NON-VOLATILE FLASH MEMORY CHARACTERISTICS IMPLEMENTING HIGH-*K* BLOCKING LAYER

SEUM BIN RAHMAN

UNIVERSITI TEKNOLOGI MALAYSIA

Replace this page with form PSZ 19:16 (Pind. 1/07), which can be obtained from SPS or your faculty.

*Replace this page with the Cooperation Declaration form, which can be obtained from SPS or your faculty. This page is **OPTIONAL** when your research is done in collaboration with other institutions that requires their consent to publish the finding in this document.]*

NON-VOLATILE FLASH MEMORY CHARACTERISTICS IMPLEMENTING HIGH-*K* BLOCKING LAYER

SEUM BIN RAHMAN

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master's In Electrical Engineering

Faculty of Electrical Engineering
Universiti Teknologi Malaysia

JUNE 2018

This work is dedicated to my respected parents and UTM authority for having faith on me.

ACKNOWLEDGEMENT

I want to thank DR. NURUL EZAILA BINTI ALIAS for her great guidance of the project. I also thank all my examiners for asking creative questions that helped improving the report. Besides, my gratitude goes to my fellow research group members, specially Muhammad Afiq Nuruddin and Farah Aqilah for their kind and patient support.

ABSTRACT

An Erasable Programmable Read Only Memory (EPROM) is a special kind of memory chip, that can retain the memory even when the power is turned off. This type of memory is known as non-volatile memory (NWM) cell. An EPROM, as a non-volatile memory is widely accepted for its excellent reliability and data storage capability for a large scale of time without noticeable data degradation. It is an implementation of floating gate charge storage system, where a conductive polysilicon layer traps electrons and shifts the threshold voltage. But research on EPROM has become very insufficient recently due to its poor memory characteristics and bulky size. To overcome this issue, an implementation of high-k as a blocking layer of an EPROM instead of SiO_2 is proposed. The proposal includes the reduction of size by using split floating gate, which was applied in FinFET device structure. The best high-k material (Si_3N_4, HfO_2, and ZrO_2) are implemented in this work and floating gate structure is chosen based on the literature review studies and applied on the EPROM to yield its excellent retention characteristics with better memory window. The dependency of EPROM characteristics and high-k blocking layer is also hypothesized, complemented with its physical and tunneling model. The EPROM device has been simulated using Silvaco TCAD Tools. An EPROM with high-k blocking layer shows much improvement in memory characteristics compared to conventional SiO_2 blocking layer. It shows that, the relationship between k value and memory window is in exponential behavior. The higher the k value, the larger the memory window obtained. The proposed device shows 5.6V of memory window which is 3 times larger than the existing devices. For the reliability, almost no retention degradation after 10 years of extrapolation, which is about 50% improvement than existing devices. These improvements also have been validated with the literature review.

ABSTRAK

An EPROM, or Erasable Programmable Read Only Memory is a special kind of memory chip, that can store memory even when the power is turned off, and can retain data after the supply is available. This type of memory is termed as non-volatile memory cell. A non-volatile memory (NVM) is a type of semiconductor based. An EPROM, as a non-volatile memory is widely accepted for its excellent reliability and data storage capability for a large scale of time without noticeable data degradation. It is an implementation of floating gate charge storage system, where a conductive policilicon layer traps electrons and shifts the threshold voltage to the right. But research on EPROM has become very insufficient recently. The poor memory characteristics shown by the device has created hindrance towards development of the device. Besides, the bulky size is another disadvantage of the device. To overcome this issue, implementation of High "k" blocking layer instead of Silicon Di-Oxide is proposed. The proposal also include the reduction of size by using split floating gate, which was applied in FinFET before. The best High k material and best floating gate structure is selected from the Literature Review, and implemented on EPROM to combine its excellent retention characteristics with good memory characteristics and come up with the best existing device. Later, the reason behind the memory improvement due to introduction of high k blocking layer is also hypothesized, accompanied with physical and mathematical modeling. It shows that, the relationship between value k and memory window is exponential. This relationship is graphically shown and mathematically and physically explained. The proposed device shows 5.6V of memory window which is 3 times than the best exsisting devices. For the reliability, almost no retention degradation after 10 years, which is 50% improvement than existing devices. At the end of the thesis, these improvements are validated with Literature Review and proven to be best device among the existing ones.

TABLE OF CONTENTS

CHAPTER	TITLE		PAGE
	DECLARATION		ii
	DEDICATION		iii
	ACKNOWLEDGEMENT		iv
	ABSTRACT		v
	ABSTRAK		vi
	TABLE OF CONTENTS		vii
	LIST OF TABLES		x
	LIST OF FIGURES		xi
	LIST OF ABBREVIATIONS		xiv
	LIST OF APPENDICES		xv
1	**INTRODUCTION**		1
	1.1	Problem Background	1
		1.1.1 Memory Cell	5
		1.1.2 Different types of NVM cells	6
		1.1.3 Flash Memory Cell	9
		1.1.4 MOSFET based Flash Memory Cell	10
		1.1.5 Silicon on Insulator based FinFET	16
		1.1.6 High *k* Material	19
		1.1.7 EPROM	19
	1.2	State-of-the-Arts	20
	1.3	Problem Statement	21
	1.4	Objective and Scope	22
		1.4.1 Objectives	22
		1.4.2 Scopes	23

1.5 Organization 23

2 LITERATURE REVIEW 24

2.1 Bench-marking Terminologies 24

2.1.1 Retention 25

2.1.2 Memory Window 26

2.2 Review on High *k* Materials in MOSFET 27

2.2.1 Summary 31

2.3 Review on FinFET Based flash NVM until 2007 32

2.4 Review on Combinational Blocking Layer of High *k* Si_3N_4 an Al_2O_3 40

2.5 Review on NVM until 2017 41

2.6 Literature Review on EPROM 46

2.7 Literature Review Summary 47

3 RESEARCH METHODOLOGY 49

3.1 Top-level View 49

3.2 Project Workflow 49

3.3 Simulation Works 57

4 RESULT AND DISCUSSION 59

4.1 Introduction 59

4.2 Device Simulation: 60

4.3 I-V characteristics curve 63

4.4 Electrical Characterization of EEPROM Device structure 64

4.4.1 DIBL 67

4.4.2 Electrical Properties Summary 67

4.5 Memory Characteristics 68

4.5.1 Memory window 69

4.5.2 Program Behavior versus different high-k material 71

4.5.3 Erase Behavior versus different high-k material 72
4.5.4 Retention Behavior versus different high-k material 73
4.5.5 Dependence on value "*k*" 74
4.5.6 Mathematical Modeling 75
4.5.7 Physical Modeling 79
4.5.8 Validation With Literature Review 80
4.5.8.1 Validation With Respect to Program Characteristics 80
4.5.8.2 Validation With Respect to Erase 81
4.5.8.3 Validation with respect to Retention 82
4.5.8.4 Validation Summarization 83

5 CONCLUSION 85

REFERENCES 86

LIST OF TABLES

TABLE NO.	TITLE	PAGE
1.1	Different High *k* Materials	19
2.1	Citation Table for Figure 2.5and Figure 2.4	38
2.2	Citation Table for Figure 2.9	44
2.3	Summary of the Literature Review	48
3.1	Device parameter of the proposed device	58
4.1	Summary of Preliminary Results	68
4.2	Validation Summary	83

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
1.1	Usage rate of memory devices [ITRS]	2
1.2	Trend Semiconductor Materials According to Moore's law [ITRS](1)	2
1.3	Introduction of ITRS with near-term and long-term projection [ITRS]	4
1.4	Introduction of "More Moore"[ITRS]	4
1.5	Ferromagnetic NVM	7
1.6	Pictorial view of NeMM	7
1.7	Spin Transfer Torque Memory	8
1.8	Macromolicular NVM	9
1.9	MOSFET structure	10
1.10	Generic view of a MOSFET Memory cell	11
1.11	Charge trap Process and Labeling of Memory cell Layers	12
1.12	Graphical Representation of V_t shift	13
1.13	A Band diagram showing the Fowler Nordheim tunneling current	15
1.14	Graphical Representation of V_t shift (8)	16
1.15	Introduction to SOI- Silicon on Insulator	17
1.16	From MOSFET to FInFET(Threshold System)	18
1.17	3D image of a FinFET(EEtimes)	18
1.18	Converting from FinFET to EPROM	20
1.19	Generation wise development of Memory cells	21
2.1	Example of Retention Characteristic Trend of a Non-Volatile Memory	26
2.2	Example of Memory Window of a Non-Volatile Memory	27

2.3	Summary of a review on non-volatile memory devices	31
2.4	Memory Window of FinFET Based NVM	37
2.5	Retention and Endurance	37
2.6	Retention Analysis of MANOS and MANOS (21)	39
2.7	Comparative Discussion Between MANOS and MONOS (21)	39
2.8	Retention and memory window improvement of Memory Cell (2)	41
2.9	Summary of NVM flash memory until 2017	44
2.10	Fabricated Device VS Control Device (14)	45
2.11	Fabricated Device VS Control Device (14)	45
2.12	Schematic of the single poly-Si JL-FinFET GAA NVM with ten NWs (40)	46
2.13	Programming and erasing characteristics of the EEPROM cell with devices(40)	47
3.1	: Research Workflow	50
3.2	Simulation window of SILVACO TCAD tool	52
3.3	Sample Output of Atlas Output	53
3.4	Program Methodology	55
3.5	Program Methodology	56
3.6	Program Methodology	57
3.7	Cross Section View of Simulated EPROM	58
4.1	3D view of Simulated EPROM	60
4.2	Silicon Bulk	61
4.3	Determination of oxide thickness	61
4.4	Determination of control gate layer thickness	63
4.5	I-V Characteristics Curve of FG EEPROM	64
4.6	Calculating Subthreshold Swing	65
4.7	Calculating Threshold Voltage	66
4.8	Calculation of I_{ON}	66
4.9	Calculation of DIBL	67
4.10	V_t Shift due to charging: simulated using SiO_2 as Blocking Layer	69

4.11	V_t Shift due to charging: simulated using Si_3N_4 as Blocking Layer	70
4.12	V_t Shift due to charging: simulated using ZrO_2 as Blocking Layer	70
4.13	V_t Shift due to charging: simulated using HfO_2 as Blocking Layer	71
4.14	Program Characteristics of FG EPROM	72
4.15	Erase Characteristics of FG EEPROM	73
4.16	Retention Characteristics of FG EPROM	74
4.17	Relationship between high-k value and memory window	75
4.18	(a) Most metal-semiconductor contacts result in Fermi level pinning to mid-gap on Si and the valence band edge on Ge; (b) fermi level pinning at mid-gap results in a large SchoJ̃ky Barrier Height (SBH) adding resistance; (c) inserting a dielectric layer at the interface reduces MIGS penetration resulting in less Fermi level pinning and SBH can be further tuned by interface dipoles (7)	79
4.19	Validation with respect to program	81
4.20	Validation with respect to Erase	82

LIST OF ABBREVIATIONS

XML	-	Extensible Markup Language

LIST OF APPENDICES

APPENDIX	TITLE	PAGE

CHAPTER 1

INTRODUCTION

1.1 Problem Background

In an electronic system, although the processor performance is prioritized first, still, memory is the one that defines the actual speed and performance, computation expertise and storage capacity of the system. Memory devices are primarily categorized according to the storage characteristics and storage purpose. The usage and implementation of memory in different devices are explained in Figure 1.2. It is clearly depicted that, usage of memory devices has become extremely versatile. Memory required for the applications of a system to run and store temporary set of data for the sake of their purposed execution is referred as volatile memory. This type of memory requires high read and write speed, and very fast data transaction, which is considered as main component that defines the speed of the whole system. These types of devices require frequent flash and extensive re-usability for the fast and high-speed application execution without any hazardous issues. These types of memories are demonstrated as Volatile Memory. Volatile memory is not meant for permanent storage, but to facilitate the application to store temporary files, which are immediately lost once the application is closed or the device is under shut down condition. Dynamic-RAM(DRAM) and Static-RAM (SRAM) are the most commonly known volatile memory. On the contrary, there is another category, which is referred as Non-Volatile Memory (NVM), intended to be used as long-term storage purpose, where slower speed compared to Volatile Memory can be traded off with secured data deposition within an ample time period.

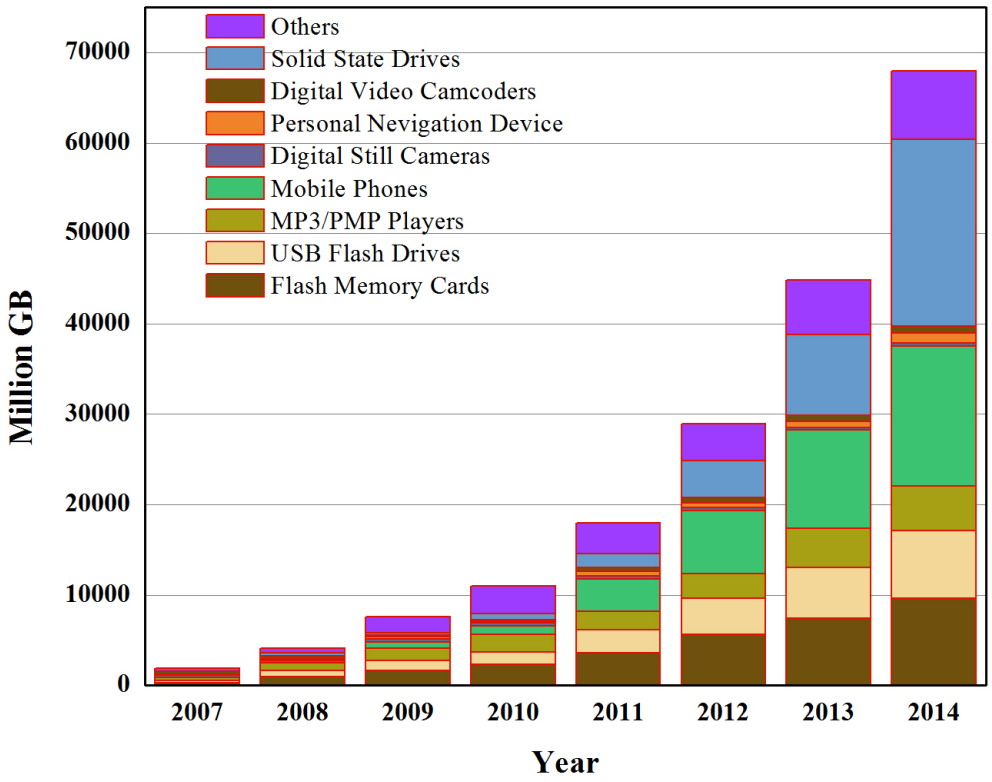


Figure 1.1: Usage rate of memory devices [ITRS]

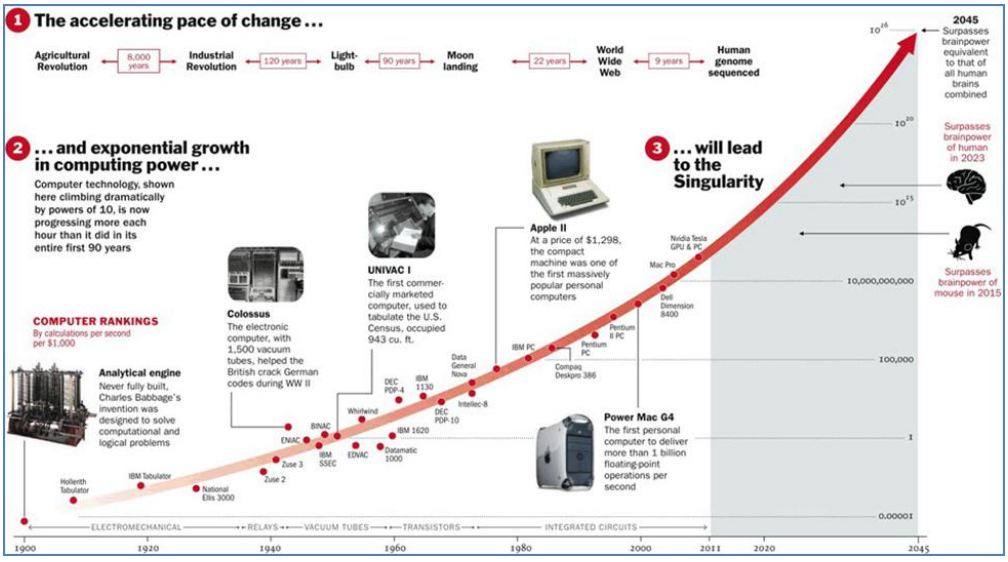


Figure 1.2: Trend Semiconductor Materials According to Moore's law [ITRS][1]

Moore's law has curved the path of advancement of high speed logic computing. This law mentions that number of transistors in a chip doubles with every 18-24months, which ultimately refers to continuous scaling of every electronic components, including the memory storage transistors within the device level. This law can be expressed in different ways. Since the law states about usage of more transistors at same place, it is evident that, each 18-24 month the size has to scaled drastically, and also the price need be static to survive in the competitive market. Figure 1.1shows the extensive use of memory in different devices over the years. Although the usage of memory has increased at an amazing speed, still, the price and the size is has not swelled up, instead dropped for certain cases.

So, right now, the semiconductor researchers are mainly up to optimize in all level. Therefore, here are the actual meaning of "Optimization". By the term "Optimization", scientists and researchers refer to another term, which is called PPA [32]. This term is very popular in Intel. The components of PPA are: Power, Performance (Speed and reliability), Area (resources). Improving all these three criteria are is extremely difficult challenge. Most of the times, all are not achieved, and some are traded off to achieve the most important goal, which depends on the purpose of the chip itself. This optimization is performed in all levels. From device level to RTL and system level, quest for optimization of PPA is always the most primary objective. In device level, the problem becomes more crucial, because here the engineers need to deal with extremely small area, within micron range, extremely small current flow. Figure 1.3 shows the development process of memory cell each year. Year 2010 is the most significant period in the history. At this year, memory cell development entered a new era. This year, memory cell industries stated 26nm Memory cell. This is also the beginning of ITRS, which was planned 2009, and implemented from 2010. PIDS of flash projection also began at 2010, that projected to develop 8nm within 2020. Production of volatile DRAM was fabricated in 2012, and PIDS of flash 4 bits/cell is within 2019. T

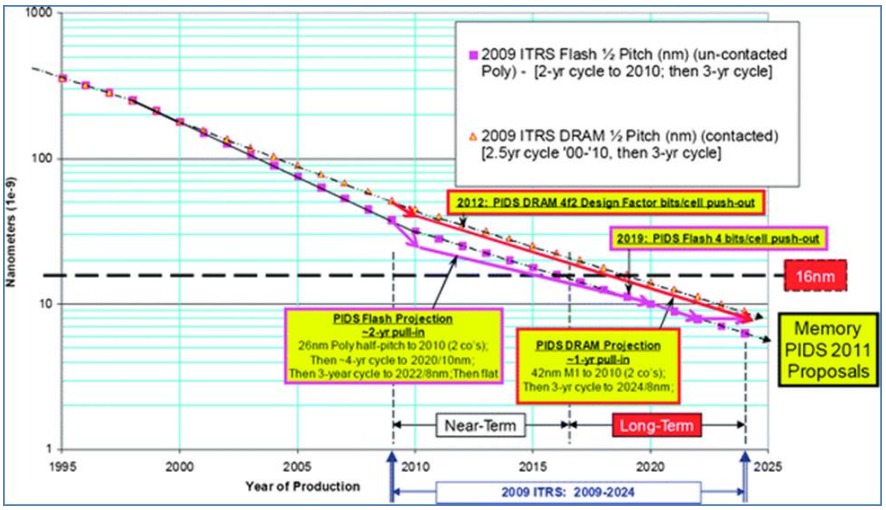


Figure 1.3: Introduction of ITRS with near-term and long-term projection [ITRS]

The nearterm projection was building 42nm, within 2017, which has already met by using FinFET. Actually, FinFET has gone beyond that, its now 32nm, that refers to new trend named "More Moore". The description is provided in Figure 1.4:

Figure 1.4: Introduction of "More Moore"[ITRS]

The International Technology Roadmap for Semiconductors (ITRS) is a set of documents produced by a group of semiconductor industry experts. These experts are representative of the sponsoring organizations which include the Semiconductor Industry Associations of the United States, Europe, Japan, South Korea and Taiwan. The documents produced carry this disclaimer: "The ITRS is devised and intended for

technology assessment only and is without regard to any commercial considerations pertaining to individual products or equipment". The documents represent best opinion on the directions of research and time-lines up to about 15 years into the future for some few well defined areas of technology. The broad area include driver design, testing environment, front end process integration, communication through radio frequency, micro-mechanical, photo lithography and much more.

As of 2017, ITRS is no longer being updated. Its successor is the International Roadmap for Devices and Systems.

Prior to aforementioned systematic procedure, the whole semiconductor level is under continuous research. But this thesis work will concentrate in the device level, deal with the problems related to unit cell of any semiconductor system, which is named as “Memory Cell”. Like all other semiconductor devices, Flash Memory has shown its tremendous evolution with a very short time span. The capacity of data storage has increased in an unbelievable rate. The techniques and theories are developed at an extreme high speed within the last few decades. For achieving this purpose, the transistors need to be exibit better performance and greater reliability, that requires introduction of newer technologies every year. This research work will demonstrate combination of different ideas to come up with the best device available at present.

1.1.1 Memory Cell

The memory cell is the most basic part of any memory based digital system. There are different types of memory cells.

Basic classifications of memories are:

1. **Volatile:** This type of memory cells are usually used for data processing, data computation. The memory cell data cannot be regained after the system is turned off, is called volatile memory. For example, SRAM, DRAM, PRAM

etc. These memory cells are related to speed of any electronic device. Right now in our laptop and PCs and Smart Phones, RAM up to DDR4 is available.

2. **Non-Volatile**: This is the opposite of volatile memory. The memory that can store data even after the device is switched off is called non-volatile memory. This memory is actually used for data storage, program and operand storage. This type of non-volatile memory also classified as the follows.

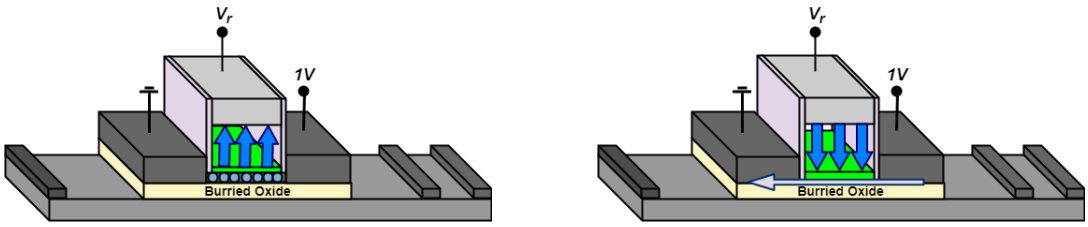
   (a) **Magnetic Memory Storage**: In this case, magnetic materials are used to store data and program. A disc is used in this case. At the beginning, only audio recordings were used to store at to these kinds of discs. But later, these kinds of discs were also used in volatile memory cell.

   (b) **Flash Memory Storage**: This is a total electronic technology, where MOSFET based memory cell is used. This is our main point of discussion. Optimization of this type of memory cell is the purpose. Before that, first the detailed description on the flash memory is required. Flash memory can be further subdivided into 'Floating Gate' based and 'Charge Trap' based memory. In floating gate, charge storage layer is a conductive material, the trapped electron is resided in conduction layer of the material. On the other hand, in charge trap layer, insulating material is used where charge is trapped between conduction band and valance band.

1.1.2 Different types of NVM cells

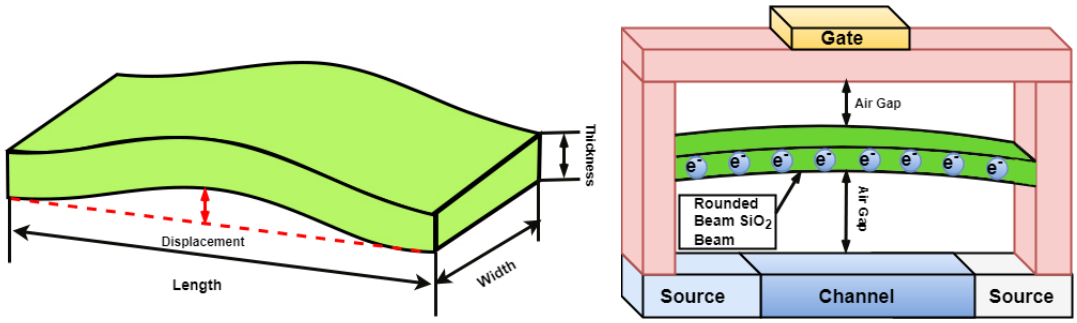
Moore's law has curved the path of advancement of high speed logic computing. This law mentions that number of transistors in a chip doubles with every 18-24 months, which ultimately refers to continuous scaling of every electronic components, including the memory storage transistors within the device level. Figure 1.1Scaling down non-volatile memory is faced with one key challenge. The challenge is that, the scaling down process cannot be traded off with the electrical and reliability performances, and the performance need to be improved along with the scaling down process. In this challenging path, many researchers came up with lots of novel ideas and demonstrated excellent results in their publications. Ferroelectric Field Effect

Transistor (Figure 1.5) is one of the possible candidates, that can be used as non-volatile memory. The idea is using ferroelectric material that causes the memory cell to turn on (Figure 1.5a) and turn off (Figure 1.5b). The device is based on dual states, which can refer to bit 0 and bit 1 (Figure 1.5a1.5b). Another possible candidate is Nanoelectromechanical Memory (NeMM). When this device is charged, the charge storage layer deforms. As both ends of the charge storage layer is clamped with the sidewall, the middle part is buckled upward or downward, which decides closing or opening of a circuit (Figure 1.6a, 1.6b).



(a) State 1 **(b)** State 2

Figure 1.5: Ferromagnetic NVM



(a) NeMM deformation Layer **(b)** NeMM full view

Figure 1.6: Pictorial view of NeMM

Besides these two, Spin Transfer Torque Memory was also proposed and investigated. Current flowing out of the fixed layer is spin-polarized(Figure 1.7). When it reaches the free layer the majority spins relax into lower-energy states of opposite spin, applying a torque to the free layer in the process. In this process, two states

are partitioned, which refers to bit 1 and 0. Nanowire phase change memory has also been presented, where phase changing materials (PCMs) are used as the heart of the component. These materials can reversibly and rapidly change their structure from a stable crystalline to a metastable amorphous phase under the influence of an optical or electrical pulse, which defines two stated that can be referred as bit 1 and bit 0. Macromolecular memory is another kind of non-volatile memory, which come with two different states, high state and low state, which are defined according to a distinct range of current

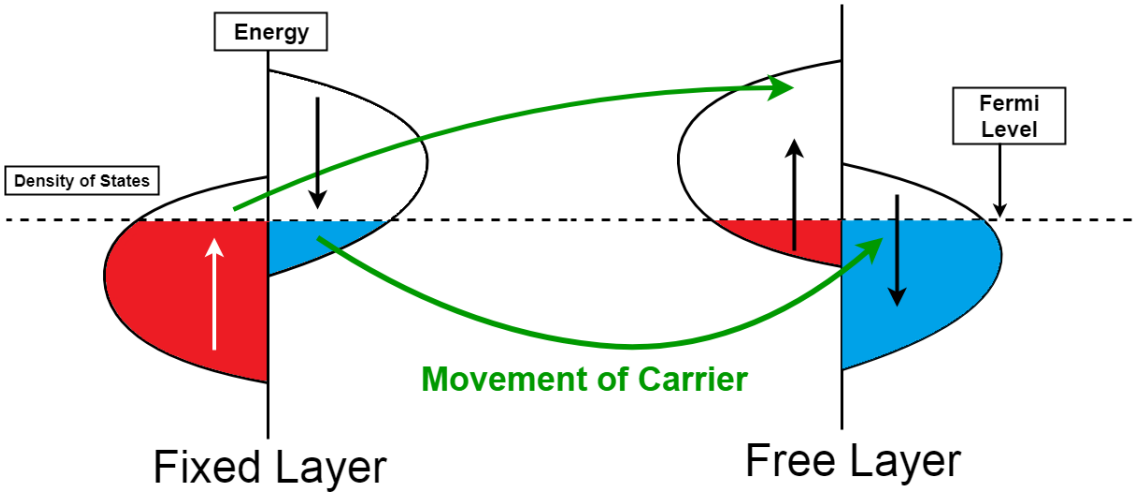
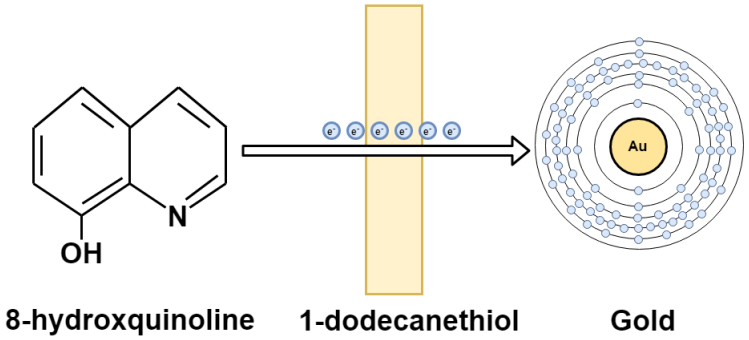
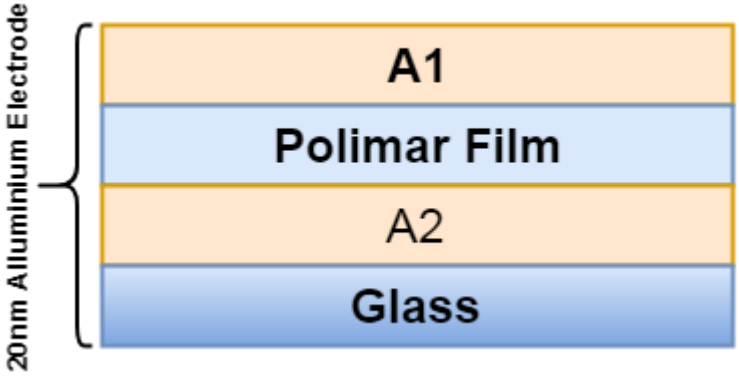


Figure 1.7: Spin Transfer Torque Memory

Figure 1.8 shows a generic cross section depicting layer structure of the stated device. A composite polymer film is placed between two Aluminum electrodes. In the polymer, resides 8-hydroxquinoline and gold molecules, which are separated by partition built with 1-dodecanethiol. The low state is defined as small current flow, caused by impurities residing inside the layers, where the materials are at chemically non-reactive state with each other. But at high state, electron is liberated from the 8-hydroxquinoline and passes the barrier of 1-dodecanethiol and reaches to gold molecules, and thus creates negative and positive domains on the two sides of the barrier. This state is termed as high state, that causes higher current conduction. These two states can define bit 0 and bit 1.



(a) Operation



(b) Structure

Figure 1.8: Macromolicular NVM

1.1.3 Flash Memory Cell

For the sake of long term data retention and superior defect handling capability, charge trap NVM is implemented in planar CMOS devices, and came up with non-mechanical charge store flash memory. This kind of device has proven it efficiency and widespread opportunities to be upgraded with new inventions. There are two types of MOSFETs. For the research, only n type MOSFET will be used. General construction of P-type MOSFET is depicted in Figure 1.9.

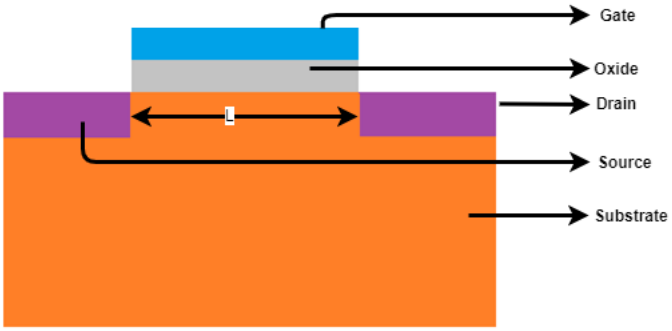


Figure 1.9: MOSFET structure

The memory transistor is structured similarly as MOSFET or FinFET, with two additional layers integrated for the memory storage purpose. Figure 1.9 shows the structural orientation of a generic charge trap memory, integrated within a planar MOSFET transistor. According to the illustration, the charge trap layer is isolated by oxide dielectric from all sides, but the top side is deposited with conductive material comprised of polysilicon or metal which is separated from the charge trap layer by a layer of dielectric oxide. This conductive region is termed as gate, and the isolating dielectric region is called blocking layer. Besides,according to Figure 1.10, there exist another insulating dielectric oxide layer separating the charge trap layer and inversion channel, named as tunneling oxide. The name is originated from its purpose, which is to facilitate accumulated electrons to travel from inversion channel to charge trap layer.

1.1.4 MOSFET based Flash Memory Cell

The MOSFET usually acts like a switch. First, assuming an n-channel MOSFET, a high voltage is provided to the drain terminal, and low voltage to the source terminal. Then, positive voltage is imposed on to the gate terminal. Under the oxide layer, holes are repelled by the voltage and minority carrier electrons are gathered. The channel is inverted from p-type to n-type. This channel is called as an inversion channel. Now if the inversion channel is broad enough, current starts to flow flowing from drain to source.The voltage at which the inversion channel is strong

enough to create the current flow is called threshold voltage. This is very important parameter in terms of Power and Performance. This is referred as V_t in short form.

To convert it from general MOSFET to Memory cell, an additional layer of Gate-Oxide is attached. Nothing new is added later on. Figure 1.10 shows the MOSFET memory cell.

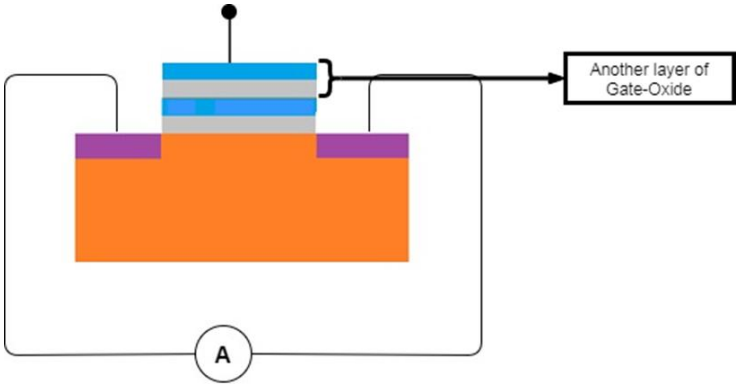


Figure 1.10: Generic view of a MOSFET Memory cell

The operation is quite simple. If a voltage is provided on at the gate terminal and it equals to the Vt, an inversion layer is formed. But when higher voltage is applied, like 12-16 V, electron will go through the first layer of oxide and will reach the layer which was previous known as the gate terminal. Now if the voltage is removed, the charge is trapped inside the layer. This is generally how the memory cell operates.

According to the Figure 1.11, the control gate, blocking layer, charge trap layer and tunneling layer is depicted. These terms are very important for further discussion.

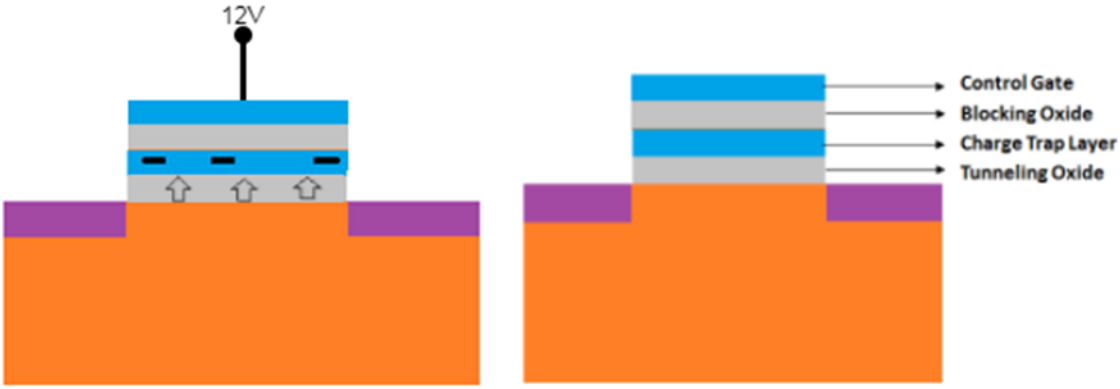


Figure 1.11: Charge trap Process and Labeling of Memory cell Layers

Read and write process is very important part of cell operation. These processes are the key parts that determines the Power and Performance of any memory cell. The faster the operation, the better the performance. At the same time, this process must absorb less power. The read process is based on the V_t shift due to the stored Voltage. The process is described as in Figure 1.12. The first image shows a charged memory cell. This has electron trapped inside it. So, when the voltage is applied on gate, some of the voltage is used up to nullify the trapped voltage. That's why, to create the inversion channel, more voltage is necessary. Which means that, the threshold voltage has increased. According to the Figure 1.12, if the voltage is applied at 5V, there will be no current flow for the charged cell, but will show current in the uncharged cell. If there is current flow, means that data is 0, or else, data is 1. This is the most used read process of memory cell.

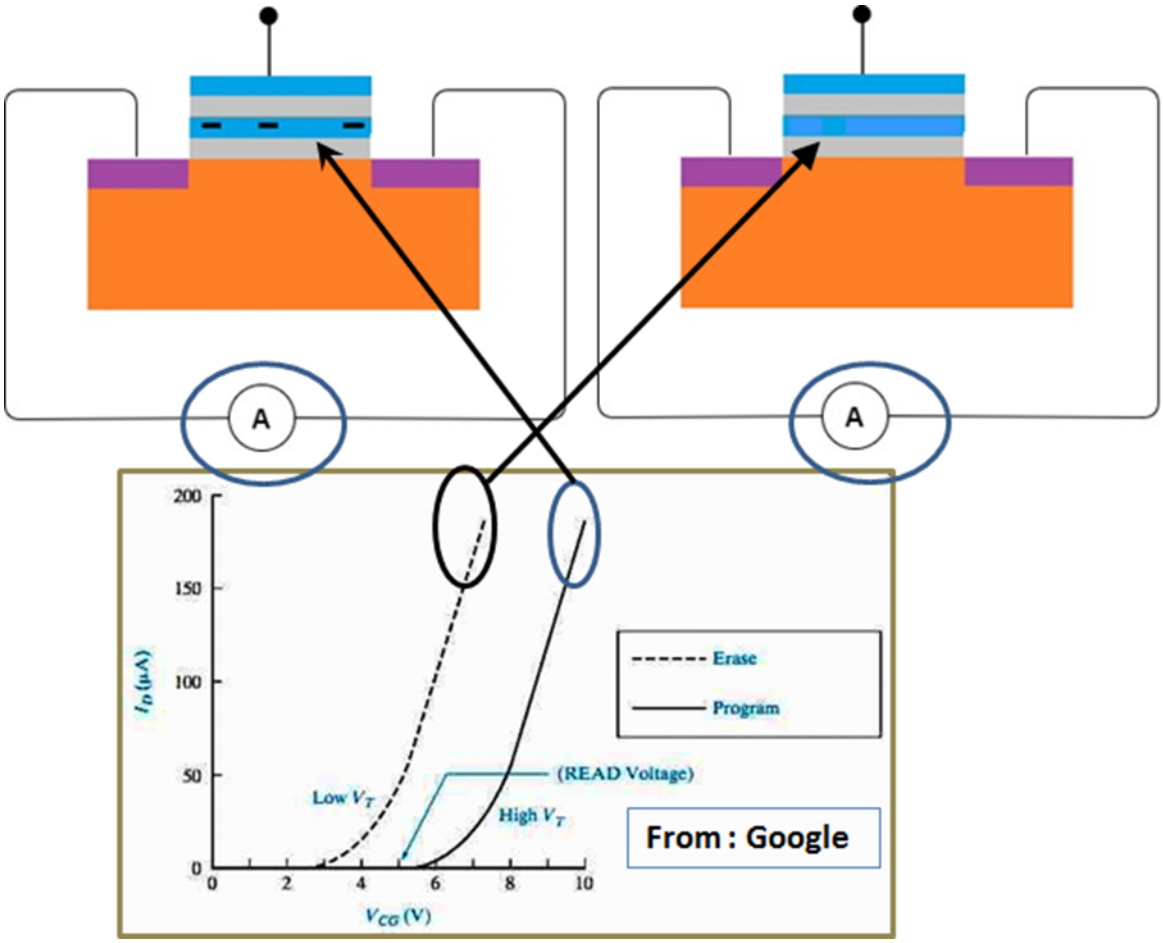
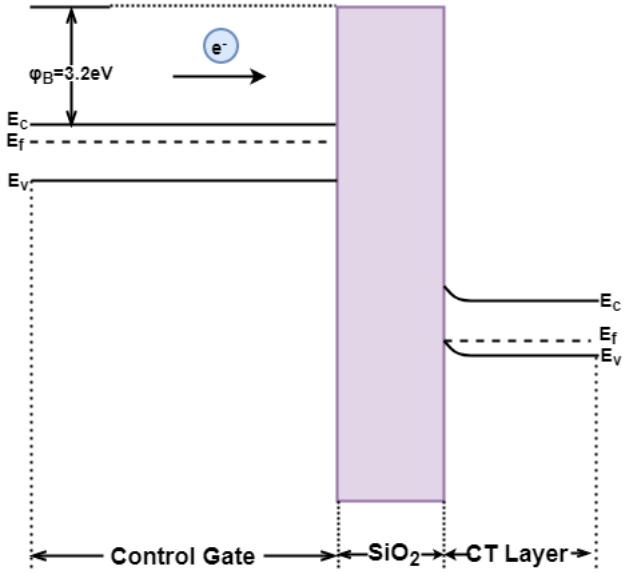


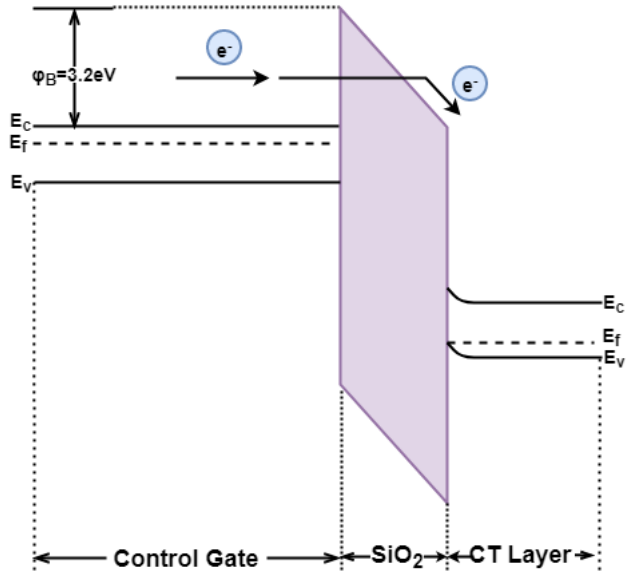
Figure 1.12: Graphical Representation of V_t shift

The non-volatile flash memories are inevitably expected to have wide range of re-usability. To have a memory cell that can toggle from one state to the other (from bit "0" to bit "1") and that can store the information regardless to external conditions, the storing element needs to be a device whose conductivity can be varied in anondestructive way for a several number of time. In the year 1929, Fowler et al presented the final corrected version of Fowler Nordheim tunneling current. According to this theory, the probability of electron tunneling through any oxide barrier is exponentially proportional to the applied electric field. Applied electric field deforms the potential barrier, that turns the barrier from rectangular shape to a triangular shape, which eventually reducing the tunneling path and eases electron flow to the intended charge trap destination. The process is explained in Figure 1.13a1.13b. In Figure 1.13a, while the potential barrier is rectangular, electrons face a long tunneling path. While in the Figure 1.13b, the barrier is deformed to a triangular form, which eases the

electron tunneling through the tunneling layer into the CTL in a quite non-destructive way. This process is very convenient because it allows to obtain the time to program (<1ms)twelve orders of magnitude shorter than usual retention-time (>10 years), which is a fundamental requirement for all NVM technologies. Although the classic form of FN current fits quite well experimental data, many features have been still undervalued: the temperature dependence of the phenomenon; the correct electron statistics, Fermi-Dirac and not Maxwellian ; and the quantum-effects at the silicon-oxide interface. There is another method of programming, which is called Carrier Hot Electron (CHE). In this process, while an electron in the inversion channel which is in equilibrium state due to equal Electric Field and Lattice Vibration [19-20], is exposed to increasing electric field. When the electric field exceeds 100 KV/cm, the electron slides from its equilibrium state and Surmount's tunneling oxide and reaches the charge trap layer. But this process comes with some detrimental effects, such as drain current reduction, small signal performance degradation, threshold voltage shift and I_D sub-threshold slope lowering etc. Erase process is conducted by applying high voltage pulse to the source, while the control gate is grounded, and drains are floating. Electrical erase is achieved via tunneling of charge from the charge-trapping layer to the source.



(a) Before providing Electric Field



(b) After providing Electric Field

Figure 1.13: A Band diagram showing the Fowler Nordheim tunneling current

According to the data storage capability of each cell, flash memory cell can be divided into two types; Single Level Cell (SLC) and Multi Level Cell (MLC). When

the memory window is very small, then according to the $V_t$shift, only 2 bits can be represented, 1 and 0. On the other hand, if the memory window is big enough, it can be subdivided into smaller memory windows, which can represent different bit orientations. Such as, in the Figure 1.14, it can be seen that, a very wide band gap can be obtained in certain memory cell. The $V_t$varies from -2.5V to 7V. So, the V_t=-2 represents 01, V_t=0V represents 10, V_t=4V represents 11. In this process, a big memory window may facilitate storage of multiple bits in one memory cell.

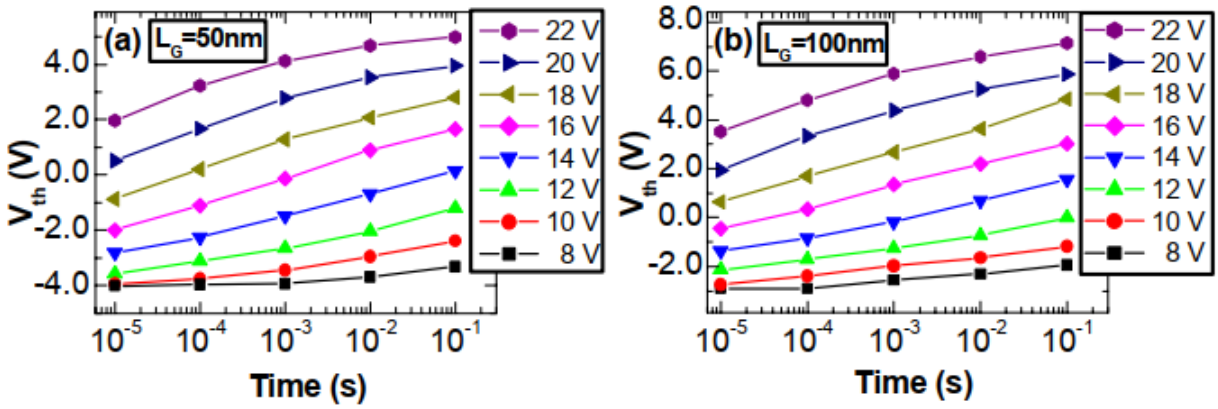


Figure 1.14: Graphical Representation of V_t shift [8]

1.1.5 Silicon on Insulator based FinFET

When a silicon layer is placed on an insulating material, then it is called SOI or Silicon on insulator. Figure 1.15 shows the details on how it is created.

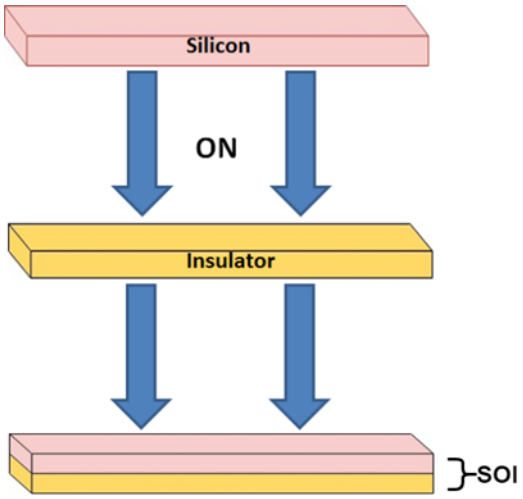


Figure 1.15: Introduction to SOI- Silicon on Insulator

. Planar CMOS devices are integrated with strain silicon materials, graphene and High "*k*" dielectric materials, which paved the the way to sustain from 100 nm until 32 nm node technology. At this point, planar CMOS faces its saturation level at the end of the tunnel. The short channel effect started taking its toll, as the gate of the MOSFET looses the control of the current flow through inversion layer. The subthreshold voltage dominates over the actual threshold voltage of the layer, which results into charge loss, low retention, less protection against P/E cycles and wrong decision during read process. At this point, new devices with more complicated nature, that shows better gate control and suppress the short channel effect is felt necessary. A 3D MOSFET device is proposed, where the gate covers the inversion channel all around. This device uses thin silicon fin as inversion channel. This new 3D device is termed as FinFET. This device shows superior gate control that results into bettersuppression of short channel effect, DIBL, superior electrical and memory characteristics (Figure 1.16). Shows generic structure of an SOIFinFET. This gate all around 3D structure is the next generation of flash memory, that can be scaled down below 32nm node technology without sacrificing electrical and memory performances. In this paper, introduction and development of FinFET as memory cell will be reviewed. First, the development of planar MOSFET as NVMwill be illustrated until it reaches saturation in scaling down, which is 32 nm technology node. Afterwards, the paper will review the role of FinFET as superior candidate of non-volatile memory cell. The results and gathered data will be graphically represented and critically analyzed in terms of electrical and

memory characteristics. Finally, a way to carry on future investigationandresearch direction for the next-generation flash memory will be suggested.

In normal MOSFET as shown in Figure 1.16, the gate is in-between source and gate. But to improve the gate control, the gate is made 3D, and it surrounds the drain and the source. According to the Figure 1.11 first the image is of a planar MOSFET. The gate is then risen up. The interconnection is made by using thin silicon fin. So when the gate voltage voltage is provided, the gate has better control over the whole inversion channel. The FinFET device structure in shown as in Figure 1.17 and Figure 1.17

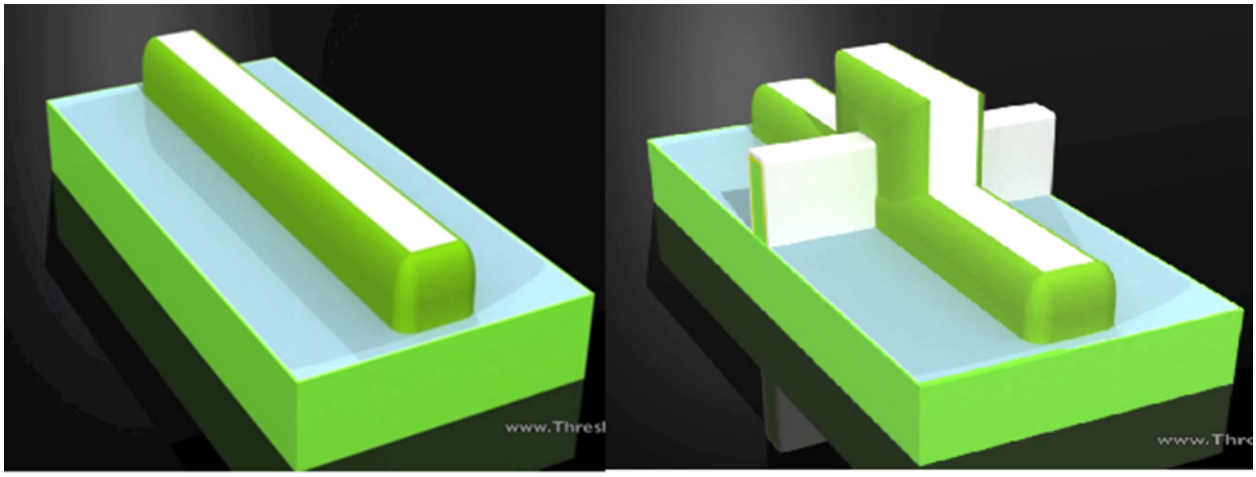


Figure 1.16: From MOSFET to FInFET(Threshold System)

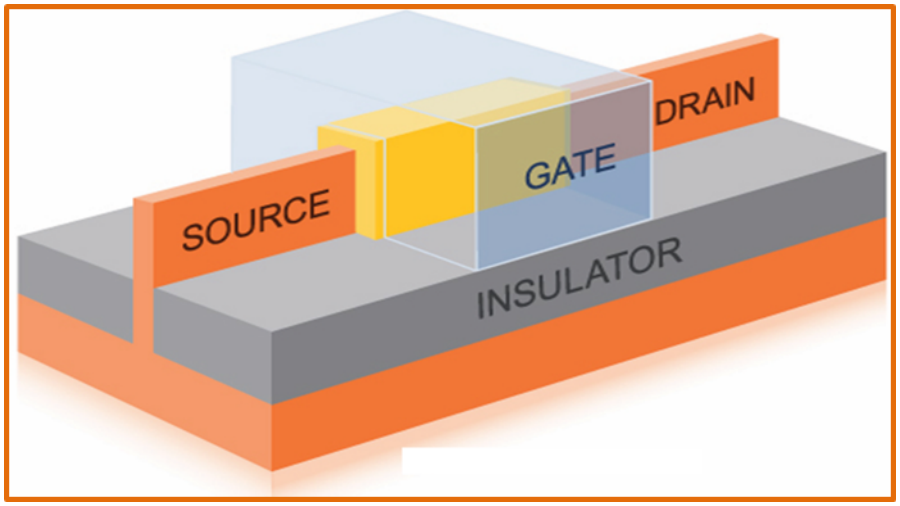


Figure 1.17: 3D image of a FinFET(EEtimes)

1.1.6 High *k* Material

Materials having high dielectric constant is called High *k* material. Dielectric constant, property of an electrical insulating material (a dielectric) equal to the ratio of the capacitance of a capacitor filled with the given material to the capacitance of an identical capacitor in a vacuum without the dielectric material. The insertion of a dielectric between the plates of, say, a parallel-plate capacitor always increases its capacitance, or ability to store opposite charges on each plate, compared with this ability when the plates are separated by a vacuum.

The list of High *k* dielectric materials are given at Table 1.1

Table 1.1: Different High *k* Materials

Material	Value of K
$Al2O_3$	(K=20)
ZrO_2	(K=23)
Y_2O_3	(K=15)
HfO_2	(K=20)

These high k materials are often used as a single layer, and as stacked multi-layer. The structure orientation depends on the purpose of the device.

1.1.7 EPROM

EPROM is another type of flash memory, that can provide high memory window and very good retention trend compared to the other devices. EPROM is a floating gate based structure, where the charge is stored in the conductive band of a polisilicon layer. The structure can be described more simply in the Figure 1.18 , here it is shown that, whenever the thin fin of a FinFET is removed, and extended, and

afterwards the $SiO_2$is replaced with doped Silicon, then it becomes roughly a structure of EPROM.

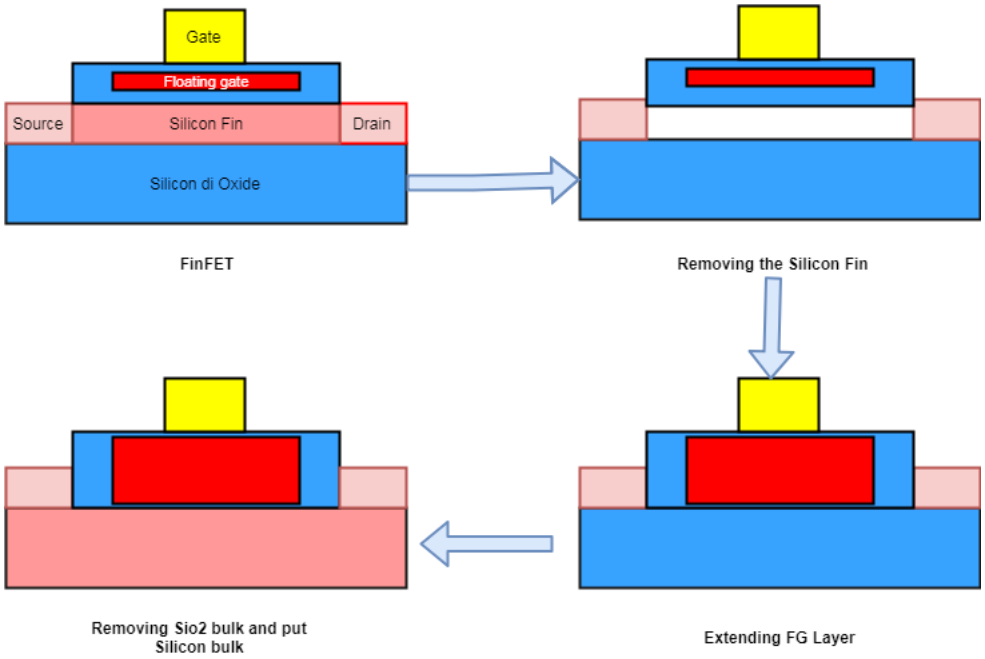


Figure 1.18: Converting from FinFET to EPROM

1.2 State-of-the-Arts

In the heart of Flash Memory Cell, lies MOSFET. In full form, MOSFET stands for means Metal Oxide Semiconductor Field Effect Transistor. MOSFET is a very reliable and low powered device, which has almost zero power loss in static condition. On the way of MOSFET scaling down, different materials are introduced as ingredient of MOSFET recipe. Primarily, The MOSFET consisted of p or n type substrate, highly doped n or p type drain and source (respectively), Silicon Di-Oxide (SiO2) as Oxide, Metal or Polycrystalline Silicon was used. This MOSFET served semiconductor devices for a very long time. But later on, when optimization of area was investigated, some drawbacks were found. The drawbacks are: Short Channel Effect (SCE), DIBL, Subthreshold Voltage, Less Reliability.

Then the doping was increased to solve the problem that ended up with tunneling effect of the electrons and holes. Therefore, later on, when the MOSFET

of first generation, 2003 and second generation, 2005, strained silicon is used. By the change of material, the disadvantages were minimized at a sustainable level. The problem was solved until 2007. Now the drawback seemed to be coming back to being. For this reason, strained silicon was replaced by High *k* materials, which permitted scaling down up to fourth generation memory cell. This technology was up to the mark until 2009. The process is described in Figure 1.19

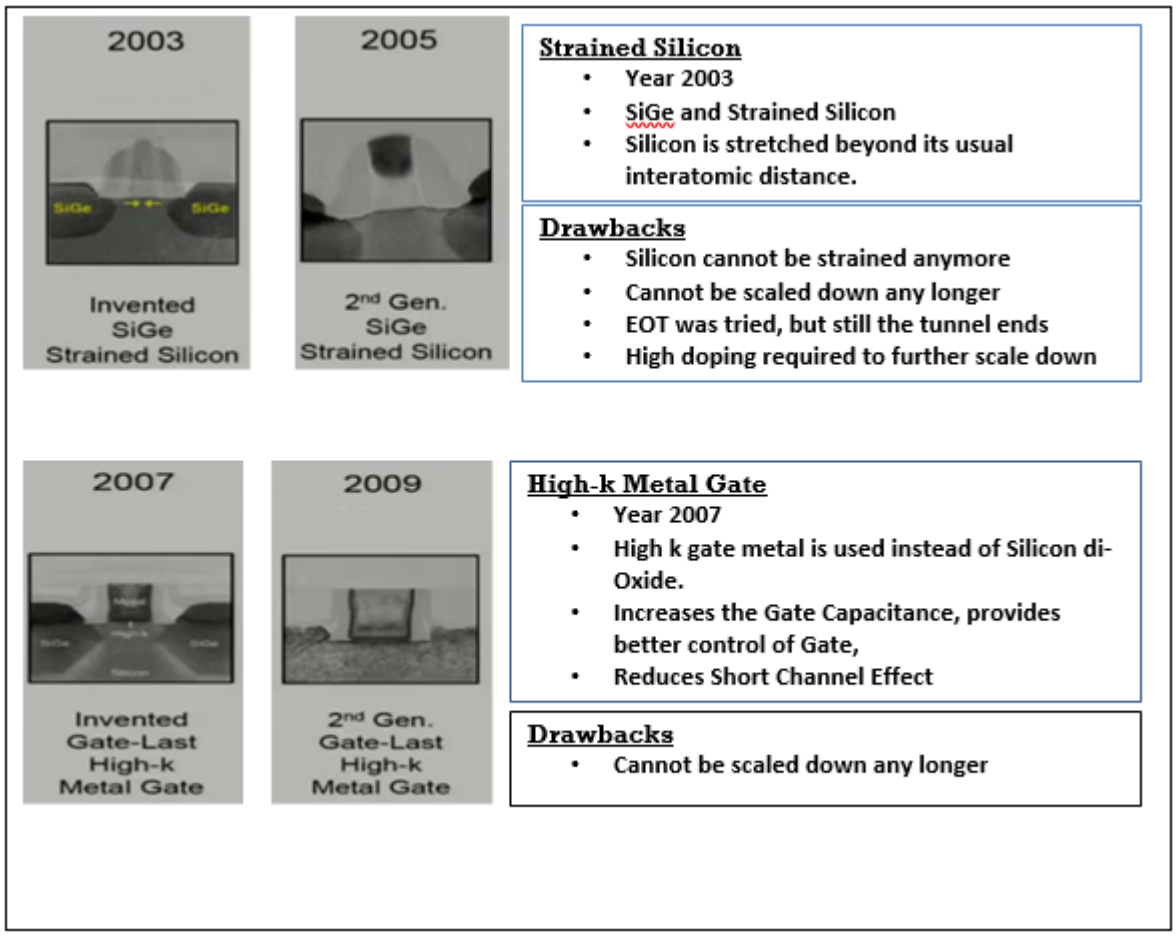


Figure 1.19: Generation wise development of Memory cells

1.3 Problem Statement

No matter what the device is, the problem with both performance and reliability is always a critical issue. When we talk about performance, we mainly refer to the P/E characteristics and memory window which refers to the amount of threshold voltage

shift after program operation. The memory window of the state-of- the art flash memories are showing very small figure. The researches were done until now, rarely shows memory window of more than 5V in small Program/Erase Voltage. All these will be discussed in detail in the Literature Review Chapter. The main parameter of memory reliability is Retention, which is also very critical due to continuous stress condition is applied to the memory device. The retention reliability is very poor in the existing flash memory devices.

Lots of researches are going on to solve these problems. According to the literature review, which will be explained in chapter two, following problem statements can be declared.

1. Memory window beyond 5V at small P/E Voltage and P/E time is not possible in semiconductor based NVM.
2. Recent Research on EPROM is very limited. Most recent researches of EPROM is extremely bulky and of low performance
3. In MOSFET and FinFET, only Al_2O_3 is experimented as High *k* blocking layer
4. Relationship between memory performance and value of dielectric constant is not hypothesized yet.

1.4 Objective and Scope

1.4.1 Objectives

The objectives of the research are:

1. To simulate the EEPROM device and characterize the P/E behavior.
2. To determine the best high-k material as a blocking layer and optimum design for EPROM device.

3. To hypothesize the dependency of dielectric constant "*k*" on memory window of an EPROM.

1.4.2 Scopes

The research scopes are described as follows,

1. Develop floating gate based EPROM implementing High *k* blocking layer by simulation. Use SILVACO Athena Simulator for this purpose
2. Characterize I/V, Retention and Endurance Reliability. Use SILVACO Atlas Simulator for the purpose.
3. Use Origin Pro to plot the data.

1.5 Organization

The structure of the report is organized as follows. Chapter 2 discusses the details on the literature review. Chapter 3 describes the research Methodology, which will be followed by Results and Discussion at chapter 4. The whole thesis will end with a conclusion.

REFERENCES

1. Rapid growth of artificial intelligence. In *http://www.practicalpd.com.au/blog/ideas-news/the-rapid-growth-of-artificial-intelligence/*.

2. Y Ahn, J-D Choe, J Lee, Donguk Choi, E Cho, B Choi, S-H Lee, S-K Sung, C-H Lee, S Cheong, et al. Trap layer engineered finfet nand flash with enhanced memory window. In *VLSI Technology, 2006. Digest of Technical Papers. 2006 Symposium on*, pages 88–89. IEEE, 2006.

3. Y Ahn, J-D Choe, J Lee, Donguk Choi, E Cho, B Choi, S-H Lee, S-K Sung, C-H Lee, S Cheong, et al. Trap layer engineered finfet nand flash with enhanced memory window. In *VLSI Technology, 2006. Digest of Technical Papers. 2006 Symposium on*, pages 88–89. IEEE, 2006.

4. Guoxing Chen, Zongliang Huo, Lei Jin, Dong Zhang, Shengjie Zhao, Yulong Han, Su Liu, and Ming Liu. Low temperature atomic layer deposited hfo2 film for high performance charge trapping flash memory application. *Semiconductor Science and Technology*, 29(4):045019, 2014.

5. Eun Suk Cho, Choong-Ho Lee, Tae-Yong Kim, Suk-Kang Sung, Byung Kyu Cho, Chul Lee, Hye Jin Cho, Yonghan Roh, Donggun Park, Kinam Kim, et al. Hf-silicate inter-poly dielectric technology for sub 70nm body tied finfet flash memory. In *VLSI Technology, 2005. Digest of Technical Papers. 2005 Symposium on*, pages 208–209. IEEE, 2005.

6. J-D Choe, YJ Ahn, S-H Lee, D Jang, Y-B Yoon, JJ Lee, I Chung, K Park, and D Par. Hybrid nanocrystal finfet with large p/e window for mlc nand flash memory application. *Electronics Letters*, 43(9):545–546, 2007.

7. Robert D Clark. Emerging applications for high k materials in vlsi technology. *Materials*, 7(4):2913–2944, 2014.

8. C Friederich, M Specht, T Lutz, F Hofmann, L Dreeskornfeld, W Weber, J Kretz, T Melde, W Rosner, E Landgraf, et al. Multi-level p+ tri-gate sonos nand string arrays. In *Electron Devices Meeting, 2006. IEDM'06. International*, pages 1–4. IEEE, 2006.

9. Changjie Gong, Qiaonan Yin, Xin Ou, Xuexin Lan, Jinqiu Liu, Chong Sun, Laiguo Wang, Wei Lu, Jiang Yin, Bo Xu, et al. The dominant factors affecting the memory characteristics of (ta2o5) x (al2o3) 1- x high-k charge-trapping devices. *Applied Physics Letters*, 105(12):123504, 2014.

10. Changjie Gong, Qiaonan Yin, Xin Ou, Xuexin Lan, Jinqiu Liu, Chong Sun, Laiguo Wang, Wei Lu, Jiang Yin, Bo Xu, et al. The dominant factors affecting the memory characteristics of (ta2o5) x (al2o3) 1- x high-k charge-trapping devices. *Applied Physics Letters*, 105(12):123504, 2014.

11. Digh Hisamoto, Wen-Chin Lee, Jakub Kedzierski, Hideki Takeuchi, Kazuya Asano, Charles Kuo, Erik Anderson, Tsu-Jae King, Jeffrey Bokor, and Chenming Hu. Finfet-a self-aligned double-gate mosfet scalable to 20 nm. *IEEE Transactions on Electron Devices*, 47(12):2320–2325, 2000.

12. XD Huang, PT Lai, L Liu, and JP Xu. Nitrided srtio 3 as charge-trapping layer for nonvolatile memory applications. *Applied physics letters*, 98(24):242905, 2011.

13. XD Huang, Johnny KO Sin, and PT Lai. Improved charge-trapping characteristics of batio3 by zr doping for nonvolatile memory applications. *IEEE Electron Device Letters*, 34(4):499–501, 2013.

14. 2 Seung-Won Ko 2 Chang-Hoon Kim 2 Jee-Yeon Kim 2 Moon-Seok Kim 2 Myeong-Lok Seol 2 Joon-Bae Moon 2 Ji-Min Choi 2 Jae-Sub Oh 3 Sung-Jin Choi 4 Jae-Hyuk Ahn, 1 Dong-Il Moon and Yang-Kyu Choi2. A sonos device with a separated charge trapping layer for improvement of charge injection. *American Institute of Physics*, 4(3):243, 2005.

15. K Jiang, X Ou, XX Lan, ZY Cao, XJ Liu, W Lu, CJ Gong, B Xu, AD Li, YD Xia, et al. Remarkable charge-trapping efficiency of the memory device with (tio2) 0.8 (al2o3) 0.1 composite charge-storage dielectric. *Applied Physics Letters*, 104(26):263506, 2014.

16. Takahiro Kamei, Yongxun Liu, Takashi Matsukawa, Kazuhiko Endo, Shinichi O'uchi, Junichi Tsukada, Hiromi Yamauchi, Yuki Ishikawa, Tetsuro Hayashida, Kunihiro Sakamoto, et al. Demonstration of split-gate type trigate flash memory with highly suppressed over-erase. *IEEE Electron Device Letters*, 33(3):345–347, 2012.

17. Xuexin Lan, Xin Ou, Yanqiang Cao, Shiyu Tang, Changjie Gong, Bo Xu, Yidong Xia, Jiang Yin, Aidong Li, Feng Yan, et al. The effect of thermal treatment induced inter-diffusion at the interfaces on the charge trapping performance of hfo2/al2o3 nanolaminate-based memory devices. *Journal of Applied Physics*, 114(4):044104, 2013.

18. Se Hoon Lee, Jong Jin Lee, Jeong-Dong Choe, Eun Suk Cho, Young Joon Ahn, Won Hwang, Taeyong Kim, Woo-Jung Kim, Young-Bae Yoon, Donghoon Jang, et al. Improved post-cycling characteristic of finfet nand flash. In *Electron Devices Meeting, 2006. IEDM'06. International*, pages 1–4. IEEE, 2006.

19. Tackhwi Lee and Sanjay K Banerjee. Device characteristics of hfon charge-trap layer nonvolatile memory. *Journal of Vacuum Science Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, 28(5):1005–1010, 2010.

20. SH Lin, Albert Chin, FS Yeh, SP McAlister, et al. Good 150 degrees c retention and fast erase characteristics in charge-trap-engineered memory having a scaled si3n4 layer. In *IEEE INTERNATIONAL ELECTRON DEVICES MEETING 2008, TECHNICAL DIGEST*, pages 843–+, 2008.

21. Yongxun Liu, Toshihide Nabatame, Takashi Matsukawa, Kazuhiko Endo, Shinichi O'uchi, Junichi Tsukada, Hiromi Yamauchi, Yuki Ishikawa, Wataru Mizubayashi, Yukinori Morita, et al. Comparative study of charge trapping type soi-finfet flash memories with different blocking layer materials. *Journal of Low Power Electronics and Applications*, 4(2):153–167, 2014.

22. Yongxun Liu, Toshihide Nabatame, Takashi Matsukawa, Kazuhiko Endo, Shinichi O'uchi, Junichi Tsukada, Hiromi Yamauchi, Yuki Ishikawa, Wataru Mizubayashi, Yukinori Morita, et al. Comparative study of charge trapping type soi-finfet flash memories with different blocking layer materials. *Journal*

of Low Power Electronics and Applications, 4(2):153–167, 2014.

23. Yongxun Liu, Toshihide Nabatame, Takashi Matsukawa, Kazuhiko Endo, Sinichi O'uchi, Junichi Tsukada, Hiromi Yamauchi, Yuki Ishikawa, Wataru Mizubayashi, Yukinori Morita, et al. Experimental study of charge trapping type finfet flash memory. In *Nanoelectronics Conference (INEC), 2014 IEEE International*, pages 1–4. IEEE, 2014.

24. YX Liu, T Kamei, T Matsukawa, K Endo, S O'uchi, J Tsukada, H Yamauchi, Y Ishikawa, T Hayashida, K Sakamoto, et al. Experimental study of tri-gate soi-finfet flash memory. In *SOI Conference (SOI), 2012 IEEE International*, pages 1–2. IEEE, 2012.

25. YX Liu, T Kamei, T Matsukawa, K Endo, J Tsukada, H Yamauchi, Y Ishikawa, T Hayashida, K Sakamoto, A Ogura, et al. Comparative study of tri-gate-and double-gate-type poly-si fin-channel split-gate flash memories. In *Silicon Nanoelectronics Workshop (SNW), 2012 IEEE*, pages 1–2. IEEE, 2012.

26. YX Liu, T Nabatame, T Matsukawa, K Endo, S O'uchi, J Tsukada, H Yamauchi, Y Ishikawa, W Mizubayashi, Y Morita, et al. Charge trapping type finfet flash memory with al 2 o 3 blocking layer. In *SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S), 2013 IEEE*, pages 1–2. IEEE, 2013.

27. Toshihide Nabatame, Akihiko Ohi, Kazuhiro Ito, Makoto Takahashi, and Toyohiro Chikyo. Role of the (ta/nb) ox/al2o3 interface on the flatband voltage shift for al2o3/(ta/nb) ox/al2o3 multilayer charge trap capacitors. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 33(1):01A118, 2015.

28. Tung-Ming Pan and Wen-Wei Yeh. High-performance high k sonos-type flash memory. *IEEE Transactions on Electron Devices*, 55(9):2354–2360, 2008.

29. J Razafindramora, Luca Perniola, C Jahan, P Scheiblin, M Gely, C Vizioz, C Carabasse, F Boulanger, Barbara De Salvo, S Deleonibus, et al. Low voltage hot-carrier programming of ultra-scaled soi finflash memories. In *Solid State Device Research Conference, 2007. ESSDERC 2007. 37th European*, pages 414–417. IEEE, 2007.

30. S-K Sung, S-H Lee, B Choi, J Lee, J-D Choe, E Cho, Y Ahn, Donguk Choi, C-H Lee, D Kim, et al. Sonos-type finfet device using p+ poly-si gate and high-k blocking dielectric integrated on cell array and gsl/ssl for multi-gigabit nand flash memory. In *VLSI Technology, 2006. Digest of Technical Papers. 2006 Symposium on*, pages 86–87. IEEE, 2006.

31. Yan Ny Tan, Wai Kin Chim, Wee Kiong Choi, Moon Sig Joo, and Byung Jin Cho. Hafnium aluminum oxide as charge storage and blocking-oxide layers in sonos-type nonvolatile memory for high-speed operation. *IEEE Transactions on Electron Devices*, 53(4):654–662, 2006.

32. S Tsuno, T Kaneko, Yoshimasa Kurihara, Shigeru Odaka, and K Kato. Gr@ ppa 2.7 event generator for collisions. *Computer physics communications*, 175(10):665–677, 2006.

33. Xuguang Wang, Jun Liu, Weiping Bai, and Dim-Lee Kwong. A novel monos-type nonvolatile memory using high-/spl kappa/dielectrics for improved data retention and programming speed. *IEEE Transactions on Electron Devices*, 51(4):597–602, 2004.

34. CY Wei, B Shen, P Ding, P Han, AD Li, YD Xia, B Xu, J Yin, and ZG Liu. Ta 2 o 5-tio 2 composite charge-trapping dielectric for the application of the nonvolatile memory. *Scientific Reports*, 7(1):5988, 2017.

35. D.A.; Bu-J. White, M.H.; Adams. On the go with sonos. In *IEEE Circuits Devices Mag*, pages 16,22â31. IEEE, 2000.

36. Yung-Hsien Wu, Lun-Lun Chen, Jia-Rong Wu, Min-Lin Wu, Chia-Chun Lin, and Chia-Hsuan Chang. Nonvolatile memory with nitrogen-stabilized cubic-phase zro2 as charge-trapping layer. *IEEE Electron Device Letters*, 31(9):1008–1010, 2010.

37. Peiqi Xuan, Min She, Bruce Harteneck, Alex Liddle, Jeffrey Bokor, and T-J King. Finfet sonos flash memory for embedded applications. In *Electron Devices Meeting, 2003. IEDM'03 Technical Digest. IEEE International*, pages 26–4. IEEE, 2003.

38. HJ Yang, CF Cheng, WB Chen, SH Lin, FS Yeh, Sean P McAlister, and Albert Chin. Comparison of monos memory device integrity when using trapping

layers with different n compositions. *IEEE Transactions on Electron Devices*, 55(6):1417–1423, 2008.

39. Mu-Shih Yeh, Yung-Chun Wu, Min-Feng Hung, Kuan-Cheng Liu, Yi-Ruei Jhan, Lun-Chun Chen, and Chun-Yen Chang. Fabrication, characterization and simulation of -gate twin poly-si finfet nonvolatile memory. *Nanoscale research letters*, 8(1):331, 2013.

40. Mu-Shih Yeh, Yung-Chun Wu, Min-Feng Hung, Kuan-Cheng Liu, Yi-Ruei Jhan, Lun-Chun Chen, and Chun-Yen Chang. Fabrication, characterization and simulation of ω-gate twin poly-si finfet nonvolatile memory. *Nanoscale research letters*, 8(1):331, 2013.

41. Hee-Wook You and Won-Ju Cho. Charge trapping properties of the hfo 2 layer with various thicknesses for charge trap flash memory applications. *Applied Physics Letters*, 96(9):093506, 2010.

42. Hao Zhu, Hui Yuan, Haitao Li, Curt A Richter, Oleg Kirillov, Dimitris E Ioannou, and Qiliang Li. Design and fabrication of ta2o3 stacks for discrete multibit memory application. *IEEE Transactions on Nanotechnology*, 12(6):1151–1157, 2013.