ANALYSIS OF CHARGE TRAP NAND FLASH MEMORY FOR IMPROVED RELIABILITY

A THESIS

Submitted in partial fulfillment of the
requirements for the award of the degree

of

DOCTOR OF PHILOSOPHY

by

DEEPIKA GUPTA

DISCIPLINE OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JUNE 2017

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled "**ANALYSIS OF CHARGE TRAP NAND FLASH MEMORY FOR IMPROVED RELIABILITY**" in the partial fulfillment of the requirements for the award of the degree of **DOCTOR OF PHILOSOPHY** and submitted in the **DISCIPLINE OF ELECTRICAL ENGINEERING**, **Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from January 2014 to June 2017 under the supervision of Dr. Santosh Kumar Vishvakarma, Associate Professor, Indian Institute of Technology Indore, India.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Signature of the Student with date

(DEEPIKA GUPTA)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Signature of Thesis Supervisor with date

(Dr. SANTOSH KUMAR VISHVAKARMA)

DEEPIKA GUPTA has successfully given her Ph.D. Oral Examination held on 21 September 2017.

Signature of Chairperson (OEB) Signature of External Examiner Signature of Thesis Supervisor

Date: Date: Date:

Signature of PSPC Member:1 Signature of PSPC Member:2 Signature of Convener, DPGC

Date: Date: Date:

Signature of Head of Discipline

Date:

ACKNOWLEDGEMENTS

Foremost, I thank God Almighty for enlightening my mind to undertake this research work. Then, I would like to express my sincere gratitude to my supervisor Dr. Santosh Kumar Vishvakarma for his sustained inspiration, guidance, and support throughout my research work. During the entire research process, he constantly encouraged me to focus on most important and challenging problems, and stimulated me to find their innovative and creative solutions. It was my pleasure to work under the esteemed guidance of such an enthusiastic and responsible supervisor.

I am grateful to my PSPC committee members Dr. M. Anbarasu and Dr. Abhishek Shrivastava for their valuable comments and suggestions during my research. Special thanks to the Director Dr. Pradeep Mathur, all faculty members and the staffs at Indian Institute of Technology Indore for their cooperation throughout my thesis work. In addition, my appreciation goes to all the members of nanoscale devices, circuits and system design lab for creating a conducive environment. It was my privilege to share my working space with responsible colleagues Mr. Ankur Beohar, Mr. Abhishek Upadhyay, Ms. Pooja Bohara, Mr. M. Gopal, Mr. Gaurav Singh, Mr. Ravi Kumar, Mr. Mahesh Kumawat, Mr. Vishal Sharma, Mr. Ambika Prasad Shah and Mr. Sajid Khan . Also, I would like to thank my seniors Dr Vikas Vijayvargiya, Dr Bhupendra Singh Reniwal, Dr. Chandrabhan Kushwah, Mr. Pooran Singh, Dr. Nand Kishore Yadav who helped me in several ways during my thesis work.

Above all, I am indebted to my parents Mr. Suresh Chand Gupta and Mrs. Sharda Gupta, parents-in-law Mr. Roop Kishore Sharma and Mrs. Bhagwan Devi Sharma, brother Mr. Manoj Gupta and sisters-in-law Mrs. Bhawana Gupta and Ms. Pavitra Sharma for their constant encouragement and assistance. Most importantly, this journey would have not been possible without the love and support of my husband Dr. Pankaj Kumar Sharma. His constant motivation through this time inspired me to achieve greatest milestones in life.

DEEPIKA GUPTA

Dedicated to

My Husband

Dr. Pankaj Kumar Sharma

&

My Parents

Mr. Suresh Chand Gupta & Mrs. Sharda Gupta

ABSTRACT

Hard disk drives (HDD) have immensely been used as external storage device in computer system for around half the century. However, HDDs are being lesser appealing these days due to its long read/write access latency, high power consumption and fragility. To provide the alternative method for external storage, flash memory was invented by Toshiba in 1980's. Flash Memories are more reliable and power efficient than HDD which makes it more suitable for battery operated hand-hold devices such as smart phones, tablets, notebooks etc. Importantly, the flash memory cell stores the electrical charge in floating gate (FG) or charge trap (CT) layer. Further, categorisation of NAND and NOR flash memory depends on the circuit connections of flash memory cells. Here, NOR flash memory can provide fast random access hence shows its suitability for code storage. However, with NOR technology, larger cell size and difficulty to scale the flash cell size makes it less suitable for bulk data storage. On the other hand, smaller cell size with NAND technology allows bulk data storage with small cost per bit as compared to NOR technology for same die area. Therefore, now a days, NAND flash memory is used in majority of consumer applications such as digital video/music player, Flash drives, MP3 players, multi-function cell phones, digital cameras and USB drives etc for bulk data storage. The large application market of NAND flash memory makes it the most significant non-volatile memory solution for the next decade.

Further, to satisfy its demand in handhold applications, the NAND flash memory must be able to provide high data storing capacity with low cost per bit. For the realisation of such cost effective and high density NAND flash memory, scaling of memory cell size is important. Owing to this scaling of cell size, high capacity NAND flash memory can be obtained on the same die area. In this regard, it has already been predicted that half pitch of the NAND flash memory cell in two dimensional (2D) integration would be scaled below 10nm by 2025. These extremely scaled memory cells will impose several limitations on the performance of the flash memory such as short channel effects (SCEs) and reduced data retention etc. In fact, while scaling, a flash memory cell has been shown to be more prone to the SCE than any other logic device due to the requirement of thick gate oxide stack (Tunnel Oxide + FG/CT + Inter-Poly

Dielectric thickness). High thickness of oxide stack is required to achieve high data retention. However, small thickness of gate oxide stack is needed to have improved SCEs. This trade-off between SCEs and data retention presents bottleneck for future scaling of flash memory. Therefore, in this thesis, we explore a flexible method that can improve SCEs as well as data retention without altering the gate oxide stack thickness. At the same time, by source/drain engineering, we show that junction boost leakage current can be reduced for CT flash memory cell with improved SCEs. Further, in this thesis, we investigate effect of lightly doped drain (LDD) depth variation on the reliability of CT based NAND flash Memory. Here, to monitor the reliability, we investigate residual charge in CT layer after erase operation and then its effect on the endurance performance of the flash memory cell.

In the same walk, to scale the NAND flash memory efficiently with the reduced cost per bit, several three dimensional (3D) NAND flash memory structure have been proposed in the literature. Here, to obtain high density NAND flash memory on the same area, memory array is realised in the vertical direction. These 3D NAND flash memories are faster in operation, has improved wear life and has lower bit error rate (BER) per KB of data than 2D flash. These 3D integration with NAND flash memory serves as the basic idea for the system known as Solid State Drives (SSD). A 3D structure can be implemented with two configurations i.e. Vertical Gate (VG) and Vertical Channel (VC) according to the direction of current flow. In this thesis, we use a single string of NAND flash memory cell having VG configuration which seems more appealing in terms of pitch scaling, selection of channel material and degradation of read current with large number of stacking layers, as compared to the VC configuration. In addition, in this thesis, we explore junction-free architecture of NAND flash memory string. Further, this thesis analyses the effect of channel engineering method to improve the data retention-SCE tradeoff in CT based VG junction-free NAND flash memory string without altering the gate oxide stack. Also, we explore the effect of V_{PASS} voltage to other adjacent cells in NAND string.

Above all, the theoretical findings in this thesis provide useful insights and guidelines for the design of reliable flash memory systems.

Contents

List of Figures

List of Symbols

Notation	Definition
C_{ONO}	= CG to FG Capacitance
C_{TOT}	= Total Capacitance of FG
$V_{th,programmed}$	= Threshold Voltage of Programmed Cell
$V_{th,erased}$	= Threshold Voltage of Erased Cell
$V_{th,read}$	= Threshold Voltage of Read cell
Ψ_{GS}	= Gate to Source Work Function
Ψ_s	= Fermi Potential
d	= LDD Depth
Q_{SS}	= Fixed Charge at Substrate-TO Interface
Q_S	= Charge of Silicon Substrate
Q_T	= Charge Trapped at Trapping Layer
d_T	= Distance of Trapped Charge
n_{pe}	= Number of Program/Erase Cycle Until the Breakdown
Q_{bd}	= Amount of Charge Crossed the Tunnel Oxide till Breakdown
ΔV_{th}	= Threshold Voltage Roll-Off
σ_L	= Source/Drain Lateral Straggle
V_{select}	= Voltage applied to select memory cell in NAND string
V_{BL}	= Voltage applied to BL
V_{PASS}	= Voltage applied to PASS memory cells in NAND string

List of Abbreviations

2D	Two dimensional
3D	Three dimensional
EEPROM	Electrically Erasable Programmable Read only Memory
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
TO	Tunnel Oxide
IPD	Inter Poly Dielctric
CG	Control Gate
k	Dielectric constant
PolySi	Poly Silicon
PDA	Personal Digital Assistance
USB	Universal Serial Buffer
SD	Secure Digital
MMC	Multi Media Card
HHD	Hybrid Hard Drives
FG	Floating Gate
CT	Charge Trap
HDD	Hard Disk Drive
SNVM	Semiconductor Non-volatile Memory
SSD	Solid State Drive
SILC	Stress Induced Leakage Current
SONOS	Silicon Oxide Nitride Oxide Silicon
ITRS	International Technology Roadmap for Semiconductor
SCE	Short Channel Effect
EOT	Effective Oxide Thickness
USJ	Ultra Shallow Junction
P-I	Program-Inhibit
S	Source
D	Drain
SS	Subthreshold Swing
DIBL	Drain Induced Barrier Lowering

FN	Fowler Nordheim
CHE	Channel Hot Electron
WL	Word Line
BL	Bit Line
LDD	Lightly Doped Drain
SSL	Source Select Line
GSL	Ground Select Line
SRH	Shockley-Read-Hall

Chapter 1

Introduction

In recent years, the communication over internet have experienced revolutionary rise in the number of users, applications and services. This evolution of machine to machine interaction produces high amount of data that needs to be stored for future purposes. Therefore, a non-volatile bulk storage memory solution is required to assist with data storage. Also, the applications such as wireless phones, Personal Digital Assistants (PDAs), portable music players, USB drives, digital cameras, Secure Digital (SD) cards, MultiMedia Cards (MMCs) and Flash drives etc require non-volatile bulk data storage memory for their operation. In this regard, the Hard Disk Drives (HDD) have immensely been used as external bulk data storage device with such applications for around half the century now. HDD consist a platter coated with magnetic material to store the data. In fact, a mechanical moving arm is used to find the location to be addressed. Importantly, the capability of HDD to deliver high storage capacity with reduced per bit cost made it popular option for external data storage [1]. However, HDDs are being lesser appealing these days due to its long read/write access latency, high power consumption and fragility. Therefore, to provide the alternative method for external data storage, flash memory was invented in 1980's by Toshiba [2]. Flash memory consist a Electrically Erasable Programmable Read Only Memory (EEPROM) cell for electronic data storage and gets implemented on semiconductor based Integrated Circuits (IC's). Compared to the HDD, flash memories are more reliable with greater performance, provides almost instantaneous data access, with quicker boot ups, higher

shock resistance, faster file transfers, and an overall snappier computing experience. Thus, despite their higher cost per bit with respect to magnetic HDD, flash memories result as the winner in all the consumer products requiring light weight, low size, low power consumption and high reliability non-volatile memory solution.

1.1 Flash Memory: Brief Overview

As a type of the non-volatile storage, flash memory is able to retain the data even when the power supply of the device is switched off. As we know, a flash memory cell is a type of EEPROM cell [2] however, the EEPROM is a byte erasable memory whereas flash memory erases data in the block. A group of flash memory cells is called as block in flash memory array. This block erasure characteristics of flash memory makes it suitable for the bulk data storage applications which needs to get frequently updated. Basically, a flash memory cell architecture closely aligns to the Metal Oxide Semiconductor Field Effect Transistor (MOSFET)[3]. The only difference is the additional layer to trap the electronic charge in between Control Gate (CG) and channel (substrate). The difference of the MOSFET and flash memory cell can be clearly seen from Figure 1.1 (a) and (b).

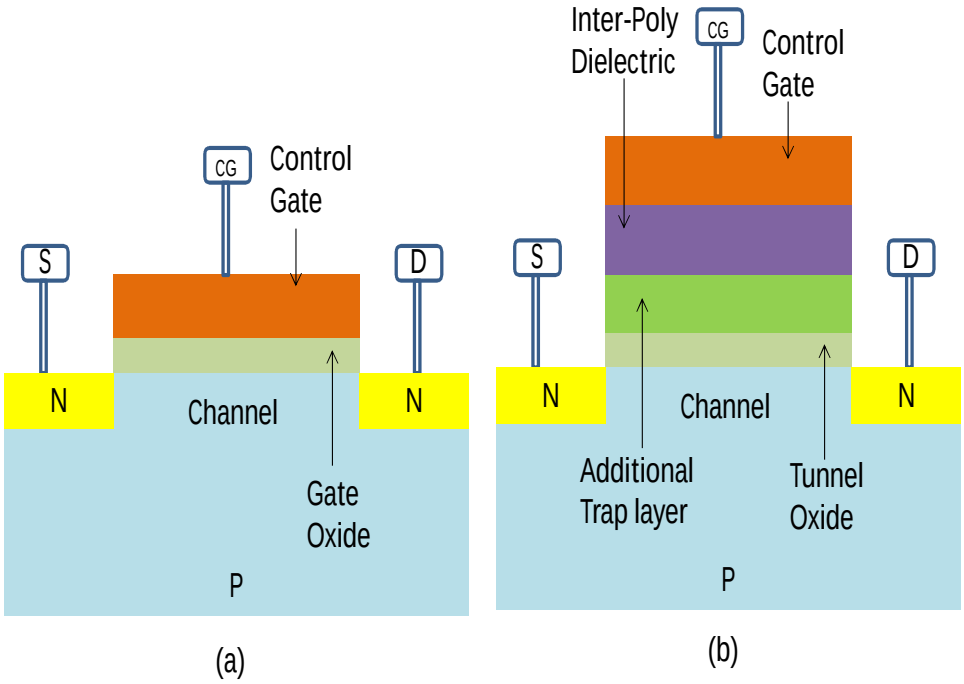

Figure 1.1: (a) A MOSFET cell (b) A flash memory cell

Here, the MOSFET cell consist a thin dielectric material layer between CG and channel which is called as the gate oxide. However, the flash memory cell consist a stack between the CG and channel. This stack consist a Tunnel Oxide (TO) at the bottom, additional layer to trap the charge in the middle, and Inter-Poly Dielectric (IPD) at the top. The materials for the dielectric stack should be chosen such that it helps to enhance electric field at TO. These layers have following operations:

- **TO:** TO can be made up of any dielectric material having small dielectric constant (k). Further, TO has two primary functions:

  1. It is used to tunnel electrons during the program and erase operations. Importantly, the thickness of the TO determines the tunneling current density.
  2. It also determines the threshold voltage of the cell during the read operation.

- **Additional Trap Layer:** Additional trap layer can be made up of either polysilicon (PolySi)or Metal and any high-k dielectric material. Further, The trap layer is completely surrounded by the dielectric materials (TO & IPD) and is used to store charge which alters the threshold voltage of the memory cell.

- **IPD:** This is simply an insulator with high k value which is used to reduce the back tunneling of stored charge from trap layer into the CG. Thickness and k value of IPD layer also determine the electric field at TO.

- **CG:** This is the conductive material used to access the flash memory cell.

Further, low power consumption, high shock resistance (physical resilience/ruggedness), small size, light weight, and fast data access are few of the reasons for the popularity of flash memory and its demand as a non-volatile data storage solution for portable electronics such as wireless phones, PDAs, portable music players, USB drives, digital cameras, SD cards, MMCs, and many other applications. In addition, flash memory is also used in Hybrid Hard Drives (HHDs). Basically, HHDs employ a large buffer of non-volatile flash memory used to cache data during normal use due to which the platters of the hard drive are at rest almost at all times, instead of constantly spinning as is the case in HDDs. This feature offers several benefits such as decreased power consumption, improved reliability, and a faster boot process compared to HDD.

Moreover, flash memory cell can be realized with two cell architecture viz., Floating Gate (FG) flash memory cell [3] and Charge Trap (CT) flash memory cell [4], based on the materials used for additional trap layer.

1.1.1 FG Flash Memory Cell

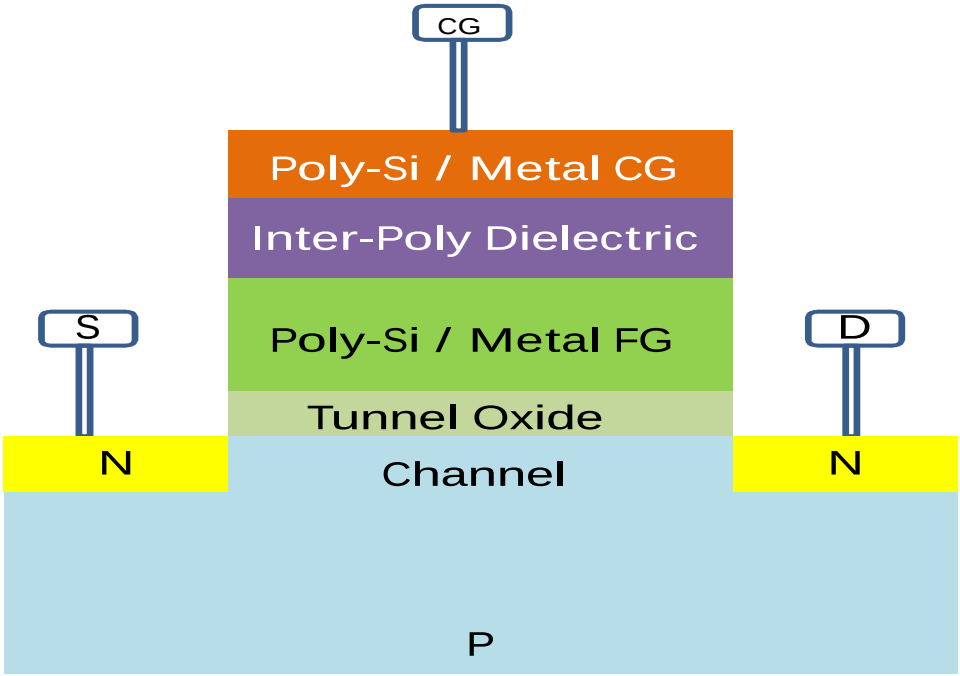

Figure 1.2: A FG flash memory cell

A FG flash memory cell uses metal/polySi as material for additional trap layer, as shown in Figure 1.2. Here, the additional trap layer is called as FG because of its metallic nature same as CG. Also, the FG (between CG and channel) is electrically insulated from all the inputs hence it is considered as floating. Here, Electrical insulation of FG is achieved through the gate oxides (TO & IPD). The CG is placed above the FG, which decide the threshold voltage of the flash cell by modifying the amount of charges in the FG. Importantly, a FG flash memory cell stores a bit of the information by injecting a certain amount of electrical charge into FG. Here, with the application of high positive voltage at CG terminal, a high electric field is produced at the TO of the gate stack. Under the influence of this high electric field, electrons from the channel get attracted vertically to CG. While travelling to CG, electrons get trapped at the

FG due to high IPD barrier. By injecting charge in the FG of the flash memory cell, the threshold voltage of the cell is increased, and by removing it, the previous state can be restored [3]. Detailed discussion about different modes of operations in flash memory is carried out later in this chapter. Here, following the previous discussion, the threshold voltage of the FG cell is used to store a bit of information in flash memory cell. Importantly, the FG forms a deep well for charges trapped in it due to the gate dielectrics around it when power is shut down. In this manner, FG flash memory works as a non-volatile memory. Moreover, in order to get electron de-trapped from the FG, application of negative voltage is required to repel the electron back into the channel.

However, FG flash memory has following disadvantages when used commercially:

- FG flash memory is unable to reduce unit cell area. Here, due to stored mobile charges at FG, scaling of FG with TO and IPD thickness gets limited from the amount of leakage through TO and IPD [5].

- Retention of FG cell is highly sensitive to defects in TO. While trying to scale TO below 7nm, severe Stressed Induced Leakage Current (SILC) is observed in FG flash memory cell [6].

- High cell to cell interference is observed from adjacent memory cells in flash memory array due to metallic nature of FG. The cell to cell interference causes shift of a cell threshold voltage proportional to the change of the threshold voltage of the adjacent cell. It results from capacitive coupling via parasitic capacitors around the FG [7].

- Reduced Gate Coupling Ration (GCR) between CG and FG. The gate coupling ration defines the amount of FG voltage as compared to the CG voltage. It is can be given as [8]

$$GCR = \frac{C_{ONO}}{C_{TOT}}$$

  where C_{ONO} is the CG to FG capacitance and C_{TOT} is the total capacitance of the FG. Importantly, Higher gate coupling ration is required for power and

performance efficient FG flash memory cell.

1.1.2 CT Flash Memory Cell

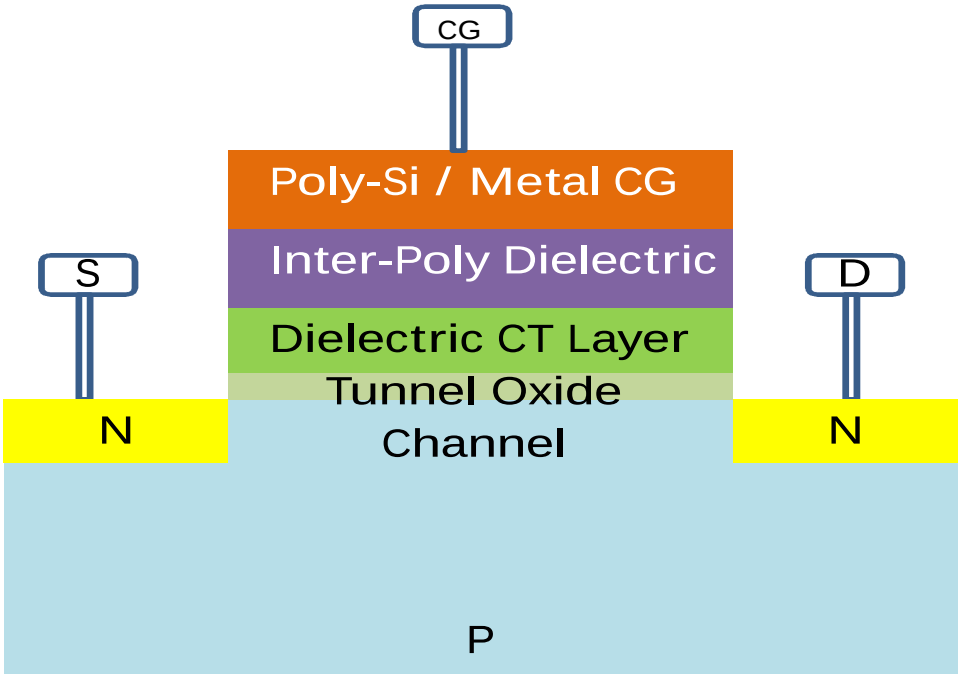

Figure 1.3: A CT flash memory cell.

A CT flash memory cell utilises high k dielectric material as additional trap layer, shown in Figure 1.3. Here, the dielectric CT layer stores the injected charge in the traps/defects present in dielectric [4]. Importantly, CT flash memory cell, precisely, Silicon Oxide Nitride Oxide Silicon (SONOS) has attracted a great deal of recent research to replace FG flash memory cell due to its MOSFET compatible structure in terms of involved fabrication process, low voltage operation and the immunity to extrinsic charge loss [5]. Recently, market for CT flash memory cell has increased with evolution of 3D NAND flash memories. 3D NAND flash memories mostly employs CT flash memory cell because of its high scalability and capability to stack over its counterpart FG flash memory [9]-[11]. Therefore, CT flash is a promising candidate to increase the scaling limit of flash memory [12]. Further, the operating principle for CT flash memory is similar as already discussed with FG flash memory cell. In CT flash memory, voltage at CG attracts charge from the channel which get trapped within the traps/defects of

CT layer. Interestingly, the CG voltage requirement is lenient in CT flash memory cell as compared to the counterpart. This effect is attributed to the reduced gate oxide thickness with CT flash memory as compared to FG. Further, the deep traps in CT layer along with the energy well of gate oxide stack provides the required data retention for non-volatility.

Advantages of CT Flash Memory

- In CT flash memory, TO thickness can be scaled down along with CT and IPD thickness, hence allows aggressive vertical scaling as compared to FG flash memory cell. In FG flash memory, thick FG is required to store the sufficient charge along with the thick TO to avoid SILC. However, in CT flash memory, charge in CT layer are stored in discrete traps/defects sites. Hence the thickness of CT and TO layer can be relieved compared to the FG one. The vertical scaling comparison between CT and FG flash memory cells is shown in Figure 1.4.

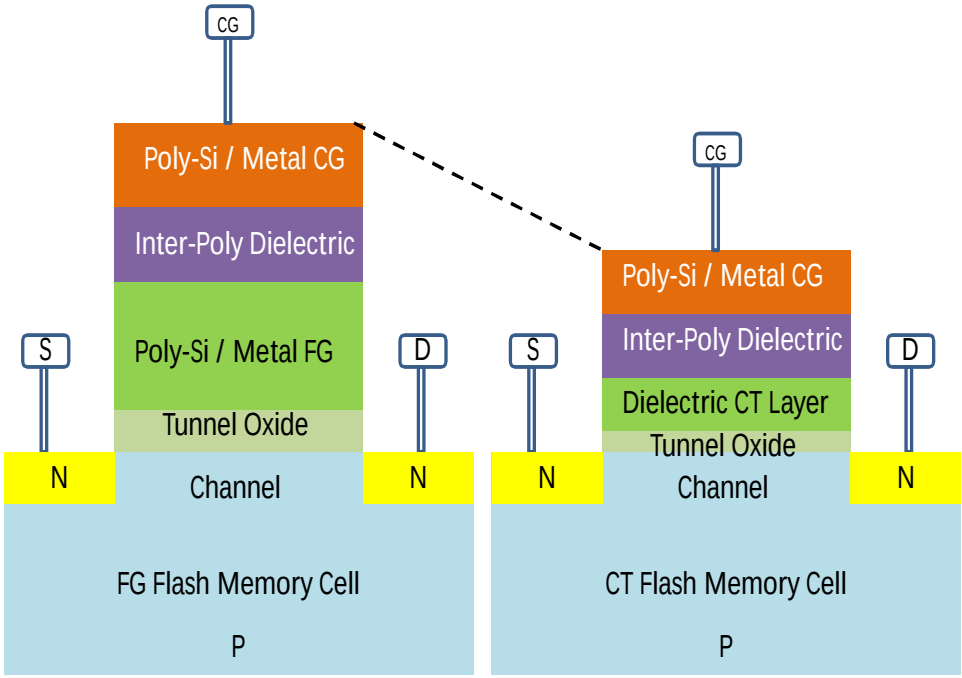

Figure 1.4: Comparison of vertical scaling between FG and CT flash memory cell

- The CT flash memory cell has higher immunity to TO defect and SILC, hence more reliable than FG.

- Due to discrete trap storage, CT flash memory is suitable for multi level cell programming hence, helps to obtain higher density.
- CT flash memory offers reduced coupling interference between adjacent cells due to absence of metallic FG. Owing to this, CT flash cells can be placed closer to each other than the FG fellow. Hence, CT flash cell aids in further optimising the area for flash memory.
- Due to almost negligible cell to cell interference, CT flash memory cells are best suitable for 3D architectures (Gate All Around, Vertical Gate)[13].

1.2 Flash Memory Cell: Operations

A flash memory cell may operate in 4 modes, i.e., 1. Program operation, 2. Erase Operation, 3. Read Operation & 4. Program-inhibit Operation.

1.2.1 Program Operation

A flash memory cell is programmed (set to a 0 state) by applying a large positive voltage on the CG [14]. This provides an high positive electric field at TO which is strong enough to tunnel electrons from the channel (substrate) into the FG/CT layer. Injection of charge into FG/CT layer changes threshold voltage of the flash memory cell according to the following equation.

$$V_{th} = \phi_{GS} + 2\phi_F + \frac{Q_{SS} + Q_S}{C_I} - \frac{Q_T d_T}{\varepsilon_I} \tag{1.1}$$

Here, V_{th} is Threshold voltage of the flash memory cell, ϕ_{GS} and ϕ_F are the gate to source work function and fermi potential respectively, Q_{SS} and Q_S are the fixed charge at substrate-TO interface and charge of silicon body respectively, Q_T is the Trapped Charge in trapping layer at d_T distance from Gate and C_I and ϵ_I are the capacitance and dielectric constant of tunneling layer respectively. Hence, any increase in charge

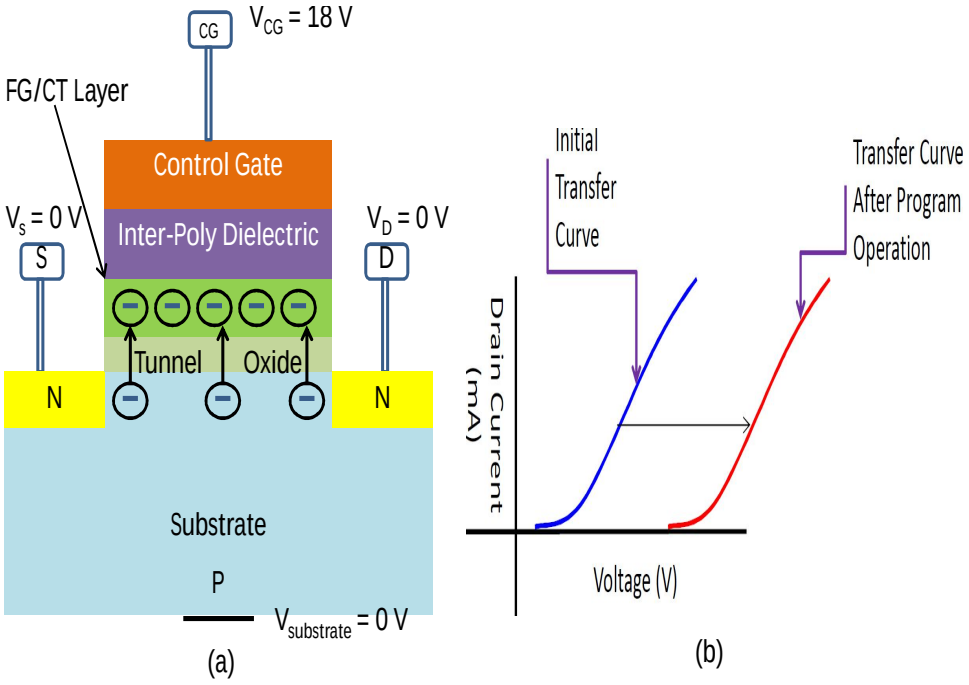

Figure 1.5: (a) Program operation (b) Modified transfer characteristics after program operation

Q_T at FG/CT layer will increase the threshold voltage of the cell to $V_{th-program}$. The operating condition during program operation and modified transfer characteristics of NAND flash cell after program operation is shown in Figure 1.5. Here, it is important to note the program threshold voltage ($V_{th-programmed}$) must be as high as possible to attain the maximum trapping of charge carrier at FG/CT layer. Further, with program operation, the tunneling of charge carrier across the TO, may exploit either Channel Hot Electrons (CHE) injection or Fowler-Nordheim (FN) tunneling mechanism [14].

CHE Injection

The CHE injection originates from the heating of the channel carriers, travelling from the source to the drain, due to the lateral electric field ($E_{||}$) produced by the applied drain-source voltage. Figure 1.6 depicts the bias condition of flash memory cell with energy bands orientations during CHE injection. Here, the influence of the drain bias (V_D) on the surface potential (φ_s) reduces the conductivity of the channel near the drain edge, thus increasing the lateral potential drop in the drain region. Therefore, $E_{||}$ presents a larger value in proximity of the drain diffusion and the channel carriers

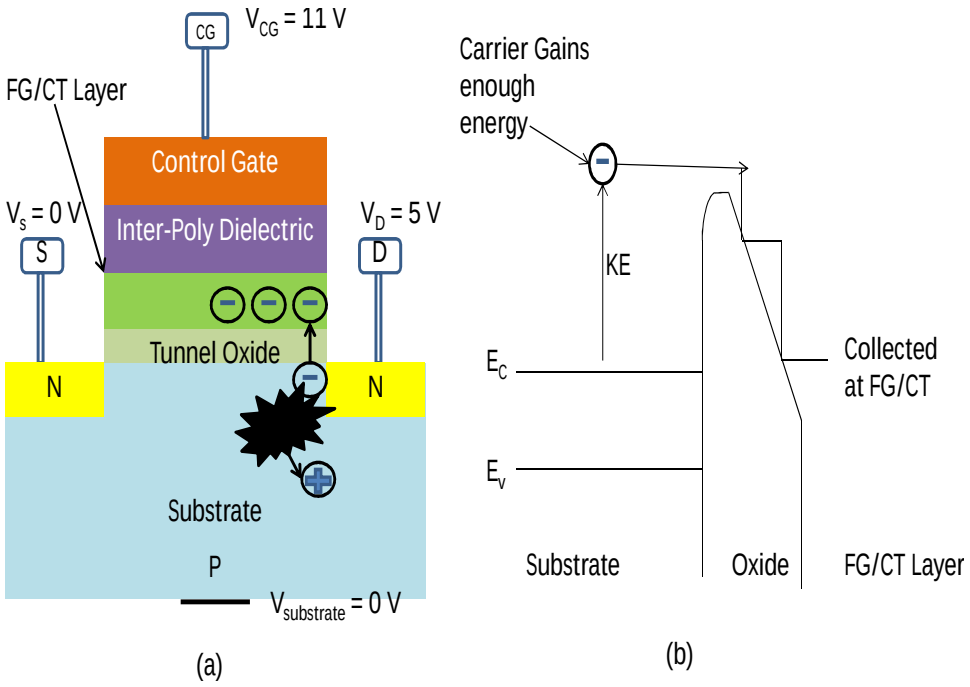

Figure 1.6: (a) Operating conditions with CHE injection (b) Band diagram during CHE injection.

reach very high energy values at the end of the channel. This high energy near the drain end causes the breakdown effect near drain region and generated electron-hole pair. Here, the generated hole contributes to the substrate current whereas the electron as charge carrier try to overcome the TO barrier to reach the FG/CT layer. Note that a carrier, to overcome the TO barrier, must have kinetic energy higher than the potential barrier of TO and the velocity towards the CG. However, only a small amount of hot electrons gain enough kinetic energy to overcome the TO barrier. Rest of electrons get collected at the drain and accord the power consumption of cell. Importantly, the main advantage of CHE injection lies in its tighter program threshold voltage distribution and faster single bit operation. However, high interface state and oxide trap generation with CHE causes high risk to the reliability of the flash memory cell. Apart from this, higher power consumption due to small injection efficiency and large currents makes it unsuitable for battery operated devices.

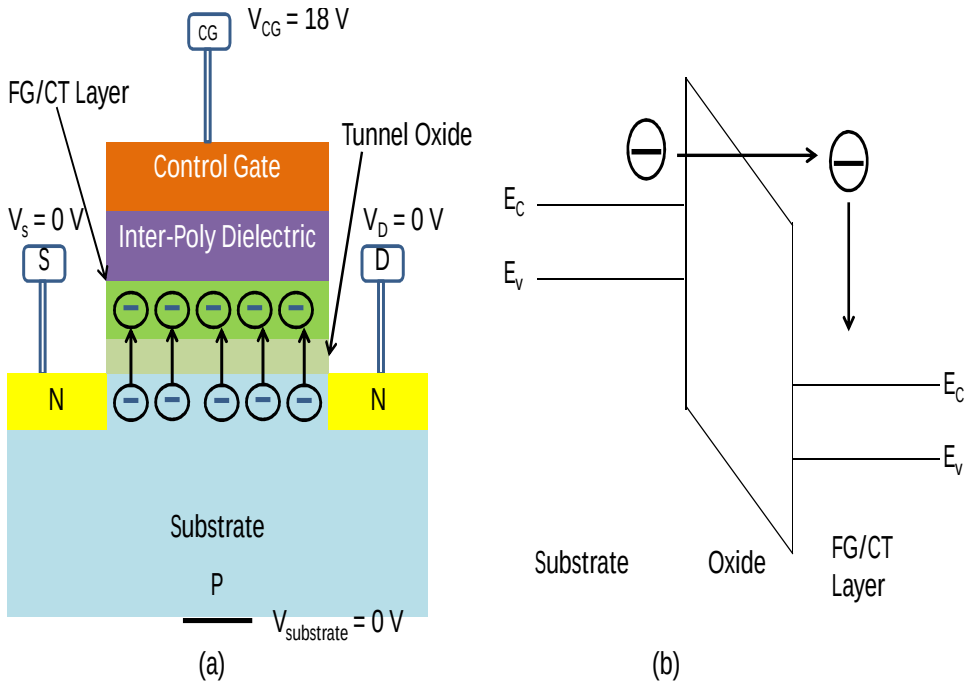

Figure 1.7: (a) Operating conditions with FN tunneling (b) Band diagram during FN tunneling

FN Tunneling

In FN tunneling, a large positive voltage, is applied to the CG while the substrate is grounded [14]. When the oxide barrier between the substrate and the FG/CT layer becomes thin enough, charge carrier tunnel from the substrate to the FG/CT layer through the TO. In FN tunneling, charge carrier travels from the vicinity of the substrate fermi level through the forbidden energy gap of TO into the conduction band of the FG/CT layer in the presence of a high electric field. Figure 1.7 depicts the bias conditions of flash memory cell with energy bands orientations during FN tunneling operation. Basically, the FN tunneling provides following advantages compared to CHE injection.

- Lower Power consumption due to small cell current.
- Compared to CHE, FN tunneling has smaller probability of TO damage. Hence, improves reliability.
- Small TO thickness is allowed, Hence provides efficient vertical scaling.

However, the exponential dependence of FN tunnel current on the oxide field causes some critical problems of process control. This is because even a very small variation of oxide thickness among the cells in a memory array results in great difference in programming and erasing currents, thus spreading the threshold voltage distribution.

1.2.2 Erase Operation

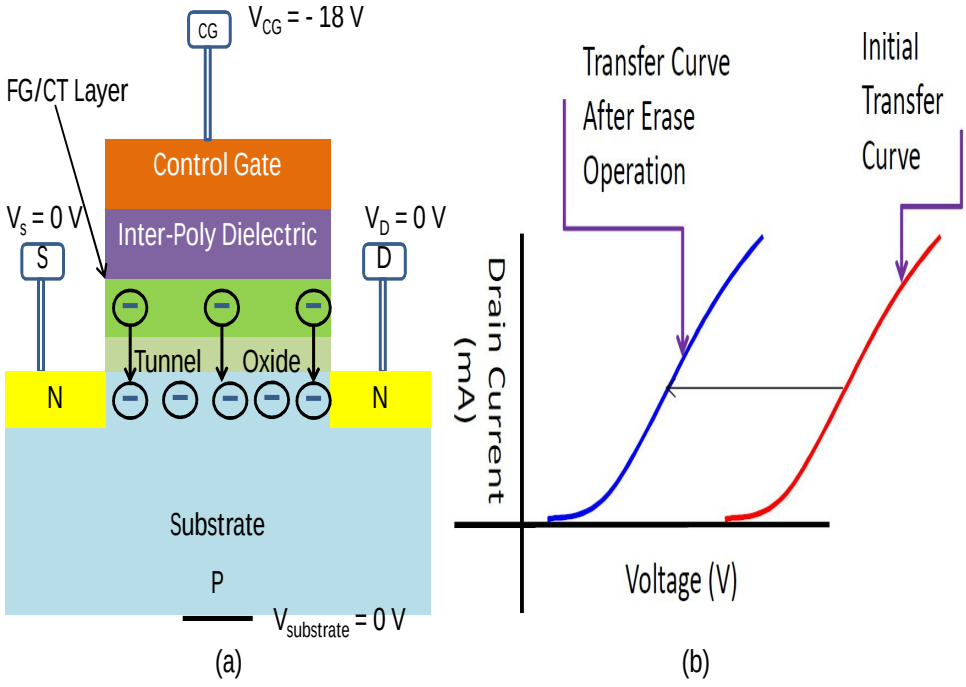

Figure 1.8: (a) Erase operation (b) Modified transfer characteristics after erase operation

A flash memory cell is erased (set to a 1 state) by applying a large negative voltage on the CG [14]. This provides a negative electric field at TO strong enough to tunnel electrons from the FG/CT layer to the substrate. This process makes FG/CT layer void of charge carrier and according to the equation 1.1 threshold voltage of the flash memory cell decreases to $V_{th-erased}$. Figure 1.8 shows the operating condition of flash memory cell and resultant change in transfer characteristics of cell after erase operation . Most commonly, FN tunneling mechanism is used for the erasing process of flash memory cell. Here, it is important to note the erase threshold voltage must be as small as possible to attain the maximum de-trapping of charge carrier from FG/CT layer.

1.2.3 Read Operation

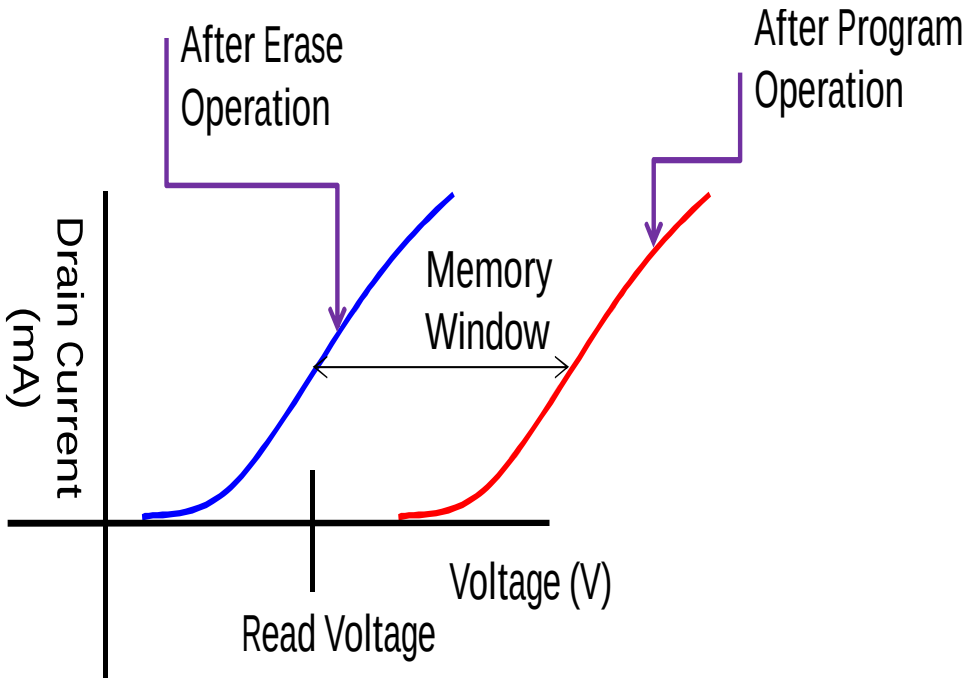

Figure 1.9: Different operations of NAND flash memory.

The read operation is performed to determine whether the cell is in a programmed (state 0) or erased (state 1) condition[14]. A reference voltage V_{read}, is applied to the CG such that, where $V_{th-erased} < V_{th-read} < V_{th-programmed}$ [3]. Here, $V_{th-erased}$ is the threshold voltage of an erased cell and $V_{th-programmed}$ is the threshold voltage of a programmed cell. Also, $V_{th-read}$ represent the threshold voltage of the flash memory cell during read operation. Specifically, if the flash memory cell to be read is in the un-programmed or in erased state, current flows through the flash cell during read operation. This effect is attributed to fact that $V_{th-erased} < V_{th-read}$, which causes easy inversion of channel to assist the flow of cell current between source and drain terminals. further, if the cell to be read is in the programmed state, no current flows through the cell during read operation. This is because $V_{th-programmed} > V_{th-read}$, which prohibits inversion of channel and hence, no current flows between source and drain terminal. Apart from this, Figure 1.9, depicts the memory window, an important performance metric of flash memory cell, is the threshold voltage difference between program and erase operations. In fact, large memory window represents the high reliability of the flash memory cell.

Several factors such as number of program and erase cycles, temperatures, defects in TO and IPD etc causes memory window closure in flash memory cell and try to affect the performance of flash cell.

1.2.4 Program-Inhibit Operation

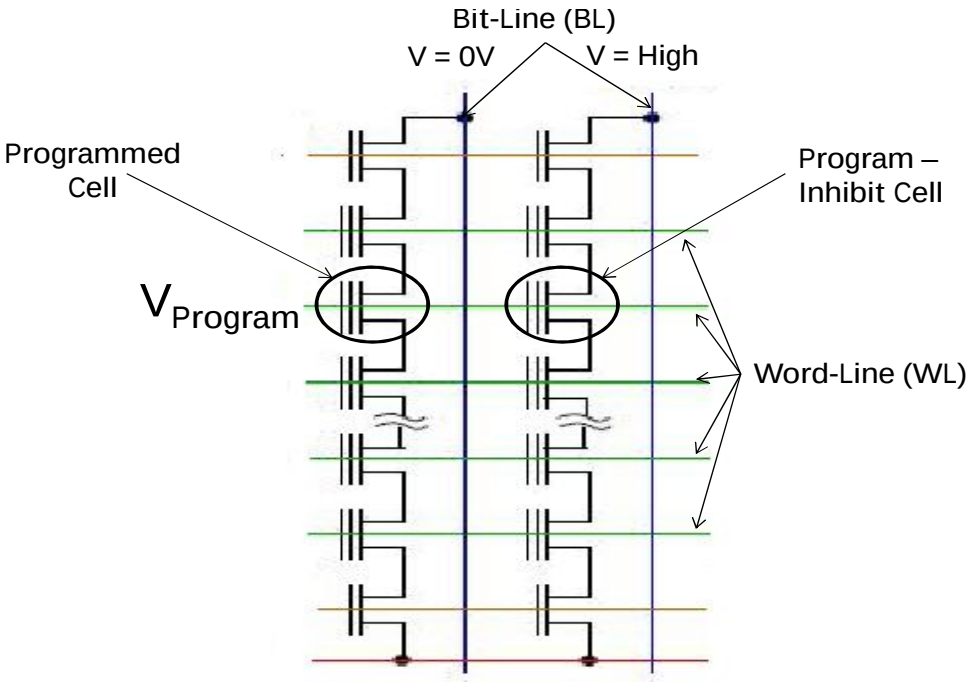

Figure 1.10: Program-Inhibit NAND flash memory operation

In flash memory array, the CG's of flash memory cells are connected through WL, as shown in Figure 1.10. In this regard, if any voltage is applied to the selected cell for programming/erasing/reading process, that will be carried to all cells sharing the that WL. During Program operation, with a high potential at the selected CG, all cells sharing the same WL would have the tendency to get programmed instead of the programmed cell only. All the cells in the string which get this unwanted operational CG voltage are called as Program-Inhibit (P-I) cells and the process to prevent unselected cells from being programmed is called as program-inhibition [15]-[16]. The program inhibit process is also often referred to as boosting. During P-I mode of operation, the value of stored charge in P-I flash memory cell must not be changed despite of high CG voltage. This phenomenon is fulfilled when the channel/substrate voltage of P-I

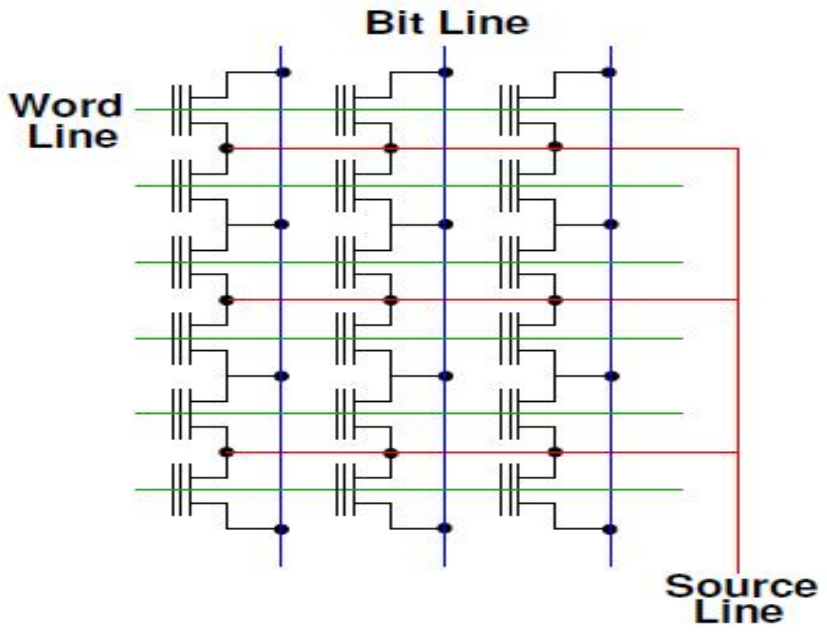

Figure 1.11: A NOR flash memory.

flash memory cell is increased to avoid the FN tunneling between substrate and FG/CT layer.

1.3 Flash Memory: Circuit Classification

By the circuit level organisation, a flash memory cell can be divided into two parts:

- NOR Flash Memory
- NAND Flash Memory

Though they are both considered leading non-volatile flash memory technologies, NAND and NOR flash meet completely different design needs based on their individual attributes.

1.3.1 NOR Flash Memory

The placement of flash memory cells in NOR flash memory configuration is shown in Figure 1.11. With NOR flash memory technology, each cell is connected to its word

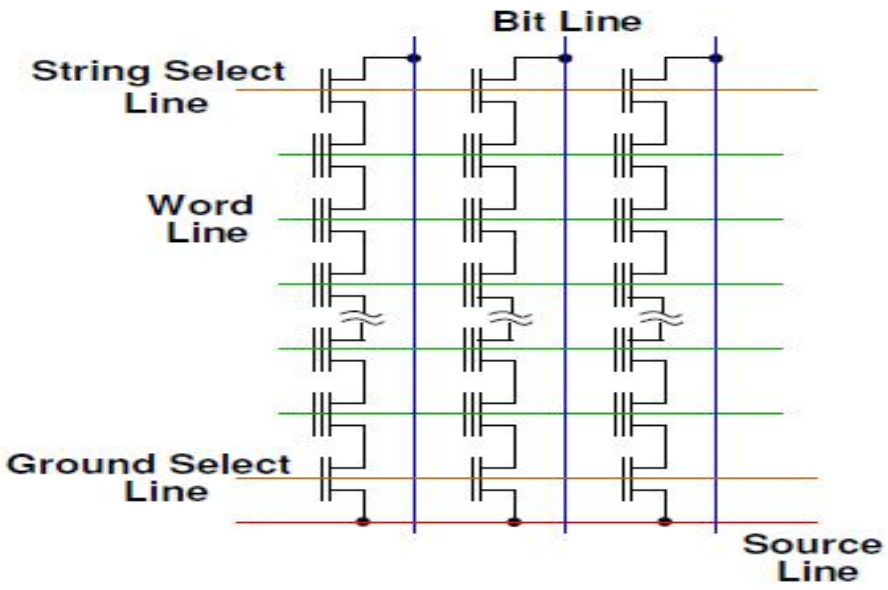

Figure 1.12: A NAND flash memory

line, bit line and source line. Here, due to the parallel arrangement of flash cells, NOR flash memory offers faster read speed and random access capabilities, making it suitable for code storage in devices such as PDAs and cell phones. This random access capability gives NOR flash its Execute-In-Place (XIP) functionality (code execution), which is often required in embedded applications. Another advantage of NOR Flash is its byte write capability. However, with NOR flash technology, write and erase functions are slow compared to NAND flash technology. Also, the NOR flash memory has a larger memory cell size than NAND flash, limiting its scaling capabilities and therefore achievable bit density compared to NAND flash. Since code storage tends to require lower density memory than file storage, NORs larger cell size and smaller capacity is not considered a concern when used in these applications. A few of the most popular applications of NOR flash today include BIOS, Mobile Phones, Routers, Set-Top Boxes, and Video Games.

1.3.2 NAND Flash Memory

The placement of flash memory cells in NAND flash memory configuration is shown in Figure 1.12. With NAND flash technology, cells are connected in a series to form a string and for each string there is only one bit line and one source line. The most prominent advantage of NAND flash memory over NOR Flash is its faster program and erase times. Moreover, NAND flash cells are 60% smaller in area as compared to NOR flash cells providing higher densities required for todays low-cost hand hold consumer devices. NAND flash offers smaller number of contact per every cell than NOR flash. This results in smaller cell size, and greater bit density at a lower cost. The fast program and erase times, lower cost per bit, and small size make NAND flash ideal for large amount of data storage. NAND Flash is very similar to a HDDs. It has sector-based operation (page-based) and well suited for storage of sequential data such as pictures, video, audio, or computer data. Importantly, NAND flash is used in virtually all removable cards, including USB drives, SD cards, memory stick cards, CompactFlash cards, and MMCs. Some of the important applications of NAND Flash memory include mobile phones, digital audio/voice storage, still & video Cameras, mass data storage in SSD. Therefore, the large application market of NAND flash memory makes it the most significant non-volatile memory solution for the next decade [17]. The main drawback of NAND Flash is its slower random access as compared to NOR flash memory. Also presence of Source Select Line (SSL) and Ground Select Line (GSL) for the decoding of physical address or individual access of flash memory cell increases the area overhead as compared to the NOR flash Memory.

1.4 Performance Measures

In the following, we briefly describe key performance metrics that we have investigated in this thesis.

1.4.1 Short Channel Effects

To meet with the customer demand of small area and low cost handhold devices, scaling of NAND flash memory have become inevitable. The scaling of the flash memories has continued for about 35 years leading to a market that, according to IC Insight generated total market revenue of 76 billions in 2017. The size of the flash memory cell is expected to be scaled below 10nm by 2025. The scaling of flash memory cell introduces short channel effects (SCEs) which falsely modify the cell characteristics [18]-[19]. Specifically, following parameters defines the SCEs in any MOS based device.

Threshold voltage roll-off

The threshold voltage roll-off (ΔV_{th}) is an effect where there is a reduction in threshold voltage with decrease in channel length. Therefore, at shorter channel length, the device may turn on erroneously and may degrade its performance. Importantly, ΔV_{th} occurs in MOS based devices due to charge sharing between source/ drain and gate terminals of the device. Further, this effect will cause the reduced gate control over the channel under the presence of the higher lateral electric field between source and drain terminal. This lateral electric field forces the threshold voltage reduction with scaling, even without gate intervention. Therefore, smaller change in threshold voltage of MOS based devices with technology reduction may serve the purpose of scaling.

Subthreshold Swing

subthreshold conduction is a phenomenon which represent drain current below the threshold voltage. This current appears due to the weak inversion in the channel between flat band and threshold voltage. Hence, a diffusion current flows from source to drain, even below the threshold voltage. This subthreshold conduction must be minimum in any MOS device. To investigate the subthreshold conduction, subthreshold swing (SS) is used. The SS defines requirement of the minimum power supply voltages and minimum power dissipation of a device. Also, the SS is defined as the gate voltage required to change the drain current by one order of magnitude per decade, defined by

the following equation.

$$SS = \frac{V_T - V_{OFF}}{log(I_{ON} - I_{OFF})}$$

Here, V_{OFF} and V_T are the voltages OFF and threshold voltage of the flash memory cell respectively. Similarly, I_{OFF} and I_{ON} are the 'OFF' and 'ON' current of the flash memory cell. Small value of the SS is important to turn ON the considered device as fast as possible. Cearly, with the small value of SS, the MOS device will act as better switch. In contrast, with the scaling, SS of MOS based devices increases which poses a limitation and slows down the speed of the device. While scaling, smaller SS is important to obtain for the reliability of the considered MOS device.

Drain Induced Barrier Lowering

Drain Induced Barrier Lowering (DIBL) is a secondary effect in MOS based devices referring originally to a reduction of threshold voltage of the transistor at higher drain voltage. Here, as the drain voltage is increased, the depletion region of the PN-junction between the drain and the body increases in size and extends under the gate. So, the drain occupy a larger portion of the burden of balancing depletion region charge, leaving a smaller burden for the gate. As a result, the charge present near the gate retains charge balance by attracting more carriers into the channel. It lowers the threshold voltage of the device. Hence, the DIBL effect could be defined as the decrease in threshold voltage when the drain voltage is increased from a low value $V_{d,low}$ to a high value $V_{d,high}$. However, for fully depleted device, the potential profile underneath the gate terminal is controlled by field lines originating from the source and drain instead of front gate. Also, the field lines inside the buried oxide changes the DIBL with drain bias. Therfore, the source-drain lateral field coupling causes DIBL in fully depleted MOS devies. In conclusion, for a MOS based device, DIBL is a negative phenomenon and it must be as small as possible to improve the reliability. Also, the DIBL increases with the scaling of MOS device which presents the provocation for device performance.

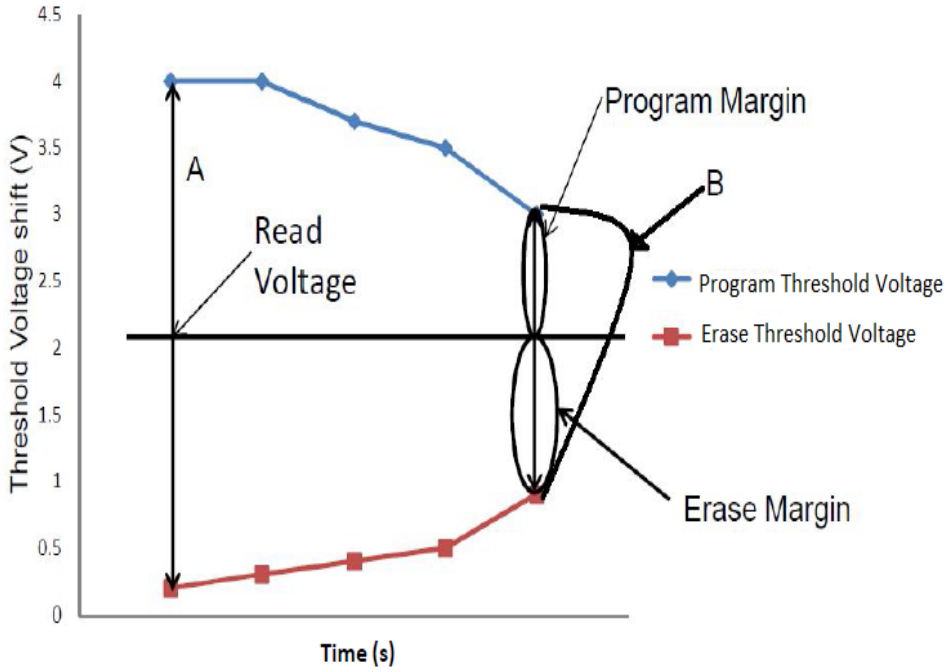

Figure 1.13: Data Retention of NAND flash memory

1.4.2 Data Retention

The ability of keeping the stored information i.e. the charge trapped into the storage layer unaltered for a long time is known as data retention [13]. A non-volatile memory should have high data retention for improved reliability. However, even with no bias applied, charge loss occurs from FG/CT layer. This is called as data retention loss. This data retention loss can lead to a read failure i.e. a programmed cell can be read as erased if its threshold voltage shifts below read voltage due to charge loss from FG/CT layer. Importantly, lower channel energy level, high temperature and large number of program/erase stress, oxide degradation causes loss of data retention. Figure 1.13 shows the 10 year data retention characteristics of a NAND flash memory cell.

In this figure, A represents the memory window of cell at t= 1 sec and B represent the memory window at 3×10^8 sec (i.e. 10 year). Importantly, data retention is calculated using following relation:

$$\%DataRetention = \frac{B}{A} \times 100$$

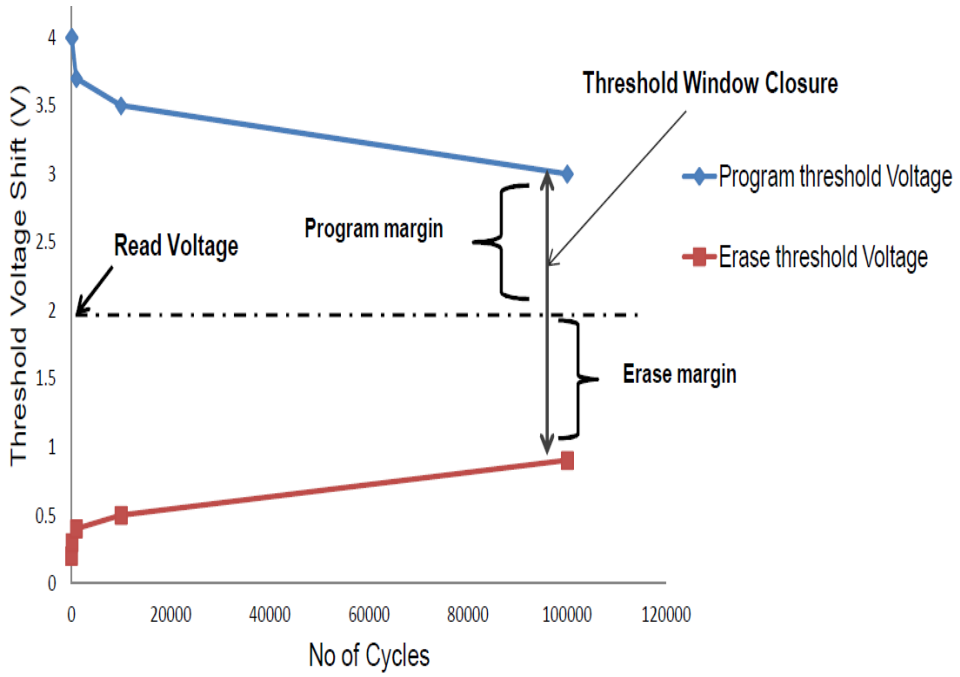

Figure 1.14: Endurance of NAND flash memory

Hence, the % data retention evaluate the charge remaining in FG/CT layer after 10 year of time. Higher the remaining charge, improved will be the data retention and reliability.

1.4.3 Endurance

In NAND flash memory cells, program and erase operations depends on charge transport through the TO. This charge transport is accomplished using FN tunneling into/from a trapping layer. This electron tunneling through TO is responsible for a slow, but continuous TO degradation because of traps generation in TO and interface damages [20]. As a consequence, there could be charge trapping/detrapping into the TO or undesired charge flowing into/from the CT layer. As the number of Program/Erease (P/E) cycles rises, the above introduced effects strongly impact writing operation of NAND flash memory cell. Importantly, electron trapping in TO reduces the tunneling efficiency so that, under constant voltage and time conditions, the charge injected into/from the CT layer decreases cycle after every P/E cycle. This results in the reduced memory window of the cell, as shown in Figure 1.14.

From the figure, it is clear that with 10^5 the initial memory window has reduced. This reduction of memory window may result in the read failure. Therefore, one can deduce that with every flash memory cell, a limited number of P/E cycles are allowed. This phenomenon presents the main drawback of flash memory. In fact, the maximum number of P/E cycles that a flash memory cell is able to withstand can be given by using following equation

$$n_{pe} = Q_{bd} A_{inj} / \triangle V_{th} C$$

Here,n_{pe} is the number of program erase cycles until breakdown of oxide, Q_{bd} is the amount of charge crossed the TO till breakdown, A_{inj} represent the area of charge injection in TO and $\triangle V_{th}$ is the threshold voltage difference between 0 and 1 state.

1.5 Simulation Tool

Simulations are required to optimize the device performance when hands on calculation and fabrication methods become too complicated or impose unacceptable assumptions. Sentaurus TCAD tool from Synopsys is an advanced commercial computational environment with a collection of tools which is used for performing simulations of electronic devices and to understand advanced-device physics. It also helps to investigate scaling analyses of device and may provide different design rules. In addition, it also allows us to interact with the fabrication methodology using process manufacturing. Importantly, with a Sentaurus TCAD, the device behavior is obtained from the solution of the appropriate differential equations describing the device physics on a given geometrical domain. Here, the physics of the considered device is obtained from different models already available with the tool. After including the appropriate models, the considered device can be ramped with necessary electrical stimulation to get the final results.

Furthermore, Synopsys Sentaurus TCAD provides two methods for the device design :

1. Device TCAD : It deals with the modeling of electrical, thermal, optical and

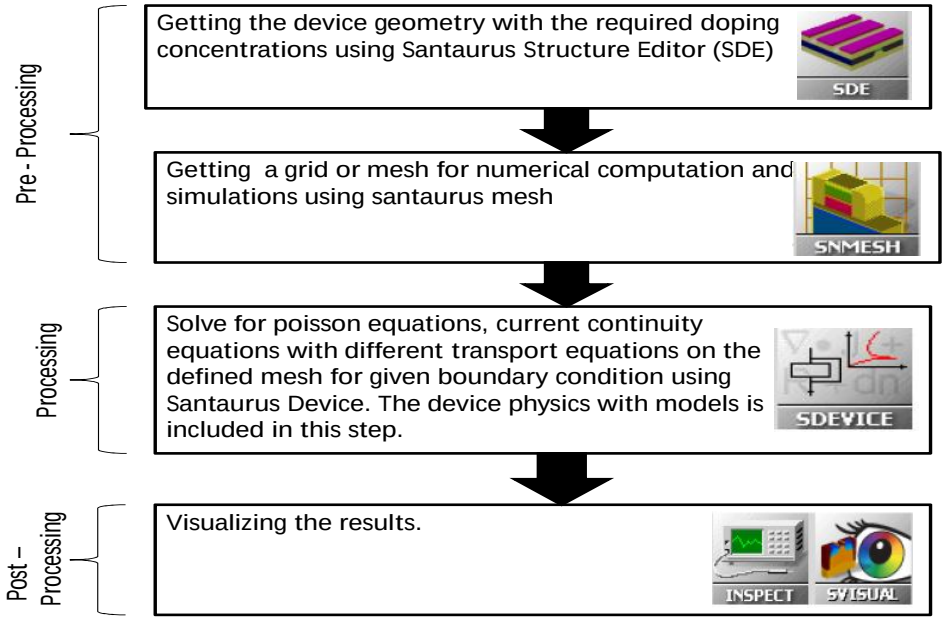

Figure 1.15: Device simulation steps with Synopsys Sentaurus Device TCAD.

mechanical behavior of semiconductor devices.

2. Process TCAD : It aims to the modeling of semiconductor-chip process-manufacturing steps like lithography, deposition, etching, ion implantation, diffusion, oxidation, silicidation, mechanical stress, etc..

In this thesis, we have worked with device TCAD tool of Synopsys Sentaurus for the modelling of electrical behavior of NAND flash memory devices. Further, different steps for Sentaurus simulations with device TCAD are given in Fig. 1.15.

1.5.1 Sentaurus Device

The description of different tools used for simulation with device TCAD can be given as follows:

Sentaurus Structure Editor (SDE)

Sentaurus Structure Editor can be used as a two-dimensional (2D) or three-dimensional (3D) structure editor. Here, the required structures are generated or edited interactively using the graphical user interface (GUI). Doping profiles strategies for different regions of the considered device can also be defined with structure editor tool. In addition, Sentaurus Structure Editor provides an interface to call the Synopsys meshing tool i.e. Sentaurus Mesh to generate required grid points. Importantly, it provides the necessary input files (the .tdr boundary file and mesh command file) for the meshing tool and provide the way for the device simulation using Sentaurus Device.

Sentarus Mesh (SNMESH)

Sentaurus Mesh is a mesh generator that produces rectangular or hexahedral elements for use in applications such as semiconductor device simulation, process simulation and electromagnetic simulation. The points where these elements intersects each other are known as the grid points. The physical equations of the considered device are solved corresponding to these grid points only. The mesh generation tools is composed of two mesh generators:

1. Sentarus Mesh: Sentaurus Mesh is a robust mesh generator capable of producing axis-aligned meshes or grid points in 2D and 3D. In the 2D MOS type of devices, Sentaurus Mesh works is recommended. Importantly, the simulated devices where the most important surfaces are the axis aligned surfaces, sentaurus mesh is used. Although Sentaurus Mesh works very well with MOS type devices, the effective and optimized numerical meshes in which the problem can be solved assuring convergence and, at the same time, with the reasonable simulation times, is a difficult to obtain. However, some general rules can be applied:

A. The grid spacing must be sufficiently dense so that all the relevant features of the geometry are accurately represented. However, it will increase the grid points and consequently the simulation time. Therefore, It is recommended that to create the most

suitable mesh, the mesh must be densest in those regions of the device where High current density, High electric fields and High charge generation event occurs.

B. Points must be allocated to accurately approximate the physical quantities of interest.

2. Noffset3D: Noffset3D is an advancing front mesh generator, capable of producing triangular meshes in 2D and tetrahedral meshes in 3D. For devices where the main surfaces are nonaxis-aligned or curved, the recommendation is to use Noffset3D. The meshes produced by Noffset3D can contain layers of elements that are nearly parallel to given surfaces of a semiconductor device structure. The Noffset3D uses a series of algorithms to generate the final meshing. The main algorithm contains of layer generation, surface refinement and doping interpolation.

Importantly, the overall computation time depends on the total number of grid points during meshing, therefore grid point number must be minimized for computational efficiency.

Sentarus Device (SDEVICE)

Sentaurus Device simulates numerically the electrical behavior of a semiconductor device in isolation or several physical devices combined in a circuit. The terminal currents, voltages, and charges are computed based on a set of physical device equations that describes the carrier distribution and conduction mechanisms. A real semiconductor device, such as a transistor, is represented in the simulator as a virtual device whose physical properties are discretized onto a nonuniform grid points. Therefore, a virtual device is an approximation of a real device in Sentarus TCAD. For this purpose, the poisson equation with electron and hole continuity equation are solved using newton iteration method for the whole range of electric stimulation with different physical models.

The SDEVICE has extensive set of models of physics for semiconductor devices,

general support for different device geometries and mixed-mode support of electrothermal net lists with mesh-based device models and SPICE circuit models. Importantly, in SDEVICE, continuous properties such as doping profiles are represented on the mesh and therefore, are only defined at a finite number of points. The doping at any point between these grid points (or any physical quantity calculated by SDEVICE) can be obtained by interpolation.

The Sentaurus Device command file can be organized in commands or sections that can be in any order. Also, Sentaurus Device keywords are not case sensitive. A Sentaurus Device command file has following sections :

1. File Section: File section consist "input files" that define the device structure i.e. .tdr file along with parameter .par file. In addition, the "output files" such as .plt, .log file are also defined in the file section for the simulation.

2. Electrode Section: With Sentaurus Device, it is necessary to specify which of the contacts are to be treated as electrodes. Electrodes in Sentaurus Device are defined by electrical boundary conditions and contain no mesh. Any contacts that are not defined as electrodes are ignored by Sentaurus Device.

3. Physics Section: The physics section of sentaurus command file allows a selection of the physical models to be applied in the considered device simulation. The selection of physical model strongly influences the accuracy of the simulation result. The physical models used are explained in detail later in this thesis.

4. Plot Section: The Plot section specifies all of the solution variables that are to be saved in the output plot files (.tdr). Only data that Sentaurus Device is able to compute, based on the selected physics models, is saved to a plot file.

5. Math Section: Sentaurus Device solves the device equations (which are es-

sentially a set of partial differential equations) self-consistently, on the discrete mesh, in an iterative fashion. For each iteration, an error is calculated and Sentaurus Device attempts to converge on a solution that has an acceptably small error. For this purpose, one need to define a few settings for the numeric solver in Math section.

6. Solve Section: The Solve section defines a sequence of solutions to be obtained by the solver. The Quasistationary command is used to ramp a device from one solution to another through the modification of its boundary conditions or parameter values in the solve section.

Sentaurus Visual (SVISUAL)

It is a plotting software for visualizing data output from simulations. Sentaurus Visual enables users to work interactively with data using both a graphical user interface and a scripting language for automated tasks.

1.5.2 Simulation Methods

In device TCAD, following methods are available for electrical simulations of the device.

Study State Simulations

In steady-state conditions, for each property of the systems, its partial derivative with respect to time is zero, i.e. , nothing is changing with time. To perform steady-state simulations, the Sentaurus TCAD keyword is Quasistationary. The Quasistationary command is used to ramp a device from a solution to another through the modification of the boundary conditions that can be Voltage, Current, or Temperature.

Transient Simulation

A transient response or natural response is the time-varying response of a system to a change from equilibrium. In Sentaurus, the keyword that must be used to perform transient simulation is "Transient". The command must start with a device that has

already been solved under stationary conditions. The simulation then proceeds by iterating between incrementing time and re-solving the device.

AC simulations

Performing a small signal or AC analysis means simulate the behavior of system when a relatively small harmonic signal is superimposed to a steady-state condition or DC bias point. The keyword for AC analysis in Sentaurus SDEVICE is ACCoupled.

1.5.3 Physical Models

Synopsys Sentaurus TCAD tool offers an extensive set of models to represent the virtual device as an approximation of the actual device. These models are included in the SDEVICE script for the considered device simulation. Some of the important models are listed as follows :

Carrier Transport Model

All the carrier transport model supported by Sentaurus TCAD follows the charge conservation law in active region of the considered device. Here, carrier concentrations in any region of the device must never be negative during the newton iteration. If during a Newton iteration a concentration erroneously becomes negative, the tool provides the message that the newton is not able to converge. After this, SDEVICE applies damping procedures (i.e. using the smaller step size) to make it positive. If the newton iteration converges for the whole range of electrical stimulation, simulation gets complete. Here, following model for carrier transport is used in Sentaurus TCAD:

1. Drift Diffusion: The drift diffusion model is very important for the semiconductor device simulation with Sentaurus Device. This model is used for isothermal simulation and is suitable for low-power density devices with long active regions. Also, the drift-diffusion model is the default carrier transport model in Sentaurus Device. The drift diffusion equation mainly consist the poisson's equation, the electron and

hole current equation and the electron and hole current continuity equation. Therefore, with drift-diffusion model, current densities of electron and holes can be as follows:

$$J_n = \mu_n(n\nabla E_C - 1.5nkTln(m_n)) + D_n(\nabla n - n\nabla ln(\gamma_n))$$

$$J_p = \mu_p(p\nabla E_C - 1.5pkTln(m_p)) + D_p(\nabla p - p\nabla ln(\gamma_p))$$

Here, The first term of the above two equations comes the spatial variations of the electrostatic potential, the electron affinity, and the band gap. The remaining terms take into account the contribution due to the gradient of concentration, and the spatial variation of the effective masses m_n and m_p. Also, the diffusion constants i.e D_n and D_p are calculated using Einstein relation.

To activate drift diffusion model, keywords *Electron* and *Hole* is used in the solve section of Sentarus Device command file with poisson equation.

Further, in this thesis, only drift diffusion model is used for the carrier transport of the charge carrier because the simulation of the considered have been obtained for room temperature.

2. Thermodynamics : This model accounts for self-heating of the active area. It is suitable for devices with low thermal exchange, particularly, high-power density devices with long active regions. The model differs from drift-diffusion when the lattice temperature equation is solved. However, It is possible to use the drift-diffusion model together with a lattice temperature equation, but it is not mandatory. To activate the thermodynamic model, specify the *Thermodynamic* keyword in the physics section of the Sentarus Device command file.

3. Hydrodynamic : It accounts for energy transport of the carriers and is suitable

for devices with small active regions. Simlar to the drift-diffusion model, the hole and electron current densities in hydrodynamic model takes into account the contribution due to the spatial variations of electrostatic potential, electron affinity, the band gap, the contribution due to the gradient of concentration, the carrier temperature gradients, and the spatial variation of the effective masses. Along with this, the thermal diffusion constants are used with electron and hole to encounter the carrier temperature. To activate the hydrodynamic model, the keyword *Hydrodynamic* must be specified in the physics section. If only one carrier temperature equation is to be solved, *Hydrodynamic* must be specified with an option, either *Hydrodynamic(eTemperature)* or *Hydrodynamic(hTemperature)*.

Generation Recombination Model

The drift diffusion model is used to calculate the electrostatic potential and electron/hole concentration. Apart from that the generation recombination processes are the methods that exchange carriers between the conduction band and the valence band. For each individual generation or recombination process, the electrons and holes involved appear or vanish at the same location. Following generation recombination models are used in this thesis:

1. Shockley Read Hall (SRH): In general, recombination of charge carrier through deep defect levels in the gap is depicted as ShockleyReadHall recombination. Net recombination with SRH is dependent on the energy difference between defect level and intrinsic level, the carrier lifetime for electrons and holes, intrinsic concentration and electron/hole charge carrier density in the considered device. Hence, SRH model takes many important features of the simulated device into consideration. In addition, the electron and hole concentration with SRH is doping dependent, field dependent, and temperature dependent.Therefore, the SRH can be made to handle variability accounted from doping, electric field and temperature variation also.

The generationrecombination model can be selected in the physics section of Sen-

taurus Device command file as an argument to the *Recombination* keyword i.e. Physics {(Recombination(SRH (dopingdependence..)...)).

2. Poole Frenkel: The PooleFrenkel model used for the interpretation of transport effects in dielectrics and amorphous films. The model predicts an enhanced emission probability for trap where the potential barrier is decreased because of the high external electric field. The model is selected by the *PooleFrenkel* keyword in the command file.

Mobility Model

Sentaurus Device uses a very modular approach for the explanation of the carrier mobilities in the Sentarus Device command file. For the simplest case, a constant mobility model is used with the undoped materials. For doped materials, the carriers scatter with the impurities and this leads to the degradation of the mobility. To justify this effect, following mobility model can be used with doped materials:

1. Doping Dependent Mobility Degradation: In doped semiconductors, the scattering of charge carriers by impurity ions leads to the degradation of the carrier mobility. The model to justify this mobility degradation due to impurity scattering are activated by specifying the *DopingDependence* flag to *Mobility* in the physics section of Sentarus Device command file i.e. *Physics(Mobility(DopingDependence ...))*.

If *DopingDependence* is specified without options, Sentaurus Device uses a material dependent default model. In silicon material, the default doping dependent mobility model is the Masetti model where the mobility is given by the following equation:

$$\mu_{dop} = \mu_{min1} exp(-\frac{P_C}{N_{A,0} + N_{D,0}} + \frac{\mu_{const}\mu_{min2}}{1 + ((N_{A,0} + N_{D,0})/C_r)^{\alpha}} - \frac{\mu_1}{1 + (\frac{C_s}{N_{A,0}+N_{D,0}})^{\beta}} \quad (1.2)$$

Here, the mobilities μ_{min1}, μ_{min2}, and μ_1 are the reference mobilities which along with doping concentrations P_C, C_r, and C_s, and the exponents α , β are accessible in the parameter set *DopingDependence* in parameter file.

2. Field Dependent Mobility Degradation In high electric fields, the carrier drift velocity is no longer proportional to the electric field, instead, the velocity saturates to a finite speed v_{sat} . In Sentarus device, actual field dependent mobility model is selected by heyword *HighFieldSaturation*. To include high field saturation effect, default canali model is used in this thesis with default parameters applicable to the silicon material.

Traps and Fixed Charges

Traps are important in device physics to induce conductivity in the dielectric material. These traps provide doping, enhance recombination, and increase leakage through insulators. These traps are available for both bulk and interfaces. Further, Sentaurus Device provides traps definitions as follows:

1. Trap types i.e. electron trap, hole traps, fixed charges, Acceptor and Donor type.

2. Five types of trap energy distribution i.e. level, uniform, exponential, Gaussian, and table.

3. Tunneling model to trap such as nonlocal tunneling.

The specification of trap distributions and trapping models appears in the Physics section. Traps can be defined for specific material, regionwise, region interface, and material interface.

Non-Local Tunneling

In non-local tunneling, The tunneling current depends on the band edge profile along the entire path between the points connected by tunneling. This means the electric field at each point in the tunneling path is dynamically changing. This makes tunneling a nonlocal process, hence model is dynamic non-local tunneling model. The local tunneling models given in the past by Ken, Hurkx and schenk assumed a constant electric field throughout the tunneling path. This assumption fails to deliver the accurate results

with tunneling. Therefore, the non-local tunneling model:

- Handles arbitrary barrier shapes.
- Includes carrier heating terms.
- Allows to describe tunneling between the valence band and conduction band.
- Offers several different approximations for the tunneling probability.

Further, the specification of trap distributions with non-local tunneling model arrive in the physics section of the device command file. In contrast to other models, most model parameters with non-local tunneling are specified in the device command file only. Importantly, using the physics section of the Sentaurus device command file, traps can be coupled to nearby interfaces and contacts by tunneling. In this thesis, the non-local tunneling model is used to interface traps of nitride charge trap layer with channel-dielectric interface. In general, the non-local tunneling model requires the following:

1. Defining Non-local Mesh The non-local mesh consist of non-local lines that represent the tunneling paths for the charge carriers. To control the construction of the nonlocal mesh, Sentaurus Device uses the keyword *NonLocal* in math section of the device command file. The construction of non-local mesh can be controlled by using *Length* and *Permeation* parameters. Further, for the definition, the keyword *NonLocal* specifies the barrier region over which the non-local mesh has to be defined. The non-local lines forms a box over the barrier, and connect the upper vertices and lower interface for tunneling. Further, Sentaurus device introduces a coordinate along each nonlocal line. The interface is at coordinate zero, the vertex for which the nonlocal line is constructed is at a positive coordinate.

2. Specifying Non-Local Tunneling Model The nonlocal tunneling model is activated and controlled in the Physics section of the device command file. Non-local tunneling model is specified by keyword *eBarrierTunneling* and *hBarrierTunneling*.

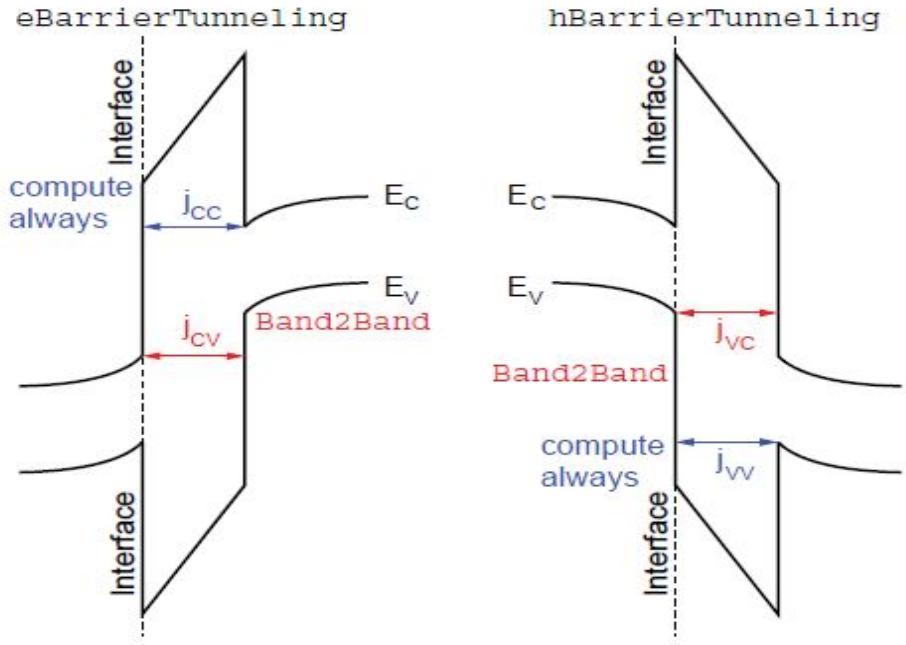

Figure 1.16: Non-local Tunneling.

Here, *eBarrierTunneling* defines all electron tunneling to the conduction band at the lower point from the conduction band at the upper point. Similarly, *hBarrierTunneling* causes all hole tunneling to the valance band at the lower point from the valance band at the upper point. Here, valance to conduction band component has be discarded.

3. Non-Local Tunneling Parameter The nonlocal tunneling model has several fit parameters. They are specified in the *BarrierTunneling* parameter set in parameter file. Specifically, The parameter pair m_t determines the tunneling masses m_c and m_v. These masses are properties of the materials that form the tunneling barrier. Therefore, tunneling masses are either specified in region-specific or material-specific parameter sets. The definition of tunneling masses can take the following form

$$Material = "Oxide"(BarrierTunneling(m_t = 0.42, 1.0))$$

This equation specifies the electron tunneling mass as $.42m_0$ and hole tunneling mass as $1m_0$. Here, m_0 represents the free tunneling mass of charge carrier. These

values must be taken from the practical fabrication processes. In this thesis, as noted in the tunneling mass (m_t) of the carrier is used as a fitting parameter to calibrate the simulation results with existing experimental data [21].

1.6 3D NAND Flash Memory

Three Dimensional (3D) memories have gathered huge attention as future ultra-high density non-volatile memory technology to keep a trend of rising bit density and reducing bit cost. Conventional Two Dimensional (2D) NAND flash memory is facing limitations for increased manufacturing cost and scaling difficulties. 3D NAND flash memory has been accepted as most promising path to increase the NAND flash memory capacity and to scale the NAND flash memory below 15nm node. Contrary to the pitch scaling of 2D NAND flash, 3D NAND flash stacks up multiple layers to increase the memory density so it is free of both lithography and device scaling limitations. On this path, according to the current flow direction through channel, 3D NAND flash memory can be two types: Vertical Channel (VC) NAND flash memory and Vertical Gate (VG) NAND flash memory, as shown in Figure 1.17 [22]. In the VG structure, the channels are stacked horizontally while gates of the memory cells are vertically shared. Here, double gate structure is the only possible gate type. This VG architecture seems more appealing in terms of pitch scaling, selection of channel material and degradation of read current with large number of stacking layers, as compared to the VC configuration. However, the memory array design and address decoding methods for VG NAND flash memory is difficult due to horizontally parallel bit lines. Apart from this, in the VC configuration, the channels are stacked vertically parallel while gates of the memory cells are horizontally shared. Here, Gate All Around structure is the only possible gate type. Specifically, due to relatively easy fabrication process and address decoding methods, VC NAND flash memory (VNAND) is readily being used for commercial purposes. Recently, mass production of VNAND has been announced. However, reduced channel current and increased tolerable electric field with scaling is a major issue for VC configuration. Similar to convention 2D NAND flash memory utilises GSL and SSL

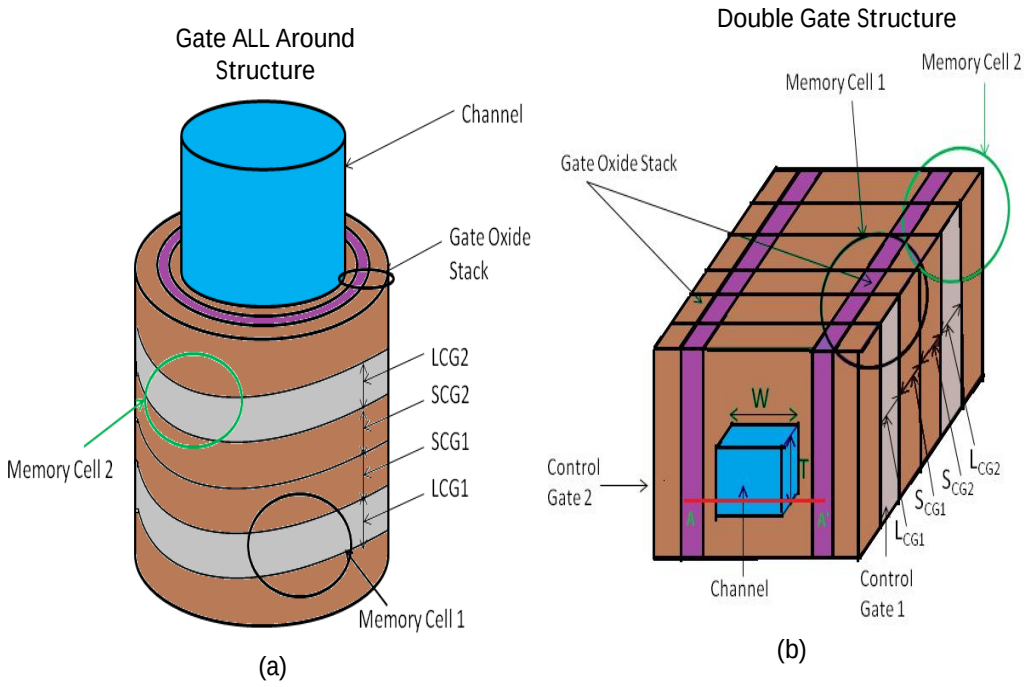

Figure 1.17: 3D NAND Flash Memory Configurations (a) Vertical Channel (VC) (b) Vertical Gate (VG)

for the decoding of the physical address.

1.7 Motivation

In the past decade, memory chips with smaller area and low cost have got an attention due to the booming market of portable handhold electronic devices such as MP3 players, digital cameras, smart phone, laptops etc. In these applications, non-volatile memories (NVM) have become indispensable in order to keep the data for long time. Compared to other emerging NVMs, charge based NVM is favoured because it offers cost and area effective memory solution for future applications. Therefore, the probability of NAND flash memory to drive the non-volatile market for coming years is very high compared to other NVMs. Further, in order to satisfy the area requirement for portable electronics, the NAND flash memory cell must be scaled both horizontally (pitch) and vertically. To optimise the vertical scaling, the CT based NAND flash memories are presently considered as the most suitable choice for the implementation of flash memory. Importantly, it has been predicted that the pitch scaling of 2D NAND flash

memory cell will go below 10-nm by 2025 [17]. Due to this fact, the industry is facing severe scaling challenges such SCEs, reduced data retention, high junction leakage etc. In fact, while scaling, a flash memory has been shown to be more prone to the SCE than any other logic device due to the requirement of thick gate oxide stack (TO + CT + IPD thickness). In flash memory, high thickness of oxide stack is required to achieve high data retention time [23]-[24]. In fact, scaling of vertical gate stack in NAND flash memory has reached its limits but Effective Oxide Thickness (EOT) used is still very high (15 - 20 nm). Importantly, the scaling of gate oxide stack in CT based NAND flash memory cell may improve program/erase speed and SCE but thinning gate oxide stack may result in poor data retention capability. This fact proposes a severe bottleneck or trade-off for the scaling of NAND flash memory [24]. Therefore, a flexible method that can improve short channel effects as well as data retention without scaling the gate oxide stack thickness is of great importance in extreme short channel CT NAND flash memory cells. Therefore, in this thesis, we will explore methods to improve SCE and data retention of flash memory simultaneously for the reliability of the future scaled NAND flash memory.

Moreover, to suppress the SCEs, high channel doping concentration has been considered as a possible solution for NAND flash memory cells. However, this method raises electric field at S/D and channel junction which causes high junction leakage current to flow during Program-Inhibition (P-I) mode of NAND flash memory operation [25]. This junction leakage current reduces boosted channel potential for Program inhibited memory cells and may results in unwanted program/erase operations. Thus, high channel doping concentration may improve SCE but it may not be a suitable approach in sub-20nm NAND flash memory cells. Therefore, investigating a method to improve SCE and channel boosting together will be important for future scaled flash memory cells. Apart from this, Ultra Shallow Junction (USJ) has also shown to eliminate the problem of SCE [26] in flash memory cells. Note that the USJ results in reduced dopant impurities spread from S/D into channel which in turn decreases lateral electric field of device to counter SCE [27]-[28]. In addition, the different doping

condition of source/drain region has been shown to affect the SCEs and junction boost leakage current in flash memory[29]. Therefore, it can be concluded that the doping condition of source/drain region affect the SCEs and memory characteristics of NAND flash memory cells. Motivated from the fact discussed, in this thesis, we try to explore the effect of different source/drain lateral straggle (σ_L) for the improvement of SCEs, data retention and junction boost leakage current simultaneously. Practically, the variation of lateral straggle originates from the different implantation energy. Here, the smaller implantation energy for source/drain dopants results in reduced σ_L and vice versa.

Furthermore, the LDD based structures are known to reduce SCEs and drain disturbance [30] in MOS devices. The same effect is observed with LDD based structure in NAND flash memory cells. However, the lower doping concentration of LDD results in the reduced 'ON' current through the memory cell. Also, with overlapping structure, the lower electric field at the LDD region does not allow electrons in CT layer to completely release during the erase operation. Therefore, a significant amount of electron trap charge remain in the CT layer after the erase operation. This remaining charge is called as residual charge. This residual charge may cause increase in erase threshold voltage (due to partial removal of charge carrier), degraded subthreshold swing, slower erase speed. Therefore, presence of LDD in NAND flash memory cell may cause mendacious memory operation due to residual charge. Additionally, the residual charge in the nitride layer may introduce the reliability issues in charge trap flash memory cells [29]. In addition, the residual charge in CT layer after erase operation may present a barrier for the subsequent program operation, degrading program performance of NAND flash memory cell. Although, the LDD based memory cells are known to provide improved flash memory cell characteristics but the practical knowledge of trapped residual charge and its effect on performance characteristics and reliability of short channel CT memory cell with the variation of LDD depth[1] (d) is not known. Hence, this work captures the

[1]Particularly, LDD depth variations in short channel NAND flash cell occur due to random implantation scattering at junctions. This may alter the LDD depth arbitrarily and cause reliability issues [31]. In fact, in large array, this effect is more pronounced as the devices in array may have different

effect of LDD depth variations on the reliability of NAND flash memory.

Recently, 3D architecture of NAND flash memory has recently gained massive research interest owing to its application in SSD. Here, SSD is the most suitable approach to replace HDD in bulk storage application. It is predicted to achieve upto 3 terabyte of capacity with SSD and its utilisation in computer systems instead of HDD. As already discussed, 3D structure can be resulted from either VG or VC configuration. In this thesis, we have adopted VG configuration, because VG configuration seems more appealing in terms of the pitch scaling. Importantly, a double gate cell architecture is the only possible option with VG configuration. Therefore, in this thesis, we consider a VG NAND flash memory string to implement 3D NAND flash memory. Further, It is to highlight that to achieve high gate electrostatic integrity with NAND flash memory, it requires a super steep doping profile at the channel-Source/Drain (S/D) junction [32]. Albeit, such fabrication process needs highly controlled and sophisticated environment and difficult to achieve. Also, in stacked 3D NAND flash memory architecture, it is quite difficult to dope the space body region between adjacent flash memory cells. The facts highlighted above stimulates the need of a junction-free NAND flash memory to enable achieve scaling towards the extreme short dimensions. Consequently, junction-free NAND flash memory has been investigated in [33]-[34]. Following the same, in this thesis, we consider a junction-free NAND flash memory string. Here, to our best knowledge, very few works have directed their attention towards channel engineering methods to improve the performance of such junction-free NAND flash memory [35] string which remains an open problem. To fill this gap in literature, we have focussed on channel engineering approach for improving short channel performance of 3D junction-free NAND flash memory string. We have comprehensively investigated various reliability metrics such as V_{PASS} interference and data retention ability for such memory and presented useful insights.

Furthermore, it is known that to improve SCEs in junction-free NAND flash memory

LDD depth and hence different performance.

string, employing high channel doping presents a possible solution. [36]. But, the higher channel doping opposes the fringing field penetration from the adjacent CG in junction-free flash memory that reduces the drive current. As a consequence, the memory operation may become ambiguous. Therefore, the low doping of channel seems desirable for the proper operation of 3D NAND flash memory [34]. However, the low channel doping makes the flash device in the string more vulnerable to the fringing interference from neighboring devices [11]. Furthermore, the low doping in the channel region may degrade the programming speed, data retention time for the NAND flash memory [36]. Hence, the combination of high and low doping in the channel is necessary to provide better SCEs with improved memory characteristics. Hsaio et. al. have illustrated in [36] that proper channel engineering for 3D NAND flash memory can optimize its operation beyond 15nm technology node. Such channel engineering may be harnessed in order to improve the SCEs for flash memory without altering the TO thickness. So, relying on this, both the SCE and the data retention time can be simultaneously improved for the flash memory.

1.7.1 Objectives

The main objectives of this thesis are

- To analyse the NAND flash memory cell with different source/drain lateral straggle to optimise short channel effects, program performance, data retention and junction leakage current of the flash memory cell.
- To study the effect of LDD depth variation on the reliability of the NAND flash memory cell.
- To analyse the 3D NAND flash memory string with uniform and retrograde channel doping to optimise short channel effects, program performance, data retention of the memory string.

With above goals, this thesis presents extensive performance analysis of NAND flash memory from cell level to array level. For instance, we first evaluate the performance

of NAND flash memory cell with different S/D lateral straggle. Based on our investigations, we have shown that adjusting the S/D doping lateral straggle σ_L appropriately not only improves SCE but also the program speed and data retention time for NAND flash memory without any need of altering the gate oxide stack. We further investigate the impact of LDD depth variations on the reliability and erase performance of the NAND flash memory cell. In this work, we investigate a new physical phenomenon for short channel NAND flash memory cell. Here, it is shown that reduced doping concentration with higher value of d results in increased residual trap charge concentration after erase operation. According to the results presented for short channel SONOS memory cell, the effect of residual charge on the 'ON' current after erase operation become small, despite of low doping concentration of LDD region. Above all, in this thesis, we study the effect of channel engineering on the SCEs and data retention trade-off and program performance of the NAND flash memory string. Here, we investigate the program characteristics, data retention capability and SCEs for junction-free NAND flash memory string with half pitch range from 35nm to 12nm. Based on our analysis, we highlight that the retrograde channel doping approach can improve not only the SCEs but also the program speed and data retention time for 3D junction-free NAND flash memory.

1.8 Thesis Outline and Contributions

This thesis aims to investigate the performance of NAND flash memory from cell level to array level. In particular, Chapter 2 and 3 analyze the performance of NAND flash memory cell, while Chapter 4 examines the performance of NAND flash memory string, whose specific details are as follows.

- In Chapter 2, we investigate [2] the impact of S/D doping lateral straggle (σ_L) on the program characteristics, data retention, and SCEs for sub-25nm NAND

[2]The contribution of this chapter are presented in the following paper:

1. **Deepika Gupta** and Santosh Kumar Vishvakarma, "Improved Short Channel Characteristics with Long Data Retention Time in Extremely Short Channel NAND Flash Device," *IEEE Transactions on Electron Devices*, Vol. 63, No. 2, pp. 668-674, February 2016.

flash memory cell. Further, in this work, we consider threshold voltage roll-off($\triangle V_{th}$), Subthreshold Swing (SS), and Drain-Induced Barrier Lowering (DIBL) parameters to study the SCE for the considered NAND flash memory cell. We also examine the effect of varying lateral straggle on the junction boost leakage current [during the Program-Inhibition (P-I) mode] for the same. Based on our investigations, we have shown that adjusting the S/D doping lateral straggle σ_L appropriately not only improves SCE but also the program speed and data retention time for NAND flash memory without any need of altering the gate oxide stack. In addition, smaller lateral spread of the dopants in the channel with reduced S/D lateral straggle has shown to reduce the junction electric field ($E_{junction}$) during P-I mode of NAND flash operation.

- In Chapter 3, we investigate [3] a new physical phenomenon for short channel NAND flash memory cell. In particular, it is shown that reduced doping concentration with higher value of LDD depth results in increased residual trap charge concentration after erase operation. According to the results presented for short channel SONOS NAND flash cell, the effect of residual charge on the 'ON' current after erase operation become small, despite of low doping concentration of LDD region. This phenomenon occurs due to the drift of holes from channel into LDD region under the influence of strong negative gate electric field. Due to holes drift into LDD, high electric field is created at LDD-drain junction which accelerates the channel electrons. Therefore, it is concluded that presence of hole channel into LDD region affects only lateral electric field of short channel SONOS flash cell but still the vertical electric field is governed by LDD doping concentration. Additionally, degradation of program operation is found due to the presence of residual trapped charge.

- In Chapter 4, we investigate [4] the impact of retrograde channel doping on the

[3]The contribution of this chapter are presented in the following paper:

1. **Deepika Gupta** and Santosh Kumar Vishvakarma, "Impact of LDD depth Variation on the performance characteristics of SONOS NAND Flash Memory Device," *IEEE Transactions on Device and Material Reliability*, Vol. 16, No. 3, pp. 298-303, June 2016.

[4]The contributions of this chapter are presented in the following papers:

performance and reliability of 3D junction-free NAND flash memory string. Here, we mainly focus on the channel stacked vertical gate 3D junction-free NAND flash memory for the investigation of the different performance metrics. In this chapter, Specifically, we study the program characteristics, data retention capability and SCEs for junction-free NAND flash memory string with half pitch range from 35nm to 12nm. Here, a thin layer of high doping concentration of the order of 25% of total channel area may cause the improved performance. We consider $\triangle V_{th}$, SS and DIBL metrics to examine the SCEs. Further, we study the effect of variation in retrograde channel doping on SS and 'ON' current of the flash memory. Based on our analysis, we highlight that the retrograde channel doping approach can improve not only the SCEs but also the program speed and data retention time for 3D junction-free NAND flash memory, without altering the oxide stack in charge trap based flash memory. This mainly results from better gate controllability over the channel with retrograde channel doping.

1. **Deepika Gupta** and Santosh Kumar Vishvakarma, "Improvement of Short Channel Performance of Junction-free Charge Trapping 3-D NAND Flash Memory," *IET Micro & Nano Letters*, Vol. 12, No. 1, pp. 64-68, December 2016.
2. **Deepika Gupta** and Santosh Kumar Vishvakarma, "Improvement of Short Channel Performance and Memory Characteristics of Junction-Free Charge Trapping 3-D NAND Flash Memory,", Under Review.

Chapter 2

Analysis of NAND Flash Memory Cell with Source/ Drain (S/D) Lateral Straggle Variation

As discussed in previous chapter, memory chips with smaller area and low cost have got an increasing attention due to the booming market of portable electronic devices such as MP3 players and digital cameras smart phones, tablets etc [17]. As the scaling of NAND flash memory is going below 10nm, this industry is facing more reliability challenges to fulfill the requirements of handhold devices [37]. Conventional FG flash memory uses a polySi/metal layer as FG which stores information in the form of injected charge from the channel. FG flash memory, while scaling suffers from serious issues like cell to cell interference, lower coupling ratio and Stress Induced Leakage Current (SILC) [3],[38]. CT flash memory, precisely, Silicon Oxide Nitride Oxide Silicon (SONOS) has attracted a great deal of recent research to replace FG flash memory due to its MOSFET compatible structure in terms of involved fabrication process, low voltage operation and the immunity to extrinsic charge loss [4],[39]. Therefore, CT flash is a promising candidate to increase the scaling limit of NAND flash memory [12]. Recently, market for CT flash has increased with evolution of 3D NAND flash memories. 3D NAND flash memories mostly employs CT memory cell because of its high scalability and capability to stack [9]-[11].

Apart from this, a flash memory has been shown to be more prone to the Short Channel Effects (SCE) than any other logic device due to the requirement of thick TO to achieve high data retention [24],[40]. In particular, scaling of vertical gate stack in

CT NAND flash memory has reached its limits but Effective Oxide Thickness (EOT) used is still very high (15 - 20nm). This large EOT results in further degradation of short channel performance of NAND flash memory cell with the scaling [41]. Here, it is interesting to note that CT flash memory offers smaller EOT because it allows scaling of TO well below 7nm as compared to FG flash memory that suffers from SILC below 7nm [4]. Although this gate stack scaling improves speed and SCE for CT flash device but ultrathin TO results in poor data retention. Therefore, a flexible method that can improve short channel effects as well as data retention is of great importance in extreme short channel CT flash memory cell. Further, data retention of the flash memory cell has been shown to improve through the device width reduction [42]. This method for data retention enhancement may helps to improve SCE by allowing scaling of gate oxide stack.

Further, Ultra Shallow Junction (USJ) has also shown to eliminate the problem of SCE in flash memory [26]. Specifically, USJ results in reduced dopant impurities spread from Source/Drain (S/D) into channel which in turn decreases lateral electric field of device to counter SCE. However, fabrication of USJ requires highly controlled and sophisticated environment which is difficult. Also, USJ is sensitive to the defects and suffers from junction leakage [27]-[28]. In addition, high channel doping concentration is also considered as a possible solution for reducing SCEs in NAND flash memory cells. However, this method raises electric field at S/D and channel junction which causes high junction leakage current to flow during Program-Inhibition (P-I) mode of NAND flash memory operation [25]. This junction leakage current reduces boosted channel potential for P-I memory cells and may results in unwanted program/erase operation. Thus, high channel doping concentration may improve SCE but it may not be a suitable approach in sub 20nm NAND flash memory cells. In a recent study, improved channel boosting has been obtained during P-I mode with the lightly doped S/D region for CT flash memory cells [29]. However, for the same study, it has also been displayed that SCE degrades with lighter S/D doping. Therefore, exploring a method to improve SCE and channel boosting simultaneously will have great significance for future flash memory cells.

Furthermore, several works have devoted their efforts to improve SCE of MOSFET and Tunnel Field Effect Transistor (TFET) by controlling lateral straggle of S/D doping (σ_L) profile [43]-[45]. Here, it has been shown that reduced σ_L enhances the gate electrostatic integrity over channel region and helps to reduce SCE. As flash memory has different operating principle, its tunneling behavior with different σ_L can not be directly predicted from the available results of MOSFET and TFET. Hence, investigating the effect of σ_L on the various performance metrics of NAND flash memory such as program speed, data retention, SCE and junction boost leakage current during P-I mode is crucial for short channel flash memory cells. In addition, authors in [24] have investigated the effect of variations of S/D implantation energy on Random Telegraph Signal (RTS) for NAND flash memory cells. It has been shown that variation in S/D implantation energy which in turn alters the σ_L of S/D doping profile changes the RTS performance of extremely scaled flash memory cell. RTS inverse scaling behavior with SCE is also shown for 25nm flash memory cells. However, understanding of other short channel characteristics with different σ_L of S/D doping and its influence on NAND flash memory cell performance is still insufficient.

Motivated by the facts highlighted above, in this chapter, we study the effect of σ_L on program characteristics, data retention and short channel performance metrics of considered SONOS NAND flash memory cell. For analysing SCE, metrics such as $\triangle v_{th}$, SS and DIBL are considered. Moreover, channel energy band analysis is carried out to justify the results presented for data retention. Further, we investigate the variation of junction electric field and junction leakage current during P-I mode. This provides an explanation for the significance of σ_L in bulk NAND flash memory. Here, we also consider various gate lengths to provide more insights related to the influence of lateral abruptness on memory cell performance.

2.1 Device Design and Simulation Setup

In this work, we consider a SONOS NAND flash memory cell as shown in Figure 2.1. High voltage is applied to gate terminal (V_G) during program or Erase operation so as to

generates sufficient high electric field at TO for desired tunneling of charge carrier into or out the CT layer. S/D terminals are set to ground to ensure FN tunneling mechanism for NAND operation of flash memory. Here, substrate terminal is also kept at ground. In the considered memory cell, silicon di-oxide (SiO_2) is used both as TO and IPD whereas silicon nitride (Si_3N_4) is used as CT layer with thickness 2.5nm, 6nm, 6nm respectively in accordance with the experimental result explained in [46]. Further, electron and hole tunneling mass is considered as a fitting parameter to achieve the desired result as shown in Table 2.1 for 40nm generation. It is difficult to predict the exact value of the tunneling mass because This is because, while tunneling, charge carrier is placed neither at the bottom of Conduction Band (CB) nor at the top of the Valance Band (VB) of the tunneling dielectric. However, for F-N tunneling, authors in [21] have suggested the tunneling mass value as $.45m_0$ through SiO_2 barrier and, hence, adopted in our study. Here, m_0 represents the free tunneling mass of charge carrier. In addition, a P type silicon substrate with starting uniform boron concentration of $5 \times 10^{17}/cm^3$ is used. Further P^+ poly-silicon is used as gate material to reduce back tunneling and to achieve the required threshold voltage with doping concentration $1 \times 10^{20}/cm^3$.

The Gaussian doping profile in the S/D region is modeled as [43]

$$N_{SD}(x) = N_{SD}(p)e^{-(x^2/2\sigma_L^2)}$$

Where $N_{SD}(p)$ is the peak of Gaussian S/D doping profile which is set to $1 \times 10^{20}/cm^3$. Here, we have considered the underlap arsenic doped S/D region to minimise the effect of overlap capacitance on the performance of NAND flash memory cell [47]. Considered S/D doping profile for SONOS CT flash memory cell is shown in Figure 2.2. In this figure, the value of σ_L varies[1] from 8 nm/dec to 16 nm/dec.

Results are also obtained for uniform doping for comparison purpose. Here, simulations are carried out using Synopsys Sentaurus [50] simulation tool to investigate the memory cell performance parameters with different σ_L. In this analysis, variations in

[1]Here, to calculate σ_L, peak of S/D doping is set at a distance of 60nm from the center of the channel. Note that, Higher word-line (WL) to WL spacer distance is necessary to minimise interferences from neighboring memory cells in array at smaller technology node [48]-[49]. Further, σ_L of doping is calculated for the distance where Gaussian doping drops to the minimum value of $5 \times 10^{17}/cm^3$. Since, the appropriate value for σ_L of S/D doping is undefined in the ITRS roadmap hence, its range has been chosen to satisfy more practical doping condition which is quite acceptable.

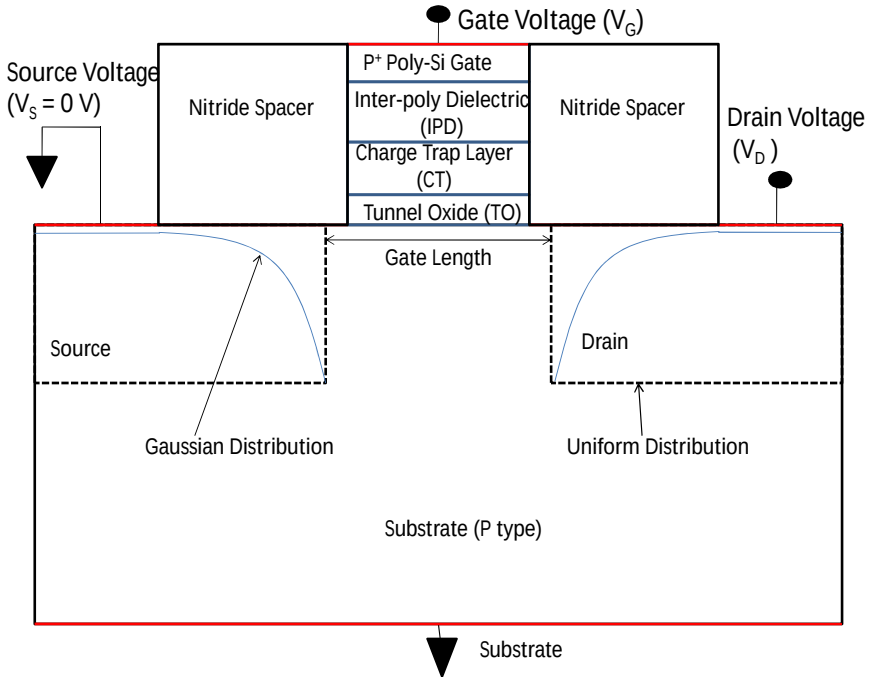

Figure 2.1: SONOS NAND flash memory cell

threshold voltage shift is monitored by using constant current method ($I_D = 10\mu A$ per μm width at $V_{DS} = 0.1$ V). Here, the program/erase operation is performed with gate voltage 17 V/-18 V with program/erase time set to 2.5 ms/7.5 ms respectively. Further, nonlocal model for band to band tunneling is used for tunneling of charge carrier across barrier to reproduce the experimental results. In non-local tunneling, the electric field at each point in the tunneling path is dynamically changing. This makes tunneling a nonlocal process, hence model is dynamic non-local tunneling model. Hereby, the transport of charge carrier in substrate is performed with conventional drift diffusion model along with other physical model such as SRH generation-recombination model, Poole Frenkel model with concentration and field dependent mobility. Simulation also comprises transient analysis to fulfill trapping and release of charge carrier across the oxide barrier.

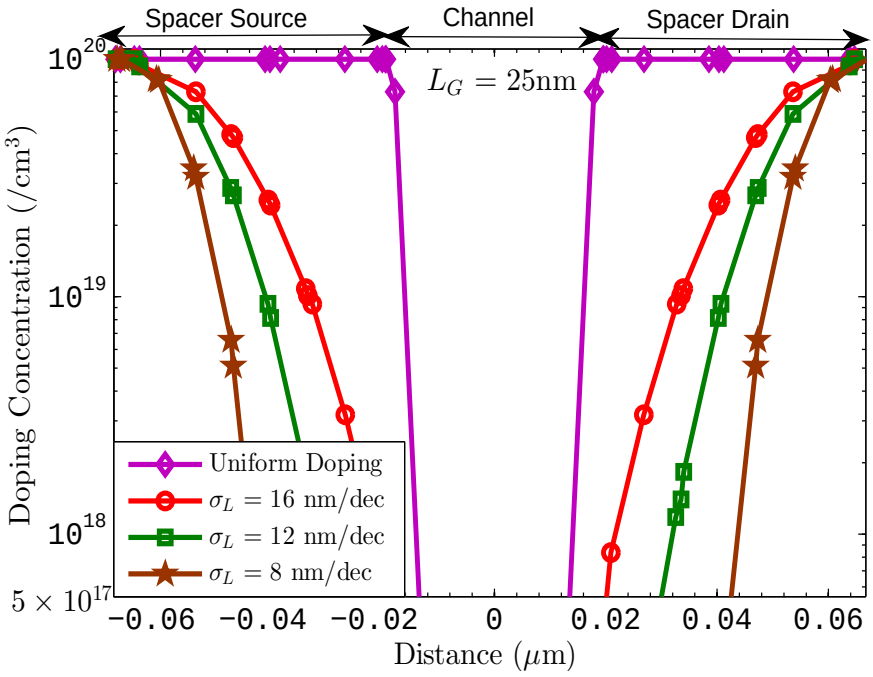

Figure 2.2: Variation of doping concentration with varying σ_L of S/D doping. This Figure also shows the comparison of transfer characteristics of NAND flash memory cell when S/D region are doped with gaussian and uniform doping

2.2 Results and Discussion

In this section of the chapter, we present simulation results for the considered SONOS flash memory cell. Firstly, we present the variation of program characteristics and data retention with different σ_L. Secondly, this chapter investigates the effect of varying σ_L on different short channel parameters such as $\triangle V_{th}$, SS and DIBL. Further, values of junction leakage current is explored during P-I mode for various values of σ_L. Note that, throughout this section the distance 0 in graphs is aligned with the center of the memory cell channel while the negative and positive distance point toward the source and drain regions of the cell respectively.

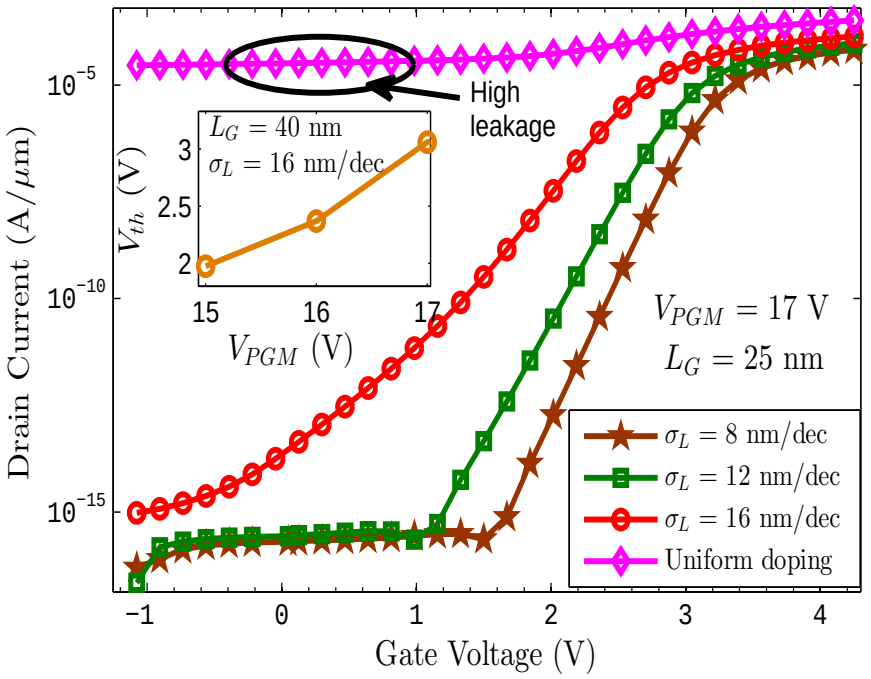

Figure 2.3: Variation of program characteristics with varying σ_L of S/D doping. Inset shows the Incremental Step Pulse Programming (ISPP) characteristics of 40nm SONOS NAND flash memory [46]. Here, V_{th} = program threshold voltage and V_{PGM} = program voltage with $V_{initial}$ set to 0 V

2.2.1 Effect of S/D Lateral Straggle (σ_L) on Program Characteristics of Flash Memory Cell

Figure 2.3 shows the transfer characteristics ($I_D - V_G$) of considered SONOS memory device with 25nm technology for different σ_L under programmed state. A clear increase of threshold voltage shift ($V_{th} - V_{initial}$) is observed with smaller σ_L. Therefore, one can infer that with reduced σ_L, memory cell can be programmed with smaller programming time to achieve equal threshold voltage shift. Here, initial threshold voltage ($V_{initial}$) for virgin device has been set to 0 V. Further, it is interesting to mention that increase in threshold voltage shift with smaller σ_L is attributed to larger vertical electric field at TO. Further, with inset of the Fig.2.1, we have provided the incremental step pulse programming (ISPP) curve of our considered device, which is well aligned with the experimental result presented in [46] and shows the slope of the curve less than 0.6.

To obtain ISPP curve pulses of increasing amplitude are used to inject a constant, controlled amount of charge into the CT layer at each step, interleaved with program verify phases in which V_T is monitored.

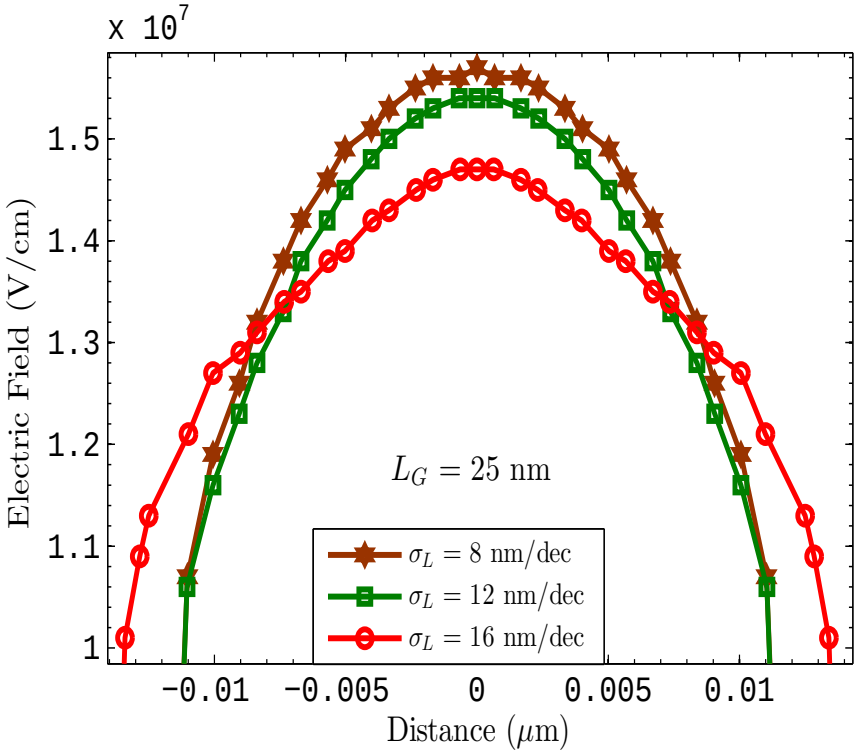

Figure 2.4: Variation of electric field in TO with varying σ_L (Increase in electric field at TO with reduction of σ_L clarifies that the threshold voltage shift is resulting from electron trapping in CT layer instead of increase in channel resistance)

Figure 2.4 depicts the variation of electric field in TO with different value of σ_L for 25nm technology. To plot this figure, a cut plane is inserted in the direction of channel length (X - direction) at TO layer of the 2D NAND flash memory device. The cut-plane is taken near the substrate-insulator interface to maximize the obtained electric field. Here, one can observe that electric field at TO increases with the reduction of σ_L. Note that, higher carrier confinement occurs in channel under the influence of higher vertical electric field with smaller σ_L. As a consequence, the offset between conduction bands of channel and TO tends to reduce, resulting into higher carrier tunneling. This eventually increases trapping of charge carriers in CT layer, as a result, threshold voltage shift increases. Importantly, increase in electric field at TO with reduction of σ_L clarifies

that the threshold voltage shift is resulting from electron trapping in CT layer instead of increase in channel resistance. Further, higher trapping in CT layer increases its potential which in turn results in reduced channel conduction, following the condition [42]

$$\sigma/\sigma_0 = e^{-q\delta/K_BT} \tag{2.1}$$

where σ and σ_0 are the channel conductivity when CT layer is charged or not charged, respectively. Here, δ is the potential increase of CT layer due to trapping of carriers. To confirm the hypothesis presented above, we investigate the program speed of the memory cell with different σ_L.

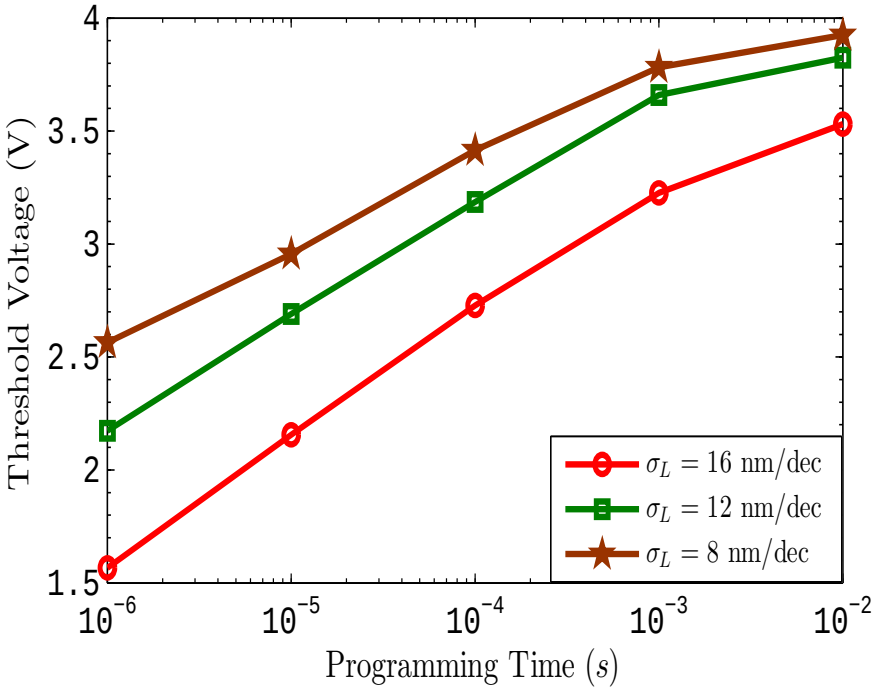

Figure 2.5: Variation of program speed with different σ_L for $L_G = 25nm$

Further, Fig. 2.5 depicts that as the program time reduces for the NAND flash memory device, the threshold voltage of the device also decreases. This effect arises due to reduced number of electron crosses the barrier and get trapped in CT layer with

Table 2.1: Comparison of Threshold Voltage Shift ($V_{th} - V_{initial}$) with Varying σ_L (nm/dec) for different L_G.

σ_L (nm/dec)	$V_{th} - V_{initial}$ (V)		
	$L_G = 40nm$	$L_G = 25nm$	$L_G = 20nm$
16	3.05	2.69	2.49
12	3.25	3.02	2.90
8	3.36	3.20	3.10

smaller program time. Hence, observing high memory window with small program time is important to improve the speed of the flash memory device. Also, One can observe from Figure 2.5 that higher speed is obtained with reduced σ_L. This phenomenon result from the above mentioned higher electric field in TO with reduced σ_L. Also, it can be deduced from Figure 2.5 that the considered NAND flash memory cell with smaller σ_L can maintain higher program threshold voltage for lower program time of $1\mu s$ as well. As a result, NAND memory cell with smaller σ_L may tolerate larger threshold voltage roll off while scaling. This leads to the improved SCE characteristics with smaller σ_L device.

Further, for comparison, values of threshold voltage shift by varying σ_L in program state is presented in Table 2.1 for different gate length. However, to completely erase the memory cell longer erasing times may be required. Over to that, we make a note here that the amount of leakage current in SONOS memory cell with uniformly doped S/D region is very high (refer to Figure 2.3) as compared to the Gaussian doped S/D region. Therefore, it suggests that uniformly doped S/D region is less suitable for NAND flash memory cell with extremely small gate length as compared to Gaussian doped S/D region. One can observe from Figure 2.3 that leakage current with gaussian doping decreases with the reduction of σ_L. As already discussed, this leakage current reduction is due to smaller channel conductivity derived from high trapping in CT layer with smaller σ_L.

2.2.2 Effect of S/D Lateral Straggle (σ_L) on Data Retention Characteristics of Flash Memory Cell

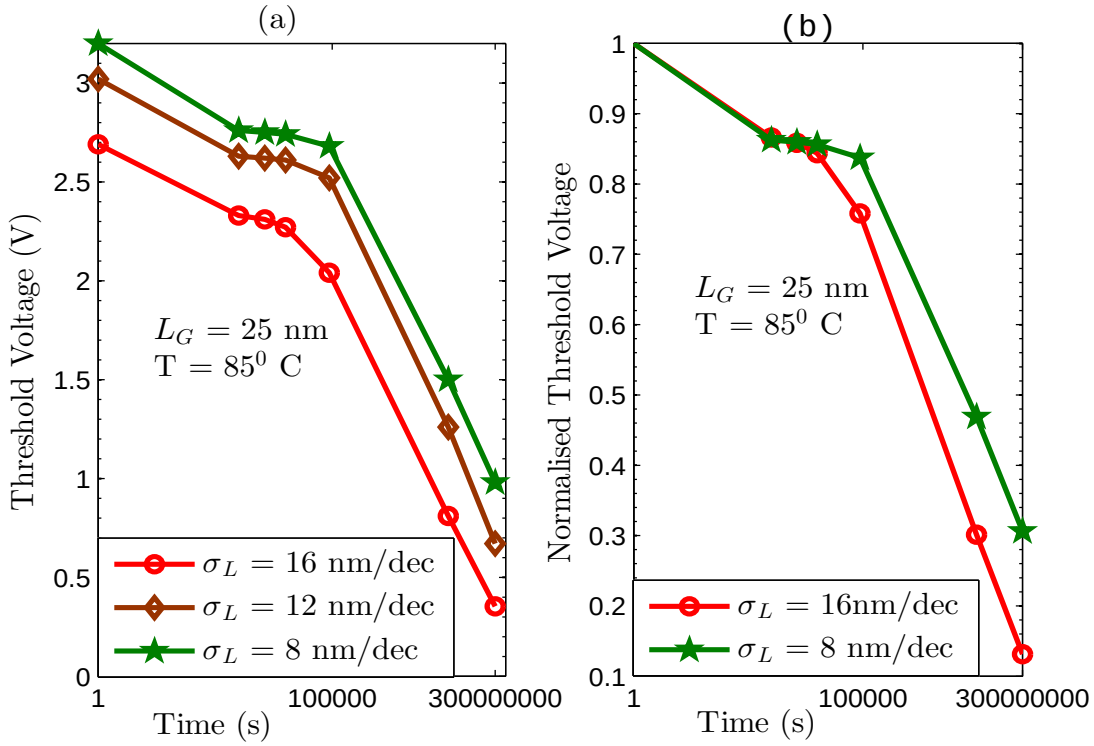

Figure 2.6: (a) Variation of data retention with varying σ_L (b) Comparison of normalized retention for different σ_L.

As we know, the charge retaining capacity for 10 years, while the cell is idle characterises the cell retention. Figure 2.6(a) shows the simulated data retention characteristics for the considered 25nm SONOS NAND flash memory cell with different σ_L. To find retention, we first programmed the 25nm NAND flash memory cell with 17 V gate voltage. After that, the gate voltage is changed to 0 V and then the monitored the considered memory device behavior for 3×10^8 s. It is to note that no external field is applied to the considered flash device during the retention operation. The charge and hence, the threshold voltage of the flash memory device reduces with time. Here, the charge leakage is observed due to the leakage through the dielectric TO and IPD. Specifically, the leakage of the charge increases with temperature. Hence, the simulation for retention operation is carried out for 85^0C using Santaurus TCAD tool. From

the Figure 2.6(a), it is confirmed that the data retention capability of the memory cell is high with reduced σ_L. Further, as shown in Figure 2.6(b), the normalized V_{th} at $3 \times 10^8 s$ with σ_L = 8 nm/dec is still 2.5 times larger than compared to V_{th} with σ_L = 16 nm/dec. Additionally, at the scale of time, the retention time with smaller σ_L is approximately 10 times higher than larger σ_L. High retention with reduced σ_L signifies small charge leakage from CT layer.

This phenomenon is again attributed to the above mentioned higher charge trapping in CT layer with smaller σ_L. Here, high charge trapping in CT layer with smaller σ_L attracts high hole concentration towards the $Si - SiO_2$ interface during retention operation. Note that, this high carrier confinement in channel lifts up the energy level in the channel with reduced σ_L, as clearly shown in Figure 2.7.

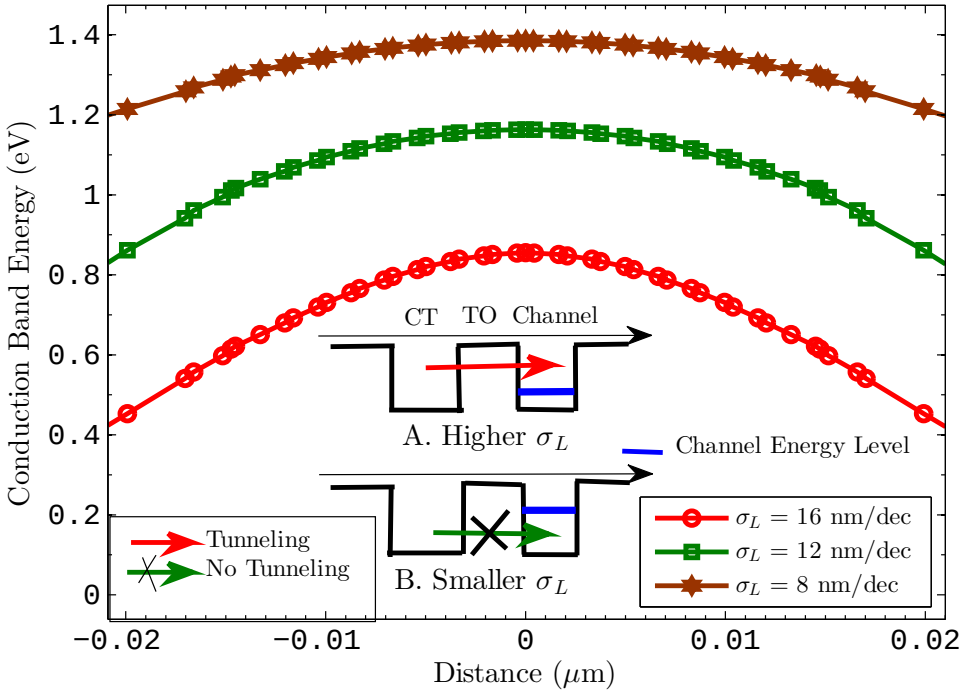

Figure 2.7: Variation of conduction band energy in channel with varying σ_L. Inset shows the effect of energy bands on data retention characteristics of NAND flash memory cell.

Further, in this figure, inset shows the effect of channel energy band diagram on tunneling of charge carrier from CT layer to channel during retention mode [51]. This clearly depicts that smaller channel energy level results in possible tunneling of electrons

which causes loss of data from CT layer. However, amount of tunneling reduces as channel energy level goes up which results in improved retention. Therefore, increased channel energy level with smaller σ_L results in improved data retention characteristics of NAND flash memory cell without altering the thickness of TO. Importantly, to obtain the Fig. 2.7, a cut-plane is inserted in the channel near substrate-insulator interface in the channel length direction.

2.2.3 Impact of S/D Lateral Straggle (σ_L) on Short Channel Effects of Flash Memory Cell

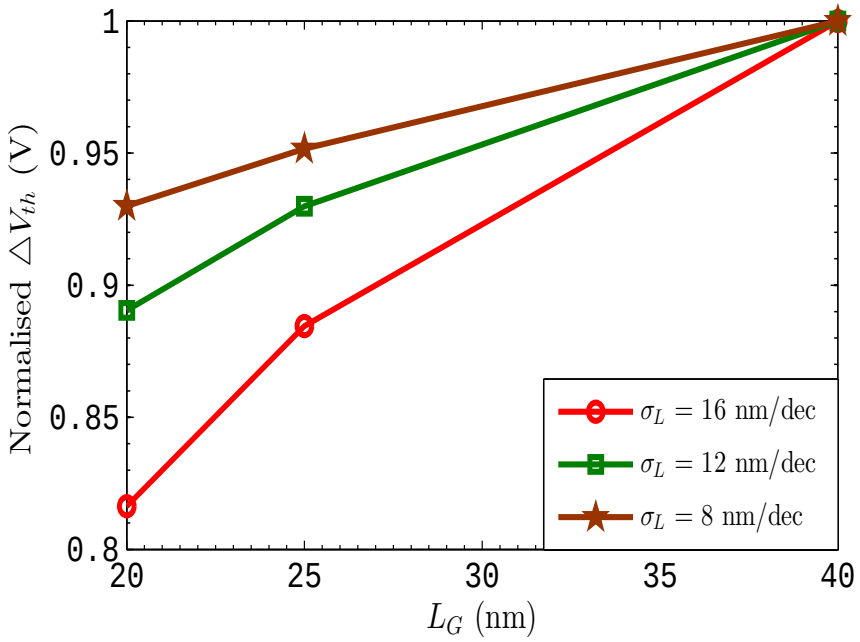

Figure 2.8: Variation of threshold voltage roll-off with varying L_G for different σ_L.

In this section, we investigate the variations of different short channel parameters with σ_L. For this, Figure 2.8 plots the normalized $\triangle V_{th}$ (for different σ_L) with gate length varying from 40nm to 25nm. Value of $\triangle V_{th}$ has been normalised to unity for all value of σ_L at 40nm gate length. One can see that, $\triangle V_{th}$ from 40nm to 25nm is found to be minimum for σ_L = 8 nm/dec. This phenomenon is attributed to the reduced spreading of donor impurities from S/D, which efficiently enhances the gate

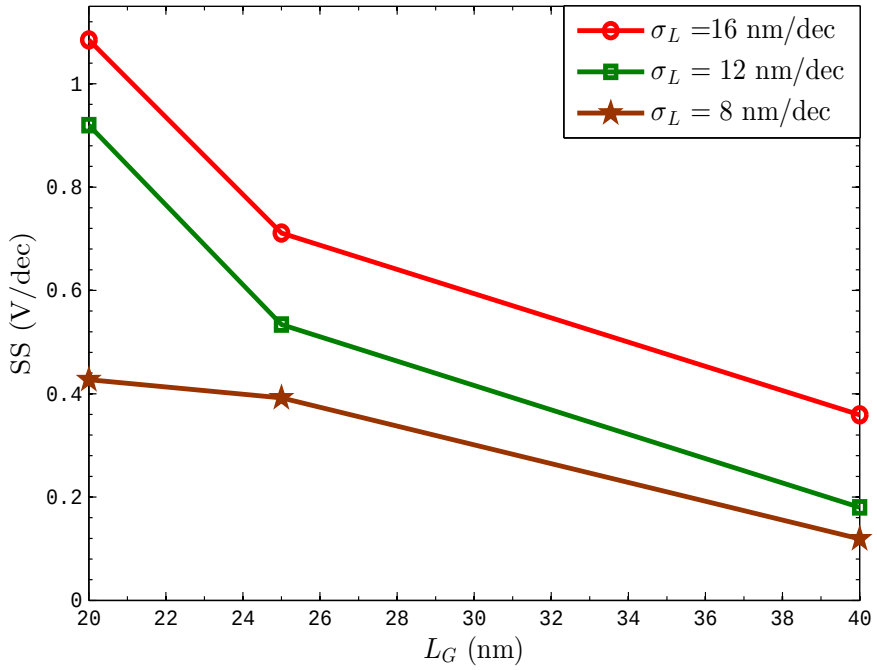

Figure 2.9: Variation of SS with varying L_G for different σ_L.

Table 2.2: Comparison of Different Short Channel metrics with Varying σ_L (nm/dec) for Different L_G.

σ_L (nm/dec)	$\triangle V_{th}$ (V)		SS (mV/dec)			DIBL (V/V)		
	L_G=25nm	L_G=20nm	L_G=40nm	L_G=25nm	L_G=20nm	L_G=40nm	L_G=25nm	L_G=20nm
16	0.45	0.20	359	711	1085	0.388	0.778	1.00
12	0.23	0.12	180	534	920	0.22	0.404	0.529
8	0.16	0.10	119	392	427	0.143	0.239	0.292

controllability over channel area through high TO electric field. Further, it can be seen that smaller σ_L may result in reduced channel capacitance due to the small number of electrons intruding from S/D into channel. This consequently reduces lateral channel electric field and hence, SCE characteristics for NAND flash memory cell improves. Therefore, reduced σ_L improves the $\triangle V_{th}$ for NAND flash memory.

Further, for CT NAND flash memory cell $\triangle V_{th}$ results in SS degradation and increased DIBL [52]. Hence, understanding the variation of SS and DIBL with σ_L may be helpful for further scaling of bulk NAND flash cell. Figure 2.9 and 2.10 depicts the variation of SS and normalised DIBL (DIBL normalised to unity for all values of σ_L at

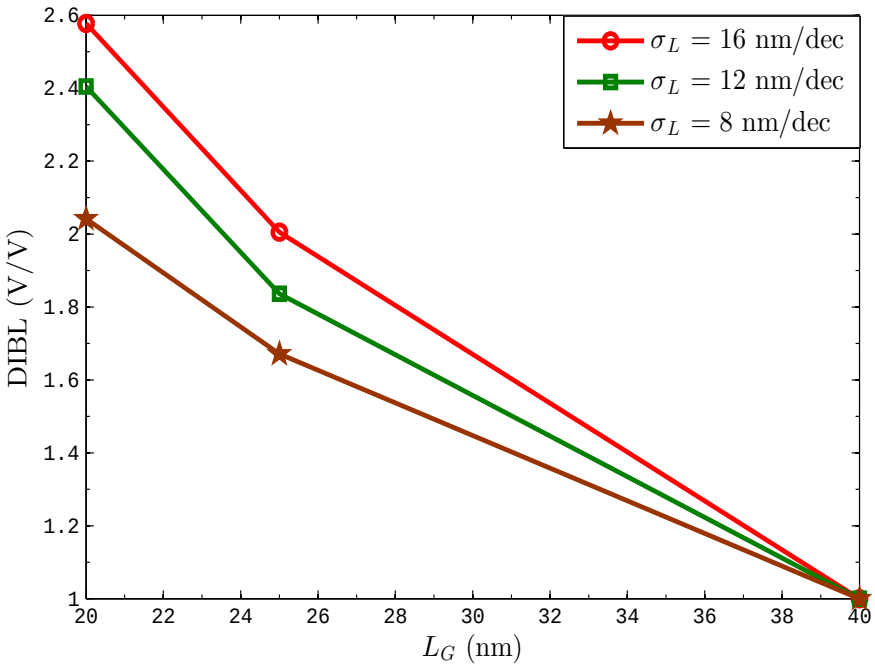

Figure 2.10: Variation of DIBL with varying L_G for different σ_L.

L_G = 40nm) of NAND flash memory cells for different σ_L respectively. Figure clearly depicts smaller SS and DIBL with reduced σ_L for all gate lengths. This is again related to the reduction of impurity atoms in the channel with reduced σ_L. Note that, it reduces interaction between vertical and horizontal electric field during program/erase operation, hence results in better gate voltage control over channel. Better gate electrostatic integrity results in lower channel leakage current, hence, improves SS and DIBL for NAND flash memory devices. Numeric results for the comparison of $\triangle V_{th}$, SS and DIBL with different values of σ_L has been presented in Table 2.2.

2.2.4 Effect of S/D Lateral Straggle (σ_L) on Memory Cell Characteristics During Program-Inhibit Mode

During program operation, channel potential of a P-I[2] NAND flash memory cell in memory array must be boosted to a high voltage. This voltage is normally around

[2]For brevity, here we have omitted the details of P-I mode of NAND flash operation which can be found in [29], [41].

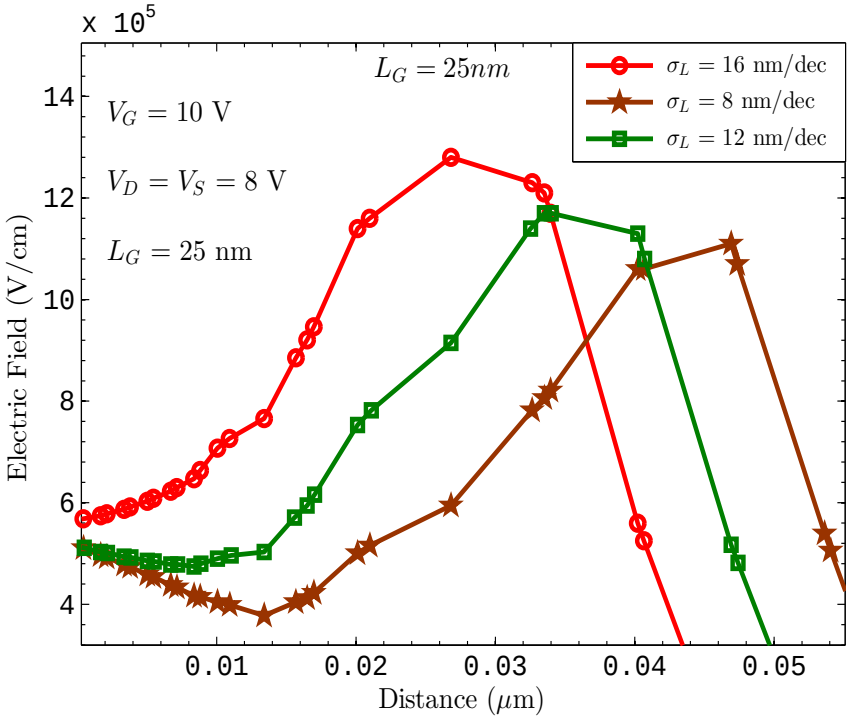

Figure 2.11: Variation of electric field at drain-channel junction during program-inhibit mode of flash memory operation with varying σ_L.

8-10 V with V_G=10 V applied to P-I flash flash memory cell, to prevent F-N tunneling between channel and trap layer [53]. As a result, during P-I mode, S/D and channel junction must be able to withstand high voltage stress. Otherwise, high junction electric field ($E_{junction}$) at S/D and channel junction will cause larger junction boost leakage current ($I_{leakage}$). This $I_{leakage}$ reduces the boosted channel potential and in turn results in failed P-I operation. The program disturbance due to high $I_{leakage}$ has been accounted as a main concern for such sub-20nm NAND flash memory cell. Many authors have devoted their efforts to minimise the channel boost leakage current [54]-[55].

Here, to examine the P-I mode of NAND flash operation, required voltage (V_G = 10 V and $V_D = V_S = 8$ V) is applied to a single bulk NAND flash memory cell. This method has been used to investigate the device characteristics during P-I mode for NAND flash memory array [41],[56]. Note that, Fig. 2.11 shows the electric field at drain channel junction with various σ_L for 25 nm technology during P-I mode. It is found that memory cell with reduced σ_L exhibits reduced $E_{junction}$ at channel-drain/source

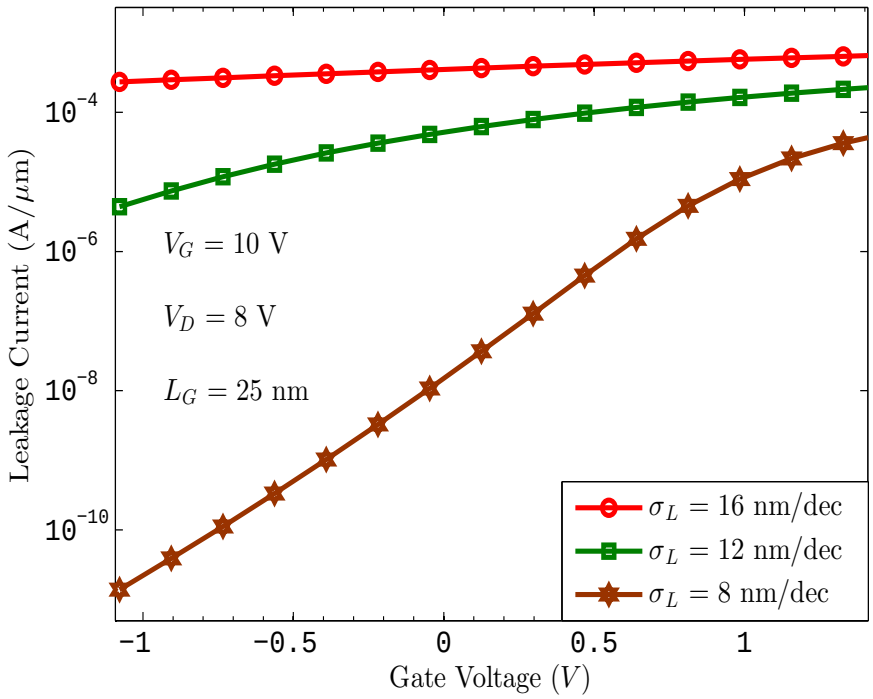

Figure 2.12: Variation of channel boost leakage current during program-inhibit mode of flash memory operation with varying σ_L for different L_G.

junction similar to P-N junction operation. Following the P-N junction electric field profile, $E_{junction}$ reduces on both sides of channel-drain junction. In Fig. 2.11, $E_{junction}$ has approximately reduced by 20% as σ_L has decreased from 16nm/dec to 8nm/dec. Consequently, lower $I_{leakage}$ flows through the memory cell during P-I mode with smaller σ_L of S/D doping, as shown in Fig. 2.12.

One can observe from Fig. 2.12 that $I_{leakage}$ has reduced from 2.72×10^{-4} $A/\mu m$ to 1.39×10^{-11} $A/\mu m$ with σ_L changing from 16nm/dec to 8nm/dec respectively. Therefore, proper choice of lateral spread of S/D doping may improve the flash memory cell characteristics during P-I mode.

In light of the above, the proposed approach leverages the development of high density future NAND flash memories with faster speed at extremely scaled technology.

2.3 Summary

In this chapter, we have investigated the effect of S/D doping lateral straggle σ_L on flash memory cell program characteristics, data retention performance and SCE. We have also studied the variation of channel boost leakage current during program-inhibit mode with σ_L. Here it is shown that, smaller value of σ_L results in better program speed and 10 year data retention. This results from the higher carrier confinement within channel with smaller value of σ_L of S/D doping. Further, utilising this junction engineering in NAND flash memory also suppresses SCE such as $\triangle V_{th}$, SS, DIBL for sub-25 nm generation. Improved SCE, without any reduction of gate stack, results from better gate controllability over channel region with reduced σ_L. Further, smaller σ_L reduces channel boost leakage current in sub-25 nm NAND flash and suppresses any chances of program disturbance with bulk NAND flash memory. Therefore, our proposed approach to control S/D σ_L may be applied to develop future high density NAND flash memory.

Chapter 3

Analysis of NAND Flash Memory Cell with LDD Depth Variation

In the chapter 2, we analyzed the performance of 25nm NAND flash memory cell with different S/D lateral straggle. Thereby, we understood the impact of S/D doping and operating condition on the performance of the NAND flash memory cell. However, in this study, we resorted to the performance analysis of lightly doped drain (LDD) based NAND flash memory cell.

As discussed, FG flash memory suffers from serious challenges while saling [3], [40]. CT flash memory, precisely, SONOS has attracted recent research interest to replace FG flash memory due to its MOSFET compatible structure in terms of involved fabrication process, low voltage operation and the immunity to extrinsic charge loss [38], [39], [40]. Therefore, CT flash is a promising candidate to increase the scaling limit of flash memory [12]. However, scaling of the CT flash memory cell causes the electric field to increase around the drain region which give rise to the SCEs. Specifically, a high electric field at the drain results in the impact ionization, by virtue of which hot carrier injection occurs from substrate into the gate oxide stack. This may degrade the reliability and the performance of the CT based NAND flash memory cell [57].

Importantly, the LDD based structures have been shown able to reduce the short channel effects and drain disturbance in MOS devices [30],[58]. LDD based devices, in general, provides larger gradient of energy band between channel and the drain regions which opposes gate electric field above LDD region. This in turn lowers the breakdown

voltage and suppress the impact ionization process at the drain. As a consequence, hot carrier injection in the device greatly reduces. For the same region, the LDD based structures are used in flash memory cell. Here, presence of LDD region may help to improve the reliability of the NAND flash memory cell. However, in a recent study [31], authors have shown the performance of LDD based SONOS NAND flash memory cell where it is presented that lower electric field at LDD region does not allow electrons in CT layer to release during an erase operation. This phenomenon occurs in the region above the LDD in CT layer. This causes a significant amount of electron trap charge to remain in the CT layer after the erase operation. This residual charge may cause increase in erase threshold voltage, degraded subthreshold swing, slower erase speed. Therefore, presence of LDD in NAND flash memory, apart from giving some advantages, may cause mendacious memory operation due to residual charge. Additionally, the residual charge in the nitride layer may introduce the reliability issues in charge trap flash memory cells [29]. Particularly, LDD depth variations in short channel NAND flash memory cells occur due to random implantation scattering at junctions. This may alter the LDD depth arbitrarily and cause reliability issues [31]. In fact, in large array, this effect is more pronounced as the devices in array may have different LDD depth and hence different performances. Although, the LDD based devices are known to provide improved flash memory performance but the practical knowledge of trapped residual charge and its effect on performance characteristics and reliability of short channel CT based NAND flash memory cell with the variation of LDD depth is not known. Hence, this chapter captures the effect of LDD depth variations on the reliability of NAND flash memory array.

Motivated by the above, we first investigate the effect of varying LDD depth (d) on the 'ON' current while reading after erase operations and then quantify the residual charge in the nitride layer for a short channel CT based NAND flash memory cell. Further, effect of residual charge on the program operation of CT based memory cell is shown. To the best of our knowledge, no previous works addressed the performance, behavior and physics of a short geometry CT based memory cell with varying LDD depth. Based on our study, we found a peculiar relation between program/erase per-

formance and residual trapped charge for short channel CT based flash memory cell. Here, we also present several useful insights for the performance evaluation of the short channel NAND flash memory cell.

3.0.1 Device Description and Simulation Details

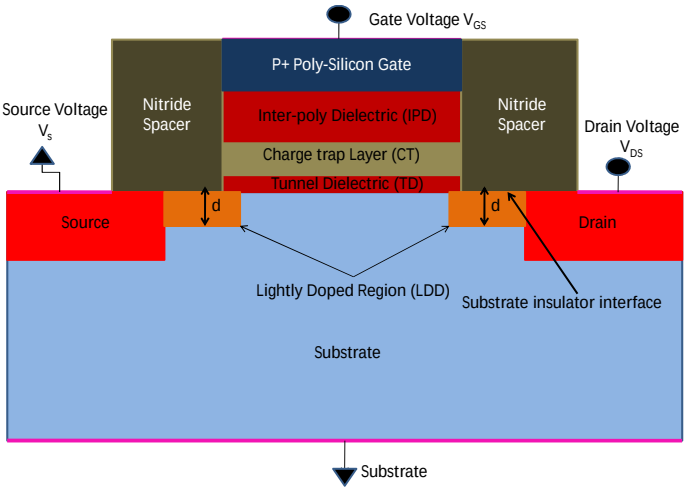

Figure 3.1: Schematic diagram of LDD based SONOS NAND flash memory cell

In this chapter, we consider a SONOS NAND flash memory cell [59] with lightly doped source and drain regions as shown in Figure 3.1. Here, junction based SONOS flash memory cell is critically observed which is readily used as a basic building block for NAND flash memory array [29], [60]. Further, to accurately model the considered SONOS flash memory cell, we use conventional drift diffusion phenomenon for the transport of charge carriers. Later, we take in account the Shockley-Read-Hall model for the charge carrier generation and recombination. In addition, the Poole-Frenkel model is selected for the charge transport in the dielectrics. We also include nonlocal band-to-band tunneling model at the substrate-oxide interface with high gate electric field.

Herein, the memory cell structure consist of a P-type silicon substrate with the starting boron concentration of $5{\times}10^{17}/cm^3$. We consider a Gaussian doping profile for

each of the LDD, source (S) and the drain (D) region to show more practical fabrication conditions. Here, the Gaussian doping profile is modeled as

$$N_{SD}(x) = N_{SD}(p)e^{-(x^2/2\sigma_L^2)}$$

$$N_{SD-LDD}(x) = N_{SD-LDD}(p)e^{-(x^2/2\sigma_L^2)}$$

where $N_{SD}(p)$ and $N_{SD-LDD}(p)$ are the peak of Gaussian doping concentration in S/D region and LDD region near S/D respectively. The S/D regions are identical with the peak doping concentration of $1{\times}10^{20}/cm^3$ for each region. Similarly, identical LDD regions with the peak doping concentration of $8{\times}10^{17}/cm^3$ is implemented near S/D. Here, for all the comparisons, value of the Gaussian doping profile straggle (σ_L) from peak position is kept constat. Further, we have used SiO_2 as TO and IPD with Si_3N_4 as CT layer. The thickness for each of the tunnel dielectric, IPD and CT layer is taken as 2.5nm, 6nm and 6nm respectively, according to the experimental results presented in [59]. We also use the P^+ polySi as the gate material where a doping concentration of $1{\times}10^{20}/cm^3$ is required to obtain the desired threshold voltage shift. We have an initial depth for lightly doped source and drain structure as 5nm while it is 40nm for each of the source and the drain regions. In our present study we vary the LDD depth within the range of 5-11nm, according to the results presented in [61], for the considered short channel memory cell.

Now, by varying d, the change in threshold voltage is monitored according to the constant current method (I_D = 50nA/μm at V_{DS}= 1V). Here, we apply a gate voltage of 17V to perform the program operation while -18V for the erase operation. During program/erase operation, the S and D terminals are set to ground potential (0V) to ensure NAND operation for flash memory. Furthermore, we set the time of 2.5ms and 7.5ms for the program and the erase operation respectively. For our experiments, the region of probe is considered at a distance of 2nm below from the $Si - SiO_2$ interface. We obtain performance characteristics and results for our memory cell by simulating through the Sentaurus TCAD device simulation tool from Synopsys technology [50]. In

the simulations, transient analysis is carried out for effective trapping and release of the charge carriers.

3.1 Results and Discussion

Throughout this section the distance 0 is aligned with the center of the channel while the negative and positive distance point toward the source and drain regions respectively.

3.1.1 Effect of Varying LDD Depth (d) on the 'ON' Current of SONOS Memory Cell After the Erase Operation

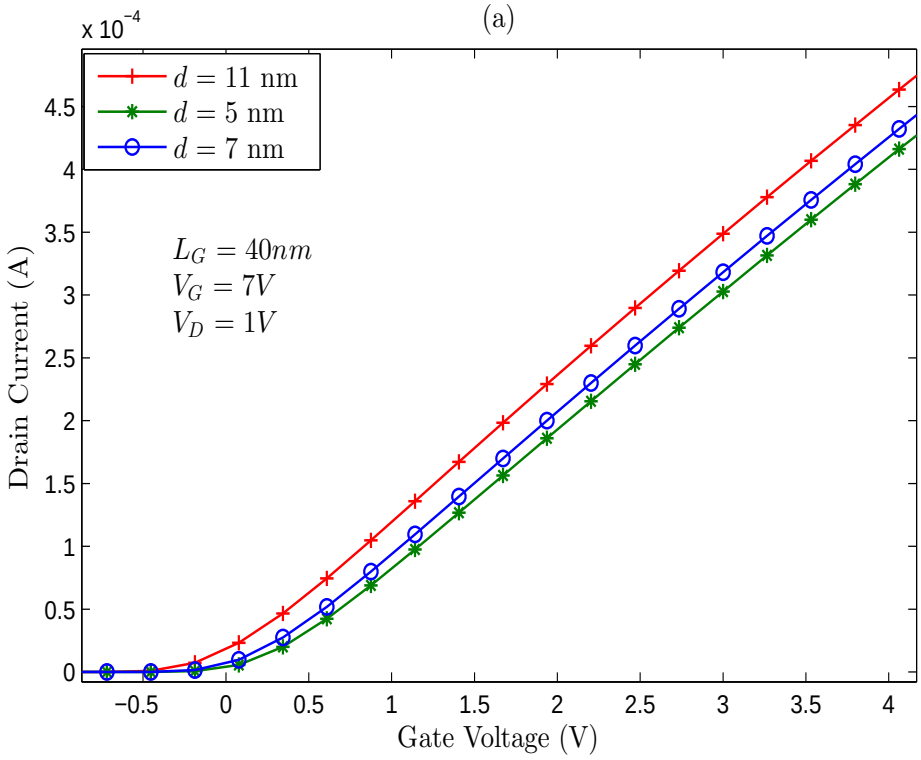

Figure 3.2: Transfer characteristics of memory cell after erase operation

Figure 3.2 shows the transfer characteristic ($I_D - V_{GS}$) of considered flash memory cell after the erase operation. The transfer characteristics is obtained during read operation with V_G = 7V and V_D = 1V to monitor changes in erase state. Note that, the 'ON' current during read operation flows in accordance with the charge released

from the CT layer during erase operation. Here, the plots for various values of d has obtained. We consider equal erase time for each set of data. It is clear from the various curves that the 'ON' current after erase operation rises with an increase in the d from 5nm to 11nm. Interestingly, it is important to note that, increase in d reduces the doping concentration within LDD region according to the following relation

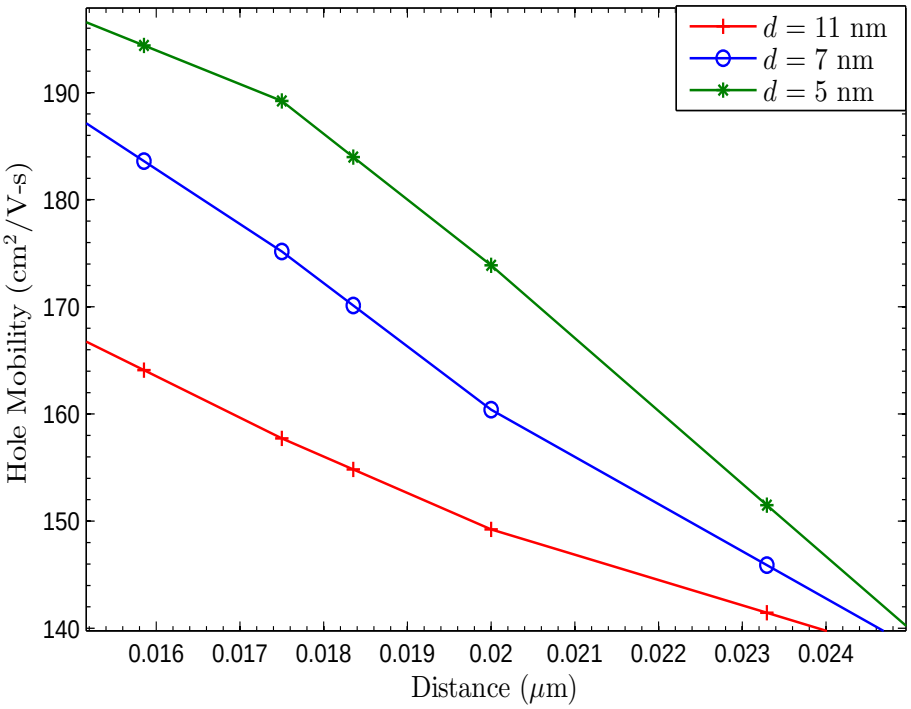

Figure 3.3: Variation of hole mobility in LDD region near drain with LDD depth

$$\bar{N}_D = N_D/d,$$

where, $\bar{N}_D$ is the average doping concentration in LDD region and N_D is the peak doping concentration. From the equation, it is clear that as d increases, average doping concentration in LDD reduces. This reduced doping concentration must have resulted in decreased 'ON' current, as shown in [29]. However, it is important to note that formation of a thin hole channel near the substrate-insulator interface in LDD region in the short channel SONOS flash memory cell is responsible for the obtained higher 'ON' current after erase operation. Further, presence of hole channel in LDD region occurs due to the drift of charge carriers from the channel into the LDD region under

the influence of a strong negative gate electric field for short channel NAND flash cells. Authors in [62] has also shown the similar phenomenon for thin film transistor with negative gate electric field.

To examine the drift of holes into LDD region under the influence of negative gate electric field, we plot the hole mobility in the LDD region for different values of d in the Figure 3.3. As such, the hole mobility in the LDD region reduces as the d increases from 5nm to 11nm at a distance of 2nm from $Si - SiO_2$ interface in channel length direction because of the high hole concentration near the insulator interface. Here, the reduction in the hole mobility near the substrate-insulator interface in the considered LDD based SONOS device proves the presence of hole channel in LDD region.

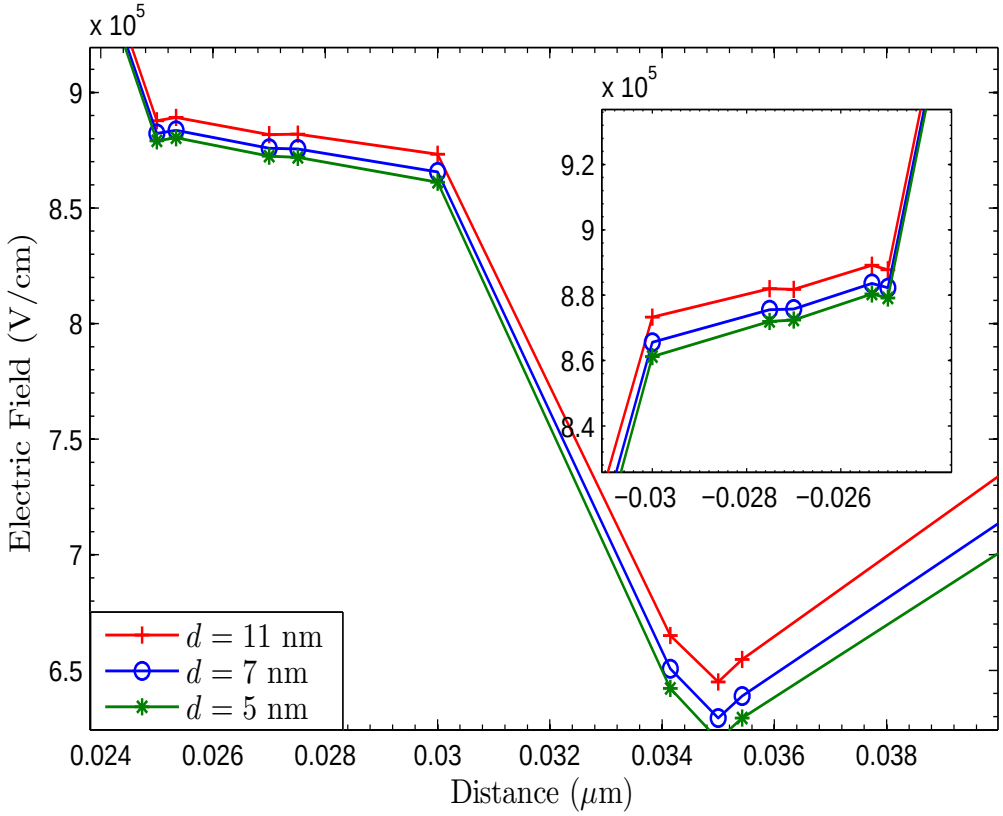

Figure 3.4: Variation of electric field at drain-LDD junction during erase operation (Inset: Lateral electric field at LDD-source junction during erase operation)

Further, Figure 3.4 plots the electric field at the LDD-drain junction during the erase operation. To draw the figure, a 2D cut plane is introduced in the channel near

the substrate-insulator interface in the channel length direction. Herein, it can be noted that the formation of a hole channel in the LDD region from the negative electric field results in the abrupt LDD-drain junction. Consequently, a higher electric field is built-up at the LDD-drain junction of the device. Also, inset of Figure 3.4 depicts that the same effect happens at LDD-source junction. This phenomenon lead to the reduction in the erase threshold voltage of SONOS flash memory cell with higher d. Note that, with higher LDD depth, excess hole concentration in LDD increases electron release rate from CT layer and result in smaller CT layer potential. As a result, the channel conduction increases according to the relation given in equation 2.1.

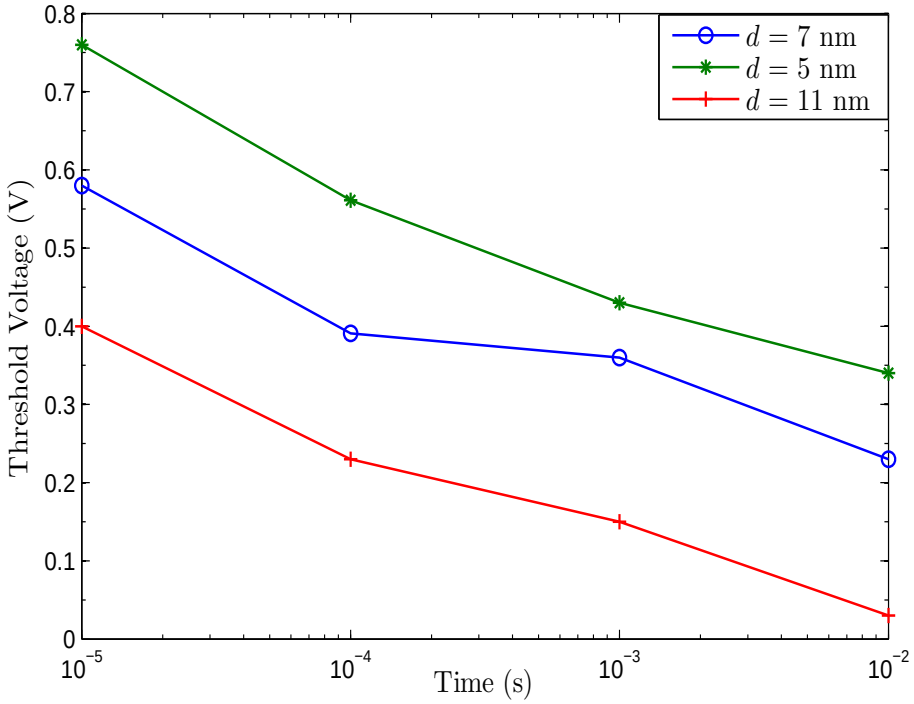

Figure 3.5: Variation of erase speed of SONOS NAND flash memory cell with LDD depth

From here, higher 'ON' current after erase operation (during subsequent read operation) flows due to increases channel conduction with higher d. Importantly, the presence of hole channel in LDD reduces the erase threshold voltage of the device but it provides advantage to the device performance by offering a better erase speed to the SONOS flash memory cell. Figure 3.5 shows the erase speed comparison of SONOS

flash memory cell for various values of d. One can observe that SONOS cell with higher d has better erase speed. Moreover, faster erase speed is attributed to higher hole accumulation with higher LDD depth during erase operation.

In conclusion, for short channel SONOS NAND flash memory cell, high vertical electric field causes holes drifts from channel into LDD region. This increases electric field at LDD-drain/source junction and in turn results in higher 'ON' current after erase operation.

3.1.2 Effect of Varying LDD Depth (d) on the Residual Electron Trapped Charge in the CT Layer of SONOS Memory Cell

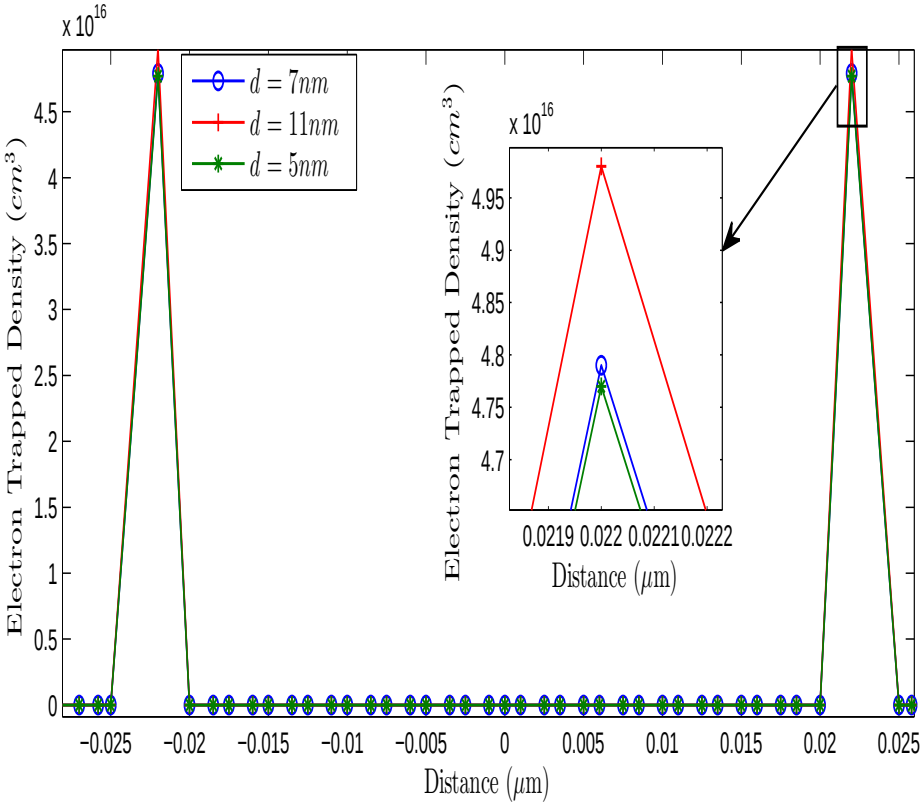

Figure 3.6: Residual electron trapped charge in CT layer after erase operation

Figure 3.6 plots the residual charge in CT layer with distance for the various values of d. To draw the figure, a 2D cut plane is introduced in the channel near the substrate-

insulator interface in the channel length direction. Herein, the residual charge referred to the electrons left in charge trap layer after erase operation. The residual charge in the device mainly occurs because of the weaker electric field above the LDD region. The residual charge in CT layer affects adversely to the 'ON' current and not desirable. One can see from the figure that the residual trapped charge increases with a LDD depth change from 5nm to 11nm. The reason for that could be a decrease in the carrier concentration in the LDD region with increase in d, as already explained. During the erase operation, the reduced carrier concentration of LDD region will offer increased barrier to the de-trapping of electrons in the CT layer, hence residual charge will increase. From here, we present an important cell characteristics that presence of hole channel in LDD affect only lateral electric field profile due to abrupt LDD/drain junction but it does not alter the trend of the vertical electric field with LDD depth.

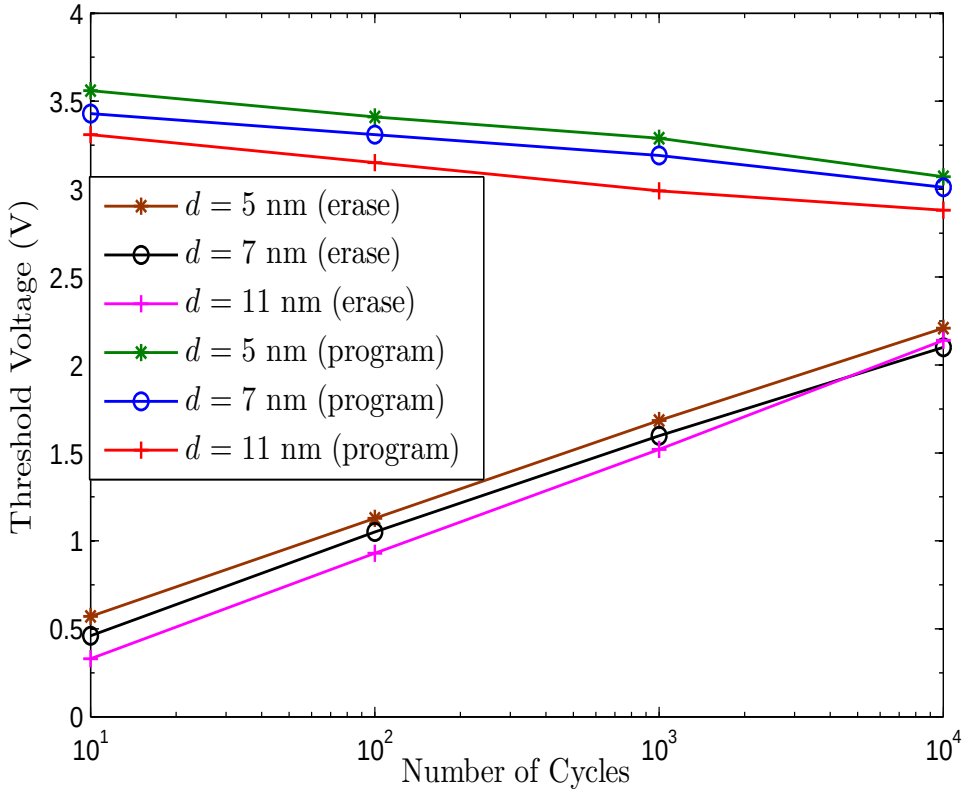

Figure 3.7: Effect of LDD depth on endurance of SONOS flash memory cell

Figure 3.7 shows the endurance characteristics of the considered SONOS flash mem-

ory cell. Here, we simulate for 10^4 program/erase (P/E) cycles. Endurance can be defined as the degradation in memory window with the increase of program/erase cycles. It is revealed from Figure 3.7 that threshold voltage variation with number of program/erase cycles is large with higher value of d. This phenomenon can be explained from the presence of higher residual charge as the value of d increases. Additionally, from Figure 3.7, one can observe another important aspect of this investigation that the variation of program threshold voltage is relatively small as compared to the case of erase operation. From here, one can infer that threshold voltage variation during erase operation is more responsible for window closure as compare to the program threshold voltage variation for all values of d. This peculiar characteristics of flash memory cell occurs because erase performance is affected by residual charge in nitride layer whereas it does not affect program performance much because of high channel potential. We will discuss the details of program operation in the next subsection. In conclusion, this large threshold voltage variation may poses some limitation to the life time of the short channel SONOS flash memory cell.

3.1.3 Effect of Varying LDD Depth (d) on the Program Performance of the SONOS NAND Flash Memory Cell

Figure 3.8 shows the transfer characteristics ($I_D - V_{GS}$) of considered flash memory cell after the program operation. The transfer characteristics is obtained during read operation with V_G = 7V and V_D= 1V to monitor change in program state. Here, we obtain plots for various values of d. We consider equal program time for each set of data. Clearly, one can observe that the program threshold voltage decreases with an increase in the d from 5nm to 11nm. This phenomenon during program operation is observed due to the weaker vertical electric field at tunnel dielectric in gate stack with increase in d. Hence, LDD depth variation to higher value degrades the performance of the SONOS flash memory cell.

In Figure 3.9, we plot the vertical electric field at tunnel dielectric of gate stack with distance. To draw the figure, a 2D cut plane is introduced in the channel near the substrate-insulator interface in the channel length direction. Note that, the presence

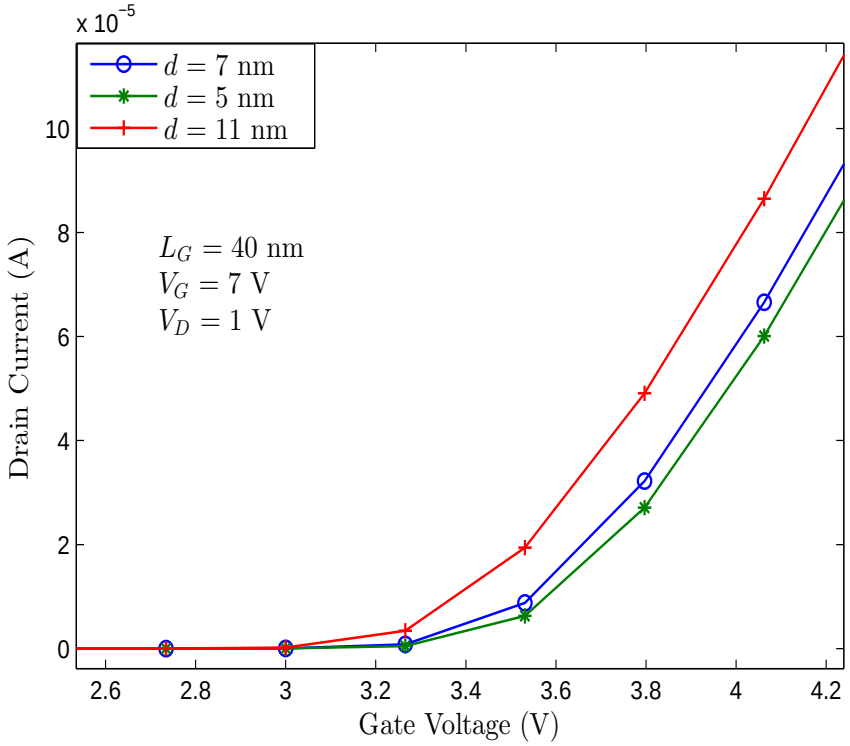

Figure 3.8: Transfer characteristics of memory cell after program operation

of residual charge in CT layer reduces the vertical electric field for subsequent program operation. Hence, as the residual charge increases with d, vertical electric field at tunnel dielectric becomes weaker. This weaker vertical electric field directly influences the surface potential in the channel as per the following relation for N-channel MOSFET

$$V_{GS} = \varphi_S + E_{OX} \cdot t_{OX} \tag{3.1}$$

where φ_S, E_{OX}, t_{OX} represent surface potential, vertical electric field and gate oxide thickness respectively. From the above equation, one can infer that for the constant gate voltage the lower vertical electric field results in a higher surface potential. As a consequence, reduced trapping of charge carriers occurs in CT layer with subsequent program operation (after erase operation due to presence of residual charge). This smaller charge trapping shows smaller CT layer potential hence results in increased channel conduction. This higher channel conduction causes drive current less sensitive

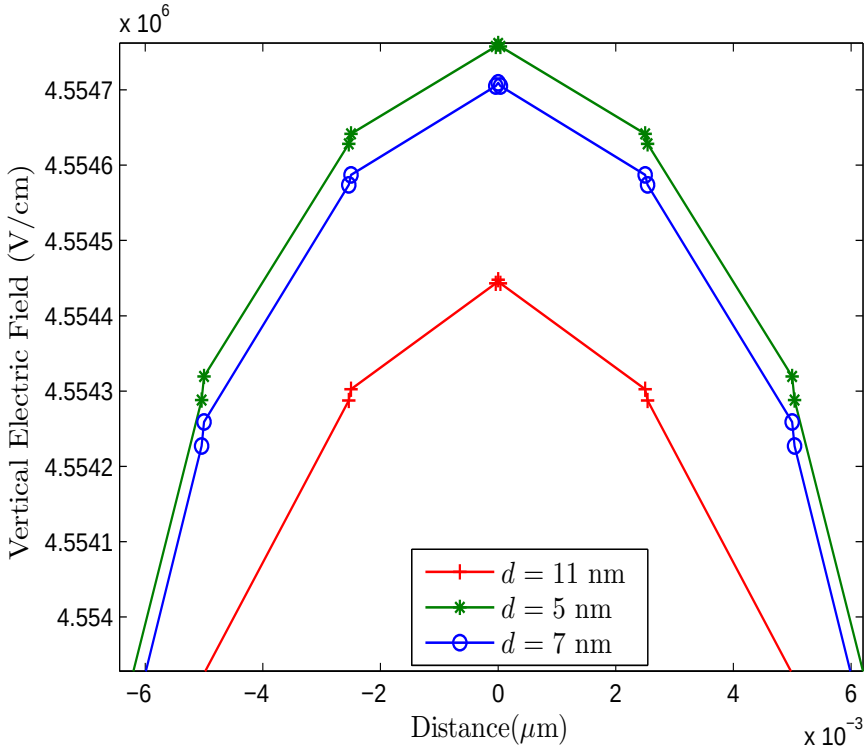

Figure 3.9: Variation of electric field at bottom oxide with LDD depth

to the oxide defects during V_{th} measurement and results in reduced program threshold voltage. Therefore, presence of residual charge in charge trap layer degrades the subsequent program operation.

Apart from this, from Figure 3.10, one can note that the 'ON' current during erase operation is found to be impeded more as compared to program operation by the presence of residual charge in nitride layer, according to [31]. Interestingly, the higher junction electric field increases the 'ON' current after erase operation but the current is still get impeded from residual charge.

3.2 Summary

This chapter reports a new physical phenomenon which may affect the reliability and P/E performance of short channel NAND flash memory cell. Here, it is shown that reduced doping concentration with higher value of d results in increased residual trap charge concentration after erase operation. According to the results presented for short

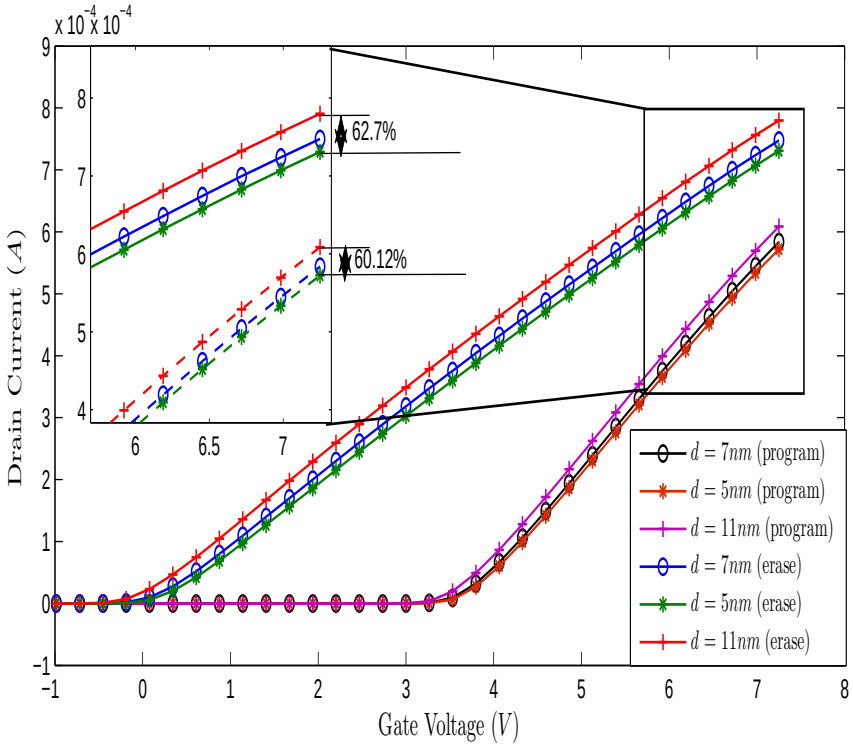

Figure 3.10: Variation of 'ON' current after program and erase operation with LDD depth

channel SONOS NAND flash memory cell, the effect of residual charge on the 'ON' current after erase operation become small, despite of low doping concentration of LDD region. This phenomenon occurs due to the drift of holes from channel into LDD region under the influence of strong negative gate electric field. Due to holes drift into LDD, high electric field is created at LDD-drain junction which accelerates the channel electrons. Therefore, it is concluded that presence of hole channel into LDD region affects only lateral electric field of short channel SONOS memory cell but still the vertical electric field is governed by LDD doping concentration. Additionally, degradation of program operation is found due to the presence of residual trapped charge.

Chapter 4

Analysis of NAND Flash Memory String with Retrograde Channel Doping

So far, we have comprehensively studied the NAND flash memory cell with variation of S/D lateral straggle and LDD depth in previous chapters. In this chapter, we investigate the effect of retrograde channel doping on the SCEs and the memory characteristics of NAND flash memory string having 16 memory cells. Further, flash memory string with uniform channel doping is opted for the comparison purpose.

As discussed previously, the NAND flash memory must have high data storing capacity with low cost per bit to satisfy its demand in handhold applications. As we know that bit cost for the NAND flash memory depends on the area of the die used. So, for the realisation of such cost effective and high density NAND flash memory, scaling of memory cell size is important so that high capacity NAND flash memory can be obtained on the same die area. It has been predicted that half pitch of the 2D NAND flash memory cell would be scaled below 10nm by 2025 [37]. These extremely scaled memory cells will impose several limitations on the performance of the flash memory. Most recently, to scale the NAND flash memory efficiently with the reduced cost per bit, several three dimensional (3D) NAND flash memory structure have been proposed in the literature. Here, to obtain high density NAND flash memory on the same die area, memory array is realised in the vertical direction [22]. It is worth mentioning that 3D NAND flash memories are faster in operation than 2D flash memory due to parallel access of large number of chips packed densely together. In addition, 3D NAND flash

has improved wear life over the 2D one. It follows that 3D NAND flash has lower bit error rate (BER) per KB of data and require less robustness for error-correcting codes. These 3D devices with NAND flash memory serves as the basic building block for the system known as Solid State Drives (SSD).

In addition, the market for CT based flash memories has accorded a great surge due to the evolution of 3D NAND flash memories such as bit cost scalable memory [64]-[65], vertical-stacked-array-transistor memory [66], and terabit cell array transistor memory [63]. It is important to highlight that 3D NAND flash memory mostly employs CT based memory cell due to their MOSFET compatible fabrication process, low voltage operation, immunity to extrinsic charge loss, small cell to cell interference, compact stacking, and high scalability as compared to FG counterpart [4], [39]. Hence, in this chapter, we will consider CT based flash memory cell in the NAND string. Apart from this, the 3D NAND flash memory can be divided into two parts according to the current flow direction through channel: Vertical Channel (VC) NAND flash memory and Vertical Gate (VG) NAND flash memory, as shown in Figure 1.15. The VG structure consist channels which are stacked horizontally while gates of the memory cells are vertically shared. The VG architecture seems more appealing in terms of pitch scaling, selection of channel material and degradation of read current with large number of stacking layers, as compared to the VC configuration. However, the memory array design and decoding methods for VG NAND flash memory is difficult due to horizontally parallel bit lines. Apart from this, due to relatively easy fabrication process and decoding method, VC NAND flash memory (which is also known as VNAND) is readily being used for commercial purposes. However, reduced channel current and increased tolerable electric field with scaling is a major issue for VC configuration [67]. Owing to the advantages of CT based VG architectures, in this chapter, we focus on the CT based VG configuration of NAND flash memory cell to analyse the performance of NAND flash memory string.

Further, the flash memory cell requires a super steep doping profile at the channel-source/drain (S/D) junction to achieve high gate electrostatic integrity [32]. Albeit, such fabrication process needs highly controlled and sophisticated environment, which is difficult to achieve. Further, in stacked 3D flash memory structure, it is quite difficult

to dope the space body region between adjacent flash memory cells. This stimulates the need of a junction-free NAND flash memory to achieve scaling towards the extreme short dimensions. Consequently, junction-free NAND flash memory has been investigated in [11],[33]-[34] for achievable memory performance. It is worth noting that compared with the junction-based memory, a junction-free flash memory can be easily fabricated as it is devoid of S/D junction. Also, the junction-free devices has shown smaller SCEs with respect to junction-based memory [34]. Therefore, in this chapter, we consider a CT based junction-free NAND flash memory string with VG configuration.

Moreover, the requirement of thick gate oxide stack to achieve high data retention time is obvious with NAND flash memory cell. This demand makes flash memories more prone to SCEs when scaled [32]. Although the scaling of vertical gate-stack in NAND flash memory has already been approached to its limit, the effective oxide thickness (EOT) remains still very high (15-20 nm). This large EOT deteriorates the short-channel performance of a NAND flash memory with scaling. Note that, with further scaling of gate stack by reducing gate oxide stack thickness, the SCEs and program performance may improve but it severely deteriorates the data retention. This presents the most crucial trade off for the scaling of NAND flash memory. In view of the above, an effective methodology is needed to be discovered that can simultaneously improves the SCEs as well as the data retention for future NAND flash memory. Further, it is known that the high channel doping concentration can be a possible solution [36] for the improvement of SCEs in NAND flash memory cells. However, in junction-free NAND flash memory string, the higher channel doping opposes the fringing field penetration from the adjacent CG into the selected flash memory cell. In junction-free NAND, the S/D regions are formed from the fringing field of the adjacent memory cell. Smaller fringing interference into the channel with NAND string results in the reduced drive current with high channel doping. Consequently, the program/erase memory operation become ambiguous due to smaller obtained memory window. Owing to this, the low doping of channel seems desirable for the proper operation of 3D NAND flash memory [34]. However, the low channel doping makes the flash memory cell in the string more vulnerable to the fringing interference from adjacent memory cells [11]. Also, the low

doping of the channel region may degrade the programming speed, data retention time for the NAND flash memory [36]. Therefore, the combination of high and low doping in the channel is necessary to provide better SCEs with improved memory characteristics. Importantly, Hsaio et. al. have illustrated in [36] that proper channel engineering for 3D NAND flash memory string can optimize its operation beyond 15nm technology node. Such channel engineering may be harnessed in order to improve the SCEs for flash memory without altering the TO thickness. So, relying on this, both the SCE and the data retention time can be simultaneously improved for the flash memory.

Apart from this, some works which have directed their attention towards improving the SCEs of junction-less TFT and multi-gate FET by using both high and low doping concentration in the channel [68], [69]. In these references, it is shown that the presence of both high and low doping in the channel reduces the active channel width and improves the gate electrostatic effect over channel. Moreover, very few works in literature have investigated channel engineering-based approaches to improve the SCEs for Barrier Engineered (BE) CT (SONOS) based junction-free 3D NAND flash memory. Hereby, the effectiveness of channel engineering approach for such memory devices and the lack of literature in this direction led us to enrol for the present work. In this chapter, we have considered retrograde doping, i.e. smaller doping concentration near the surface and higher doping concentration at the centre, in the channel for the improvement of SCEs and memory characteristics simultaneously. Primely, the retrograde channel doping was proposed to improve the scalability of junction-based devices [70]-[71]. However, the junction-free NAND flash memory has different operating principle hence, its behavior with such channel engineering can not be predicted straightforwardly. Hereby, it is important to investigate the impact of such retrograde channel engineering on the SCE of 3D junction-free NAND flash memory [72]. Most importantly, in this chapter, we have shown to improve the SCE-data retention trade off for scaling NAND memory cell into the sub-10-nm generation. Also, the impact of retrograde channel doping on the other SCEs, program operation and data retention capability should be studied extensively in NAND flash memory.

Motivated by the above, in this chapter, we examine the impact of retrograde doping

in channel on SCEs such as $\triangle V_{th}$, SS and DIBL for junction-free NAND flash memory. Additionally, $\triangle V_{th}$ of NAND flash memory is also studied with the different pass gate voltage to investigate the V_{PASS} interference with retrograde channel doping. Moreover, we also investigate the impact of retrograde doping in the channel on the program speed and data retention of short channel NAND flash memory. Further, we compare the SCEs of NAND flash memory having retrograde channel doping with the one having uniform channel doping.

4.1 Device Description and Simulation Details

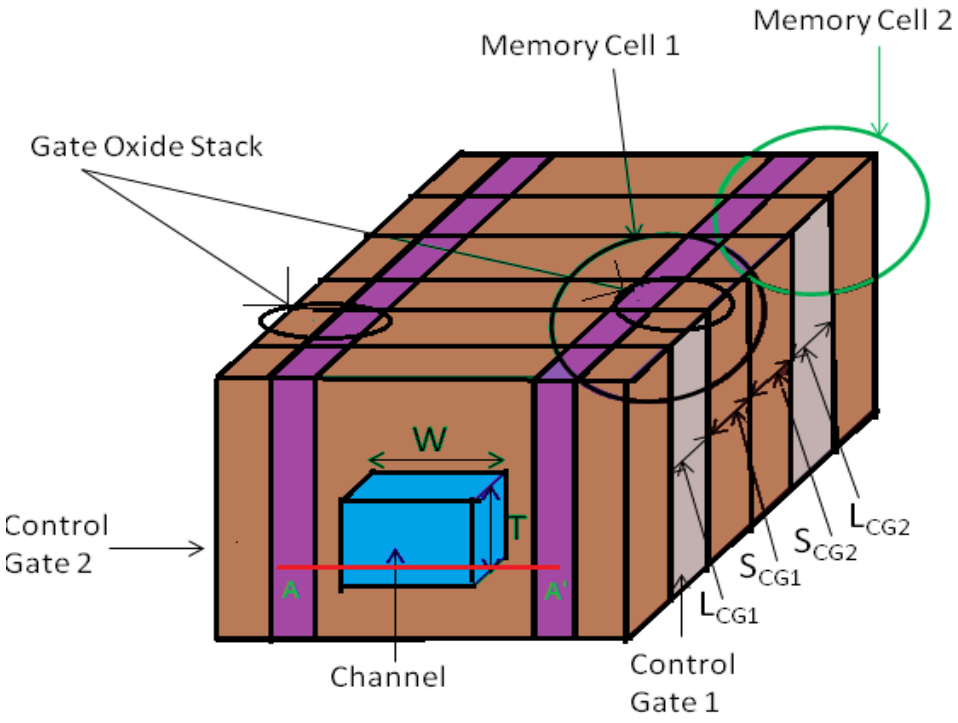

Figure 4.1: Schematic of the 3D vertical gate NAND flash memory. Figure shows two double gate charge trap flash memory cell along the channel length direction with uniform channel doping

Figure 4.1 presents the two adjacent memory cells of the vertical gate NAND flash memory string. Here, each memory cell in the string has double gate (control gate 1 and

2) device architecture. The gate oxide stack of all NAND flash memory cells consist silicon dioxide (SiO_2) and silicon nitride (Si_3N_4) as IPD and CT layer respectively. Further, barrier BE tunnel layer with oxide-nitride-oxide (O-N-O) structure is used to achieve required data retention. Note that the barrier engineering approach is used to enhance the data retention and operation speed of NAND flash memory cell [73]. In this chapter, the thickness of each layer is used in accordance with the experimental result provided in [34]. In addition, pitch in the NAND flash memory string can be given as

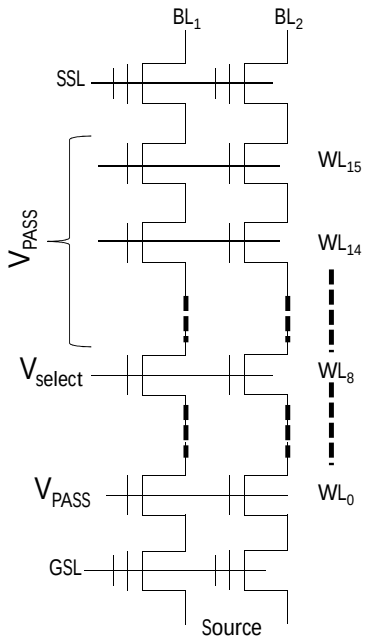

Figure 4.2: Schematic diagram of the NAND string with SSL and GSL. V_{select} voltage is applied to select transistor (WL_8) and V_{PASS} voltage is used with all other cells in the string to read through it

$$Pitch = L_{CG1} + S_{CG1} + S_{CG2},$$

where L_{CG1} denotes the length of control gate for memory cell 1, S_{CG1} and S_{CG2} represent respectively the length of spacer dielectric of memory cell 1 and 2. Here, S_{CG1}

Table 4.1: Bias Conditions Used for the Simulation of 3D NAND Flash Memory with Uniform and Retrograde Channel Doping.

Node	Bias Condition
BL_1	1 V
SSL	3.3 V
GSL	3.3 V
V_{PASS}	6 V
V_{select} (Read)	7 V
V_{select} (P/E)	20/-20 V
Source	0 V

$+ S_{CG2}$ denotes the total spacing (S) between two adjacent cells in the memory string. From here, half pitch can be expressed as

$$Half\ Pitch = Pitch/2 = (L_{CG1} + S_{CG1} + S_{CG2})/2.$$

Therefore, in the analysis, we have taken $S_{CG1} + S_{CG2}$ equal to the S_{CG1} which makes half pitch length same as the gate length. It is note that, half pitch is most commonly used term to represent technology node instead of gate length accepted by ITRS. Further, Figure 4.2 shows the schematic circuit diagram of the considered NAND flash memory string. The string contains 16 memory cells from WL_0 to WL_{15} along with string select line (SSL) and ground select line (GSL). Note that, high voltage (V_{select}) is applied to the gate terminal of the select flash memory cell (WL_8) during program and erase operation so as to generate sufficient high electric field at TO for desired tunneling of charge carriers.

Also, to investigate the characteristics of the select memory cell, all the other cells along with SSL and GSL are supplied with sufficient voltage to read through the string. Voltage applied to memory cells other than the select gate memory cell is denoted as V_{PASS}. Table 4.1 depicts the bias conditions to read through the NAND string.

In addition, Figure 4.3 depicts the 2D horizontal cut plane of the considered junction

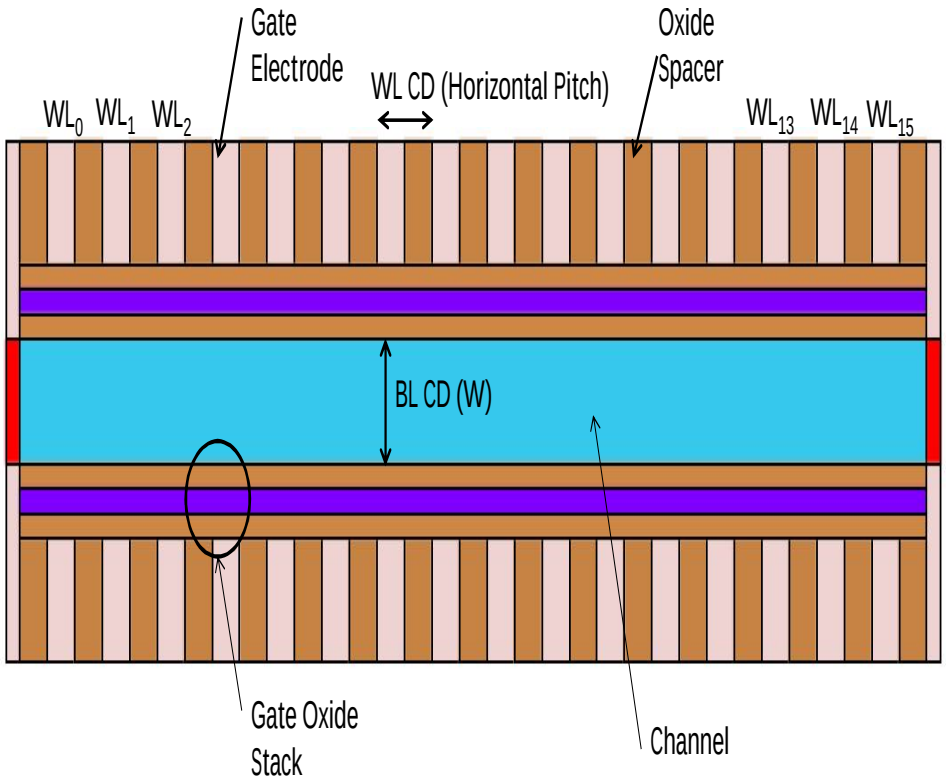

Figure 4.3: 2D cut plane of NAND flash memory string with 16 memory cells. Cut plane is in the AA' direction as shown in Figure 4.1. Here, the junction-free NAND flash memory with uniform channel doping is shown

free NAND flash memory string with 16 memory cells. The considered NAND memory string is junction-free i.e. no junction is available inside the NAND string. However, to obtain the desired 'ON/OFF' characteristics junctions are present at the outer ends. Also, it is clear that the considered structure is the VG configuration of 3D NAND where the channels are horizontal and gates are vertically shared. Apart from this, FN tunneling mechanism is used for NAND flash memory operation. Importantly, for the authentication of our simulation results, in this chapter, we have calibrated the transfer characteristics of the considered NAND flash memory from [34].

In this regard, Figure 4.4 shows the simulated I-V curve of the select memory cell (WL_8) in a 16 cell NAND flash memory string. Here, the obtained curve is well aligned with the results presented in [34]. Importantly, the SS for the 15nm memory cell is found to be smaller than 400 nm/dec along with $\sim$ 600 mV/V DIBL. Here, a channel with starting uniform boron concentration of $5 \times 10^{17}/\text{cm}^3$ is used for simulation. P^+

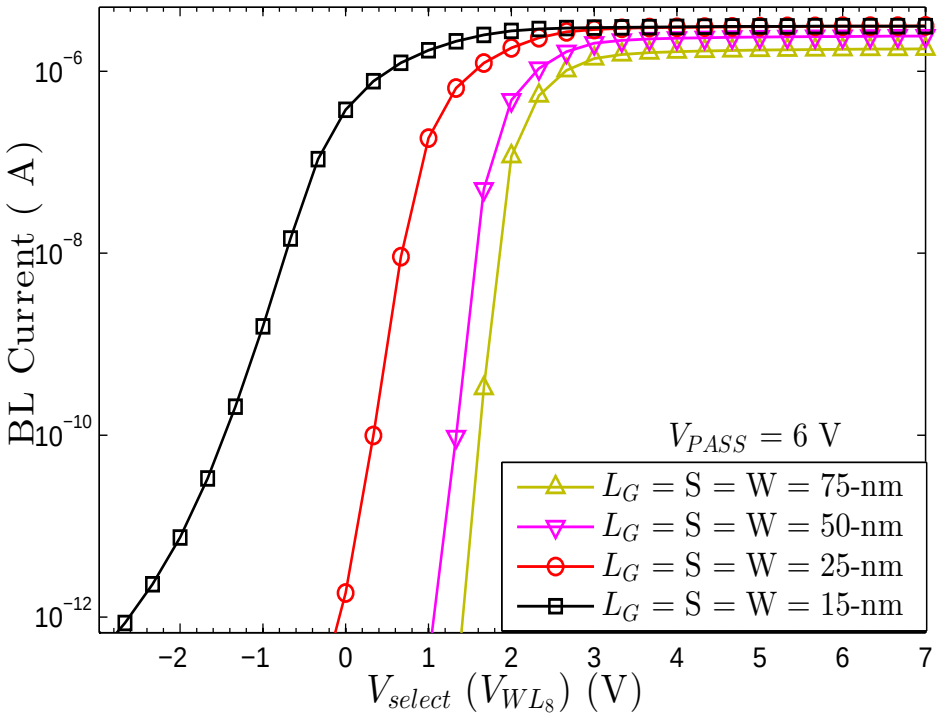

Figure 4.4: Simulated I-V characteristics of the select memory cell (WL_8) in a 16 cell NAND flash memory string [34]

poly-silicon is used as gate material to reduce back tunneling and to achieve the required threshold voltage with doping concentration $1 \times 10^{20}/\text{cm}^3$. Uniform doping is used for the source/drain region.

Further, in this chapter, the simulation of the junction-free NAND flash memory is carried out using Synopsys Sentaurus 3D TCAD tool at 300 K [50]. Figure 4.5(a) and Figure 4.5(b) represent the uniform and proposed retrograde channel doping condition for considered NAND flash memory respectively. As mentioned previously, the research in this direction has gained attention very recently and the literature on exact fabrication process for retrograde channel doping in 3D NAND flash memories is currently unavailable. However, for conventional devices, to physically achieve retrograde channel doping profile ion implantation method is employed [74]. During implantation, the dose and energy of the corresponding implantation process is controlled so as to locate the peak doping concentration in the desired area of the channel [74]. However, ion implantation method does not provide sharp retrograde profile. Hence, the epitaxially

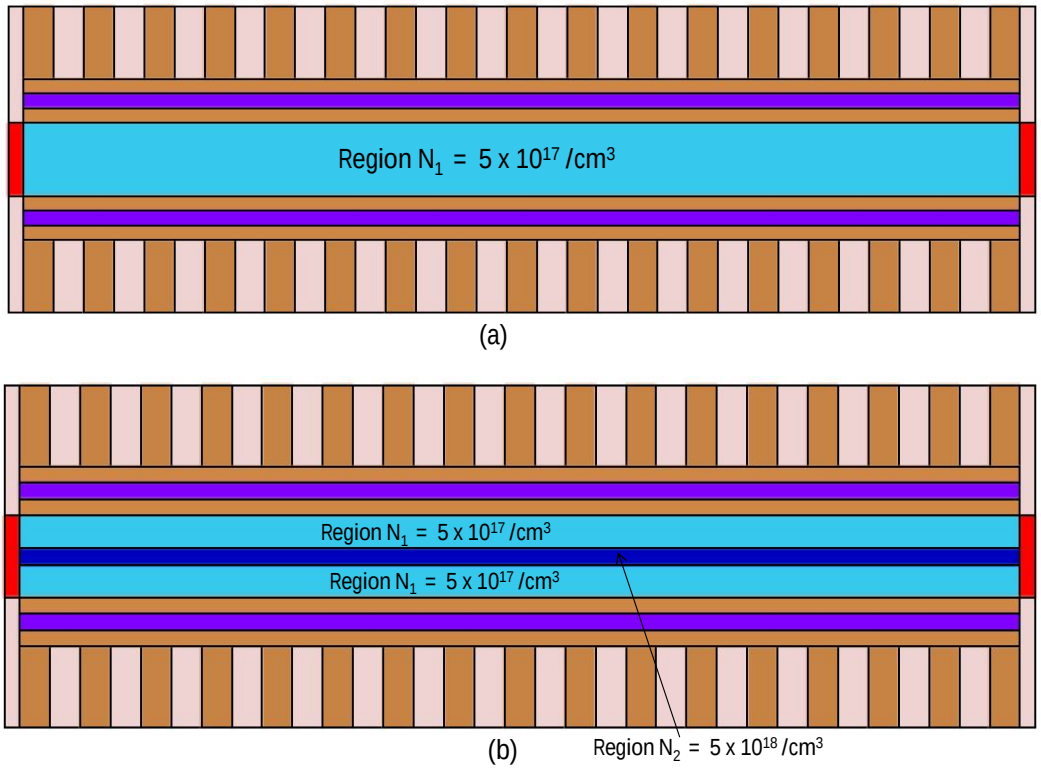

Figure 4.5: (a) Uniform channel doping (b) Proposed retrograde channel doping in junction-free 3D NAND flash memory

grown channel technology is used to achieve the abrupt retrograde channel doping profile [75]. The detailed fabrication process for attaining retrograde channel doping with epitaxially grown channel technology can be found in [75]. In this chapter, to imitate retrograde channel doping profile, the channel consist of two different doping concentrations: low (Region $N_1 = 5 \times 10^{17}/\text{cm}^3$) at the surface of the channel and high (Region $N_2 = 5 \times 10^{18}/\text{cm}^3$) at the center of the channel. Note that, for fair comparison, both the structure have same doping concentration at the surface i.e. channel-TO dielectric interface. Further, area for the high doping concentration is taken as 25% of the total channel region. This particular dimension has been taken to maximize the advantages of both high and low channel doping in junction-free NAND flash memory. In this analysis, variation in threshold voltage shift are monitored using constant current method ($I_D = 100$ nA at $V_{BL} = 1$ V). In addition, all the cells in the string are maintained in the erased state during the SCEs analysis. Moreover, transport of charge carrier is performed with conventional drift diffusion model along with other physical model

such as SRH generation recombination model, Poole Frenkel model with concentration and field dependent mobility. Furthermore, non-local band to band tunneling model is used for the tunneling of charge carriers across the barrier during the program operation. Simulation during program/erase operation involves transient analysis to fulfill trapping and release of charge carriers.

4.2 Results and Discussion

In this section of the chapter, we present the simulation results for the considered junction-free NAND flash memory with retrograde and uniform channel doping. First, we present the effect of retrograde doping on SCEs such as $\triangle V_{th}$, SS and DIBL. Further, we investigate the effect of varying doping concentration and PASS gate voltage with retrograde channel doping on the select device (WL_8) performance in the flash memory array. Furthermore, we monitor the effect of retrograde channel doping on program and data retention characteristics. Note that, in all the simulations, the channel width of considered junction-free NAND flash memory is kept at constant at 40nm with both retrograde and uniform channel doping.

4.2.1 Effect of Retrograde Channel Doping on SCEs of Junction-Free NAND Flash Memory String

Here, we investigate the variation of different SCE parameters i.e. δV_{th}, SS and DIBL in considered NAND flash memory for uniform and retrograde channel doping. In this regard, Figure 4.6(a) depicts the normalised $\triangle V_{th}$ for the two type of channel doping with select gate (WL_8) under observation. In this analysis, the gate length varies from 35nm to 12nm. In figure, the $\triangle V_{th}$ value has been normalised to unity for uniform and retrograde channel doping at 35-nm gate length for comparison. One can note that $\triangle V_{th}$ from 35nm to 12nm is found to be small for retrograde channel doping as compared to the uniform doping. This phenomenon is attributed to the reduced effective channel width of the flash memory cell with retrograde channel doping. Importantly, the presence of low to high doping area in the channel with retrograde

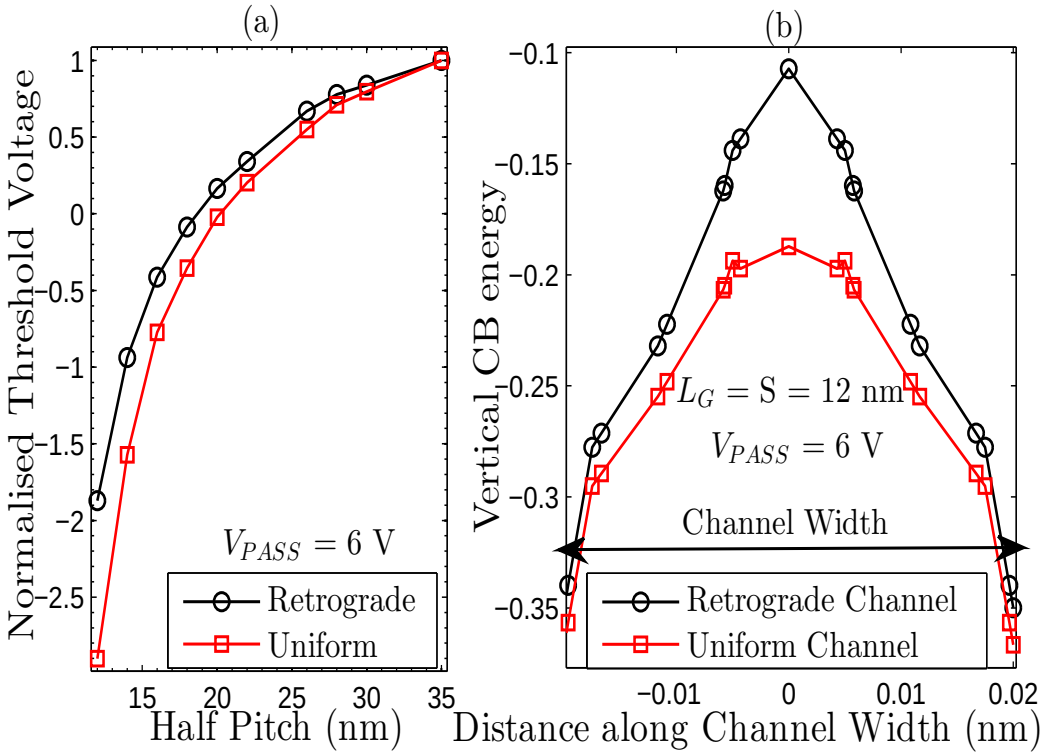

Figure 4.6: (a) Variation of threshold voltage roll-off with technology scaling (b) Vertically cut channel conduction band energy during read operation.

channel doping creates a junction between region N_1 and region N_2. This, in turn, generate a depletion layer between the two region. As a consequence, the effective channel width reduces which improves the gate control over the channel. To confirm this hypothesis, Figure 4.6(b) shows the normalised channel conduction band energy diagram cut vertically through the center of the channel of the select memory cell (WL_8). One can see that with retrograde channel doping, the energy barrier from surface to centre is high as compared to the counterpart. Hence, higher energy barrier is produced at the junction of region N_1 and N_2. Consequently, thinner effective width is obtained with retrograde channel doping. This improves the SCEs in considered NAND flash memory with retrograde channel doping. Again, from Figure 4.6(a), one can deduce that threshold voltage difference between retrograde and uniform channel string is larger for the lower technology. It is obvious because of the well known leakage current increase with technology scaling. Here, the significant improvement of $\triangle V_{th}$ at lower technology with retrograde channel doping proves its practicality in reducing

leakage current for scaled devices.

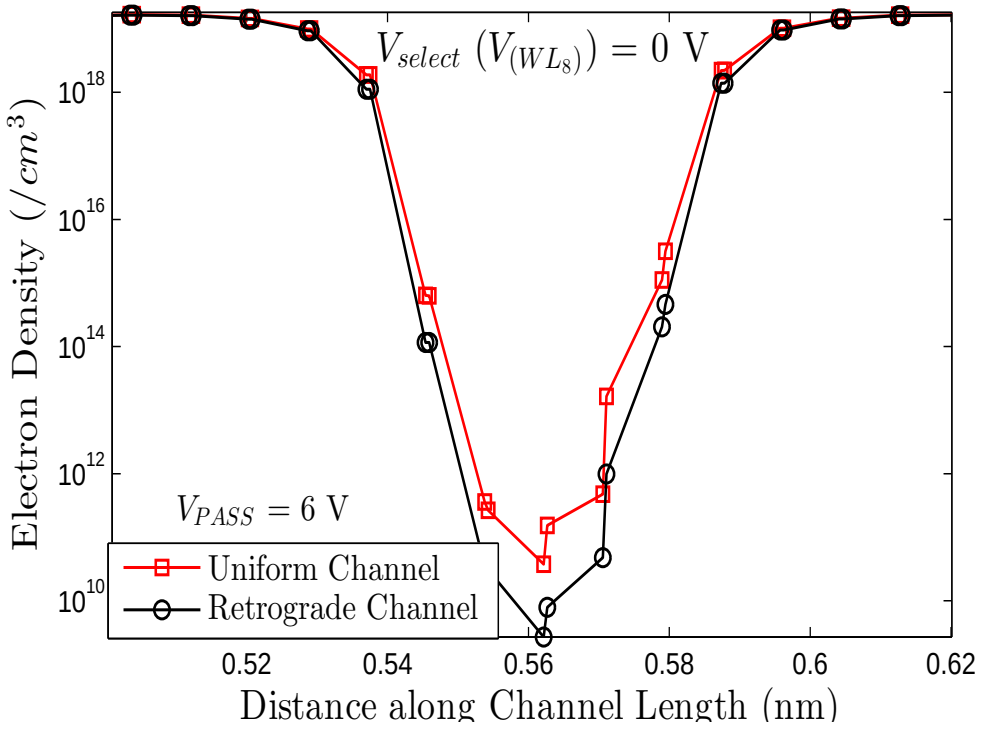

Figure 4.7: Variation of electron density under the select memory cell (WL_8) with retrograde and uniform channel doping

In addition, the reduced intensity of fringing field interference is observed into the select memory cell from adjacent memory cells with retrograde channel doping, as shown in Figure 4.7. In this figure, for the equal V_{PASS} with two channel doping condition, one can see that the electron penetration from adjacent memory cell into the select gate is more with uniform channel doping. To draw this figure, a 2D cut plane is introduced in the channel near the substrate-insulator interface in the channel length direction. This phenomenon appears due to the presence of high average doping density in the channel with retrograde doping which opposes the field penetration into the select gate. As a result, the spreading of the charge carrier from virtual S/D into the effective gate region reduces. This, further, enhances the gate controllability over the channel through high gate electric field.

Furthermore, for the NAND flash memory, $\triangle V_{th}$ results in SS degradation and DIBL. Therefore, analysing the variations of SS and DIBL with retrograde and uniform channel

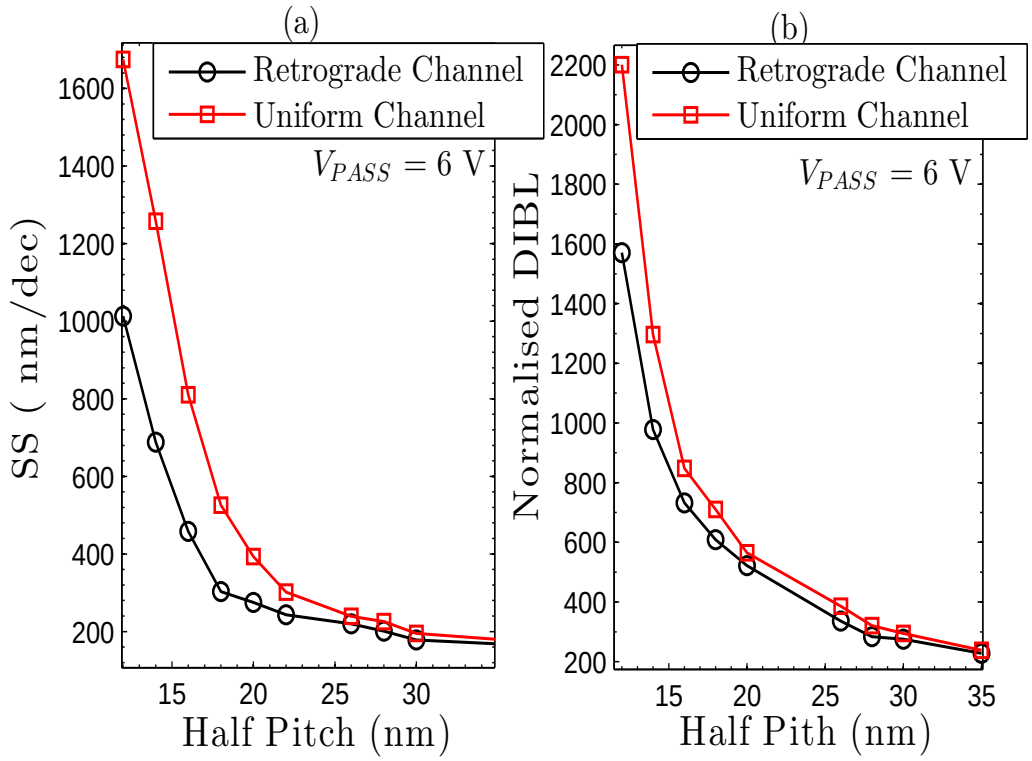

Figure 4.8: (a) Variation of SS with technology scaling (b) Variation of DIBL with technology scaling

doping may be beneficial for future scaling of junction-free NAND flash memory. With this regard, Figure 4.8(a) and 4.8(b) shows the variation of SS and DIBL with retrograde and uniform channel doping in NAND flash memory with select memory cell. Here, the extraction of SS is performed in the current range from 0.1nA to 100nA. The DIBL and SS value has been normalised to unity for uniform and retrograde channel doping at 35nm gate length. From figures, it is clear that smaller SS and DIBL is obtained with retrograde channel doping for gate length varying from 35nm to 12nm. Once again, as already explained, this effect is associated with the reduced effective channel width under the control of the select gate with retrograde channel doping. Numerical results for the correlation of $\triangle V_{th}$, SS and DIBL with retrograde and uniform channel doping are shown in Table 4.2.

Moreover, high V_{PASS} voltage with unselected cells in the NAND string is necessary to read through it. However, while scaling, the high V_{PASS} voltage increases the interference from adjacent (PASS gates) memory cells into the select memory cell. Con-

Table 4.2: Comparison of Different Short Channel Metrics with Retrograde and Uniform Channel Doping.

$L_G\ (nm)$	$\triangle V_{th}\ (V)$		SS (mV/dec)		DIBL (mV/V)	
	Retrograde Doping	Uniform Doping	Retrograde Doping	Uniform Doping	Retrograde Doping	Uniform Doping
12	-2.81342	-3.88389	1013	1675	1571	2201
14	-1.41006	-2.1044	688	1258	978	1296
18	-.128995	-.475102	303	526	609	710
20	.249161	-.0290273	275	393	522	565
26	1.00744	.736	220	239	336	386
28	1.16882	.951276	201	226	284	321
35	1.50347	1.33906	168	180	228	238

sequently, the current through the select cell will change. Here, it is to mention that the measurement of threshold voltage is, generally, obtained by constant current method. Since, the V_{PASS} interference is able to change the current through the select memory cell hence, the threshold voltage will change. This will adversely affect the performance of the extremely scaled NAND flash memory.

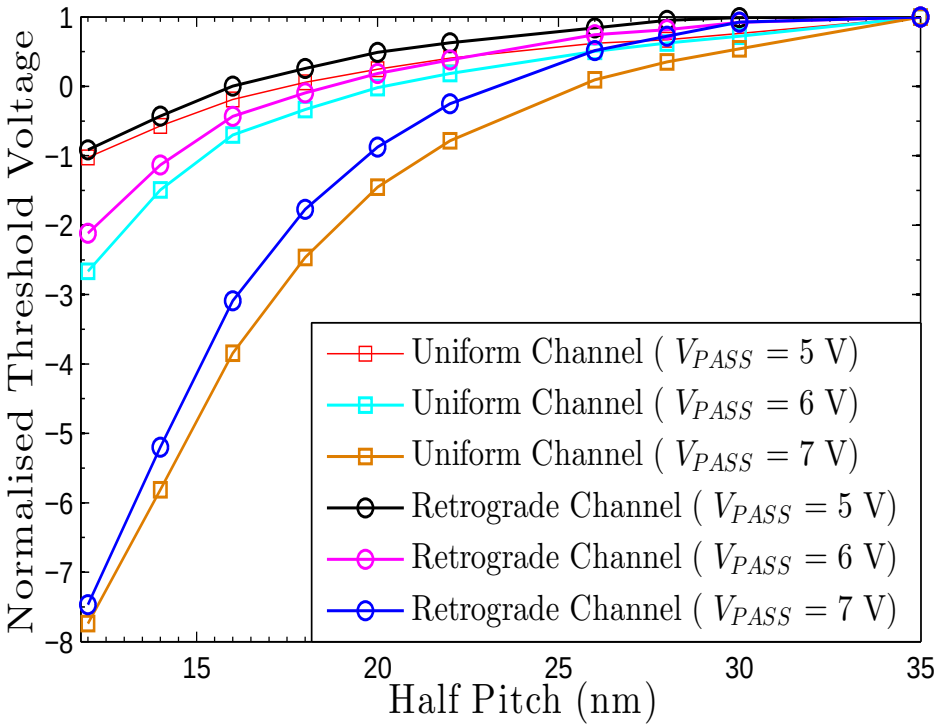

Figure 4.9: Variation in $\triangle V_{th}$ with retrograde and uniform channel doping and different V_{PASS} voltages.

With this regard, Figure 4.9 shows the normalised $\triangle V_{th}$ for the NAND flash memory with retrograde and uniform channel doping for different V_{PASS} voltage. Here, $\triangle V_{th}$

with different V_{PASS} voltage is normalised to unity at 35nm for better understanding of results. Here, It is clear that NAND flash memory with retrograde channel doping shows improved $\triangle V_{th}$ than the counterpart for the all values of V_{PASS} voltages. From figure, one can deduce that NAND flash memory with retrograde channel doping may help to improve V_{PASS} interference. Importantly, as already explained in Figure 4.6, the effect of suppressed V_{PASS} interference is attributed to the reduced effective channel width with retrograde channel doping.

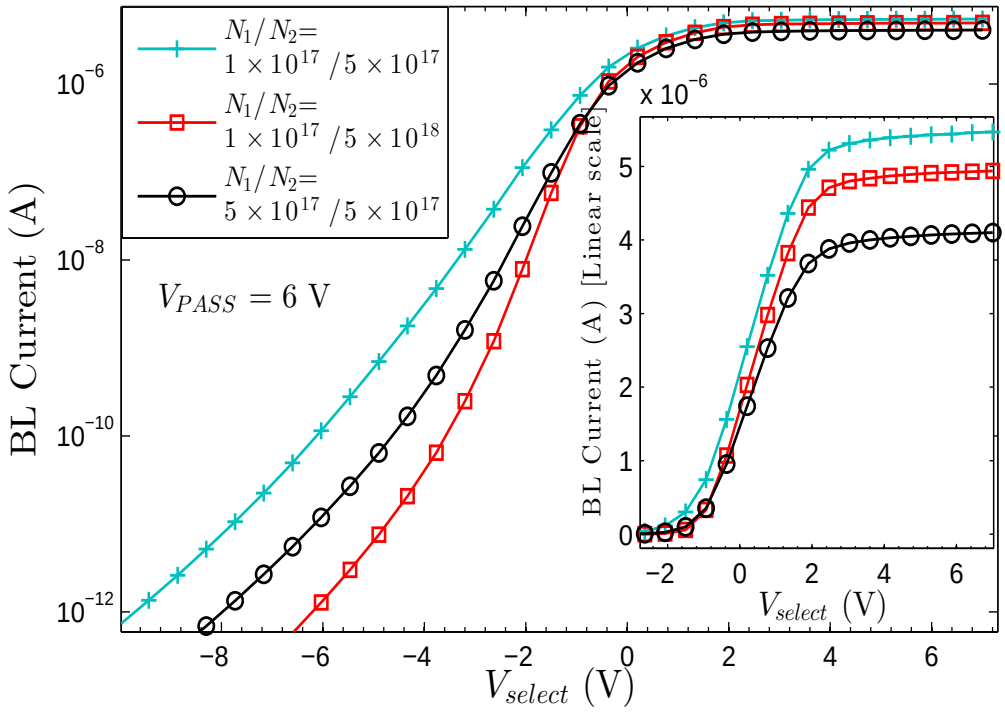

Figure 4.10: Comparison of SS of the select cell in NAND flash memory with different doping concentrations for region N_1 and region N_2 in retrograde channel doping. Inset: Comparison on 'ON' current through select cell in NAND flash memory with different doping concentration for region N_1 and region N_2 in retrograde channel doping

Apart from the above, Figure 4.10 compares the variation of transfer characteristics with change in SS of the considered NAND flash memory cell with various doping concentrations of region N_1 and region N_2 for retrograde channel doping. Also, results for retrograde channel are compared with the uniform channel doping (5×10^{17} doping in region N_1 and N_2). Here, inset of Figure 4.10 depicts the 'ON' current comparisons of junction-free NAND flash memory with retrograde channel in contrast to uniform

channel. From Figure 4.10, one can deduce that as the channel surface doping reduces to 1×10^{17} in comparison with uniform channel doping, rise in 'ON' current is observed but with poor SS. Evidently, to achieve high 'ON' current, the surface doping of NAND flash memory channel must be reduced but it may result in poor SS due to weaker gate control. Further, the improvement in SS along with higher 'ON' current as compared to the uniform channel doping is observed when the doping concentration at the centre of the channel is increased to 5×10^{18} with reduced doping concentration of region N_1. Here, in comparison with the uniform channel doping, improved SS results from the reduction of channel width due to depletion region at the interface of region N_1 and region N_2 and higher 'ON' current is acquired from the lower surface channel doping. Also, it is important to mention that retrograde channel doping results in increased number of dopant atom which may reduce the drive current.

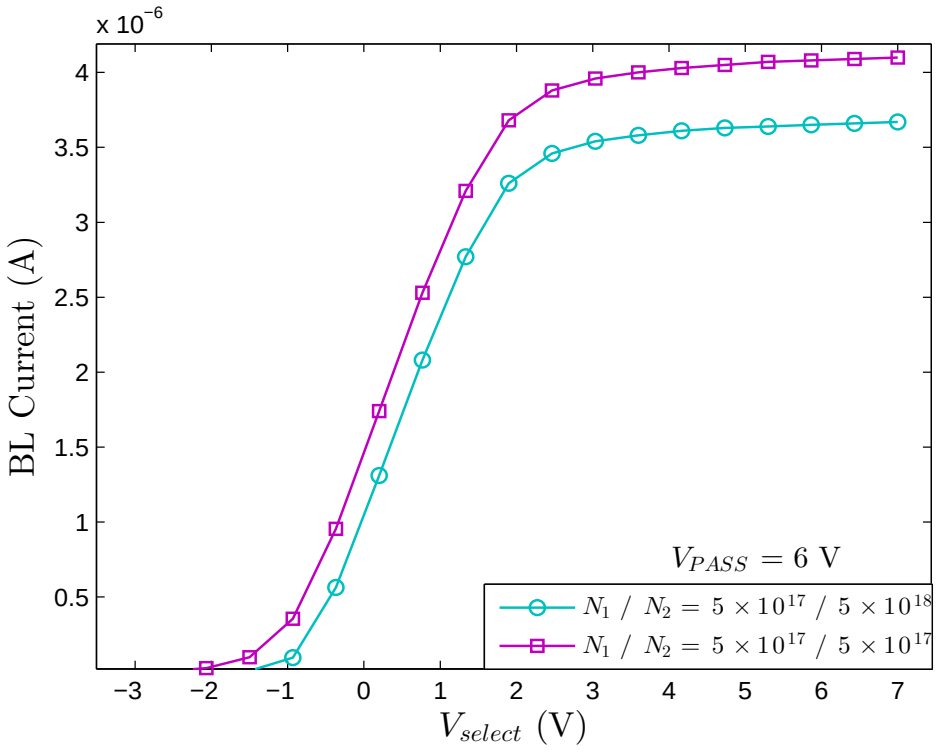

Figure 4.11: Comparison of 'ON' current through select memory cell in junction-free NAND flash memory with retrograde and uniform channel doping concentration

However, as shown in Figure 4.11, the reduction in 'ON' current from uniform channel doping to retrograde channel doping is by 10.3% whereas improvement in $\triangle V_{th}$ and

SS is observed as 38.04% and 46.81% respectively. Hence, it can be deduced that the presence of high doping concentration layer, covering 25% of the channel area in the centre, improves the short channel performance of the considered NAND flash memory with appropriate drive current available in the channel for operation.

4.2.2 Effect of Retrograde Channel Doping on Program Performance of Junction-Free NAND Flash Memory

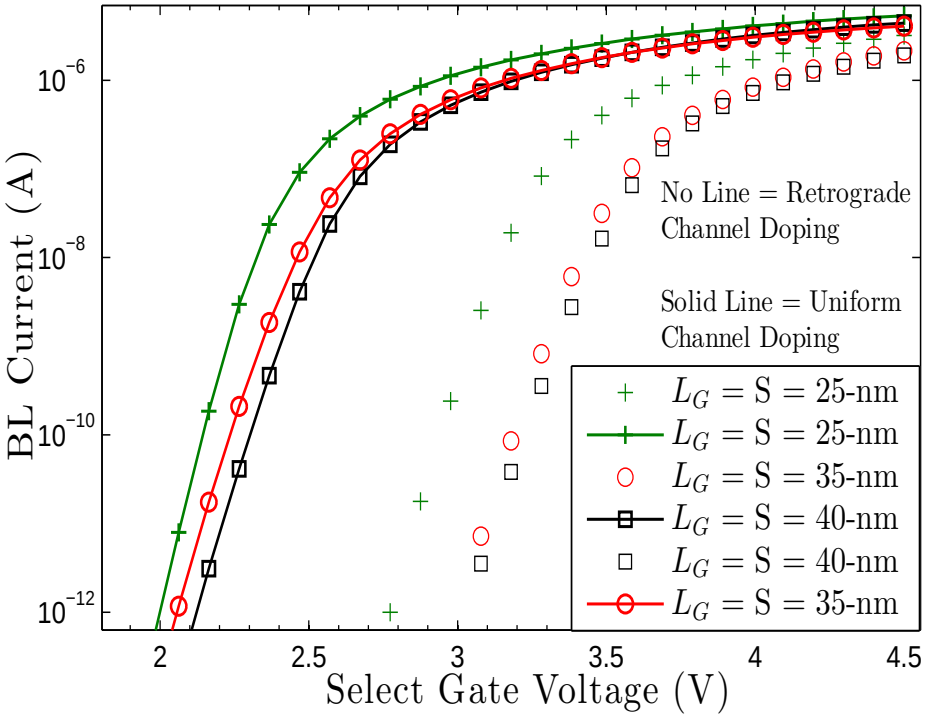

Figure 4.12: Variation of program characteristics with uniform and retrograde channel doping with different technology node

Figure 4.12 shows the transfer characteristics ($I_{BL} - V_{Wl_2}$) of the select memory cell (WL_2) in the considered junction-free NAND flash memory. Here, to save the simulation time, only 3 cells are considered in the NAND flash string with SSL and GSL for the analysis of the program performance. The operating condition applied to the string is same as shown in Table I. One can see the clear increase of program

threshold voltage with retrograde channel doping. Therefore, junction-free NAND flash memory with retrograde channel doping can be programmed by smaller programming time to attain the equal threshold voltage shift. Furthermore, it is important to mention that the obtained improvement in program threshold voltage with retrograde channel doping is attributed to higher number of dopants in the channel as compared to the uniform channel doping. Here, high doping concentration in the channel reduces the fringing field interferences between the adjacent devices in the NAND memory string. This helps to confine the electric field at TO within the select memory cell in the string and hence, helps to improve the program performance.

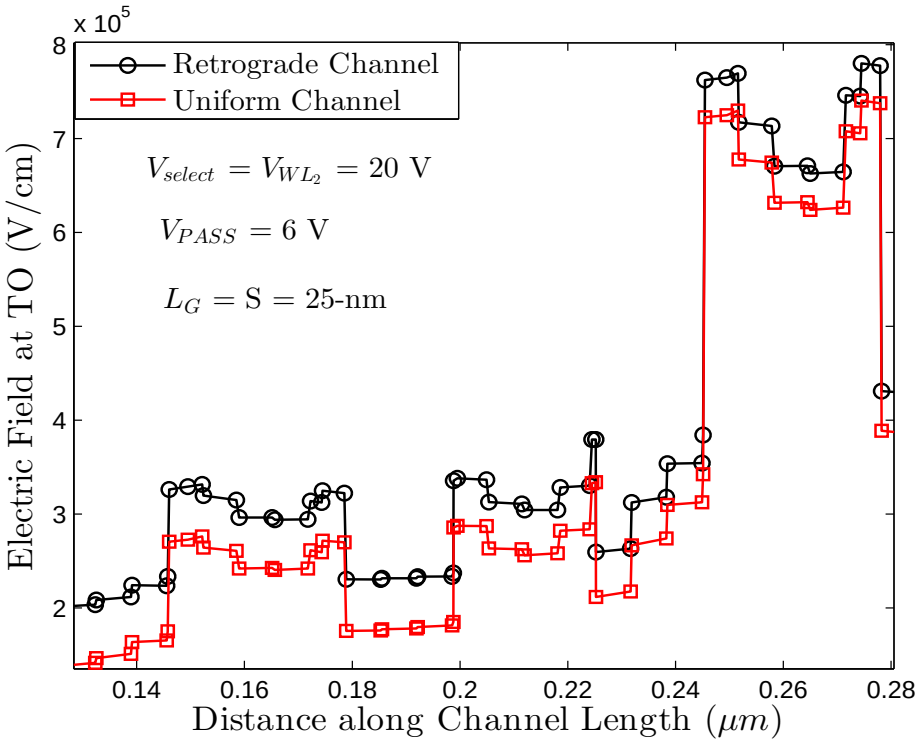

Figure 4.13: Vertical electric field in the tunnel layer corresponding to retrograde and uniform channel doping

The speculation presented above could be verified from Figure 4.13. Here, Figure 4.13 plots the variations of TO vertical electric field for the the select device in the string with 25-nm half pitch technology node. From figure, one can observe that vertical electric field at TO of the select device is higher with retrograde channel doping as compared to the uniform channel doping. Higher carrier accumulation occurs in the

channel under the influence of higher vertical electric field with retrograde channel doping. As a consequence, the offset between the conduction bands of channel and TO tends to reduce resulting into higher tunneling of electrons vertically. This eventually increases the trapping of the charge carriers in the CT layer, as shown in Figure 4.14.

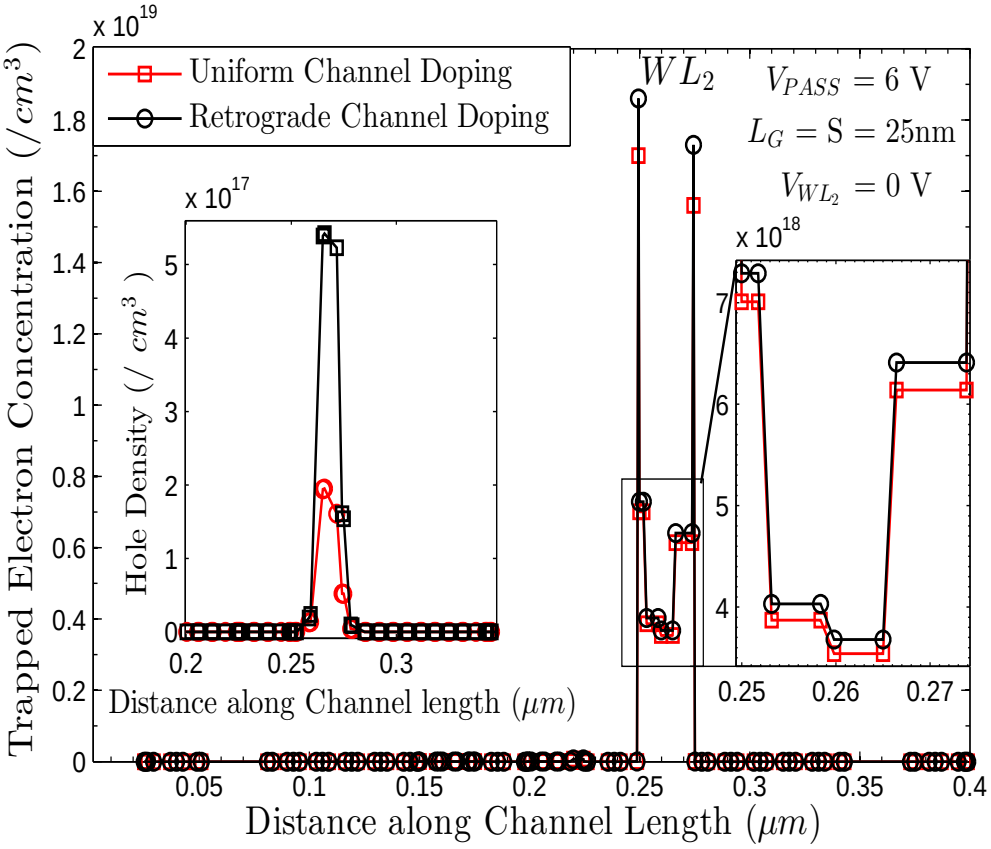

Figure 4.14: Electron trapped density in CT layer with retrograde and uniform channel doping after program operation. High value at the edges of the cell is due to device geometry. Inset: density of holes under the select cell (WL_2) after programm operation

This figure depicts the change in the trapped electron concentration in the CT layer of the select device after program operation with retrograde and uniform channel doping. It is to note that before program operation, all three cells of the string were in the erased state. From figure, it is clear that electron trapping in the CT layer is higher with retrograde channel doping. Further, the inset of Figure 4.14 presents the density of holes under the select device after program operation. One can observe from figure that after program operation, higher hole density is observed under the select memory cell with retrograde channel doping. This represents higher trapping in CT layer of

the select memory cell with retrograde channel doping and hence, provides the higher program threshold voltage.

From the above discussion, one can deduce that NAND flash memory string with retrograde channel doping may attain higher program threshold voltage for similar program time as compared to the uniform channel doping. However, for complete erasing of the select memory cell, longer erasing time may be required.

4.2.3 Effect of Retrograde Channel Doping on Data Retention Characteristics of Junction-Free NAND Flash Memory

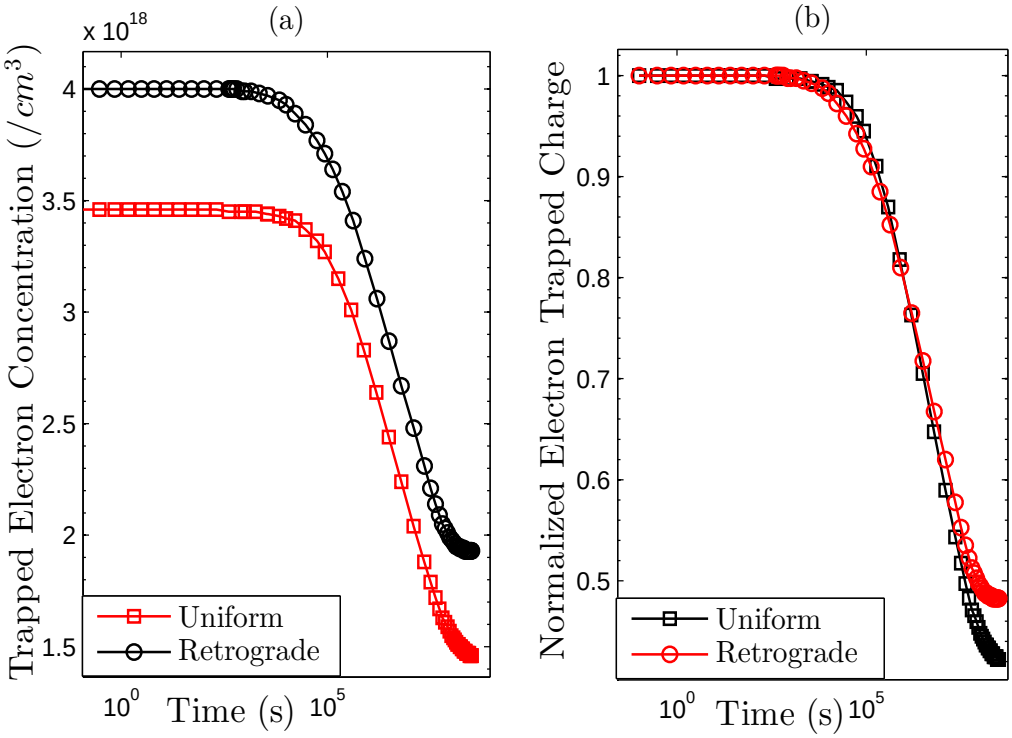

Figure 4.15: (a) Variation of trapped electron concentration with data retention. (b) Variation of normalised trapped electron concentration with data retention

In addition, in this chapter, we investigate the effect of retrograde channel doping on data retention characteristics of the charge trap based NAND flash memory string. Herein, the charge retention capacity for ten years, when the cell is idle, characterises the flash device data retention capability. In this regard, Figure 4.15(a) shows the

simulated data retention characteristics for select flash memory cell with retrograde and uniform channel doping. Importantly, to find retention capability, first the select flash memory cell is programmed in the NAND string with 20 V gate voltage. Now, gate voltage of the programmed memory cell is changed to 0 V from 20 V. After that, we monitored the programmed memory cell behavior for 3×10^8 s. Precisely, we measure the amount of charge in CT layer of programmed flash memory cell at various time instant from 1s to 3×10^8 s. This procedure allow us to know which channel doping profile causes small charge loss from CT layer in the given time. From Figure 4.15(a), it is confirmed that the charge retention capability of CT layer for select memory cell in NAND string increases with retrograde channel doping. Furthermore, as shown in Figure 4.15(b), the normalised charge in CT layer at 3×10^8 s with retrograde channel doping is still 1.14 times larger than the charge with uniform channel doping. Here, the trapped electron concentration in CT layer has been normalised to unity at 1 s.

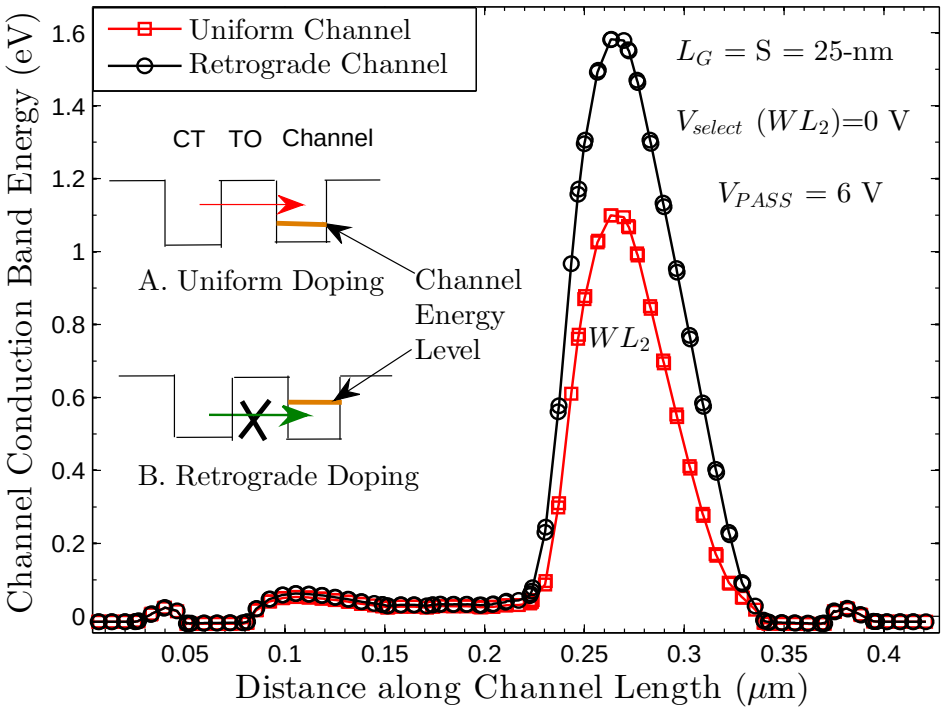

Figure 4.16: Channel conduction band energy under the programmed cell with retrograde and uniform channel doping. Inset shows the effect of energy bands on data retention characteristics of junction-free NAND flash memory

Importantly, with considered 3D NAND flash memory, The phenomenon of improved retention with retrograde channel doping is again attributed to the above mentioned higher charge trapping in CT layer with retrograde channel doping. Here, this high charge trapping in the CT layer with retrograde channel doping attracts high hole concentration towards the interface during the data retention operation. Note that, this high carrier confinement in channel with retrograde channel doping lifts up the channel energy level during retention operation, as shown in Figure 4.16. Furthermore, Figure 4.16 (inset) shows the effect of channel energy band diagram on the tunneling of charge carrier from the CT layer to the channel during the retention mode [51]. Importantly, higher channel energy level reduces the tunneling of carriers from CT layer to channel and therefore, improves the retention capability. Hence, the amount of tunneling from CT layer to channel reduces with retrograde channel doping due to high channel energy level and improves retention. Note that, with retrograde channel doping in 3D junction-free NAND flash memory, improved data retention is obtained without altering the oxide stack.

4.3 Summary

In this chapter, we investigated the impact of retrograde channel doping on the program characteristics, data retention and SCEs for 3D junction-free NAND flash memory string. We demonstrated that the considered NAND flash memory string with retrograde channel doping can attain improved program performance as well as 10 year data retention than that of with uniform channel doping. This is owing to the presence of relatively higher electric field at TO of the select device in NAND string with retrograde channel doping. Further, such channel engineering approach help suppressing the SCEs (i.e., $\triangle V_{th}$, SS and DIBL) for the considered NAND flash memory string. It is worth noting that with this approach the improvement in SCEs can be achieved without making changes to TO stack. This mainly results from better gate controllability over the channel with retrograde channel doping. Here, a thin layer of high doping concentration of the order of 25% of total channel area may cause the improved performance.

In addition, this enhanced memory characteristics with retrograde channel doping in NAND flash memory string could be used to optimise future SSD procurement.

Chapter 5

Conclusions and Future Works

5.1 Conclusions

This thesis investigated the effect of S/D doping lateral straggle σ_L on NAND flash memory cell program characteristics, data retention performance and SCE. We have also studied the variation of channel boost leakage current during program-inhibit mode with σ_L. Here it is shown that, smaller value of σ_L results in better program speed and 10 year data retention. This results from the higher carrier confinement within channel with smaller value of σ_L of S/D doping. Further, utilising this junction engineering in NAND flash memory also suppresses SCE such as $\triangle V_{th}$, SS, DIBL for sub25 nm generation. Improved SCE, without any reduction of gate stack, results from better gate controllability over channel region with reduced σ_L. Further, smaller σ_L reduces channel boost leakage current in sub-25nm NAND flash and suppresses any chances of program disturbance with bulk NAND flash memory. In conclusion, our proposed approach to control S/D σ_L may be applied to develop future high density NAND flash memory.

In addition, this thesis report a new physical phenomenon which may affect the reliability and P/E performance of short channel NAND flash memory cell. Here, it is shown that reduced doping concentration with higher value of d results in increased residual trap charge concentration after erase operation. According to the results presented for short channel SONOS NAND flash memory cell, the effect of residual charge on the

'ON' current after erase operation become small, despite of low doping concentration of LDD region. This phenomenon occurs due to the drift of holes from channel into LDD region under the influence of strong negative gate electric field. Due to holes drift into LDD, high electric field is created at LDD-drain junction which accelerates the channel electrons. Therefore, it is concluded that presence of hole channel into LDD region affects only lateral electric field of short channel SONOS memory cell but still the vertical electric field is governed by LDD doping concentration. Additionally, degradation of program operation is found due to the presence of residual trapped charge.

Furthermore, this thesis also investigates the impact of retrograde channel doping on the program characteristics, data retention and SCEs for 3D junction-free NAND flash memory string. We demonstrated that the considered junction-free VG NAND flash memory string with retrograde channel doping can attain improved program performance as well as 10 year data retention than that of with uniform channel doping. This is owing to the presence of relatively higher electric field at TO of the select cell in NAND string with retrograde channel doping. Further, such channel engineering approach help suppressing the SCEs (i.e., $\triangle V_{th}$, SS and DIBL) for the considered NAND flash memory string. It is worth noting that with this approach the improvement in SCEs can be achieved without making changes to TO stack. This mainly results from better gate controllability over the channel with retrograde channel doping. In addition, this enhanced memory characteristics with retrograde channel doping in NAND flash memory string could be used to optimise future SSD system.

5.2 Future Works

In this thesis, the schemes are presented for the improvement of reliability and performance of NAND flash memory which can be further analyzed with emerging 3D NAND flash memory. Also, many problems related to the topics of this thesis are still open. In the sequel, we discuss some future directions. In 3D NAND flash memory, channel of flash memory string is generally made with polySi material to improve the power performance of flash system. Importantly, PolySi material consist grain boundaries where

trapping of charge carrier may occur. This charge trapping can affect the reliability and performance of NAND flash memory. Therefore, It will of great interest to investigate the effect of grain boundaries in polySi channel on the performance and reliability of junction-free VG NAND flash memory.

Moreover, apart from the VG configuration, 3D integration may also utilise the VC configuration. Owing to its easier fabrication process and decoding method, VC configuration is widely accepted for commercial use in industry for the production of SSD. However, VC flash memory suffers from smaller scaling ability. In fact, reduced channel current and increased tolerable electric field with scaling is a major issue for VC configuration. With this regard, suitable method such as gate stack engineering, channel engineering may be adopted to resolve the concern.

Furthermore, emerging non-volatle memories have emerged as a new paradigm that provide not only the non-volatility for data retention but also the high operational speed for accessing data. It is to note that static RAM, dynamic RAM and flash memories are the key players for the memory of the computing systems these days. These memories are based on the charge storage mechanisms. All these charge based memories face challenges while scaling down the memory below 10nm node and beyond. Easy loss of stored charge at nanoscale results in the degradation of performance and the reliability of the memory system. In this context, several resistance based emerging non-volatile memories (E-NVM) have been investigated to achieve ideal memory characteristics. The E-NVM candidates include: Resistive Random Access Memory (ReRAM), Spin Transfer Torque RAM (STT-RAM), Phase Change RAM (PCRAM). Here, the E-NVM could be useful to improve the performance of computer subsystem in terms of speed and power consumption as compared to the conventional one. Further, low power consumption of E-NVM can be exploited in different IoT applications. Therefore, it would be interesting to explore the properties for E-NVM to add new functionality and features to the system.

With aforementioned future research directions, one can design and formulate new efficient future non-volatile memories to help in the advancement of next generation handhold applications such as smart phones, tablets and personal computer systems.

On the other, the future non-volatile memory also has huge scope in IoT systems for the data storage.

Bibliography

[1] Y. Shiroishi, K. Fukuda, I. Tagawa, H. Iwasaki, S. Takenoiri, H. Tanaka, H. Mutoh, and N. Yoshikawa, "Future Options for HDD Storage," *IEEE Transactions On Magnetics*, Vol. 45, No. 10, pp. 3816-3822, October 2009.

[2] F. Masuoka, A. Momodomi, Y. Iwata and R. Shirota, "New ultra high density EPROM and flash EEPROM with NAND structure cell," *IEEE Electron Device Meeting*, pp. 552-555, August 1987.

[3] R. Bez, E. Camerlenghi, A. Modelli, and A. Visconti, "Introdction to flash memory", *Proeedings of the IEEE* , Vol. 91, No. 4, pp. 489-502, April 2003.

[4] M. H. White, D.A. Adams, and J. Bu, "On the Go with SONOS, " *IEEE Circuits and Devices Magzine*, Vol. 16, No.4, pp. 22-31, July 2000.

[5] K. Kuesters, C. Ludwig, T. Mikolajick, N. Nagel, M. Specht, V. Pissors, N. Schulze, E. Stein, and J. Willer, "Future trends in charge trapping memories," *Proceedings of International Conference on Solid-State Integrated Circuits Technology*, pp. 740743, July 2006.

[6] S. Kamohara, D. Park, C. Hu, "Deep-Trap SILC (Stress Induced Leakage Current) model for nominal and weak oxides," *Proceedings of IEEE International Annual Reliability Physics Symposium*, pp. 57-61, August 2002.

[7] J.D. Lee, S.H. Hur, and J.D. Choi, "Effects of floating-gate interference on NAND flash memory cell operation, " *IEEE Electron Device Letters*, Vol. 23, No. 5, pp. 264-266, May 2002.

[8] Y. S. Hisamune, K. Kanamori, T. Kubota, Y. Suzuki, M. Tsukiji, E. Hasegawa, A. Ishitani, and T. Okazawa, "A high capacitive-coupling ratio (HiCR) cell for 3 V-only 64 Mbit and future flash memories," *IEEE International Electron Devices Meeting*, pp. 2.3.1-2.3.4, 1993.

[9] P. Y. Wang and B. Y. Tsui, "A novel approach using discrete grain-boundary traps to study the variability of 3-D vertical-gate NAND flash memory cells devices," *IEEE Transactions on Electron Devices*, Vol. 15, No.2, pp. 2488-2493, August 2015.

[10] M. K. Jeong, H. J. Kang, S. M. Joe, S. K. Park, B. G. Park and J. H. Lee, "Trap profiling in nitride storage layer in 3-D NAND flash memory using retention characteristics and AC-gm method," *IEEE Electron Device Letters*, Vol. 36, No.6, pp. 561-563, June 2015.

[11] Y. H Hsiao, H. T. Lue, W. C. Chen, K. P. Chang, B. Y. Tsui, K. Y. Hsieh, C. Y. Lu, "The impact of V_{PASS} interference on charge trapping NAND flash memory devices," *IEEE Transactions on Device and Material Reliability*, Vol. 62, No.8, pp. 136-141, January 2015.

[12] (2011). *Internaltional Technology Roadmap for Semiconductors (ITRS)* [online]. Available: http//www.itrs.net

[13] D. Oh, B. Lee, E. Kwon, S. Kim, G. Cho, S. Park,S. Lee, S. Hong, "TCAD simulation of data retention characteristics of charge trap devices for 3-D NAND flash memory," *Proceedings of the IEEE International Memory workshop*, pp. 113-116, May 2015.

[14] P. Paolo, R. Bez, P. Olivo, E. Zanoni, "flash memory cells : An overview," *Proceedings of the IEEE*, Vol. 85, No. 8, pp. 1248-1271, August 1997.

[15] H. W. Chang, C. Y. Liu, C. S Lin, Y. P. Chao, Y. W. Chang, T. C. Lu, K. C. Chen, C. Y. Lu, "Advanced spice modelling for program-inhibit operation in NAND flash memories," *Internation Symposium on VLSI Technology, System and Applications*, pp. 13-14, June 2013.

[16] J . K. Kim, "Flash memory program inhibit operation," *US Patent*, US8300468B2, October 2012.

[17] Paolo Cappelletti, "Non volatile memory evolution and revolution," *IEEE International Electron Device Meeting*, pp. 241-244, December 2015.

[18] K. Takeuchi, "Scaling challenges of NAND flash memory and hybrid memory system with strage class memory and NAND flash memory," *IEEE Custom Integrated Circuit Conference*, pp. 13-19, September 2013.

[19] S. G. Jung, K. W. Lee, K. S. Kim, S. W. Shin, S. S. Lee, J. C. Om, G. H. Bae, J. H. Lee, "Modelling of V_{th} shift in NAND flash memory cell device considering crostalk and short channel effects," *IEEE Transactions on Electron Devices*, Vol. 55, No. 4, pp. 1020- 1026, April 2008.

[20] R. Shirota, B. J. Yang, Y. Y. Chiu, Y. T. Wu, P Y Wang, J. H. Chang, M. Yano, M. Aoki, "Improvement of oxide reliability in NAND flash memories using tight endurance cycling with shorter idling period," *IEEE International Reliability Physics Symposium* , pp. 15-19, June 2015.

[21] S. H. Gu, C. W. Hsu, T. Wang, W. P. Lu, Y. H. J. Ku and C. Y. Lu, "Numerical Siulation of Bottom Oxide Thickness Effect on Charge Retention in SONOS Flash Memory Cells, " ,*IEEE Trans on Elec. Dev.*, Vol. 54, No. 1, pp. 90-97, Jan. 2007.

[22] P. Y. Du, S. H Chen, Y. H. Shih, K. Y. Hsieh,C. Y. Lu, "Overview of 3D NAND flash and progress of split page 3D vertical gate NAND architecture," *IEEE International Conference on Solid State and Integrated Circuit Technology*, pp. 1-4, October 2014.

[23] S. Aritome, R. Shirota, G. Hemink, T. Endoh, F. Masuoka, *et. al.*, "Reliability issues of flash memory cells," *IEEE International Conference on Solid State and Integrated Circuit Technology*, Vol. 81, No. 5, pp. 776-788, May 1993.

[24] T. Kim, N. Franklin, C. Srinivasan, P. Kalavade, and A.Goda, "Extreme short-channel effect on RTS and inverse scaling behavior: source-drain implantation effect

in 25-nm NAND flash memory," *IEEE Electron Device Letters*, Vol. 32, No. 9, pp. 1185-1187, September 2011.

[25] A. Torsi, Y. Zhao, H. Liu, T. Tanzawa, A. Goda, K. Parat, "A program disturb model and channel leakage current study for sub-20 nm nand flash cells," *IEEE Transactions on Electron Devices*, Vol. 58, No. 1, pp. 11-16, November 2010.

[26] S.K. Saha, "Scaling considerations for sub-90 nm split-gate flash memory cells," *IET Circuit Device and Systems*, Vol. 2, No. 1, pp.144-150, February 2008.

[27] A. Shima, H. Ashihara, T. Mine, Y. Goto, M. Horiuchi, Y. Wang, S. Talwar, and A. Hiraiwa, "Self-limiting laser thermal process for ultra shallow junction formation of 50 nm gate CMOS," *IEEE International Electron Device Meeting*, pp. 493-496, December 2003.

[28] J. Gelpey, S. McCoy, A. Kontos, L. Godet, C. Hatem, D. Camm, J. Chan, G. Papasouliots, and J. Scheuer, "Ultra-shallow junction formation using flash annealing and advanced doping techniques," *Proceedings International Workshop on Junction Technology*, pp. 82-86, May 2008.

[29] Y. J. chen, L. H. chong, S. Lin, T. H. Yeh, K. F. Chen, J. Huang, C. Cheng, S. ku, N. Zous, T. Han, T. Hsu, H. Lue, M. Chen, W.Lu, K. chen, C. Lu, "Source/Drain dopant concentration induced Reliability issue in charge trapping NAND flash cells," *IEEE International Reliability Physics Symposium*, pp. 634-638, May 2010.

[30] S. Fang, C. Chen, "High voltage gate formation," U. S. Patent 8822289 B2, September 2014.

[31] G. W. Chang, T. C. Chang, Y. E. Syu, Y. H. Tai, and F. Y. Jian, "On-Current decrease after erasing operation in the nonvolatile memory device with LDD structure," *IEEE Electron Device Letters*, Vol. 32, No. 8, pp.1038-1040, August 2011.

[32] D. Gupta and S. K. Vishvakarma, "Improved short channel characteristics with long data retention time in extreme short channel flash memory devices," *IEEE Transactions on Electron Devices*, Vol. 63, No. 2, pp. 668-674, Febuary 2016.

[33] H. T. Lue, E. K. Lai, Y. H. Hsiao, S. P. Hong, M. T. Wu, F. H. Hsu, N. Z. Lien, S. Y. Wang, L.W. Yang, T. Yang, K.C. Chen, K.Y. Hsieh, Rich Liu, C. Y. Lu, "A noval junction-free BE-SONOS NAND Flash Memory," *Symposium on VLSI Technology*, pp. 140-141, June 2008.

[34] Y. H. Hsiao, H. T. Lue, K. Y. Hsieh, R. Liu and C. Y. Lu, "Modelling and scaling evaluation of junction-free charge trapping NAND flash devices," *International Symposium on VLSI Technology, System and Applications*, pp. 103-104, April 2009.

[35] H. T. Lue, Y. H. Hsiao, P. Y. Du, S. C. Lai, T. H. Hsu, S. P. Hong, M. T. Wu, F. H. Hsu, N. Z. Lien, C. P. Lu, J. Y. Hsieh, L. W. Yang, T. Yang, K. C. Chen, K. Y. Hsieh, R. Liu, C. Y. Lu, "A Novel burried channel FinFET BE -SONOS NAND flash with improved memory window and cycling endurance," *Symposium on VLSI Technology Digest of Technical Papers*, pp. 224-225, August 2009.

[36] Y. H. Hsiao, H. T. Lue, K. Y. Hsieh, R. Liu, C. Y. Lu, "A study of stored charge interference and fringing field effects in sub-30nm charge trapping NAND flash," *IEEE International Memory Workshop*, pp. 1-2, May 2009.

[37] S. Aritome, "NAND flash memory revolution," *IEEE International Memory Workshop*, pp. 1-4, June 2016.

[38] B. Eitan, P. Pavan, I. Bloom, E. Aloni, A. Frommer and D. Finzi, "Can NROM, a 2 bit trapping storage NVM cell, give a real challenge to floating gate cells?", *Proceedings Solid State Devices and materials*, pp. 522-524, September 1999.

[39] M. H. White, Y. Yang, A. Purwar,and M. L. French, "A low voltage SONOS nonvolatile semiconductor memory technology, " *IEEE Transactions on Components,Packaging and Manufacturing Technology Part A*, Vol 20, No 2, pp. 190-195, June 1997.

[40] M. She, "Semiconductr flash memory scaling," *University of California Berkeley (PhD)*, 2003.

[41] K. Miyaji, C. Hung and K. Takeuchi, "Scaling trends and tradeoffs between short channel effect and channel boosting characteristics in sub-20 nm bulk/silicon-on-Insulator NAND flash memory," *Japanese Journal of Applied Physics*, Vol. 51, 04DD12, April 2012.

[42] J. Brault, M. Saitoh, T. Hiramoto, "Channel width and length dependence in silicon naocrystal memories with ultra-nanoscale channel," *IEEE Transactions on Electron Devices*,Vol. 4, No. 3, pp. 349-354, May 2005.

[43] A. Nandi, A.K. Saxena, S. Dasgupta, "Analytical modeling of double gate MOSFET considering source/drain lateral gaussian doping profile," *IEEE Transactions on Electron Devices*, Vol. 60, No. 11, pp. 3705-3709, November 2013.

[44] V. Vijayvargiya and S. K. Vishvakarma, "Effect of drain doping profile on double gate tunnel field effect transistor and its influence on device RF performance," *IEEE Transactions on Nanotechnology*, Vol. 13, No. 5, pp. 974-981, September 2014.

[45] M. S. Alam, T. C. Lim and G. A. Armstrong, "Analog performance of double gate SOI transistors," *International Journal of Electronics*, Vol. 93, No. 1, pp. 118, January 2006.

[46] T. H. Hsu, H. T. Lue, S. C. Lai, Y. C. King, K. Y. Hsieh, R. Liu, and C. Y. Lu, "Reliability of planar and FinFET SONOS devices for NAND flash applications - field enhancement Vs. barrier engineering," *VLSI Technology, Systems and Applications*, pp. 154-155, April 2009.

[47] C. S. Lin, Y. C. Chen, T. C. Chang, F. Y. Jian, H. W. Li, S. C. Chen, Y. S. Chuang, T. C. Chen, Y. H. Tai, M. H. Lee, J. S. Chen, "Charge trapping-induced parasitic capacitance and resistance in SONOS TFTs under gate bias stress," *IEEE Electron Device Letters*, Vol. 32, No. 3, pp. 321-323, March 2011.

[48] D. Oh, C. Lee, S. Lee, T. K. Kim, J. Song, J. Choi, "A new self-boosting phenomenon by source/drain depletion cut-off in NAND flash memory," *IEEE Non-Volatile Memory Workshop*, pp. 39-41, August 2007.

[49] I. H. Park, W. G. Hahn, K. W. Song, K. H. Choi, H. K. Choi, S. B. Lee, C. S. Lee, J. H. Song, J. M. Han, K. H. Kyoung, Y. H. Jun, "A new GIDL phenomenon by field effect of neighbouring cell transistors and its control solutions in sub-30nm NAND flash device," *Symposium on VLSI Technology*, pp. 23-24, June 2012.

[50] *Sentaurus TCAD Manuals, Synopsys Inc., Mountain View, CA, USA, 2012.*

[51] M. Saitoh, E. Nagata and T. Hiramoto, "Large memory window and long charge retention time in ultranarrow-channel silicon floating-dot memory," *Applied Physics Letter*, Vol. 82, No. 11, pp. 1787-1789, March 2003.

[52] K.P. Chang, H.T. Lue, Y.H. Hsio, K.Y.Hsieh and C.Y. Lu, "Effect of junction engineering for 38nm BE-SONOS charge trapping NAND flash device," *VLSI Technology, Systems and Applications (VLSI-TSA)*, pp. 1-2, April 2011.

[53] K. Takeuchi, S. Satoh, K.I. Imamiya, and K. Sakui, "A source-line programming scheme for low-voltage operation NAND flash memories," *IEEE Journal of Solid State Circuits*, Vol. 35, No. 5, pp. 672-681, May 2000.

[54] C.H. Lee, J. Choi, Y. Park, C. Kang, B.I. Choi, H. Kim ,H. Oh and W.S. Lee, "Highly scalable NAND flash memory with robust immunity to program disturbance using symmetric inversion-type source and drain structure," *Symposium on VLSI Technology*, pp. 118-119, June 2008.

[55] C. H. Lee, I. C. Yang, Chienying Lee, C. H. Cheng, L. H. Chong, K. F. Chen, J. S. Huang, S. H. Ku, N. K. Zous, I. J. Huang, T. T. Han, M. S. Chen, W. P. Lu, K. C. Chen, Tahui Wang, and Chih-Yuan Lu, "Junction optimization for reliability issues in floating gate NAND flash cells, " *Proceedings of International Reliability Physics Symposium*, pp. 6B3.1-6B3.5, April 2011.

[56] K. Miyaji, C. Hung, K. Takeuchi, "Pushing scaling limit due to short channel effects and channel boosting leakage from 13nm to 8nm with SOI NAND flash memory cells," *International Conference on Solid State Devices and Materials*, pp. 128-129, September 2011.

[57] A. Chimenton, A. S. Spinelli, D. Ielmini, A. L. Lacaita, A. Viscanti, and P. olivo, "Drain accelerated degradation of tunnel oxides in flash memories," *IEEE International Electron Device Meetings*, pp. 167-170, December 2001.

[58] A. Bonfiglietti, M. Cuscuna, A. Valletta, L. Mariucci, A. Pecora, G. Fortunato, S. D. Brotherton, and J. R. Ayres, "Analysis of electrical characterstics of gate overlapped lightly doped drain (GOLDD) polysilicon thin-film transistors with different LDD doping concentration," *IEEE Tranactions On Electron Devices*, Vol. 50, No. 12, pp. 2425-2433, December 2003.

[59] T. H. Hsu, H. T. Lue, S. C. Lai, Y. C. King, K. Y. Hsieh, R. Liu, C. Y. Lu, "Reliability of planer and FinFET SONOS devices for NAND flash applications-field enhancement vs. barrier engineering," *IEEE International Electron Device Meeting*, pp.154-155, April 2009.

[60] K. P. Chang, H. T. Lue, Y. H. Hsaio, K. Y. Hsieh and C. Y. Lu, "Effect of junction engineering for 38nm BE-SONOS charge trapping NAND flash device," *VLSI Technology, Systems and Applications*, pp. 1-2, April 2011.

[61] E. Murakami, T. Yoshimura, Y. Goto, and S. Kimura, "Gate length scalability of n-MOSFET's down to 30 nm: comparison between LDD and non-LDD structures," *IEEE Tranactions on Electron Devices*, vol. 47, No. 4, pp. 835-840, April 2000.

[62] M. Kimura, A. Nakashima, and Y. Sagawa, "Mechanism analysis of off leakage current in poly-Si TFTs with LDD structure," *Electrochemical Solid-State Letter*, Vol. 13, No. 12, pp. 409-411, September 2010.

[63] J. Jang, H. S. Kim, W. Cho, H. Cho, J. Kim, S. I. Shim, Younggoan, J. H. Jeong, B. K. Son, D. W. Kim, Kihyun, J. J. Shim, J. S. Lim, K. H. Kim, S. Y. Yi, J. Y. Lim, D. Chung, H. C. Moon, S. Hwang, J. W. Lee, Y. H. Son, U. I. Chung, W. S. Lee, "Vertical cell array using TCAT (terabit cell array transistor) technology for ultra high density NAND flash memory," *Symposium of VLSI Technology*, pp. 192-193, August 2009.

[64] H. Tanaka,M. Kido, K. Yahashi,M. Oomura, R. Katsumata, M. Kito, Y. Fukuzumi, M. Sato, Y. Nagata, Y. Matsuoka, Y. Iwata, H. Aochi, A. Nitayama, "Bit cost scalable technology with punch and plug process for ultra high density flash memory," *Symposium of VLSI Technology*, pp. 14-15, June 2007.

[65] Y. Fukuzumi, Y. Matsuoka, M. Kito, M. Kido, M. Sato, H. Tanaka, Y. Nagata, Y. Matsuoka, Y. Iwata, H. Aochi, A. Nitayama, "Optimal integration and characteristics of vertical array devices for ultra-high density, bit-cost scalable flash memory, " *IEEE Internation Electron Device Meeting*, pp. 449-452, December 2007.

[66] J. Kim, A. J. Hong, S. M. Kim, E. B. Song, J. H. Park, J. Han, S. Choi, D. Jang, J. T. Moon, K. L. Wang, "Novel vertical-stacked-arraytransistor (VSAT) for ultra-high-density and cost-effective NAND flash memory devices and SSD," *VLSI Symposium Technology Digest*, pp. 186-187, June 2009.

[67] D. W. Kwon, W. Kim, D. B. Kim, S. H. Lee, J. Y. Seo, E. Choi, G. S. Cho, S. K. Park, J. H. Lee, B. G. Park, "Novel program method of string select transistors for layer selection in channel stacked NAND flash memory," *IEEE Transactions on Electron Devices*,Vol. 63, No. 9, pp. 3521 - 3526, August 2016.

[68] Y. Song, X. Li, "Scaling junctionless multigate field-effect transistors by step-doping," *Applied Physics Letters*, No. 105, 223506, 2014.

[69] S. Gundapaneni, S. Ganguly, A. Kottantharayil, "Bulk planer junctionless transistor (BPJLT): An attractive device alternative for scaling," *IEEE Electron Device Letters*, Vol. 32, No. 3, pp. 261-263, March 2011.

[70] I. De and C. M. Osburn, "Impact of super-steep-retrograde channel doping profiles on the performance of scaled devices," *IEEE Transactions On Electron Devices*, Vol. 46, No. 8, pp. 1711-1717, August 1999.

[71] A. Asenov, S. Saini, "Suppression of random dopant-induced threshold voltage fluctuations in sub-0.1-m MOSFETs with epitaxial and d-doped channels," *IEEE Transactions on Electron Devices*, vol. 46, no. 8, pp. 1718-1724, August 1999.

[72] D. Gupta and S. K. Vishvakarma, "Improvement of short channel performance of junction-free charge trapping 3-D NAND flash memory," *IET Micro and Nano Letter*, Vol. 12, No. 1, pp. 64-68, December 2016..

[73] H.T. Lue, S. C. Lai, T H Hsu, P Y Du, S Y Wang, K Y Heish, R. Liu, C. Y. Lu, "Modeling of barrier-engineered charge-trapping NAND flash devices," *IEEE Transactions on Device and Materials Reliability*, Vol. 10, No. 2, pp. 222-232, July 2010.

[74] K. Wieczorek, M. Horstmann, R. Stephan, "Semiconductor device having a retrograde dopant profile in a channel region and method for fabricating the same," *U.S. Patent, 6,881,641 B2*, April 19, 2005.

[75] J. Cai, A. Majumdar, T. H. Ning, Z. Ren, "Method for fabricating super-steep retrograde well method on SOI or bulk silicon substrate,and device fabricated in accordance with the method," *U.S. Patent, US 2009/0302388 A1*, December 2009.

List of Publications

Journals

1. **Deepika Gupta** and Santosh Kumar Vishvakarma, "Improved Short Channel Characteristics with Long Data Retention Time in Extremely Short Channel NAND Flash Device," *IEEE Transactions on Electron Devices*, Vol. 63, No. 2, pp. 668-674, February 2016.

2. **Deepika Gupta** and Santosh Kumar Vishvakarma, "Impact of LDD depth Variation on the performance characteristics of SONOS NAND Flash Memory Device," *IEEE Transactions on Device and Material Reliability*, Vol. 16, No. 3, pp. 298-303, June 2016.

3. **Deepika Gupta** and Santosh Kumar Vishvakarma, "Improvement of Short Channel Performance of Junction-free Charge Trapping 3-D NAND Flash Memory", *IET Micro & Nano Letters*, Vol. 12, No. 1, pp. 64-68, December 2016. (IF : 0.9)

4. **Deepika Gupta** and Santosh Kumar Vishvakarma, "Improvement of Short Channel Performance and Memory Characteristics of Junction-Free Charge Trapping 3-D NAND Flash Memory ", Under Review.

Conferences

1. **Deepika Gupta** and Santosh Kumar Vishvakarma, "A New Approach to Suppress GIDL in NAND Flash Memory ", 18^{th} *International Workshop on Physics of Semiconductor Devices (IWPSD)*, Banglore (India), December 2015.

Publications besides thesis work

Journals

1. Sonal Jain, **Deepika Gupta**, Santosh Kumar Vishvakarma and Vaibhav Neema, "Investigation of BE-SONOS Flash Memory along with Metal Gate and High-k Dielectrics Tunnel Barrier and its Impact in Charge Retention Dynamics," *Journal of Semiconductors*, Vol. 37, No. 3, January 2016.

2. Sonal Jain, Vaibhav neema, **Deepika Gupta**, Santosh K. Vishvakarma, "Investigation of BE-SONOS Flash Memory with High-k Dielectrics in Tunnel Barrier and its Impact on Charge Retention Dynamics," *Journal of Nanoelectronics and Optoelectronics (American Scientific Publishers)*, Vol. 11, No. 6, pp. 663-668, December 2016.

Conferences

1. Mansimran Kaur, **Deepika Gupta** and Santosh Kumar Vishvakarma, "Rare Earth Materials and LaON in Tunnel Barrier Engineering With Enhanced Memory Characteristics," *IEEE Prime Asia Conference*, pp. 1-5, December 2015.

2. Anupreet Gupta, Hajra Anwar, **Deepika Gupta**, Vikas Vijayvargiya and Santosh Kumar Vishvakarma, "A Non-Volatile Memory MONOS Device for Improved Stability Applications," *IEEE International Conference on Devices, Circuits and Communications (ICDCCom)*, pp 1-3, September 2014.