DESIGN AND ANALYSIS OF NANOCRYSTAL FLASH MEMORY CELL BASED ON SINGLE GATE AND DOUBLE GATE MOSFET STRUTURES

Dissertation submitted in partial fulfillment of the requirements
for the award of the degree of

Master of Technology

In

VLSI Design

Submitted by

ANURAG SINGLA

Roll. No. 601261009

Under the supervision of

Mr. Arun Kumar Chatterjee

Assistant Professor, ECED

Department of Electronics & Communication Engineering

Thapar University, Patiala

INDIA

July, 2014

CERTIFICATE

I hereby declare that the work which is being presented in the dissertation entitled, **"Design and Analysis of Nanocrystal Flash Memory Cell Based on Single gate and Double gate MOSFET Structures"** in partial fulfillment of the requirement for the award of degree of Master of Technology (VLSI Design) at the department of Electronics and Communication Engineering, Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of Mr. Arun Kumar Chatterjee, Assistant Professor, ECED.

The matter presented in this dissertation has not been submitted in any other University/Institute for the award of any other degree.

Date: 9/7/14

Anurag Singla
Roll No. 601261009

It is certified that the above statement made by the student is correct to the best of my knowledge and belief.

Arun Kumar Chatterjee

Mr. Arun Kumar Chatterjee
Assistant Professor
ECED, Thapar University

Counter signed by:

Dr. Sanjay Sharma
Professor & Head
ECED, Thapar University
Patiala-147004

Dr. S. K. Mahopatra
Dean of Academic Affairs
Thapar University
Patiala-147004

ACKNOWLEDGEMENTS

I take this opportunity to express my profound sense of gratitude and respect to all those who helped me through the duration of this dissertation. I acknowledge with gratitude and humility my indebtedness to **Mr. Arun Kumar Chatterjee, Assistant Professor,** Department of Electronics and Communication Engineering, Thapar University, Patiala, under whose guidance I had the privilege to complete this dissertation. I wish to express my deep gratitude towards him for providing individual guidance and support throughout the dissertation work.

I convey my sincere thanks to **Head of the Department, Dr. Sanjay Sharma** as well as **PG Coordinator, Dr. Kulbir Singh, Associate Professor, ECED**, **and Program Co-ordinator Dr. Anil Arora, Assistant Professor, ECED** and entire faculty and staff of Electronics and Communication Engineering Department for their encouragement and cooperation.

I am also thankful to Ms Madhu Kushwaha for her regular guidance and support throughout this dissertation work.

I wish to thank Department of Electronic Science, Kurukshetra University, Kurukshetra for providing lab facility to perform the simulation work.

My greatest thanks are to all who wished me success especially my family. Above all I render my gratitude to the Almighty who bestowed ability and strength in me to complete this work.

Anurag Singla

ABSTRACT

Flash memory is the most widely used non-volatile information-storage device today. Flash memories are ubiquitous in their use as portable storage media in cell phones, cameras, music players, and other portable electronic devices.

In conventional flash memory charges are stored on the continuous floating gate. If there is any oxide defect in the tunnel oxide, all the charges on the floating gate will leak back to either source/drain or channel through oxide defect. This not only limits the scalability, but also affects the retention characteristics of the flash memory.

With scaling many short channel effects like DIBL, GIDL etc. come in to picture. Due to charge sharing effect gate have less control on the channel region. Due to SCE's the subthreshold leakage current increases exponentially hence stand by power dissipation increases. In order to overcome the problems associated with conventional flash memory new structure has been proposed in which continuous floating gate is replaced with array of discrete charge storage regions i.e. nanocrystals which are electrically isolated from each other.

Further to enhance the performance of the nanocrystal flash memory cell, new structure has been proposed which combines the concept of Double gate MOSFET and nanocrystal flash memory.

In this work simulation study of single gate nanocrystal flash memory cell and double gate nanocrystal flash memory cell has been studied and simulated using SILVACO (ATLAS) TCAD tool. The programming and erasing characteristics of memory cell has been analysed. Also the impact of write-pulse's time duration and value of control gate voltage on programming characteristics of the memory cell has been observed. Finally the results of double gate nanocrystal flash memory cell and single gate flash memory cell are compared.

In order to facilitate continued scaling of the control dielectric, we explore replacement of the conventional silicon di oxide dielectric with high-k dielectric materials in single gate nanocrystal flash memory cell. In this work Sapphire (Al_2O_3) is used as the high-*k* dielectric. Sapphire is sandwiched between two SiO_2 layer and this stacked combination is used as control oxide and observed the programming characteristics of single gate nanocrystal flash memory cell.

It is found that the performance of double gate nanocrystal flash memory cell is much more enhanced in comparison to single gate nanocrystal flash memory cell.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVATIONS

MOS	Metal Oxide Semiconductor
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
SCE	Short Channel Effect
SS	Subthreshold Slope
SOI	Silicon on Insulator
DIBL	Drain Induced Barrier Lowering
GIDL	Gate Induce Drain Leakage
DG MOSFET	Double gate MOSFET
FG	Floating Gate

CHAPTER 1

Introduction

1.1 MOTIVATION

The conventional floating gate flash memory structure will be ideal with advancement of technology if the stored charge remains indefinitely on the floating gate, but actual case is different, if there is defect chain in the tunnel oxide, all the charge on the floating gate will leak back to either source/drain or channel through defect chain which limits the scalability, this will also affects the retention characteristics of the flash memory To retain charges for long time tunnel oxide is designed in such a way that its barrier height should high and its thickness is large to prevent the tunneling of charges from floating gate to the substrate [1]

With thick tunnel oxide leakage of charge from floating gate reduces and retention time will increase but thick tunnel oxide leads to higher programming voltage and programming time. Thin tunnel oxide lowers the barrier height and thickness at the tunnel oxide /Si interface which leads to charge loss from the floating gate

There is a trade-off relation between tunnel oxide thickness if increased, programming voltage and programming time will be high and if decreased, floating gate charge will tunnel through the weak spots [2]. Fig 1.1 illustrate this problem.

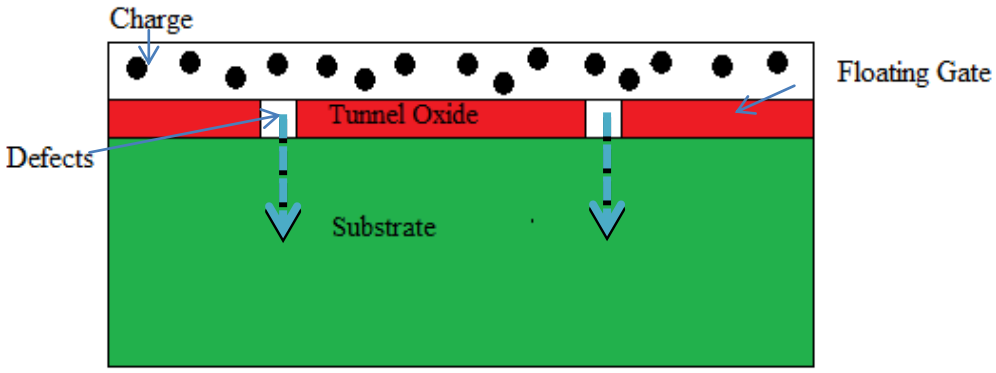

Figure 1.1: Issue of lateral conduction in floating gates

The main motivation is the restriction of lateral conduction with distributed charge storage, which promises to provide better data retention capabilities and the other parameters like

programming time and operating voltage of existing flash memory device are dependent on the barrier height and oxide thickness of the tunnel oxide.

1.2 Introduction to Nanocrystal Memory Device

As mentioned above, continuous floating gate is not good charge storage layer due to the leakage problem, as stored charge leak back in to the substrate through the oxide defects. The problem of the leakage can be overcome by replacing the continuous floating gate with array of discrete charge storage regions (nanocrystals) that are electrically isolated from each other can be used. Although the regions that are directly above the oxide defects can still experience charge leakages, but the charges on other nanocrystal will remain same, as the nanocrystal are electrically isolated from each other. It is shown in the figure below

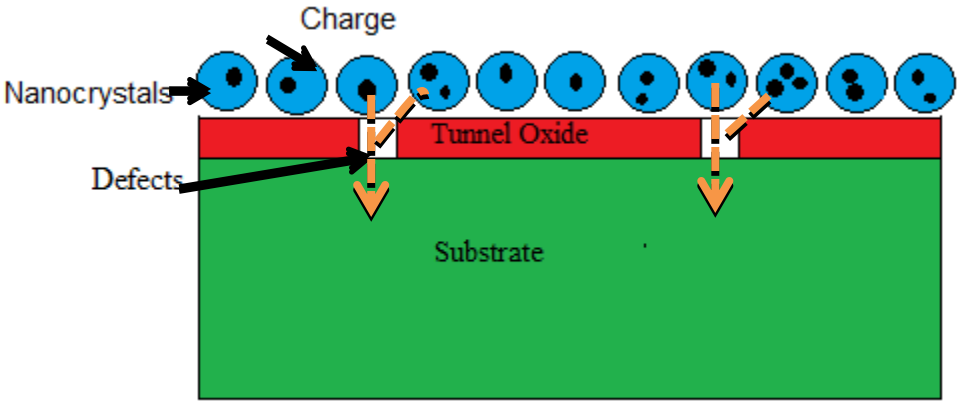

Figure 1.2: Reduction of lateral conduction.

Semiconductor flash memory scaling is far behind CMOS logic device scaling. For example, the EOT of the gate stack in semiconductor flash memory is still more than 10 nm. Moreover, semiconductor flashes memory still requires operation voltages of more than 10 V, while the operation voltage of CMOS logic has been scaled to about 1V or even less.

It is important to scale the EOT of the gate stack to achieve a small memory cell size, and also prolong battery life. A floating gate flash memory structure is shown in Fig 1.3

The gate stack consists of an 8nm thermal oxide as the tunnel layer, a 150 nm poly-silicon floating gate and a 13 nm (EOT) inter-poly oxide layer [10]. The EOT of the whole gate stack is 21nm. A typical drain bias is 2 V in the reading mode and 4.5 V in the programming mode. This memory cell suffers from serious short channel effects when the channel length is scaled to sub 100 nm, since the EOT of the gate stack is very thick and the drain bias is relatively

large. Both the drain-induced barrier lowering (DIBL) effect and the sub-surface punch-through effect induce significant leakage current during reading and programming.

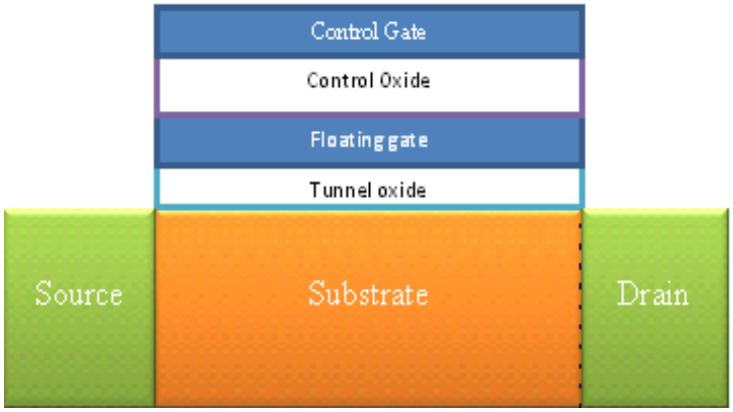

Figure 1.3: Schematic of floating gate Flash memory cell

Programming characteristics of nanocrystal depends on the time duration of programming pulse and programming voltage, design of the structure and type of dielectric. Increase coupling between the floating nanocrystals and control gate enhance the performance of the device and the same can be achieved by high-k dielectric.

With the advancement in VLSI industry concept of multigate is used to reduce various short channel effects to enhance the performance of the device. Double gate MOSFET is the potential device to minimize the SCE's. In advanced technology, concept of double gate and nanocrystal flash memory can combine and new device structure is designed to enhance the programming efficiency of the flash memory and reduce the short channel effects.

1.3 Outline of Thesis

The organization of the thesis and brief synopsis of the various chapters of this thesis are provided as follows:

Chapter 1: Introduction

This chapter covers some introductory information pertaining to the subject matter of this study. It also describes the scope and objectives of this work

Chapter 2: Theory of Flash Memory

This chapter provides the background and relevant theories of floating gate memory devices, the advent of distributed floating gate memory devices and the relevant technology scaling trends and problems. Brief overview of flash memory history, operation and scaling are presented. The "floating-gate transistor" cell is introduced with its operation principle. The types of flash memory are described and the main charge injection mechanisms are discussed. Then, reliability considerations of flash memory are briefly described, followed by the scaling trends and the main scaling challenges of flash memories.

Chapter 3: Literature Survey

This chapter provides the brief overview of the research work done on flash memory cell by different researchers

Chapter 4: Design of nanocrystal memory

This chapter describes all the necessary steps required to design and simulation of nanocrystal flash memory cell in SILVACO TCAD tool.

Chapter 5: Result and Discussion

In this chapter simulation result obtained in chapter 4 for programming characteristics of single gate nanocrystal flash memory cell and double gate nanocrystal flash memory cell are discussed and compared.

Chapter 6: Conclusion and Future Scope

After simulation study the programming characteristics are summarized and based on analysis future research possibilities are also given.

CHAPTER 2

Theory of Flash Memory

2.1 Introduction

In this chapter, brief overview of flash memory operation and scaling are presented, including its short history, the floating-gate transistor‟ cell and its operation. The types of flash memory and the main charge injection mechanisms are discussed. Then, reliability considerations of flash memory are briefly described, with scaling trends and main scaling challenges.

2.2 What is Flash Memory?

There are two types of memory one is volatile and other is non-volatile. A non-volatile memory is one in which stored data remain in the absence of power. Read only memory hard disk optical disk are example of non-volatile memory

ROM (Read Only Memory) in which data once written permanently during manufacturing. Data cannot be altered once written during manufacturing, that's why ROM lack in flexibility.

First Floating Gate memory device was introduced by kahng And Sze [3] in 1967, after that In 1970 Frohman – Bentchkowsky developed a polysilicon floating gate transistor in which programming is done by the technique of hot electron emission and erasing by exposing device in ultraviolet photoemission. This device was called the erasable programmable read only memory, as the flexibility increases due the facility of erasing by exposing to ultraviolet light. Following this many researches was done to develop electrical erasable programmable rom to increase the flexibility [4, 5]. As EEPROM cell required two transistors for their operation hence the size of device increases.
First commercial flash memory chip was produced by Intel in 1988[6, 7].

2.3 The Floating-gate Cell

A simple floating transistor cell is shown in fig 2.1

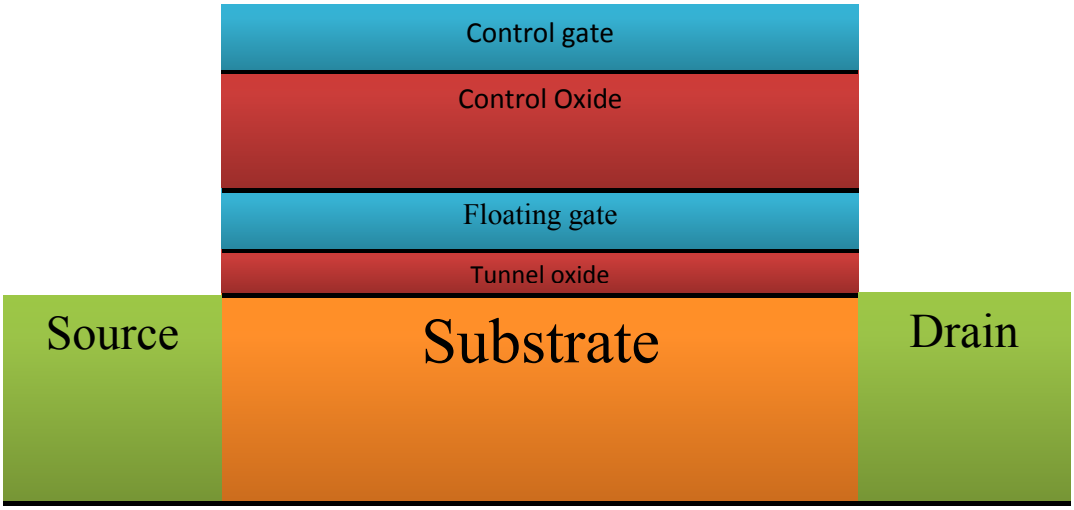

Figure 2.1: Structure of floating gate transistor cell

A floating gate transistor is similar to conventional MOS transistor, except for the additional electrode called floating gate and an additional dielectric layer between the floating gate and control gate. Dielectric between the substrate and floating gate is known as tunnel oxide. Floating gate is electrically isolated from the control gate and substrate by the layer of dielectric, layer of dielectric between control gate and floating gate is known as control oxide. Substrate is of p type and source and drain are highly doped n+ type. Floating gate which acts as the charge storage node capactively coupled to the control gate and other nodes. Programming cam done by two methods one by hot electron emission and second by F-N tunneling mechanism and erasing can be performed by F-N tunneling mechanism. Storage and removal of charge can be reflected in to the transfer characteristics of the device, after write operation negative charge is stored on the floating gate, threshold voltage of the device increases and after erasing negative charge tunnel back to the channel and the threshold voltage changes back to the original value. Fig. 2.1 shows the band diagram for the charged and neutral states of the floating gate transistor.

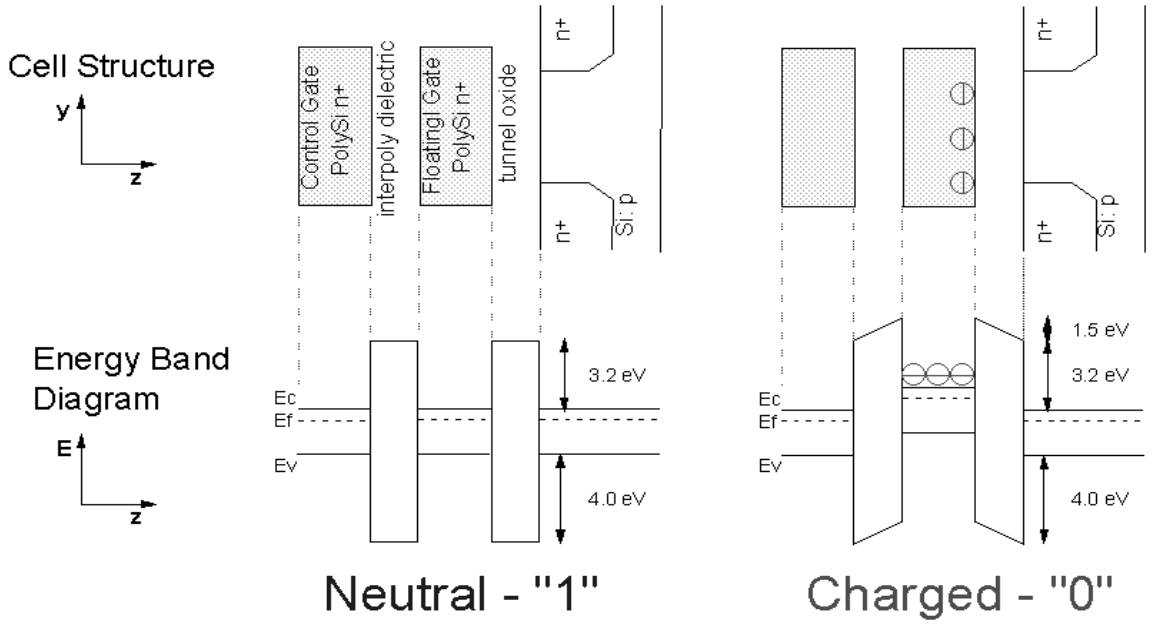

Figure 2.2: Band diagrams for the neutral state (left) and charged state (right) of the floating-gate transistor. [1]

In neutral state no electron on the floating gate. Electron pushed from substrate to floating gate during write operation which increases the potential of the floating gate and the potential well formed by the control oxide and tunnel oxide enables it to retain the charge.

2.4 Basic Physics of Floating gate non-volatile devices

The floating gate memory element consists of an electrode that is completely encased in an insulator and adjacent conductors are capacitively coupled to the floating gate modulate potential on the floating gate. The potential on the floating gate can be sensed by its effect on the current, which alters the logical output by changing the charge state on the floating gate is referred to as “writing” to the device. Physical mechanisms necessary for memory operation are described below.

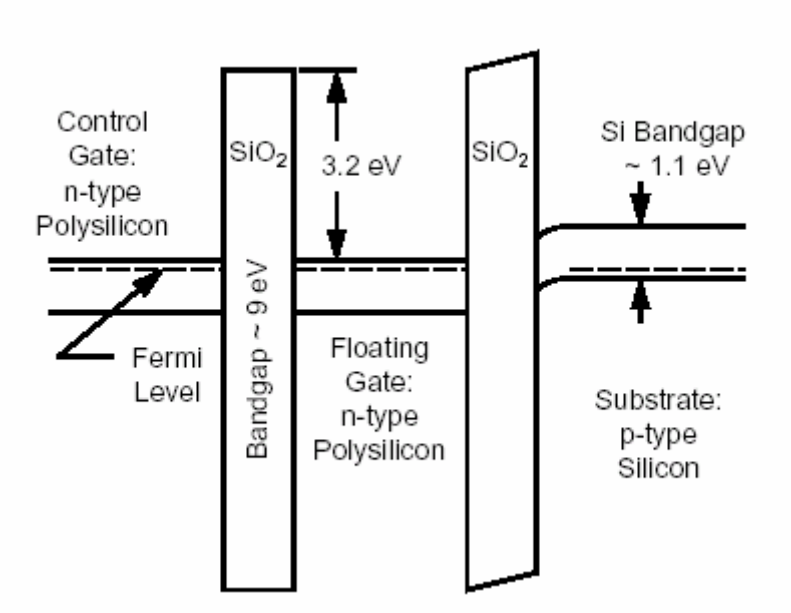

Figure 2.3: Band structure diagram of floating gate structure showing the n-type poly silicon control gate on the left, the n-type poly-silicon floating gate in the centre sandwiched between two SiO2 layers, and the p-type substrate on the right. [2]

2.4.1 Hot Carrier Injection

The wide- band gap of the insulator provides a potential barrier to prevent electrons from tunneling onto or from the floating gate as shown in Fig.2.3.Hot carriers are injected when carriers are accelerated to the energy level to cross the barrier. There are several mechanisms that may accelerate the carriers, but acceleration of electrons by lateral fields in the channel is the only hot electron mechanism that is currently used by floating gate memories for writing. The channel hot electron must reach at the oxide interface barrier with high energy so as to cross barrier height and be stored in the floating gate.

2.4.2 Drain-side programming

Drain-side programming is commonly employed because of simplicity of the structure. The drain bias is applied with a direct connection to a voltage source, while the gate bias depends upon capacitive coupling. The potential that is induced on the floating gate by a change in one of the surrounding electrodes depends on the capacitive coupling between the floating gate and the electrode in question [1-4]. In Fig. 2.4 (a) a gate is shown to be biased in such a state. The channel inversion layer is wider near the source and narrows as it approaches the pinch-off point. As the electrons pass through the pinch-off point, they are strongly

accelerated in the high field of the drain depletion region [1-4]. The shape of the SiO2 barrier varies along the channel length because the potential along the channel varies from source to drain while the gate potential remains constant [1-4]. Fig. 2.4 (b), 2.4 (c), and 2.4 (d) show schematically the oxide barrier shape at three selected points [1-4].

The highest density of hot electrons is toward the drain, where the oxide field is repulsive and near the source, where the oxide field is very strongly collective, there are almost no hot electrons .The programming efficiency is rather low as only a small portion of channel length is effective for programming.

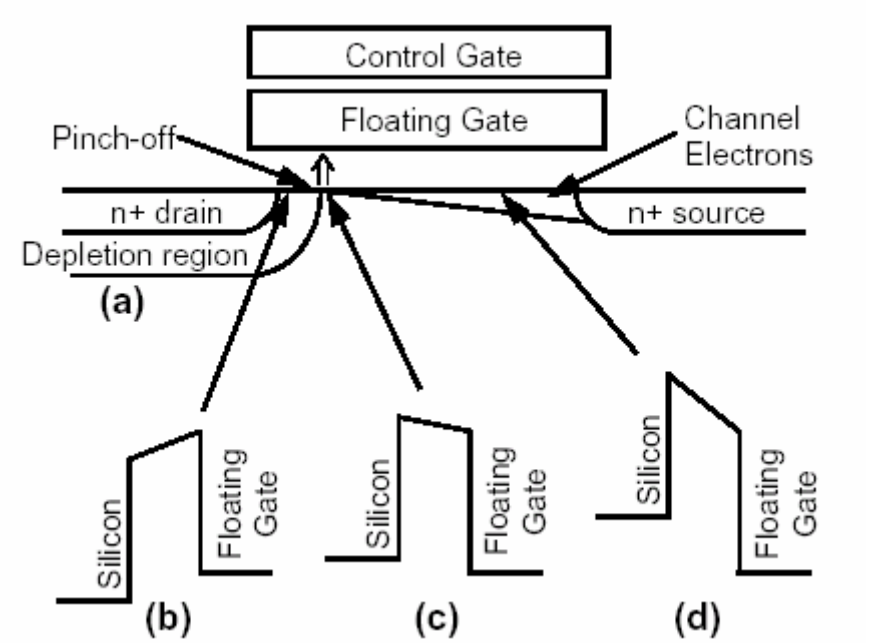

Figure 2.4: (a) illustrates the physical layout and indicates the injection of electrons near the pinch-off point. (b)- (d) show the potential diagrams of the conduction bands at the point's labelled b–d. [2]

2.4.3 F-N tunneling

F-N tunneling was introduced by Fowler and Nordheim in 1928 and was named after them. This is an approach that demonstrates for writing data. This method is realized by tunneling between an electrically isolated electrode and another conducting electrode. This effect can be employed for performing the floating gate programming or erasure. Some cells uses F-N

tunneling for both programming and erasing, while others engage CHE injection , for programming and F-N tunneling for erasure only.

The concept of F-N tunneling is presented in Fig.2.5. The energy discrepancy between the conduction band and valence bands in Si is about 1.1 eV while in SiO_2 is approximately 9 eV. When these two substances are merged together, the conduction band in SiO_2 is higher than that in Si by 3.25 eV. The difference in valence band and conduction band energies is even greater, i.e. more than 4 eV. Since the thermal energy of an electron averages only 0.025 eV at room temperature, the possibility of electron in silicon gaining sufficient thermal energy to break through the barrier and enter the conduction band in SiO_2 is extremely low. The case is demonstrated in Fig. 2.5 (b)

However in the presence of high electric field, the energy bands will be distorted as exemplified in Fig. 2.5 (b). Under these conditions, there is a still small but finite probability that an electron in the conduction band in the silicon will tunnel through the energy barrier and appear in the conduction band of SiO_2.

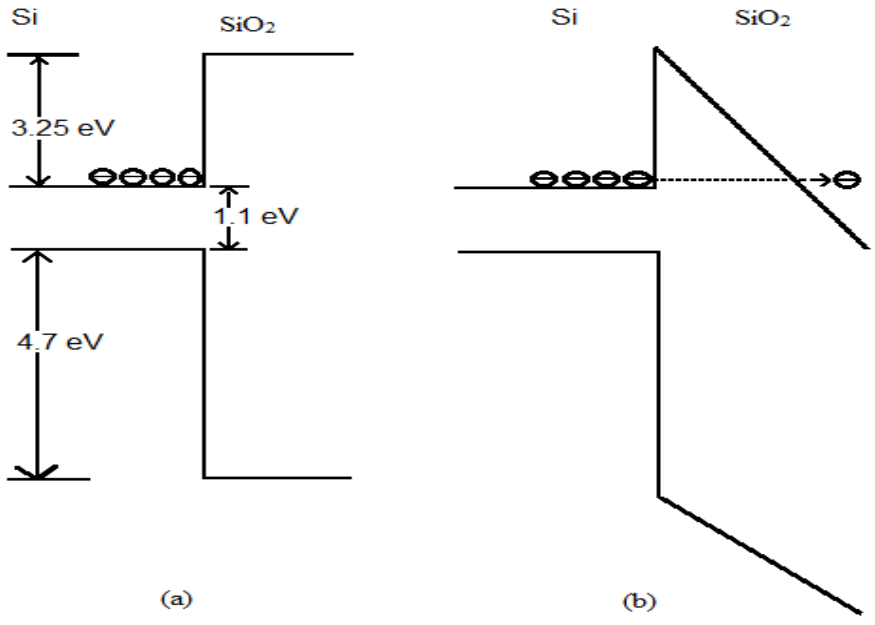

Figure 2.5: Potential diagram of the Si/SiO2 interface (a) with no electric field and (b) with a strong applied electric field. The electric field applied in (b) is strong enough to allow electrons to tunnel from the conduction band in the Si into the conduction band in the SiO_2. [2]

2.5 Programmable ROM-Information Storage and Access

To have a memory cell that can commute from one state to the other and that can store the information independently of external conditions, the storing element needs to be a device whose conductivity can be changed in a non-destructive way [4]. Transistor cells can be "written" into either state "1" or "0" by either "programming" or "erasing" methods. One of the two states is called "programmed," and the other "erased". The "read" operation is performed by applying to the cell a gate voltage that is between the values of the thresholds of the erased and programmed cells and senses the current flowing through the device [1-4].

The threshold voltage of a MOS transistor can be written as

$$V_T = K - \frac{Q}{C_{ox}} \tag{2.1}$$

Where

V_T is the threshold voltage of the device

K is a constant that depends on substrate and gate material, doping, and gate oxide thickness

Q is the charge weighted with respect to its position in the gate oxide,

Cox is the gate oxide capacitance

As can be seen, the threshold voltage of the memory cell can be change by changing the charge present between the gate and the channel.

Change in threshold voltage is

$$\Delta V_T = -\frac{Q}{C_{ox}} \tag{2.2}$$

There are many ways to obtain the threshold voltage shift. Two are the most common solutions used to store charge:

1. **In traps that are present in the oxide, more precisely at the interface between two dielectric materials:** The most commonly used interface is the silicon oxide/nitride interface. Devices obtained in this way are called metal-nitride-oxide silicon (MNOS) cells.

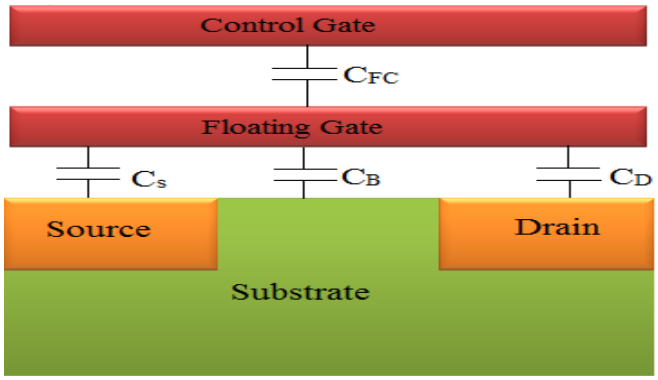

Figure 2.6: Schematic cross section of an FG transistor. [9]

2. **In a conductive material layer between the gate and the channel and completely surrounded by insulator:** The basic concepts and the functionality of an FG device are easily understood if it is possible to determine the FG potential [1-4].The schematic cross section of a generic FG device is shown in Fig. 2.4; the upper gate is the control gate and the lower gate, completely isolated within the gate dielectric, is the FG [1-4]. The FG acts as a potential well, as shown schematically in Fig. 2.5, if a charge is forced into the well, it cannot move from there without applying an external force, the FG stores charge [1-4]. The simple model shown in Fig. 2.4 helps in understanding the electrical behaviour of an FG device; C_{FC}, C_S, C_D, and C_B are the capacitances between the FG and control gate, source, drain, and substrate regions, respectively.

Consider the case when no charge is stored in the FG, *i.e., Q = 0*

Figure shows various capacitive couplings to the floating gate. FG is capacitively coupled to control gate, source, drain and body of the transistor. As we know

Charge = Capacitance × Voltage

For floating gate equation becomes

$$Q = C_{FC}\left(V_{FG} - V_{CG}\right) + C_S\left(V_{FG} - V_S\right) + C_D\left(V_{FG} - V_D\right) + C_B\left(V_{FG} - V_B\right) \quad (2.3)$$

When there is no charge on the floating gate

$$Q = 0 = C_{FC}\left(V_{FG} - V_{CG}\right) + C_S\left(V_{FG} - V_S\right) + C_D\left(V_{FG} - V_D\right) + C_B\left(V_{FG} - V_B\right) \quad (2.4)$$

Where

C_{FC} is the capacitance between floating-gate and control-gate

C_S is the capacitance between floating-gate and source

C_D is the capacitance between floating-gate and drain

C_B is the capacitance between floating-gate and body

V_{FG} is the potential of the floating-gate
V_{CG} is the potential of the control gate
V_D is the potential of the drain
V_S is the potential of the source
V_B is the potential of the body

If we name $C_T = C_{FC} + C_S + C_D + C_B$ the total capacitance of the FG and we define the coupling coefficient $\alpha_J = C_J/C_T$ relative to the arbitrary J electrode. It should be pointed out that

$$V_{FG} = \alpha_G V_{GS} + \alpha_D V_{DS} + \alpha_S V_S + \alpha_B V_B \quad (2.5)$$

Equation (2.5) shows that the FG potential does not depend only on the control gate voltage but on the source, drain, and bulk potentials. If the source and bulk are both grounded, (2.5) can be rearranged as

$$V_{FG} = \alpha_G \left(V_{GS} + \frac{\alpha_D}{\alpha_G} V_{DS} \right) = \alpha_G (V_{GS} + f V_{DS}) \quad (2.6)$$

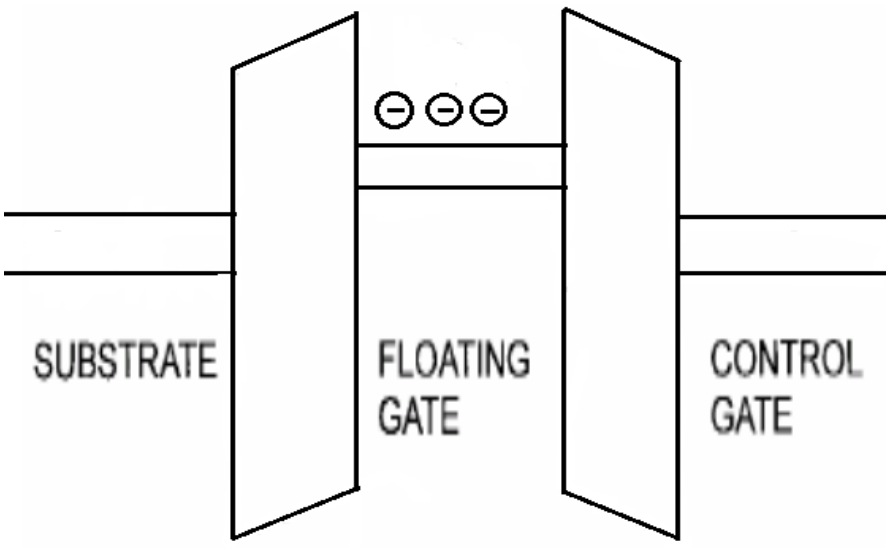

Figure 2.7: Energy band diagram of an FG transistor [9]

2.6 RELIABILITY

Non-volatility of flash memory means retention of data for at least 10 years and data must be retain after many read/write/erase cycle. The basic structure and operation of the floating-gate device has been described. These descriptions have established the mechanisms through which charge is injected and removed from the floating-gate and the means of reading the state of the device. In this section, a very important issue in flash memory technology is discussed – the reliability.

2.6.1 Endurance (Cycling)

The flash memory device is required to retain its properties on being subjected to repeated program/erase cycles. When thin dielectrics are repeatedly stressed at high fields, interface and bulk traps develop in the dielectric [22]. Charge is trapped and released from these traps, and this modifies the fields across the dielectric. This tends to modify the program/erase characteristics over time, as damage is induced in the dielectric.

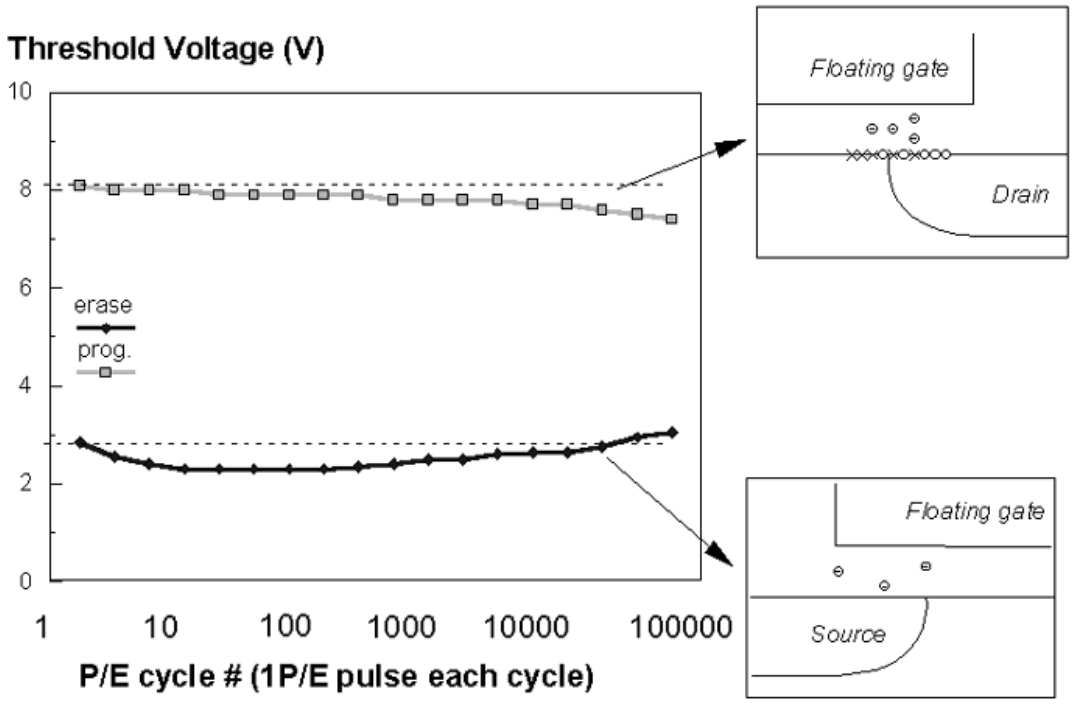

Figure 2.8: Threshold voltage window closure as a function of program/erase cycles on a single cell [9]

2.6.2 Retention

The retention [24] is critical in any non-volatile memory device. The ability to retain charge without supplied power is the definition of non-volatile memory. Retention‟ is a benchmark used to quantify the extent of time for which the stored charge is retained in a device.

The charge-loss in the retention state is determined by tunneling leakage under weak fields through adjoining dielectrics. This charge-loss would be greatly amplified if the dielectrics contained defects, since that would enhance trap-assisted tunneling [23].

A typical retention benchmark for flash memories is 10 years, i.e. the floating-gate device should not lose more than a small fraction of its carriers for 10 years. In practice, it is not possible to quantify retention by doing measurements for such a long period of time. Hence, retention tests are accelerated at higher temperatures. Typically, a 24-hour benchmark at 150 °C is utilized. Loss of charge with time at higher temperatures is measured to make a comparison of retention properties of the devices under study. Mean times to failure and defect density are other metrics used for quantifying reliability of flash memory.

2.7 Nanocrystal Formation

Generally, a metal-oxide-nanocrystal-oxide-silicon non-volatile memory structure is fabricated as follows. First, a silicon (100) wafer is cleaned by standard cleaning process, which removes the native oxide and micro-particles from the wafer surface. A high quality oxide layer of desired thickness is then grown as a tunneling oxide. Afterwards, a charge trapping layer is deposited, and then, depending on the material, a treatment (e.g., primarily thermal) may be necessary to produce nanocrystals. Next, a thick oxide layer is deposited to serve as a blocking oxide. Finally, gate electrodes are deposited and patterned. The process flow and the structure of the memory are shown in Figure 2.9

There are numerous methods that can be used to form nanocrystals as storage centres for non-volatile memory applications. The most commonly used are self-assembly, precipitation, and chemical reaction, as described below.

2.7.1 Self-assembly

The basic procedures of self-assembly for nanocrystal formation are shown in Figures. A trapping layer of 1 – 5 nm is deposited and then the film is annealed at a temperature close to its eutectic temperature in an inert ambient gas to transform the trapping layer into a nanocrystal structure. The diameter of the nanocrystal is influenced by the thickness of the

trapping layer, as well as the temperature and duration of the thermal treatment. Dispersion forces and the electrical double layers affect the nanocrystal size and location distributions [25].

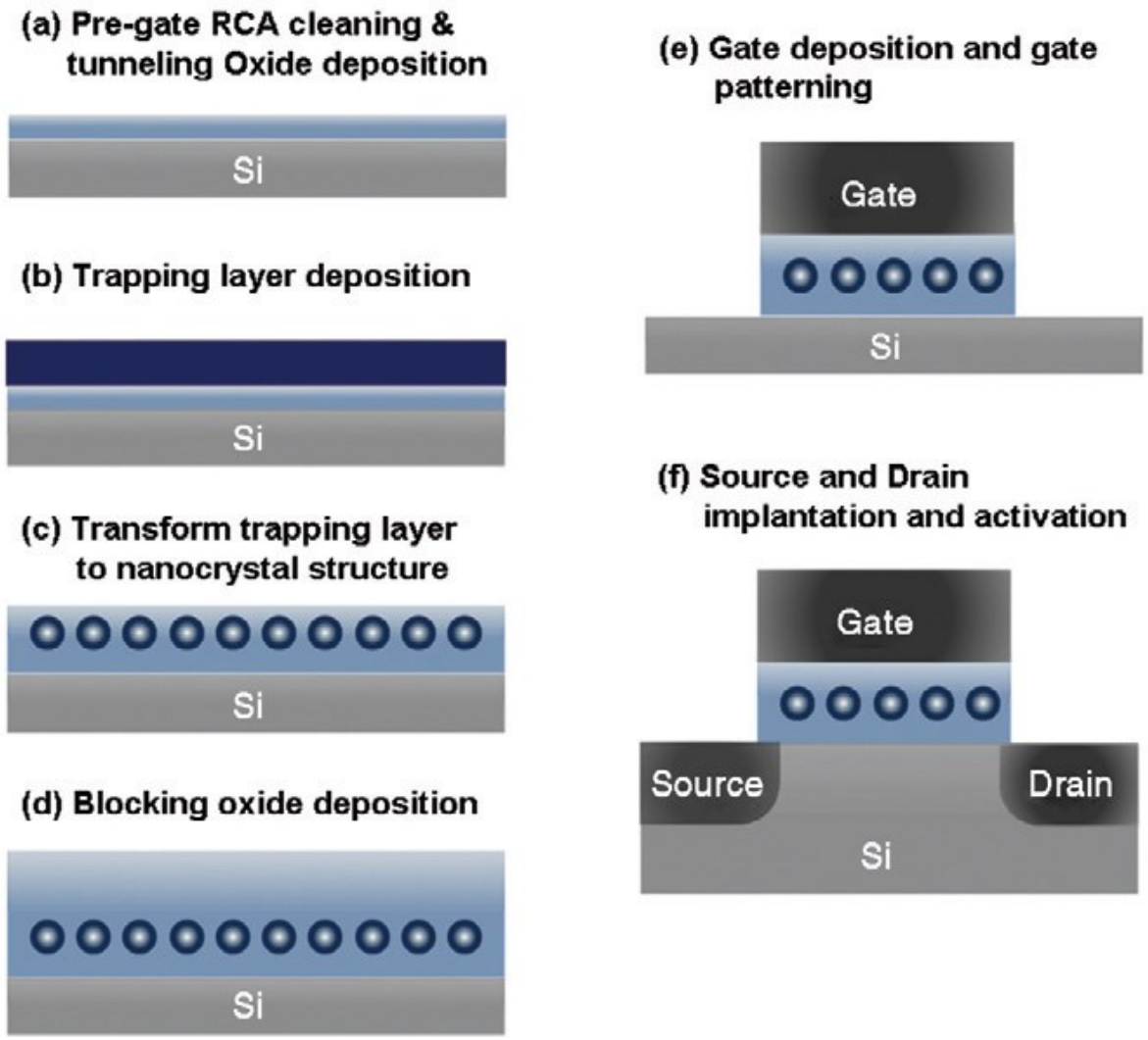

Figure 2.9: Process flow and structure of nanocrystal non-volatile memory [27]

This process is accomplished through the relaxation of film stress and is limited by the surface mobility. During the thermal treatment, these atoms gain enough surface mobility, allowing the film to self-assemble into the more thermodynamically and energetically stable state. The film breaks into “islands” along the initial perturbation to reduce the elastic energy carried by the stress built into the film during the deposition process. Moreover, minimization of the surface energy and the dispersion force between the top and bottom interfaces is conducive to stabilizing the film. Therefore, the final geometry depends on the balance between these driving forces. The forces between those regions helps stabilize the

nanocrystals and keep a uniform distance between them. However, using self-assembly for nanocrystal formation cannot ensure that the trapping layer is completely discrete.

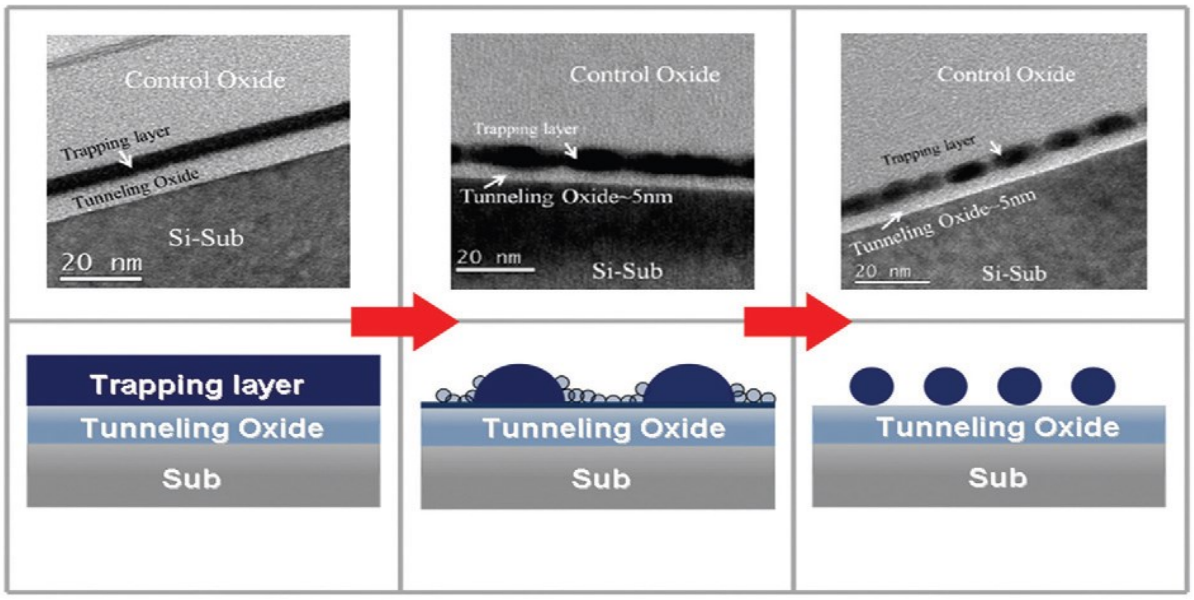

Figure 2.10: TEM image and schematic drawing of nanocrystal formation by self-assembly with increased duration of thermal treatment [27]

2.7.2 Precipitation

An oversaturated or mixed trapping layer is prepared by ion implantation into a deposited insulator layer or co-deposit system to form nanocrystals by further thermal annealing. The nucleation of nanocrystal formation during rapid thermal annealing (RTA) process can be described as following. Initially, oversaturated solid-solutes separate during thermal annealing. As the temperature and duration of the thermal annealing increases, these educts can obtain enough energy to leave their initial sites and diffuse through the mixed film. Educt diffusions with a large enough number of collisions will result in nuclei formation. With the increase in annealing temperature, more educts tend to bond to the nuclei and form the nanocrystal structure in the trapping layer, forming a high density distribution of nanocrystal structures, as shown in Figure 2.11. However, employing traditional high-energy ion implantation for nanocrystal memory applications has revealed some obvious short comings. The ion distribution (following a Gaussian distribution) is wide with high injection energy (~30 – 150 keV), such that controlling the nanocrystals close to the tunneling oxide is difficult. In addition, even when the implanted ion reaches the interface between the silicon

oxide and silicon substrate, it may cause damage to the tunneling oxide, resulting in degradation of the device performance.

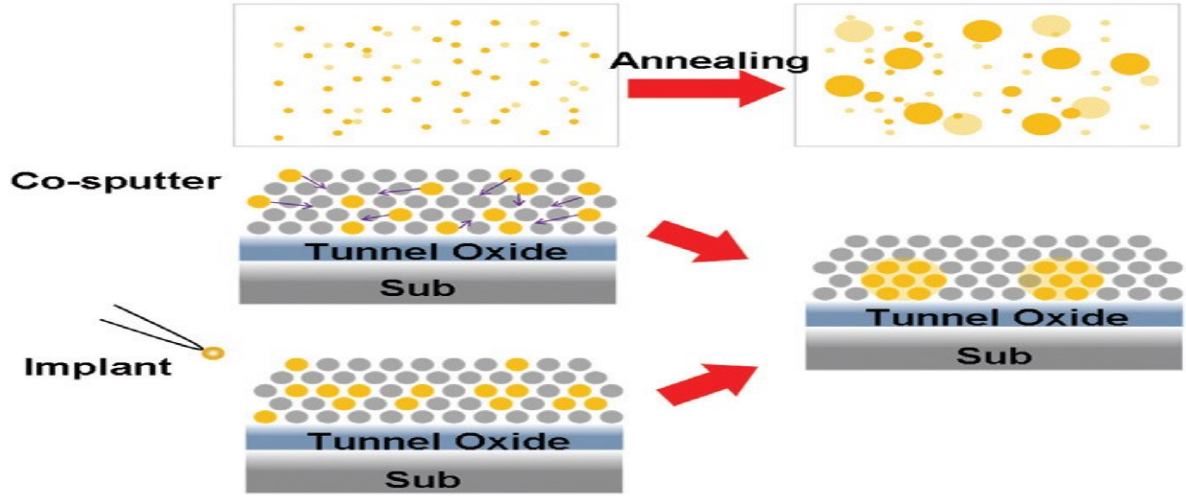

Figure 2.11: Schematic drawing of nanocrystal formation by precipitation [27]

2.7.3 Chemical reaction

The chemical reaction method is widely used to form a nanocrystal trapping layer and is illustrated in Figure 2.12 Initially, a binary or tertiary mixed layer is co-deposited by different material systems, and then the layer is oxidized by RTA under an oxygen flow. Different materials have different oxidation capabilities due to their different Gibbs free energies; it is easier to form a stable oxide compound when it has lower enthalpy. Therefore, during RTA oxidation, one material comprising the binary mixed layer is typically more easily oxidized while the other material tends to conjugate with other atoms of the same material to form the nanocrystal.

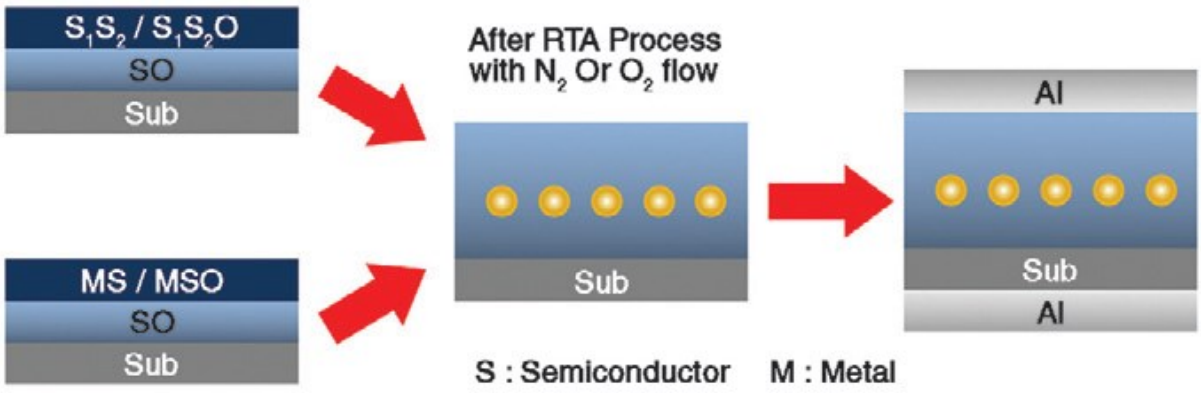

Figure: 2.12 Schematic drawing of nanocrystal formation by chemical reaction [27]

2.8 Scaling Limitation of Memory Devices

The cost for semiconductor memories (*i.e.*, DRAM, SRAM, ROM, and flash) is largely determined by the amount of silicon area that takes to store a data bit of information [5, 6]. As with other semiconductor memories, flash memory, which retains its data even when the power is removed, achieves higher density and lower cost through traditional silicon process scaling techniques, such as feature size reduction? However, as the industry begins manufacturing at smaller geometries – 90nm and smaller –manufacturing floating gate-based flash becomes impractical. At those dimensions, the chip area spent on the 9-12V high voltage transistors needed to write and erase the flash becomes too expensive. It was difficult for Engineers to scale down the high voltage in floating-gate based flash without compromising reliability, at the risk of memory failures and loss of data. Chip giant Intel illustrates the limit of tunnel oxide thickness is 8nm due to Stress Induced Leakage Current (SILC). It was explained that the operating voltages cannot be reduced below 10-12V for programming and 20V for erase. Further, it was reported that the cell scaling below 0.13 μm is very difficult to meet unless there is some technology breakthrough.

With the contemporary technologies scaling down of thickness of tunnel oxide in a flash EEPROM is inevitable to reduce cell size and operating voltage. The following considerations indicate that scaling has a physical limit. Because of direct tunneling and an increase in SILC currents, tunnel oxide cannot be scaled much below 60Å, while yield considerations limit oxide thickness to 80Å. The electron barrier height of 3.2eV prevents significant reduction in programming voltages. Without a reduction in tunnel oxide thickness, the erase voltage cannot be reduced. Since operating voltages cannot be reduced, in order to prevent high voltage breakdown/punch through and parasitic FET turn-on the minimum channel length and isolation spacing between devices cannot be reduced. On the other hand multilevel storage technology provides alternate solution to process scaling. Feasibility studies were done up to 4 bit or 16 levels with stacked gate technology. Technological breakthrough with nanocrystals is not only sensitive to the charge of a single electron, high density, low power, fast write/erase, long retention time but also multilevel storage [7]. Si based nanoscale devices are strong contenders due to the existing Si process infrastructure as well as the nearly perfect interface between SiO_2. Motorola has demonstrated that nanocrystal memory has significant advantage over SONOS memories when operating with HCI for programming and F-N for erase having tunnel oxide as 5nm. Nanocrystal memories allow reduction of module area by a factor of 2. Hence with NC based memory devices there is

tremendous. Si floor and module area saving and hence the cost can be reduced with existing Si manufacturing process, see Fig. 2.7 [7].

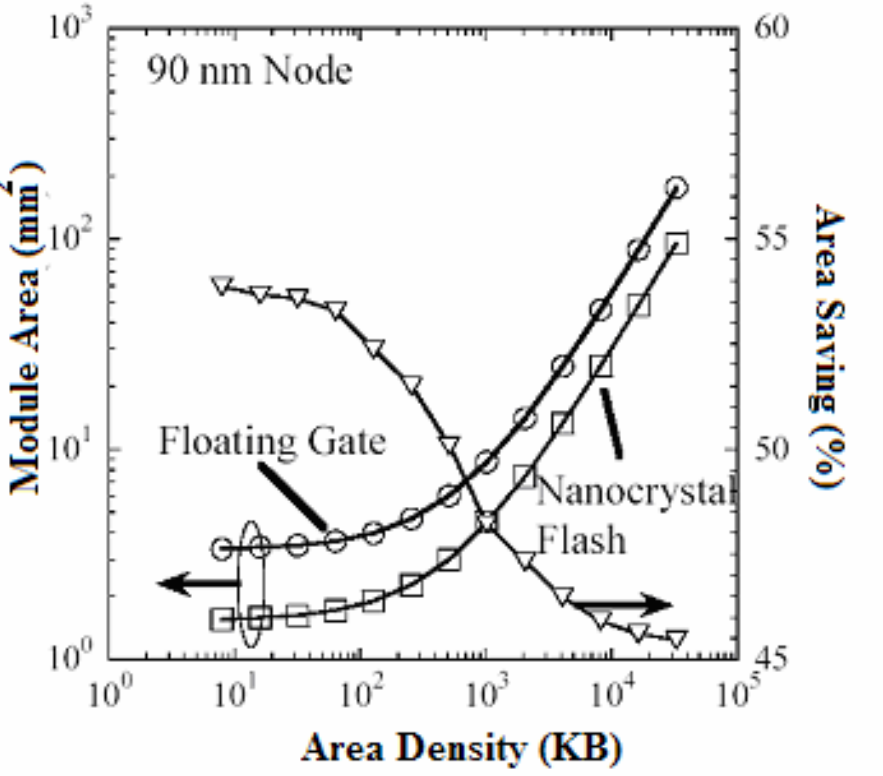

Figure 2.13: Area saving & module area comparison between nanocrystal memory and conventional floating gate. [6]

2.9 High-k dielectric as Control Oxide

In order to improve GCR, the conventional control oxide needs to be replaced by a high-k dielectric. Due to the increased dielectric constant, there is greater coupling between the control gate and the floating-gate, and hence the GCR is improved. According to ITRS, target high-k EOT is 8-10 nm from 2012-2024.

Gate coupling ratio GCR = C_c / C_{TOT},

Where

C_c = Capacitance between Control Gate and Floating Gate

C_{TOT} = Total capacitance to Floating Gate

Since C_c = k * €o * Area / t

Where

k = dielectric constant of control oxide

€o = vacuum permittivity

Area = Area between Control Gate and Floating Gate

t = thickness of control oxide dielectric

We have

$$GCR \alpha k$$

Hence the use of high-k control oxide to improve GCR

2.9.1 Choice of high-k dielectric for Flash memory application

Requirements of high-k control oxide

- Effectively block charge loss from the floating-gate: This implies a high-quality dielectric without defects, and good conduction band barrier height.
- Must not trap charge: The high-k dielectric must not be trap charges as this will lead to undesired level shifts, and retention issues related to detrapping.
- Large conduction-band barrier height: Among other things, a large conduction band barrier height is important to prevent gate-injection. Gate injection is the undesirable injection of carriers from the control gate into the floating gate during the erase operation. When gate injection is equal to the charge flowing out from the floating-gate into the substrate, the cell cannot be erased anymore. This condition is called erase-saturation. A large conduction band barrier height helps alleviate the erase-saturation issue because it reduces the amount of gate-injection.
- Physically thin: The high-k dielectric IPD cannot be physically thick because fringing field lines through the high-k dielectric will cause interference between adjacent cells.

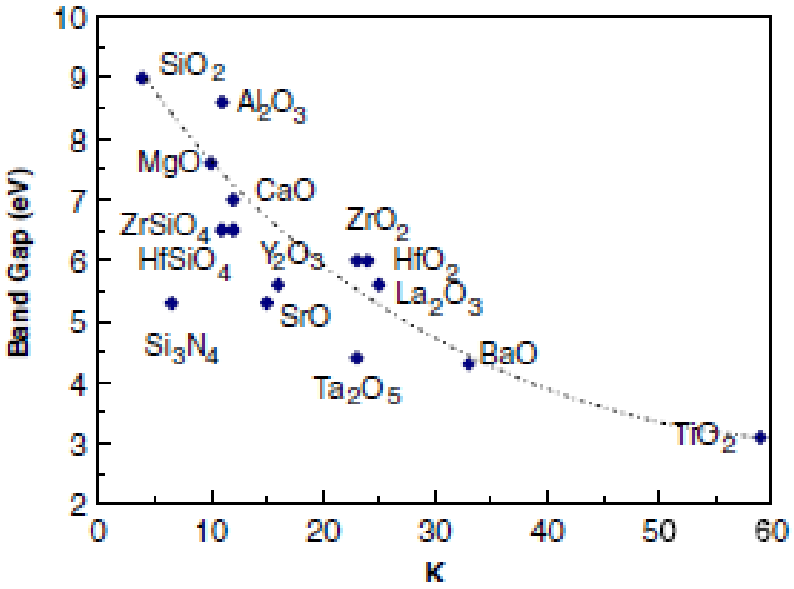

Figure 2.14: Common high-k dielectrics. In general, larger the dielectric constant, smaller the band-gap [29]

2.9.2 Problems of high-k control oxide

- Most high-k dielectrics have small band-gap and barrier height. This leads to gate-injection and charge loss issues.
- High-k materials good enough for logic may not be suitable as control oxide since they may be trappy at the large thicknesses (8-10 nm EOT) used.
- High-k dielectrics tend to have earlier dielectric breakdown.
- High-k dielectrics are known to form a poor interface with poly-silicon

Chapter 3

Literature Survey

D. Kahng and S.M. Sze in 1967 [3] have described the basic structure of floating gate transistor and physics of the floating gate transistors. Concept of band bending in floating gate transistor was also presented in this work. Different concept of injection of charge like hot electron emission and F-N tunneling was also presented by authors

M. Lenzlinger and E. H. Snow in 1969 [11] explained the electronic conduction in thermally grown SiO_2. Fowler-Nordheim characteristics have been observed over more than five decades of current for emission from Si, Al, and Mg. If previously measured values of the barrier heights are used, the slopes of the Fowler-Nordheim characteristics ($\log J/E^2$ vs. $1/E$) imply values of the relative effective mass in the forbidden band of about 0.4. These values take into account corrections for image-force barrier lowering and for temperature effects. The absolute values of the currents are lower by a factor of five to ten than the theoretically expected values, probably due to trapping effects.

A.Thean and J.Pierre Leburton in 2002 [10] described the concept of flash memory and the scaling issues related to flash memory. Complete description of conventional continuous floating gate memory was given and the future scope was also explained by the authors. As VLSI industry moving towards the advanced technology nodes new concept is added in the flash memory as scaling the tunnel oxide leads to leakage problem and the retention time is less. The main problems are related to high field stressing and the leaky scaled-down oxide barrier. Currently, there are efforts to find high-permittivity (high-k) dielectric materials to replace SiO_2 as the gate dielectric. High-K dielectric materials would be able to maintain an equivalent potential difference between the floating gate and the device body for a greater thickness compared to SiO_2. The leakage through the dielectric would be minimized and the scaling limits would be extended. To solve the above problem concept of the nitride memory was described. After that Controlling the location and the distribution of charge storage nodes in the dielectric found ground with nanocrystal memory. IBM researchers first proposed a flash memory with a granular floating gate made out of silicon nanocrystals. This high-speed

flash memory was based on direct quantum tunneling. Also, compared to the atomic-size nitride traps, electron or hole energy states are energetically deeper in the nanocrystal wells. For instance, the energy levels of traps are typically 1-2 eV below the Si_3N_4 conduction band, while the electron ground state in nanocrystals could exceed 3 eV below the SiO_2 conduction band. Hence, if the nanocrystals are well separated in the 2-D layer, then conduction between the nodes is totally prevented. Besides that, the transport of charges is restricted to carrier exchanges with the Si channel. Conceptually, the increased confinement and the reduction in leakages imply that the tunnel oxide can be more aggressively scaled down. With a thin tunnel oxide, direct quantum-mechanical tunneling can be exploited as a transport mechanism for programming and erasing the charges in the nodes.

S. Tiwari, *et al.* in 1996 [3] described a new memory structure using threshold shifting from charge stored in nanocrystals of silicon (5nm in size) is described. The devices utilize direct tunneling and storage of electrons in the nanocrystals. The limited size and capacitance of the nanocrystals limit the numbers of stored electrons. The retention times are also measured and found no degradation in the performance. This nano memory exhibits characteristics necessary for high density and low power

P. Pavan, *et al.* in 1997 [1] demonstrated the comparative analysis of different type of MOS memories is discussed. This paper will mainly focus on the development of the NOR Flash memory technology, with the aim of describing both the basic functionality of the memory cell used so far and the main cell architecture consolidated today. The NOR cell is basically a floating-gate MOS transistor programmed by channel hot electron and erased by Fowler–Nordheim tunneling. The main reliability issues, such as charge retention and endurance, will be discussed, together with the understanding of the basic physical mechanisms responsible. Most of these considerations are also valid for the NAND cell, since it is based on the same concept of floating-gate MOS transistor. Concept of multilevel approach is also discussed in which multiple bits can store at a time. The exploitation of multilevel concept at advanced technology increases the efficiency and reduces the cost per bit. Issue of scaling is also discussed in this work. The crucial issues to be solved to allow cell scaling below the 65-nm node is the tunnel oxide thickness reduction, as tunnel thinning is limited by intrinsic and extrinsic mechanisms.

R. Bez, *et al.* in 2003 [2] gives a thorough overview of Flash memory cell. Basic operations of flash memory and charge-injection mechanisms that are most commonly used in actual Flash memory cells are reviewed to provide an understanding of the underlying physics and principles in order to appreciate the large number of device structures, processing technologies, and circuit designs presented in the literature. New cell structures and architectural solutions have been surveyed to highlight the evolution of the Flash memory technology, oriented to both reducing cell size and upgrading product functions. The subject is of extreme interest: new concepts involving new materials, structures, principles, or applications are being continuously introduced. The worldwide semiconductor memory market seems ready to accept many new applications in fields that are not specific to traditional non-volatile memories.

J. D Blauwe in 2004 [28] in provides an overview of nanocrystal memories—a nascent nonvolatile memory technology that promises to extend the scaling of more conventional charge storage devices to nanometer-scale dimensions. The concept of ONO stack as the control oxide is disscussed in this paper. Different fabrication techniques of the formation of nanocrystal memory cell disscused in this paper

C.C Monzio, *et al.* in 2007 [30] worked on nanocrystal flash memory, the silicon nanocrystal memory technology has received widespread interests from the scientific community working in the field of non-volatile solid-state memories, considering it as a feasible candidate for the post-Flash scenario. The immunity to stress-induced leakage current and the reduction of parasitic floating-gate capacitive couplings make the nanocrystal technology very attractive, especially when considering the CMOS compatible process flow. However, many open issues still exist for its development, first of all concerning its scaling perspectives. Starting from the discussion of the basic principles of nanocrystal storage, in this paper review of the major benefits and the open challenges of the silicon nanocrystal memory technology.

Kaushik Roy, *et al.* in 2003 [31] discussed different scaling issues in advanced technology nodes. As high leakage current in deep-submicrometer regimes is becoming a significant contributor to power dissipation of CMOS circuits as threshold voltage, channel length, and gate oxide thickness are reduced. Consequently, the identification and modelling of different leakage components is very important for estimation and reduction of leakage power,

especially for low-power applications. In this work various transistor intrinsic leakage mechanisms, including weak inversion, drain-induced barrier lowering, gate-induced drain leakage, and gate oxide tunneling. Channel engineering techniques including retrograde well and halo doping are explained as means to manage short-channel effects for continuous scaling of CMOS devices. Different circuit techniques are also discussed which can used to reduce the leakage power consumption.

Anurag Chaudhry, *et al.* in 2004 [32] the performance degradation of a MOS device fabricated on silicon-on-insulator (SOI) due to the undesirable short-channel effects (SCE) as the channel length is scaled to meet the increasing demand for high-speed high-performing ULSI applications. The review assesses recent proposals to circumvent the SCE in SOI MOSFETs and a short evaluation of strengths and weaknesses specific to each attempt is presented. A new device structure called the dual-material gate (DMG) SOI MOSFET is discussed and its efficiency in suppressing SCEs such as drain-induced barrier lowering (DIBL), channel length modulation (CLM) and hot-carrier effects, all of which affect the reliability of ultra-small geometry MOSFETs, is assessed.

P. M. Solomon, *et al.* in **[]** worked on different MOSFET structure, which is used to suppress the short channel effect in MOSFET with advancement in technology. Simulation study of different structure like double gate, MOSFET with ground plane and single gated ultra-thin SOI MOSFET is also discussed. Two dimensional effects in gate insulator and buried insulator are also discussed.

B. Govoreanu, *et al*. in 2005 **[]** discussed the main issues in scaling down the interpoly dielectric (IPD) for future floating gate Flash memory technology generations has been reviewed. The equivalent oxide thickness (EOT) of the IPD must reach the sub-10 nm range to enable lowering of the operating voltages and further scale device feature sizes. Additionally, the loss of control gate wrap around the floating gate for high density memories as device dimensions scale down will require a drastic reduction in IPD EOT to maintain the same capacitive coupling. As the scalability of the conventional oxide-nitride-oxide (ONO) IPD's is limited, new solutions that exploit the opportunities offered by the high-k dielectrics. Their effectiveness increases when midgap and p-type metals are considered, instead of the conventional polysilicon control gate. The optimal approach depends on the most critical requirement that the control oxide has to fulfil, which in turn is application or device-

structure dependent. The non-ideal nature of the dielectric materials, however, may severely reduce the design window, calling for a sustained effort to improve their electrical properties by process optimization.

J. Robertson in 2004 [29] explained the concept of high-k dielectrics. As the scaling of complementary metal oxide semiconductor (CMOS) transistors has led to the silicon dioxide layer used as a gate dielectric becoming so thin (1.4 nm) that its leakage current is too large. It is necessary to replace the SiO2 with a physically thicker layer of oxides of higher dielectric constant or 'high-*k*' gate oxides such as hafnium oxide and hafnium silicate. Little was known about such oxides, and it was soon found that in many respects they have inferior electronic properties to SiO_2, such as a tendency to crystallise and a high concentration of electronic defects. Intensive research is underway to develop these oxides into new high quality electronic materials. This review covers the choice of oxides, their structural and metallurgical behaviour, atomic diffusion, their deposition, interface structure and reactions, their electronic structure, bonding, band offsets, mobility degradation, flat band voltage shifts and electronic defects. The use of high-*k* oxides in capacitors of dynamic random access memories is also covered.

Ni Henan, *et al*. in 2009 [26] worked on the MOS (metal oxide semiconductor) capacitor structure with double-layer heterogeneous nanocrystals consisting of semiconductor and metal embedded in a gate oxide for non-volatile memory applications has been fabricated and characterized. By combining vacuum electron-beam co-evaporated Si nanocrystals and self-assembled Ni nanocrystals in a SiO2 matrix, an MOS capacitor with double-layer heterogeneous nanocrystals can have larger charge storage capacity and improved retention characteristics compared to one with single-layer nanocrystals. The upper metal nanocrystals as an additional charge trap layer enable the direct tunneling mechanism to enhance the flat voltage shift and prolong the retention time.+

G. Chakraborty, *et al.* in 2011 [17] In this work, a comparative performance of the silicon (Si) and germanium (Ge) nanoparticles embedded SiO_2 floating gate MOS memory devices. In such devices for low applied fields, the tunneling current is dominated by the direct tunneling mechanism, whereas for higher electric fields, the Fowler–Nordheim tunneling mechanism dominates. As the device dimensions get smaller, problem arises in the

conventional MOS memory devices due to the leakage through the thin tunnel oxide. This leakage can be reduced via charge trapping by embedding nanoparticles in the gate dielectric of such devices.To reduce the write voltage, by lowering the onset voltage of the Fowler–Nordheim tunneling mechanism. Results showed that both the Si and the Ge nanoparticles embedded gate dielectrics offer reduction of the leakage current and a significant lowering of the writing or programming onset voltage, compared to the pure SiO_2 gate dielectric. In terms of the comparative performance, the Germanium nanoparticles embedded gate dielectric showed better results compared to the silicon nanoparticles embedded one.

CHAPTER 4

Design of Nanocrystal Flash Memory Cell

4.1 Introduction

In this chapter, the design of following nanocrystal flash memory cell has been discussed. Following two device structure designs will discuss in this chapter.

1. Single gate nanocrystal Flash memory
2. Double gate nanocrystal flash memory

The above structures are designed and characterized on SILVACO TCAD tool using ATLAS framework. SILVACO modules can be invoked by the use of the ***go*** command and the simulator name. Here work has been done on ATLAS framework. ATLAS simulator is used for the electrical characterization of the device

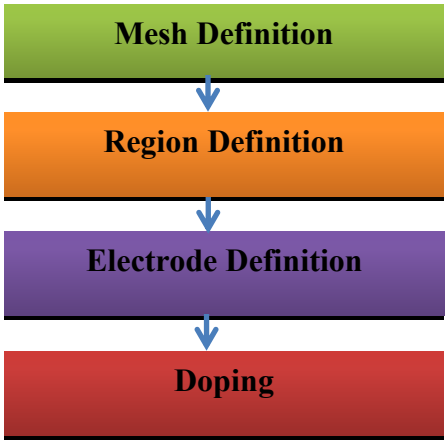

Figure 4.1: Flow of design in SILVACO Atlas

New proposed device, in which concept of double gate and nanocrystal memory is combined is known as double gate nanocrystal flash memory. The difference between two structure can predict from their names .Single gate nanocrystal memory cell have only one control gate which control the control conduction of the device . Double gate nanocrystal have two gates on opposite side of the channel, both the gates control the conduction in channel.

Steps required to design both the structures is explained below is in sequence.

4.2 Design of Single gate nanocrystal flash memory cell

S.no	Parameter	Value
1.	Channel length	130 nm
2.	Nanocrystal Size	5 nm
3.	Distance between nanocrystals	5 nm
4.	Source / Drain Doping	1e20 cm^{-3}
5.	Substrate Doping	3e17 cm^{-3}
6.	Gate Material	Aluminium
7.	Tunnel Oxide	5 nm
8.	Control Oxide	13 nm

Table 4.1: Design Parameter for Single Gate Nanocrystal Flash Memory

4.2.1 Mesh Definition

After invoking ATLAS with the command of **go atlas**, the x-line, setting the mesh with too few lines can cause the calculation to be inaccurate, while setting the grid with too dense mesh can consume more time for ATLAS to calculate in the characterization part. At the vertical line or y-line, the bigger the size of the lines could contribute to the increment of the weight of the actual fabricated device.

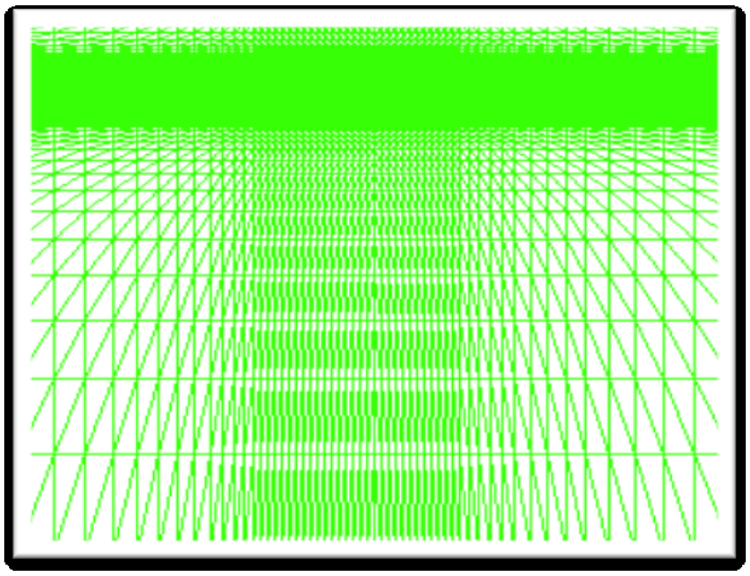

Figure 4.2: Grid lines for single gate nanocrystal flash memory cell

3.2.2 **Region Definition**

After defining the mesh, region is defined, in the desired structure following regions are defined 1) Oxide Region 2) Substrate region. Oxide region is divided in two parts one is tunnel oxide and second is interpoly oxide or control oxide.

Figure 4.3: Substrate and Oxide Region for single gate nanocrystal flash memory

3.2.3 Electrode Definition

After the formation of body, gate, source and drain, the structure is ready for the next step, which is depositing and patterning the contact. When contact is deposited, the layer of metal is electrically interconnected the device fabricated on the silicon substrate. The material used for the contact is Aluminium, since it has a very low resistivity (high conductivity) and its adhesion compatibility with SiO2. Electrode for single gate nanocrystal memory cell is shown in figure 4.4

3.2.4 Doping

Doping is done to change the conductivity of device different region have different doping concentration which is given in the table 4.1. In this work substrate have different doping than drain and source. Drain and source have same doping concentration. Source/Drain is

highly doped n-type, whereas substrate is of p type. Doping profile of different region is shown in the figure given below

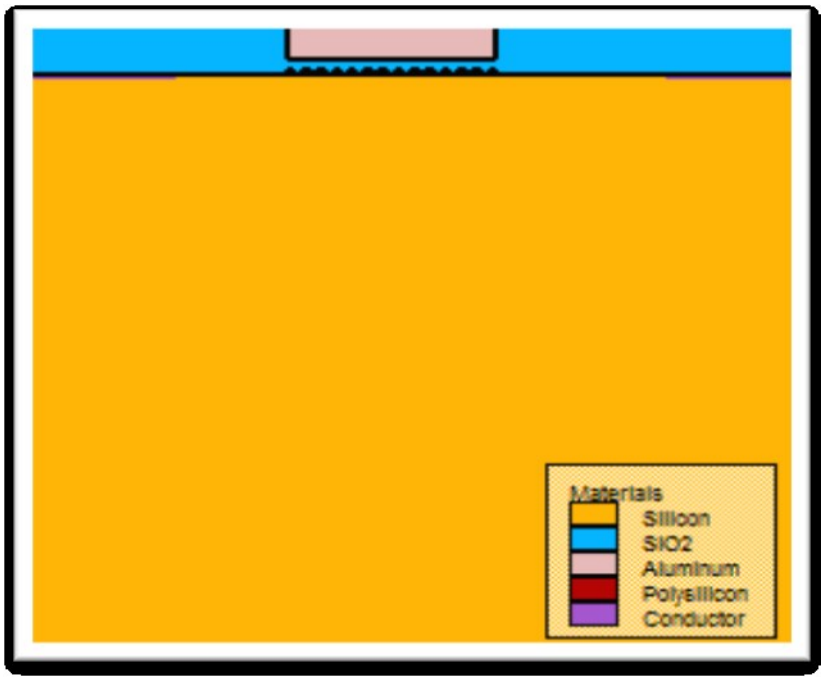

Figure 4.4: Electrodes in single nanocrystal memory cell

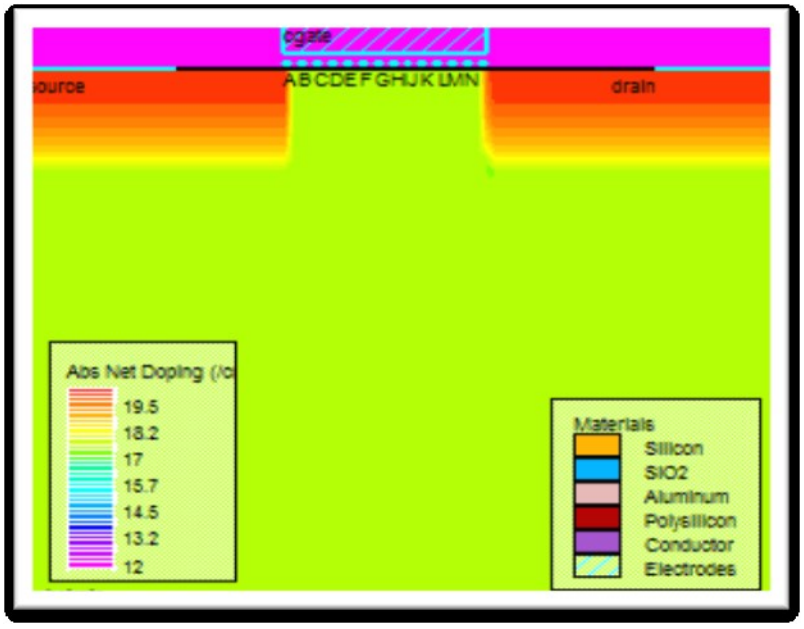

Figure 4.5: Doping of different regions in single gate nanocrystal memory cell.

Above parameters of Table 4.1 are used to define the structure of the single gate flash memory cell. Control gate is represented by cgate. Source and drain are marked in the Fig.

4.6. Nanocrystals are marked by A,B,C,D,E,F,G,H,I,J,K,L,M,N from source to drain respectively in the figure below.

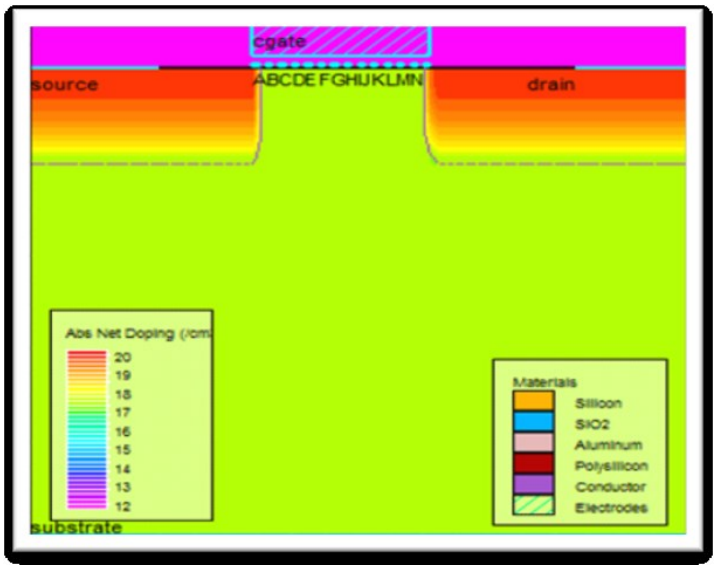

Figure 4.6: Structure of nanocrystal memory cell.

4.3 Design of double gate nanocrystal flash memory cell

All the steps in design the double gate nanocrystal flash memory cell is same as single gate nanocrystal flash memory. Difference between these two structures is the addition of control gate 2 and nanocrystals marked as A1 to N1 in double gate nanocrystal flash memory cell Parameters for double gate nanocrystal flash memory is showing in Table 4.2

S.no	Parameter	Value
1.	Channel length	130 nm
2.	Nanocrystal Size	5 nm
3.	Distance between nanocrystals	5 nm
4.	Source / Drain Doping	1e20 cm^{-3}
5.	Body Doping	3e17 cm^{-3}
6.	Gate Material	Aluminium
7.	Tunnel Oxide	5 nm
8.	Control Oxide	13 nm

Table 4.2: Design Parameter for Double Gate Nanocrystal Flash Memory Cell

Design flow of double gate nanocrystal memory is shown in figures given below.

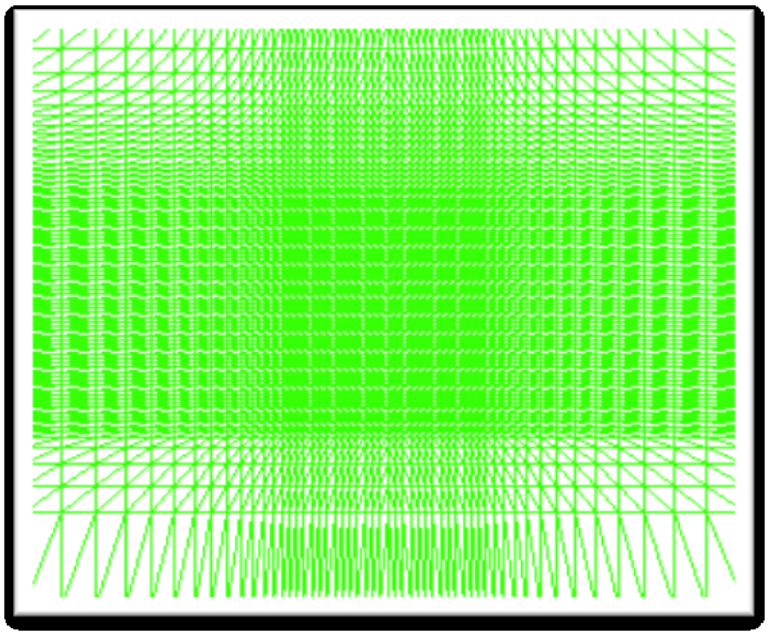

Figure 4.7: Grid lines for double gate nanocrystal flash memory cell

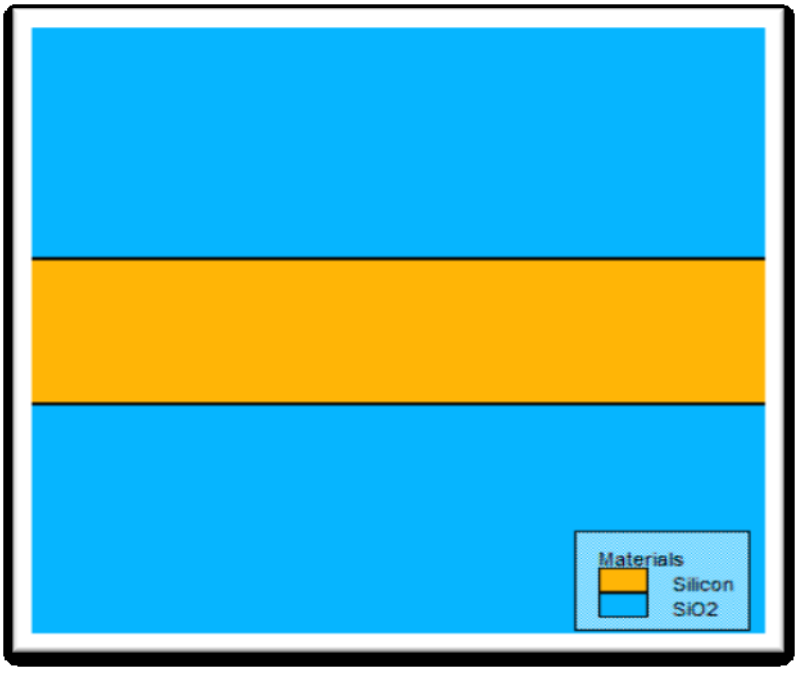

Figure 4.8: Oxide and body region of double gate nanocrystal flash memory cell.

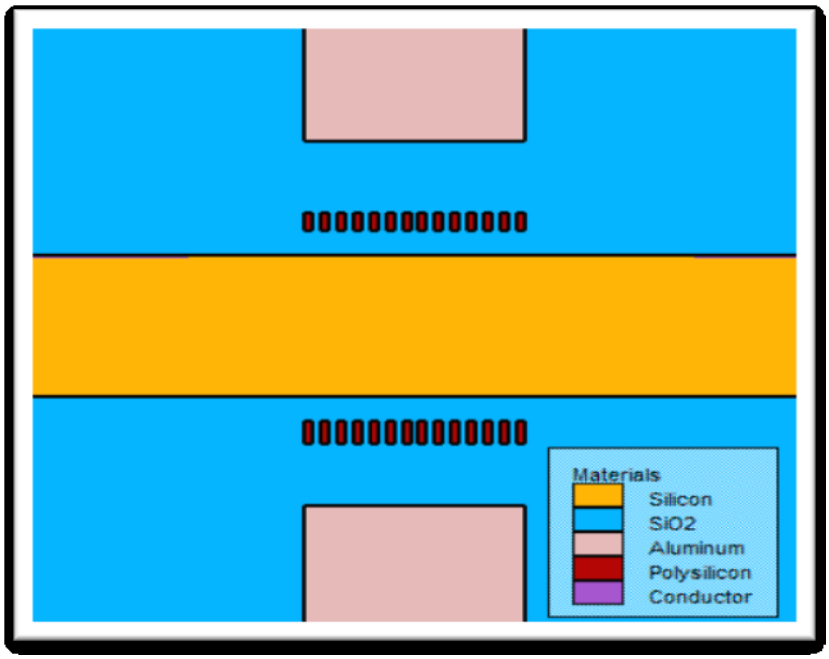

Figure 4.9: Electrodes for double gate nanocrystal flash memory cell.

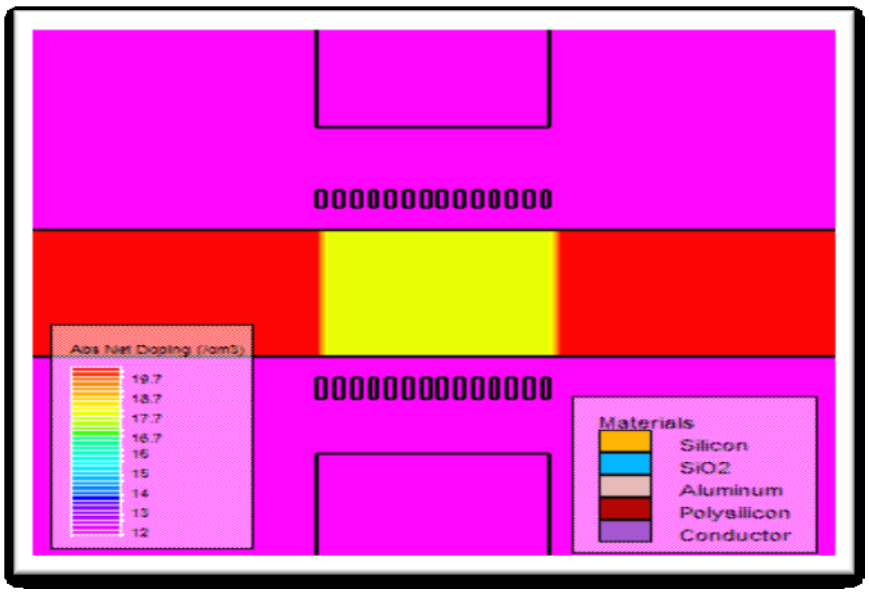

Figure 4.10: Doping in different regions of double gate nanocrystal flash memory cell.

Fig 4.11 shows the complete struture of double gate nanocrystal flash memory cell . From fig it can be seen that there are two control gates one is represented by cgate and other is represented by cgate1 in Fig. 4.11. There are 14 nanocrystals above the body region and 14 below the body region . Source and drain are highly doped n type and body is of p type semicondutor. In this work silicon di oxide is used as control oxide or interpoly dielectric and tunnel oxide .

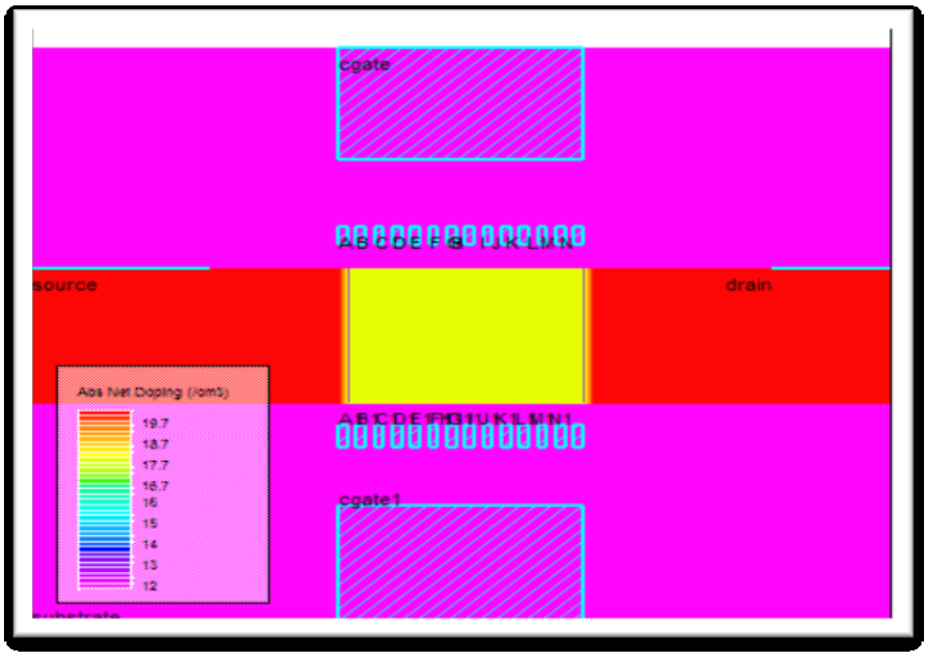

Figure 4.11: Structure of double gate nanocrystal flash memory cell.

CHAPTER 5

Result and Discussion

Simulation study of single gate and double gate flash memory shows that performance of the double gate nanocrystal is better than single gate nanocrystal flash memory in term of faster write and erases operation. On applying same gate voltage for fixed duration of time, charge stored on the double gate nanocrystal memory is more than single gate flash memory which can be clearly seen by observing the change in threshold voltage of the devices. For single gate flash memory observed change in threshold voltage is 4.158 V and for double gate nanocrystal flash memory change in threshold voltage is 5.807 V i.e. an increase of 1.649 V.

Write Operation

Write operation of flash memory is done by applying a postive gate pulse voltage on the control gate. On applying the postive gate voltage electron from channel start tunneling from channel to nanocrystals.The charge store on the nanocrystal with write pulse of +18 V for 0.1 seconds is shown in the figure 5.1. Charge storage on the nanocrystal depends upon the time of write pulse and voltage applied on the control gate. As the time increases, charge on the nanocrytsals increase which result in increase in threshold shift.

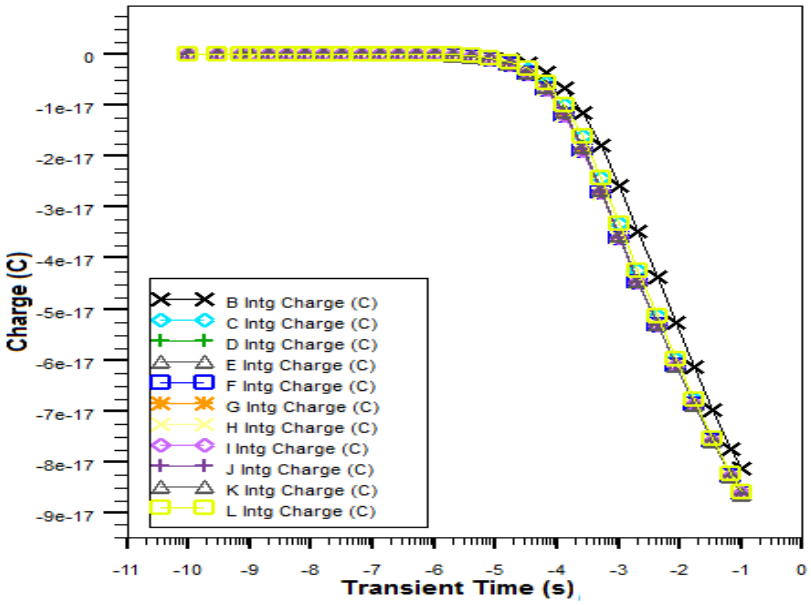

Figure 5.1: Charge during write operation in single gate nanocrystal memory cell

Erase Operation

Erase operation of flash memory is done by applying a neagtive gate pulse voltage on the control gate . Electron which store on the nanocrystals can tunnel back to substate or body of the device and change in the threshold backs to its original value. F-N tunneling techinque is used for erasing of nanocrystal memory cell. Figure 5.2 shows the charge on the different nanocrystals during erase operation in case of single gate nanocrystal memory cell. Initially charge stored on the nanocrystal is about 10^{-17} cm^{-3} , after applying neagtive voltage pulse , tunneling of electron start from nanocrystals to the channel and charge on nanocrystal starts decreasing and finally charge on the nanocrystal become approximately zero.

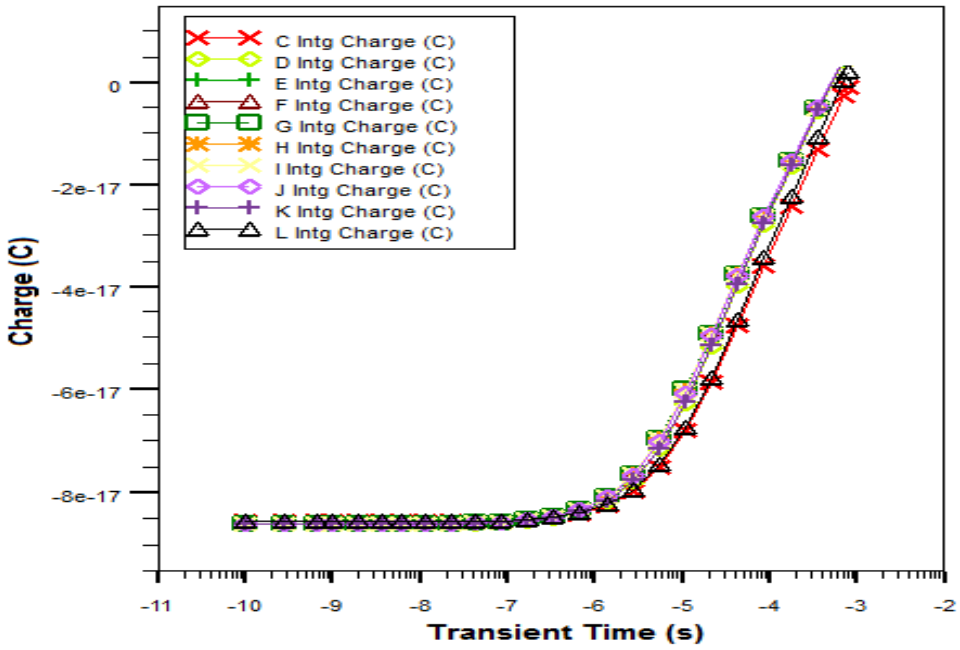

Figure 5.2: Charge during erase operation in single gate nanocrystal memory cell

This change in threshold voltage is reflected by shift in drain current and control voltage. Two different states of transistor form state ‘1’ and state ‘0’ of nanocrystal memory. The difference between threshold voltage for program and erase at particular gate voltage is known as program/erase window.

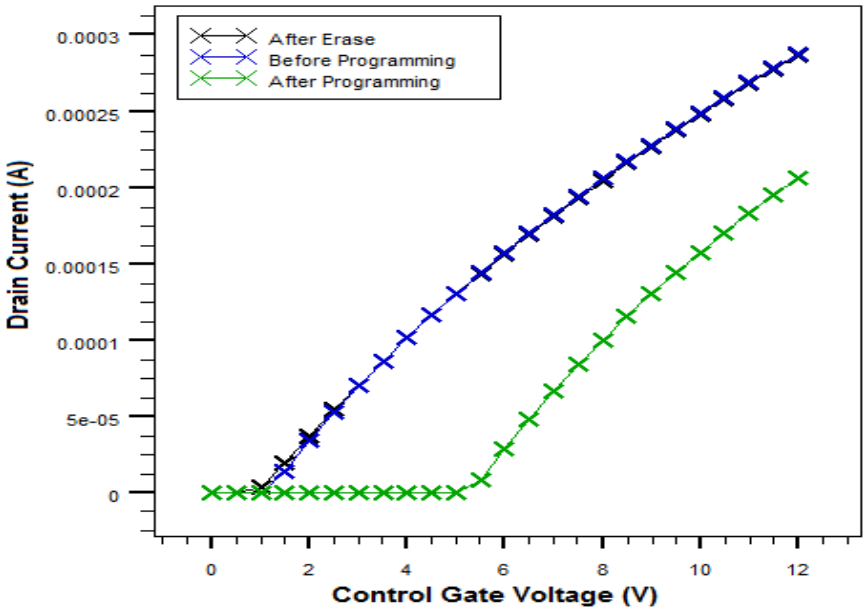

Figure 5.3: Threshold voltage shift in double gate nanocrystal memory cell

Charge on nanocrystals can increase by increasing the time duration of write pulse or by increasing the gate voltage of the device, fixed amount of charge can induce by increasing the time of write pulse for fixed voltage or by increasing the control gate voltage for fixed duration of write pulse. Hence there is trade off between power and speed.

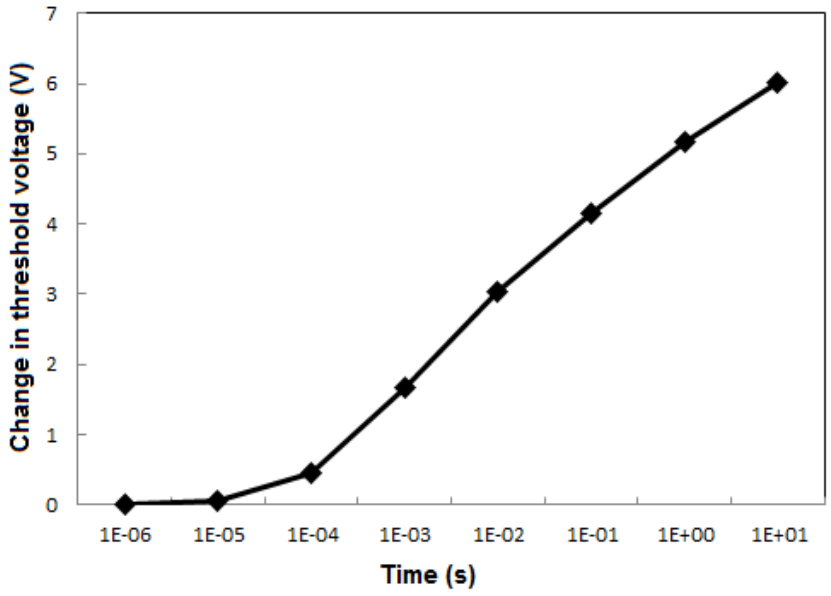

Figure 5.4: Change in threshold voltage of single gate nanocrystal flash memory cell with change in duration of write pulse.

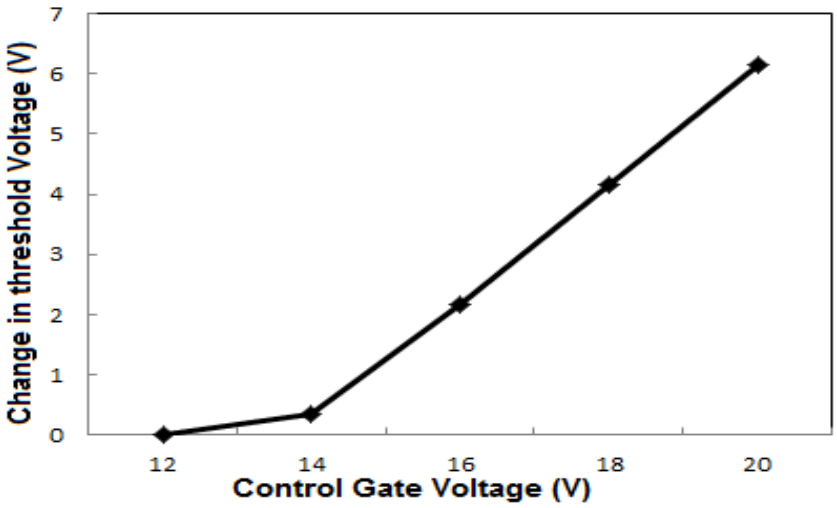

Figure 5.5: Change in threshold voltage of single gate nanocrystal flash memory cell with change in control gate voltage.

In double gate memory, both the gates are short and write pulse of 18 V for 0.1 second is applied for programming; Electrons from the channel starts tunneling to both sides of nanocrystals. It is found that charge in the double gate increases as compared to the single gate nanocrystal flash memory which is reflected in transfer characteristics of the device which shows increase in the memory window of the device which is more than single gate flash memory. Due to two gates, more coupling associated between the channel and control gates will reduce the short channel effects. As the number of nanocrystals increases, the charge storing nodes increase and hence capacity of the flash memory.

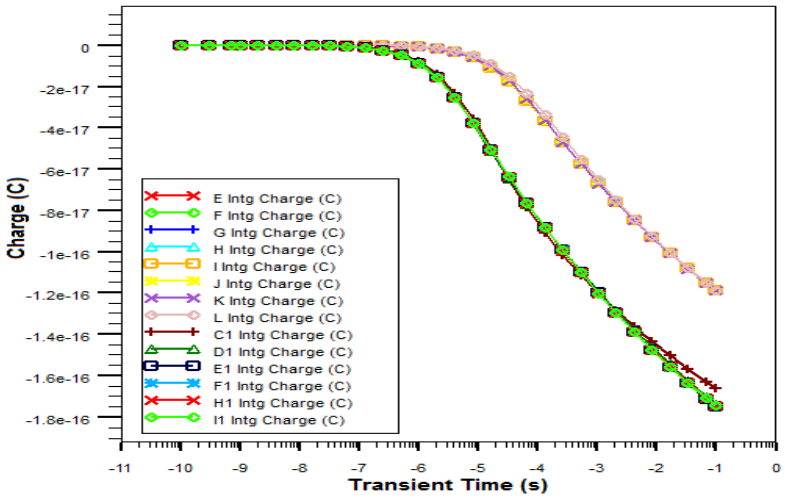

Figure 5.6: Nanocrystals charge during write operation in double gate nanocrystal flash memory cell

As shown in the figure 5.3 and figure 5.7, change in threshold voltage in double gate nanocrystal cell is more than single gate nanocrystal memory cell. Change in threshold voltage in case of single gate is 4.158 V and in double gate change in threshold voltage is 5.807

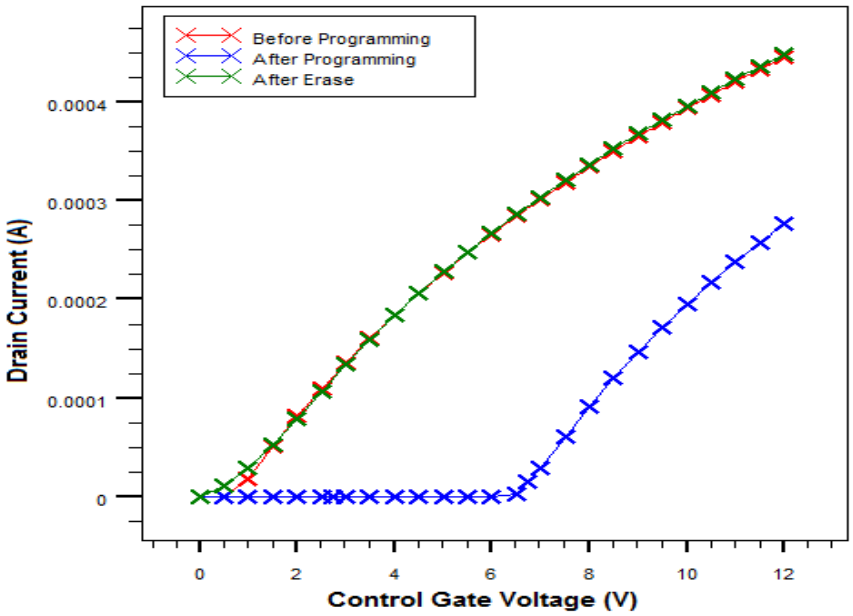

Figure 5.7: Threshold voltage shift in double gate nanocrystal flash memory cell

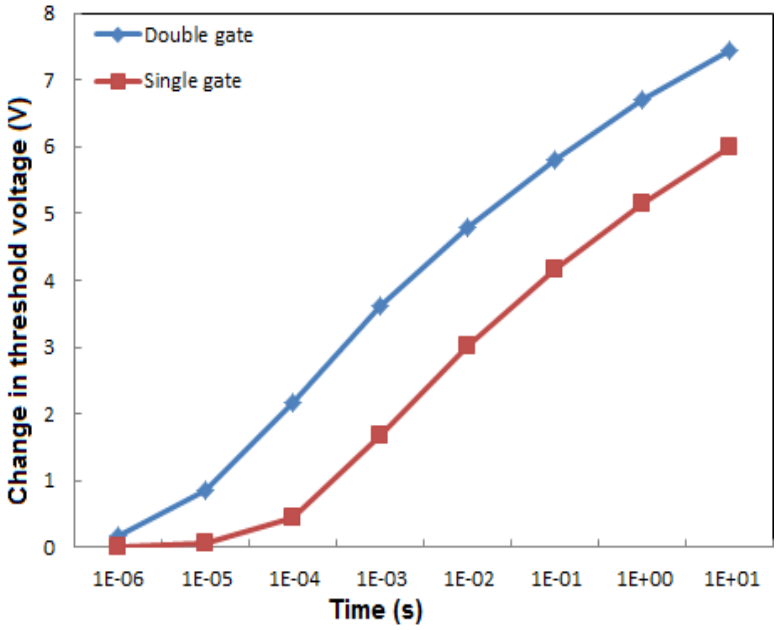

Figure 5.8: Change in threshold voltage of memory cell with time duration of write pulse.

Figure 5.8 demsotrated the effect of duration of write pulse on the nanocrystal flash memory. It is observed that change in threshold voltage in case of double gate flash memory is more as compared to single gate flash memory time which shows more charge is stored in case of double gate as compared to single gate for same duration of write pulse,more electrons tunnel in case of double gate memory, i.e. fixed value of charge in case of double gate is store in less time as compared to single gate nanocrystal flash memory.

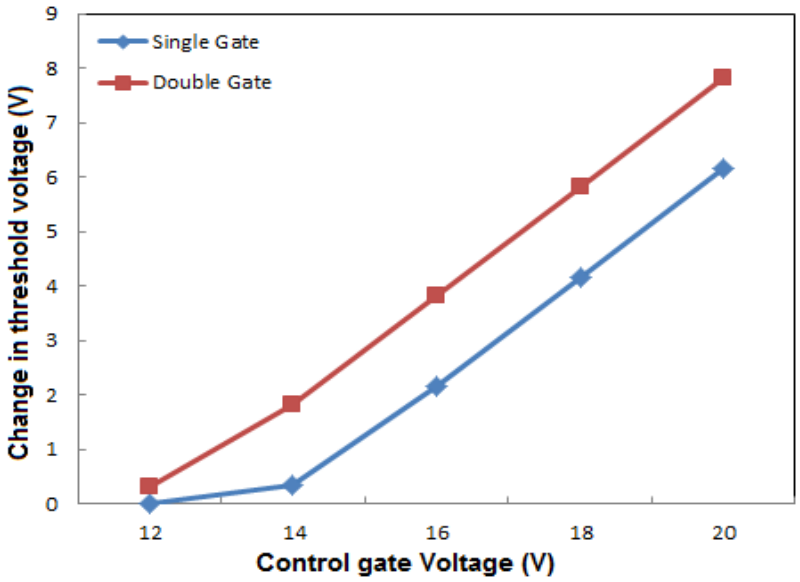

Figure 5.9: Change in threshold voltage of memory cell with control gate voltage.

The change in threshold voltage increases linearly with applied voltage, the result demonstrates that the transfer of carriers between channel and nanocrystal is very difficult at low electric field and transfer become significant as electric field increases and transfer of charges in case of double gate memory cell is more as compared to single gate .Hence more charge is store in case of double gate flash memory with same control gate voltage.

High–*k* dielectric as Control gate oxide

Gate coupling ratio is very important parameter in the design of flash memory [2-3]. It measures how strongly the control gate coupled to the nanocrystals which act as floating gates. Field across the tunnel oxide decides the speed of the device. Gate coupling ratio depends on the capacitance between the control gate and the nanocrystals, more the capacitance between the control gate and the floating gate more will be the gate coupling ratio and this capacitance can be increased by decreasing the control gate oxide thickness, replacing the dielectric between control gates and floating gate by high-k dielectric. Reducing

the control gate oxide thickness will increase the risk of tunneling of electrons from nanocrystals to the control gate which is undesirable. So in order to improve gate coupling ratio high k dielectric is used [12-15]. Due to increased dielectric constant capacitance increases hence gate coupling ratio increases. In this work Sapphire (Al_2O_3) used as the high-*k* dielectric. Sapphire is sandwiched between two SiO_2 layer and this stacked combination is used as control.

As shown in figure 5.10 with gate voltage of +18 V for 0.1 seconds, the charge induces on the nanocrystal is increased as compared to conventional SiO_2. Using High-K dielectric, as control gate oxide increases the gate coupling and hence electric field increases, thus more electrons will tunnel from channel to nanocrystals for a fixed gate voltage. Hence, using high-*k* dielectric scaling of the voltage is possible

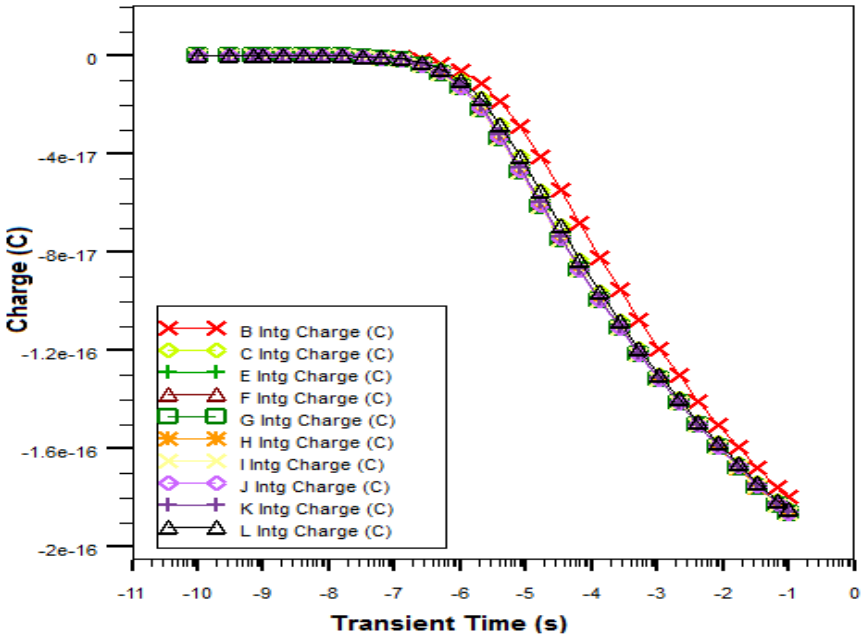

Figure 5.10: Charge during write operation in single gate with high–*k* control oxide

CHAPTER 6

Conclusion and Future Scope

The present study of flash memory cell indicates that the amount of charge storage depends on cell structure and value and duration of read-write pulse applied . There is a trade-off between the programming efficiency and the data retention in continuous floating gate memory. Nanocrystal flash memory is the potential device to overcome the problems associated with conventional flash memory.

Combining the concept of double gate MOSFET and nanocrystal flash memory enhanced the charge storing capacity of the double gate nanocrystal flash memory. Even the short channel effects introduced in advanced technology nodes due to charge sharing are reduced in double gate nanocrystal flash memory. Programming efficiency of the memory cell also increases in double gate nanocrystal flash.

Using high-k dielectric as control oxide overcomes the main obstacles in scaling in nanocrystal flash memory by increasing the gate coupling ratio of the device for reduced gate leakage.

Though the double gate nanocrystal flash memory cell structure have so many advantages over the available cell structures, its performance can further increased by using other materials having higher work function in place of polysilicon nanocrystals. Also the advanced nano MOSFET's structures can be employed to further improve the performance of nanocrystal flash memory cell.

REFERENCES

[1] R.Bez, E.Camerlenghi, A.Modelli, and A.Visconti, "Introduction to flash memory", Proceedings *of the IEEE*, Vol. 91, No. 4, April 2003.

[2] IEEE Standard Definitions and Characterization of Floating Gate Semiconductor Arrays Standards Committee of the IEEE Electron Devices Society, 1998

[3] S. Tiwari, F. Rana, H. Hanafi, A. Hartstein, E. F. Crabbe, and K. Chan,"Floating gate memory cell", *Appl. Phys. Lett.* 68, 1377, 1996.

[4] D. Kahng and S.M. Sze, "A floating gate and its application to memory devices*",* in *IEEE Trans. On Electron Devices,* Vol. 9, No. 9, pp. 629, 1967.

[5] D. Frohman-Bentchkowsky, "Memory behaviour in a floating-gate avalanche-injection MOS (FAMOS) structure", *Appl. Phys. Lett.*, Vol. 18, pp. 332–334, 1971.

[6] H.Iizuka, F. Masuoka, T.Sato, and Ishikawa, "Electrically alterable avalanche-injection type MOS read-only memory with stacked-gate structures", *IEEE Trans. on Electron Devices*. Vol.23, pp. 379. 1976.

[7] D. C. Guterman, I. H. Rimawi, T. L. Chiu, R. D. Halvorson, and D. J. McElroy, "An electrically alterable non-volatile memory cell using a floating gate structure", *IEEE Trans. on Electron Devices*, Vol. 26, p. 576, 1979.

[8] F. Masuoka, M. Asano, H. Iwahashi, T. Komuro, and S. Tanaka, " A new Flash E_2PROM cell using triple polysilicon technology", *IEDM Tech. Dig.,* p. 464, 1984.

[9] P. Pavan, R. Bez, P. Olivo, and E. Zanoni,"Flash memory cell-An overview", *Proceedings of IEEE*, Vol. 85, pp. 1248–1271, 1997

[10] A. Thean and J. P. Leburton, "3-D computer simulation of single-electron charging in silicon nanocrystal floating gate flash memory devices," *IEEE Electron Device Lett.*, Vol. 22, pp. 14, 2001

[11] M.Lenzlinger and E.H Snow, "Fowler-Nordheim Tunneling into Thermally Grown SiO_2," in *J. Appl Phys.*, Vol. 40, pp. 278, 1968

[12] S. Lai, "Flash memories: Where we were and where we are going", *IEDM Tech. Dig.*, pp. 971–973, 1998.

[13] S. Aritome, "Advance Flash memory technology and trends for file storage application," *IEDM Tech Dig.*, pp. 763–766, 2000

[14] D. Ielmini, A. Spinelli, A. Lacaita, and A. Modelli, "Statistical model of reliability and scaling projections for Flash memories", *IEDM Tech. Dig.*, pp. 32.2.1–32.2.4, 2001.

[15] S. Mukherjee, T. Chang, R. Pang, M. Knecht, and D. Hu, "A single transistor EEPROM cell and its implementation in a 512 K CMOS EEPROM", *IEDM Tech. Dig.*, pp. 616–619, 1985.

[16] C. Calligaro, A. Manstretta, A. Modelli, and G. Torelli, "Technological and design constraints for multilevel Flash memories", Proc. *3rd IEEE Int. Conf. Electronics, Circuits and Systems*, pp.1003–1008, 1996.

[17] G. Chakraborty, A. Sengupta, F. G. Requejo, and C. K. Sarkar, "Study of the relative performance of silicon and germanium nanoparticles embedded gate oxide in metal-oxide-semiconductor memory devices," *J. Appl. Phys*., Vol. 109, pp. 064504 (1-6) , March 2011.

[18] M. Yang, T. P. Chen, W. Zhu, J. I. Wong, and S. Zhang, "Comparison of Charge Storage Behaviour of Electrons and Holes in a continuous Ge Nanocrystal Layer," Nanosci. Nanotechnol. Lett. Vol. 2, No. 1, pp. 7-10, 2010.

[19] M. Sadd, R. Muralidhar, S. Madhukar, K. Scheer, D. Gentile, B. Hradsky, M. Rossow, R. Rao, M.Ramon, A. Konkar, J. Conner, S. Bagchi, B. E. White, A. Thean, and J.P. Leburton, "Physical Modeling of Reliability of a Nano-crystalline Memory Cell," *Proceedings of IEEE Non-Volatile Semiconductor Memory Workshop*, pp.82-83, Aug. 2001.

[20] S. Lai, "Tunnel Oxide and ETOX Flash Scaling Limitation," IEEE International Non – Volatile Memory Conference, pp. 6, 1998.

[21] Y. Yang, A. Purwar, and M. H.White, "Reliability considerations in scaled SONOS non-volatile memory devices," in *Solid-State Electronics*, Vol. 43, pp. 2025- 2032, May 1999.

[22] P. Olivo, T. Nguyen and B. Ricco , "High-field-induced degradation in ultra-thin SiO2 films," *IEEE Trans. on Electron Devices*, Vol. 35, No. 12, pp. 2259–2267, 1988.

[23] L. Larcher, "Statistical simulation of leakage current in MOS and flash memory devices with a new multiphonon trap-assisted tunneling model", *IEEE Trans. on Electron Devices*, Vol. 50, No. 5, pp. 1246-1253, 2003.

[24] C. M. Compagnoni, D. Ielmini, A. S. Spinelli, A. L. Lacaita, C. Previtali, and C.Gerardi, "Study of data retention for nanocrystal flash memories", *IEEE 41st Annual International Reliability Physics Symposium Proceedings*, pp. 506-512, 2003.

[25] Z. Liu, "Metal nanocrystal memories-part II: electrical characteristics," in *IEEE Trans. on Electron Devices*, Vol.49, No. 9, pp. 1614 – 1622, 2002

[26] Ni Henan, Wu Liangcai , Song Zhitang and Hui Chun, “Memory characteristics of an MOS capacitor structure with double-layer”, *Journal of Semiconductors*, Vol. 3, No.11, 2009.

[27] Ting-Chang Chang, Fu-Yen Jian, Shih-Cheng Chen, and Yu-Ting Tsai, “Developments in nanocrystal flash memory”, *Journal of Materials today,* Vol. 14, No.12, 2011

[28] J.D Blauwe ,“Nanocrystal Non-volatile Memory Devices”, *IEEE Trans. on Nanotechnology,* Vol. 1, No. 1, March 2002.

[29] J. Robertson, “High dielectric constant oxides”, *Eur. Phys. J. Appl. Phys*., Vol. 28, pp. 265–291, 2004

[30] C.C Monzio, G. Riccardo, I.Daniele, S. Alessandro, L. Andrea , “Silicon Nanocrystal Memories: A Status Update”, J. *Nanosci. Nanotechnol*., Vol. 7, No.1, pp. 193,2007

[31] K.Roy, S. Mukhopadhyay and H.Mahmoodi-Meimand , “Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer CMOS Circuits,” in *Proceedings of the IEEE*, Vol. 91, No. 2, 2003.

[32] A.Chaudhry and M.J Kumar, "Controlling Short-channel Effects in Deep Submicron SOI MOSFETs for Improved Reliability: A Review”, in *IEEE Trans. on Device and Materials Reliability*, Vol.4, pp.99-109, 2004.