SIMULATION AND MODELING OF SONOS NON-VOLATILE MEMORY

by

ASHA RANI

A Thesis
Submitted to the
Graduate Faculty
of
George Mason University
In Partial fulfillment of
The Requirements for the Degree
of
Master of Science
Electrical Engineering

Committee:

_____	Dr. Qiliang Li, Dissertation Director
_____	Dr. M.Rao Mulpuri, Committee Member
_____	Dr. Alok Berry, Committee Member
_____	Dr. Andre Manitius, Department Chair
_____	Dr. Lloyd J. Griffiths, Dean, The Volgenau School of Information Technology and Engineering
Date:_____	Fall Semester 2010 George Mason University, Fairfax, VA

Simulation and Modeling of SONOS Non-Volatile Memory

This thesis is submitted in partial fulfillment of the requirements for the degree of Master of Science at George Mason University

By

Asha Rani
Master of Science
George Mason University, 2010

Director: Dr. Qiliang Li, Assistant Professor,

Department of Electrical and Computer Engineering

Fall Semester

George Mason University

Fairfax, VA

Acknowledgments

There are many people who have helped me through various stages of my thesis work. First of all I am heartily thankful to my advisor, Dr. Qiliang Li, whose encouragement, guidance and support from initial to the final level enabled me to develop an understanding of the subject.

I am thankful to XiaoXiao Zhu and Yang Yang for their guidance in learning the Synopsys TCAD tools. The discussions and cooperation with Aveek Gangopadhyay and Anindya Nath have contributed substantially to this work.

I would like to thank Dr. M. Rao Mulpuri for his support and suggestions. I would also like to thank Dr. Alok Berry who agreed to serve on my examining committee on very short notice.

Finally I take this opportunity to express my profound gratitude to my beloved parents for their moral support during my study in George Mason University.

Table of Contents

List of Figures

Abstract

SIMULATION AND MODELING OF SONOS NON-VOLATILE MEMORY

Asha Rani M.S.E.E.
George Mason University, 2010
Thesis Director: Dr. Qiliang Li

Silicon-Oxide-Nitride-Oxide-Silicon (SONOS) memory is one of the best non-volatile, FLASH-like memories for the next-generation electronic devices which require high-level portability, stand-alone capability and low-power consumption for extremely long battery life. As the dimensional CMOS scaling for better performance, the application of high-k dielectric (e.g., oxide-nitride layer) to replace current floating polysilicon gate is the most attractive strategy to meet the challenges of conventional floating-gate FLASH memory: reliability issue and lateral interaction.

In this work, SONOS memory cells with varying dielectric thickness have been systematically and analytically studied by using a SYNOPSIS Technology CAD simulation and modeling tool. The mechanisms of carrier tunneling and memory retention of the memory cells have been tested by simulation and studied in detail. In addition, Hafnium oxide (HfO_2) as charge trapping layer has also been analyzed by using TCAD simulation.

From the simulation results we have the following conclusions: (i) as the thickness of charge-storage dielectrics (e.g., Si_3N_4 and HfO_2) increases, the threshold voltage shift increases due to the higher capture probability of electrons; (ii) as the thickness of tunneling oxide decreases, the programming speed increases due to the higher electric field across the tunneling oxide; (iii) as the channel length decreases from 210 nm to 70 nm, threshold voltage shift increases with increasing program time; (iv) as the temperature increases, threshold voltage shift decreases.

In summary, the SONOS-like nonvolatile memory is a strong candidate for future high-density, high-performance, portable electronics.

Chapter 1

Introduction and Background of FLASH Memory Technology

1.1 Introduction of recent development of Flash memory

Definition of Flash Memory:

Flash memory is a non-volatile memory that can be programmed and erased with electric pulses. A block, sector or page consisting of a large number of Flash memory cells can be electrically programmed or erased at the same time. The word “flash “ itself is related to the fact that since whole memory can be erased at once, erase times can be very fast. Flash technology combines the high density of the UV EPROM with electrical in-system erasability of EEPROMs.There has been continuous increase in the market of Flash Memories.

Since the introduction of first device in early 1980s, the market demand of FLASH memory is continuously expanding in a fast pace. Recently, one of the main developments and application of FLASH memory is flexible, low-cost and reliable solid-state memory, such as removable memory, non-volatile memory for portable electronics and solid-state hard drive. The three major market demands of Flash memories are related to Personal Computers (PC), wireless and telecommunication applications, and automotive electronics. For the next-generation electronics, the research of non-volatile FLASH-like memory focus on fast programming/erasing and low-power applications for excellent portable electronics.

Schematic of the evolution of the whole semiconductor and Non-Volatile Memory (NVM) markets is shown in Fig 1.1 (a) and (b) respectively. As shown in Fig

1.2, devices for application in computers have dominated the market for years. Among different computer devices, market share for NVM has recently been expanding in faster pace, mainly to the high demand for high-capacity non-volatile memory devices for portable applications.

As also shown in Fig 1.1[1-2], the fastest growing Flash memory focuses on, NAND architecture, which grows exponentially in market sales, while the demand for-NOR FLASH memory have almost been consistent since 2001.As can be seen in figure 1.2, it is estimated that Flash NVM market will grow with a higher average annual rate than DRAM and SRAM, reaching $50 billion in 2011.

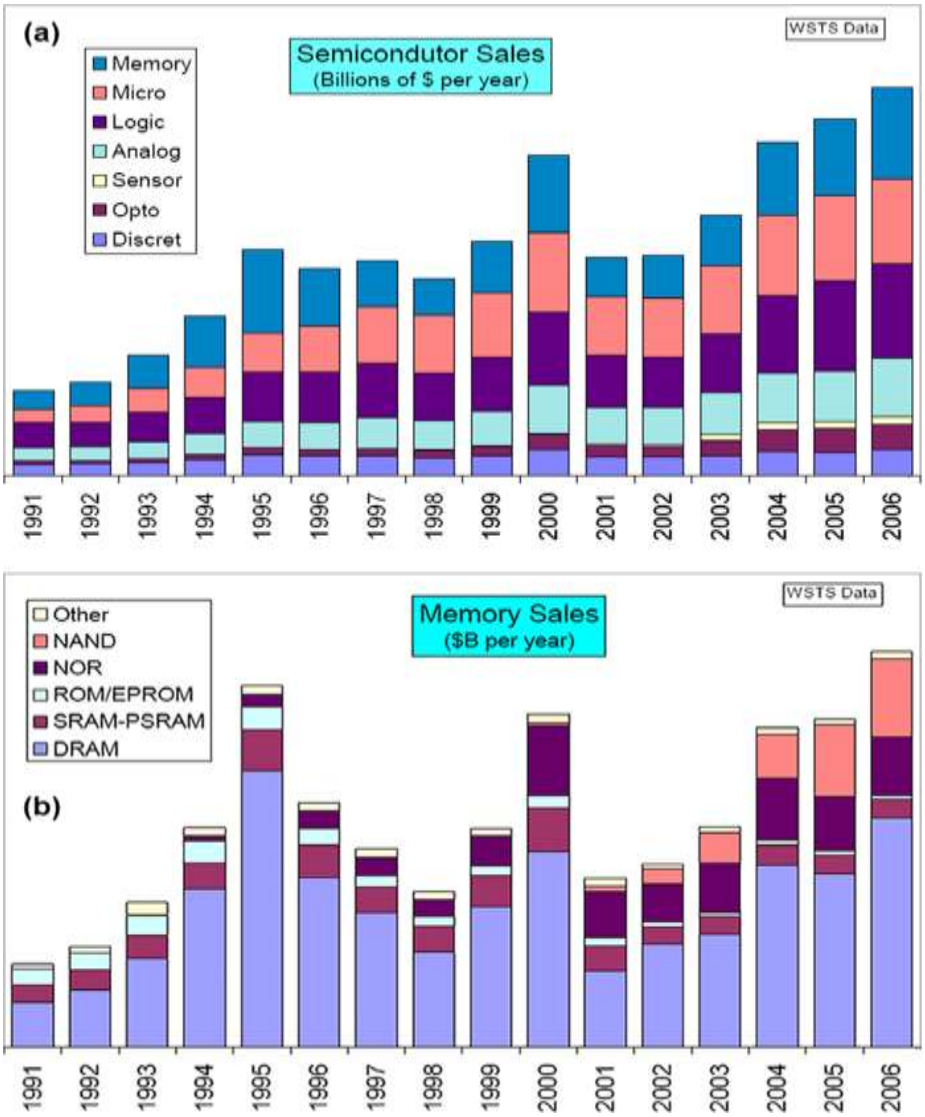


Figure1.1. ***Recent evolution of (a) semiconductor and (b) memory markets***

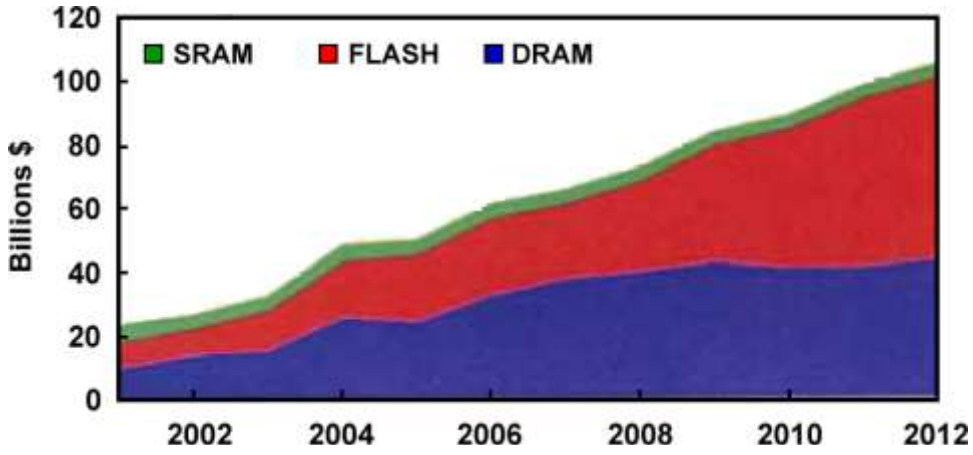


Figure 1.2 .CMOS memory market evolutions.

1.2 Applications:

Flash memories have two major application segments[3-4]: code and data applications:

Code applications:

In this, the program or operating system is stored and is processed by microprocessor or microcontroller. Possibility of non volatile memory integration in logic systems to allow software updates, store identification codes, or reconfigure the system on the field.

In this sense, Flash devices are widely used in several fields. In the computer environment they allow to store and update the operation system in PC BIOS and Hard – Disk Drives (HDDs), in almost all peripherals like printers and DVD –readers, and in most add-on boards like video and sound cards. On computer network equipments, they allow to quickly upgrade the software in modems, interface cards and network routers. In the automotive electronics field they are used in vital function such as Engine Control Units (ECUs) and Global Positioning Systems (GPS).Finally, popularity of cellular phones has resulted in increased demand for reliable and low-power memory devices.

Data applications:

Data (or mass) storage where data files for image, music, and voice are recorded and read sequentially. This is to create storing elements like memory boards or solid-state hard disks, made by Flash memory arrays, which are configured to create large size memories. Presently USB storage devices have reached 64GB.

1.3 Classification of Semiconductor devices:

There are two parameters that describe how "good" and reliable a nonvolatile memory cell is [4]:

a. Endurance: Capability of maintaining the stored information after erase/program/read cycling.
b. Retention: Capability of keeping the stored information in time.

To have a memory cell which can commute from one state to other, and which can store the information independently of external conditions, the storing element needs to be a device whose conductivity can be changed in a non-destructive way. Figure 1.3 shows the classification of semiconductor memory devices.

First classification of Semiconductor memory is on the basis of data loss due to disconnected power supply.

- Non-Volatile Memory (NVM),like EPROM,EEPROM or Flash, that are able to balance the less-aggressive (with respect to SRAM and DRAM) programming and reading performances with nonvolatility, i.e., with the capability to keep the data content even without power supply
- Volatile Memory, like SRAM or DRAM, that although very fast in writing and reading (SRAM) or very dense(DRAM),lose the data content when the power supply is turned off.

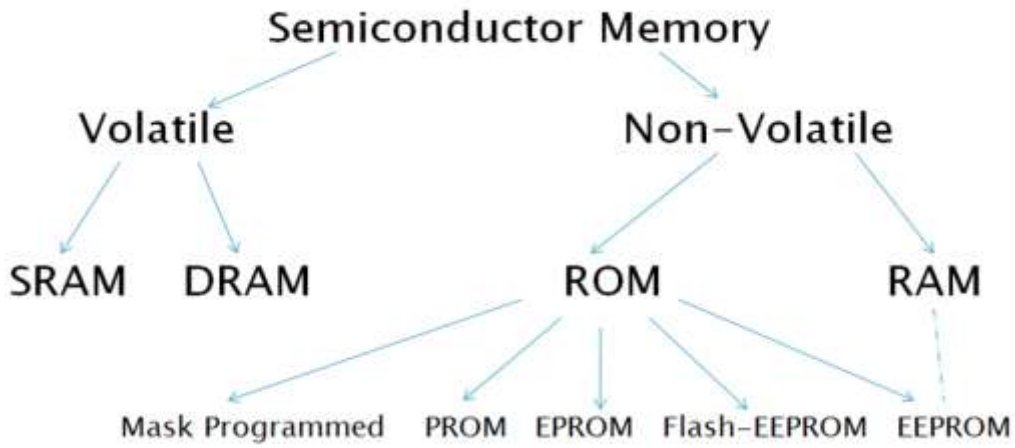


Figure 1.3 Classification of Semiconductor Memory

Before getting into each type of nonvolatile memories, it is necessary to know the difference between RAM and ROM.RAM is random-access memory having x-y address for each cell which distinguishes it from other serial memories such as magnetic memory.ROM (read-only memory) also has random access capability since the addressing architecture is similar. In fact the read process of RAM and ROM are identical.RAM is sometimes also called read-write memory. To some extent ROM also has started to develop rewriting capability. So the main difference between ROM and RAM lies is the ease and frequency of erasing and programming.RAM has equal opportunity of reading and writing. A ROM has much more frequent read than rewrite.ROM itself has the spectrum of rewriting capability, ranging from a pure ROM without any writing capability to full-featured EEPROM. Because ROM is smaller in size and more cost-effective than RAM, it is used whenever frequent rewriting is not required.

With this background different types of NVM are described below.

a) Mask-programmed ROM:

The memory content is fixed by the manufacture and is not programmable once it is

fabricated.

b) PROM:

It is also called field-programmable ROM or fusible-link ROM. It is a form of digital memory where the setting of each bit is locked by fuse or antifuse. Such PROMs are used to store programs permanently. Main difference from ROM is that programming is applied after the device is constructed.

c) Erasable Programmable Read Only Memory (EPROM):

EPROM can be electrically programmed without the need of electricity to erase the stored charges. EPROM must be subjected to expose under ultra-violet (UV) radiation in order to be erased.

d) Electrically Erasable Programmable Read Only Memory (EEPROM):

EEPROM are considered to be electrically erasable and programmable in system, byte by byte. Not only can it be erased electrically, but also selectively by byte address. However, occupy larger areas than that of EPROM, because it requires a select transistor for each cell and a floating gate transistor, leading to two-transistor cell. Therefore, EEPROM have disadvantage of higher cost and lower densities.

e) Flash EEPROM:

Flash memories emerged as combined features of EPROM and EEPROMs. They show similar advantages of single transistor as EPROM and tunnel oxide as EEPROM. Flash memories can electrically program and erase large scale of cells, which is referred to as a block erasing. It loses byte selectivity but maintains one-transistor cell.

f) Non volatile RAM:

This memory can be viewed as a nonvolatile SRAM or EEPROM with short Programming time as well as high endurance.

When the gate electrode of conventional MOSFET is modified such that semi-permanent charge storage inside gate is possible, new structure formed is nonvolatile memory device. Since the nonvolatile memory device was proposed by Kahng and Sze in 1967, various device structures has been made and nonvolatile memory devices has been used

commercially. In the development of non-volatile memory devices, two different structure types appear at the same time.

i) Charge trapping device,
ii) Floating gate device.

Although their structures are different the theorems of the device operating conditions are similar. These two different types of device can be differentiated by the material of the stored charge. The data storing capability of NVMs is due to charge trapped in floating gate or in charge trap dielectrics .Figure 1.4 shows the basic operating principle of NVMs. If one can store charges in the insulator of a MOSFET, the threshold voltage can be modified to switch between two distinct values, conventionally between "0" and "1". "0" is referred to as erase state and "1" is referred to as program state. Depending on the existence of trapped charges, the NVM device represent itself as either logical "1" or "0" as shown in figure1.5.

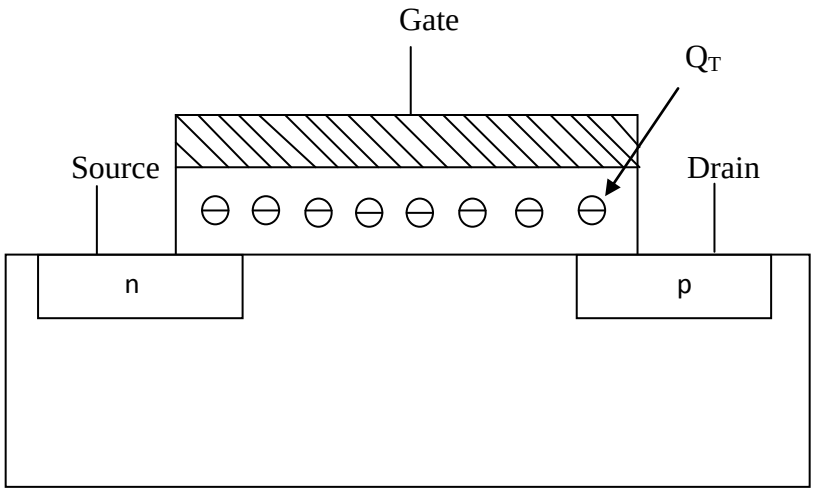


Figure 1.4 Basic operating principle of NVM: the storage of charges in gate insulator of a MOSFET

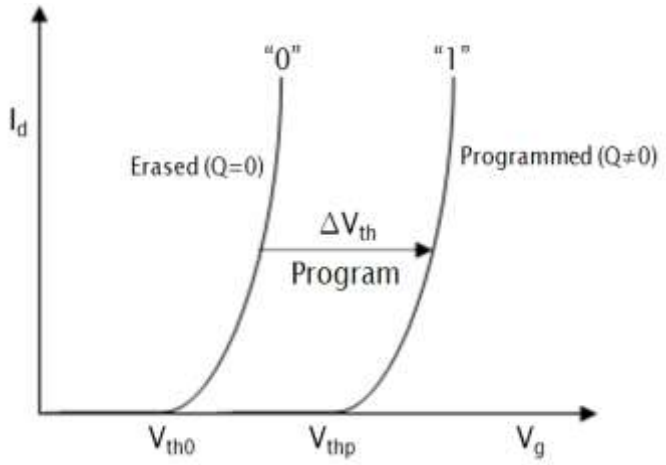


Figure 1.5(a). **Threshold Voltage shifted once the charge is trapped in floating gate or charge trap material.**

From the basic theory of the MOS transistor, the threshold voltage is given by

$$V_{th} = 2\Phi_f + \Phi_{ms} - Q_I/C_I - Q_D/C_I - (Q_T/\epsilon_I)\, d_I \qquad (1.1)$$

Where,

Φ_{ms} = the work function difference between the gate and the bulk material

Φ_f = the Fermipotential of the semiconductor at the surface

Q_I = the fixed charge at the silicon/insulator interface

Q_D = the charge in the silicon depletion layer

Q_T = the charge stored in the gate insulator at a distance d_I from the gate

C_I = the capacitance of the insulator layer

ϵ_I = the dielectric constant of the insulator

In programmed state, V_{th} is given as equation 1.1. In erased state V_{th} is given by

$$V_{th} = 2\Phi_f + \Phi_{ms} - Q_I/C_I - Q_D/C_I \qquad (1.2)$$

Thus the threshold voltage shift, caused by the storage of the charge Q_T is given by

$$\Delta V_{TH} = -(Q_T/\epsilon_I)\, d_I \qquad (1.3)$$

The storage of charges in the gate insulator of MOSFET is realized in two ways resulting in subdivision of Nonvolatile semiconductor memory devices into two main types:

1) Floating gate device: In these devices the charge is stored on a conducting or semiconducting layer surrounded completely by a dielectric, usually thermal oxide. As this layer acts as completely isolated gate, this type of device is referred to as floating gate device.

2) Charge – trapping devices: In these devices, the stored charge is stored in discrete trap centers of an appropriate dielectric layer. Most successful in this category is MNOS device (Metal-Nitride-Oxide-Semiconductor) structure, in which the insulator consists of silicon nitride layer on top of very thin silicon oxide layer.

The device state can be read by applying an appropriate "sensing" voltage to the control gate, as shown in Figure 1.5(b).When the FG device I-V curve corresponds to curve a (Q=0),the $V_{cs} > V_{th0}$ and the device is ON; when the device has been previously programmed (curve b),$V_{CS} < V_{thp}$ and the device is OFF.

Full description of floating gate device and charge-trapping device are mentioned in section 1.4 and 1.5.

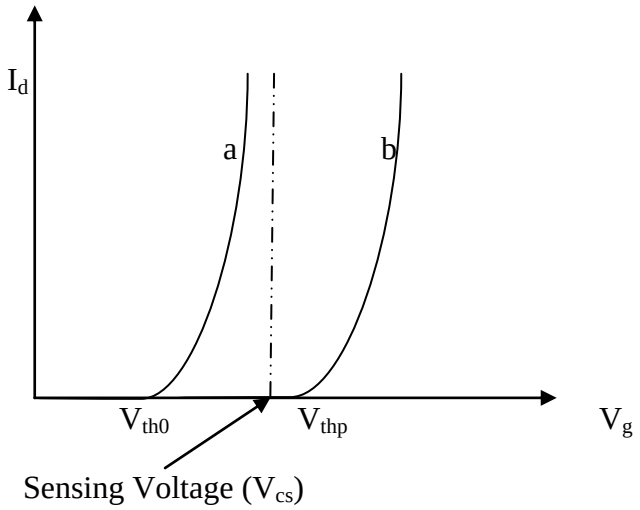


Figure 1.5 b. Reading operation of a FG device: a suitable control voltage ($V_{th0} < V_{cs} < V_{thp}$) is applied to the device to determine whether it is conductive or not.

1.4. Concept of Floating gate devices:

A flash cell is basically a floating-gate MOS transistor (figure 1.6) i.e. a transistor with a floating-gate (FG) completely surrounded by dielectrics, and electrically governed by a capacitively coupled control gate (CG). Due to lack of electrical connection it is referred as "floating gate". The quality of the dielectrics guarantees the non-volatility, while the thickness allows the possibility to program or erase the cell by electrical pulses. Usually, the gate dielectric, i.e. the one between the transistor channel and the FG, is an oxide in the range of 9-10 nm and is called 'tunnel oxide' since FN electron tunneling occurs through it.

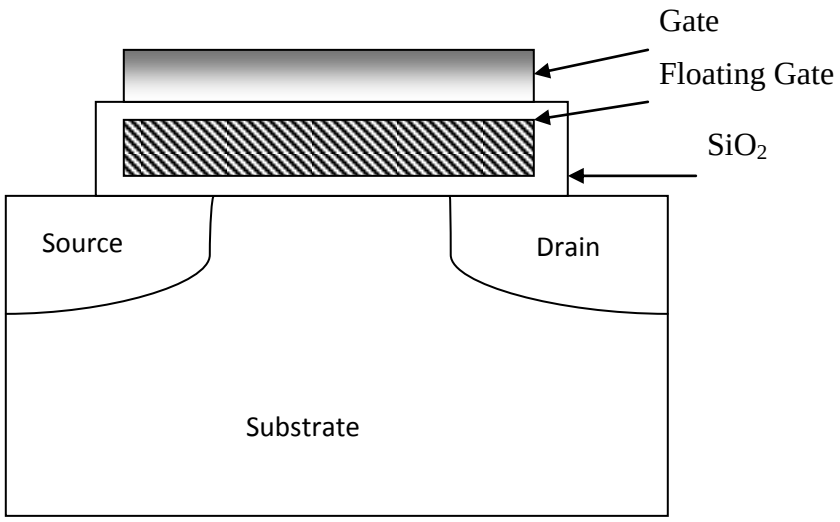


Figure 1.6 Cross Section of Floating Gate Device, it contains 2 stacked gates; the bottom gate surrounded by dielectric material and is used to store charges (Floating Gate).Top gate is used to supply operating bias called "Control Gate".

Since the charges stored in the floating gate are not affected by the electric field and temperature, floating gate device became the most popular and important kind of NVMs. Charge is injected to the floating gate to change the threshold voltage.
The two modes of programming are:

a. Hot-carrier injection
b. Fowler-Nordheim tunneling.

Figure 1.7a shows the mechanism of hot-carrier injection .Near the drain, the lateral field is at its highest level. The channel carriers (electrons) acquire energy from the field and become hot carriers. When their energy is higher than the barrier of the Si/SiO_2 interface, they can be injected to the floating gate. At the same time, the high field also induces impact ionization. These generated secondary hot electrons can also be injected to the floating gate. Figure 1.7b shows the original method of hot-carrier injection using drain-substrate avalanche. In this scheme, the floating-gate potential is more negative such that hot holes are injected instead. This injection scheme is found to be less efficient and is no longer used in practice.

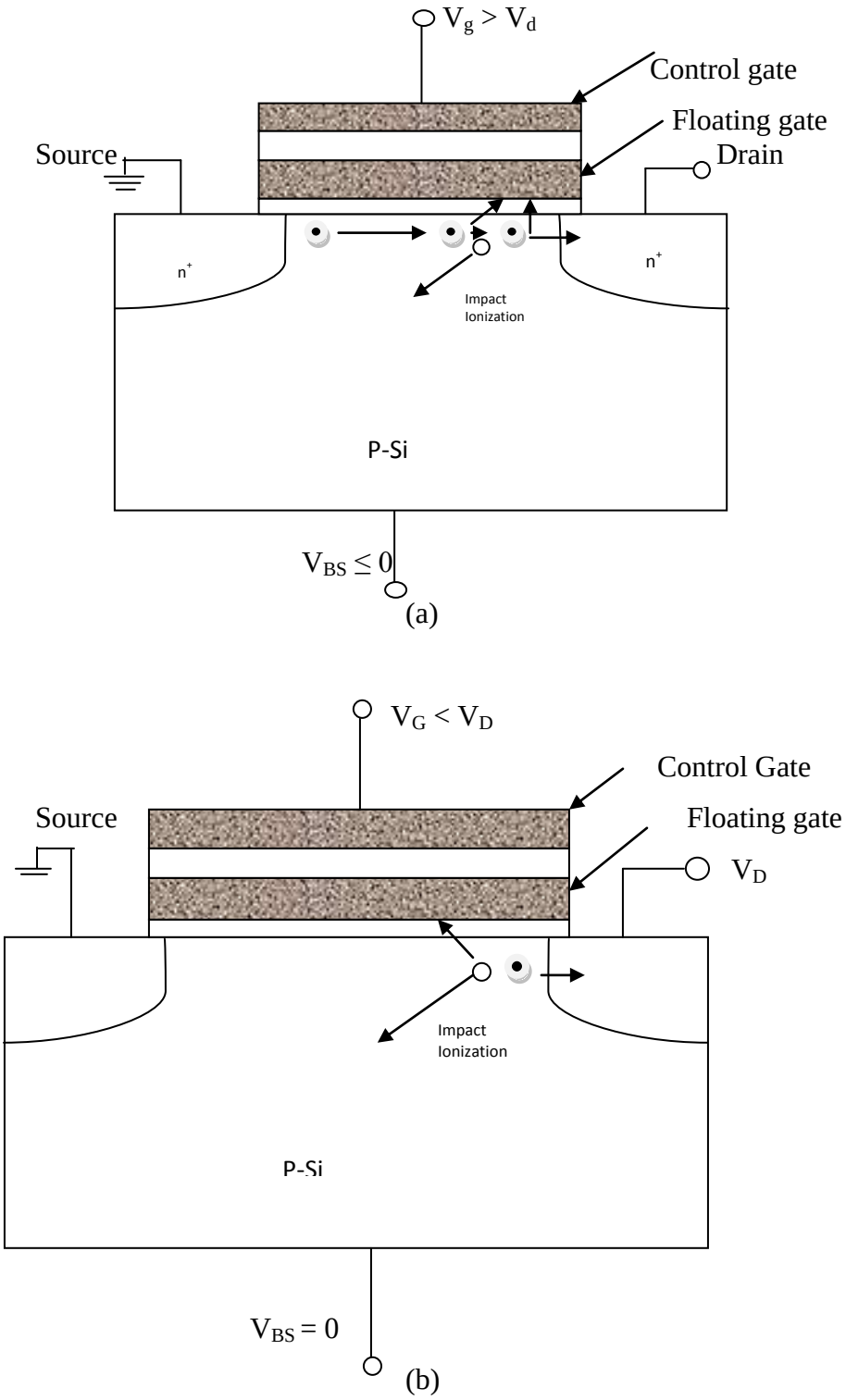


Figure 1.7 Charging of the floating gate by hot carriers. (a) Hot electrons from channel and impact ionization.(b) Hot holes from drain avalanche. Note difference in gate bias between the two figures.

1.5 Basic Concept of charge trapping devices:

The metal-nitride-oxide-silicon (MNOS) devices were invented in 1967[5] and were first electrically alterable semiconductor (EAROM). This charge trapping device became another potential candidate of NVM due to its storage capability of charges in discrete traps in the nitride layer. The schematic cross-section structure of a poly-Si gate MNOS memory is shown in figure 1.8.Electrons or holes are injected from the channel region into the nitride by quantum mechanical tunneling through ultra-thin oxide (UTO, typically 1.5 to 3 nm). Charges are stored in deep level traps in Si_3N_4 .Nitride layer is to increase the density and probability of capturing electrons and holes. These trapped charges causes a significant shift in the threshold voltage of the transistor [equation 1.2, Q_T the trapped charge in the nitride layer].In programming process, a large positive bias is applied to the gate. Current conduction is due to electrons that are emitted from the substrate to the gate by the tunneling. During erase, negative bias is applied to the gate and current conduction is due to tunneling of holes from substrate to neutralize the trapped electrons. The advantages of the MNOS transistor include reasonable speed for programming and erasing, making it a suitable candidate for NVM. Drawbacks of MNOS are:

a. MNOS transistor has large programming and erasing voltages.
b. Data retention and erase/write endurance of MNOS device must be considered due to the electron escape through the top metal gate.

These drawbacks result in narrow threshold voltage window after many cycles of programming and erasing. Because of their low endurance and retention, MNOS devices are used only in specific applications (such as military due to their radiation hardness).Therefore, MONOS (metal-oxide-nitride-oxide-silicon) and SONOS (Silicon-oxide-nitride-oxide-silicon) have been developed to achieve high-reliability and high-

yield EEPROM products. Modern counterpart of MNOS, the SONOS (Silicon-Oxide-Nitride-Oxide-Silicon) transistor is similar to MNOS transistor except that it has an additional blocking oxide layer placed between the gate and nitride layer, forming and ONO (Oxide-nitride-oxide) stack. This top oxide layer is usually similar in thickness to the bottom oxide layer. The function of the blocking oxide layer is to prevent electron injection from the metal to the nitride layer during erase operation. As a result, a thinner nitride layer can be used, leading to lower programming voltage as well as better charge retention.

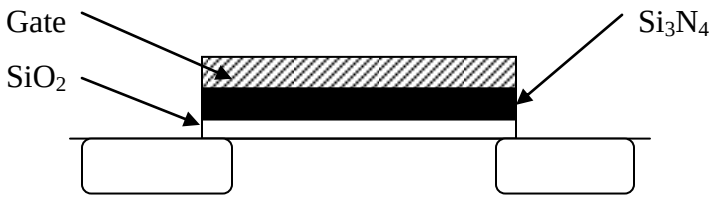


Figure 1.8 Charge Trapping Devices (MNOS device)

Chapter2.

Design of SONOS Non-Volatile Memory Devices

2.1 Evolution of SONOS

The very first idea of using a floating gate device to obtain a nonvolatile memory device was suggested by D.Kahng and S.M.Sze in 1967[5].This was also the first time that possibility of nonvolatile MOS memory device was recognized. The memory transistor proposed by them was similar to MOS structure, except the gate structure was replaced by a layered structure of a thin oxide I_1,a floating gate but conducting metal layer M_1,a thick oxide I_2, and an external metal gate M_2 as shown in figure 2.1.This device was referred to as MIMIS(metal-insulator-metal-insulator-semiconductor) cell.

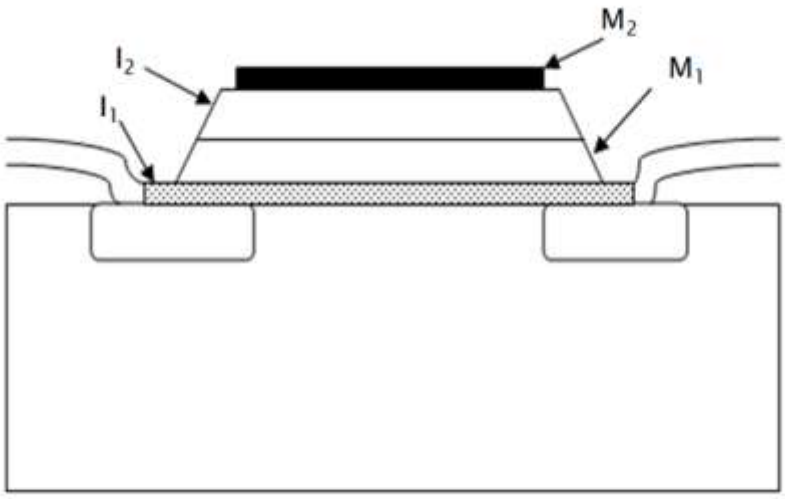


Figure 2.1 Introduction to floating gate principle: the MIMIS structure, introduced by Kahng and Sze.

The first dielectric I_1 has to be sufficiently thin in order to obtain a high electric field to allow tunneling of electrons toward the floating gate. These electrons are then captured to the conduction band of the floating gate M_1, if the dielectric I_2 is thick enough to prevent discharging. When the gate voltage is removed, the field in I_1 is too small to allow

backtunneling. The injection mechanism to bring electrons to the floating gate is direct tunneling. To discharge the floating gate, a negative voltage pulse is applied at M_2, removing the electrons from the floating gate by same direct tunneling mechanism. Direct tunneling programming mechanism imposes the use of very thin oxide layers (<5 nm), which was difficult to achieve because any pinhole in I_1 will cause all the charge stored on M_1 to leak off. To overcome these technological constraints of MIMIS cell, the first solution was introduction of MNOS cell in 1967 by Wegener. In MNOS cell, the M_1 and I_2 was replaced by a nitride layer as shown in figure 2.2, which contains high density of trapping centers to capture holes and electrons. These traps fulfill the storage function of M_1 with the important difference that any pinhole in thin oxide layer (I_1) will not result in complete discharge of the cell since individual traps are isolated from each other by the nitride. The MNOS device has the intrinsic advantage that both programming and erasing operations can be performed electrically.

Figure 2.2 illustrates the progression of device cross-section, which has led to SONOS device structure. Initial device structure in early 70's were p-channel MNOS structures with aluminum gate electrodes, thick (i.e.45 nm) silicon nitride storage layers and program/erase voltage of 25-30V.In the late 70's and early 80's scaling moved to n-channel SNOS devices with program/erase voltages of 14-18V.In the late 80's and early 90's, both n and p-channel SONOS device emerged with write/erase voltage of 5-12V.

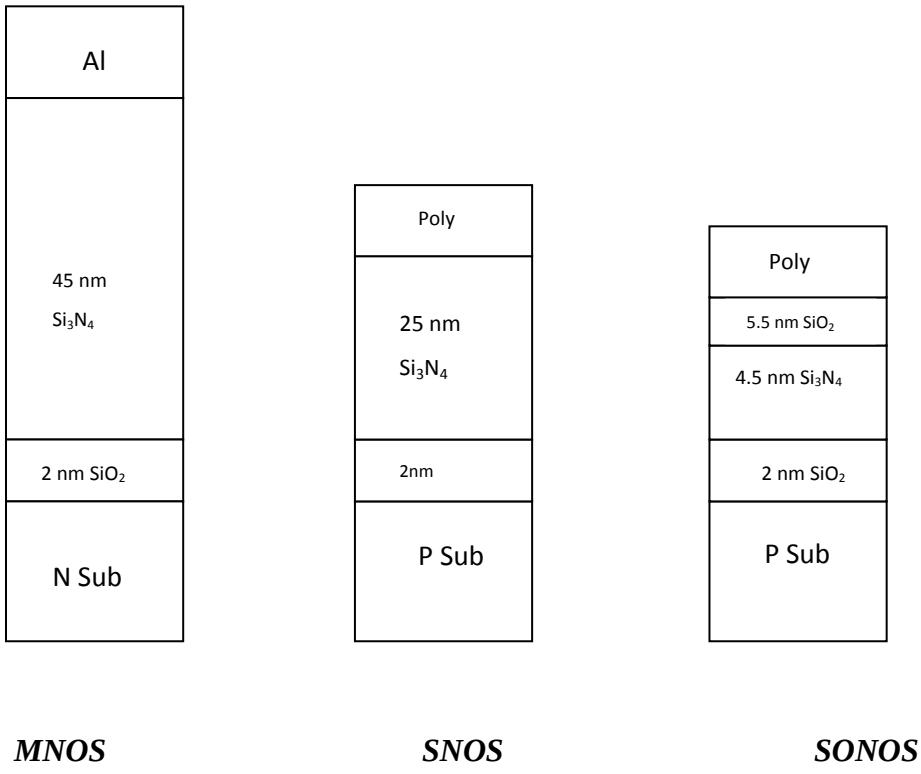


Figure 2.2 Evolution of SONOS nonvolatile memory device

MNOS employed a 1 transistor per bit configuration based on the tri-gate transistor cell concept [8]. In this transistor only the center part of the cell channel contained the programmable UTO-nitride sandwich structure. At both drain and source, a thicker oxide-nitride sandwich was used, which induced a fixed threshold voltage in the erased state and prevented the device from going into the depletion mode. These memory devices suffered from low-speed, limited-density, inherent read disturbance and the need for 2 to 3 voltage supplied to operate the memory. Erasure in MNOS is obtained by tunneling of holes from the semiconductor to the nitride traps when V_G is negative and sufficiently high.

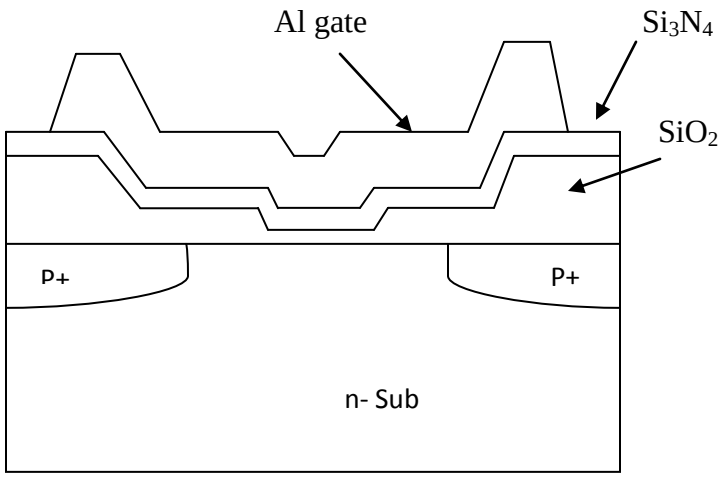


Figure 2.3 Cross Section of the p-channel tri-gate MNOS device. The thin tunneling oxide (1.5-3 nm) is presented only at the center of the channel. At source and drain, a thicker oxide-nitride sandwich acts as a select transistor.

An important breakthrough for MNOS occurred in 1980, called SNOS (Silicon-Nitride-Oxide-Silicon), which improved the charge retention of MNOS memories. Use of an aluminum gate in an MNOS device does not allow the manufacture of dense and high speed circuits because Al deposition is not a self-aligned process. Also the integration of the MNOS technology in a VLSI or ULSI process is not easy. Thus, it was desirable to develop a polysilicon gate technology. Replacing aluminum by polysilicon seems a simple approach but it is an unpractical one because, silicon/silicon nitride interface is unstable. This instability, observed in a SNOS [9] (poly Silicon-Nitride-Oxide-Silicon) memory capacitor has been remedied by employing a SONOS (poly Silicon-Oxide-Nitride-Oxide-Silicon) structure. Cross-section of SNOS is shown in figure 2.4. It is a two transistor per bit configuration in which MOS acts as a select device whose implementation completely eliminated the problem of read disturbance [10]. The SNOS consists of a silicon nitride layer (20-40 nm) on top of the UTO on silicon. The programming of the cell is as follows: during the write operation, a high (positive)

voltage is applied to the gate with the well grounded. Electrons tunnel from the silicon conduction band into the nitride conduction band

through Modified Fowler-Nordheim tunneling process and are trapped in the nitride traps, resulting in a positive threshold voltage shift. Erasing is achieved by grounding the gate and applying a high (positive) voltage to the well. This induces direct tunneling of holes from the silicon valence band into the nitride valence band, or the nitride traps [11,12], resulting in a negative threshold voltage. During the off-state, the gate is grounded and the select transistor is required for proper operation within the array. Reading of the cell is accomplished by addressing the cell through the select transistor and by sensing the state of the SNOS transistor. The charge content within the nitride will be modified in time due to backtunneling of charges through the UTO. Hole injection from the gate limit the memory window, a problem that becomes more severe for thinner nitride layers as charge stored in it is widely distributed. To cope with this problem, one solution is to use silicon oxynitride layer instead of a pure silicon nitride layer. But this requires increased programming voltages because of their higher energy barriers. And another solution is SONOS devices.

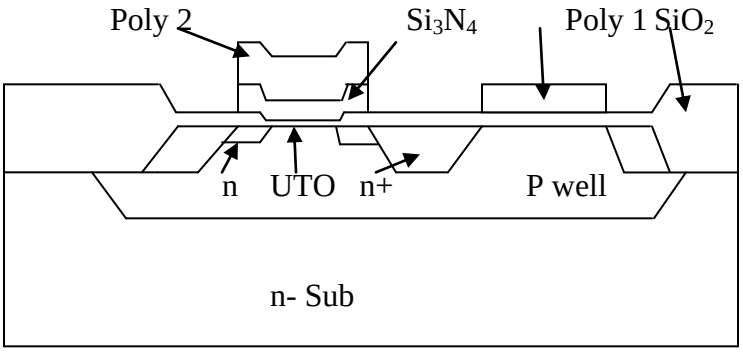


Figure 2.4 Cross Section of the two –transistor n-channel SNOS memory cell consisting of a MOS select transistor and a SNOS memory transistor, both located in a p-well

In the SONOS structure, the charge centroid is closer to the silicon surface. The presence of a SiO_2 layer between the nitride and the gate electrode modifies the conditions which limit the amount of charge which can be stored in the nitride during writing or erasing. It also modifies the decay of the stored charge during retention.

2.2 Advantages of SONOS over general flash memory:

1. In general flash memory, charges are stored in the floating gate, whereas in the SONOS, charges are stored in the nitride layer in terms of structure.

2. In general flash memory, floating gate is formed using polysilicon. Thus if any one defect exist in polysilicon, it'll result in discharge of stored memory charge due to conductive properties of polysilicon gate. In contrast, in the SONOS, nitride layer is used instead of polysilicon. Hence, SONOS is less sensitive to polysilicon defect and has improved endurance [11].

3. In the flash memory, tunnel oxide having a thickness of about 70 Å is formed under the floating gate, limiting the implementation of low voltage and high speed. Whereas in SONOS direct tunnel oxide is formed under the nitride, making it possible to implement low-voltage and high speed operation [13].

4. SONOS offers radiation hardness [14] improvements over floating gate EEPROM technology. This is due to difference in charge storage method. SONOS technology relies on trapped charge that is stored in nitride dielectric, which is not easily removed in a total ionizing dose environment. In floating gate memories, charge is stored in conducting floating gates that is separated from the silicon substrate by thin (<100 Å) tunnel oxides leading to rapid charge loss. Heavy ion strikes can discharge floating gate devices. Both these effects severely limit the radiation hardness of floating gate devices which is not observed with SONOS devices.

5. Scaling of floating gate is limited in the lateral [15] (gate-to-gate) direction also. Electrons stored on the polysilicon gate exert an electric field on adjacent gates. As the geometry shrinks, this electric field becomes so strong that it unintentionally programs or erases adjacent cells. This is main reason for floating gate to reach scaling limit at a channel length of 45nm.In SONOS, the charges are electrically trapped in the nitride layer. So they do not interfere with adjacent cells. Though scaling always presents challenges, but the scaling limit of SONOS device is not as serious as the floating gate Flash memory.

6. In terms of process complexity and manufacturing steps, SONOS is advantageous over floating gate. Floating gate adds two poly gates, two oxide (one tunnel oxide, one thick oxide for the high-voltage pump circuits) and upto 10 masking steps. This increases cycle time and manufacturing costs. SONOS has only one poly gate, one (unpatterned) nitride, one oxide (pump voltages are lower) and three masking steps [16].

2.3 Structure and Theory of SONOS nonvolatile memory device.

2.3.1 Structure of SONOS device

Figure 2.5 shows the cross sectional view of SONOS transistor [17]. The device is similar to SNOS transistor except for an oxide layer between Silicon Nitride and polysilicon gate electrode. Structure of SONOS device is similar to that of n-MOSFET with ONO stack replacing the tunnel oxide.

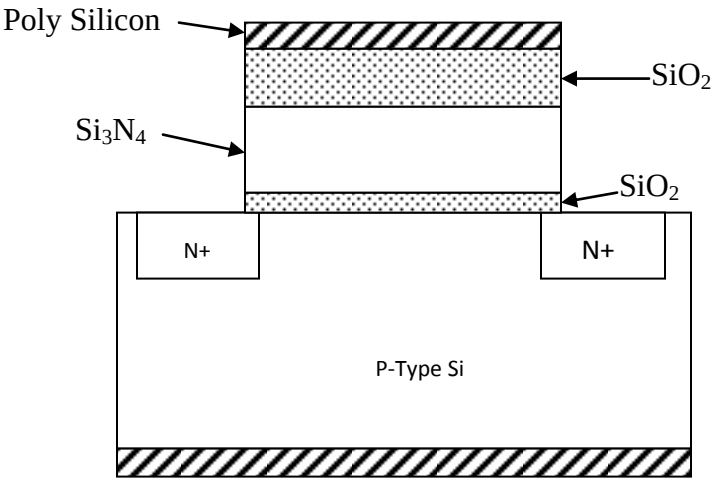


Figure 2.5 Cross Sectional View of SONOS

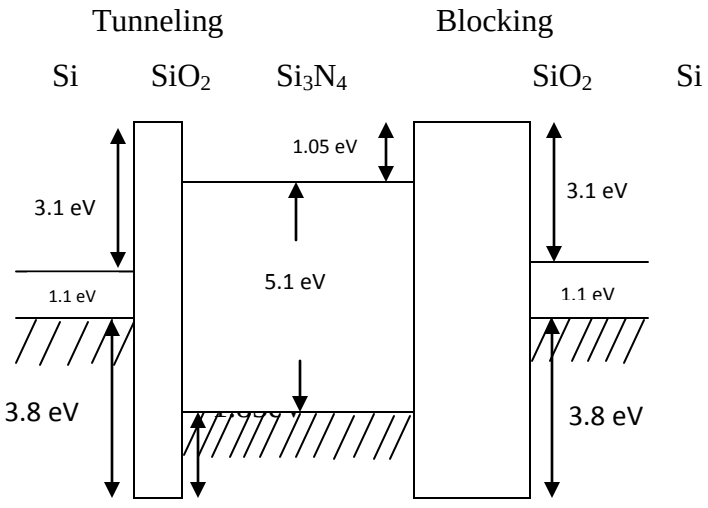


Figure 2.6 Cross Sectional View of SONOS ideal energy band diagram

If we assume that the electric charge is nil everywhere and that the difference in work function is also nil, the energy band diagrams of SONOS structures at equilibrium, without any applied voltage, is shown in figure 2.6.

2.3.2 Theory of SONOS structure

Relationship linking charges and potential:

Consider an MNOS structure under bias [17]. Voltage V_G is applied to the gate electrode, e.g. positive with respect to the silicon substrate. Generally speaking, a volume distribution ρ of charges exits within each insulator and a surface density Q of charge exists at each interface as shown in figure 2.7.

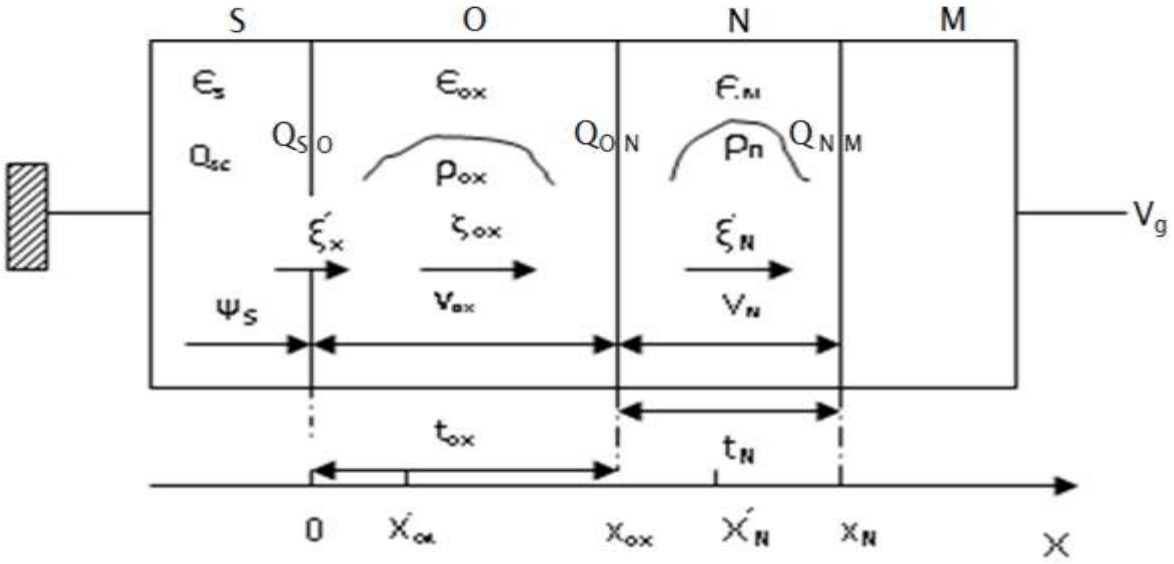


Figure 2.7 Cross Section of an MNOS capacitor and definition of symbols used.

The relationships linking charges and voltage or charges and electric fields are obtained from the Poisson's equation given as:

$$\Delta V + \rho/\epsilon = 0 \quad (2.1)$$

which is solved for each insulator. For the oxide, the first integration yields;

$$\xi_{ox}(x) = -\delta V/\delta x\,|_x = 1/\epsilon_{ox} \int_0^x \rho_{ox}\, dx + \xi_{ox}(0) \quad (2.2)$$

The second integration is used to obtain potential V(x) in the oxide.

$$V(x) - V(0) = -1/\epsilon_{ox} \int_0^x \left(\int_0^x \rho_{ox}\, dx\right)dx + x\, \xi_{ox}(0) \quad (2.3)$$

The voltage drop in the oxide is given by

$$V_{ox}=V(x_{ox}) - V(0) = -1/\epsilon_{ox} \int_0^{Xox} \left(\int_0^x \rho_{ox}\, dx\right)dx - x_{ox}\, \xi_{ox}(0) \quad (2.4)$$

The same calculation procedure applied to the nitride yields

$$\xi_N(x) = 1/\epsilon_N \int_{Xox}^x \rho_N\, dx + \xi_N(x_{ox}) \quad (2.5)$$

$$V_N=V(x_N) - V(x_{ox}) = -1/\epsilon_N \int_{Xox}^{xN} \left(\int_{Xox}^x \rho_N\, dx\right)dx - (x_N - x_{xox})\xi_N(x_{ox}) \quad (2.6)$$

By performing a partial integration of the following two terms $\int_0^{Xox}(\,)\,dx$ of equation 2.4 and $\int_{Xox}^{xN}(\,)\,dx$ of equation 2.6 we obtain:

$$V_{ox} = -Q_{ox}/\epsilon_{ox}\,(x_{ox} - x'_{ox}) - x_{ox}\,\epsilon_{ox}(0) \quad (2.7)$$

$$V_N = -Q_N/\epsilon_N\,(x_N - x'_N) - t_N\,\epsilon_N(x_{ox}) \quad (2.8)$$

Where $t_N = x_N - x_{ox}$

$$Q_{ox} = \int_0^{X ox} \rho_{ox}\, dx \quad (2.9)$$

$$Q_N = \int_{X ox}^{xN} \rho_N\, dx \quad (2.10)$$

Barycenters x'_N and x'_{ox} measured from silicon/semiconductor interface is given as

$$x'_N = 1/Q_N \int_{X ox}^{xN} \rho_N\, dx \quad (2.11)$$

$$x'_{ox} = 1/ Q_{ox} \int_0^{X ox} \rho_{ox}\, dx \quad (2.12)$$

Using Gauss's theorem, the electric fields $\xi_{ox}(0)$ and $\xi_N(x_{ox})$ can be written as

$$\xi_{ox}(0) = 1/\epsilon_{ox}(Q_{SO} + \epsilon_S\xi_S) \quad (2.13)$$

$$\xi_N(x_{ox}) = 1/\epsilon_N\,(Q_{so} + Q_{ox} + Q_{oN} + \epsilon_S\xi_S) \quad (2.14)$$

The relationship linking $\xi_{ox}(0)$ and $\xi_N(x_{ox})$ is obtained similarly,

$$\epsilon_N\, \xi_N(x_{ox}) = \epsilon_{ox}\xi_{ox}(0) + Q_{ON} + Q_{OX} \quad (2.15)$$

The voltage drop across both insulators, i.e.

$$(V_{ox}+V_N) = -Q_{ox}/\epsilon_{ox}\,(x_{ox} - x'_{ox}) - x_{ox}\,\epsilon_{ox}(0) - Q_N/\epsilon_N\,(x_N - x'_N) - t_N\,\xi_N(x_{ox})$$

$$= -Q_{ox}/\epsilon_{ox}\,(x_{ox} - x'_{ox}) - x_{ox}/\epsilon_{ox}(Q_{so} + \epsilon_s\xi_s) - Q_N/\epsilon_N\,(x_N - x'_N) -$$

$$t_N/\epsilon_N(Q_{so}+Q_{ox}+Q_{oN}+\epsilon_s\xi_s) \quad (2.16)$$

$$= -Q_{ox}/\epsilon_{ox}\,(x_{ox} - x'_{ox} + t_N(\epsilon_{ox}/\epsilon_N)) - Q_N/\epsilon_N\,(x_N - x'_N) - Q_{so}\,(x_{ox}/\epsilon_{ox} + t_N/\epsilon_N) - Q_{oN}\,t_N/\epsilon_N -$$

$$\epsilon_s\xi_s(x_{ox}/\epsilon_{ox} + t_N/\epsilon_N) \quad (2.17)$$

The total voltage applied to the structure can be expressed as

$$V_G = \Phi_{MS} + V_{ox} + V_N + \Psi_S \quad (2.18)$$

Φ_{MS} is internal voltage difference corresponding to the difference in work function between the metal and the semiconductor, divided by q.

$$V_{ox}+ V_N = -Q_{ox}/\epsilon_{ox} (x_{ox} - x'_{ox} + t_N(\epsilon_{ox}/\epsilon_N))- Q_N/\epsilon_N (x_N - x'_N)- Q_{oN}\, t_N/\epsilon_N - \epsilon_{ox}\xi_{ox}(0)(t_{ox}/\epsilon_{ox} + t_N /\epsilon_N) \quad (2.19)$$

$$V_{ox}+ V_N = (Q_{ox}/\epsilon_{ox})x'_{ox} - Q_N/\epsilon_N (x_N - x'_N) + Q_{oN}\, x_{ox}/\epsilon_{ox} - \epsilon_N\xi_N(0)(t_{ox}/\epsilon_{ox} + t_N /\epsilon_N) \quad (2.20)$$

Substituting $V_{ox}+ V_N$ in equation (2.18) electric field in the oxide at the Si-SiO_2 interface $\xi_{ox(0)}$ and $\xi_{N(Xox)}$ can be obtained as:

$$V_G - \Phi_{MS} - \Psi_S = -Q_{ox}\,[(X_{ox} - X'_{ox})/\epsilon_{ox} + t_N/\epsilon_N] - Q_N\,[(X_N - X'_N)/\epsilon_N] - Q_{ON}/C_N - \epsilon_{OX}\,\xi_{OX(0)}/C_e \quad (2.21)$$

$$V_G - \Phi_{MS} - \Psi_S = -Q_{ox}\,(X'_{ox}/\epsilon_{ox}) - Q_N\,[(X_N - X'_N)/\epsilon_N] - Q_{ON}/C_{OX} - \epsilon_N\,\xi_{N(0)}/C_e \quad (2.22)$$

Likewise, the flat band voltage of the structure, which corresponds to that value of V_G which must be applied to obtain $\Psi_s = 0, \xi_s = 0$ and $Q_{sc} = 0$,can be deduced:

$$V_{FB} = \Phi_{MS} - Q_{ox}\,[(X_{ox} - X'_{ox})/\epsilon_{ox} + t_N/\epsilon_N] - Q_N\,[(X_N - X'_n)/\epsilon_N] - Q_{SO}/C_e - Q_{ON}/C_N \quad (2.23)$$

Equation 2.23 shows that the flat band voltage depends on the charge distribution in the insulator volume. The calculation of V_{FB} requires the knowledge of atleast the charge barycenter (given as 2.11 and 2.12), shows that the charges making the largest contribution to the flat band voltage are the charges located nearest to the Silicon. A charge Q_{NM} located at the interface between the metal and the nitride makes no contribution to the flat band voltage. Similarly, a bulk distribution of charges in the nitride with a barycenter x_N located near the interface between the metal and the nitride yields a very low contribution to V_{FB}. In memory structures, the charges whose

magnitude varies with the applied voltage stress are: the charges trapped at the oxide/nitride interface (Q_{ON}) and the charges trapped in the nitride (Q_N).If all the terms in equation 2.23 which neither contain Q_{ON} nor Q_N contribute to the flat band voltage (V_{FB}) a quantity symbolized by V'_{FB}, the flat band voltage can be expressed as,

$$V_{FB} = V'_{FB} - Q_N [(X_N - X'_N)/Є_N] - Q_{ON}/C_N \quad (2.24)$$

In above equations C_{ox}, C_N, C_e are respectively the capacitance per unit area: of the oxide layer, of the nitride layer and of their series association.
C_N is given by

$$C_N = Є_N/t_N \quad (2.25)$$

$$C_{ox} = Є_{ox}/t_{ox} \quad (2.26)$$

$$1/C_e = 1/C_{OX} + 1/C_N \quad (2.27)$$

A variation of (ΔQ_{ON}) in the charge at the interface on the two insulators leads to a flat band voltage variation (ΔV_{FB}) given by

$$\Delta V_{FB} = - \Delta Q_{ON} (t_N /\epsilon_N) = - \Delta Q_{ON}/C_N \quad (2.28)$$

In measurements performed on transistors, instead of capacitors, the threshold voltage V_T is obtained more easily than flat-band voltage.V_T is gate that is just necessary to obtain a strong inversion channel at the semiconductor surface ($\Psi_s \approx 2\Psi_B$).V_T and V_{FB} are related by the equation.

$$V_T \approx V_{FB} + 2\Psi_B - Q_{SC} (t_{ox}/\epsilon_{ox} + t_N/\epsilon_N) \quad (2.29)$$

$$\Delta V_T = \Delta V_{FB} = - \Delta Q_{ON}/C_N \quad (2.30)$$

MNOS structures have several limitations. These are overcome by introducing an additional oxide layer between the nitride layer and metal electrode. There are two major reasons to introduce an additional oxide layer.

1. Use of an aluminum gate in an MNOS does not allow the manufacture of dense and high speed circuits because Al deposition is not a self-aligned process. The integration of MNOS technology in a VLSI or ULSI process is not easy. It thus became desirable to develop a polysilicon gate technology. Replacing Al with polysilicon seems a simple approach but it is an impractical [19]. This instability in a SNOS (poly Silicon-Nitride-Oxide-Silicon) memory capacitor has been remedied by employing a SONOS (Silicon-Oxide-Nitride-Oxide-Silicon) structure.

2. MNOS require high operating voltage pulses (15V-20V), for writing and erasing. It is desirable to decrease this voltage for future applications in VLSI technology. This can be performing by reducing the nitride thickness. However, the fact that the charge stored in the nitride is widely distributed, limits the possibilities of scaling down the nitride thickness. Here also the alternative solution is to introduce an oxide layer between metal gate and nitride [20]. In MONOS/SONOS structures, the charge centroid is closer to the Silicon surface. The presence of a SiO_2 layer between the nitride and the gate electrode modifies the condition which limits the amount of charge stored in the nitride during writing or erasing. It also modifies the decay of the stored charge during retention.

The basic relationships of the MONOS/SONOS structures are as follows. When a MONOS memory operates, the only charge distributions which vary are: Q_{ON}, ρ_N and its integral over t_N (i.e. Q_N) and Q_{NBO}. The flat band voltage can be written as:

$$V_{FB} = V'_{FB} - Q_{ON}\,(t_N/\epsilon_N + t_{BO}/\epsilon_{OX}) - Q_N\,[(t_N - t'_N)/\epsilon_N] - Q_{NBO}\,(t_{BO}/\epsilon_{OX}) \qquad (2.31)$$

V_{FB} is a function of t_N and t_{BO} for given Q_{ON}, Q_N and Q_{NBO}. However for a given V_{FB} value, there still exits an infinity of possible value for the (t_N, t_{BO}) couple. For a given writing or erasure pulse, the values of the charges Q_{ON}, Q_N and Q_{NBO} stored in the insulators also depend on t_N and t_{BO} values giving an expected V_{FB} is thus difficult.

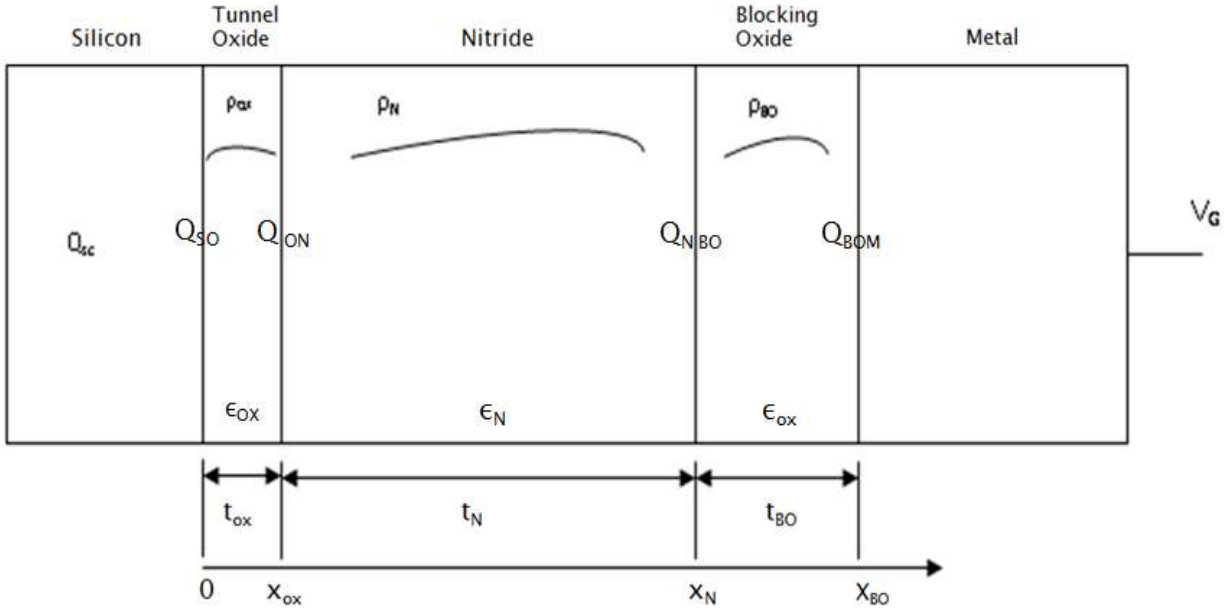


Figure 2.8 Cross Section of a MONOS capacitor and definition of symbols used.

Practically, two technological approaches have been proposed.

1. [21] Scale down the nitride thickness for a low voltage operation while memory charge is stored in the nitride. Blocking oxide is thick enough to just prevent the exchange of carriers by direct tunneling between the gate and the nitride. Since t_{BO} is as thin as possible, the effect of Q_{NBO} i.e. the effect of Q_{NBO} (charge stored at the N/BO interface, on V_{FB} (and thus on V_T) is rather small (from equation 2.9).The V_T shift is mainly due to the Q_N. Charge as in an MNOS structure. In this approach, no special effort is needed to obtain a high density of traps at the N/BO interface.

2. Other approach [20] is to make a thinner nitride layer while using a thicker blocking oxide in order to increase the V_T shift produced by charge Q_{NBO}. In that case, the fabrication method is devised to enhance the density of traps at the N/BO interface.

Whatever, approach is considered; it does not appear that a precise model of the V_T shift can be developed by taking into account the variations of only one type of charge in equation above. An interface trap density can also be considered as a large increase in bulk trap density in a very narrow region near this interface.

Relationship linking currents and charges:

In the case of a one-dimensional MNOS structure, the conservation of the total currents (i.e. the sum of the conduction and displacement currents) is expressed as :

$$\text{div } \bar{\imath}_{tot} = \text{div } [\bar{\imath}_c (x,t) + d/dt \check{D}(x,t)] = 0 \qquad (2.32)$$

By using the Poisson equation linking the electric displacement vector $\check{D}$ (with $\check{D}=\epsilon\xi$) to the charge volume density, relation (2.10) applied to the nitride becomes:

$$\delta J_N (x,t)/\delta_x = \delta\rho_N (x,t)/\delta t \qquad (2.32)$$

where ρ_N is the charge density (expressed per unit volume) in the nitride and J_N is the current density in the nitride. Finally, by using Gauss law at the O/N interface, we obtain from relation (2.32):

$$J_{ox} (x_{ox,t}) - J_N (x_{ox,t}) = dQ_{ON(t)} /dt \qquad (2.33)$$

Equations 2.32 and 2.33 express that a spatial variation in carrier current causes a temporal variation in the amount of charge, respectively in the bulk and at an interface.

The continuity equations (2.32 and 2.33) also apply to the case of a SONOS structure. SONOS requires additional continuity equation for N/BO interface, given as

$$J_N (x_N , t) - J_{BO} (x_N , t) = dQ_{NBO}/dt- \quad (2.33)$$

J_{BO}: current density in blocking oxide.

If blocking oxide is perfect, J_{BO} is not met. Similar equation can be written for other interfaces of the structures, practically corresponding charges in these areas can be neglected. During erase/write operation, significant currents flow through the device. In the general case, the charge density must include both the free carriers and the trapped carriers. The charge due to free carriers may not be negligible in the nitride especially at N/BO interface. If the trapping efficiency is low and the blocking action is good, free carriers can accumulate and contribute significantly to the charge density.

Charge Trapping by Amphoteric Traps in Silicon Nitride

From the research on the origin of nitride traps, memory effect in the nitride layer in MNOS and MONOS devices are due to silicon dangling bonds (SDB).Silicon dangling bonds possess three states of charge of amphoteric traps denoted as D^+, D^0, D^- .

A trap in D^+ state contains a positive charge and can be considered to be devoid of electrons.

A trap in D^0 state is neutral and D^- state contains a negative charge and can be considered to be occupied by electrons. The different interactions between an amphoteric trap, with its three states of charge, and free electrons and holes are illustrated in figure 2.9a.

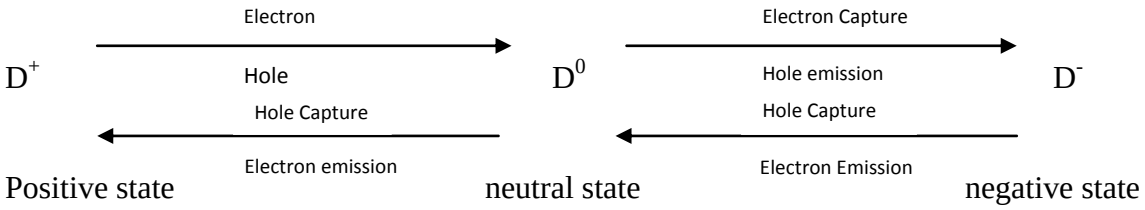


Figure 2.9 (a) Illustration of different states of charge

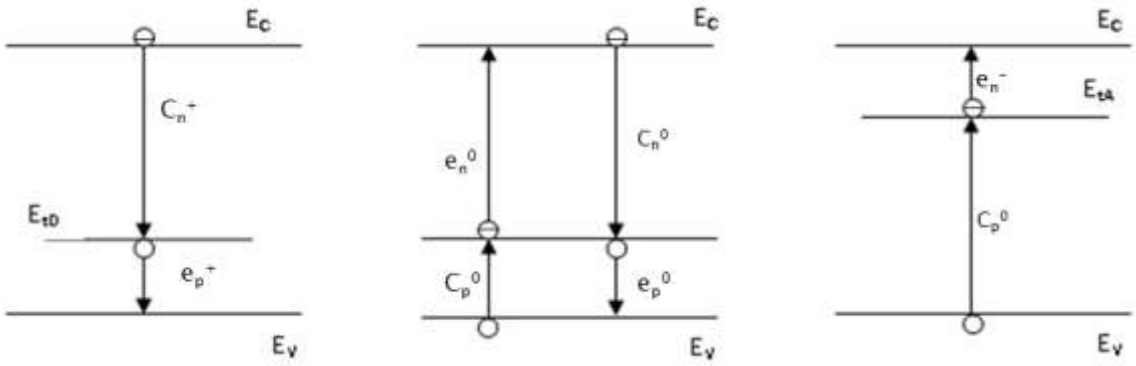


Figure 2.9(b) : Carrier exchange processes of an amphoteric trap

By considering that the amphoteric trap possesses two energy levels, E_{tA} and E_{tD} ,the possible transitions are shown in figure 2.9 (b) on an energy band diagram of the nitride bulk. In this representation the neutral state is associated with energy level E_{tD} and $E_{tA}>E_{tD}$. The fraction of traps encountered in the D^+, D0, D^- states are denoted respectively by f^+, f^0, f^- .The variations of these occupancy functions can be expressed by considering that the elementary capture and emission processes defined in figure 2.9(b) obey the Shockley, Read, Hall (SRH) theory.

$$df^+/dt = -C_n^+ n f^+ - e_n^+ f^+ + e_n^0 f^0 + C_p^0 p f^0 \quad (2.34)$$

$$df^+/dt = -C_n^0 n f^0 - e_n^0 f^0 + e_n^- f^- + C_p^- p f^- \quad (2.35)$$

$$f^0 + f^+ + f^- = 1 \quad (2.36)$$

In equation 2.33, 2.34, 2.35, n and p are the densities of electrons and holes in the nitride conduction and valence band respectively, C and e are the capture coefficients and emission probabilities respectively. The upped script denotes the initial states of charge (+,0,-) and the subscript indicates the carrier type (p or n) involved in the process.
The density of charge trapped in the nitride

$$\rho_N = q N_t (f^+ - f^-) \quad (2.37)$$

If the nitride traps are located near the O/N interface, a carrier exchange can take place not only between the traps and the nitride bands but also between the traps and either the silicon bands or the fast states of the Si/SiO_2 interface. In these cases, the e and C coefficients, which now involve tunnel exchange mechanisms, differ from those of SRH theory which only imply thermal processes. This is illustrated in Figure 2.10 for two elementary emission processes.

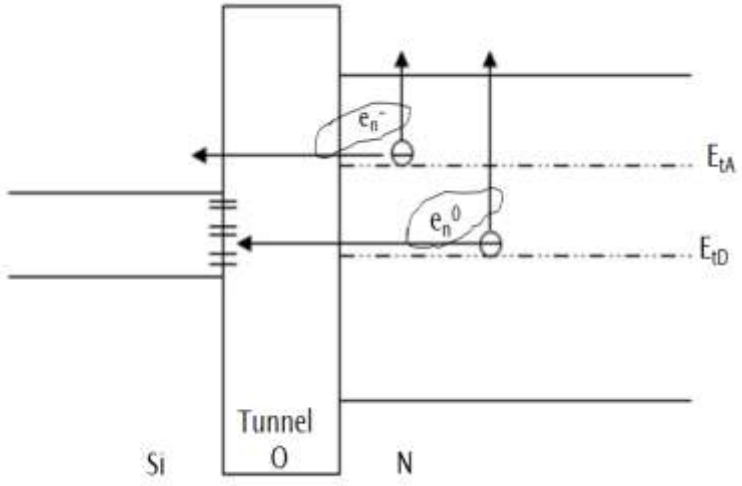


Figure 2.10 Representation of two types of electron process for both D^- state (E_{tA}) and D^0 state (E_{tD})

Modeling of the Memory Behavior of MONOS and SONOS devices:

Modeling of the behavior of an MONOS cell while taking into account all elementary carriers exchange processes appears as a complex task. This can be simplified by considering specific conditions regarding the trap energy levels, the initial state of charge and the applied bias. A complete model of the switching of MONOS/SONOS memories is given in figure 2.11.Both free electrons and holes are assumed to exist in the nitride. They are injected into or escape from the nitride by Fowler-Nordheim or modified Fowler-Nordheim effects as indicated in figure 2.11(a) (J_{On} ,J_{Bo} , J_{Bop} , J_{Op}) for a negative gate bias. A trap-assisted tunneling current is also considered for electron injection from

the silicon under positive gate bias and low field condition (J_{TA} on figure 2.11(b).During negative bias, back tunneling of electrons is also considered. The total nitride charge includes: trapped holes, trapped electrons and the free carriers, i.e. electrons in the conduction band and holes in the valence band. This can be written:

$$\rho_N = q [N_t (f^+ - f^-) + p – n] \quad (2.38)$$

For the write/erase process, the equations describing the variations of the occupancy function neglect the emission (detrapping) processes but retain the four capture processes of holes and electrons.

Hence relation 2.34 and 2.35 becomes:

$$df^+ /dt = -C_n^+ n f^+ + C_p^0 p f^0 \quad (2.39)$$

$$df^- / dt = - C_n^0 n f^0 + C_p^- p f^- \quad (2.40)$$

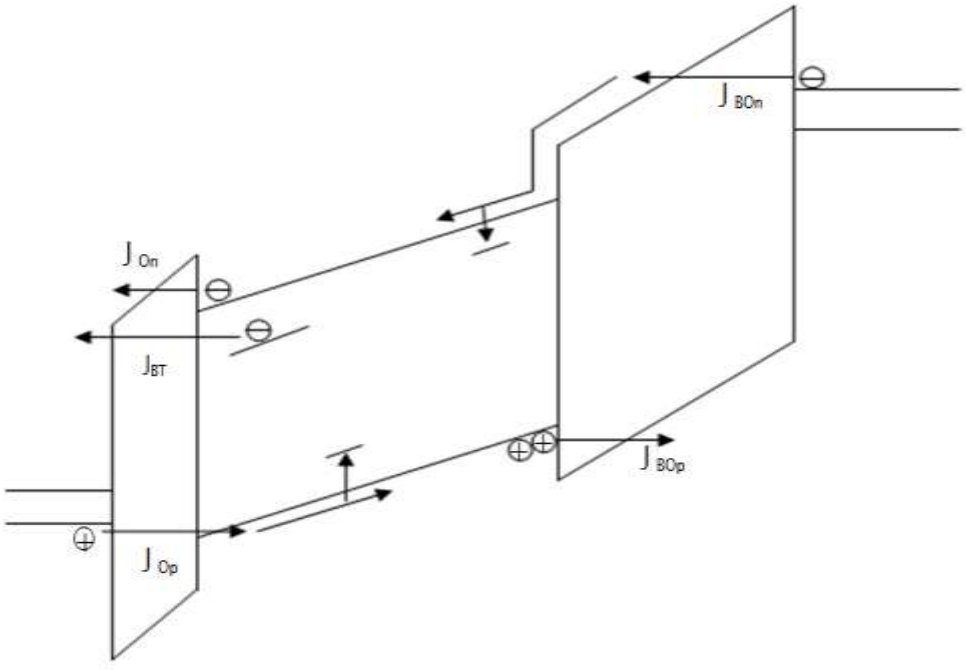


Figure 2.11 (a): SONOS structure under negative bias

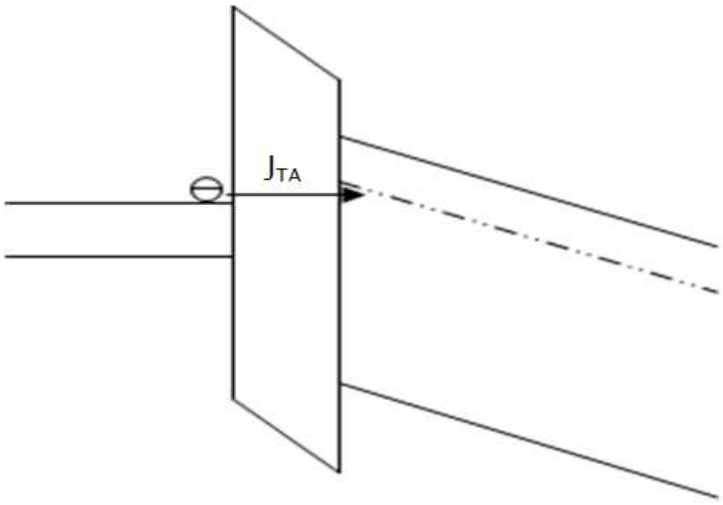


Figure 2.11 (b) : band-to-trap currents under positive gate bias

Figure 2.11: Electron and hole currents flowing through a SONOS structure

Modeling of the Retention Behavior of MONOS/SONOS devices:

Discharge [22] of (previously negatively charged) amphoteric traps is considered to be due to back tunneling of electrons from the traps in the D^- state(figure 2.12)Dominant processes of nitride discharge are considered in this section. Consider energy distribution of traps with a constant energy difference $U = E_{tA} – E_{tD}$ is applied.

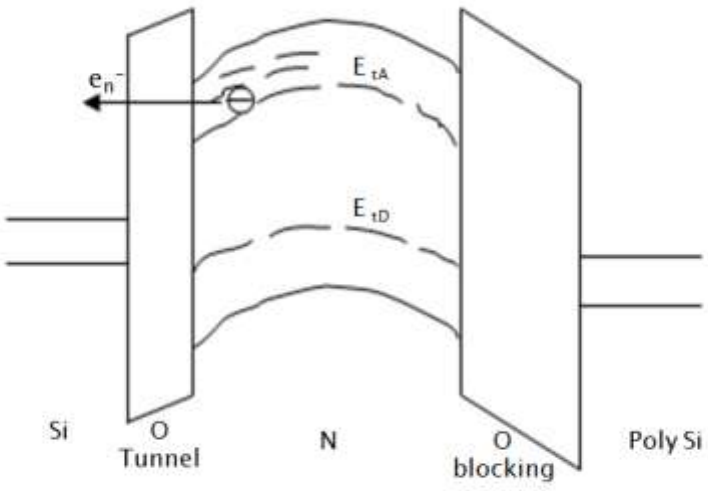


Figure 2.12 (a): with a large negative charge stored in the nitride.

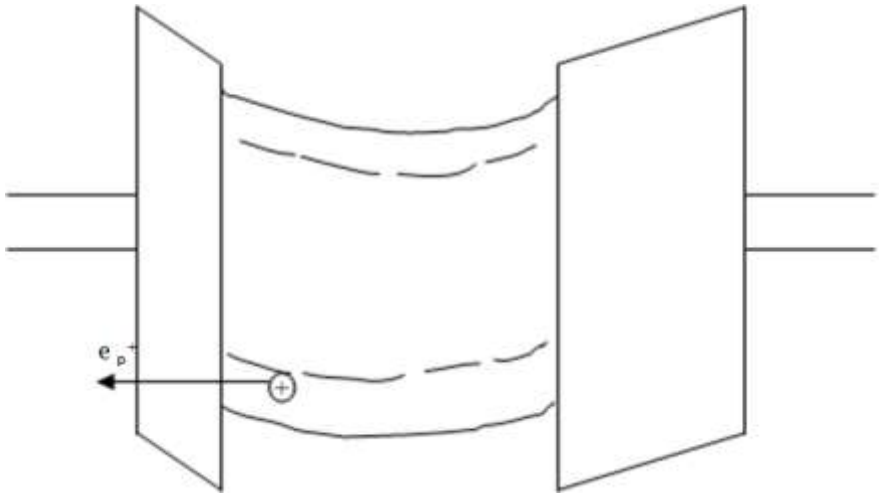


Figure 2.12 (b): with a large positive charge stored in the nitride.

Figure 2.12 – Band diagram of a MONOS structure in the retention mode.

The discharge [23] of (previously negatively charged) amphoteric trap is assumed to take place through the first three elementary processes shown in figure 2.13 which contribute to the nitride discharge.

Process 1: back tunneling of electrons from the traps in the D^- state (e_n^-).

Process 2: back tunneling of electrons from the traps in the D^0 state to the fast states of the Si/SiO_2 interface (e_n^0).

Process 3: Capture of holes from the silicon valence band by the traps in the D^0 state (C_p^0).

In such condition equations expressing the variations in occupancy functions become:

$$df^+ /dt = + e_n^0 + C_p^0 f^0 \quad (2.41)$$

$$df^- / dt = -e_n^- f^- \quad (2.42)$$

In figure 2.13, p is free hole density in silicon at the Si/SiO_2 interface. Only two single levels, E_{tA} ,E_{tD}, are considered and all traps are assumed to be initially in the D^- state.

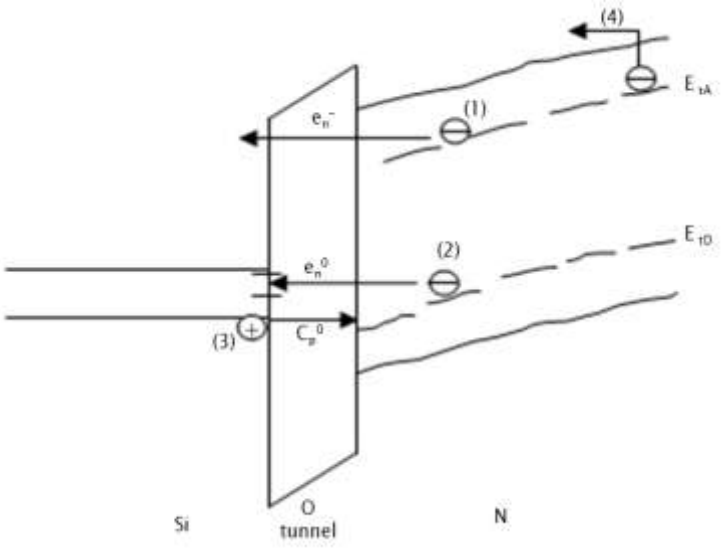


Figure 2.13 – Illustration of various exchange mechanisms which take part in the discharge of the nitride when a large negative charge is stored.

Figure 2.13 shows important retention properties. The short term decay of the stored charge is governed by process (1) while the charge loss occurring during long term retention is mainly governed by processes (2) and (3).This model does not take into account the thermal (Poole-Frenkel) emission [24] of electrons trapped in the bulk of the nitride (4).

2.4 Physical operation of SONOS device:

2.4.1 Programming:

During program, positive bias is applied to the poly gate resulting in a band diagram as shown in figure 2.9.

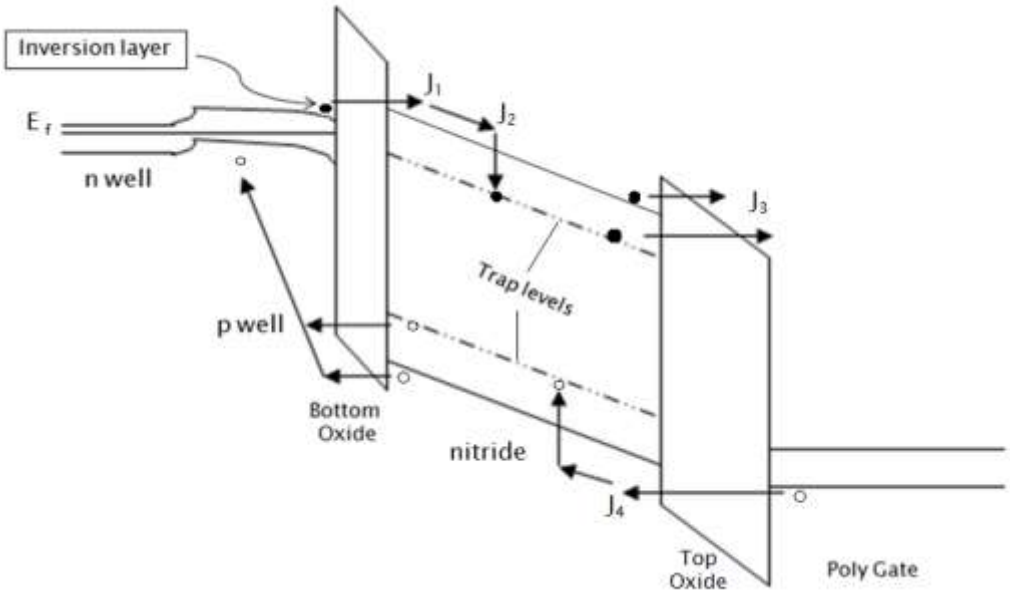


Figure 2.14 SONOS program

Following four tunnel phenomena is held responsible for change in threshold voltage, V_t of the device.

- J1: The electron tunneling current from substrate to nitride conduction band.

- J2: In nitride electrons are trapped and drift towards the top oxide by Poole-Frenkel conduction.
- J3: The electrons reaching the top oxide may tunnel through it to be collected by the polysilicon gate.
- J4: The hole injection current from the poly-silicon gate to the Nitride Valence band. There is no hole leakage current corresponding to tunneling current from the Nitride valence band to the substrate.

Contribution from the holes is assumed to be negligible due to two reasons:

a) thick blocking oxide and
b) Higher tunneling current barrier for holes than electrons. That is in oxide the potential barrier for holes is less than electrons.

A positive shift is observed due to capture of tunneled electrons in the nitride traps.

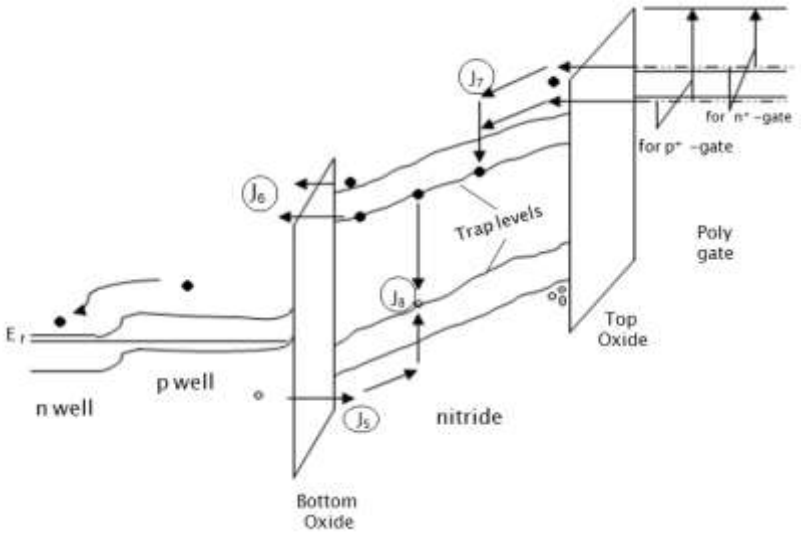
In the program of write operation, electrons quantum-mechanically tunnel from a silicon inversion layer through an ultra thin oxide. The electrons arrive in the silicon nitride film, where they are stored in deep-level traps, which lie about 2.5eV below the edge of the silicon nitride conduction band. The electrons, which are not trapped in the nitride film, tunnel through a blocking oxide into the gate electrode.

2.4.2 Erase:

During erase process, negative bias is applied to the gate.

Under erase mode, following phenomena is observed:

- J_5 : Holes are injected from semiconductor into the nitride.
- J_6 : Electrons previously injected by a write-erase pulse back tunnel into the substrate.
- J_7: Back injection of electron from the poly gate to nitride conduction band.
- J_8 : Recombination of holes and electrons at trap levels in nitride.



2.15 SONOS erase

2.4.3 Retention :

A SONOS cell once programmed does not retain all trapped electrons for longer period of time, even when the applied gate voltage is zero. Different mechanisms explain the retention loss occurring in programmed state as shown in figure 2.10

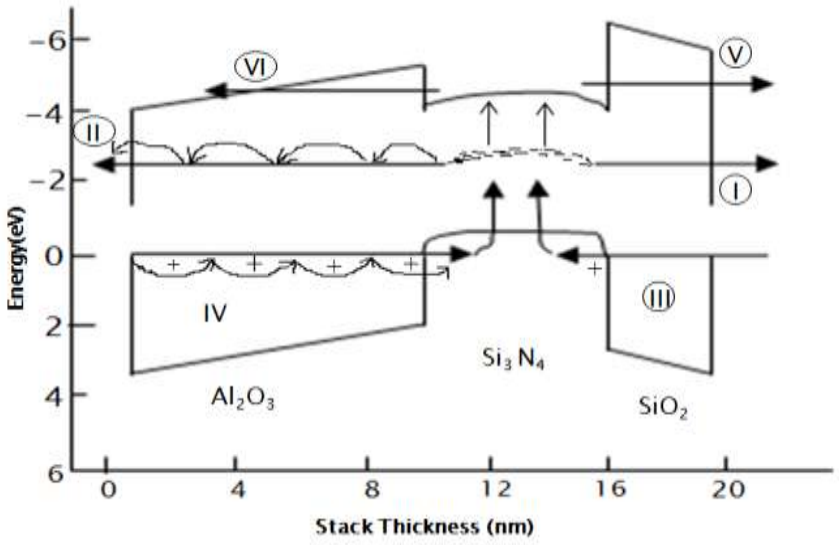


Figure 2.16 Programmed state retention loss mechanism

(I) Direct tunneling of electrons from nitride to substrate

(II) Direct tunneling of electrons from nitride to gate.

(III) Direct tunneling of holes from substrate to nitride.

(IV) Direct tunneling of holes from gate to nitride.

(V) (VI) Thermal detrapping of trapped electrons into the nitride conduction band and subsequent tunneling into the substrate and gate [25].

This phenomena results in decrease of V_t of the device. Contrarily in erase state it has been found that V_t of the device increases due to hole leakage current from nitride to substrate [25-26]. The positive shift in V_t is found to be independent of temperature.

2.5 Device Physics:

2.5.1 Tunneling Process:

This section involves various tunneling mechanism in SONOS device operation. Tunneling is a quantum mechanical effect by which a particle can penetrate an energy

barrier that would be forbidden by the classical laws of particle mechanics. This is the most basic phenomena occurring during SONOS device operation. The computation of the tunneling probabilities is, by default, based on the WENTZEL-KARMERS-BRILLOUIN (WKB) approximation .Within the WKB approximation, the transmission coefficient TC can be written as

$$TC(\xi) = \exp\left(-\frac{2}{\hbar}\int_{x_1}^{x_2}\sqrt{2m_{diel}(W(x)-\xi)}dx\right) \qquad (2.34)$$

In this expression, the integration is performed only within the region defined by $\xi \leq W(x)$.ξ represents the particle energy and W(x) is the variation of the potential energy barrier as a function of the tunneling distance x in the dielectric. Since W(x) is a function of the electric field present across the dielectric, three tunneling regimes arise depending upon the magnitude of the applied electric field or voltage and oxide thickness, as shown in figure 2.10.The specific case of electron tunneling from the silicon substrate to the nitride conduction band through the bottom oxide (tunnel oxide) (current component J1 in figure 2.8) will be used to explain the tunneling phenomena.

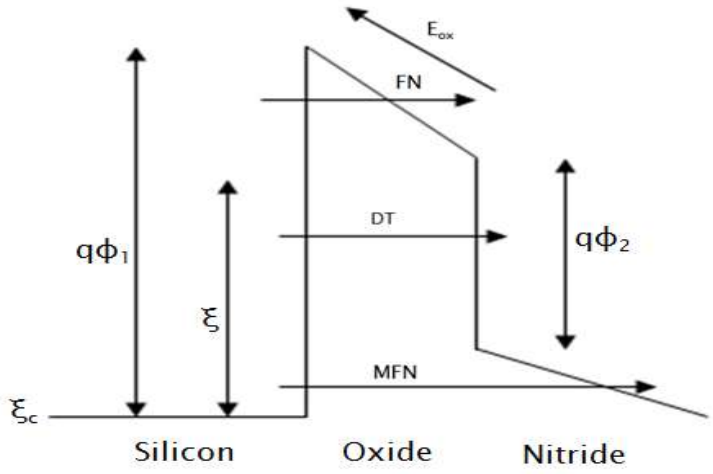


Figure 2.17 Different Tunneling Regimes: (a) FN (b) DT (c) Modified FN

2.5.1.1 Fowler Nordheim (FN) Tunneling:

For and electric field E_{ox} in the oxide

$$E_{OX} \geq \Phi_1 / d_{ox} \tag{2.35}$$

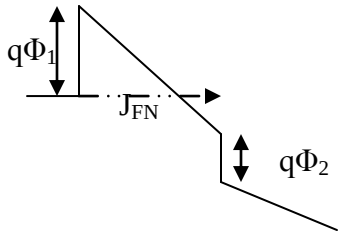


Figure 2.18 Conduction band of SONOS structure for Fowler Nordheim Tunneling

Tunneling is said to occur under the FN regime when W(x) is a linear function of x i.e. the tunneling barrier is triangular in shape as shown in figure 2.10 .For a given Electric Field is the oxide E_{ox} ,FN tunneling occurs for particles with energy $\xi \geq q\Phi_1 - qE_{ox}d_{ox}$ where d_{ox} is the tunnel oxide thickness. Consequently, $W(x) = q\Phi_1 - qE_{ox}\, d_{ox}$ and upper limit of the integration in equation of TC (ξ) reduces to $x_2 = (q\Phi_1 - \xi)/qE_{ox}$ .Hence the equation for the case of FN tunneling reads as:

$$TC(\xi) = \exp\left(-\frac{2}{\hbar}\int_0^{x_2}\sqrt{2m_{ox}(q\phi_1 - qE_{ox}x - \xi)}dx\right) \tag{2.36}$$

This evaluates to

$$TC(\xi) = \exp\left(-4\frac{\sqrt{2m_{ox}}}{3\hbar qE_{ox}}(q\phi_1 - \xi)^{3/2}\right) \tag{2.37}$$

The tunneling current density J_{FN} is directly proportional to the transmission coefficient TC, and can be written as [27]

$$J_{FN} = \frac{m_o q^3}{8\pi m_{ox} h q \phi_1} E_{ox}^2 \exp\left(-4\frac{\sqrt{2m_{ox}}}{3\hbar q E_{ox}}(q\phi_1)^{3/2}\right) \qquad 2.38$$

2.5.1.2 Direct Tunneling (DT):

For an electric field

$(\Phi_1 - \Phi_2)/d_{ox} < E_{ox} < \Phi_1 /d_{ox}$ (2.39)

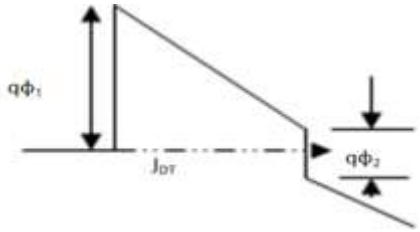


Figure 2.19 Conduction band diagram of Direct Tunneling.

In the DT regime, the electrons face a trapezoidal barrier as shown in figure 2.10.In this case the particle energy $\xi \in (q\Phi_1 - q\Phi_2 - qE_{ox}\, d_{ox}\,,\, q\Phi_1 - qE_{ox}\, d_{ox})$ and the equation of TC (ξ), under the assumption of a trapezoidal barrier, reduces to

$$TC(\xi) = \exp\left(-\frac{2}{\hbar}\int_0^{d_{ox}} \sqrt{2m_{ox}(q\phi_1 - qE_{ox}d_{ox} - \xi)}dx\right) \qquad (2.40)$$

This further evaluates to [27]

$$TC(\xi) = \exp\left(-4\frac{\sqrt{2m_{ox}}}{3\hbar q E_{ox}}\left(\left(q\phi_1 - \xi\right)^{3/2} - \left(q\phi_1 - qE_{ox}d_{ox} - \xi\right)^{3/2}\right)\right) \qquad (2.41)$$

The DT current density J_{DT} is directly proportional to the transmission coefficient TC and given by the equation

$$J_{DT} = \frac{m_o q^3 \exp\left(-4\frac{\sqrt{2m_{ox}}}{3\hbar qE_{ox}}\left(\left(q\phi_1\right)^{3/2} - \left(q\phi_1 - qE_{ox}d_{ox}\right)^{3/2}\right)\right)}{8\pi m_{ox} h\left(\sqrt{q\phi_1} - \sqrt{q\phi_1 - qE_{ox}d_{ox}}\right)^2} E_{ox}^2 \quad (2.42)$$

d_{ox} is thickness of oxide, q is the electron charge,m_0 is the mass of free electron, T is the temperature ,ħ is the reduced Planck's constant, q Φ_1 is the tunneling oxide barrier height and m_{ox} is the effective mass of a an electron in the oxide.

2.5.1.3 Modified FN Tunneling:

For an electric field

$$(\Phi_1 - \Phi_2)/ (d_{ox} - \gamma d_N) < E_{ox} < (\Phi_1 - \Phi_2)/d_{ox} \quad (2.43)$$

modified Fowler Nordheim Tunneling occurs.

γ = ratio of oxide and nitride dielectric constant= ϵ_N/ϵ_{ox}

d_N = thickness of the nitride.

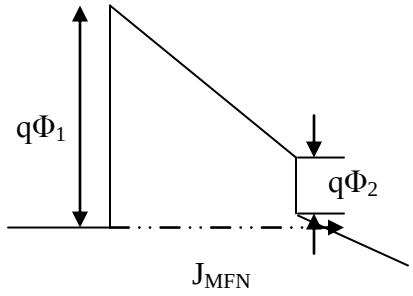


Figure 2.20 Conduction band diagram of Modified Fowler Nordheim Tunneling.

In this tunneling regime, the electron tunnels through a trapezoidal energy barrier in the oxide and a triangular energy barrier in the nitride before reaching the nitride conduction

band as shown in figure 2.15.The transmission coefficient can be approximately calculated as the product of TCs for oxide DT tunneling and nitride FN tunneling.MFN tunneling will be observed if, for the given electric field in the tunnel oxide E_{ox}, the electron energy $\xi < q\Phi_1 - q\Phi_2 - q\,E_{ox}\,d_{ox}$.

$$TC(\xi) = \exp\left[-4\frac{\sqrt{2m_{ox}}}{3\hbar qE_{ox}}\left(\begin{array}{l}\left(q\phi_1-\xi\right)^{3/2}-\left(q\phi_1-qE_{ox}d_{ox}-\xi\right)^{3/2}\\ +\frac{\varepsilon_n}{\varepsilon_{ox}}\sqrt{\frac{m_n}{m_{ox}}}(q\phi_1-q\phi_2-qE_{ox}d_{ox}-\xi)^{3/2}\end{array}\right)\right] \quad (2.44)$$

Where ε_n and ε_{ox} represent nitride and oxide dielectric constants respectively.
The tunneling current density J_{MFN} is then written as [27]:

$$J_{MFN} = \frac{m_o q^3 \exp\left[-4\frac{\sqrt{2m_{ox}}}{3\hbar qE_{ox}}\left(\begin{array}{l}\left(q\phi_1\right)^{3/2}-\left(q\phi_1-qE_{ox}d_{ox}\right)^{3/2}\\ +\frac{\varepsilon_n}{\varepsilon_{ox}}\sqrt{\frac{m_n}{m_{ox}}}(q\phi_1-q\phi_2-qE_{ox}d_{ox})^{3/2}\end{array}\right)\right]}{8\pi m_{ox}h\left(\left(\sqrt{q\phi_1}-\sqrt{q\phi_1-qE_{ox}d_{ox}}\right)+\frac{\varepsilon_n}{\varepsilon_{ox}}\sqrt{\frac{m_n}{m_{ox}}}\sqrt{\left(q\phi_1-q\phi_2-qE_{ox}d_{ox}\right)}\right)^2}E_{ox}^2 \quad (2.45)$$

Hence, it is evident that electrons at different energy levels will undergo tunneling through different tunneling mechanisms (FN/DT/MFN) depending upon the magnitude of the electric field in the oxide and the energy level.

2.5.1.4 Trap Assisted Tunneling:

Electric field for trap assisted tunneling is given as

$$E_{ox} \leq (\Phi_1 - \Phi_2 - \Phi_t)/d_{ox} \quad (2.46)$$

Where $q\Phi_t$ is the trap energy below the conduction band.

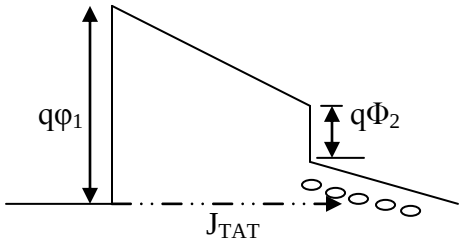


Figure 2.21 Conduction band diagram of Trap Assisted Tunneling

2.5.2 Emission mechanism:

Emission of trapped electrons and holes in nitride are the key processes involved during SONOS erase and retention. Some of important emission models are as follows:
Several discharge mechanisms may be responsible for time and temperature dependent retention behavior of SONSO devices. Figure 2.17 shows a bandgap diagram of a SONOS device in the excess electron state, illustrating trap-to-band tunneling, trap-to-trap tunneling, band-to-trap tunneling, thermal excitation and Poole-Frenkel emission mechanisms. These mechanisms may be classified into two categories.

i) tunneling processes that are not temperature dependent [28]:
During retention, trapped electrons can ‘back-tunnel’ to the conduction band of the silicon substrate (trap-to-band tunneling), under the influence of an internal self-built electric field. Meanwhile, holes from the substrate may tunnel through

the thin tunneling-oxide and become trapped in the nitride (band-to-trap tunneling).

ii) This category contains those mechanisms that are temperature dependent [29]: Trapped electrons may redistribute vertically inside the nitride by Poole-Frenkel emission, which will give rise to a shift in the threshold voltage. Also at elevated temperatures, trapped electrons can also be thermally excited out of the nitride traps and into the conduction band of the nitride (thermal excitation), and drift towards the tunnel oxide, followed by subsequent tunneling into the silicon substrate.

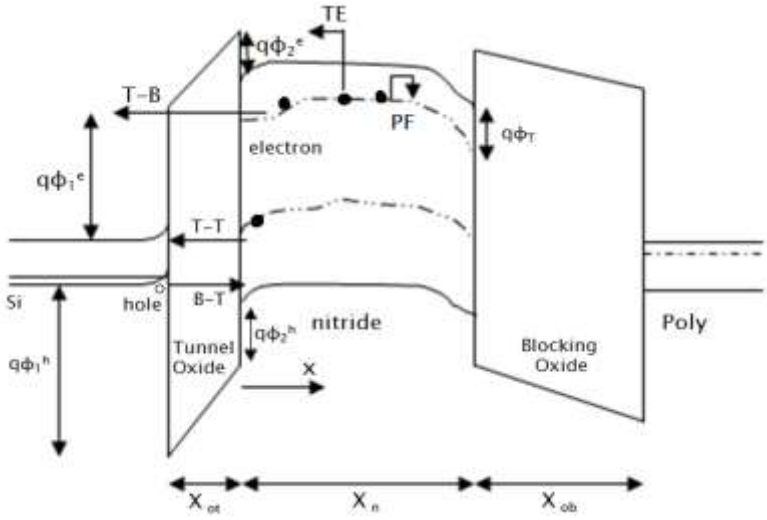


Figure 2.22 Bandgap diagram of a SONSO device in the excess electron state, showing retention loss mechanisms: trap-to-band tunneling (T-B), trap-to-trap tunneling (T-T), band-to-trap (B-T), thermal excitation (TE) and Poole-Frenkel emission (PF)

2.5.2.1 Poole Frenkel Effect

Poole Frenkel effect [18] is one of the most important retention loss mechanisms. Trapped particles may eject out of the tarp if they have sufficient energy to overcome the trap depth Φ_t .Thermal emission can be modeled by the Arrhenius equation, given as

$$e_{TH} = \upsilon_0 \exp\left(\frac{q\phi_t}{\kappa T}\right) \quad (2.47)$$

Where v_o is the attempt to escape frequency.
This frequency is given as

$$\upsilon_0 = A T^2 = 2\sigma\sqrt{\frac{3\kappa T}{m_{eff}}}\left(\frac{2\pi m_{eff}\kappa T}{h^2}\right)^{3/2} \quad (2.48)$$

σ is the capture cross section associated with the trap.

In the presence of high electric field, the emission process by this mechanism is enhanced due to the lowering of the barrier of a columbic trap as shown in figure 2.17

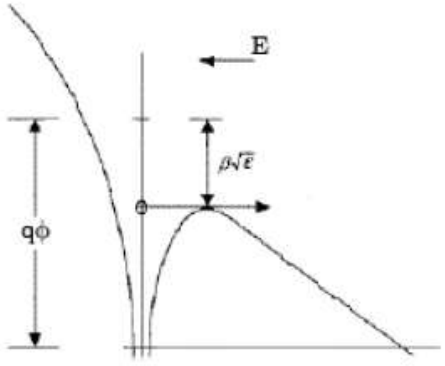


Figure 2.23 Poole-Frenkel effect: Field assisted barrier lowering

The extent of reduction in the effective trap depth is given by,

$$\Delta\Phi_t = \beta\,(E)^{0.5} \quad (2.49)$$

Where,

$$\beta = [q/\,(\pi\varepsilon_\infty\,\varepsilon_0)]^{1/2} \quad (2.50)$$

ε_∞ is the high frequency dielectric constant of the material (Nitride).

Barrier lowering due to Poole Frenkel effect diminishes during the later stage of retention, as the internal electric field relaxes after electron discharge.

2.5.2.2 Trap-to-Band Emission:

It is possible for trapped electrons (holes) in the nitride to tunnel out of the traps into either the nitride or the substrate/Poly-silicon gate conduction (valence) band [30]. These possibilities are illustrated in figure 2.18.In the presence of large electric fields (during program/erase), it is highly likely that tunneling occurs to a point inside the nitride, especially for particles distant from the interface. The tunneling process can be regarded as a FN type tunneling in the nitride with the barrier height equal to the trap depth.

$$TC(\xi) = \exp\left[-4\frac{\sqrt{2m_{ox}}}{3\hbar qE_{ox}}(q\phi_1 - \xi)^{3/2}\right] \qquad (2.51)$$

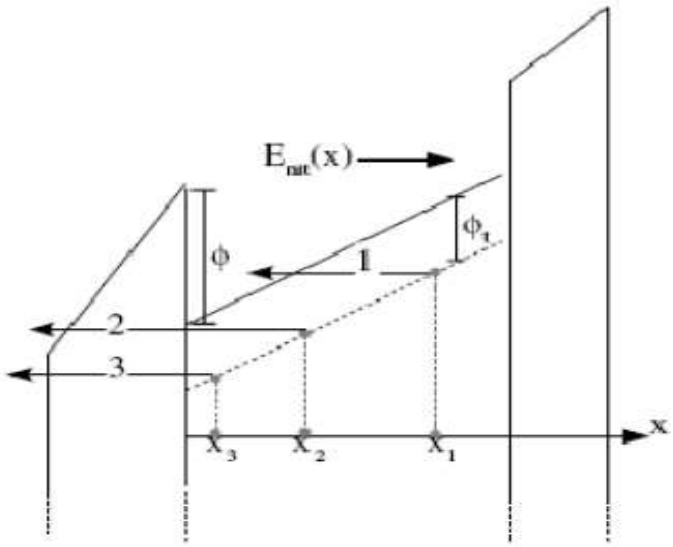


Figure 2.24 Trap-to-Band Emission Mechanisms

Modifying equation 2.31 to get the tunneling probability gives,

$$T = \exp\left[-4\frac{\sqrt{2m_{nit}}}{3\hbar qE_{nit}(x_1)}(q\phi_t)^{3/2}\right] \quad (2.52)$$

The particle may also tunnel directly to the substrate /poly-silicon gate conduction. This process can be adequately modeled as two consecutive tunneling processes, the first being a direct tunneling to the interface, followed by a DT/FN tunneling through the oxide to the substrate/gate. The probability is given as

$$T = T_{nit}\, T_{ox} \quad (2.53)$$

Where,

$$T_{nit} = \exp\left[-4\frac{\sqrt{2m_{nit}}}{3\hbar qE_{nit}}\left(\left(q\phi_t\right)^{3/2} - \left(q\phi_t - qE_{nit}(x_2)x_2\right)^{3/2}\right)\right] \quad (2.54)$$

And,

$$T_{ox} = \exp\left[-4\frac{\sqrt{2m_{ox}}}{3\hbar qE_{ox}}(q\phi_t + q\phi)^{3/2}\right]...(FN)$$
$$= \exp\left[-4\frac{\sqrt{2m_{ox}}}{3\hbar qE_{ox}}\left(\left(q\phi_t + q\phi\right)^{3/2} - \left(q\phi_t + q\phi - qE_{ox}d_{ox}\right)^{3/2}\right)\right]...(DT) \quad (2.55)$$

The emission coefficient can then be written as,

$$e_{TTB} = v_{TTB}\, T \quad (2.56)$$

Where, e_{TTB} is the frequency with which the particles collide with the barrier.

2.5.2.3 Temperature dependent electron discharge:

The time-dependent electron discharge through trap-to-band tunneling and thermal excitation can be described by a rate function for the occupancy function $f(x,\Phi_T ,t)$

$$df / dt = -f/ \zeta_{T-B} - f/ \zeta_{TE} \quad (3.57)$$

Solving Eq. 3.36 with the initial condition, $f(x,\Phi_T ,t)=1$,we have,

$$f (x,\Phi_T ,t) =\exp [- (1/\zeta_{T-B} + 1/\zeta_{TE})t] \quad (3.58)$$

The threshold voltage shift due to the trapped electrons can be written as

$$\Delta V_{TH} (t) = q \int \int D(x)\, g(\Phi_t)\, f(x,\Phi_T ,t)dx\, d\varphi_T , \quad (3.59)$$

Where

$$D(x) = (X_n - x)/\epsilon_n + X_{ob} /\epsilon_{ox} \quad (3.60)$$

ϵ_{ox} and ϵ_n are the permitivities of the nitride and oxide, respectively, $g (\Phi_T)/q$ is the trap density (eV^{-1} cm^{-3}).

Applying the following operator identity

$$\delta/\delta (\log t) \Xi \ln(10)t (\delta/\delta t) \approx 2.3t (\delta/\delta t)- \quad (3.61)$$

2.5 Characteristics

Characteristics shown in the following section represent typical present-day device characteristics of commercially available memory arrays.

2.6.1 Non-volatility:

SONOS is nonvolatile semiconductor memory device because it does not require supply voltage to maintain the data.

2.6.2 Erase/Write:

A SONOS device allows the insertion and removal of charge from nitride layer by application of electrical signals to the device terminals. As has been discussed before ,in the write mode, a positive gate to substrate bias is applied, which causes electrons to be injected by modified Fowler-Nordheim tunneling from the device channel, through the tunnel oxide layer, and into the traps in the Si_3N_4 layer. The net negative charge causes a positive shift in threshold voltage and places the n-channel SONOS transistor in the low-conduction state. In the erase mode, a negative gate to substrate bias is applied, which causes holes to be injected by direct tunneling from the device channel through the tunnel oxide layer and into traps located in the Si_3N_4 layer. The resulting net positive charge causes a negative shift in the threshold voltage and placed the n-channel SONOS transistor in the high-conduction state.

2.6.3 Excellent scalability:

Due to its charge trapping mechanism within an insulating nitride layer, SONOS is considered as most promising for vertical scaling on Flash memory. These traps are isolated with each other and, thereby, provide immunity to the deleterious effects of single pinhole defects. Vertical scaling corresponds to decreasing either the tunnel oxide, nitride, or blocking oxide thickness. Tunnel oxide can be made much thinner because the charge transfer mechanism doesn't required high voltage. Lateral scaling does not have much effect on SONOS device because charges are electrically trapped in the nitride so they don't interfere with adjacent cells. A typical gate stack in the SONOS memory consists of 2.7nm tunnel oxide, 5 nm charge trap nitride and 5 nm of control oxide, the EOT of the gate stack if about 10nm.

2.6.4 Low programming voltage

Low programming voltage of SONOS is due to scaling of nitride layer. Blocking action of top oxide not only inhibit gate injection, but also blocks the charges injected from the silicon at the top oxide-nitride interface, resulting in a higher trapping efficiency and

thus, reduces problem related to nitride layer reduction. In this way, the total thickness of the insulator structure can be reduced, and, consequently, the programming voltage can be reduced.

2.6.5 High density

SONOS devices are designed specifically for high-density EEPROMs operating at high temperatures due utilization of low power and low voltage. Low voltage operation, down to 5V, has been demonstrated for a nitride of 3 nm thickness and a blocking oxide thickness of 5.5 nm. Application of SONOS cell concept has allowed realization of memories with densities in the Mbit range.

2.6.6 Radiation hardness

SONOS use a layer of charge-trapping dielectric such as silicon nitride to store information that is not easily removed in a total ionizing dose environment. Solid-state devices may be susceptible to gamma, neutron, electron, proton, and alpha source radiation. SONOS devices are inherently hard because, unlike silicon dioxide, the nobilities of electrons and holes are not much different in nitrides. Thus, when exposed to ionizing radiation, both generated carriers can be swept out of nitride, resulting in a negligible amount of trapped charge. In the thin tunnel and blocking oxide layer, the density of electron hole pairs generated upon irradiation is quite small because of the small generation volume. Even if a few Si-H bonds are broken at the interface by electron hole pairs that have escaped recombination, the formation of a stable dangling bond (interface trap) requires that hydrogen, released in the process, diffuse away from the generation site. The low ion diffusivity of the nitride layer prevents the hydrogen, released as a byproduct of the interface reaction, from diffusing away from the interface. Thus, the hydrogen, which is confined to the thin oxide region, recombines with the dangling bond to regenerate the Si-H center, and a negligible change in interface trap density is measured upon irradiation of SONOS devices. Because of these properties, it is expected that further scaling of the SONOS device can only help render a more radiation-

hardened device. The reduction of the programming voltages and the absence of thick oxide in scaled SONOS devices have improved their radiation hardness.

2.6.7 Retention

Retention is defined as the time between the storage of data (i.e. nitride charge) and the time at which it can no longer be read out correctly. A good quality of scaling is that the retention characteristic may be improved as the SI_3N_4 memory storage layer thickness is reduced. Retention time may be estimated from backtunneling of trapped nitride charge, which is nitride thickness dependent and may be expressed as

$$\zeta=\zeta_0 \exp (\alpha_{ot} X_{ot}+ \alpha_n x) \quad (2.62)$$

where the backtunneling charge depends exponentially on X_{ot}, the tunnel oxide thickness. x= distance of the trapped charge in the silicon nitride from the nitride/tunnel oxide interface for 10 year storage.

ζ_0 = semiconductor conduction band to trap tunneling time constant (10^{-13}) sec.

$\alpha_{ot} = 1.07 \times 10^{10}$ m^{-1}

X_{ot} =tunnel oxide thickness.

In scaled SONOS structures, the tunneling of trapped nitride charge to the gate electrode is also possible through the blocking oxide.

2.6.8 Compatible with CMOS

The SONOS fabrication process is based on standard polysilicon gate technology and is CMOS compatible. The ONO triple dielectric film of the SONOS device has been adapted into submicron CMOS processes for the manufacturing of high-density DRAMs and SRAMs because of increased reliability and yield realized by the inherent ability of the top (blocking) oxide film process to fill any underlying pinholes in the nitride. When fabricated on normal CMOS production lines, the SNOS-type EEPROM technology yield

is higher than that of SRAM. This high yield is attributed to the low defect density and ion-barrier properties of the Si_3N_4 film, as well as to the insensitivity of the MNOS EEPROM's device characteristics to SiO_2 failures.

2.6.9 Increased endurance.

Endurance of a SONOS device is a measure of the device's ability to meet a specified retention time as a function of accumulated erase/write cycles. Flash products are specified for 10^6 erase/program cycles. Reduction of the programmed threshold with cycling is due to following two effects;

a. Deterioration of the Si-SiO_2 interface in the form of the creation of interface states (traps) with cycling, and
b. Deterioration of the bulk Si_3N_4 layer trap density, which manifests itself in the form of increased charge centroid penetration into the nitride film as a function of write/erase cycling. Both mechanisms are linked to the breaking of bonds, which are believed to be Si-H bonds.

2.7 Scaling Issues

(I) Since the erase or write voltage pulse is not applied between the source and drain of the SONOS, a shorter channel length can be used than is possible in conventional floating gate type memory. Relationship of SONOS scaling method is given as:

$$X_N \propto V_p \propto X_{ot} \tag{2.63}$$

and

$$\zeta_e \propto \exp(X_{ot} / \lambda_{ox}) \tag{2.64}$$

where, V_P is the programming pulse voltage amplitude

X_N is the nitride film thickness

ζ_e is the erase time

λ_{ox} is the de Broglie wavelength in the oxide (approximately 0.1nm)

As can be seen in these equations, the programming voltage is reduced by decreasing the nitride thickness, and the erase time is shortened by decreasing the tunnel oxide thickness. The literature suggests that there is a larger penetration length in nitrides for holes (typically 15 to 20 nm) than for electrons (typically 5 to 10 nm).In reducing the nitride thickness, the holes, will, therefore, be trapped closer to the gate. The thin blocking oxide of 2.5 nm may not be thick enough to prevent hole loss from the nitride to the gate.

(II) Second scaling approach describes a thinner nitride thickness (< 10nm) and a thicker top oxide (>3 nm).

The goal is to inhibit gate injection as well as to stop charges injected form the silicon at the nitride-blocking oxide interface, resulting in a higher trapping efficiency, thus preserving the memory window as the nitride thickness is reduced.

Chapter 3

Electrical Characteristics of SONOS Memory Cells

3.1 Gate length effect on threshold voltage shift

Figure 3.1 shows threshold voltage shift versus gate length graph. Longer channel length has more trap charge density than shorter channel [4]. Hence, more electrons are captured to these trapped sites resulting in higher threshold voltage shift in positive direction.
The growth of interface trap charge results in positive threshold voltage shift given by

$$\Delta V_{th} = \Delta Q_{it}/C_{OX} \qquad (3.1)$$

C_{ox} = gate capacitance per unit area
Q_{it} = interface trap charge or electrons.

Device Parameter:

Toxb = 1.8nm, Tsn = 8nm, Toxt = 4 nm
Program time = 2.5 ms, Erase Time = 7.5 ms
Programming voltage = 9V, Erase Voltage =-9V

Lg	Vt
70	1.746
90	2.081
130	2.334
150	2.429
210	2.583
300	2.666

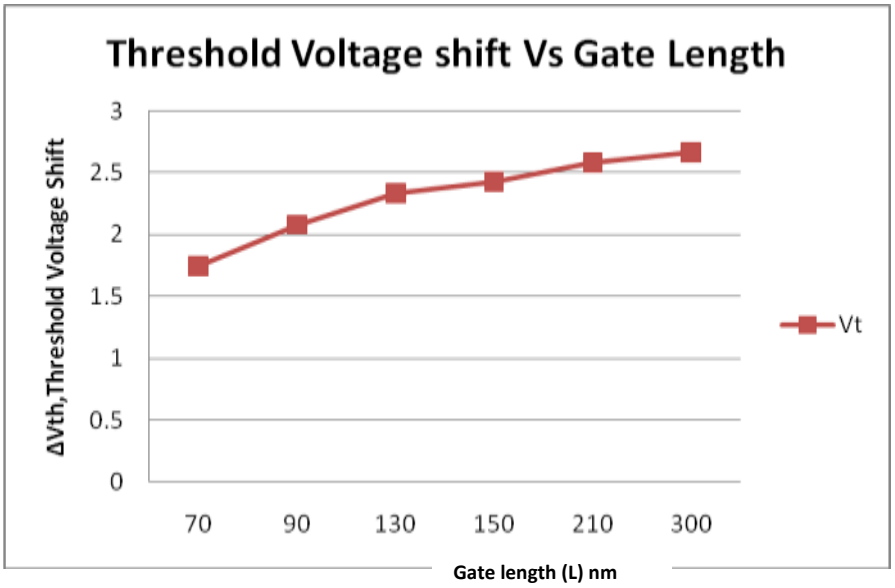


Figure 3.1 Threshold voltage dependence on gate length

In figure 3.2, we see that as gate length increases, threshold voltage shift decreases with increasing programming time. Short channel effects (DIBL, channel length modulation, source drain charge sharing) are more prone in short channel device than Long channel. So change in threshold voltage with program time will be higher in short channel compared to long channel devices.

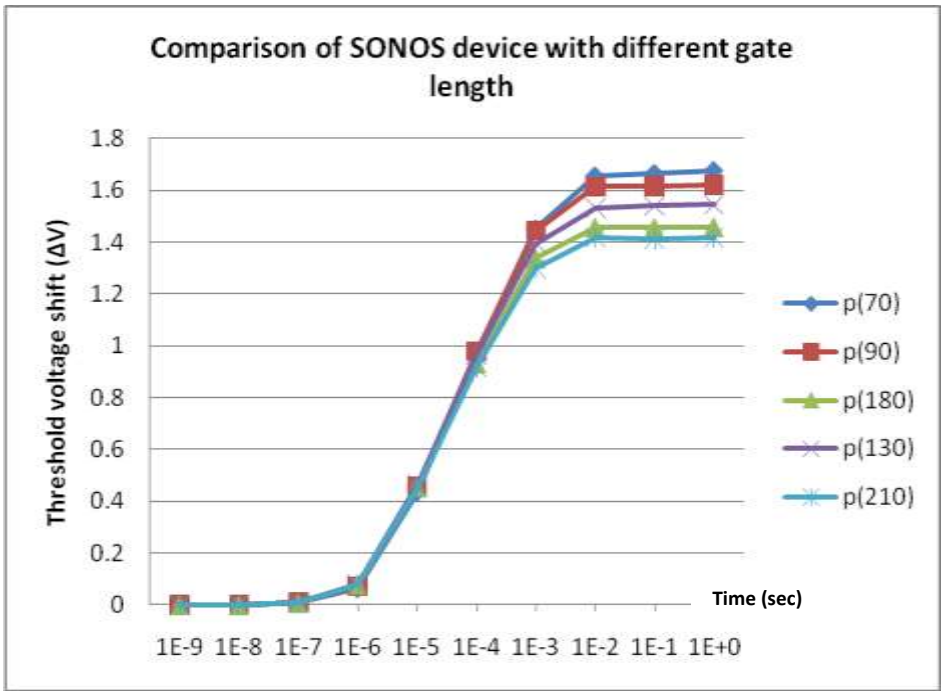


Figure 3.2 Program characteristic of SONOS device at different Gate Lengths.

3.2 Gate voltage effect on Threshold voltage shift

Threshold voltage is given as,

$$V_{th} = V_{FB} + \{[2\epsilon_S\epsilon_0 q N_A (2\Psi_B)]^{1/2} / C_{ONO} \} + 2\Psi_B \quad (3.2)$$

The threshold voltage V_{th} and V_{FB} differ by a constant term only, ϵ_s silicon dielectric constant, N_A doping concentration in the semiconductor substrate. Ψ_B is the energy difference between Fermi level and intrinsic Fermi level in the substrate, and C_{ONO} is the capacitance of the ONO multilayer. Thus for a given V_{th} the electric field in the bottom oxide E_{bottom} is given as,

$$E_{bottom} = (V_G - V_{FB} - \Psi_S)/d_{eq} \quad (3.3)$$

$$d_{eq} = d_{Bottom} + d_{Topox} + (\epsilon_{ox} / \epsilon_N) d_N \quad (3.4)$$

Device Parameter

Lg=130nm
Device :Toxb=2nm,Tsn=4nm,Toxt=5nm

Program pulse = 2.5 ms, erase pulse = 7.5 ms

Where Toxb=tunnel Oxide

Tsn = Nitride Layer

Toxt = Blocking Oxide

Figure 3.3 and figure 3.4 exhibit the gate voltage dependence of threshold voltage shift.

3.2.1 Program

Vp	ΔVt
8	0.611
9	0.727
10	0.789
11	0.798
12	0.799
13	0.799

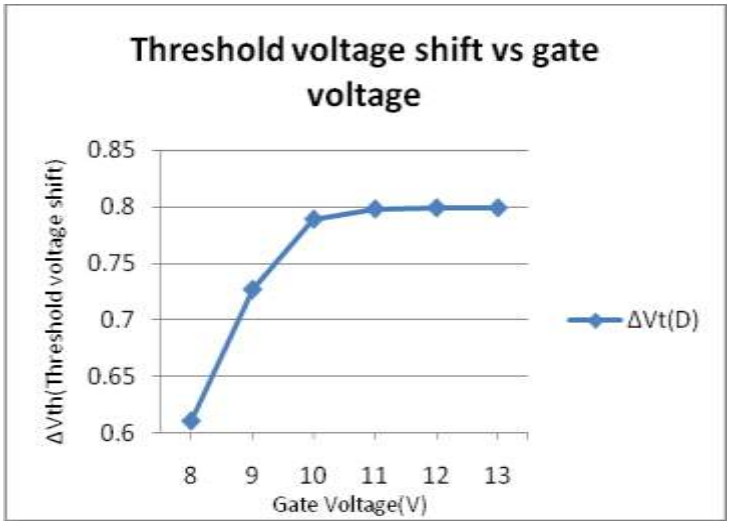


Figure 3.3 Threshold voltage dependence on positive gate voltage

For various gate voltages, current density differs because the total nitride charge and charge distributions are different. Threshold voltage increases with increasing gate voltage. Due to increasing gate voltage, tunneling electron increase in nitride from substrate,resulting in increased negative charge in the nitride. Due to this the electric field increases in blocking oxide and decreases in tunnel oxide. Thus electrons tunnel from substrate into the nitride. Also, when we apply positive gate voltage, electrons are trapped in the nitride layer. To bring more electrons to nitride layer, first the trapped electrons has to be neutralized, which require more gate voltage, resulting in increase of threshold voltage. Increase of electrons traps sites results in positive threshold voltage shift given by

$$\Delta V_{th} = \Delta Q_{it}/C_{ox} \qquad (3.5)$$

where C_{ox} is gate capacitance per unit area, Q_{it} = interface trap charge.

3.2.2 Erase

Ve	ΔVt
-8	-0.361
-9	-1.111
-10	-1.199
-11	-1.199
-12	-1.199
-13	-1.199

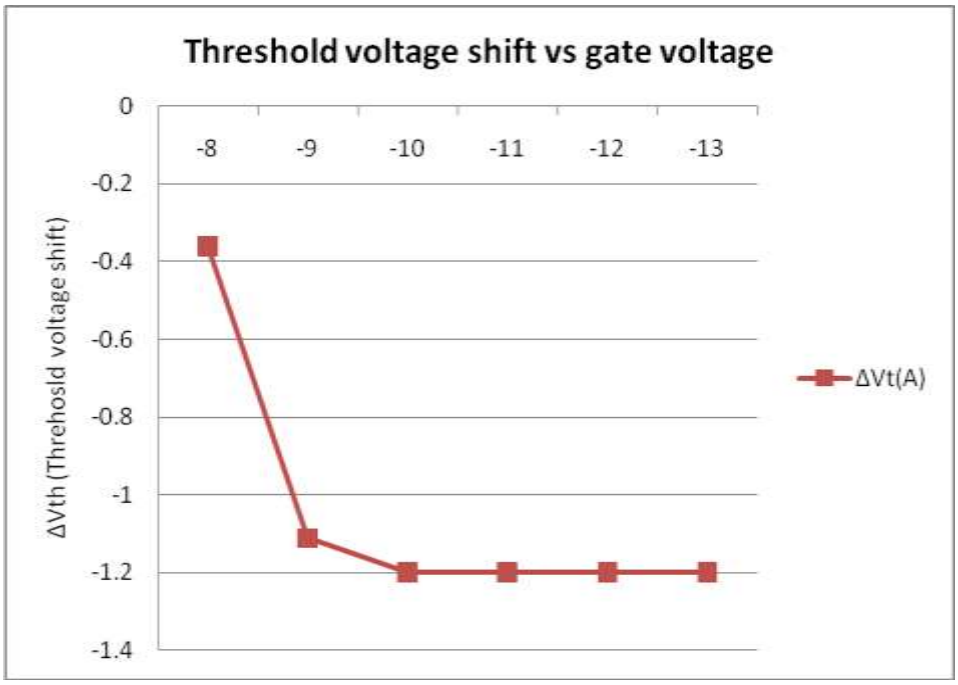


Figure 3.4 Threshold voltage dependence on negative gate voltage

Threshold voltage decreases with increasing negative gate voltage. This is due to tunnel back of electron from nitride to substrate and also due to tunneling of holes into nitride from substrate. Both the process occurs simultaneously resulting in increase in electric field in tunnel oxide and decrease in electric field in blocking oxide. Increase of holes in nitride layer results in negative threshold voltage shift given by,

$$\Delta V_{th} = -\Delta Q_{ot}/C_{ox}, \qquad (3.6)$$

where Q_{ot} is oxide trap charge and C_{ox} is gate capacitance per unit area.

Figure 3.5 shows the write characteristic at various voltages for 130 nm gate length and ONO stack as 2/8/5 nm respectively. Higher programming voltage results in higher tunneling probabilities due to reduction of effective barrier. This is observed as exponential decrease in writing time with increased applied voltages. There are small

variations and some irregularities at some of the voltage that may be related to the specific tunneling mechanisms involved in ONO trapping in very short devices.

Saturation in curve of Figure 3.5 is due to different causes. One possibility is that since larger voltages do not seem to uncover more traps for storage, there exist a finite number of available traps in the energy and thickness range accessible by the voltage applied. Another possibility is that, after a certain number of electrons have been injected, more traps becomes available by the use of higher voltages and are occupied, forming conduction paths that lead to the leakage of any further injected charge. This leakage of charge through defects results in a maximum total charge that can be stored and causes the saturation of the threshold voltage shift. Nearly, all tunneling-based injection processes have a characteristic bias/energy dependence that is inversely exponentially related to the thickness of barrier height. This occurs because the wave function is evanescent and, therefore, follows an exponential tail. A larger barrier, a larger thickness or a lower field all makes transmission smaller. Consider for example a triangular barrier, the transmission rate is

$$T_t = C \exp [\{- 8\pi (2m^*)^{1/2} t_{barrier} (Q \varphi_B)^{3/2} \}/3hqV] \quad (3.7)$$

Where m^* = effective mass of electron

$t_{barrier}$ = barrier height

V = applied bias

h = Planck's constant

q = electronic charge

The pre-factor C also has electric-field dependence. A larger bias, smaller thickness, smaller barrier height, or small mass all increases the transmission rate.

Generalized relationship is given as

$$T_t = C \exp (-E_{ch} /qV) \quad (3.8)$$

Where C is reduced to localization and reduced coupling of the defect and has a field dependence, but E_{ch} still has the effect of the barrier. Hence capture and emission time (t) for a change in threshold voltage of ΔV_t can be written as

$$t = \beta \Delta V_t \exp (E_{ch} /qV) \quad (3.9)$$

Where, t is derived through integration of injected charge trap proportional to the current density. A large V_T shift requires a large charge injection or removal, and takes a longer time at constant voltage. A larger voltage however increases the transmission rate reducing the time required to achieve the same threshold voltage shift.

β = coefficient which includes the effect of capture cross-section

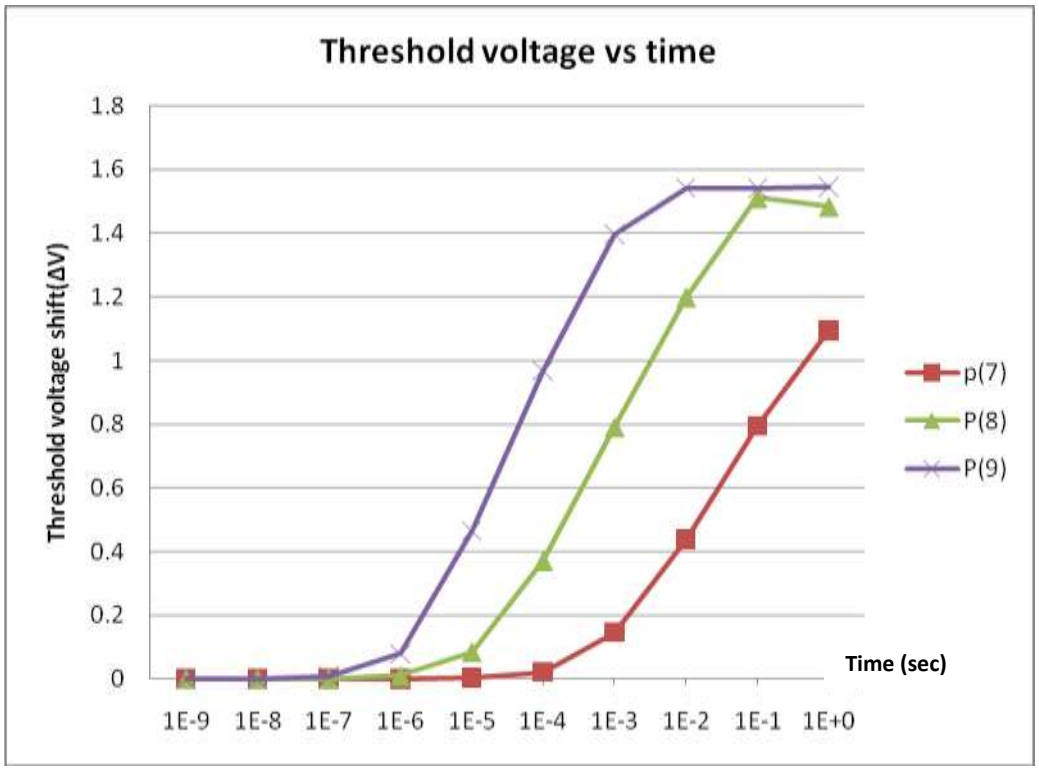


Figure 3.5 Write characteristic at different gate voltage.

3.3 Temperature effect on threshold voltage

Temperature effect [31] on threshold voltage can be explained by Poole-Frenkel effect. This effect is based on the emission of an electron from a charged trap (positive when empty) towards the conduction band as shown in figure below. The corresponding current depends on electric field. This field lowers the potential barrier resulting from the attraction of the charged trap and the emitted electron.

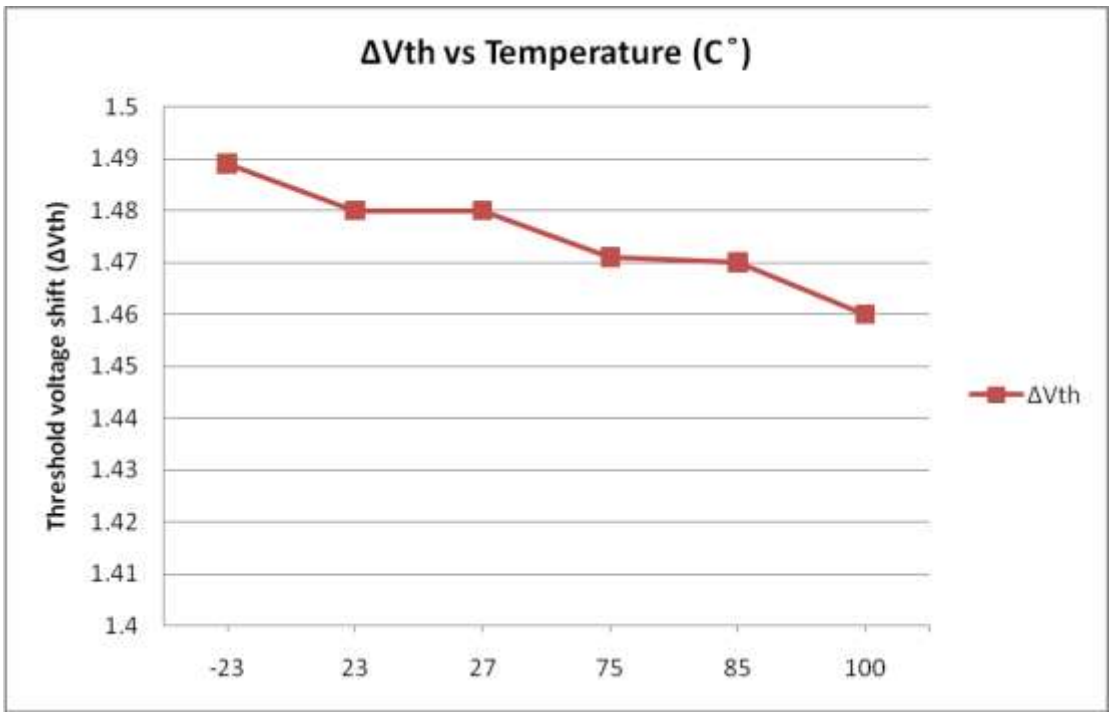


Figure 3.6 Temperature effect on threshold voltage shift

The nitride current (J_{N1} thermal current) is given by

$$J_{N1} = C_1 \xi_N \exp(1/kT) [\{(q^3 \xi_N)/(\pi \epsilon^*_N)\}^{0.5} - \Phi_t] \quad (3.10)$$

$$= C_1 \xi_N \exp [(\beta \xi_N)^{0.5} - \Phi_t/kT]$$

Where ϵ^*_N is the high frequency dielectric constant of the nitride.

As temperature is increased, current density of electron in nitride increases resulting in more tunneling of electrons from nitride to substrate and hence less threshold voltage at

high temperature. Also, this is attributed the fact that rise in temperature results in an increase in the number of free carriers which leads to the channel formation at lower gate voltage. Threshold voltage decay behavior of a scaled SONOS device for and extended time after electrons are injected is given as,

$$\frac{\delta \Delta V_{TH}(t)}{\delta \log t} \approx -2.3\, k_B\, T_g\, (\Phi^*_T) M(x) - 2.3q \int \frac{D(x^*)}{\alpha_n}\, g(\Phi_T)\, d\varphi_T \qquad (3.11)$$

Above equation indicated that the threshold voltage decay rate is a function of the logarithm of the retention time, and is affected by the properties of the ONO dielectrics and the temperature of the environment. The individual contribution of thermal excitation and trap-to-band tunneling to the threshold voltage decay is explained through above equation. The first term in equation is from the contribution of thermal excitation, which is proportional to the environmental temperature. The second term is from the contribution of trap-to-band tunneling, which depends slightly on the temperature through Φ^*_T.This small temperature dependency is due to the fact that thermal excitation reduces the number of trapped electrons available for back tunneling. Figure 3.7 shows how the two discharge processes influences the electron retention loss in SONOS devices. At retention time‘t’ the charge storage nitride contains both empty traps and filled traps that are distributed in both space and energy. As the discharge process continues, the “empty-filled” boundary marked by x^* moves toward the right (the blocking oxide).Meanwhile, the boundary marked by $q\Phi^*_T$ moves toward the bottom (the nitride valence band)-a process enhanced at elevated temperatures. The two discharge processes compete with each other and thermal excitation starts to dominate the total retention loss as temperature increases.

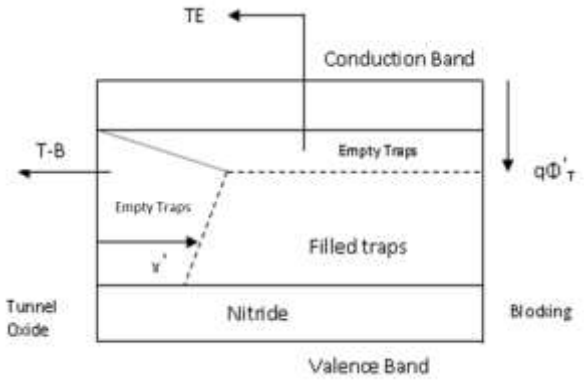


Figure 3.7.Contributions of thermal excitation and trap-to-band tunneling to electron discharge in a SONOS device.

Device Parameter:

Lg = 130 nm

Tunnel oxide (Toxb) = 1.8nm

Silicon Nitride (Tsn) = 8nm

Blocking Oxide (Toxt) = 4.5 nm

Programming Voltage = 9 V, Erase voltage = 9V

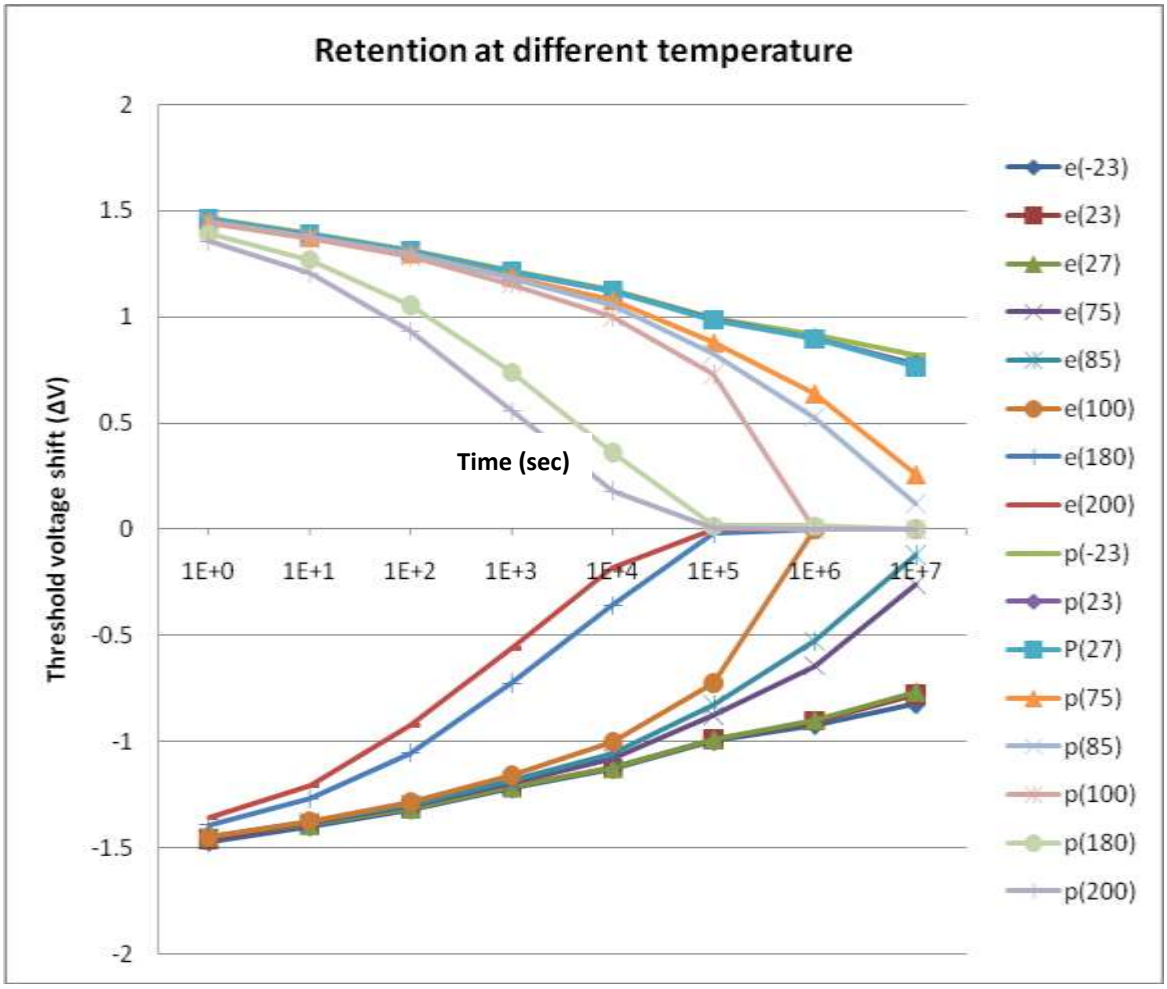


Figure 3.8 Retention characteristic of SONOS transistor at elevated temperature

The simulated retention characteristics of an n-type SONOS device after program and erase at temperatures of -23, 23,27,75,85,100 ^{0}C are shown in figure 3.8

Temperature affects device parameters and performance, in particular mobility, threshold voltage, and subthreshold characteristics. To derive temperature dependence of the threshold voltage, expression is given as,

$$V_T = \Phi_{ms} - Q_f/C_{ox} + 2\,\Psi_B + [\,(4\epsilon_s q N_A\,\Psi_B)^{1/2}/C_{ox}] \qquad (3.12)$$

Because the work-function difference Φ_{ms} and the fixed oxide charges are essentially independent of temperature, differentiating equation above with respect to temperature yields,

$$dV_T / dT = d\psi_B/dT \, [2 + (1/C_{ox}) (\epsilon_s q N_A/\Psi_B)] \quad (3.13)$$

From the basic equations of

$$\Psi_B = (kT/q)\ln (N_A /n_i) \quad (3.14)$$

$$n_i^2 \propto T^3 \exp (-E_{g0}/kT) \quad (3.15)$$

where E_{g0} is the energy gap at T=0 .

Hence,

$$d\psi_B /dt \approx 1/T \, (\Psi_B - E_{g0}/2q) \quad (3.16)$$

As temperature decreases the Id-Vg characteristic improves as shown in transfer characteristics of figure 3.9 with temperature as parameter. Note that as temperature decreases the threshold voltage V_T increases. Most important improvement is the reduction of the subthreshold swing S, as temperature is decreased. This improvement comes mainly from the kT/q term in equation given as

$$S = \ln (10) (kT/q) [(C_{OX} + C_D)/C_{OX}] \quad (3.17)$$

C_{OX} = oxide capacitance

C_D = depletion-layer capacitance.

Other improvement at low temperature is higher mobility, thus, higher current and transconductance, lower power consumption, lower junction leakage current.

Device parameter

Lg= 130 nm

Toxt (tunnel oxide) = 1.8nm,

Tsn (Silicon Nitride) = 8nm,

Toxb (Blocking oxide) = 4.5 nm

Vg=9V, Ve = -9 V

dtp = 1ms,dte = 1ms

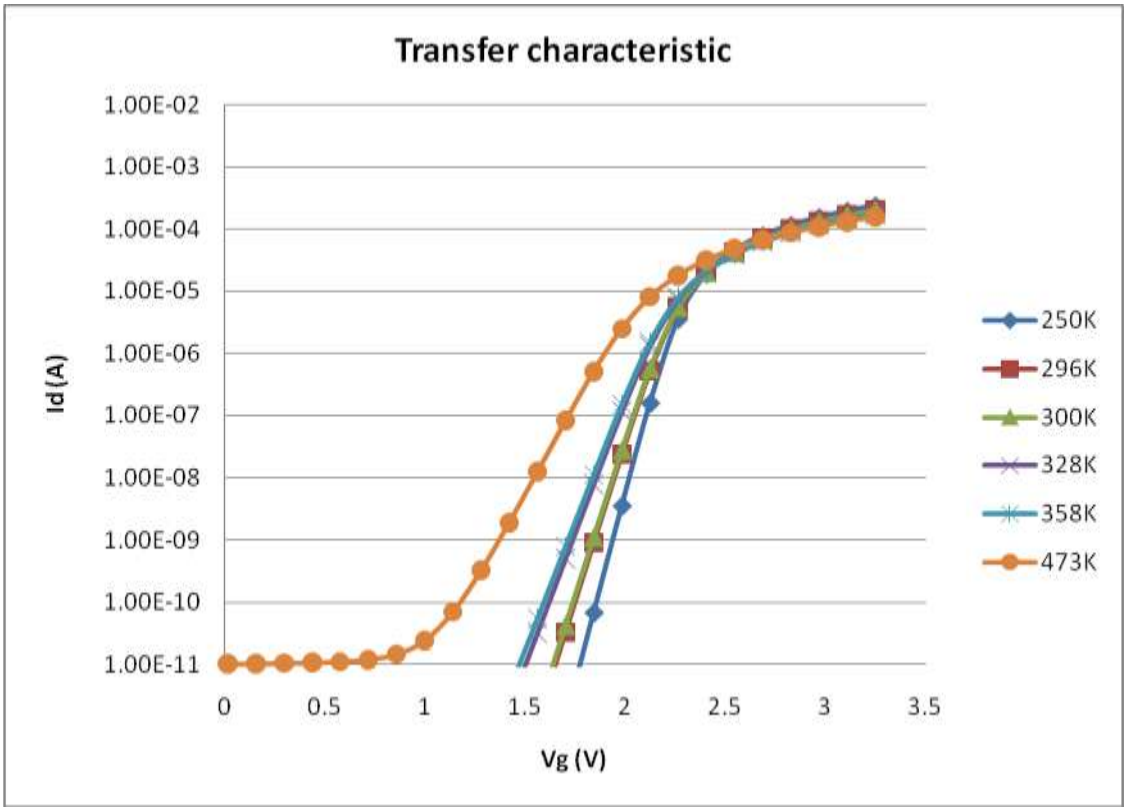


Figure 3.9: Subthreshold characteristic of SONOS with temperature as a parameter.

3.4 Tunnel oxide thickness effect on Threshold voltage shift

Parameters of Device

V_p = 9V, V_e = -8V

dtp= 2.5 ms ,dte = 7.5 ms

Lg = 130nm

Toxb (tunnel oxide) (nm)	Tsn(Nitride) (nm)	Toxt(Blocking oxide) (nm)
0.5	4	4
1.0	4	4
1.5	4	4
2.0	4	4
2.5	4	4
3.0	4	4
3.5	4	4

ΔV_{th} Vs Programming Voltage

Vp(V)	**ΔVth (0.5nm)**	**ΔVth (1nm)**	**ΔVth (1.5nm)**	**ΔVth (2nm)**	**ΔVth (2.5nm)**	**ΔVth (3nm)**	**ΔVth (3.5nm)**
7	-0.351	Failed	0.096	0.363	0.594	0.714	0.283
8	-0.187	0.072	0.271	0.471	0.614	0.716	0.45
9	-0.052	0.209	0.397	0.567	0.612	0.715	0.715
10	0.075	0.263	0.428	0.6	0.59	0.708	0.713
11	0.076	0.263	0.428	0.6	0.567	0.694	0.703
12	0.076	0.263	0.428	0.6	0.555	0.682	0.691

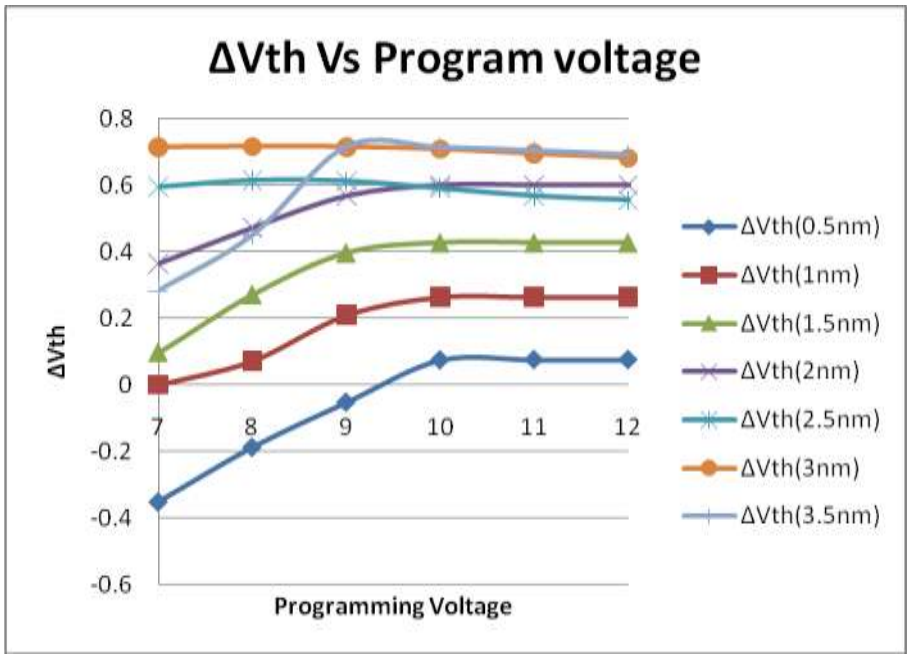


Figure 3.10 Effect of different tunnel oxide on threshold voltage shift.

The samples with thinner tunneling oxide have less threshold voltage shift (for programming voltage) can be explained from equation given below [30]

$$V_{TH}(V_p,t_p) = \Phi_{GS} + 2\Phi_F + [(4\epsilon_{si}qN_B\Phi_F)^{0.5}/C_{eff}] - Q_N(V_p, t_p)[X_{OB}/\epsilon_{OX} + X_N/2\epsilon_n] \quad (3.18)$$

Where,

Φ_{GS} = gate to semiconductor work function

Φ_F = bulk Fermi potential

N_B = bulk doping

ϵ_{ox} , ϵ_n , ϵ_{Si} = dielectric constant for an oxide, nitride and silicon respectively

C_{eff} = effective capacitance of the triple dielectric.

X_{OT} , X_N, X_{OB} = tunnel oxide, nitride oxide, blocking oxide.

The charge stored in the nitride layer Q_N depends on the programming voltage V_p and programming time t_p .In these graph programming voltage is 9V and Programming pulse is 2.5ms.Changing the tunnel oxide thickness effects the 3rd term of equation (3.15).

C_{eff} (capacitance per unit area) = $Є/x_{eff}$ (3.19)

$X_{eff} = X_{oT} + (Є_{ox}/Є_n)*X_n + X_{OB}$ (3.20)

Increasing the tunnel oxide results in increase of X_{eff} and hence decrease of C_{eff} resulting in increase of $V_{th}(V_p,t_p)$.

Based on data, conclusion is a reduction in the tunnel oxide thickness results in decreased threshold voltage shift.

Id-Vg curve of different tunnel oxide

Graph below shows that ultra-thin tunnel oxide conduct high current in comparison to thick tunnel oxide at program voltage of 9V resulting in low threshold voltage shift for ultra-thin tunnel oxide.

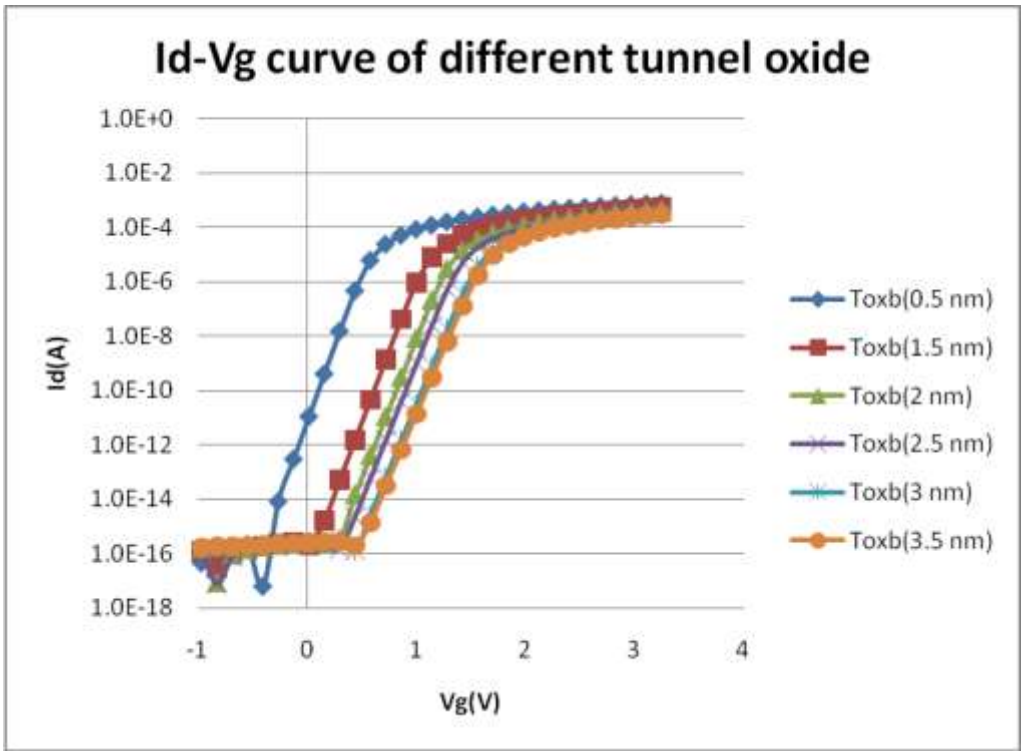


Figure 3.11 .Id-Vg curve of different tunnel oxide

Figure 3.6 shows the behavior of threshold voltage shift at different values of oxide thickness, t_{ox}. It is observed that for all values of t_{ox}, drain current increases rapidly with increasing gate voltage which is due to the gate-induced barrier lowering effects. It is observed that as gate-oxide thickness decreases drain current increases which may be due to the lowering of trap sites at the oxide-nitride interface. Doping concentration is given as

$$N = C_{eff} (V_G - V_T)/q\, x_{eff} \quad (3.21)$$

Where N =Carrier concentration in the strong inversion channel.
Barrier potential at N/TOX is given as

$$\Psi_B = q n_t^2 / 8N\epsilon_s \quad (3.22)$$

Where n_t is the trap density in Si_3N_4.Hence, after substituting the value of N, Ψ_B become

$$\Psi_B = (q^2 n^2_t\, x_{eff})/ [8\epsilon_s\, C_{ox}(V_G - V_T)] \quad (3.23)$$

From above equation it is clear that as N increases, barrier potential will decrease. This will make available carriers for transport increase hence decreasing the trap sites in nitrides, which in turn will increase drain current.

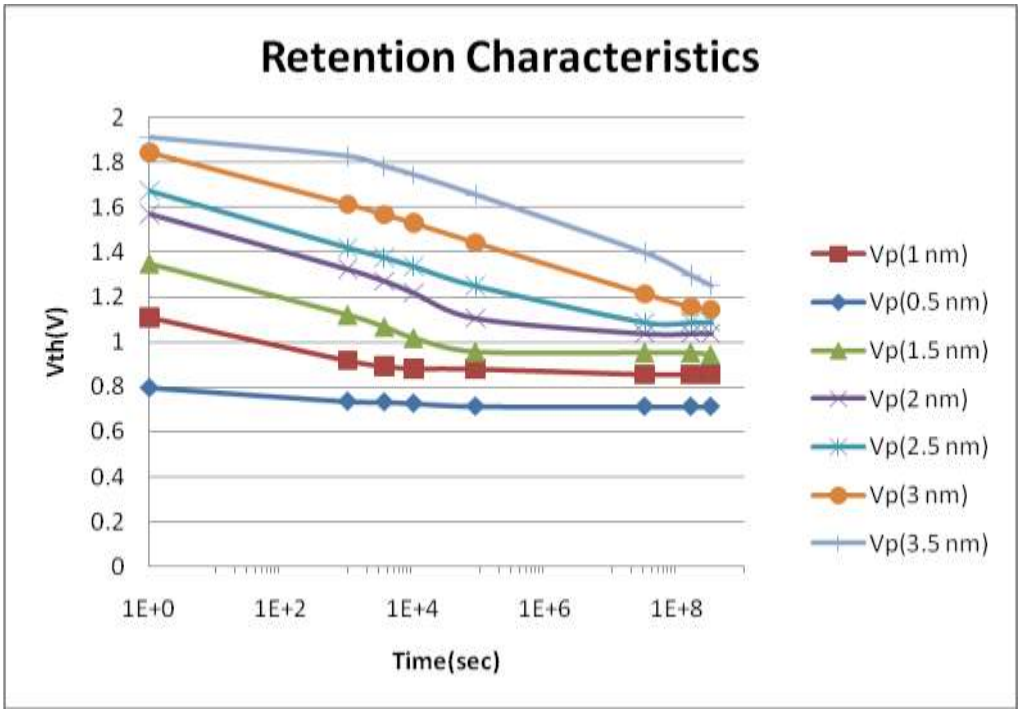


Figure 3.12 Retention characteristic of different tunnel oxide

Above retention characteristic shows that saturation for thinner oxide appears early in program state. This graph has been taken for 5 program cycles.

The threshold voltage shift is more positive for the thicker tunnel oxides, which indicates as reduction in electron backtunneling (i.e.an increase in retention).If the tunnel oxide becomes thinner, the P/E operation speed becomes faster but retention properties are degraded. Tunnel Oxide in the SONOS device structures is one of the most important factors that directly affect the device performances because hole/electron injection occurs through the tunnel oxide during P/E device operations. As the tunnel oxide gets thicker, the P/E operation speed becomes slower, but retention properties are improved. On the other hand if the tunnel oxide becomes thinner, the P/E operation speed becomes faster but retention properties are degraded. There it is hard to control the thickness of tunnel oxide to improve both P/E operations speed and charge retention characteristics. A good criterion of the tunnel oxide is to achieve low leakage current, enough breakdown

strength and long-term reliability. So, it is necessary to understand the effects of the silicon dioxide as tunnel oxide on memory characteristics.

Tunnel oxide effect on Threshold voltage shift

An increase in program speed has been observed by reducing tunneling oxide thickness. Simulation of different tunnel oxide thickness was done to observe the effect of threshold voltage shift. Programming speed increases with decreasing tunnel oxide. But this degrades the charge retention characteristic of device. Thus we cannot decrease the tunnel oxide thickness for increasing speed.

Device parameters:

Gate length = 130nm
Blocking oxide thickness = 4nm
Nitride Thickness = 10nm
Programming voltage = 9V
Erase voltage= -9 V

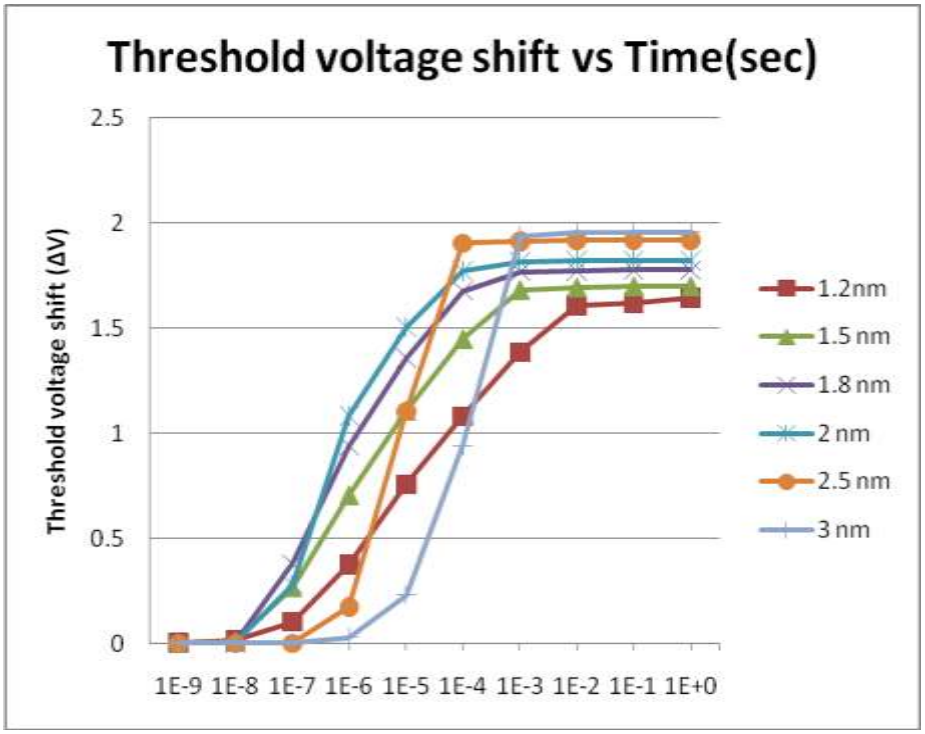


Figure 3.13: Programming speed characteristic of SONOS device with various thickness of tunnel oxide.

3.5 Effect of Nitride Thickness on threshold voltage shift

The charge trap characteristic of SONOS is shown in figure 3.13. At 2 and 2.5 nm thickness of nitride layer, the memory windows are 0.088 and 0.135 V respectively. The memory window for 3.5, 4.5 and 5.5 nm is 0.394, 0.641 and 0.854 V respectively. Most of the threshold voltage shift is caused by stored electrons in the bulk of the nitride layer rather than the interface. There is little charge trapping in Si_3N_4 layer less than 2.5 nm as less voltage shift was observed for nitride layer less than 2.5 nm. Thicker nitride gives larger threshold voltage shifts, partly due to dept in the nitride over which trapping occurs.

Device parameter

Lg = 130nm,
Blocking oxide = 4nm,
Program voltage = 9V, Erase voltage = -9V
Silicon nitride thicknesses are 2nm, 2.5nm, 3.5nm, 4.5nm, 5.5nm, 6nm, 7nm and 8nm.

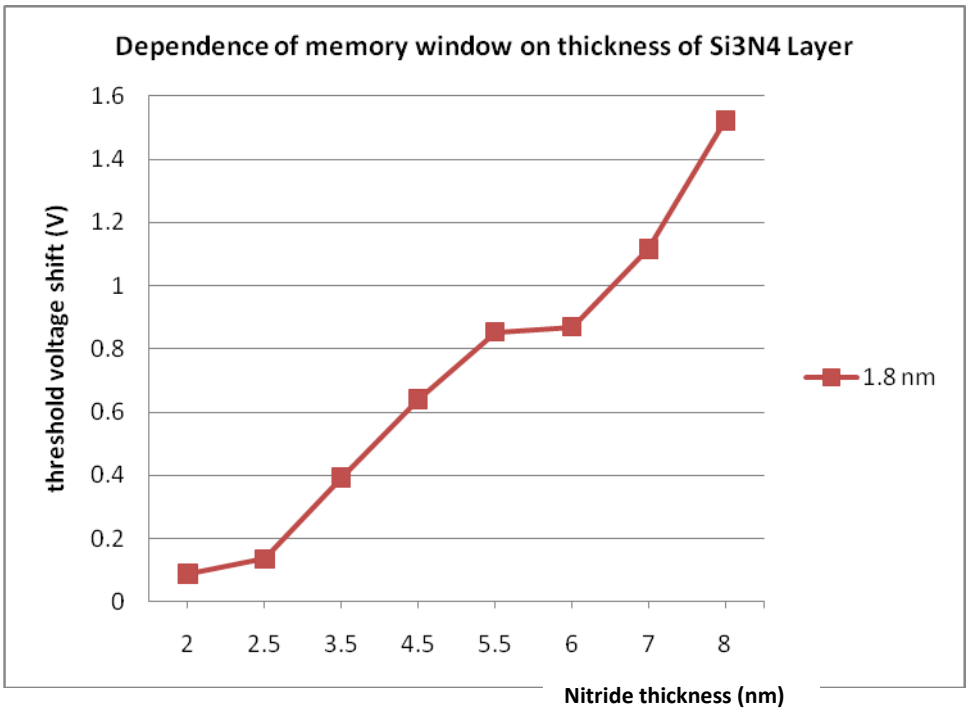


Figure 3.14: Dependence of memory window on Si_3N_4 layer

Especially no shift was observed below 2.5 nm. To understand the dependence of charge trapping and tunneling characteristics on the thickness of Si_3N_4 layer, the charge trap centroid and charge trap density should be considered.

The charge trap centroid is given by

$$X_{CENT} = t_{stack} / [1-(\Delta V^{-}_{g} / \Delta V^{+}_{g})] \quad (3.24)$$

$$C_{trap} = (\epsilon_0 \epsilon_{stack}/t_{stack}) \times [1-(\Delta V^{-}_{g} / \Delta V^{+}_{g})] \quad (3.25)$$

X_{CENT} = is measured from the gate/oxide interface.

ΔV^{-}_{g} and ΔV^{+}_{g} are negative and positive gate voltage shifts. This is caused by charge trapping in Si_3 N_4 layer. The charge trap centroid (X_{CENT}) shifts to the gate/oxide interface as the thickness of the nitride layer decreases.Therefore,the dominant charge trap changes from the bulk traps in Si_3N_4 layer to the interface states at Si_3N_4 /SiO_2 interface. Also, the trapping efficiency defines as the ratio of trapped charges to total injected charges (C_{trap} / C_{total}) considerably decreases as there is decrease in nitride layer thickness. Change in C_{trap} /C_{total} in thicker Si_3N_4 layer is small for thicker nitride layer. Therefore, it can be concluded that a rapid increase in tunneling current and a decrease in memory windows for Si_3N_4 layer thinner than 2.5 nm is due to reduction in charge trapping characteristics. Hence, for tunnel barrier application of NVM, the thickness of Si_3N_4 layer should be thinner than 2.5 nm and for charge storage application, the Si_3N_4 layers thicker than 3.5 nm are necessary. Figure 3.14 shows the charge trapping characteristics of SONOS with various thickness of Si_3N_4 layer. The voltage shift is largest at 8nm Si_3N_4 and decreased with decreasing thickness of Si_3N_4. To examine the dependence on the silicon nitride thickness, the SONOS device with various nitride thicknesses were simulated. Figure 3.14 shows the threshold voltage shift as a function of programming time. SONOS device with thick nitride layer has larger threshold shift than thin nitride layer at longer program time because of the presence of more bulk trap-sites as carrier storage node in the thick silicon nitride layer resulting in higher carrier capture probability. At smaller program time SONOS with thick nitride layer has less threshold voltage shift due to lower electron driving force. But as programming time is increased, threshold voltage shift are larger due to higher electron capture probability.

Tunnel oxide = 1.8nm

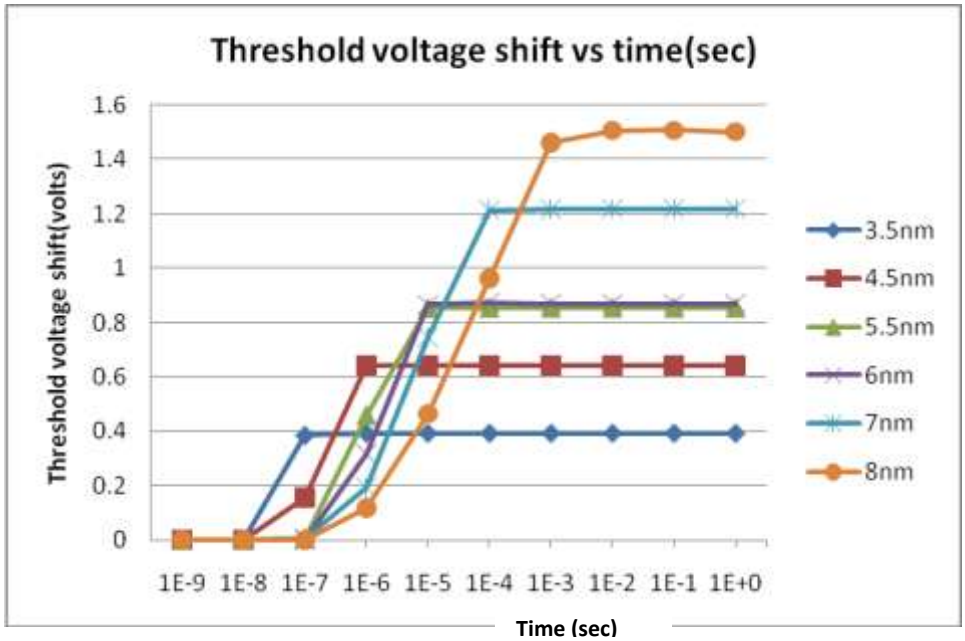


Figure3.15 (a)

Tunnel Oxide = 2nm

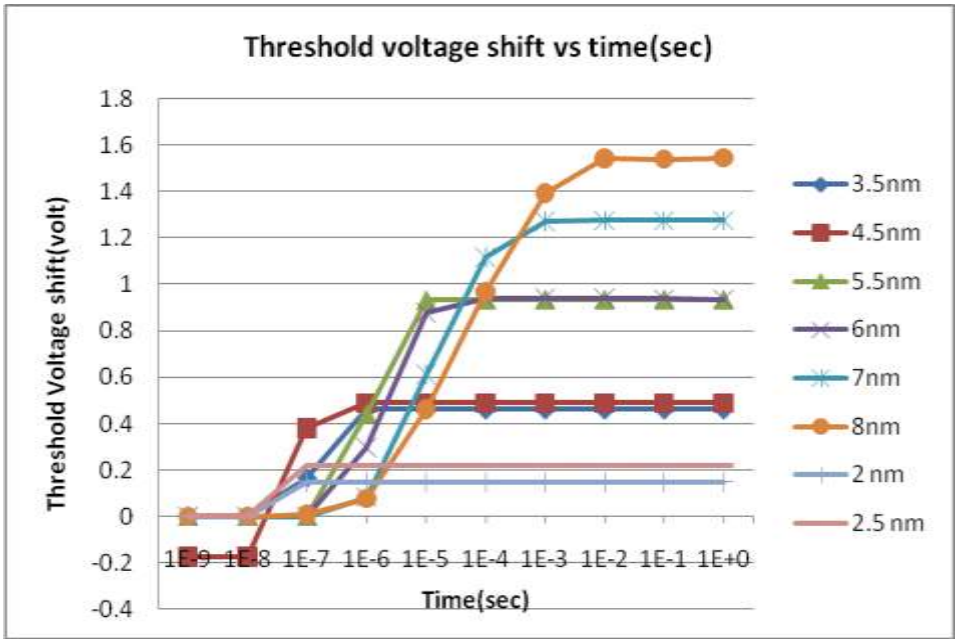


Figure 3.15 (b)

Tunnel Oxide = 2.5nm

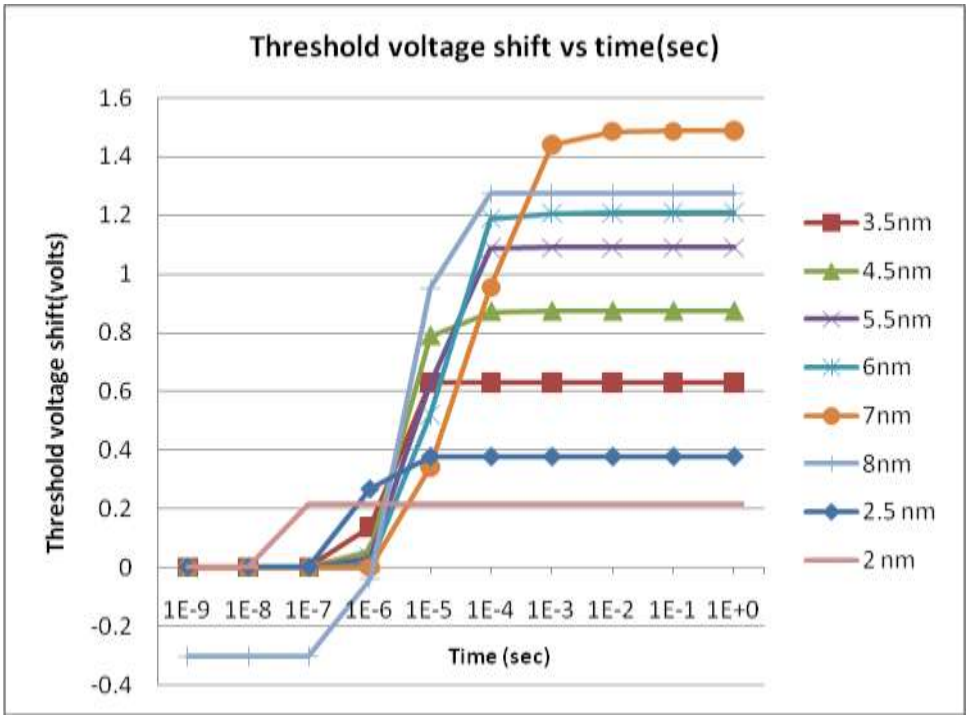


Figure 3.15 (c)

Tunnel Oxide = 3nm

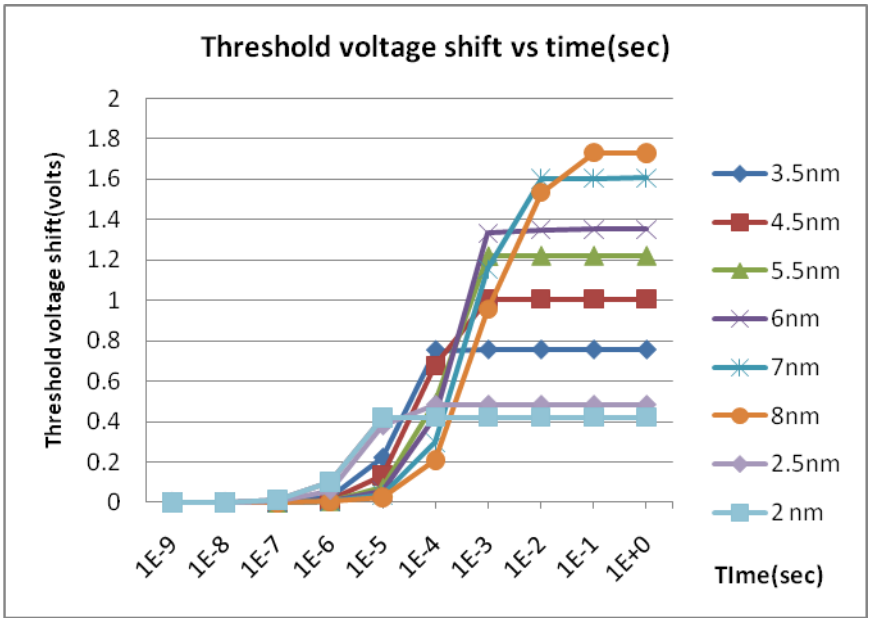


Figure 3.15 (d)

Tunnel Oxide = 4nm

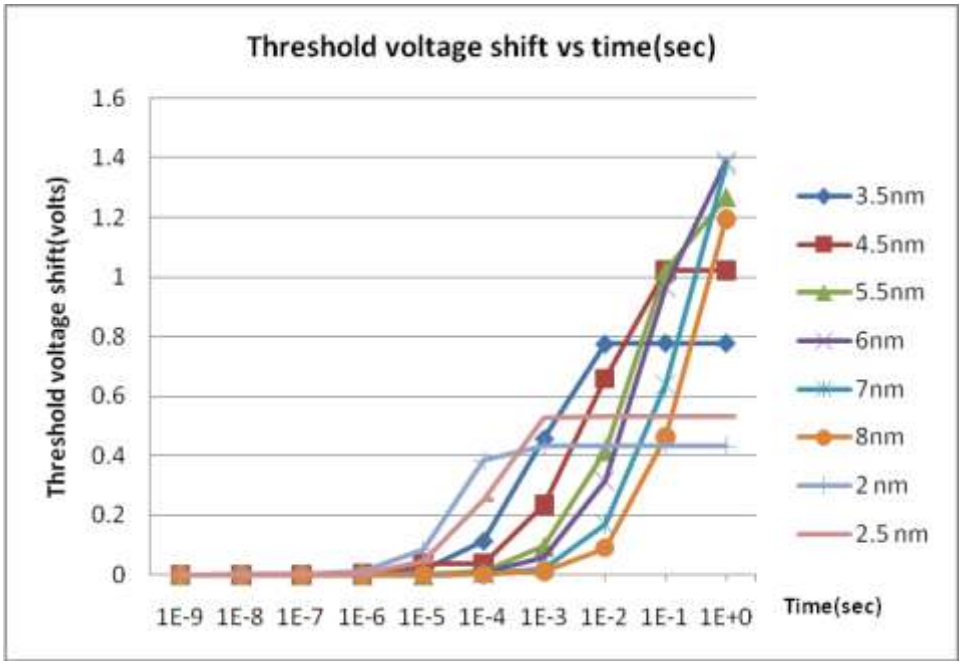


Figure 3.15 (e)

Figure 3.15: Charge trapping characteristics of SONOS with various thickness of Nitride.

In summary, an increase in tunneling current and no charge trapping characteristics under 2.5 nm is due to shift in charge trap centroid close to N/TO interface. The tunnel current rapidly increase and memory window decreases for thin Si_3N_4 layer.

3.6.2 Retention characteristic of a SONOS device at different nitride thickness

The fact that threshold voltage is narrowed over time means that charge leakage processes occurs in both states. In programmed state negative charge leaked away from the traps. The threshold voltage increase in erased state indicates positive charge leaking away from the traps.Detrapping processing in program and erase state is same because electron trap energy level and hole trap energy level in nitride is 2.94 eV.

Device Parameters:

Temperature =85 C^0

Lg = 130nm
Vp = 9v Ve=-9V
dtp = 1ms dte=1ms

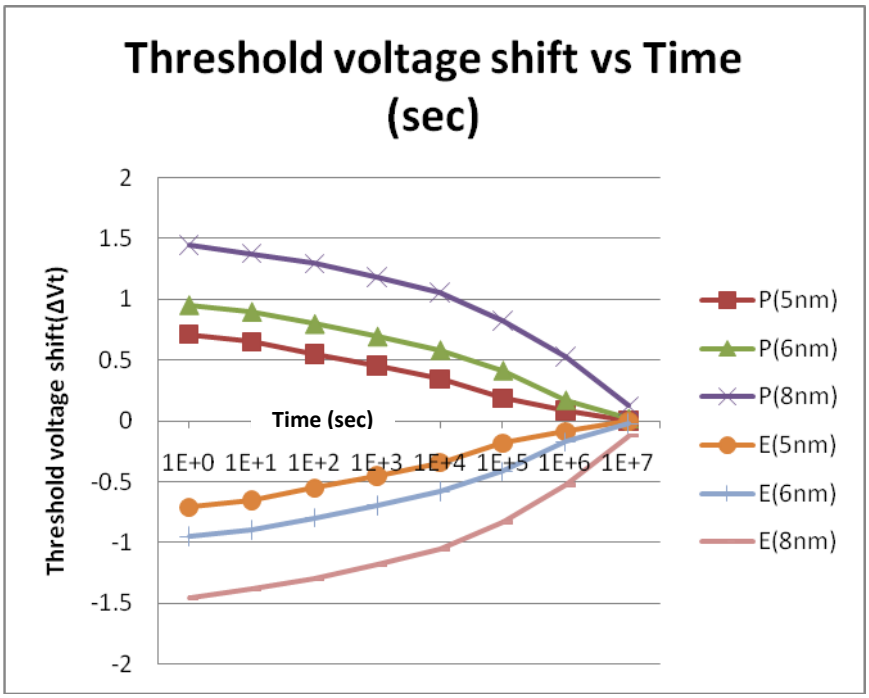


Figure 3.16 Retention characteristic of SONOS device at different nitride thickness.

3.6.1 Nitride thickness versus threshold voltage at different tunnel oxide

Device Parameter

Lg = 130 nm

Blocking Oxide = 4nm

Tunnel Oxide = 1.8nm, 2nm, 2.5nm, 3nm, 4nm

Program time = 1ns

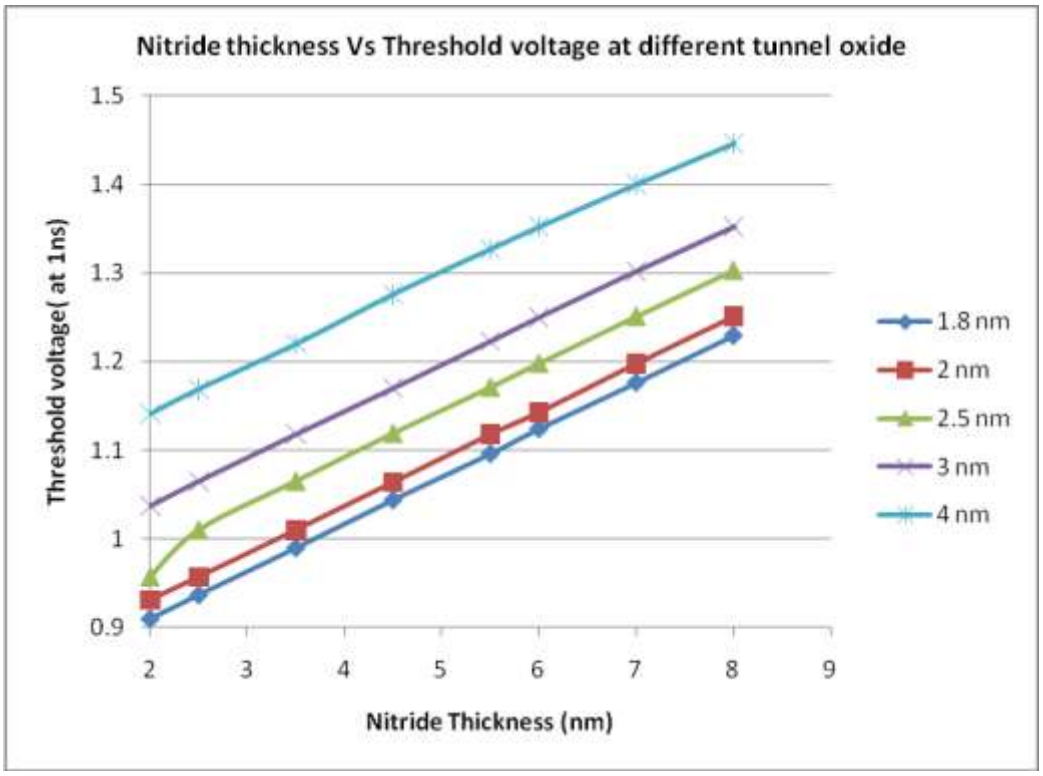


Figure 3.17 Nitride Thickness vs Threshold voltage shift at different tunnel oxide

Conclusion:
1. Turn-on threshold voltage increases with increase in nitride thickness
2. Turn-on threshold voltage increases with increase in tunnel oxide for same nitride layer thickness.

3.7 Effect of replacing Si_3N_4 layer by High-k dielectric.

The program results of SONOS and SOHOS devices are shown in figure below. The SOHOS device has faster programming speed compared to SONOS. For SOHOS device, the charges maybe trapped in electron and hole traps in the quantum well. From ideal energy band diagram, the quantum well formed by conduction band is deeper for SOHOS structure compared to SONOS. Therefore, at same gate bias; electron will tunnel through

thicker energy barrier in SONOS to the conduction band of the charge storage layer (Si_3N_4) as compared to SOHOS.SOHOS device charges up faster than SONOS device.

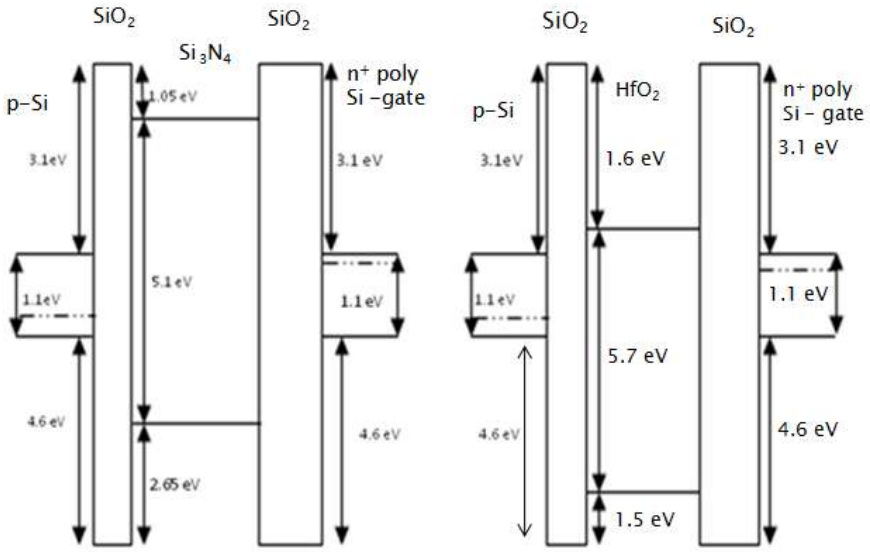


Figure3.18 Ideal energy band diagram for (a) SONOS and (b) SOHOS structure.

The conduction band offset of Si_3N_4 with respect to silicon is 2.05eV, as compared to a 1.5eV conduction band offset of HfO_2 with respect to silicon. This is shown in band diagram. Hence electron tunneling through charge storage in quantum well will be easier in SOHOS as compared to SONOS devices.

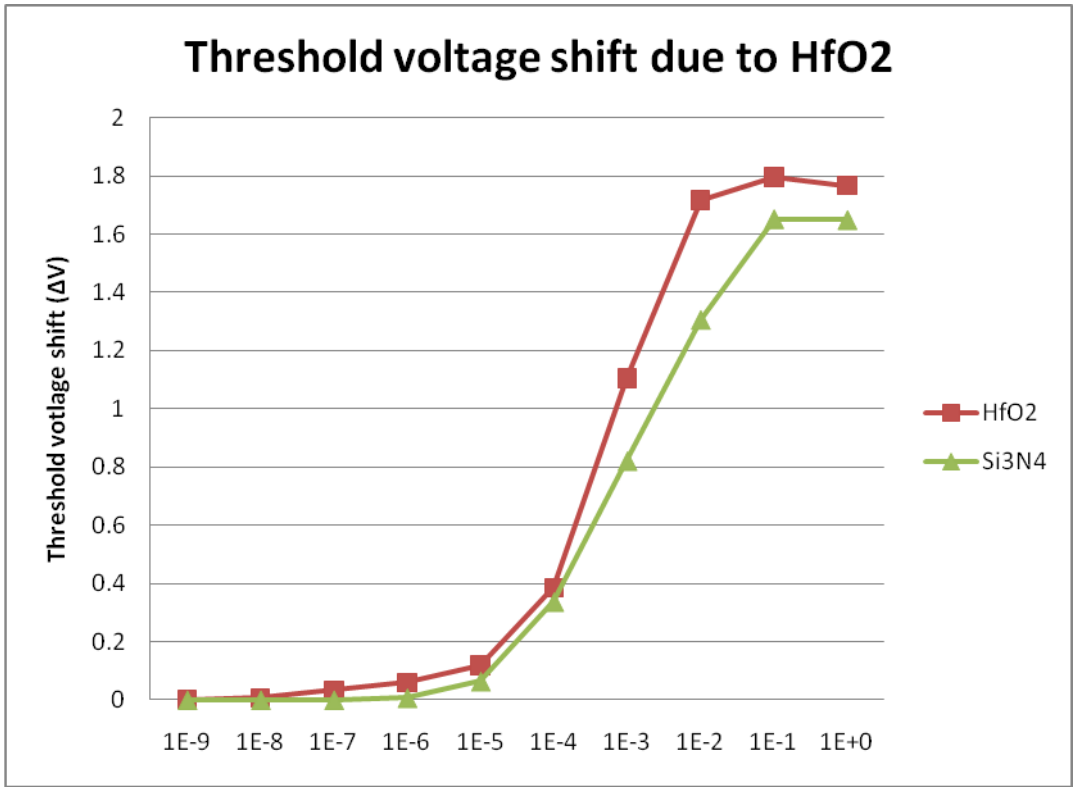


Figure 3.19: Threshold voltage shift due to HfO_2

Chapter 4

Conclusion

In this work, SONOS memory cells with varying dielectric thickness have been systematically and analytically studied by using a SYNOPSIS Technology CAD simulation and modeling tool. The mechanisms of carrier tunneling and memory retention of the memory cells have been tested by simulation and studied in detail. In addition, Hafnium oxide (HfO_2) as charge trapping layer has also been analyzed by using TCAD simulation.

From the simulation results we have the following conclusions: (i) as the thickness of charge-storage dielectrics (e.g., Si_3N_4 and HfO_2) increases, the threshold voltage shift increases due to the higher capture probability of electrons; (ii) as the thickness of tunneling oxide decreases, the programming speed increases due to the higher electric field across the tunneling oxide; (iii) as the channel length decreases from 210 nm to 70 nm, threshold voltage shift increases with increasing program time; (iv) as the temperature increases, threshold voltage shift decreases.

In summary, the SONOS-like nonvolatile memory is a strong candidate for future high-density, high-performance, portable electronics.

List of References

List of References

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

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

[3] Y.P.Tsividis, Operation and Modeling of the MOS Transistor. New York: McGraw-Hill1987.

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

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

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

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

[8] J.R. Cricchi, F.C.Blaha, and M.D. Fitzpatrick, " The drain-source protected MNOS memory device and memory endurance,"IEEE IEDM Tech. Dig., p.126, 1973.

[9] Y.Yatsuda, T.Hagiwara,R.Kondo,S.Minami,and Y.Itoh, "N-channel Si-gate MNOS device for high speed EAROM," Proc. 10th Conf. Solid State Devices,p.11,1979.

[10] T.Hagiwara, Y.Yatsuda, R.Kondo, S.Minami, T.Aoto, and Y.Itoh, "A 16kbit electrically erasable PROM using n-channel Si-gate MNOS technology," IEEE J.Solid State Circuits, vol. SC-15, p.346, 1980.

[11] Y.Yatsuda, T.Hagiwara, S.Minami, R.Kondo, K.Uchida, and K. Uchiumi, "Scaling down MNOS nonvolatile memory devices," Jap. J.Appl.Phys. vol.21,S21-1, p.85, 1982.

[12] T.Hagiwara,Y.Yatsuda,S.Minami,S.Naketani,K.Ushida, and T.Yasui, "A 5V only 64k MNOS memory device for highly integrated byte erasable 5V only EEPROMs," IEEE IEDM Tech. Dig., p.733, 1982.

[13] F.R.Libsch,A.Roy,and M.H.White, "Amphoteric trap modeling of multidielectric scaled SONOS nonvolatile memory structures," 8th NVSM, Vail, Colo., 1986.

[14] D.Adams, J.Murray, and M.H.White, "SONOS Nonvolatile Semiconductor Memories for Space and Military Applications", NVMTS, San Diego, CA, 2001.

[15] J.D.Lee, et al,, "Effects of floating-gate interference on NAND flash memory cell operations," IEEE Electron Device Letters, vlo. 23, pp., 264-266, May 2002.
[16] Minami et al. "A novel MONOS nonvolatile memory enduring 10-year data retention after 10^7 Erase/Write cycles," IEEE Trans. Elec. Dev. 10(11), pp. 2011, 1993.
[17] Chen P.C.Y. (1977) IEEE Tr. On E.D.,ED 24,584.
[18] Gentil P. (1989) Instabilities in silicon devices, Vol. 3, edited by G. Barbottin and A. Vapaille,Elsevier science publishers B.V. (North Holland)
[19] Chu T.L., Szedon J.R. and Lee C.H. (1967) Solid State Electr., 10, 897.
[20] Suzuki E., Hiraishi H., Ishii K. and Hayashi Y. (1983) IEEE Tr. On E.D., ED 30,122.
[21] Yatsuda Y., Hagiwara T., Minami S., Kondo R. and Uchida K. (1982) IEEE Tr. on E.D., 36, 1145.
[22] Roy A. and White M.H. (1991) Solid State Electr. 34, 1083.
[23] Hu Y. and White M.H. (1993) Solid State Electr., 36, 1401.
[24] Miller S.L., Mc Whorter P.J., Dellin T.A. and Zimmerman G.T. (1990) J.Appl. Phys., 67, 7115.
[25] Furnemonth, Arnaud, et al., "Physical understanding and modeling of SANOS retention in programmed state." Solid-State Electronics, 2008, pp. 577-583.
[26] Tsai, W J, et al., "Cause of Data Retention Loss in a Nitride-Based Localized Trapping Storage Flash Memory Cell." Dallas":s.n., 202. 40th International Reliability Physics Symposium.
[27] Bacchofer, H, et al., "Transient Conduction in multidielectric silicon-oxide-nitride-oxide semiconductor structures." Journal of Applied Physics, 2001, Vol. 89,pp. 2791-2800.
[28] Y. Yang and M.H. White, "Charge Retention of Scaled SONOS Nonvolatile Memory Devices at Elevated Temperatures", Solid State Elect. 44, 949 (2000).
[29] Stein H.J. and Wegener H.A.R. (1977) J. Electrochem. Soc., 124, 908.
[30] M.L. French and M.H.White, "Scaling of multidielectric nonvolatile SONOS memory structures." Sol. State Electron. vol., 37p. 1913, 1995.
[31] S.Manzini and A. Modelli, "Tunneling Discharge of Trapped Holes in Silicon Dioxide", in Insulation Films on Semicondcutors, J.F. Verwij and Wolters,Eds. Amsterdam,The Netherlands; Elsevier, p. 112, 1983.
[32] W.J.Tsai, N.K.Zous,C.J. Liu,C.C. Liu,C.H.Chen, "Data Retention Behavior of a SONOS Type Two-Bit Storage Flash Memory Cell".
[33] F. R. Libsch and M. H. White, "Charge transport and Storage of low programming voltage SONOSlMONOS memory devices," ***Solid-stateElectron.,*** vol. 33, pp. 105-126, 1990.

Curriculum Vitae

Asha Rani graduated from Holy Cross School, Bokaro Steel City, India, in 1997. She received her Bachelor of Science from Birla Institute of Technology in 2001. She was employed as a Lecturer in University of Rajasthan, India for four years and received her Master of Science from George Mason University in 2010.