristics diverge from their expected behaviour. This makes it difficult to obtain an accurate prediction for the memory window at the ten-year threshold for non-volatility. In fact, out of all the samples, only the one with the lowest amount of GaN (C2014) could give a possible estimate. The pattern that arises, however, is that as the amount of GaN embedded in the oxide increases, the charge seems to be escaping faster. This agrees with what is described in paragraph 3.3 as far as the pulse measurements are concerned. We saw there that devices with more layers of GaN were more unpredictable in how charge is stored. If, therefore, it is difficult to erase a device without much GaN, it could also be possible that this device holds charge better in time than another device, maybe one with more GaN and also one that can be erased with lower voltages.

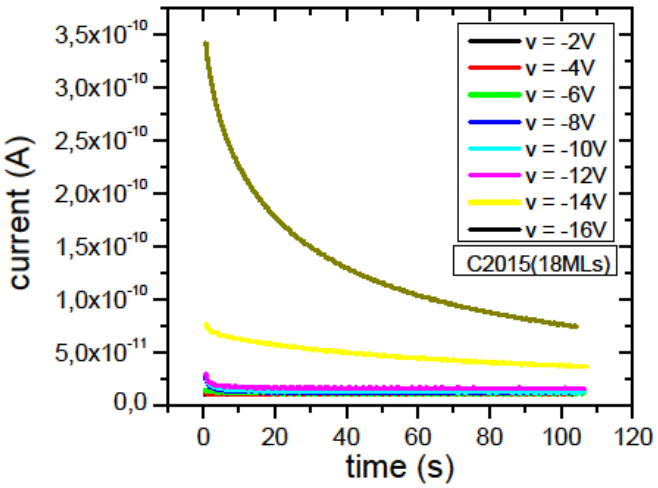
From figure 3.4-3, where transient characteristics for negative bias voltages are presented, one can see that, for most of the devices, the curves deviate from what a conventional capacitor is expected to yield as its transient response. The transient voltage for a capacitor for a simple case would follow an exponential curve which would asymptotically approach the bias voltage applied. The current would approach the value 0, also at an exponential rate. In our case, however, when a negative voltage is applied, a significantly different behaviour is observed (fig. 3.4-3a, b, c, and d). More specifically, the current initially appears to drop to a value higher than zero and retains it for a few moments before it starts to decrease again until it reaches its final value of practical 0.


(a)
current (A)
time (s)
v = -2V
v = -4V
v = -6V
v = -8V
v = -10V
C2011(8MLs)
v = -16V
v = -18V
v = -20V
C2011(8MLs)
(b)
v = -2V
v = -4V
v = -6V
v = -8V
v = -10V
C2012(10MLs)
v = -14V
v = -16V
v = -18V
C2012(10MLs)
(c)
v = -2V
v = -4V
v = -6V
v = -8V
v = -10V
v = -12V
v = -14V
C2013(14MLs)
v = -16V
v = -18V
C2013(14MLs)

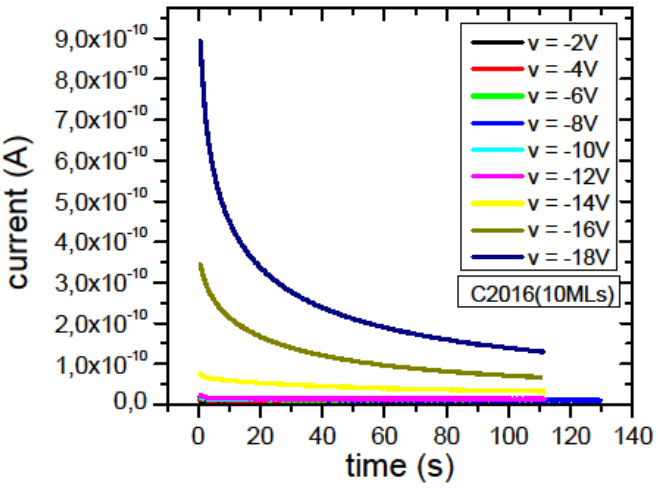
(d)


current (A)
time (s)
v = -2V
v = -4V
v = -6V
v = -8V
v = -10V
C2014(5MLs)
v = -14V
v = -16V
v = -18V
C2014(5MLs)

(e)


current (A)
time (s)
v = -2V
v = -4V
v = -6V
v = -8V
v = -10V
v = -12V
v = -14V
v = -16V
C2015(18MLs)

(f)


current (A)
time (s)
v = -2V
v = -4V
v = -6V
v = -8V
v = -10V
v = -12V
v = -14V
v = -16V
v = -18V
C2016(10MLs)

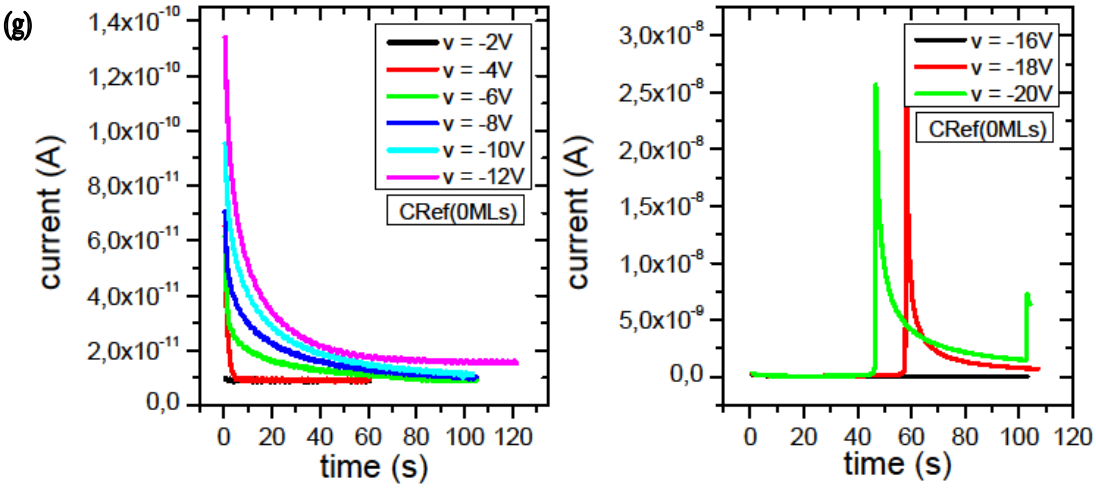


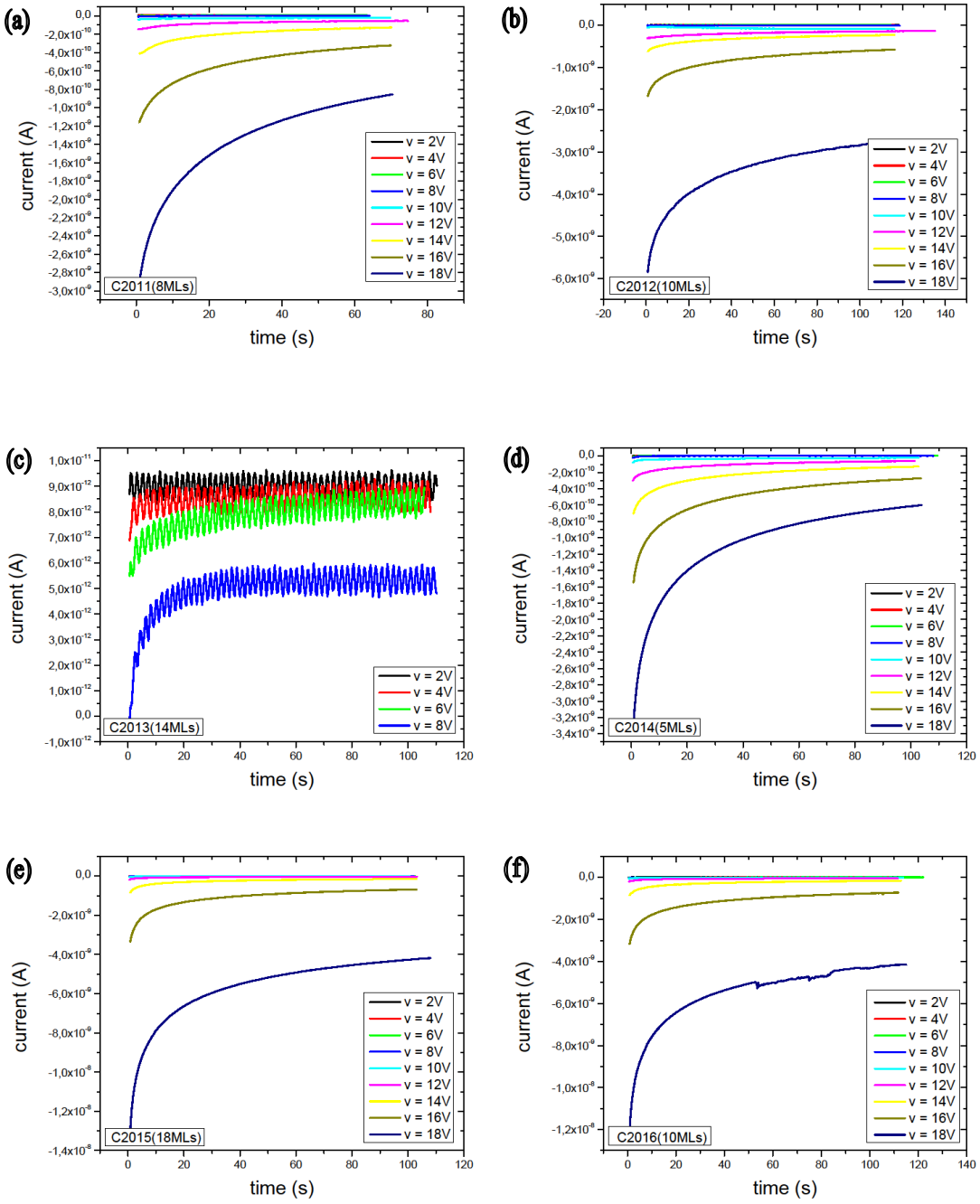
Figure 3.4-3: Current transients at different negative voltages: (a) C2011 (8MLs GaN), (b) C2012 (10MLs GaN), (c) C2013 (14MLs GaN), (d) C2014 (5MLs GaN), (e) C2015 (18MLs GaN), (f) C2016 (10MLs GaN – lower III-V), and (g) CRef (no GaN).

What we are trying to achieve by applying negative voltage at the gate of the device is see how carriers behave during discharging. The fact that we do not obtain a clean exponential curve might mean that there are indeed more than one mechanisms taking place, as was suggested by the retention measurements.

This transient behaviour is observed for four out of the six samples which contain GaN. The reference sample seems to yield simpler more predictable exponential transients. This could mean that the different discharging mechanism is related to the GaN QDs. However, neither sample C2015 with 18MLs of GaN nor sample C2016 with lower rate of deposition seem to have two separate branches in their characteristics. What is more, the transients seem to be faster for those devices with more GaN embedded in the oxide (figures 3.4-4b, and c).

It should be noted here that the measurement does not begin for a few hundred milliseconds each time the bias is applied and this could mean that some information regarding the current is lost.

At positive voltages, all of the samples seem to follow the exponential transient response better, indicating that, under conditions that correspond to charging, carriers might be behaving as expected (fig. 3.4-4). Still, there might be more than one time constants involved in these responses.


(a)
current (A)
time (s)
C2011(8MLs)
v = 2V
v = 4V
v = 6V
v = 8V
v = 10V
v = 12V
v = 14V
v = 16V
v = 18V
(b)
current (A)
time (s)
C2012(10MLs)
v = 2V
v = 4V
v = 6V
v = 8V
v = 10V
v = 12V
v = 14V
v = 16V
v = 18V
(c)
current (A)
time (s)
C2013(14MLs)
v = 2V
v = 4V
v = 6V
v = 8V
(d)
current (A)
time (s)
C2014(5MLs)
v = 2V
v = 4V
v = 6V
v = 8V
v = 10V
v = 12V
v = 14V
v = 16V
v = 18V
(e)
current (A)
time (s)
C2015(18MLs)
v = 2V
v = 4V
v = 6V
v = 8V
v = 10V
v = 12V
v = 14V
v = 16V
v = 18V
(f)
current (A)
time (s)
C2016(10MLs)
v = 2V
v = 4V
v = 6V
v = 8V
v = 10V
v = 12V
v = 14V
v = 16V
v = 18V

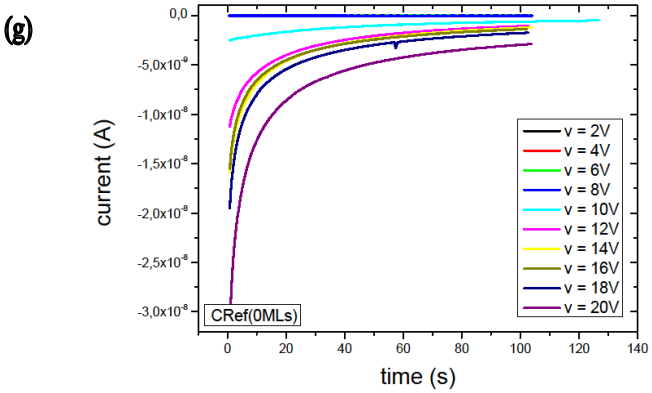


Figure 3.4-4: Current transients at different positive voltages: (a) C2011 (8MLs GaN), (b) C2012 (10MLs GaN), (c) C2013 (14MLs GaN), (d) C2014 (5MLs GaN), (e) C2015 (18MLs GaN), (f) C2016 (10MLs GaN – lower III-V), and (g) CRef (no GaN).

For higher negative voltages, peaks start to appear in the transient characteristics of the device currents (fig. 3.4-3a, b, c, and d). This may have to do with the storing of negative charges in the oxide because after such a peak, a narrow C-V sweep shows the capacitance characteristic shifted to the right (fig. 3.4-5a). For samples C2015 and C2016, where no peak is observed, the C-V curve does not exhibit significant shift (fig. 3.4-5b).

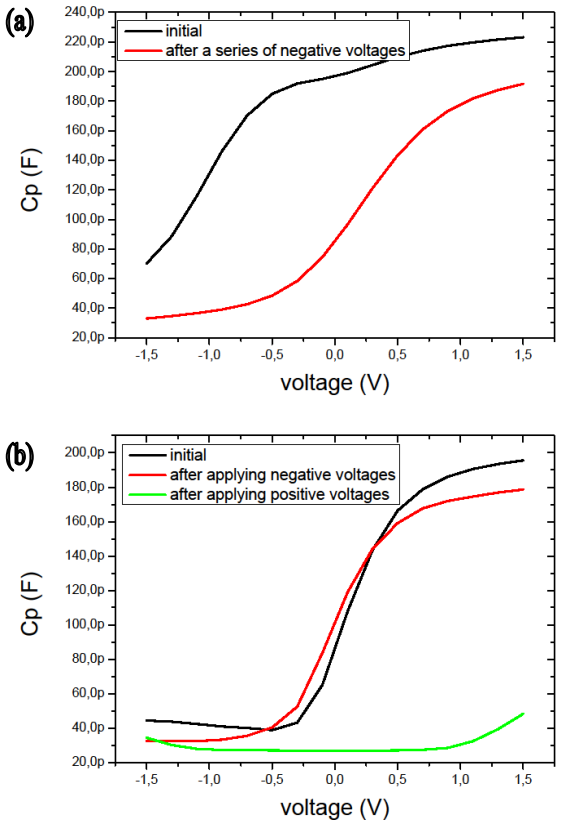


Figure 3.4-5: C-V curves before and after applying bias voltages for samples C2011 (a), and sample C2015 (b).

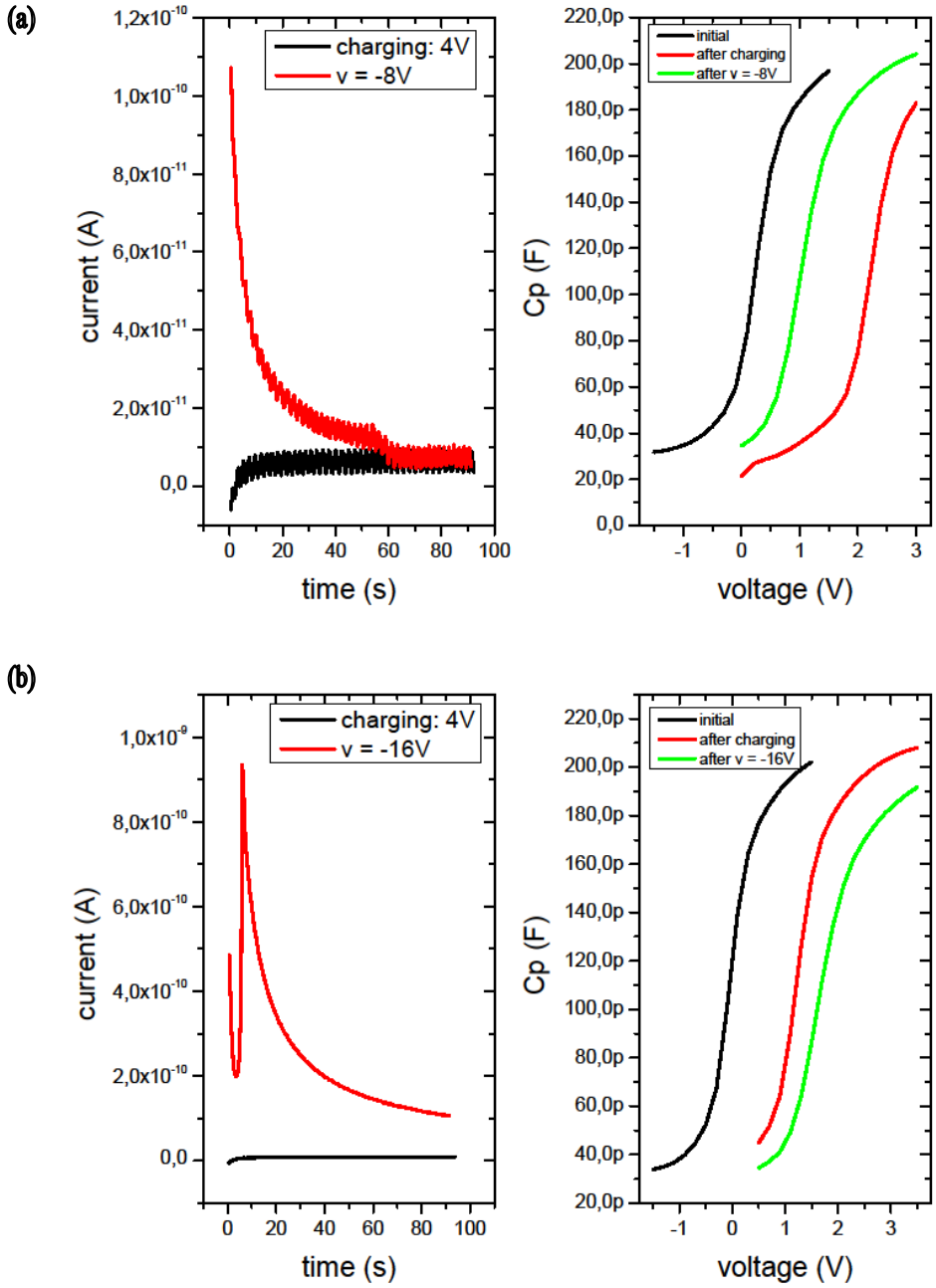


Figure 3.4-6: To the left: transient characteristics for the charging and discharging currents. To the right: capacitance shifts for the applied voltages. (a) applied voltage: -8V, (b) applied voltage: -16V (sample C2014).

Figure 3.4-6 shows the C-V curve shifts, this time for a device (C2014) that is first charged with positive voltage, which as a result, moves the capacitance characteristic to higher voltages. In the case of a low negative bias of -8V (fig. 3.4-6a) the C-V curve starts to move back to its original place, suggesting the device is being erased, as it is expected under negative voltage. When high voltage is applied, however, (fig. 3.4-6b) the C-V curve apparently moves even further to the right.

This move of the C-V curve could mean that for very high negative voltages electrons are being injected into the nanocrystals from the gate electrode instead of the substrate. However, similar observations can be made for the reference sample, which does not contain any QDs (fig. 3.4-3g) so this could mean this phenomenon is not quantum-dot-related. In fact, one way to explain it would be to consider that the semiconductor is in deep depletion because of the high negative voltage applied. As the substrate then goes into inversion, carriers are enerated and this could be the cause of the observed peaks.

3.5 CONCLUSIONS AND FURTHER STUDY

Non-volatile memories face problems following the rate of shrinkage of other integrated circuits due to inherent weaknesses of the continuous floating gate transistor with scaling. Nanoparticle memory cells are a possible solution to this problem and the purpose of this work was to examine the performance of GaN nanocrystals as storage nodes for a nanocrystal floating gate.

The successful growth of GaN quantum dots on an amorphous SiO_2 substrate was achieved. It was found that the size and density of the nanocrystals depend both on the dose of GaN and the deposition conditions. Next, the growth method was used to fabricate Metal-Oxide-Semiconductor devices with GaN QDs embedded in the oxide. These devices were characterized with regard to their memory properties. Results reveal a sufficiently wide memory window. The memory window and the amount of charge such a device can hold were studied as a function of the dos of GaM used during deposition. It can be concluded that the storage capability increases with the dose of GaN although the exact amount of charge is further affected by the conditions during deposition.

The devices show good retention properties with potential for achieving the ten-year threshold for non-volatility. However, the discharging mechanisms that seem to take place have yet to be fully understood. What is more, there appears to be difficulty with the erase operation and discharging the devices is, at the moment, quite challenging.

To further understand the operation of GaN nanocrystal memories, more information about the quantum dot size and density is essential. In this direction, the first step for future study should be to investigate how the size and density of the nanocrystals affect the charging and discharging properties of the memory cells. A study of the memory characteristics as a function of the QD properties would be very helpful. What is more, since transient responses have not given definitive results about the discharging mechanisms those could be looked into more extensively, as well.

To optimize the performance of GaN quantum dot memory devices, other parameters than the properties of the nanocrystals could be varied and their effect could be put under study. Such parameters include the thickness of the tunnel oxide, primarily. One should also consider a complementary device developed on p-type silicon instead of an n-type one. Additionally, one could experiment with the composition of the nanoparticles themselves, thus changing and possibly improving the band alignment for retention. InGaN quantum dots could be utilized to engineer the band structure of the memory cells, thus enhancing their properties.

Lastly, we could move from the two-terminal MOS-type devices to the three-terminal nanoparticle floating gate transistor which should allow for a greater accuracy in characterization. Write, erase and read operations could then be studied with more precision and the devices would be evaluated not only for their retention and memory window, but also for their speed and power consumption.

REFERENCES

[1] Muller, R. S., Kamins, T.I., (2003) *Device Electronics for Integrated Circuits*. New Jersey: John Wiley and Sons.

[2] Schroder, D. K, (2006) *Semiconductor Material and Device Characterization*. 3rd Ed. New Jersey: John Wiley and Sons.

[3] Dimitrakis, P., Schamm-Chardon S., Bonafos, C., and Normand, P. (2012) Nanoparticle-Based Memories: Concept and Operation Principles. in: Chaughule R.S., and Watawe S.C. (eds.) *Applications of Nanomaterials*. American Scientific Publlishers.

[4] Blant, A. V., Hughes. O.H., Cheng, T. S., Novikov, S.V., and Foxon, C.T., (2009) *Nitrogen Species from Radio Frequency Plasma Sources Used for Molecular Beam Epitaxy Growth of GaN*. Plasma Sources Science and Technology, vol. 9, pp. 12-17.

[5] Joo, K.-H., Moon C.-R., Lee S.-N., Wang X., Yang J.-K., Yeo, I.-S., Lee, D., Nam, O., Chung, U-I., Moon J. T., and Ryu B.-I., (2006) *Novel Charge Trap Devices with NCBO Trap Layers for NVM or Image Sensor*. IEDM Technical Digest, 10.1109/IEDM.2006.34950.

APPENDIX A:

MODELING THE IMPEDANCE OF THE DEVICES.

The most common way to measure the capacitance of a MOS structure or a MOS-type structure is to use an impedance meter. The principle that such an instrument uses to perform C-V measurements is shown in figure A.1:

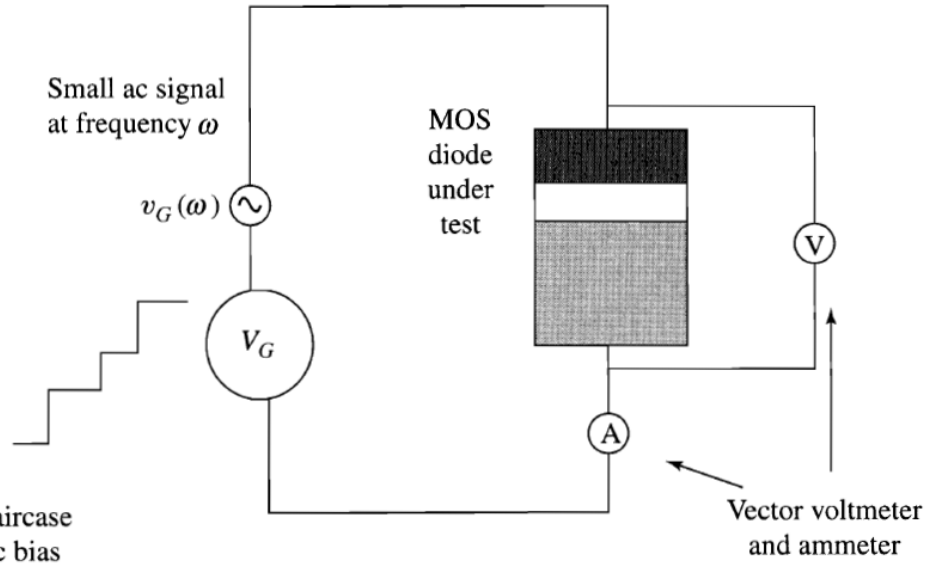


Figure A.1: MOS C-V measurement system. The voltmeter and ammeter measure both the magnitude and the phase of the voltage across the diode and the current through it.

The voltage applied to the MOS system consists of a dc voltage V_G on which a very small sinusoidal ac voltage is superposed. The dc voltage is programmed to step up and down within a certain desired range. The response is then usually measured and both the magnitude and the phase of the ac current through the structure are determined along with the voltage across it. Then, it becomes easy to extract either the conductance *G* or the impedance *Z* according to the equations:

$$Z = \frac{\upsilon}{i}$$

$$G = \frac{i}{\upsilon}$$

The next step is to model the measured impedance in some appropriate way in order to be able to extract the capacitance of the device.

The easiest, most common way to model the impedance is by using one of the two two-element models: the parallel model (fig. A.2a) or the series model (fig. A.2b).

The elements of these models are easy to extract and, furthermore, they yield quite accurate results for most cases. One need only consider which one models more closely the behavior of the device they are trying to measure. For MOS devices, the parallel model is usually preferred of the two. C is then considered to be the real dielectric capacitance and $G_p=1/R_p$ is used to model the parallel conductance due to gate leakage currents. When the series model is used, R_s represents the series resistance due to well/substrate and contact resistances.

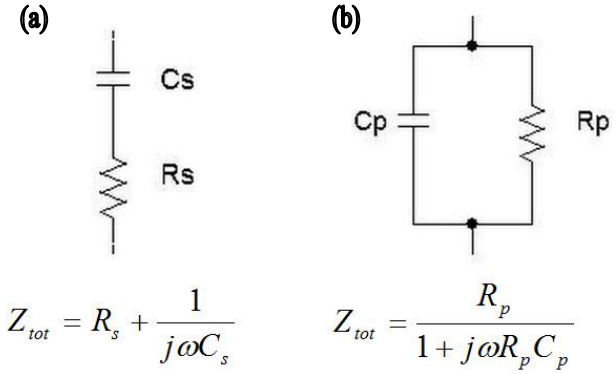


Figure A.2: Two-element models: (a) series model, (b) parallel model,

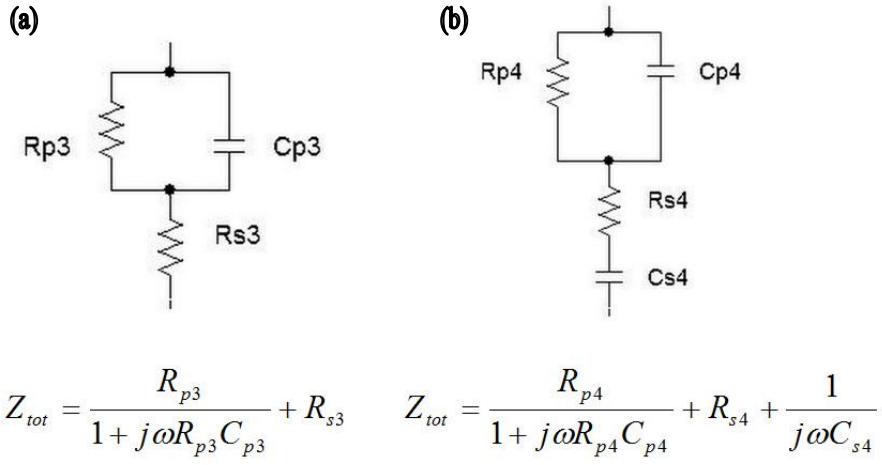
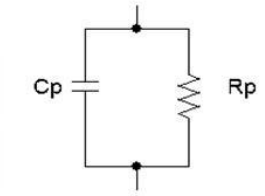


Figure A.3: Two-frequency models: (a) three-element model, (b) four-element model.

Generally, an impedance analyzer can measure a complex impedance with real and imaginary parts, and then interpret it by either the series of parallel two-element modes and thus produce a value for the capacitance. However, sometimes the two-element models may not suffice to simulate the behavior of the structure in question because, for example, a real capacitance can have leakage current (represented by the parallel resistance) and a series resistance simultaneously. In such cases more complicated models must be employed, such as the three element model presented in figure A.3a.

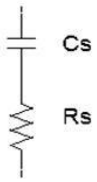
Z at frequency f=ω/2π with:

$$Z = \mathrm{Re} + j\,\mathrm{Im}$$



$$Rp = \frac{\mathrm{Re}^2 + \mathrm{Im}^2}{\mathrm{Re}}$$

$$Cp = \frac{k}{Rp}$$

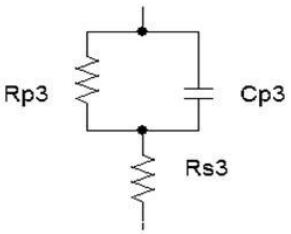


$$Rs = \mathrm{Re}$$

$$Cs = \frac{1}{\omega\,\mathrm{Im}}$$

Z_1 at frequency $f_1 = \omega_1/2\pi$, Z_2 at frequency $f_2 = \omega_2/2\pi$ with:

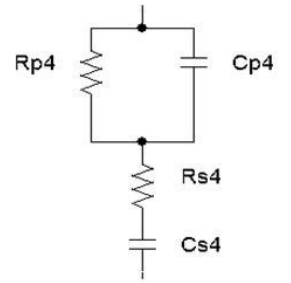
$$Z_1 = \mathrm{Re}_1 + j\,\mathrm{Im}_1 \quad \text{and} \quad Z_2 = \mathrm{Re}_2 + j\,\mathrm{Im}_2$$



$$Rp3 = -\frac{\mathrm{Im}_1}{\omega_1}\frac{1+\omega_1^2 k^2}{k}$$

$$Rs3 = \mathrm{Re}_1 - \frac{Rp3}{1+\omega_1 k^2}$$

$$Cp3 = \frac{k}{Rp3} \qquad k = -\frac{\mathrm{Im}}{\omega\,\mathrm{Re}}$$



$$Rp4 = \frac{(\mathrm{Re}_1 - \mathrm{Re}_2)}{k^2}\frac{(1+\omega_1^2 k^2)(1+\omega_2^2 k^2)}{(\omega_2^2 - \omega_1^2)}$$

$$Rs4 = \mathrm{Re}_1 - \frac{Rp4}{1+\omega_1^2 k^2}$$

$$Cs4 = \left[-\omega_1\,\mathrm{Im}_1 - \frac{\omega_1^2 k Rp4}{1+\omega_1^2 k^2}\right]^{-1}$$

$$Cp4 = \frac{k}{Rp4} \qquad k = \frac{\mathrm{Re}_1 - \mathrm{Re}_2}{\omega_1\,\mathrm{Im}_1 - \omega_2\,\mathrm{Im}_2}$$

Figure A.4: Calculating the elements for all four models.

In order to assess whether the two-element models were sufficient in our analysis, a comparative study of the three models mentioned above, as well as the four-element model shown in figure A.3b was performed. The models were solved so that their elements could be calculated and compared. The resulting equations are collectively

presented in figure A.4 and the matlab code that was used is given in Appendix B (extracts 1 and 2). It should be mentioned that more complicated models, such as the three and four-element require the impedance to be measured at more than one frequencies (otherwise we lack in information to solve the respective equations. This is the reason why the two-element models are sometimes called *one-frequency models* and the three and four-element are called *two-frequency models*. In such a case the frequencies at which the measurement is performed should be close to each other.

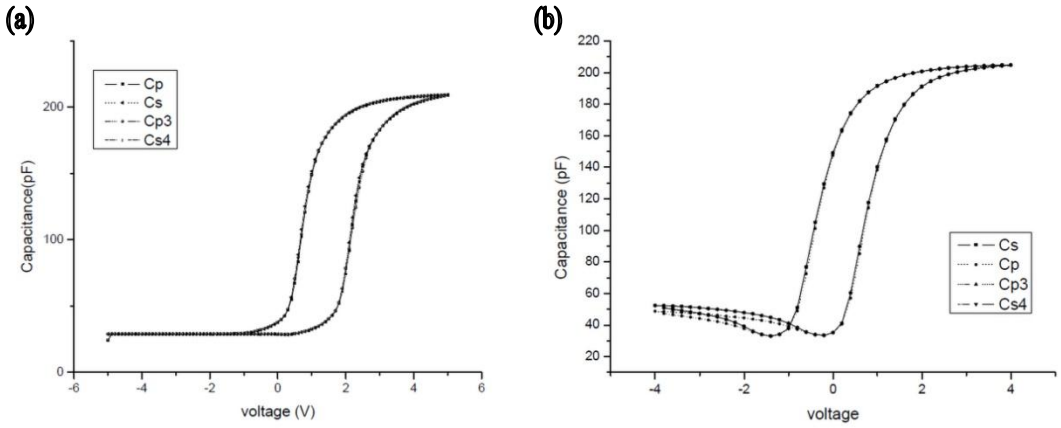


Figure A.5: The capacitances calculated from all four models are compared for two different devices: device (a) and device (b).

In the instance of the samples studied in the current thesis, all four models were found to produce very similar results as is demonstrated in figure A.5 for two distinct cases. Examples of the calculated valued of the elements for all models are given in figures A.6 to A.9. The measurement was performed at 1 MHz and for the two-frequency models the second frequency was 960 kHz.

In some cases, the three-element model was found to exhibit instabilities (fig. A.10). For this reason, out of the two frequency models the four element model has been preferred in our calculations. Additionally, the parallel model is preferred over the series model for reasons of comparability with other studies.

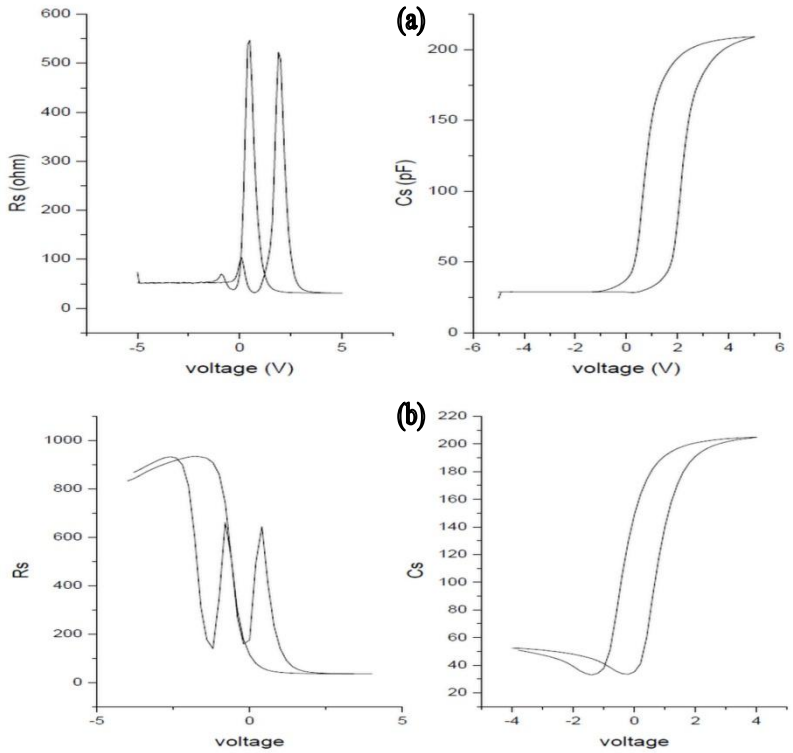


Figure A.6: Series models for device (a) and device (b).

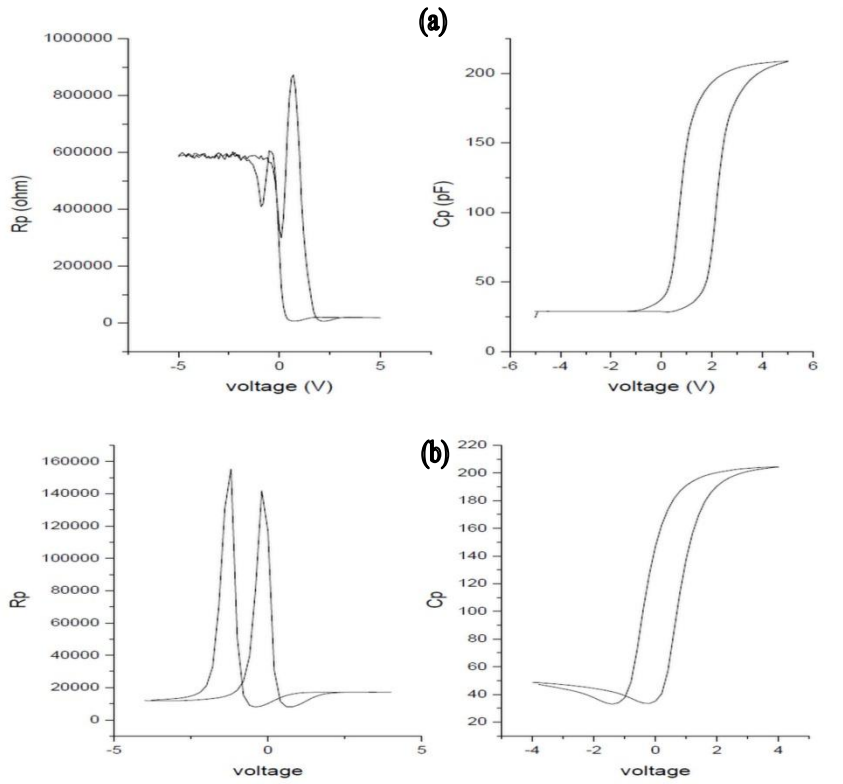


Figure A.7: Parallel models for device (a) and device (b).

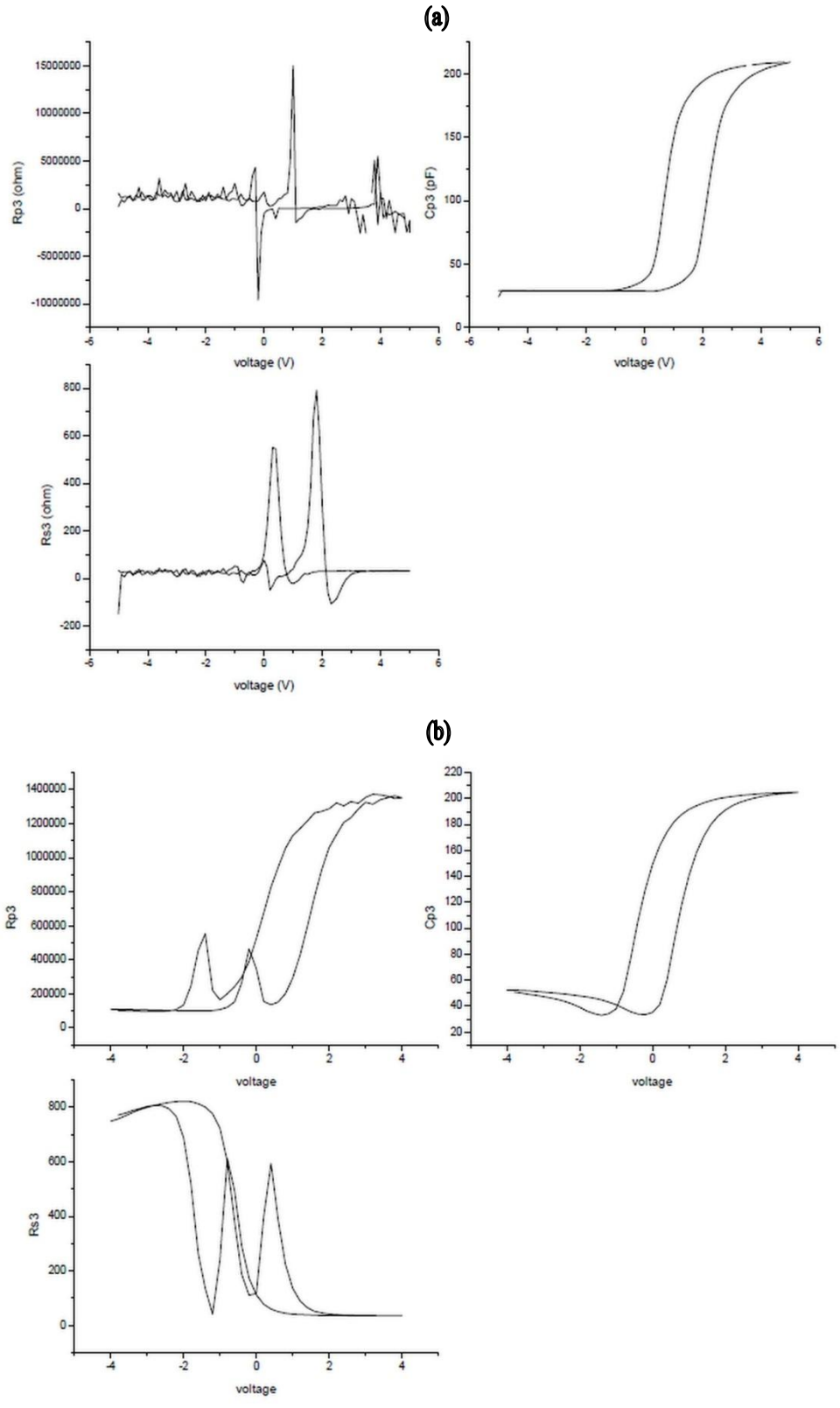


Figure A.8: Three-element model for device (a) and device (b).

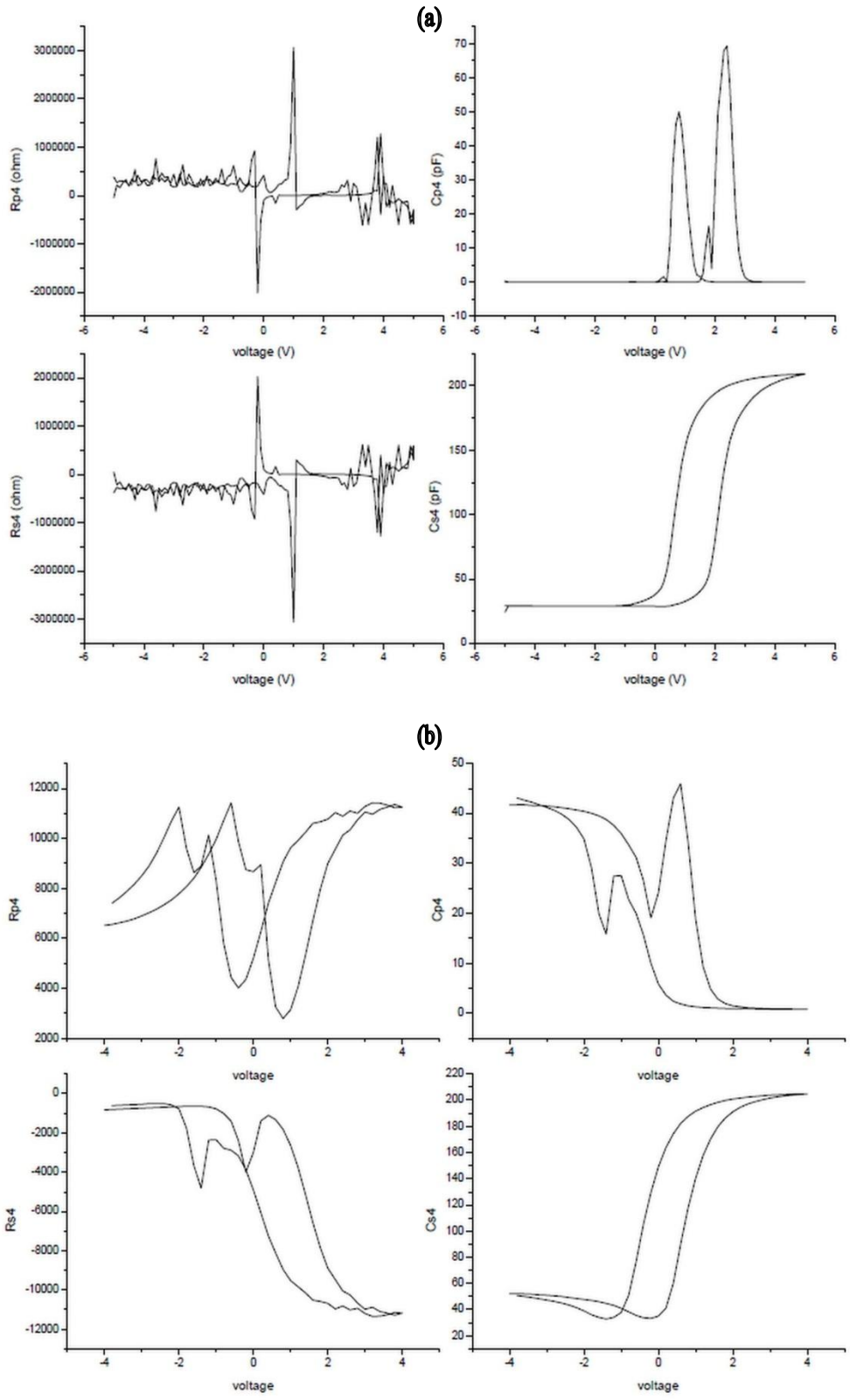


Figure A.9: Four-element models for device (a) and device (b).

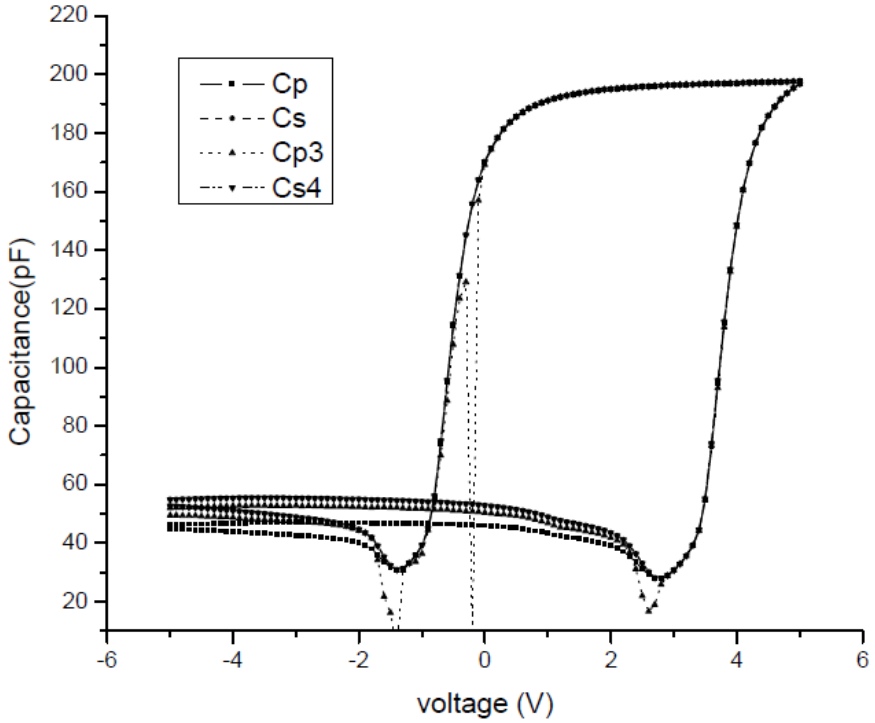


Figure A.10: The three-element model sometimes exhibits instabilities compared to the other three models.

REFERENCES

[1] Muller, R. S., Kamins, T.I., (2003) *Device Electronics for Integrated Circuits*. ^^Ed.^^ New Jersey: John Wiley and Sons.

[2] Baomin, W., Guoping, R., Yulong, J., Xinping. Q., Bingzong, L., and Ran, L., (2009) *Capacitance-Voltage Characterization of fully silicided gated MOS capacitor*, Journal of Semiconductors, vol. 30, no. 3, 034002

APPENDIX B:

MATLAB CODE.

Extract 1:
Series and Parallel Models

```
clear;

files = dir('*.dat');
FIL = length(files);

for fil = 1:1:FIL

    prog = ['A=importdata(''' files(fil).name ''');'];
    eval(prog);

    f = 1000000;                    %the frequency of the C-V test
    omega = 2*pi*f;



    % Re = A(i,2), Im = A(i,3) and A(i,1) the voltage
    % Series Model: Series(i,1)=voltage, Series(i,2)=Rs kai Series(i,3)=Cs
    % Parallel Model: Parallel(i,1)=voltage, Parallel(i,2)=Rp kai
Parallel(i,3)=Cp

    [lines rows] = size(A);
    Mseries = 0;
    Mparallel = 0;
    for i=1:1:lines
        %for the series model
        Mseries(i,1) = A(i,1);
        Mseries(i,2) = A(i,2);                              %Rs
        Mseries(i,3) = -1/(omega*A(i,3));                   %Cs

        %for the parallel model
        k = -A(i,3)/(omega*A(i,2));
        Rp = (A(i,2)^2 +A(i,3)^2)/A(i,2);
        Mparallel(i,1) = A(i,1);
        Mparallel(i,2) = Rp;                                %Rp
        Mparallel(i,3) = k/Rp;                              %Cp
    end

    voltage = 0;
    Cp = 0;
    Rp = 0;
    Cs = 0;
    Rs = 0;
    for i=1:1:lines
        voltage(i) = Mparallel(i,1);
        Cp(i) = Mparallel(i,3);
        Rp(i) = Mparallel(i,2);
        Cs(i) = Mseries(i,3);
```

```
        Rs(i) = Mseries(i,2);
    end

    figure; plot(voltage, Cp, voltage, Cs);

    prog1 = ['save ' files(fil).name '_series.out Mseries -ASCII'];
    prog2 = ['save ' files(fil).name '_parallel.out Mparallel -ASCII'];

    eval(prog1);
    eval(prog2);

end
```

Extract 2:

Three and Four Element Models

```
clear;

files = dir('*.dat');
FIL = length(files);

for fil = 1:2:FIL
    prog1 = ['A1=importdata(''' files(fil).name ''');'];
    prog2 = ['A2=importdata(''' files(fil+1).name ''');'];
    eval(prog1);
    eval(prog2);

    f1 = 1000000;         % the two frequencies of the C-V tests f1->A1
    f2 = 960000;          %                                      f2->A2
                          % note that A1 and A2 should be the same size

    omega1 = 2*pi*f1;
    omega2 = 2*pi*f2;

    % Re1 = A1(i,2), Im1 = A1(i,3) and A1(i,1) the voltage
    % Re2 = A2(i,2), Im2 = A2(i,3) and A2(i,1) the voltage
    % Three Element Model: ThreeEl(i,1)=voltage, ThreeEl(i,2)=Rp3,
    %     ThreeEl(i,3)=Cp3, ThreeEl(i,4)=Rs3,
    % Four Element Model: FourEl(i,1)=voltage, FourEl(i,2)=Rp4,
    %     FoueEl(i,3)=Cp4, FourEl(i,4)=Rs4, FourEl(i,5)=Cs5

    [lines rows] = size(A1);

    ThreeEl = 0;
    FourEl = 0;
    for i=1:1:lines
        %for the three element model
        k3 = (omega2*A1(i,3)- omega1*A2(i,3))/(omega1*omega2*(A2(i,2)-
A1(i,2)));
        Rp3 = -(A1(i,3)/omega1)*((1+(omega1^2)*(k3^2))/k3);
        ThreeEl(i,1) = A1(i,1);                                    %voltage
        ThreeEl(i,2) = Rp3;                                        %Rp3
        ThreeEl(i,3) = k3/Rp3;                                     %Cp3
        ThreeEl(i,4) = A1(i,2) - Rp3/(1+(omega1^2)*(k3^2));        %Rs3

        %for the four element model
        k4 = (A1(i,2)-A2(i,2))/(omega1*A1(i,3)-omega2*A1(i,3));
        Rp4 = ((A1(i,2)-
A2(i,2))*(1+(omega1^2)*(k4^2))*(1+(omega2^2)*(k4^2)))/((k4^2)*(omega2^2-
omega1^2));
```

```
        FourEl(i,1) = A1(i,1);                                      %voltage
        FourEl(i,2) = Rp4;                                          %Rp4
        FourEl(i,3) = k4/Rp4;                                       %Cp4
        FourEl(i,4) = A1(i,2)- Rp4/(1+(omega1^2)*(k4^2));           %Rs4
        FourEl(i,5) = 1/(-omega1*A1(i,3)-
((omega1^2)*k4*Rp4)/(1+omega2^2)*(k4^2));  %Cs4
    end



    voltage = 0;
    Cp3 = 0;
    Rp3 = 0;
    Rs3 = 0;
    Rp4 = 0;
    Rs4 = 0;
    Cp4 = 0;
    Cs4 = 0;
    for i=1:1:lines
        voltage(i) = ThreeEl(i,1);

        Cp3(i) = ThreeEl(i,3);
        Rp3(i) = ThreeEl(i,2);
        Rs3(i) = ThreeEl(i,4);

        Rp4(i) = FourEl(i,2);
        Cp4(i) = FourEl(i,3);
        Rs4(i) = FourEl(i,4);
        Cs4(i) = FourEl(i,5);
    end

    figure; plot(voltage, Cp3, voltage, Cs4);

    prog3 =['save ThreeEl_' files(fil).name '_and_' files(fil+1).name '.out
ThreeEl -ASCII'];
    eval(prog3);
    prog4 =['save FourEl_' files(fil).name '_and_' files(fil+1).name '.out
FourEl -ASCII'];
    eval(prog4);

end
```

Extract 3:

Sample Calculations

```
clear;

q = 1.602e-19;              %C
k = 1.3806e-23;             %J*K^-1
epsilon0 = 8.854e-12;       %F*m^-1
epsilonSi = 11.9;
epsilonSiO2 = 3.9;
epsilonSi3Ni4 = 7.5;
epsilonOx = epsilonSiO2;    %selecting the appropriate oxide
epsilonGaN = 8.9;

Ni = 1.01e16;         % in m^-3
T = 300;              % in Kelvin
A = 1*(1e-8);         % in m^2
```

```
%%%  For ε effective: Chakraborty, G., Sengupta, A., Requejo F.G., and
Sarkar, C.K
%%%  'Study of the relative preformance of silicon and germaniums
%%%  nanoparticles embedde gate oxide in metal-oxide semiconductor memory
%%%  devices', Journal of Applied Physics, 109, 0614504 (2011)

D = 2.9;            % depends on the sample, see chapter 3 for approximation
dox = 3.5 + 20;     % in nm
dtot = dox +D;      % in nm
f = 0.25;           % volume fraction of the ncs present in the oxide

epsilonQD = 1 + ( (epsilonGaN-1) / (1+(1.38/D)^1.37) );
epsiloneff = ( (dox/(epsilonOx*dtot)) + (1 - (dox/dtot))*( ...
    (epsilonQD+(2*epsilonOx)-(f*(epsilonQD-epsilonOx))) ...
    / (epsilonOx*( 2*f*(epsilonQD - epsilonOx)+...
    (epsilonQD + 2*epsilonOx) )) ) )^(-1);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%



%%% Importing the file with the C-V characteristic swept fro inversion to
%%% depletion and back to inversion. Sweep must be wide enough for the
%%% extraction of Cmin and Cmax. It is assumed that the first row is the
%%% voltage.

A1 = importdata('Atestsweep1.dat_parallel.out');
w = 2;          % w = 3 (series-parallel) or 4(three-element ...
                % or 5 (four-element) or other, row of the capacitance



[lines rows] = size(A1);

%%% assiging values %%%
for i=1:1:lines
    C(i) = A1(i,w);
    voltage(i) = A1(i,1);
end

[p1 r1] = size(C);
[p2 r2] = size(voltage);

r2 == r1     %%% checking the file dimensions

mesi = floor(r1/2);



%%% Splitting the back and forth sweep to two parts.
%%% If the sweep is only one way then:
%%% mesi = r1; C1temp(i)=C(i); voltage1temp(i)=voltage(i);
%%% which is a bit redundant but helps with the flow.

for i=1:1:mesi
    C1temp(i) = C(i);
    voltage1temp(i) = voltage(i);
    C2temp(i) = C(mesi+i);
    voltage2temp(i) = voltage(mesi+i);
end



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Cox = C1temp(end);
Cinv = min(C1temp);           %% or Cinv = C1temp(20)
```

```
P = (4*k*T)/(epsilon0*epsilonSi*(q^2)*(A^2));
F = ((Cox*Cinv)/(Cox-Cinv))^2;
R = Cinv/Cox;
Cap = (R*Cox)/(10000*A*(1-R));
x = 30.38759 + 1.68278*log10(Cap) - 0.03177*(log10(Cap))^2;
Nd = (10^x)*1e6


%%% Method #1: through the Debye Length.

Ld = sqrt(epsilon0*epsilonSi*k*T/((q^2)*Nd));
Cfbs = epsilon0*epsilonSi*A/Ld;
Cfb1 = (Cox*Cfbs)/(Cox+Cfbs);

%%% Method #2: through the oxide thickness, using an empirical equation for
%%% Cfb.

ti = 100*epsilon0*epsiloneff*A/Cox;    % for ε effective, see above
Nd2=Nd*1e-6;                           % se cm^-3
Cfb2 = Cox*(1/(1+(136/ti)*sqrt(T/(300*Nd2))));


%%% Selection of method, and converting farads in picofarads.

Cfb1 = Cfb1*1e12;
Cfb2 = Cfb2*1e12;
Cfb = Cfb1;



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%% Calculating the flat-band voltages

C1temp = C1temp*1e12;
C2temp = C2temp*1e12;



%%% Giving C distinct values:

l = 1;
for i=1:1:mesi
    x = C1temp(i);
    y = voltage1temp(i);
    j = 1;
    while x~=C1temp(j)
        j = j+1;
    end
    if i==j
        C1(l)=x;
        voltage1(l)=y;
        l=l+1;
    end
end

l = 1;
for i=1:1:mesi
    x = C2temp(i);
    y = voltage2temp(i);
    j = 1;
    while x~=C2temp(j)
        j = j+1;
```

```
    end
    if i==j
        C2(l)=x;
        voltage2(l)=y;
        l=l+1;
    end
end

CoxpF = Cox*1e12;
CinvpF = Cinv*1e12;
Vfb1 = interp1(C1, voltage1, Cfb);
Vfb2 = interp1(C2, voltage2, Cfb);
%%% MemWin = abs(Vfb1-Vfb2);
Vfbsmall = min(Vfb1,Vfb2);

subplot(2,2,1), plot(voltage1, C1); title('C1');
subplot(2,2,2), plot(voltage2, C2); title('C2');
subplot(2,2,3), plot(voltage1temp, C1temp); title('C1temp');
subplot(2,2,4), plot(voltage2temp, C2temp); title('C2temp');

save('C011_var.txt', 'CoxpF', 'CinvpF', 'Nd2', 'Ld', 'ti', 'Cfb1', 'Cfb2',
'Vfbsmall', '-ASCII');
```

Extract 4:

Calculations of Memory Window from C-V Measurements

```
clear;

q = 1.602e-19;              %C
k = 1.3806e-23;             %J*K^-1
epsilon0 = 8.854e-12;       %F*m^-1
epsilonSi = 11.9;
epsilonSiO2 = 3.9;
epsilonSi3Ni4 = 7.5;
epsilonOx = epsilonSiO2;    % selecting the appropriate oxide

Ni = 1.01e16;         % in m^-3
T = 300;              % in Kelvin
A = 1*(1e-8);         % in m^2


%%% Importing data from as have been previously extracted from long sweep
%%% (m-file: flat_band(for_pulse_one_way)/

Data = importdata('C012_var.txt');
CoxpF = Data(1);
CinvpF = Data(2);
Nd2 = Data(3);
Ld = Data(4);
ti = Data(5);
Cfb1 = Data(6);
Cfb2 = Data(7);
Vfbsmall = Data(8);

Nd = Nd2*1e6;          % converting from cm^-3 to m^-3.

%%% picking the row with the capacitances:
%w = 3 (series-parallel) or 4(three-element or 5 (four-element)
w = 5;
```

```
%%% every file in the folder with alphabetical order
files = dir('*.out');
FIL = length(files);

for fil = 1:1:FIL
    prog = ['A1=importdata(''' files(fil).name ''');' ];
    eval(prog);



    [lines rows] = size(A1);



    %%% assigning values: C, voltage
    C = 0;
    voltage = 0;
    for i=1:1:lines
        C(i) = A1(i,w);
        voltage(i) = A1(i,1);
    end

    [p1 r1] = size(C);
    [p2 r2] = size(voltage);

    r2 == r1        % checking the file dimensions

    mesi = floor(r1/2);

    % splitting the capacitance in two branches
    C1temp = 0;
    C2temp = 0;
    voltage1temp = 0;
    voltage2temp = 0;
    for i=1:1:mesi
        C1temp(i) = C(i);
        voltage1temp(i) = voltage(i);
        C2temp(i) = C(mesi+i);
        voltage2temp(i) = voltage(mesi+i);
    end



    % selecting flat-band capacitance
    Cfb = Cfb1;



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

    % in pF:
    C1temp = C1temp*1e12;
    C2temp = C2temp*1e12;

    % giving C1 and C3 distinct values

    C1 = 0;
    voltage1 = 0;
    l = 1;
    for i=1:1:mesi
        x = C1temp(i);
        y = voltage1temp(i);
        j = 1;
        while x~=C1temp(j)
```

```
            j = j+1;
        end
        if i==j
            C1(l)=x;
            voltage1(l)=y;
            l=l+1;
        end
    end


    C2 = 0;
    voltage2 = 0;
    l = 1;
    for i=1:1:mesi
        x = C2temp(i);
        y = voltage2temp(i);
        j = 1;
        while x~=C2temp(j)
            j = j+1;
        end
        if i==j
            C2(l)=x;
            voltage2(l)=y;
            l=l+1;
        end
    end


    % calculating the flat-band capacitance for the two branches, and the
    % memory window
    Vfb1 = interp1(C1, voltage1, Cfb);
    Vfb2 = interp1(C2, voltage2, Cfb);
    MemWin = abs(Vfb1-Vfb2);


    % calculating threshold voltages:
    sweeplimit = max(voltage);
    PhiB = (k*T/q)*log(Nd/Ni);
    Vth1 = -( ((A)/(CoxpF*(10^-12))) *
sqrt(4*epsilon0*epsilonSi*q*abs(Nd)*abs(PhiB))  + 2*abs(PhiB) ) + Vfb1;
    Vth2 = -( ((A)/(CoxpF*(10^-12))) *
sqrt(4*epsilon0*epsilonSi*q*abs(Nd)*abs(PhiB))  + 2*abs(PhiB) ) + Vfb2;


    Results(1,fil) = sweeplimit;
    Results(2,fil) = MemWin;
    Results(3,fil) = Vfb1;
    Results(4,fil) = Vfb2;
    Results(5,fil) = PhiB;
    Results(6,fil) = Vth1;
    Results(7,fil) = Vth2;

end

save results.txt Results -ASCII
figure; plot (Results(1,:), Results(2,:))
```

Extract 5:

Extraction of Flat-Band Voltage for Pulse Measurements

```
clear;
```

```
q = 1.602e-19;              %C
k = 1.3806e-23;             %J*K^-1
epsilon0 = 8.854e-12;       %F*m^-1
epsilonSi = 11.9;
epsilonSiO2 = 3.9;
epsilonSi3Ni4 = 7.5;
epsilonOx = epsilonSiO2;    %selecting the appropriate oxide

Ni = 1.01e16;        % in m^-3
T = 300;             % in Kelvin
A = 1*(1e-8);        % in m^2



%%% Importing data from as have been previously extracted from long sweep
%%% (m-file: flat_band(for_pulse_one_way)/

Data = importdata('C011_var.txt');
CoxpF = Data(1);
CinvpF = Data(2);
Nd2 = Data(3);
Ld = Data(4);
ti = Data(5);
Cfb1 = Data(6);
Cfb2 = Data(7);
Vfbsmall = Data(8);

Nd = Nd2*1e6;          % converting from cm^-3 to m^-3.

%%% picking the row with the capacitances
%w = 3 (series-parallel) or 4(three-elenent or 5 (four-element)
w = 2;



%%% every file in the folder with alphabetical order
files = dir('*.dat');
FIL = length(files);

for fil = 1:1:FIL
    prog = ['A1=importdata(''' files(fil).name ''');' ];
    eval(prog);



    [lines rows] = size(A1);



    %%% assigning values: C, voltage
    C = 0;
    voltage = 0;
    for i=1:1:lines
        C(i) = A1(i,w);
        voltage(i) = A1(i,1);
    end

    [p1 r1] = size(C);
    [p2 r2] = size(voltage);

    r2 == r1        % checking the file dimensions

    mesi = r1;

    C1temp = 0;
```

```
    voltage1temp = 0;

    for i=1:1:mesi
        C1temp(i) = C(i);
        voltage1temp(i) = voltage(i);
    end

 %%%selecting the flat-band capacitance
 Cfb = Cfb1;  %se pF



%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

    C1temp = C1temp*1e12;

    %%% giving C1 distinct values



    C1 = 0;
    voltage1 = 0;
    l = 1;
    for i=1:1:mesi
        x = C1temp(i);
        y = voltage1temp(i);
        j = 1;
        while x~=C1temp(j)
            j = j+1;
        end
        if i==j
            C1(l)=x;
            voltage1(l)=y;
            l=l+1;
        end
    end



    %%% caclulating the flat-band voltage
    Vfb1 = interp1(C1, voltage1, Cfb);

    MemWin = abs(Vfbsmall-Vfb1);

    Results(fil,1) = Vfb1;
    Results(fil,2) = MemWin;

end


save results.txt Results -ASCII

xaxix = [0.2 1 10 100 500 1000];
figure; plot (xaxix, Results(:,1)
```