Novel Floating Gate Materials for Flash Memory

Submitted in partial fulfillment of the requirements

For the degree of

DOCTOR OF PHILOSOPHY

by

Abhishek Misra

(Roll No. 08407603)

Supervisor

Prof. Anil Kottantharayil

Department Of Electrical Engineering

INDIAN INSTITUTE OF TECHNOLOGY BOMBAY

2014

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof. Anil Kottantharayil who introduced myself to a very interesting topic and giving his all kind of support to go inside the topic and understand the field. I am indebted to Prof. Anil Kottantharayil for his valuable guidance, innovative ideas, evaluation of the work and most importantly for his encouragement in my research throughout the course.

I also owe many thanks to my RPC members, Prof. J. Vasi and Prof. Udayan Ganguly for their time to time evaluation and valuable inputs on the work.

I acknowledge Prof. M. Aslam and his student Hemen Kalitha from department of Physics, IIT Bomaby for providing graphene oxide and useful discussions.

I acknowledge the "IITB Nanofabrication facility" (IITBNF), department of electrical engineering, IIT Bombay where all the work presented in this thesis has been carried out. I wish many thanks to all the IITBNF staff members for their friendly support throughout the work.

I acknowledge "Centre for Research in Nanotechnology and Science (CRNTS) and Sophisticated Analytical Instrument Facility (SAIF), IIT Bombay for HRTEM and Raman measurements. I thank Prof. Soumya Mukherjee, erstwhile head of the SAIF, for allowing me to get trained and have access to the HRTEM facility round the clock. I acknowledge "Central Surface Analytical Facility" and Scanning Probe Microscopy facility of IIT Bombay for XPS and AFM analysis respectively.

I would like to thank Department of Science and Technology, Govt. of India and The Department of Electronics and Information Technology, Govt. of India for financial assistance.

I would like to thank my present and past group members; Shaivalini Singh, Meenakshi B., Kousik Midya, Manali Khare, Amritha Janardanan, Rohini Golve, Manini Gour, Sahul Nath, Satya Suresh, Mayur Waikar, Amit Gour, Suresh Gundapaneni, Sunnay Sadana and Senthil Srinivasan for their all kind of support, discussion on several technical and non technical topics and making my stay at IITB enjoyable. I also thank and acknowledge Mayank Srivastav and other friends who associated with me during my stay at IITB.

A special thanks to Madam J. Arthi for proof reading of the thesis.

I thank my dear parents, my sisters and my wife Neha for their unconditional love, support and patience. It all was possible only because of *their* constant care and encouragement. Their contribution in this research and in my life is beyond the words of acknowledgements. With all my respect and love, I humbly dedicate this thesis to them.

Abhishek Misra **Roll No. 08407603**

Abstract

High data storage capacity and low program/erase voltage of the flash memory is obtained by the continuous scaling of these devices. It is desirable to continue this scaling trend to achieve even higher density of memory cells and hence to lower down the cost of these devices. However, increased proximity of the devices due to lateral scaling causes the increased capacitive coupling among the floating gates of neighboring cells. This capacitive coupling is undesirable as memory operations performed on one memory cell affects the threshold voltage of the neighboring cells. Floating gate height reduction is one possible way to reduce the cross cell interference of scaled flash memory cells.

In this work, charge storage capability of multilayer graphene is investigated. As the interlayer spacing between two graphene sheets in multilayer graphene is only 0.34 nm, 6-7 layers of multilayer graphene would be 2-3 nm thick. Therefore, incorporation of multilayer graphene as floating gate in flash memory structure would offer the ultimate scalability to the vertical dimension of these devices. Graphene is reported to be thermally stable up to 1500°C and hence thermal stability issues anticipated with thin metal floating gate may not be a problem with graphene floating gate devices. Multilayer graphene is chosen for several technical advantages over single layer graphene. First, work function of graphene is susceptible to the number of layers when it is less than four. Graphene work function varies from 4.2 eV for single layer graphene, it increases and saturates to 4.6 eV for more than four layers of graphene. This variation in the work function would cause a variation in the potential well depth formed by graphene floating gate layer. This can be avoided with multilayer graphene having more than four layers. In addition to this, high work function of multilayer graphene is desirable to achieving long-term data retention. Further, higher density of states in multilayer graphene compared to single layer graphene and reduced conductivity along C-axis in multilayer graphene are desirable for floating gate application.

Charge storage capability of multilayer graphene is evaluated in flash Metal Oxide Semiconductor (MOS) capacitors, flash transistors and by MOS capacitors with implanted surroundings with SiO_2 as tunnel oxide, Al_2O_3 as a blocking dielectric and TiN as a gate electrode. The multilayer graphene is obtained by the thermal reduction of the graphene oxide and thus called as reduced multilayer graphene. A memory window of 9.4 V for 1 sec.

programming pulse is obtained at ±20 V program/erase voltage with MOS capacitors having implanted surroundings. Number of electrons stored in reduced multilayer graphene sheets after 18 V programming voltage is calculated as 1 x 10^{13} cm^{-2} which is higher than the density of states in single layer graphene. Retention of the stored charges is tested at room temperature as well as at elevated temperatures. 6.9 V remnant memory window at room temperature and 2.8 V remnant memory window at 150°C after 10 years is demonstrated.

As the memory performance strongly depends on the work function of the charge storage layer, it is highly desirable to know the work function of the reduced multilayer graphene sheets used as a charge storage layer in the present work. Therefore, the work function of the reduced graphene oxide is calculated by integrating them as a gate electrode in the MOS structure under different contact metals. Experimental data reveal that work function of the reduced graphene oxide can be modulated by varying the thickness of the reduced graphene oxide sheets as well as by varying the amount of oxygen concentration attached to the reduced graphene oxide sheets. The work function of the gate electrode shows strong dependence on the number of reduced graphene oxide sheets and is seen to be nearly independent of the contact metals used. A minimum of about 4.35 eV and a maximum of about 5.28 eV work function values are obtained with very thin and very thick reduced graphene oxide sheets under Platinum/Titanium Nitride (Pt/TiN) contact metal. The observed work function modulation is attributed to the different amounts of the oxygen concentration in different thicknesses of reduced graphene oxide layers. To confirm this hypothesis, thick graphene oxide (so as to exclude the layer dependent contribution) is reduced at different temperatures. Reduction at different temperatures causes different amounts of remnant oxygen in the reduced graphene oxide sheets. High oxygen concentration in the reduced graphene oxide sheets corresponds to the high work function and the vice versa. The oxygen concentration in the reduced graphene oxide sheets is obtained by X-ray photoelectron spectroscopy (XPS).

These work function values are found to be thermally stable upto 800°C thermal annealing in nitrogen ambient. Post annealing, cross section high resolution transmission electron microscopy (X-HRTEM) analysis of the samples reveals that the diffusion of metal through reduced graphene oxide layers at higher temperature is the main cause of work function instability of the graphene gate electrodes in the MOS structure. Incorporation of reduced graphene oxide between

contact metal and the gate dielectric not only allows to modulate the work function of the gate electrode but also improves the gate dielectric reliability. It is confirmed by X-HRTEM analysis that graphene behaves as a diffusion barrier for metals and as a result prevents the metallic contamination of the dielectric.

The information on the work function of reduced graphene oxide sheets is very important to optimize the overall performance of reduced graphene oxide charge storage layer flash devices. This study, in a very distinct manner, suggests that by controlling the amount of oxygen concentration in the reduced graphene oxide sheets, depth of the potential well formed by reduced graphene oxide sheets and hence the memory performance of the reduced graphene oxide charge storage layer flash devices can be tailored.

List of Tables

List of Figures

List of Abbreviations and Chemical Symbols

AFM	Atomic Force Microscopy
AG	Air Gap
Al	Aluminum
Al_2O_3	Aluminum Oxide
AMAT	Applied Materials
BIOS	Basic Input/Output System
BHF	Buffered Hydrofluoric Acid
BE	Binding Energy
CHE	Channel Hot Electron
CMOS	Complementary Metal Oxide Semiconductor
CSL	Charge Storage Layer
CG	Control gate
C-OH	Hydroxyl
C-O-C	Epoxide
COOH	Carboxyl
C=O	Ketonic
CTF	Charge Trap Flash Memory
CVD	Chemical Vapor Deposition
DOS	Density Of States
DSA	Double Side Aligner
DRAM	Dynamic Random Access Memory
EBL	Electron Beam Lithography
EDAX	Energy Dispersive X-Ray
EEPROM	Electrically Erasable Programmable Read Only Memory
EOT	Effective Oxide Thickness
FG	Floating Gate
FGA	Forming Gas Anneal
FTIR	Fourier Transform Infrared Spectroscopy

GO	Graphene Oxide
Gr	Graphene
HDD	Hard Disc Drives
HfO_2	Hafnium dioxide
HfAlO	Hafnium Aluminum Oxide
HRTEM	High Resolution Transmission Electron Microscopy
IPD	Interpoly Dielectric
Ir	Iridium
ITRS	International Technology Roadmap For Semiconductors
KPM	Kelvin Probe Force Microscopy
LPCVD	Low Pressure Chemical vapor Deposition
MOCVD	Metal Organic Chemical vapor Deposition
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
MOSCAP	Metal Oxide Semiconductor Capacitor
MLG	Multilayer graphene
MNC	Metal nanocrystal
NC	Nanocrystal
NVM	Nonvolatile Memory
PVD	Physical Vapor Deposition
Pt	Platinum
rGO	Reduced Graphene Oxide
RAM	Random Access Memory
ROM	Read Only Memory
RCA	Radio Corporation Of America
RIE	Reactive Ion Etching
RTP	Rapid Thermal Processing
SD	Standard Deviation
SEM	Scanning Electron Microscopy
SLG	Single layer graphene

SILC	Stress induced leakage current
SiO_2	Silicon dioxide
Si_3N_4	Slilcon Nitride
STI	Shallow Trench Isolation
SRAM	Static Random Access Memory
SONOS	Silicon-Oxide-Nitride-Oxide-Silicon
TDEAH	Tetrakis [DiEthylAmino]Hafnium
TiN	Titanium Nitride
TEM	Transmission Electron Microscopy
TANOS	Tantalum Aluminum Nitride Oxide Silicon
UPS	Ultraviolet Photoelectron Microscopy
VLSI	Very Large Scale Integration
WKB	Wentzel- Kramers-Brillouin
WF	Work function
XPS	X-Ray Photoluminescence Spectroscopy
X-HRTEM	Cross-section High Resolution Transmission Electron Microscopy
X-SEM	Cross-section Scanning Electron Microscopy

Chapter 1

Introduction

Flash memories are the basic ingredients for data storage in all sorts of portable electronic devices like mobile phones, digital cameras etc. NAND flash based solid state derives are also likely to replace the hard disc drives (HDD) in future computers and laptops. The ever increasing demand for all these electronic equipments in the consumer market is the key driving force behind the efforts to increase the data storage capacity and to reduce the cost per bit stored in flash memories. Figure 1.1(a) shows the cost of data storage in NAND flash in the recent years [1]. However, still the cost per bit in NAND flash cell is higher than that of in HDD as shown in figure 1.1(b) [2]. Therefore, continuous efforts are being made to increase the data density and to lower down the cost per bit in flash memories by decreasing the size of the flash cell to smaller and smaller dimensions.

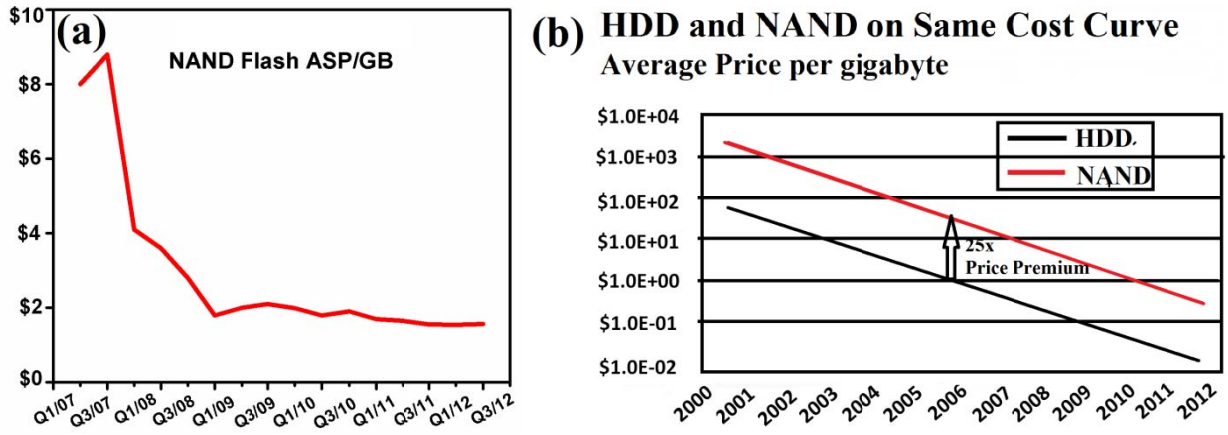


Figure 1.1: (a) Cost reduction of data storage in flash cell in the recent years (source: http://agigatech.com/blog/page/2/), (b) comparison of the cost per gigabyte for NAND flash and HDD (Source: http://agigatech.com/blog/system-uses-for-nand-flash/). Figures re-plotted from given sources.

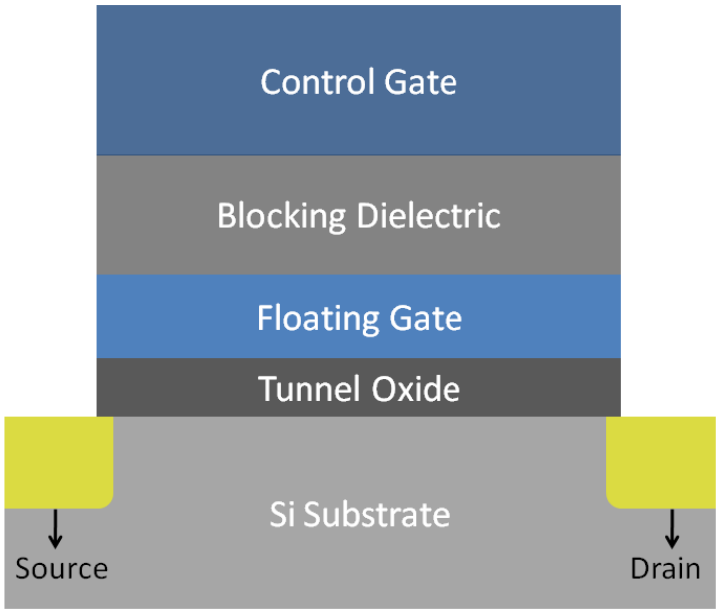


Figure 1.2: Schematic of the cross section of conventional flash memory devices.

A schematic of the cross section of conventional flash memory device is shown in figure 1.2. Gate stack of flash memory consists of tunnel oxide, floating gate (also known as charge storage layer), blocking dielectric and top gate electrode. This flash cell could be successfully scaled down to 20 nm technology node [3]. However, further scaling below this technology node with the conventional device structure is challenging due to several technical challenges like non scalability of tunnel and blocking dielectric, decrease in the number of stored electrons, increased cross talk among the neighboring devices and breakdown voltage limitation in the word line [3]. With the scaling of the memory devices, thickness of the tunnel and blocking dielectric is approaching towards its physical limitations. Further reduction in the tunnel oxide thickness (less than ~ 8 nm) [3] would lead to poor data retention due to leakage of the stored charge through the defects in the SiO_2. Apart from these issues, increased proximity among the devices due to lateral scaling also causes the following two undesirable effects on the memory operations.

- Reduced space between the two memory devices inhibits the covering of the side walls of the floating gate (FG) by blocking dielectric and control gate (CG) leading to the planar geometry of the devices [3]. This planar geometry results in the reduced FG to CG coupling

ratio from a desirable value of 60% to only 40% [3]. Reduced coupling ratio causes the early program saturation due to strongly increased electric field in the blocking dielectric during program and erases operations [4]. A 60% coupling ratio is desirable to have the program and erase mechanism without any major current flowing through the blocking dielectric.

- Increased electrical cross talk among the neighboring FGs due to parasitic capacitances [5] which affects the threshold voltage of the devices.

Considering these issues, further scaling necessitates either the inclusion of new materials in the flash gate stack or altered device structure. For example, to make the memory device more immune to charge loss through the defects in surrounding dielectrics and hence to improve the data retention, discretization of the FG was proposed. In this approach, charge storage layer, i.e. floating gate no longer remains electrically continuous and the stored charge remains well isolated. Two ways were proposed to achieve this, one is known as charge trap flash (CTF) and the other is known as the nanocrystal (NC) floating gate memory. In CTF memory, charge is stored in traps available in the dielectrics while in the NC memories, charge is stored in well isolated metal or semiconductor islands. In both the techniques, data loss through the defects in the surrounding dielectric is very localized and hence are supposed to be more immune against the charge loss through the defects in the dielectric. Different dielectrics like slilcon nitride (Si_3N_4), hafnium dioxide (HfO_2), Aluminium oxide (Al_2O_3) and some composite oxide of hafnium and aluminum (HfAlO) are investigated as a charge trap layer [6-14]. Memory performance of CTF memories strongly depends on the density and nature (i.e. shallow or deep traps) of the traps which in turn depends on the composition of the dielectric film. In silicon nitride CTF based memories, electron or hole trap depth can be varied by varying the nitrogen or silicon composition [8-10] while the HfO_2 and Al_2O_3 based CTF memories have issues like poor charge retention and slower programming speed respectively [14]. On the NC memories, both semiconductor (silicon and germanium) and metal NCs have been evaluated as a charge storage medium [15-20]. It is proposed that metal NCs are more suitable for charge storage purpose compared to their semiconductor counter parts [18]. Even though this technology owns several advantages, there are some fundamental issues associated with it. The memory performance is affected by nanocrystal size, density and distribution [21, 22]. To assemble the nanocrystals

having uniform distribution on each memory cell is technically challenging. Variation in the NC statistics causes the device to device variability [22, 23]. In order to reduce this cell to cell variability, a very high density of NCs (~ 5 x 10^{12} cm^{-2}) is required; however, with such high density of NCs, spacing between NC to NC decreases significantly (~ 2-4 nm apart). With such closely packed density even though they physically may remain separated, however, electrically they behave like continuous metal floating gate [24]. Therefore, the proposed benefits of isolated charge storage nodes may not be achieved with this technology. Further, charge redistribution among the nanocrystals also causes the threshold voltage instability [24]. Infact, the successful scaling of FG flash cell towards 20 nm node has put even more stringent requirements on the nanocrystal distribution.

To minimize the interference among the neighboring FGs of the flash memory array, reduction of the FG height and the use of low-K dielectrics for device isolations are proposed [5]. Reducing the height of the polycrystalline silicon (poly-Si), which is a conventional FG material, can be a possible solution to reduce the FG to FG interaction. However, ultrathin poly-Si (less than 7 nm) layer is not capable of preventing the vertical charge carriers of the programming current and these unscattered carriers known as ballistic carriers reach to the blocking dielectric after crossing the thin poly-Si film [25]. These carriers may cause impact ionization in the blocking dielectric layer and thus degrade the dielectric quality [25]. It will also lead to the early program saturation. To minimize this ballistic current component, metal layers as a FG are proposed to be more effective as they possess sea of free electrons and hence increased probability of scattering of incoming carriers during programming, thereby reducing the ballistic current component [25]. It is experimentally demonstrated that a good memory action can be obtained with a metal FG as thin as of 1nm thickness [25, 26]. High work function (WF) metals are suggested to be more suitable for this purpose because of the deep potential well offered by these metals prevents early program saturation. However, this improvement in the program saturation may cause degradation in the erase performance because of increased tunnel barrier for carriers tunneling from FG to tunnel oxide. High work function of FG metals may also cause the erase saturation [27, 28]. To overcome these issues, hybrid FG layer i.e. a combination of low work function material at the bottom and high work function material at the top is proposed

[27, 28]. In this technique, hybrid FG composed of n+ poly and TaN metal is investigated to improve the erase performance. Further, to reduce the overall floating gate height, combination of thin n+ poly and thin metal is also studied very recently [29, 30].

Incorporation of the metal layer as a FG in the memory gate stack, however, may impose its own device reliability issues like; (1) diffusion of metal into the tunnel and blocking dielectrics at high temperatures [31], (2) increased leakage current through the blocking dielectric deposited on metal films due to higher degree of dielectric crystallization at elevated temperatures [32] and (3) agglomeration of thin metal layers at elevated temperatures during device processing [18].

In the recent times, graphene (Gr) has attracted much attention due to its extraordinary material and electrical properties like two dimensional sheet like structure, very high carrier mobility and high thermal and mechanical flexibility [33]. In the graphene research, most of the efforts are concentrated to utilize its high mobility of charge carriers in complementary metal oxide semiconductor (CMOS) devices as a channel material. In several reports, graphene is proposed as a channel material in conventional silicon (Si) transistors to increase the drive current of the transistors [34, 35]. However, zero or negligible band gap in large area graphene is a major bottleneck in realizing the desired I_{ON}/I_{OFF} ratio. Nevertheless, the other unusual properties of Gr such as 2-dimensional sheet like structure, high thermal stability and mechanical flexibility motivate to explore new avenues for its application in VLSI devices.

Owing to the atomically thin nature, metallic character, and high thermal stability of Gr, it can be used as a charge storage layer (CSL) in conventional FG flash memory devices. Inclusion of Gr as CSL in flash memory gate stack would offer the ultimate scalability of the vertical height of the gate stack. Further, contamination issues associated with the metal FGs would be avoided. Very recently, Gr and graphene oxide (GO) have been tested as a charge storage layer in flash memory devices and promising memory performance is demonstrated [36-41]. Some of these studies utilize Gr or GO as a charge storage layer in conventional flash memory structure [36, 37] while in other studies [38-41], charge storage capability of the Gr, GO or reduced graphene oxide (rGO) is investigated with modified or unconventional flash memory structures.

1.1 Scope of the Present Work

With this discussion, it is evident that to continue the scaling of flash memory devices, material innovations in the memory gate stack and particularly in the charge storage layer are imperative. This thesis mainly focuses on the modification in charge storage layer. The thesis starts with a detailed discussion on the metal nanocrystal formation statistics and their memory behavior. Considering the recent progress in the flash memory research, issues related to the metal nanocrystal memory devices are highlighted and it is concluded that nanocrystal memory technology would not be advantageous for technology node 20 nm or below. In this wake, novel materials like graphene is identified as a replacement for metal NC or thin metal FG as a charge storage layer. Further, work function of the material which is to be used as CSL is an important consideration; therefore, work function of the graphene layers is experimentally determined using it as a gate electrode in MOS test structure.

1.2 Thesis Organization

This thesis contains 7 chapters including the introduction. The thesis is organized as follows:

In chapter 2, evolution of the flash memory over the period of time and its operation mechanism is described in brief. Reliability and performance issues arising due to the miniaturization of the flash devices are discussed. A literature survey on different kinds of charge storage layers in the NAND flash memory, their technological requirements, scaling and challenges in the further scaling are presented in brief.

In chapter 3, material requirements for the nanocrystal memories *viz*. semiconductor NC or metal NC (MNC), work function of the metal, size of the MNC, etc. are discussed followed by experimental results on the MNC formation, their statistics (size, density, coverage, etc.) and memory behaviors. This chapter ends with highlighting the issues and challenges with the MNC memories.

In Chapter 4, the need for thin FG materials, recent literature survey on thin metal FG memories and related issues are discussed. Useful properties of the graphene for charge storage layer and its benefits over metal FG are also discussed. Experimental results on the graphene FG memory with MOSCAP, long channel flash transistor and pseudo source drain (memory gate stack surrounded by implanted regions) are presented.

Chapter 5, for any material intended to be used as a charge storage layer, knowledge of its work function is very important. Therefore, the work function of graphene is experimentally determined using it as a gate electrode in MOS structure. Since graphene is atomically thin, a contact metal is required to connect it to measurement probes. Therefore, graphene WF is calculated under different contact metals like Titanium Nitride (TiN), Platinum (Pt), Iridium (Ir) and Aluminum (Al). The obtained electrical data is verified by physical characterization techniques like transmission electron microscopy (TEM), ultraviolet photoelectron microscopy (UPS), Fourier transform infrared spectroscopy (FTIR) and X-ray photoelectron spectroscopy (XPS).

In chapter 7, reliability and the stability of the work function values of the graphene gate electrode devices at elevated temperatures is experimentally studied. The obtained electrical results are verified by cross section high resolution transmission electron microscopy (X-HRTEM) analysis.

Chapter 8, summaries the thesis with future directions.

Chapter 2

Floating gate flash memory: device structure, working principle and scaling challenges

2.1 Floating Gate Flash Memory Evolution and Current Device Structure

The semiconductor memories can be categorized into two groups: (1) volatile memory i.e. they lose the stored information once the power supply is turned off. The example of this type of memory is Random Access Memory (RAM- DRAM, SRAM) and (2) the non volatile memory i.e. they retain the information in the absence of power supply. The example of this type of memory is Read Only Memory (ROM). The demand for the non volatile memories will be continuously increasing because of the ever increasing market for the portable electronic equipments like mobiles, digital camera, laptops etc. The first metal oxide semiconductor (MOS) based non volatile memory concept, comprising of a metal-insulator-metal-insulator-semiconductor structure, was presented by D. Khang and S. M. Sze in 1967 at Bell Laboratories [42]. The insulator, in direct contact with the semiconductor, was thin enough to allow quantum mechanical tunneling of electrons from the silicon substrate to the FG and vice versa. However, at that time, growth of such thin defect free oxide was not possible [42]. Therefore, a new device structure, known as the floating gate avalanche injection (FAMOS) transistor cell, with a thicker gate oxide and without control gate was proposed by Frohman-Bentchkowsky in 1971 [43]. Such memory cell was known as Erasable Programmable Read Only Memory (EPROM). Programming action in this type of memory was achieved by the avalanche injection of the carriers through oxide to the floating gate at the drain/substrate junction. As these devices did not have any control gate, the operation was very inefficient and the voltage needed to program the cell was very high. The erasure of the memory was achieved by ultraviolet photoemission.

The erase process for this memory was very time consuming as the memory had to be removed and exposed to UV radiation for long time. The FAMOS concept very soon evolved into the double poly-Si stacked gate n-channel cell and now better known as the floating gate memory cell. This became the fundamental cell of the EPROM. The programming of this cell was achieved by injecting channel hot electrons into the floating gate (poly-Si), however, the erasure was still by UV irradiation which was not an efficient technique to erase the cell. Since in-systems electrical erasure was very desirable, a new physics was proposed to achieve the in-system erasure with the same floating gate device but with a select transistor (known as Floating Gate thin oxide memory cell. These memories were programmed by either channel hot electron (CHE) or quantum mechanical tunneling of the carriers through the tunnel oxide and were known as Electrically Erasable Programmable Read Only Memory (EEPROM). The erasure of these memories was achieved by raising the drain voltage to a high value (controlled by select transistor) and control gate to be grounded and thus facilitating the tunneling of the electrons from floating gate to the drain [42]. The limitation of this type of EEPROM was low density due to the select 2 transistor. It was basically a 2 transistor cell. In 1984, the first FLASH EPROM was proposed (known as Flash because whole memory array can be erased at the same time). It was basically an EPROM cell with a possibility to be in-system electrically erased. The first FLASH product was presented in 1988 [44]. FLASH memories contain the features of both EPROM and EEPROM i.e. these can be programmed as well as erased within the system and are better than the EEPROM, in the sense, these are one transistor cell hence can be fabricated with a very high density. The device structure of floating gate flash memory transistor is very much similar to the logic MOS transistors, the only difference between the two is the addition of two extra layers in the flash memory devices; one is known as the floating gate (where the charge is stored) and the other is the blocking dielectric as shown in figure 1.2 of this thesis.

Owing to the high density of the flash devices, the cost of these memories is very low compared to EEPROMs. Figure 2.1 shows the comparison of all memories in terms of flexibility and cost. Flexibility means the possibility to be programmed and erased many times on the system with minimum granularity [44]. Flash memories, with their smaller size and having all the features of EPROMSs and EEPROMs, offer the best compromise between flexibility and cost. The first

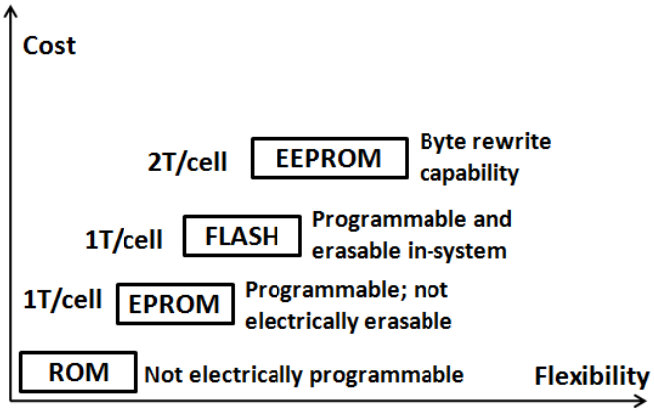


Figure 2.1: Nonvolatile memory (NVM) qualitative comparison in the flexibility-cost plane, adapted from [44].

significant high volume application for the flash memory was as BIOS memory in personal computers [45]. Here the flash memory was used for the code storage purpose (where the program of the operating system is stored and is operated by microprocessor or microcontroller). The other application of the flash memory was the data storage purpose where data files for images, music, video files, etc. can be stored.

So far in the history of flash memories different types of architectures have been proposed. Among all of these architectures, two of them can be considered as the industry standard. First one is the common ground NOR flash cell (known as NOR because it is similar to NOR logic) and is optimized for both code and data storage. The other is NAND flash cell (known as NAND because it is similar to NAND logic) which is better optimized for data storage applications. However, the structure of the single cell is identical in both memory architectures. In this thesis, single cells are mainly characterized for NAND applications. In the initial part of this chapter, a brief overview of FG flash memory evolution, current device structure and operation mechanism is discussed. The later part of the chapter focus on the scaling of the FG flash memory devices, issues arising due to the scaling and possible solutions to continue the further scaling with emphasis on the modification of FG or the charge storage layer is discussed.

2.2 Floating Gate Flash Memory: Operation Mechanism

The operation of the flash memory cell can be described by considering the threshold voltage (V_T) equation of the MOS transistor [46]

$$V_T = K - {}^{Q}/_{C_{OX}} \quad \text{..... (2.1)}$$

where K is a constant and depends on the material of the gate and substrate, gate oxide thickness and on the doping concentration. Q is the charge weighted with respect to its position in the gate oxide and C_{OX} is the gate oxide capacitance. It is evident from equation (2.1) that the threshold voltage of the MOS transistor can be changed by varying the charge between the gate electrode and the channel. There are several ways to achieved this. The most common procedure to have the charge between gate and channel was to store the charge in the traps available in the oxide or at the interface of the two dielectrics. The devices operating on this procedure were known as MNOS (Metal Nitride Oxide Silicon) devices. These were the very first flash type of devices. Here the gate electrode was in direct contact with a nitride dielectric layer and the charge was stored in the traps available at oxide - nitride interface. The limitation of such devices was the leakage of charge to the top gate electrode as the top gate was in direct contact with the charge storing layer. These devices suffered from poor endurance (capability of maintaining the stored charge after every cycle viz. read/erase/program) and retention (retaining the stored charge in time) problems. The other proposed method was to store the charge in a conductive layer between the gate and the channel. In this case, the conductive layer was separated from the gate electrode by a dielectric layer and hence the charge leakage problem to the top gate could be avoided. These types of devices are known as floating gate devices and are now universally accepted for the flash memory applications [46].

The basic floating gate device is shown in figure 2.2 with all capacitances associated with the device. The top gate is known as the control gate and the middle gate, completely surrounded by the dielectric layers, is known as the floating gate. The thickness of the bottom dielectric, known

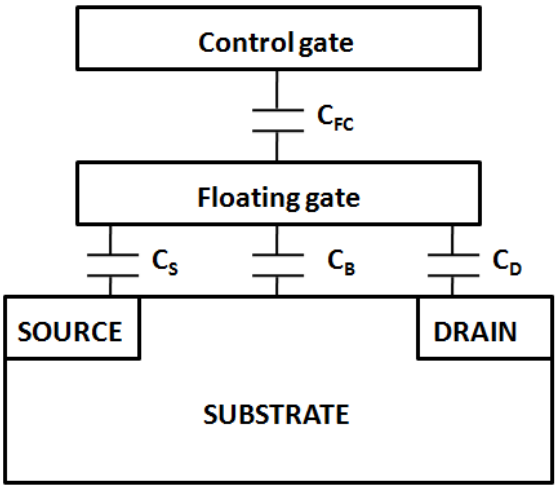


Figure 2.2: Schematic cross-section of a floating gate transistor, adapted from [46].

as the tunnel dielectric, is kept very low (8-10 nm) in order to allow the quantum mechanical tunneling of the carriers from the substrate and vice versa. The top dielectric, known as the blocking dielectric, is kept somewhat thicker (15~20 nm) so as to prevent any charge leakage to the top gate. The floating gate behaves as a potential well, i.e. once the charge is forced into the well, it cannot come out of the well without the help of any external force and thus is capable of storing the charge. The basic operation (read/write/program) of the flash memory can be understood in the following ways.

2.2.1 Read Operation

Reading operation intended to find whether there is charge stored in the floating gate or not. When there is charge in the floating gate, threshold voltage of the device is different compared to the threshold voltage when there is no charge in the floating gate. The difference in the threshold voltage is proportional to the charge stored in the floating gate. Current-voltage (transfer characteristic of the device) characteristic will be different in these two cases as shown in figure 2.3. The shift in the I_D (drain current) vs. V_{CG} (control gate potential) is proportional to the shift in the threshold voltage. When there is no charge in the floating gate, V_T of the device is low (basically un-programmed cell, say logic 1 state) which leads to the higher drain current.

Similarly in the case of charge in the floating gate, V_T of the device is high (programmed cell say logic 0 state), leading to the low drain current

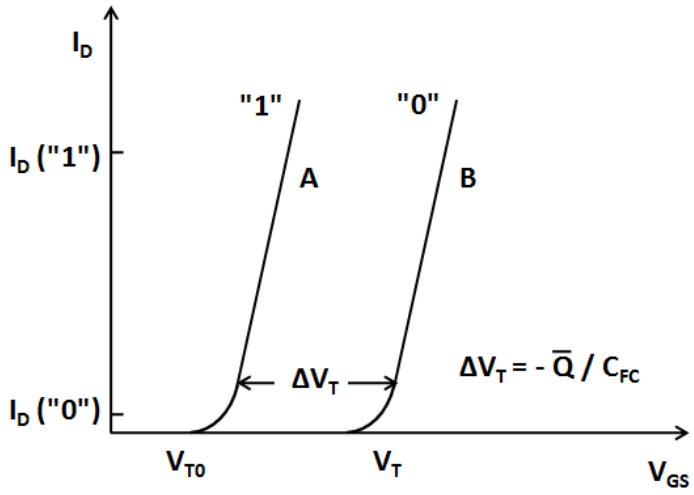


Figure 2.3: IV curve of a FG device when there is no charge stored in the floating gate (A) and when a negative charge Q is stored in the FG (B), adapted from [46].

2.2.2 Programming the cell (Write Operation)

Programming the cell means to inject the charge from the substrate into the floating gate. This can be achieved by two ways: (1) either the carriers should have the sufficient energy to surmount the Si/SiO_2 barrier (3.2 eV for electrons and 3.8 eV for holes) or (2) the barrier must be thin enough to let the carriers to tunnel through it. Based on these two possibilities, two different programming mechanisms for flash devices namely (a) channel hot electron programming and (b) Fowler Nordheim tunneling exist as described below.

A. Channel Hot Electron programming

For this type of programming to happen, there must be very high lateral electric field in the channel of the transistor so that electrons get sufficiently high energy to surmount the Si/SiO_2 barrier. In this situation, these electrons are known as hot electrons. For an electron to overcome the potential barrier, three conditions must hold [46].

(1) The kinetic energy of the electron must be higher than the potential barrier (more than 3.2 eV in case of Si/SiO_2 barrier)

(2) Electrons must be directed toward the barrier.

(3) The field in the oxide must be collecting it.

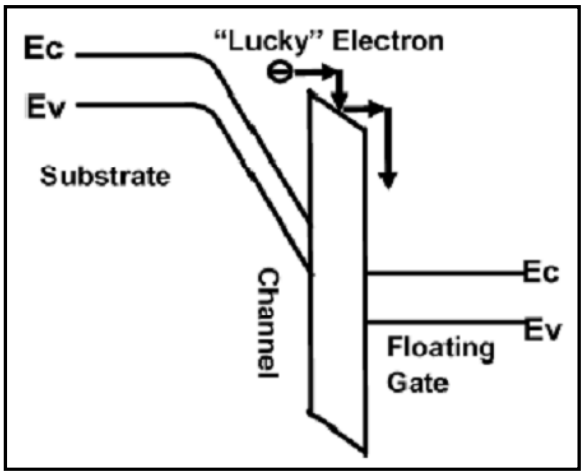


Figure 2.4: Band diagram for hot electron programming [47].

Figure 2.4 shows the band diagram for channel hot electron programming. This type of programming is used in the NOR flash memories. The drawback of the CHE programming is the high field needed in the channel of the transistor to increase the electron kinetic energy, which requires the high gate voltage as well as higher drain voltages. High drain voltage would limit the scaling of the device channel length and hence the memory density as the Si/SiO_2 barrier height is a non scalable quantity.

B. Fowler – Nordheim Tunneling

One of the most important injection mechanisms used in floating gate devices is the Fowler–Nordheim (or FN) tunneling, which is, in fact, a field assisted electron tunneling mechanism. This mechanism can be better understood with the help of band structure of poly-Si (or metal) / SiO_2 /Si structure. When a positive or negative voltage is applied to the poly-Si gate (or metal

gate), bands of these layers bend as shown in figure 2.5 [48]. Now because of the band bending, electrons see a triangular barrier instead of a rectangular barrier. This means that now electrons have to tunnel less distance and the probability of tunneling increases. Using the free electron gas model for the metal and the Wentzel-kramers-Brillouin (WKB) approximation for the tunneling probability, following expression for the current density is obtained [48].

$$J = \frac{q^3 E^2}{8\pi h\phi} \exp\left[\frac{-8\pi(2m)^{1/2}\phi^{3/2}}{3hqE}\right] \quad \ldots\ldots (2.2)$$

Where h is Planck's constant, q = electronic charge, E = electric field, Φ= barrier height and m = free electronic mass.

In equation (2.2), tunneling current depends on the electric field. As the electric field increases, band bending becomes sharper and the tunneling increases. The FN tunneling is widely used in the programming of the NVM, particularly in NAND type of memory. Basic reason for using this mechanism is the low current level required for all operations. The programming time in this type of programming is about 1ms [46].

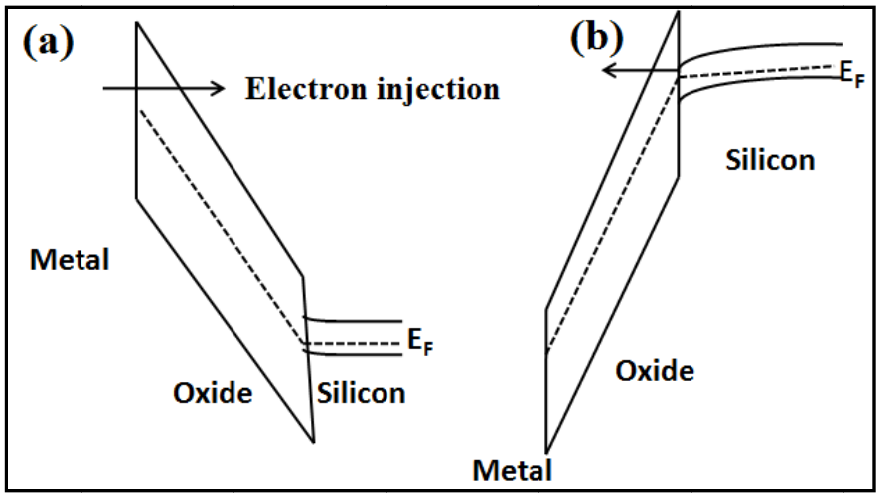


Figure 2.5: Energy band diagram of a metal-silicon dioxide-silicon structure, (a) with large negative bias on the metal electrode, (b) with large positive bias on the metal electrode, adapted from [48].

2.2.3 Erase operation

Electrical erase can also be achieved by two ways: (1) hot hole injection in the floating gate so that injected charge neutralize the electrons already present in the floating gate, (2) FN tunneling of the electrons from the floating gate to the substrate at the source side. Now FN tunneling based erase mechanism is more prevalent in both type of memory architecture (NOR and NAND).

So far in this chapter, the evolution of flash memory devices and their basic operation mechanism is discussed. In the next section current status of the flash memory devices is presented. It is well known that today's high density, high speed, and low cost flash memories are possible only because of the continuous scaling of the flash devices. However, now it seems that the scaling of these devices has reached its limit point. In the following section, challenges in the further scaling and possible solutions to overcome these challenges are discussed.

2.3 Scaling of the Flash Memory

The motivation for the scaling of the flash memory devices is to increase the density and to reduce the cost and the operating voltage. Figure 2.6 shows the scaling trend of the NAND flash memory in recent years [49].

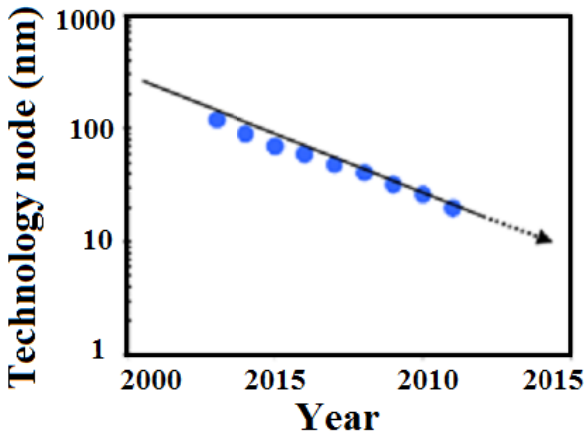


Figure 2.6: Scaling trend of floating gate NAND flash cell over the past 10 years [49].

However, further scaling down to 10 nm node would face several difficult scaling challenges [49-52]. Major concerns are noted as below:

A. Challenges in physical scaling
B. Disproportionate scaling of physical dimensions and electrical biases
C. Cell to cell interference
D. Fewer electron storage

These issues are briefly discussed below.

A. Challenges in physical scaling

Horizontal and vertical scaling of the flash cell could not be achieved in the same proportion. Laterally devices are coming close to each other, however, vertical dimensions of the flash gate stack could not be reduced substantially. Figure 2.7 shows the scaling of flash memory cell in horizontal and vertical directions. Scaling of the tunnel oxide as well as of blocking oxide is essential to achieve the low voltage operation of the flash devices. As the thickness of the tunnel

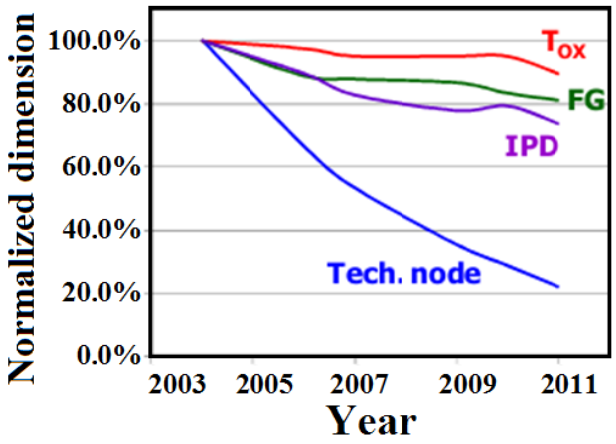


Figure 2.7: Scaling trend of floating gate NAND flash cell in the recent years. Here T_{OX} is the tunnel oxide, FG is the floating gate, IPD is the interpoly dielectric (blocking dielectric) and Tech. node is the technology node. The figure clearly demonstrates the unbalanced scaling of the flash cell in horizontal and vertical direction [49].

dielectric decreases, the interface between silicon and the silicon-dioxide plays a very important role in retaining the charge for the long time. The required tunnel dielectric thickness is reaching below 8 nm. However, for very thin SiO_2, it is theoretically impossible to have a defect free oxide as at any temperature there is an equilibrium amount of point defects [53]. Even if a single defect exists in the dielectric, the whole charge stored in the floating gate (which is a conductive layer) can leak through the tunnel oxide. Further for very thin tunnel oxide, stress induced leakage current becomes very severe. The defects generated when high voltage operations are performed on the thin oxide, are responsible for this current. Stress induced leakage current (SILC) puts the fundamental limitation on the scaling of the tunnel dielectric. Similar restriction is on the minimum thickness of the blocking dielectric. For ultra thin blocking dielectric, charge leaking through the blocking dielectric is a serious reliability issue. Due to these reliability concerns, tunnel oxide and blocking dielectric could not be scaled much. Thinner tunnel oxide would degrade the charge retention while very thin blocking dielectric cannot perform the blocking action efficiently.

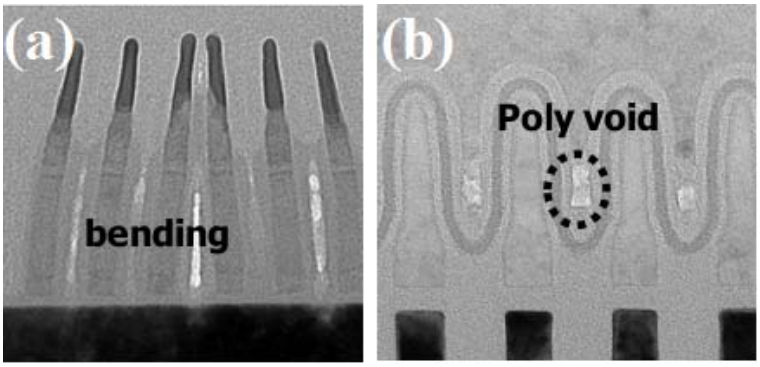


Figure 2.8: (a) NAND structure in bit line (BL) to bit line direction. Narrow and tall gate stack is susceptible to bend. (b) NAND structure in word line (WL) direction. Voids in between the FGs can be seen [49].

Unbalanced scaling in vertical and horizontal direction causes tall and narrow gate stack susceptible to bend as shown in figure 2.8(a) [49]. Further, due to reduced horizontal space between the devices, blocking dielectric and control gate cannot fill the space between the adjacent floating gates which causes the voids in space between the FGs as shown in figure

2.8(b) [49]. These voids would cause the loss of overlap area and hence the loss of CG to FG coupling ratio to fall below desirable value of 0.6. When blocking dielectric and control gate wraps around the floating gate, the covering of the side walls of the FG by the blocking dielectric and CG provides the extra side wall capacitance which keeps the coupling ratio to about 0.6. The high value of coupling ratio is desirable to have the maximum of control gate voltage to drop across the tunnel oxide.

This unbalanced geometrical scaling is causing the transition of wrapped around FG structure to planar FG structure where blocking dielectric and CG no longer cover the side walls of FG as shown in figure 2.9 [52]. As a result, effective coupling of the control gate to the floating gate reduces. Therefore better coupling of the control gate to the floating gate demands a thinner blocking dielectric. However, very thin blocking dielectric cannot block the stored charge from leaking to the top gate and the electron injection from the CG during erase operation. To achieve the high gate coupling ratio and good blocking action at the same time, a physically thicker but an electrically thin dielectric film is required. A high dielectric constant (high-K) material is required to serve this purpose [51, 52, 54]. The high-K material intended to be used as a blocking dielectric should be trap free and should provide sufficiently high electron barrier during program and erase operations.

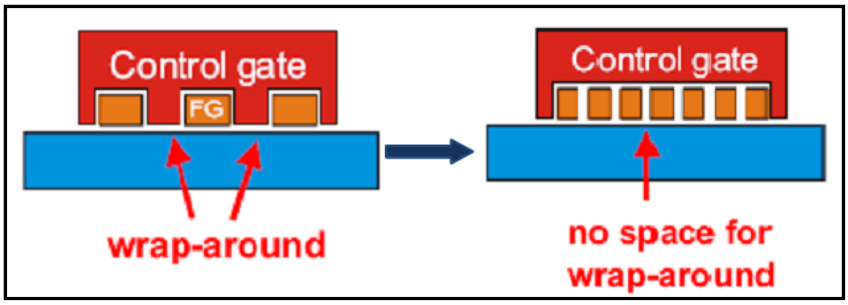


Figure 2.9: Effect of the scaling on the wrapping of control gate to floating gate [52].

B. Disproportional scaling of physical dimensions and electrical biases

When the physical dimensions of the devices are scaled down, it is imperative to scale the operating voltage of the devices in the same proportion. However, this is not happening in the

case of NAND flash cell due to unscalability of tunnel and blocking dielectric. Horizontal dimensions of the flash cell has been scaled aggressively, however, as discussed in the previous section, vertical dimensions could not be scaled accordingly. The program and erase voltage of flash cell is about ~ 15 - 20 V. At 20 nm node (WL spacing ~ 20 nm), the electric field between selected and unselected cell is about 10 MV/cm which is close to the breakdown field of the dielectrics. The high electric field between the world lines and neighboring floating gates also causes threshold voltage reduction of adjacent word line when other word line is being programmed. This is known as program interference [55]. Use of air gap between the neighboring WLs is proposed to increase the breakdown voltage and to improve the lateral charge loss [49, 50].

C. Cell to cell interference

As the cell to cell distance shrinks, one of the major concerns is the increasing cross talk or interference among the neighboring WLs and BLs. This is due to the increased capacitive coupling among the FG's of neighboring cells. This causes a wide distribution in the threshold voltage of the devices. Figure 2.10 shows the increase in the cell to cell inference as the technology shrinks.

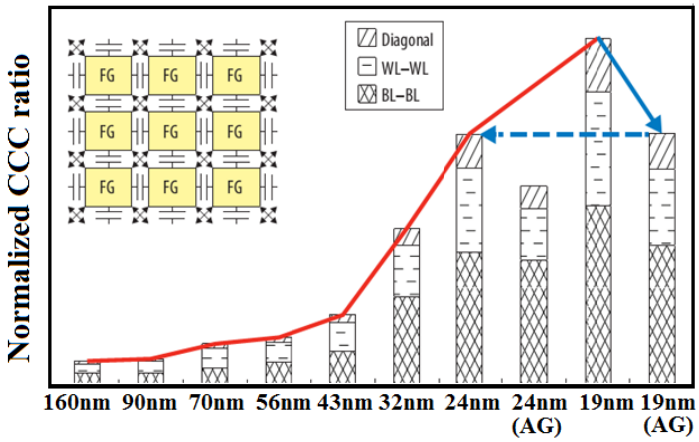


Figure 2.10: Cell to cell interference; diagonal, WL to WL (WL-WL) and BL to BL (BI-BL). Graph also shows that having air gap (AG) between the WL's can reduce the cell to cell interference [56].

Having air gap between the WLs can help to reduce the WL to WL interference [50]. Similarly, filling the shallow trench isolation (STI) with air gap can counter the BL to BL interference [49]. Infact, in a scaled cell, air gap should be used in all possible areas to reduce the cell to cell interference. However, the amount of improvement depends on the air gap shape and height [50]. Reduction of the floating gate height is also proposed to reduce the FG to FG interference [25].

D. Few electron storage

As the NAND cell size shrinks, number of stored electrons decrease significantly. Figure 2.11 shows the number of stored electrons required for 100 mV threshold voltage shift with decrease in feature size [57]. For 20 nm node, the number of stored electrons would be close to the statistical fluctuation limit. A small change in the number of electrons would cause a significant change in the threshold voltage of the device. This would be an important issue for extremely scaled flash cell; however, any viable solution could not be suggested so far to tackle this [51].

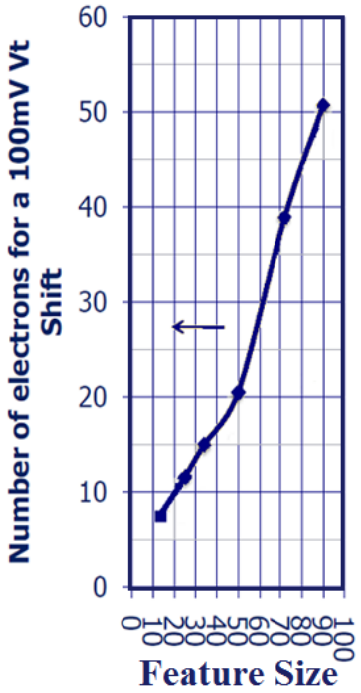


Figure 2.11: Number of electrons required for 100 mV threshold voltage shift with decreasing technology node [57]. Feature size is in nanometers.

In summary, the motivation for the scaling and challenges are:

A. Scaling of the tunnel dielectric (for faster and low operating voltages); limited by the charge retention requirements.

B. Scaling of the blocking dielectric (for better coupling of the gate voltage to the channel; basically for low voltage operation); limited by the leakage of charge from floating gate to the control gate.

C. Lateral scaling: limited by the increasing cell to cell interference and other reliability issues like non scalability of the electrical biases.

Impact of some of these issues on the memory performance can be reduced by engineering the floating gate layer. Next, some possible modifications in the floating gate layers are discussed.

2.4 Replacement of floating gate for further scaling

The first proposed flash cell had a conducting poly-Si layer as a charge storage node. This conducting layer worked well till the need for highly scalable device was felt. Now in the wake of scaling issues discussed above, it appears that if this conducting floating gate is replaced by any other charge storage layer, then some of the above mentioned issues can be addressed. The major drawbacks with conducting layer as a charge storage node are (1) complete loss of charge even if a single defect exists in the tunnel or blocking dielectric, (2) coupling of the two neighboring floating gates. Both of these issues can be resolved to some extent by replacing the conducting charge storage layer by discrete charge storage layer. The two apparent benefits of this approach are; (1) with the discrete charge storage nodes, complete charge loss because of defects in the tunnel or blocking oxide is avoided and hence it is less vulnerable to the defects in the dielectric. It would enhance the scalability of the tunnel and blocking dielectric. (2) With discrete charge storage nodes, as the charge is localized and hence cross talk among the neighboring cells FG is much reduced suggesting that devices can be more compactly packed. This discussion suggests that the successful further scaling essentially requires the modifications

in the FG or charge storage layer. Some modifications in the FG proposed in the recent time namely (1) charge trap flash memory (2) nanocrystal flash memory, are discussed in the following sections.

2.4.1 Charge trap flash memories (CTF)

In this scheme, the charge is stored in the traps (which are by nature discrete) available in the dielectric layer. This type of memory is known as the charge trap flash (CTF) memory. Continuous floating gate is replaced by some suitable dielectric having high density of traps. Several dielectrics (HfO_2, Al_2O_3, and HfAlO) have been explored for this purpose [6-14], however, silicon nitride (Si_3N_4) has shown the most promising results so far [6-8, 10, 51]. Si_3N_4 CTF is known as the SONOS (Silicon-Oxide-Nitride-Oxide-Silicon, where the poly-Si is used as a top gate electrode) or TANOS (where some metal gate electrode, mostly TaN, and Al_2O_3 as a blocking dielectric is used). Significant work has been done to explore the Si_3N_4 as a charge trap layer and it has indeed shown promising results. The most important aspect that goes in favor of the charge trap based memories is their CMOS compatible process flow and reduced FG-FG interference [6-8, 10, 51]. For CTF memories, only poly-Si floating gate needs to be replaced by some suitable charge trap layer, hence all the process flow for the existing floating gate flash can be adopted for CTF. CTF can be a possible candidate to replace the floating gate flash memory below 20 nm technology node [51]. However, CTF memories have some reliability issues which need to be addressed. The performance of CTF memory strongly depends on the chemical composition of the CTF layer. It has been demonstrated that Si rich nitride film has good erase state retention but poor program state retention [8, 58]. On the other hand, N rich nitride layer has good program state retention but it worsens the erase state retention [8, 58]. The quality of the SiN layer strongly depends on the process parameter. A slight change in the process conditions may change the composition of the nitride film and hence the memory performance. CTF memories with HfO_2 and Al_2O_3 charge trap layers suffer from the issues like poor charge retention and slower programming speed respectively [14]. Further, decreasing number of stored electrons cannot be addressed by CTF memories [51]. In summary, CTF provides planar structure which is more scalable than FG NAND flash and enhanced immunity against the

charge loss through defects in the surrounding dielectrics and FG-FG interference. However, the issues like compositional variation of CTF layer, few electron storage and WL breakdown should be resolved to make it viable. Further discussion on the CTF memories is beyond the scope of this thesis.

2.4.2 Nanocrystal flash memories

Storing the charge in nanocrystals (NCs) in place of traps in the dielectric is another approach to provide the discrete charge storage node for flash devices. The idea of storing the charge in nanocrystals was first presented in 1995 by Tiwari et al. [59]. Tiwari and his group suggested the semiconductor nanocrystals for flash memories. Semiconductor nanocrystals were the first choice for the flash application because of the proper understanding of these materials and their compatibility with standard CMOS fabrication flow. Two most important semiconductors widely available, i.e. Si and Ge were tried for this purpose [59 - 62]. However, semiconductor NCs suffer from inherent material limitations like limited size scalability due to Coulomb blockade and the quantum confinement effect [63]. Because of the Coulomb blockade, electrostatic potential of the nanocrystal increases and hence it hinders the entry of another electron in the well. Because of the quantum confinement, energy levels of the nanocrystal move upwards and as a result, the band offset between the nanocrystal and the surrounding dielectric reduces which affects the retention characteristic of the memory. It is suggested by Liu et.al [64] that the Coulomb blockade and quantum confinement will not be as severe in the metal NCs as it is in semiconductor NCs. Further, because of the wide availability of the metals with different work functions, it is rather easy to tune the potential well depth formed by the metal nanocrystals. It has been shown that energy level up shifts with decreasing size is more pronounced in semiconductor NCs rather than metal NCs. Figure 2.12 shows the conduction band up shifts for a semiconductor NC and Fermi level up shift for a metal NC as a function of NC size [63]. Figure 2.12 shows that the up shift in the Fermi level of metal NC is less compared to that in a semiconductor NC. Hence it is concluded that quantum confinement effect is less pronounced in metal NCs until very small size NCs (form the figure, a limit of 1.5 nm can be taken) are formed. It is demonstrated in [63] that for the same NC size, metal NC memory shows better retention

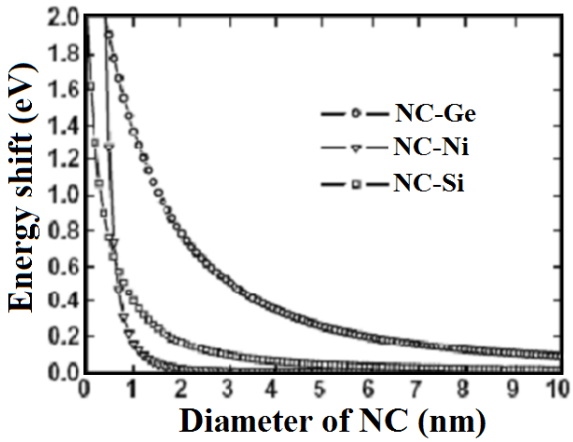


Figure 2.12: Energy band up shift in SNC and MNC of different sizes [63].

performance compared to the semiconductor NC memory which again signifies that quantum confinement is less severe in metal NCs. Further, MNC memories are reported to be faster than semiconductor NC memories. In brief, MNCs would have following advantages over their semiconductor counterparts.

(1) A wide range of available work function in the case of MNCs, which provides one extra degree of freedom to tune the write/erase and charge retention characteristics of memory devices.
(2) High density of states available in the metal NCs, which provides the stronger coupling with the conduction channel.
(3) Metal NCs suffer from smaller energy perturbation around the Fermi level caused by quantum confinement which results in better retention performance.
(4) Metal NCs offer enhanced size scalability. MNCs can be scaled down to 2 nm size without being influenced by quantum effects which is very important for ultra scaled flash memory devices.
(5) Due to high density of energy states and asymmetrically enhanced electric field, metal NCs offer better program efficiency [64].

Metal nanocrystals for memory application are discussed in detail in the next section.

2.4.3 Metal Nanocrystals

Metals of different work functions are readily available. It is, therefore, very easy in case of metal NCs to tune the programming and retention characteristic of memory devices. There are several ways by which metal NCs can be fabricated. Some of them are colloidal, aerosol, and self assembly method [66]. In colloidal method, NCs are precipitated from the chemical solution of the metal to be deposited, however, the contaminations associated with the chemical solution are the major concern with this method. In the aerosol method, NCs are formed in the gas phase condensation and falls on the substrate. Particle delivery and non uniformity in the size of the NCs are the concerns with this method. Self assembly method is the easiest and most commonly used method for the NC formation. In this method a very thin metal film is deposited on the tunnel oxide and subsequently annealed at a suitably high temperature. Size of the NCs can be controlled by optimizing the initial metal thickness, anneal temperature and anneal time.

2.4.3.1 Effect of different parameters on the NC formation

A. Effect of initial metal thickness

Initial metal thickness is the most important parameter to control the size and density of the NCs. Thinner the initial metal film, smaller the NCs can be obtained as a consequence of annealing [67]. As the thickness of the initial metal film increases, size of the nanocrystal increases and the density decreases. For the thicker film, nanocrystals grow in size and form larger agglomerates. Hence, for smaller size NCs and large density which is desirable for the ultra scaled flash memory devices, initial metal thickness should be very low. This can be achieved by reducing the deposition rate of the metal. In the deposition process, metal atoms deposit in the form of small nuclei during the very initial phase of deposition. Therefore, if the metal deposition is restricted in this regime, very little temperature spike can results in very small nanocrystal with very large density.

B. Effect of anneal temperature

Anneal temperature is another important parameter which affects the NC distribution very significantly. In the process of nanocrystal formation, the non-equilibrium state clusters reshape, attempting to obtain a local minimum energy state. The anneal temperature provides them the energy to obtain local minimum energy state. At low temperatures (~ 450°C), metal adatoms just starts moving and forms the small nuclei. As the temperature increases, more and more adatoms join each other and forms the large nuclei. As the temperature continues to increase, more and more nuclei merge and form larger nanocrystals. Hence, with increasing temperature, size of the NCs increases and the density decreases [67].

C. Effect of Anneal time

For the optimum size of the nanocrystals, anneal time depends on the anneal temperature. If the anneal temperature is very high, anneal time can be reduced and vice versa. If the annealing is performed for longer time at high temperature, density of the NCs will reduce at the cost of increased size.

D. Effect of metal melting point

Melting point of the metal also affects the nanocrystal size. Higher melting point metals form the smaller NCs as compared to the NCs formed by low melting point metal [66, 68]. For example, in case of Ag, Au and Pt metals, because of its highest melting point, Pt would form the smallest nanocrystals than Au and Ag.

2.4.3.2 Metal NC flash memory devices

So far, different metal nanocrystals like Au, Pt, W, Ni, Co, and TiS_2 [69-77] have been evaluated as a charge storage node in flash memory structure. Nanocrystals of high work function metals like Pt, Au, Co, Ni offer the deeper potential well thereby better charge retention and program efficiency. However, deeper potential well degrades the erase speed due to the increased

potential barrier between the Fermi level of the metal NC and the conduction band of the substrate [78]. Figure 2.13 depicts the band diagram of the MNC embedded between tunnel and blocking dielectric during programming and retention [23]. Figure 2.14 shows the work functions of various nanocrystal materials with respect to Si and Ge band gap [23]. To achieve the efficient program and erase at the same time, a combination of high and low work function metal NCs are proposed in the form of a dual layer memory stack [78]. Here, four combinations of nickel and gold (Ni/Ni, Au/Au, Ni/Au, and Au/Ni) were used for the top and/or bottom metal as charge storage materials. The authors of this paper concluded that top metal NCs govern the program efficiency and bottom metal NCs govern the erase efficiency. They also concluded that a metal of high WF on the top while a low WF on the bottom offers the best memory performance. This is an important conclusion in terms of optimizing the overall nanocrystal memory performance. Table 2.1 lists the memory window obtained with various metal nanocrystal memories reported in different studies.

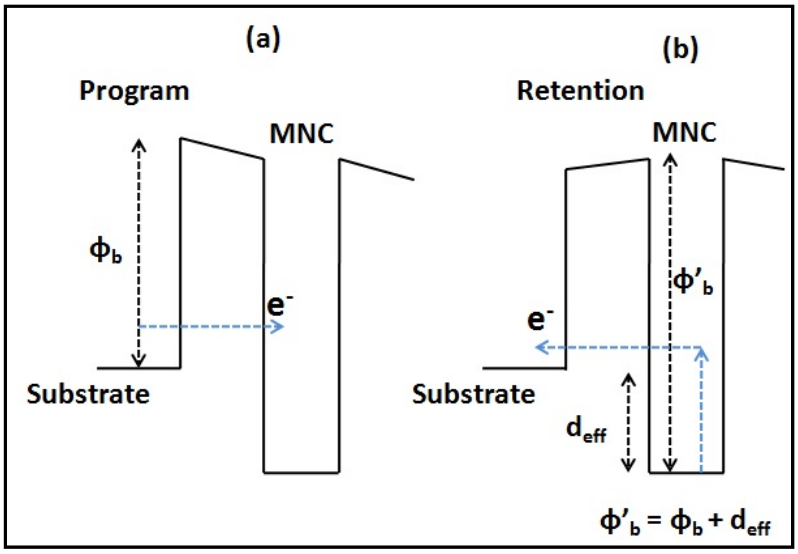


Figure 2.13: Energy band diagram for MNC embedded between a control oxide and tunnel oxide during (a) programming and (b) retention, adapted from [23].

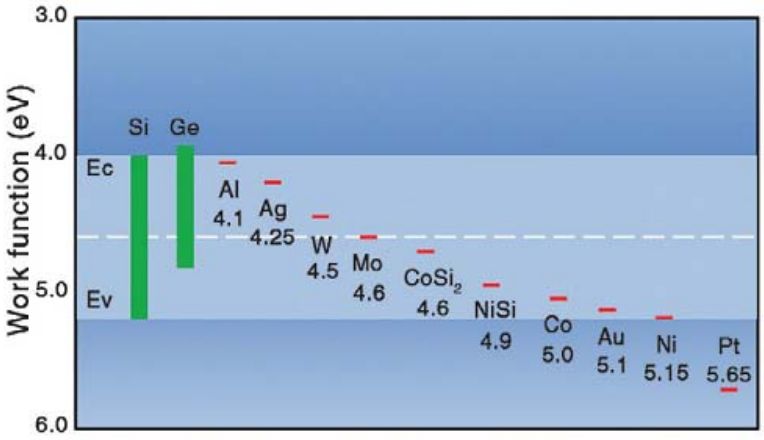


Figure 2.14: Work functions of various nanocrystal materials with respect to Si and Ge band gap [23].

Though the metal nanocrystal technology owns several advantages, there are technical issues associated with it. The memory performance is affected by nanocrystal size, density and distribution [21, 22]. As the most preferable method of nanocrystal formation is self assembly, nanocrystal distribution cannot be controlled precisely in this method. Infact, the successful scaling of floating gate flash cell towards 20 nm node has put even more stringent requirements on the nanocrystal distribution. Variation in the nanocrystal distribution would cause the device to device variability [22, 23] which would be even more serious with technology scaling where any small fluctuation in number of stored electrons would cause large change in threshold voltage. In order to reduce the cell to cell variability, a very high density of NCs (~ 5 x 10^{12} cm^{-2}) is required. However, with such high density of NCs, the spacing between the NC to NC decreases significantly (~ 2 - 4 nm apart). With such closely packed density, though they may remain physically separated, electrically they behave like continuous metal floating gate [24]. Hence the proposed benefits of isolated charge storage nodes may not be achieved with this technology. Further, charge redistribution among the nanocrystals also causes the threshold voltage instability [24].

Table 2.1: Memory window obtained from various nanocrystals in literature.

Nanocrystal material	Tunnel dielectric	Voltage Sweep	Memory Window	Reference
Si	SiO_2	9.5 V to -9 V	3.5 V	[79]
Ge	SiO_2	16 V to -16 V	6 V	[80]
Au	SiO_2	18 V to -19 V	7 V	[69]
Ni	SiO_2	10 V to – 10 V	4 V	[81]
	SiO_2	4 V	1V(Program window)	[73]
	SiO_2	8 V to – 8 V	0.5 V	[74]
W	HfAlO	4 V to – 4V	5.7 V	[72]
	SiO_2	3V to – 4 V	0.95 V	[19]
WSi_2	$Al_2O_3/HfO_2/Al_2O_3$	9 V to – 9 V	2.5 V	[82]
Ag	SiO_2	2 V to – 4 V	2.1 V	[19]
Pt	SiO_2	20 V to – 20 V	10 V	[72]
	SiO_2	20 V to – 20 V	15 V (dual layer Pt NC)	[72]
Co	SiO_2	7 V to – 7 V	1.8 V	[76]
Mo	SiO_2	9 V to –11 V	3.6 V	[83]
IrOx	SiO_2/Al_2O_3	8 V to – 8 V	7.2 V	[84]
Ru	SiO_2	11 V to – 11 V	5.5 V	[85]
TiN	Al_2O_3	16 V to – 15 V	2.3 V	[23]
Pd	SiO_2/HfO_2	17 V to – 17 V	6 V	[86]
HfO2	SiO_2	Vg 9 V, Vd 9 V	2.2 V	[87]
HfSixOy	SiO_2	15 V to –15 V (with Vd 10 V)	~ 6.5 V	[88]

2.4.4 Thin Metal as a floating gate

Increased parasitic capacitive coupling among the floating gates of adjacent cells due to aggressive lateral scaling is one of the foremost challenges that need to be addressed [5]. This undesirable phenomenon causes a wide distribution in threshold voltage of the devices [5]. Floating gate height reduction is one possible way to reduce the capacitive coupling [5]. Recently it has been demonstrated that conventional polycrystalline silicon (poly-Si) floating gate thickness can be reduced to 7 nm [25]. However, a significant fraction of electrons injected into such thin poly-Si would be ballistically transported till they reach the blocking dielectric, which would result in slower programming [25]. These carriers may also cause impact ionization in the blocking dielectric and thus degrade the dielectric quality. To alleviate this problem with thin poly-Si layer, use of continuous metal in place of thin poly-Si is proposed and a 1 nm thick metal layer as a floating gate material is found to be capable of suppressing the ballistic current component [25, 26]. High work function metals as a floating gate improves the program and retention, however, at the cost of degraded erase. To achieve efficient program and erase at the same time, a combination of thin metal films and poly-Si known as hybrid floating gate is proposed [27- 30, 54]. In this hybrid floating gate structure, a low WF poly-Si layer is deposited first on tunnel dielectric followed by high WF metal. Very recently, a hybrid floating gate as thin as 4 nm with about 8 V memory window at 22 V program/erase voltage is demonstrated [30]. Further, by the same group, thin hybrid floating gate with different high-K materials as a integrate dielectric is reported [54].

Currently, thin metal floating gate technology is under intense investigation and offers some advantages over charge trap and nanocrystal memory technology. However, this technology would also suffer from charge leakage through defects in tunnel and blocking dielectric and hence, would not offer any advantage as far as scalability of tunnel and blocking dielectric is concerned. Further, metal layer as a FG may impose its own device reliability issues at high temperatures such as, agglomeration of thin metal layers, diffusion of metal into the tunnel and blocking dielectrics [31] and increased leakage current through the blocking dielectric deposited on metal films due to higher degree of dielectric crystallization at elevated temperatures [32].

2.4.5 Graphene charge storage layer

Very recently, graphene based charge storage layers are investigated in conventional or modified flash memory devices [36-41]. Graphene, owing to its ultrathin nature, offers the ultimate scalability of the FG. Further, the issues like thermal stability and contaminations in the surrounding dielectrics would also be avoided with graphene FG devices. In [36], GO and in [37], Gr is integrated in conventional FG flash memory structure. In [36], with GO as a CSL, a hysteresis of 7.5 V at ± 14 V voltage sweep and in [37] with Gr as CSL, a hysteresis of 6 V in C–V characteristics at ± 7 V voltage sweep are reported. Program/erase (P/E) transients are not discussed in these studies [36, 37]. P/E transients are important to estimate the speed of the flash memory device.

Properties of the graphene like the number of layers, defects present in it, etc. would strongly affect the memory performance of these devices. The work on the graphene as a CSL is in its infancy and significant progress need to be done before qualifying it as a CSL.

2.5 Summary of the chapter

In this chapter, evolution of flash memory, basic operation mechanism and challenges for further scaling were discussed in brief. Modifications in the charge storage layer seem to be inevitable to continue further scaling. In this regards, recent developments in modification of charge storage layer i.e. charge trap flash memory, nanocrystal flash memory, thin metal floating gate memory and graphene FG memory with benefits and issues related to each technology were discussed.

This thesis mainly focuses on the modifications in the charge storage layer to facilitate further scaling. In this direction, the next chapter 3 is based on the metal nanocrystal formation processes, impact of different process parameters on the NC statistics and memory behavior of Pt nanocrystals. The technical challenges associated with the metal NC memories are also discussed in chapter 3.

Chapter 3

Metal nanocrystal flash memory: Nanocrystal formation, their statistics and memory behavior

As discussed in chapter 2 of this thesis, modifications in the charge storage layer can solve some of the issues due to scaling of the memory cell. Breaking the continuity of the floating gate (or the charge storage layer) in the flash memory is one possible solution to improve the retention characteristics to reduce the cell to cell interference of the memory devices. With the discrete charge storage nodes, complete charge loss through the defects in the tunnel or blocking oxide can be avoided which will also enhance the scalability of the tunnel oxide. Further, in this scheme, the stored charge would be localized, hence cross talk among the neighboring cells would be reduced. This would further enhance the packing density of the devices. There are two ways by which charge can be stored in isolated nodes; one is charge trap flash memory and the other is nanocrystal flash memory. In this chapter, metal nanocrystal formation process, different parameters affecting the nanocrystal statistics (size, density and area coverage) and memory results of Pt NCs memory are discussed. Chapter closes with the technical challenges NC memory faces when devices dimensions reduce below 20 nm node.

The results presented in this chapter were reported at the Materials Research Society Fall Meeting 2010 held in Boston. A part of the work presented was done in collaboration with Mr. Sunny Sadana, (M.Tech student in microelectronics, IIT Bombay).

3.1 Metal Nanocrystals: Pt NC formation process and statistical analysis

In the chapter 2 of this thesis, material requirements for the nanocrystals used for charge storage purpose were discussed and it was concluded that metal NCs are more effective for this

particular application compared to their semiconductor counterparts. Further, it is also argued that high work function metals like Pt, Ir, Ni, etc. would be more effective for memory application. Platinum has a very high work function (5.65 eV in vacuum and 5.3 eV on SiO_2 and melting point of 1772°C). High wok function leads to the deeper potential well for charge storage and hence provides excellent memory retention characteristic. Because of its high melting point, it would form very small size nanocrystals compared to other metals like Ag and Au. Nanocrystal distribution statistics can be controlled by controlling a number of parameters like: initial metal thickness, anneal time and anneal temperature. Further, underlying dielectric on which NCs are to be formed would also affect the NC distribution. Impact of all these parameters (initial metal thickness, anneal time, underlying dielectric) are discussed in this chapter. NC statistics is also compared between Pt and Ir metal. Pt and Ir both have high WF, however, the melting point of Ir is much higher than that of Pt, and hence Ir should form lower size NCs as compared to the size of Pt NCs. In the following sections, experimental details with result and analysis on the Pt NC formation are discussed.

3.1.2 Experimental details

Pt nanocrystals were fabricated on a 7 nm SiO_2 layer grown on 2" p-type Si wafer of resistivity 4-7 Ω-cm. Si wafer was first cleaned by standard RCA process and finally dipped in 2% HF to remove any native oxide just before the oxidation. 7 nm of SiO_2 was grown on this Si wafer in a rapid thermal processing (RTP) system at 1050°C in oxygen diluted by nitrogen ambient. To form the Pt NCs, Pt metal was deposited by physical vapor sputtering system under different conditions for different process splits. Deposition conditions (deposition power and deposition time) were optimized to obtain very thin film (~ 2 nm) of platinum at sufficiently reduced deposition power. It was found that Pt deposition rate can be reduced almost linearly with deposition power as shown in figure 3.1.

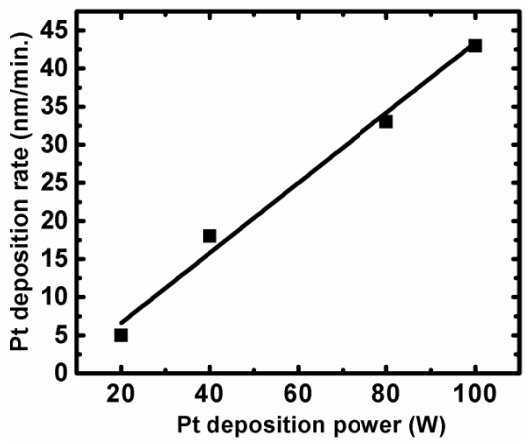


Figure 3.1: Platinum deposition rate as a function of deposition power.

To find the NC formation dependence on process conditions, three parameters *viz.* initial Pt thickness, effect of anneal time and effect of substrate material were studied. For these experiments, anneal temperature was fixed at 550°C. The aim was to form the NCs at as low temperature as possible, and it was found that for platinum metal, NCs form at temperature about 450°C, however, the shape of the NCs was not well defined, hence, a slightly higher temperature of 550°C was chosen. It was also found that as the anneal temperature was increased up to 900 °C, smaller size NCs were formed.

Nanocrystals were analyzed by secondary electron microscopy analysis using Raith 150Two system. Various statistical parameters of the nanocrystals were estimated by analyzing the images using custom made software developed using MATLAB software. For the diameter of the NCs, it was assumed that the top view of the nanocrystals is circular, in case of deviation from the circularity, major axis of the NC was assumed as the diameter. The diameter data is fitted to a Gaussian distribution and standard deviation (SD) from the average value was noted.

3.1.3 Results and analysis

3.1.3.1 Effect of initial metal thickness

Pt metal was deposited by PVD method in a sputtering system. Pt deposition rate was found to be about 15 (± 2) nm per minute at 40 watt of sputtering power in Ar plasma. For very first experiment, Pt was deposited for 15 sec. and 10 sec. at 40 watt of sputtering power. For these deposition times, expected thickness of Pt film was around 3.5 nm and 2.5 nm. The samples were annealed in N_2 ambient at 450°C, 550°C and 700°C for 1 minute. The maximum NC areal density achieved from these experiments was for 10 sec. Pt deposition and annealed at 550°C for 1 minute. The density was 4.85 x 10^{11} cm^{-2} with average size (SD) of 8 (± 5) nm and area coverage 24%. The density (size) obtained from this experiment was very low (high) which was unacceptable for scaled flash memory devices [21]. The reason for this low density and large NC size was attributed to the thick initial Pt film.

Pt deposition rate was further reduced to about 5 nm/min at 20 W deposition power. Below this deposition power (at 10W), the plasma was not sustainable. For the next experiment, Pt was deposited at 20 watt for four different times (15 sec., 30 sec., 45 sec. and 60 sec.). All these four samples were annealed at 550°C for 1 minute in N_2 ambient. Deposition conditions of different samples are listed in table 3.1. Figure 3.2 shows the SEM images of these four samples after annealing and figure 3.3 shows the NCs size, diameter and area coverage.

Table 3.1 Process conditions for the experiment to study the effect of initial Pt thickness on the NC formation.

Sample No.	Deposition Power (W)	Pt Deposition time (sec.)	Anneal Time (sec.)	Anneal Temperature (°C)
S1	20	60	60	550
S2	20	45	60	550
S3	20	30	60	550
S4	20	15	60	550

Well separated nanocrystals are visible in the SEM images of figure 3.2(a-d). The NCs for higher deposition time (i.e. 60 sec., 45 sec. and 30 sec., figure 3.2(a-c)) could be resolved very clearly in the SEM imaging. However, the NCs for 15 sec. Pt deposition could not be resolved clearly (figure 3.2(d)). Vaguely tiny NCs are visible in figure 3.2(d). Hence, NCs size, density and area coverage was calculated using the images of figure 3.1(a-d). Figure 3.3(a-c), shows the variation of size, density, and area coverage of the NCs with initial Pt thickness respectively. From these experiments, maximum NC density of 9.46 x 10^{11} cm^{-2} with average size (SD) of 4.96 (± 3) nm and area coverage of 20% for sample S3 (see table 3.1 for sample S3 deposition conditions) was achieved. These numbers are very close to the proposed value of density 1 x 10^{12} cm^{-2} and NC size 3 nm. However, the value of the area coverage is below the desirable value reported in the literature which is about 25% to 36% depending on the dielectric in which NCs are embedded [21].

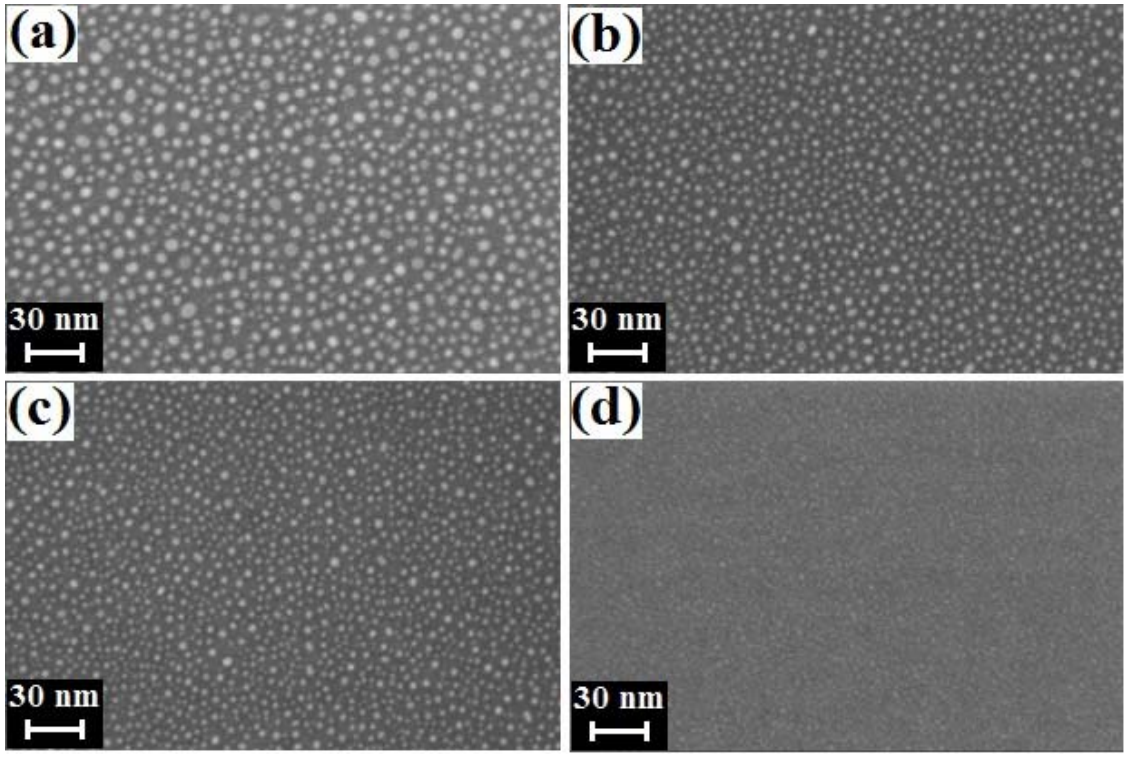


Figure 3.2: SEM images of Pt NCs for 4 different Pt deposition times; (a) 60 sec., (b) 45 sec., (c) 30 sec., (d) 15 sec. All the samples were annealed at a temperature of 550°C for 60 sec.

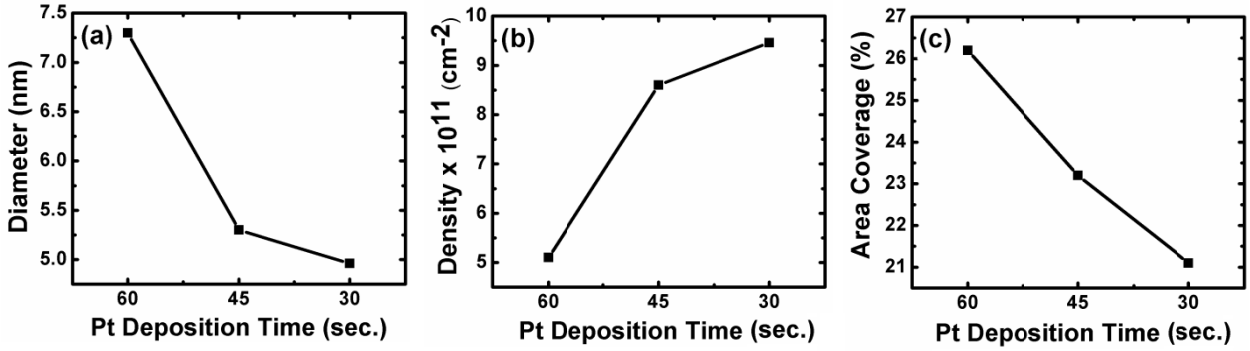


Figure 3.3: Pt NC statistics with varying Pt thickness; (a) Diameter, (b) Density and (C) area coverage.

3.1.3.2 Effect of anneal time at a fixed anneal temperature

From the previous experiment (section 3.1.3.1), it was concluded that Pt deposition at 20 Watt for 30 sec. yields the minimum size NCs with maximum density. Hence, to study the effect of anneal time at a fixed anneal temperature, the Pt was deposited at 20 Watt for 30 sec. The process conditions for these experiments are given in table 3.2.

Table 3.2: Process conditions for the experiment to study the effect of variation in anneal time on the NC formation.

Sample No.	Deposition Power (W)	Pt Deposition time (sec.)	Anneal Time (sec.)	Anneal Temperature (°C)
S1	20	30	60	550
S2	20	30	45	550
S3	20	30	30	550
S4	20	30	15	550

Figure 3.4(a-d), shows the SEM images of the different samples annealed at different times. The NCs for the sample annealed for 15 sec. could not be resolved clearly. Although NCs were visible in the image, the image analysis software could not recognize these small NCs. Hence, this image was not used to calculate the NCs' size, density and area coverage. For this experiment minimum average size was obtained for sample S2 which was annealed for 45 sec.

Nanocrystals of average size (SD) of 5 (± 3.3) nm and density 8.96 x 10^{11} cm^{-2} were obtained for this sample. Figure 3.5(a-c), shows the variation in NC size, density and area coverage with anneal time. The SEM image of figure 3.4(d) could not be included in the analysis because of the low resolution of the SEM image. Nevertheless, it is very certain that NCs of very small sizes are forming in this sample. Further, the results of the experiment also suggest that it is not required to anneal the samples for 60 sec. as it was done for experiments in section 3.1.3.1. Annealing for lower time can yield the NCs of acceptable size and density. This experiment also verified that the NC formation conditions (Pt deposition for 30 sec. at 20 watt) are very stable as the average size of the NCs and dot density in both the experiments (section 3.1.3.1 and section 3.1.3.2) were almost equal (5 nm of this experiment compared to 4.96 nm of the previous experiment). Again in this experiment, area coverage was very poor (only 18%) which is far below the theoretically proposed area coverage for good memory performance [21].

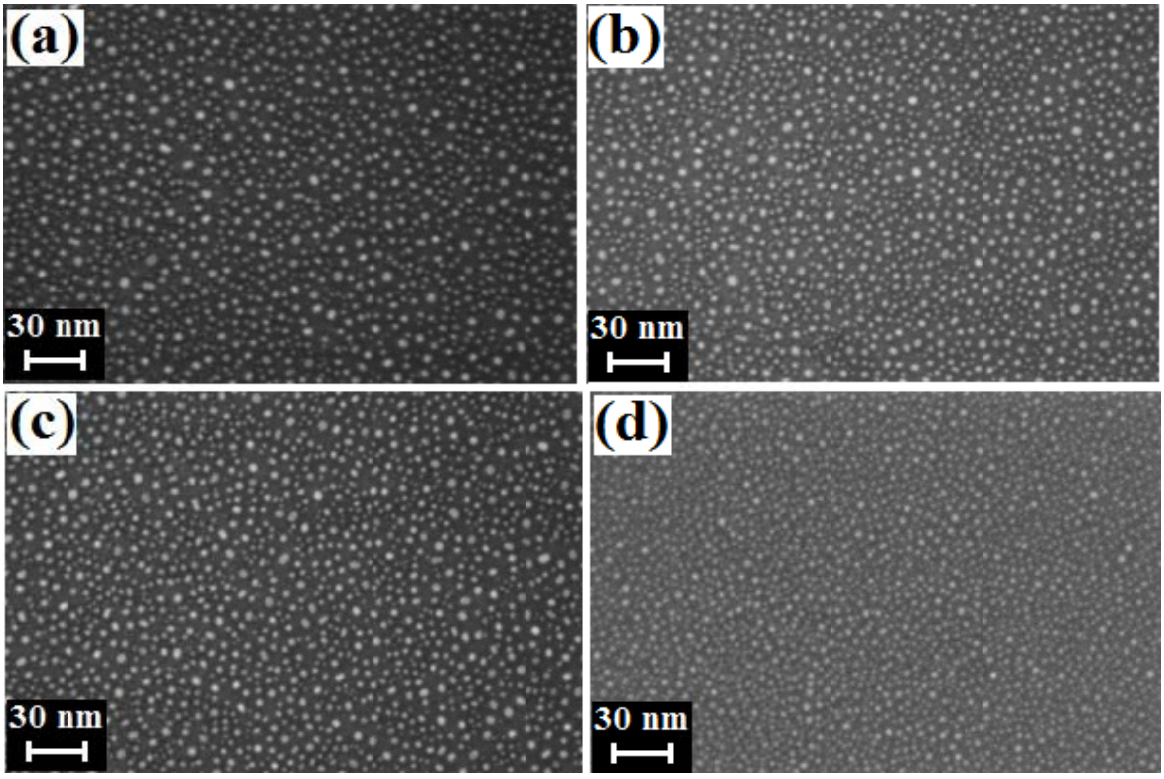


Figure 3.4: SEM images of Pt NCs for a fixed initial Pt thickness of ~ 2 nm and an anneal temperature of 550°C for different anneal time (a) 60 sec., (b) 45 sec., (c) 30 sec., (d) 15 sec.

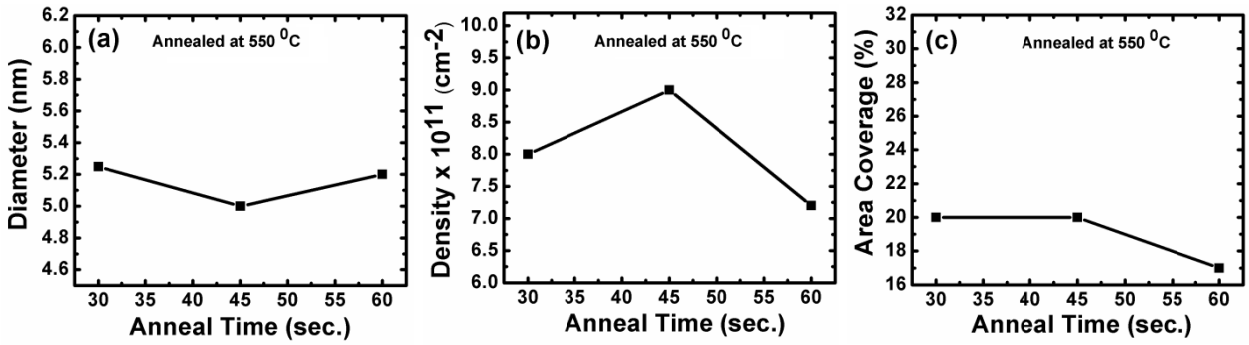


Figure 3.5: Pt NC statistics for a fixed initial Pt thickness of ~ 2 nm and an anneal temperature of 550°C for different anneal time (a) Diameter, (b) Density and (C) area coverage.

3.1.3.3 Effect of substrate temperature during the Pt deposition

In the previous experiments (section 3.1.3.1 and section 3.1.3.2), initial Pt was deposited without intentionally heating the substrate (referred to as room temperature deposition henceforth) and then annealing was performed. The experiment was also performed to study the effects of increased substrate temperature during initial Pt metal deposition and then subsequent anneal on the NC formation. The experimental conditions for this experiment are mentioned in table 3.3.

Table 3.3: Process conditions for the experiment to study the effect of substrate temperature during Pt deposition on the NC formation

Sample No.	Deposition Power (W)	Deposition time (sec.)	Substrate Temperature during Pt deposition (°C)	Anneal Time (sec.)	Anneal Temperature (°C)
S1	20	60	Room Temperature	30	550
S2	20	30	Room Temperature	30	550
S3	20	60	125	30	550
S4	20	30	125	30	550

Figure 3.6(a-d) shows the SEM images of NCs obtained. Figure 3.6(a, b) are the NCs obtained for room temperature Pt deposited at two different deposition times i.e. 60 sec. and 30 sec., while figure 3.6 (b, c) shows the NC images obtained for the Pt deposited for the same time (i.e. 60 sec.

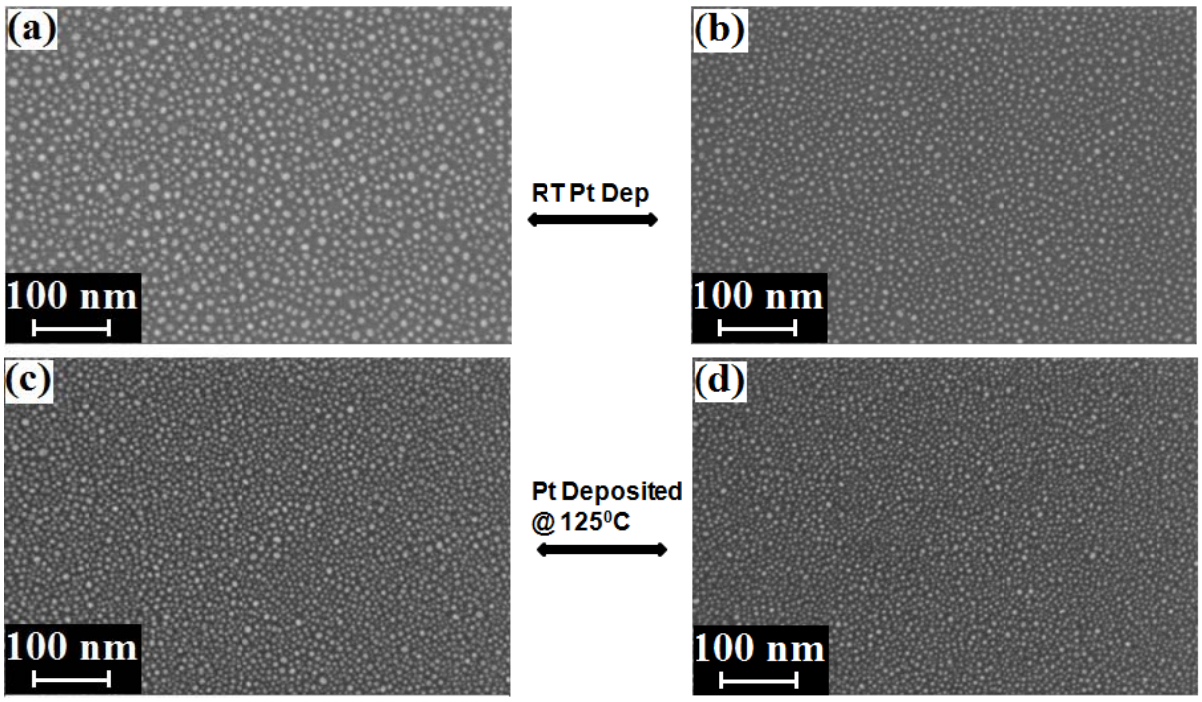


Figure 3.6: (a, b) are the SEM images of NCs obtained from room temperature Pt deposited for 60 sec. and 30 sec. respectively, while c, d are the images of NCs obtained from heated substrate (125°C) Pt deposited for 60 sec. and 30 sec. respectively. All samples were subjected to a post deposition anneal at 550°C for 30 sec.

and 30 sec.) but at 125°C substrate temperature. All samples were annealed at 550°C for 30 sec. From the SEM images shown in figure 3.6(a-d), it is clear that the density and area coverage has increased significantly for the NCs obtained under heated substrate Pt deposition condition. NCs of average size (SD) of 5.4 (± 3.5) nm and density of 1x 10^{12} cm^{-2} and the area coverage of 31.8% for sample S4 were obtained while the area coverage for sample S2 where the Pt was deposited at room temperature was only 18%. This is a huge improvement in density and area coverage without affecting the NC size much. This improvement is attributed to the increased surface migration of the Pt adatoms during initial metal deposition at higher substrate temperatures. At elevated substrate temperatures, metal adatoms move laterally on the sample due to extra thermal energy and thereby avoiding the accumulation of metal adatoms on the top of each other. This is clearly reflected in the increased surface density and area coverage without affecting the size of the nanocrystal. The areal density and the area coverage values obtained

from this experiment are very close to the desirable values of the NCs reported in [21]. Table 3.4 summarizes the NC statistics obtained from room temperature and heated substrate Pt deposited samples.

Table 3.4: summarizing the NC statistics obtained from Pt deposited at room temperature (RT) and 125°C substrate temperature. Data in bold letters is for heated substrate Pt deposited samples. SD stands for the standard deviation.

Deposition time (sec.)	Diameter (SD) (nm)		Density (10^{11} cm^{-2})		Area coverage (%)	
	RT	**125°C**	RT	**125°C**	RT	**125°C**
60	7.35 (±4)	**6.4 (± 5)**	4.3 x 10^{11}	**9 x 10^{11}**	22	**32.4**
30	5.2 (±3)	**5.4 (± 3.5)**	8 x 10^{11}	**1x 10^{12}**	18	**31.8**

3.1.3.4 Effect of the dielectric material

In all the previous experiments reported in this chapter, the NCs were formed on Si substrate having SiO_2 as a dielectric layer. However, several new materials with high dielectric constant are being explored to replace the SiO_2 as tunnel dielectric. Also, for the dual layer nanocrystal devices [72, 78], second layer of the NCs may form on the dielectric which need not be the SiO_2. Hence, the study of NC formation on other dielectrics is equally important.

A. Experimental

For Nanocrystal formation on dielectrics other than SiO_2, SiO_2 was thermally grown on bare Si wafer by dry oxidation at 850°C. Different high-K dielectrics, Al_2O_3, HfO_2 and Si_3N_4 were deposited on thermally grown SiO_2. These dielectrics were deposited on SiO_2 base layer to prevent any unwanted stress in the deposited layer because of any lattice mismatch between Si and the deposited film. For $Si/SiO_2/Al_2O_3$ sample, Al_2O_3 of 11 nm physical thickness was deposited at 300W deposition power by pulse DC sputtering technique in Applied Materials ENDURA PVD system [90, 91]. For $Si/SiO_2/Si_3N_4$ sample, Si_3N_4 layer [92] of 5 nm thickness was deposited by LPCVD process at 800°C. For $Si/SiO_2/HfO_2$ sample, HfO_2 of 7 nm thickness

was deposited by MOCVD in Applied Materials cluster tool using TDEAH precursor at 450°C. For NC formation, Pt was sputter deposited at 20W for 30 sec. on SiO_2, Al_2O_3 and Si_3N_4 samples. For the comparison of nanocrystal formation on HfO_2 and SiO_2, Pt deposition was done at 20W for 1 minute and therefore, initial Pt thickness was higher in this case. To form the Pt NCs, all the samples were subsequently annealed at five different temperatures starting from 450°C, 550°C, 650°C, 750°C and 850°C for 30 sec. in N_2 ambient. As a result of this high temperature treatment, thin Pt film converts into uniformly distributed metal islands. NC distribution on the four base layer dielectrics was analyzed by high resolution SEM. To find the effect of surface roughness on the NC formation process, Si/SiO_2 and Si/Al_2O_3 samples were prepared for atomic force microscopy (AFM) measurements.

B. Results and Discussions

Comparison of NC formation on SiO_2 and Al_2O_3

Figure 3.7(a-f) show the SEM images of NCs formed on SiO_2 and Al_2O_3 dielectrics at different annealing temperatures from 450°C to 850°C. Analysis of the NC size, density and area coverage is shown in figure 3.8(a) and 3.8(b) respectively. For both the dielectrics, as the temperature increases, NCs size decreases and density increases. This is attributed to the fact that the nucleation is a temperature dependent process and probability of nucleation increases at higher temperatures [93]. As the temperature increases, new nucleation sites continue to appear and hence new NCs form at these newly created nucleation sites leading to the increase in density with temperature. As the temperature continues to increase, nucleation process saturates and beyond a critical temperature, surface migration of metal atoms dominates on the nucleation process. Hence, after a certain temperature, NCs size increases (by coalescence of different metal atoms as a result of increased surface migration). Mean NC size at 850°C on SiO_2 and Al_2O_3 were 5.05 nm and 4.9 nm respectively. An important observation was that the NC size for all annealing temperatures was smaller on Al_2O_3 substrate when compared to the NC size on the SiO_2 substrate. Similarly area coverage of the NCs on Al_2O_3 substrate was higher (30%) than that compared to SiO_2 substrate (24%) at 850°C. These results may be explained on the basis of surface roughness of both the substrates. Mean root mean square (rms) roughness measured from

AFM for Al_2O_3 is 0.226 nm while for SiO_2 is 0.171 nm as shown in figure 3.9. Higher surface roughness leads to more nucleation sites which enable denser and smaller NC formation.

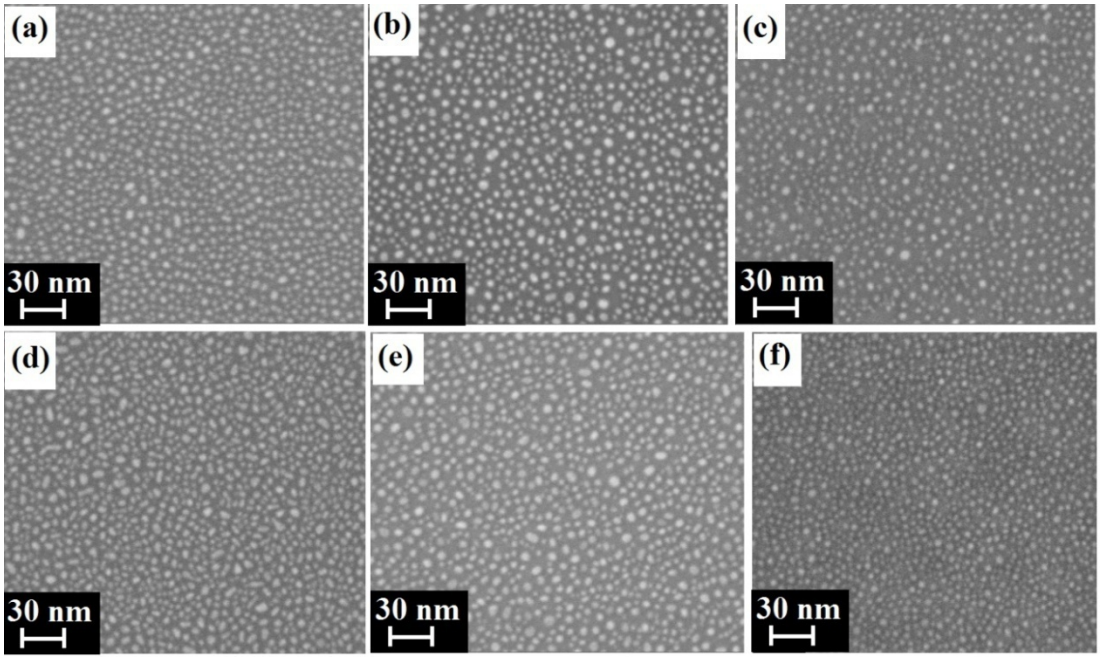


Figure 3.7: (a-c) SEM images of NCs on SiO_2, (d-f) SEM images of NCs on Al_2O_3 annealed at temperatures 450°C, 550°C and 850°C respectively.

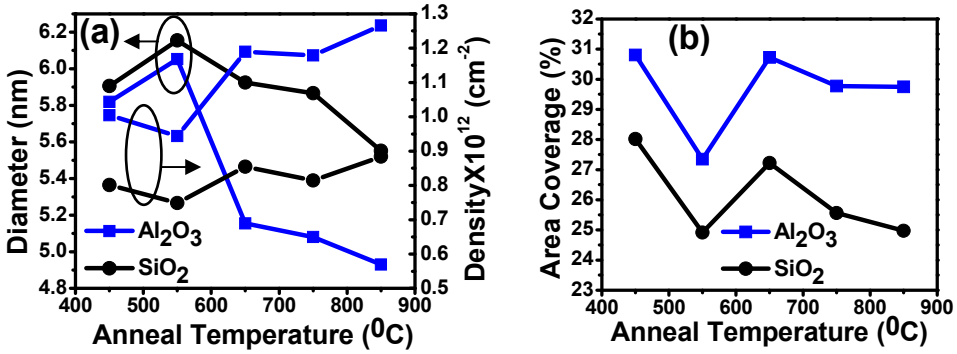


Figure 3.8: (a) Comparison of NCs diameter and density and (b) comparison of area coverage of NCs on SiO_2 and Al_2O_3 dielectrics at different annealing temperatures.

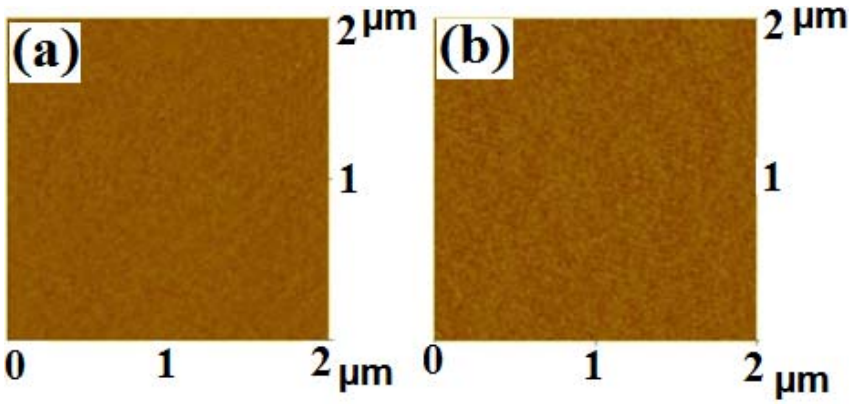


Figure 3.9: AFM image of (a) SiO_2 and (b) Al_2O_3. Surface roughness of SiO_2 is 0.17 nm while for Al_2O_3 surface roughness is 0.22 nm.

Comparison of NC formation on SiO_2 and HfO_2

The SEM images of NCs formed on SiO_2 and HfO_2 at 450°C, 550°C and 700°C respectively are shown in figure 3.10(a-f). NCs diameter and density are shown in figure 3.11. For these experiments, Pt deposition was done for longer time i.e. 60 sec., therefore, the overall NCs size is larger than the size of the NCs in the experiments on SiO_2/Al_2O_3 substrates. It is clear from the figure 3.10 that larger size NCs with lower density are forming on HfO_2 as compared to the NCs on SiO_2. Furthermore, as the temperature increases, NCs size on HfO_2 increases (density decreases), which reflects that surface migration of Pt atoms is more dominant phenomenon rather than the creation of new nucleation sites with temperature.

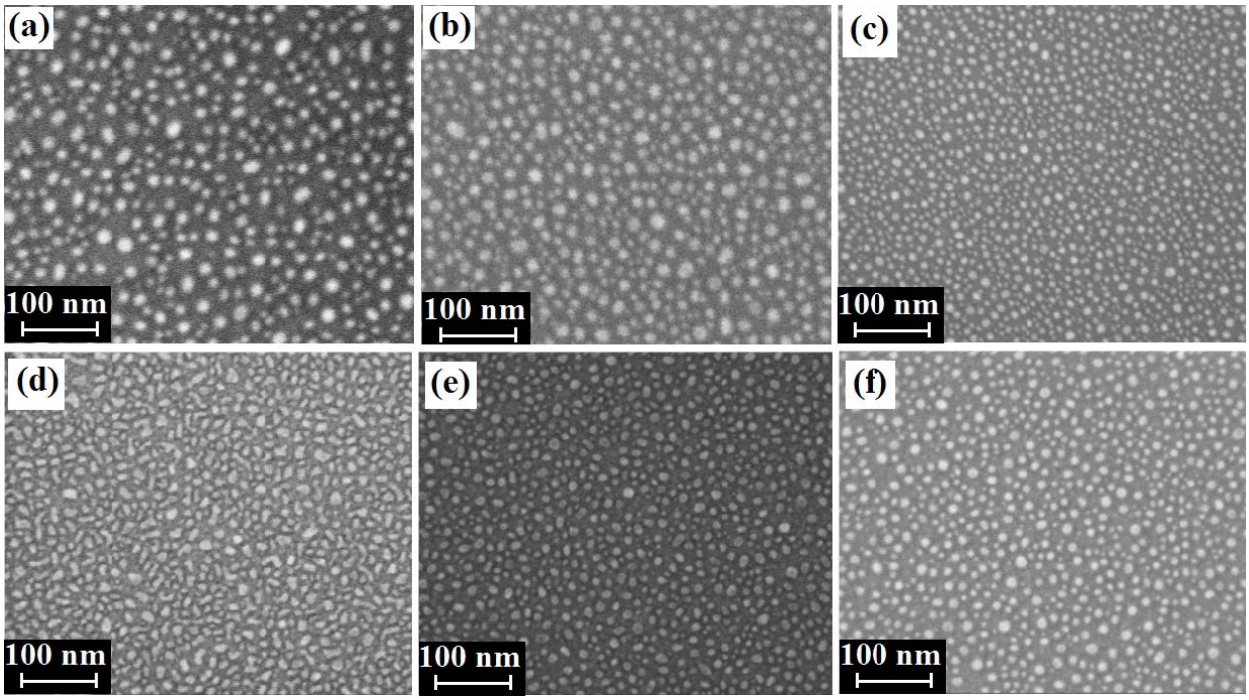


Figure 3.10: (a-c) SEM images of Pt NCs on SiO_2 annealed at temperatures 450°C, 550°C and 700°C respectively, (d-f) SEM images of Pt NCs on HfO_2 annealed at temperatures 450°C, 550°C and 700°C respectively.

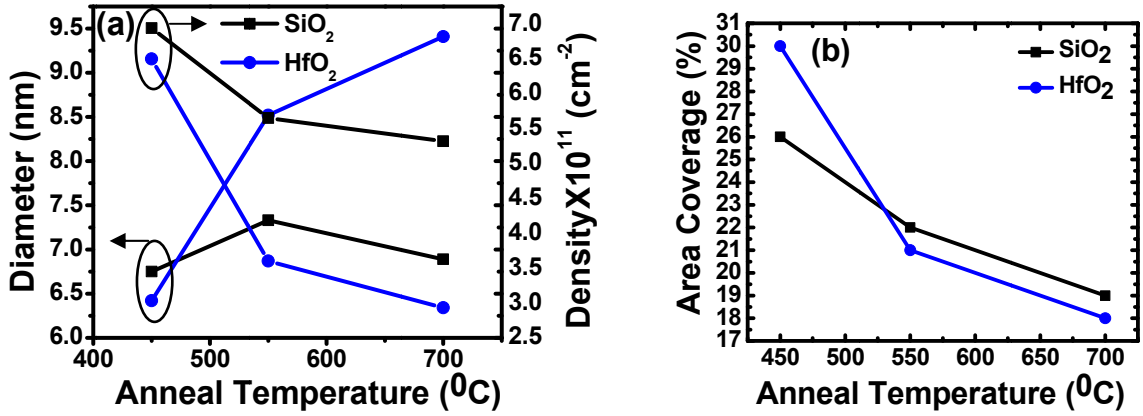


Figure 3.11: (a) Comparison of Pt NC's diameter and density and (b) comparison of area coverage of NCs on SiO_2 and HfO_2 dielectrics at different annealing temperatures.

Comparison of NC formation on SiO_2 and Si_3N_4

Figure 3.12(a - c) shows the SEM images of NCs formed on Si_3N_4 at temperatures of 450°C, 550°C and 750°C respectively. NCs size and density are shown in figure 3.13(a) and area coverage is shown in figure 3.13(b). On Si_3N_4, the nonuniformity in NCs size and distribution is clearly observed when compared to the NCs obtained on SiO_2. Overall NCs size (density) is higher (lower) for Si_3N_4 substrate compared to NCs on SiO_2 substrate.

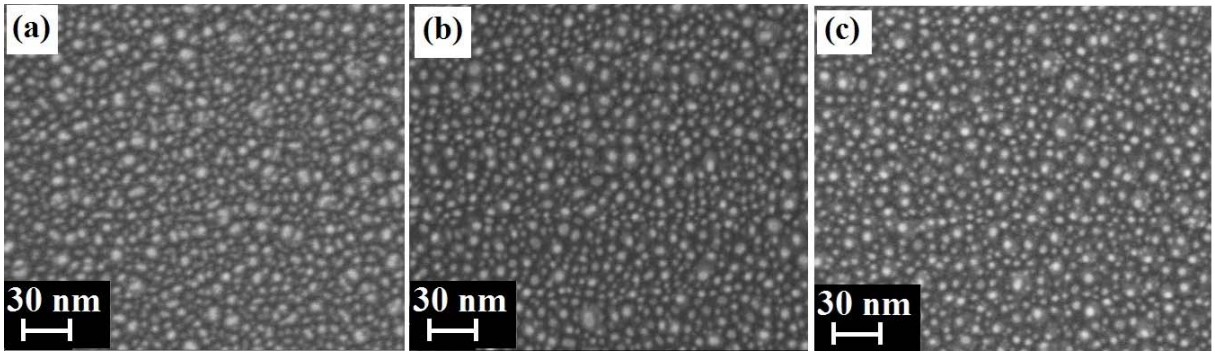


Figure 3.12: SEM images of Pt NCs on Si_3N_4 annealed at temperatures (a) 450°C, (b) 550°C and (c) 750°C respectively.

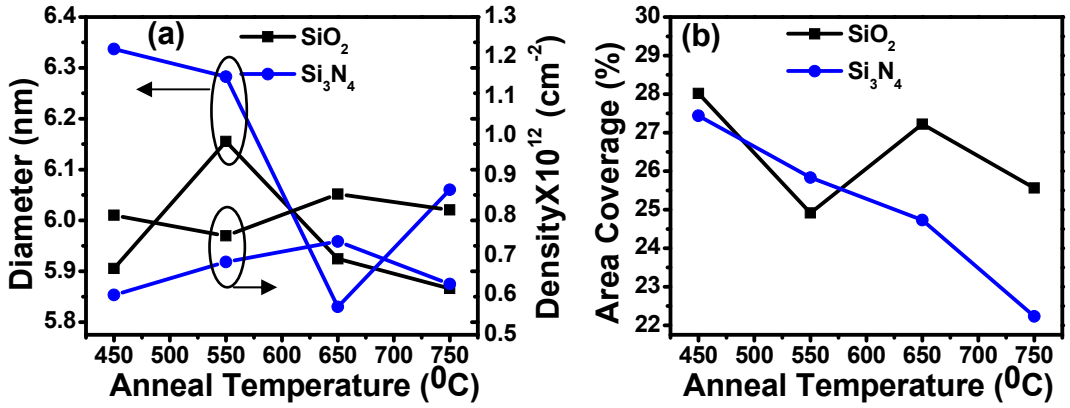


Figure 3.13: (a) Comparison of NCs diameter and density, (b) comparison of the area coverage of NCs on SiO_2 and Si_3N_4 dielectrics at different annealing temperatures.

With these experiments, it is clearly demonstrated that nanocrystal formation process is strongly dependent on the material on which the NCs are formed. Experiments performed on different

dielectrics clearly show that under identical conditions, NCs statistics significantly differ from one dielectric to other dielectric. Al_2O_3 supports the smallest diameter NCs as compared to the NCs on other dielectrics. Hence, to optimize the NCs size, density and area coverage for flash memory application, the NCs statistics obtained on one dielectric simply cannot be assumed to be applicable to other dielectrics. The results are important especially for dual layer flash memory devices where the first layer (tunnel dielectric) and the second layer (intermetal dielectric) need not to be of same dielectric material. Hence, for such devices, NCs formation optimization should be done separately for both the layers.

3.1.4 Comparison of Pt and Ir NC formation on SiO_2

Ir and Pt have almost comparable work functions (for Ir, WF is 5.27 eV and for Pt, WF is 5.3 eV) however, the melting point of Ir (2443°C) is higher than that of Pt (1772°C). Therefore, it is expected that Ir should form smaller size NCs compared to Pt [68]. In this section, NC formation statistics of these two metals are compared and found that for same initial thickness (as optimized by adjusting the deposition power and deposition time) of Ir and Pt, Ir forms the smaller size NCs as compared to Pt. One point that is worth mentioning here is, to form the Ir NCs, higher anneal temperature or longer duration anneal is required. Pt nanocrystals form very readily at 550°C for 30 sec., however, for Ir, it required 90 sec. anneal at 550°C to form the NCs. The SEM images for Ir NCs deposited at 30 sec. and annealed at 550°C for 90 sec. is shown in figure 3.14(b) and the comparison between Pt and Ir NCs statistics is given in table 3.5.

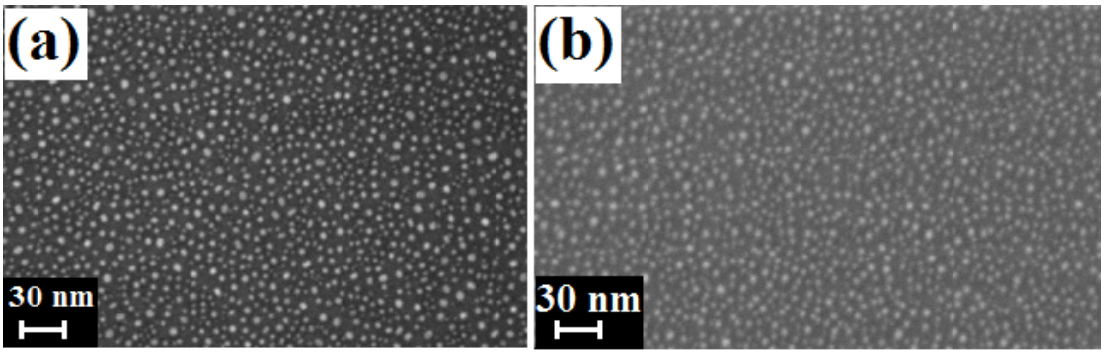


Figure 3.14: (a) SEM image of Pt NCs annealed at 550°C for 30 sec., (b) SEM image of Ir NCs, annealed at 550°C for 90 sec.

Table 3.5: Comparison between Pt and Ir NC formation statistics.

Deposition Time (sec.)	Diameter (nm)		Density $*10^{11}$ cm^{-2}		Area coverage (%)	
	Pt	Ir	Pt	Ir	Pt	Ir
25	5.4	-	8.78	-	22.73	-
30	5.97	5.04	5.89	10.6	19.29	25.44
45	6.2	5.32	6.34	10.2	22.38	25.98
60	7.35	5.83	4.37	7.22	22.98	22.42

3.2 Pt NC memory

Pt NCs were integrated in a flash memory device structure with Al_2O_3 as a blocking dielectric and SiO_2 as a tunnel dielectric. For this, Pt NCs (of average. diameter 5 nm) were fabricated on thermally grown SiO_2 (3.5 nm). Pt was deposited at 20 W for 30 sec. (thickness ~2 nm) and annealed at 550°C for 60 sec. Al_2O_3 of 12.5 nm physical thickness was deposited by PVD using pulse DC power supply in Applied Materials ENDURA PVD system. TiN as a top gate metal was deposited in the same ENDURA system in a different chamber and lithographically patterned in circular dots of diameter 80 µm. Dry plasma etching of TiN was done in Applied Materials etch CENTURA using Cl_2/BCl_3 chemistry. Control devices (without Pt NCs) were also fabricated under identical process conditions. Cross section SEM (X-SEM) image of the full memory stack and X-HRTEM image of Pt NC is shown in figure 3.15(a) and 3.15(b) respectively. Diameter of the Pt NCs obtained from the X-HRTEM image of the figure 3.15(b) is ~ 5 nm which is consistent with the size of the NCs obtained from the SEM analysis (figure 3.2(c)). Capacitance-Voltage (CV), Current density-Voltage (JV) and Weibull plot for control sample and for NC devices are shown in figure 3.16(a-c) respectively. From the CV plot, effective oxide thickness (EOT) of the stack is 9.1 nm. EOT is obtained as per equation 3.1

$$\boldsymbol{EOT = \frac{\varepsilon_{SiO_2}\varepsilon_o}{C_{OX}}} \quad \text{..... (3.1)}$$

where ε_0 is the absolute permittivity of free space, ε_{SiO2} is the relative permittivity of the SiO_2, C_{OX} is the oxide capacitance per unit area. Breakdown field of the control devices is about 17 MV/cm while for Pt NC devices, breakdown field is 11 MV/cm. For memory devices, 3.6 V memory window is observed at 10 V programming/erase voltage as shown in figure 3.16(d). These devices could not be programmed further because of low breakdown voltage.

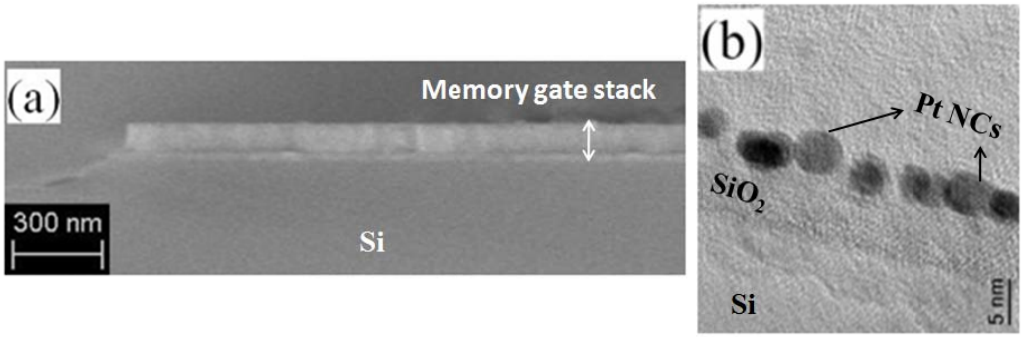


Figure 3.15: (a) X-SEM image of full Pt NC memory stack, (b) X-HRTEM image of Pt NCs.

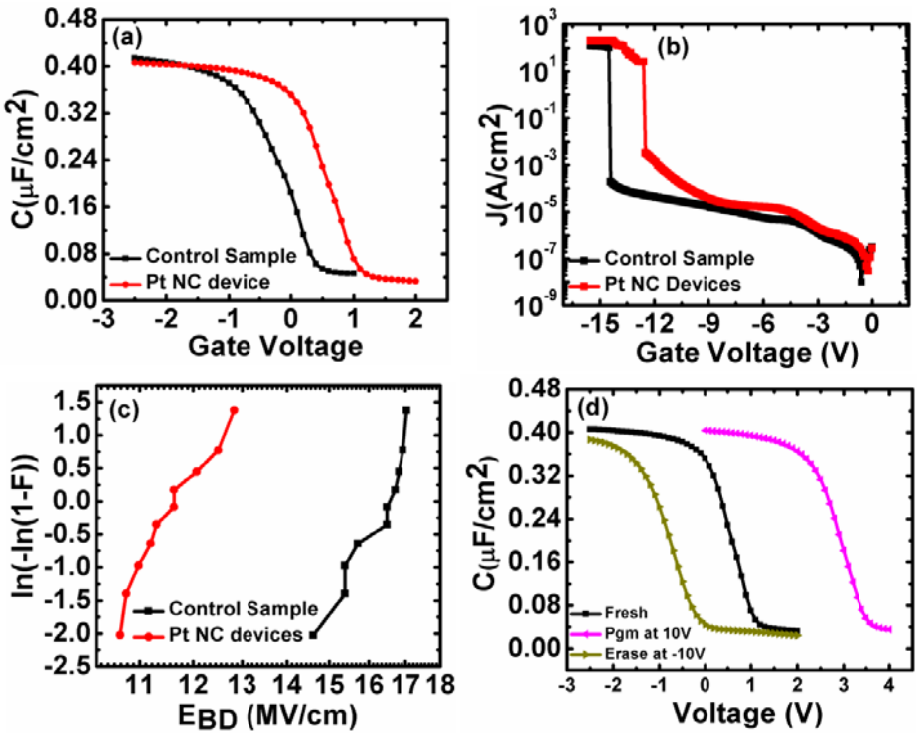


Figure 3.16: (a) CV Plot, (b) JV plot, (c) Weibull plot of control sample and Pt NC device, (d) CV Plot of Fresh, Programmed (at 10 V) and Erased (at -10 V) Pt NC device.

3.3 Shortcomings of the NC as a charge storage layer

Even though NCs, as a charge storage layer, offer several advantages over the continuous floating gate, this technology suffers from severe technical challenges. Some of these are listed below [21-24];

(1) NC memory performance strongly depends on the NC statistics i.e. size, density and area coverage. As NC formation is a self assembled process, controlling the NC statistics is not easy. This would cause device to device variability, which would be even more serious with technology scaling where any small fluctuation in the number of stored electrons would cause large change in the threshold voltage. Infact, the successful scaling of floating gate flash cell towards 20 nm node has put even more stringent requirements on the nanocrystal distribution.

(2) In order to reduce the cell to cell variability, a very high density of NCs (~ 5 x 10^{12} cm^{-2}) is required; however, with such high density of NCs, spacing between the NC to NC decreases significantly. For example, average spacing between the NCs for maximum density obtained in our study is ~2.5 nm (sample S4 figure 3.6(d)). With such closely packed density, even though they may remain physically separated, electrically they behave like continuous metal floating gate [24]. Hence the proposed benefits of isolated charge storage nodes may not be achieved with this technology.

(3) Reliability: NCs are formed at high temperature process; hence metal diffusion in the surrounding dielectric poses a serious reliability issue with this technology.

These issues are very serious when the NC memory technology is considered for 20 nm or below technology node. There is no solution for some of these technological limitations like variability in NC statistics. These issues with NC memory motivate one to explore the other possibilities of charge storage medium in floating gate flash memory devices.

3.4 Summary of the chapter

Nanocrystal formation statistics were studied at different process conditions and it was clearly demonstrated that NC distribution can be controlled by controlling initial metal thickness, anneal temperature and anneal time. Further, it was also found that substrate temperature during initial metal deposition also affects the NCs statistics. NCs formed with the Pt deposited on heated substrates were dense and had high area coverage. Minimum size of the Pt NCs was found to be about 5 nm for 30 sec. Pt deposition (~2 nm) annealed at 550ºC for 30 sec. It was also demonstrated that nanocrystal formation process is strongly dependent on the dielectrics on which it is formed. Experiments done on different dielectrics clearly showed that under identical conditions, NCs statistics significantly differ from dielectric to dielectric. It was shown that Al_2O_3 supports the smallest diameter NCs as compared to the NCs on other dielectrics. Further, comparison between Pt and Ir NCs revealed that, Ir forms the smaller size NCs as compared to Pt. Complete Pt NC memory devices were fabricated with Al_2O_3 as a blocking dielectric and TiN as a top gate electrode. Control sample showed high breakdown field (17 MV/cm) as compared to Pt NC memory devices (~ 11.7 MV/cm). For complete flash stack, a memory window of 3.7 V at ± 10 V program/erase voltage was obtained. However, the Pt NC devices could not be programmed and erased further because of the poor breakdown field of the Pt NC memory devices. Since, it was identified that NC memories are not promising for scaled technology nodes due to the technical limitations associated with them, efforts were redirected towards exploring other charge storage layers.

.

Chapter 4

Graphene floating gate flash memory

4.1 Introduction

Lateral scaling of the flash memory devices results in increased capacitive coupling among the floating gates of adjacent cells which causes a wide distribution in threshold voltages of the devices [5]. Floating gate height reduction is one possible way to reduce the capacitive coupling [5]. Recently it has been demonstrated that conventional poly-Si floating gate thickness can be reduced to 7 nm [25]. However, thin poly-Si may not be able to scatter the carriers in the programming current and hence these carriers would be ballistically transported through the floating gate which would result in slower programming [25]. These ballistic carriers may also cause impact ionization in the blocking dielectric and thus degrade the dielectric reliability. To alleviate these issues with thin poly-Si floating gate, thin metal layer can be more effective to reduce the ballistics current component. Metal layer as thin as 1 nm thickness as a floating gate material is found to be capable of suppressing the ballistic current component [25, 26]. Metal floating gate, however, may impose its own device variability and reliability issues at high temperatures such as: (i) agglomeration of thin metal layers, (ii) diffusion of metal into the tunnel and blocking dielectrics [31], and (iii) increased leakage current through the blocking dielectric deposited on metal films due to higher degree of dielectric crystallization at elevated temperatures [32].

Since graphene is the thinnest naturally stable sheet having metallic properties, it is possible to use graphene in place of poly-Si or metal as a charge storage layer (CSL) in floating gate flash memory. As the interlayer spacing between two graphene sheets in MLG is only 0.34 nm, 6-7 layers of MLG sheets would be 2 - 3 nm thick. Therefore, incorporation of MLGs as floating gate layer in flash memory structure would lead to substantial reduction in the vertical dimension

of these devices. Graphene is reported to be thermally stable upto 1500°C [94], and hence thermal stability issues anticipated with metal floating gate may not be a problem with graphene.

Previously, nonvolatile memories (NVM) based on graphene [95-97, 37] and graphene oxide (GO) [36] have been reported. In most of these reports, with the exception of [36] and [37], the device structure and memory operation is different from the existing floating gate flash technology. In [37] gate stack consists of Si / SiO_2 (5 nm) / MLG / Al_2O_3 (35 nm) / Ti/Al/Au. With this gate stack, a hysteresis of 6 V in CV characteristics of floating gate metal-oxide-semiconductor (MOS) capacitors is reported when the gate voltage was swept from -7 V to +7 V and back. While in [36], a hysteresis of 7.5 V for ±14 V voltage sweep in the CV curve is reported with MOS capacitors having 5 nm SiO_2 as a tunnel oxide, GO as floating gate and 15 nm Al_2O_3 as blocking dielectric with TaN top gate electrode. Erasure of the devices is not discussed in [37]. Owing to the ambipolar nature of the Gr sheets, possibility of hole storage and hence over-erase cannot be ruled out in these memory devices. In [36], hole storage in the GO is discussed. In [36], charge storage capability of reduced single layer GO is also demonstrated with a memory window of 1.4V. However, program/erase (P/E) transients, an important figure of merit to qualify the speed of FG flash devices, are not discussed in these studies [36, 37]. In P/E transient measurements, voltage pulses of fixed voltages are applied on the gate electrode for exponentially increasing time (e.g. 1 µsec., 10 µsec., 100 µsec., 1msec. and so on) and a corresponding change in the flatband voltage (V_{FB}) is noted. Programming speed can be assessed by noting the change in V_{FB} with various pulse durations. In this chapter, we discuss the experimental results on graphene as a charge storage layer with three different kinds of test structures and program/erase transients, memory window and retention characteristics.

The results presented in this chapter were reported at the International Memory Workshop 2011 held in Milan, Italy and in the IEEE Electron Device Letters vol. 34, no. 9, pp. 1136- 1138, 2013. The work presented in section 4.3.3 was done in collaboration with Mr. Mayur Waikar.

4.2 Multilayer graphene as a charge storage node

We investigate charge storage capability of multilayer graphene (MLG) sheets. Though, single layer graphene (SLG) has attracted more attention from scientific community due to its extraordinary charge carrier mobility, for charge storage application, MLG has several technical advantages over their single layer analogs; hence we propose that MLG is more suitable for memory application because of following reasons:

(1) Material to be used as a FG purpose should have high work function to improve the retention of the memory devices. WF of single layer graphene is 4.2 eV. WF of graphene is susceptible to the number of layers, when it is less than 4. Thicker graphene sheets (> 4 layers) WF saturates to a value of 4.6 eV [98]. Accurate control of the number of sheets on large areas required for electronic applications is a challenge as the layers can overlap. This would lead to variations in WF over the wafer. In the flash memory structure using graphene for charge storage, graphene sheets are to be sandwiched between two dielectrics thus forming a potential well. The variation in the WF of Gr sheets will cause a variation in the potential well depth from a value of 4.6 eV for MLG sheets to 4.2 eV for single layer graphene (SLG), which would cause the device to device variability. In addition to this consideration, high WF of MLG may be favorable for achieving long-term data retention.

(2) Density of states (DOS) in MLG is higher compared to SLG. Using the equations given in [99], value of DOS for single layer, bilayer and trilayer graphene can be calculated as 8 x 10^{12} $cm^{-2}eV^{-1}$, 3 x 10^{13} $cm^{-2}eV^{-1}$ and 4.4 x 10^{13} $cm^{-2}eV^{-1}$ respectively. High DOS in MLG is favorable for large memory window.

(3) Graphene has very high in plane conductivity and low conductivity in plane perpendicular to it. Reduced conductivity along C-axis in MLG [100] could be more effective in suppressing the ballistic component of the programming current through the floating gate.

4.3 Graphene flash memory device fabrication and characterization

To evaluate the charge storage capability of MLG in floating gate memory structure, three test structures viz. flash MOS-capacitor, flash MOS-capacitor with implanted surroundings and flash transistors were fabricated. The Reason for three different kinds of structures is discussed in the respective sections. All the three test structures were fabricated with gate last process. In section 4.3.1, experimental details of graphene flash MOS capacitors with the analysis of results is presented, in section 4.3.2 experimental details and analysis of results of flash MOS-capacitors having implanted surroundings are discussed while in section 4.3.3 experimental details and analysis of graphene flash transistors is discussed.

4.3.1 Graphene floating gate flash MOS Capacitor Structure

4.3.1.1 Device Fabrication

Graphene flash memory MOS-capacitors were fabricated on p-type silicon substrate with SiO_2 as tunnel oxide, reduced MLG (rMLG) sheets as charge storage layer and Al_2O_3 as blocking dielectric [91] with TiN as the top gate electrode. A schematic of the cross section of the rMLG CSL devices is shown in figure 4.1. rMLG sheets, obtained after reducing graphene oxide (GO) with water and hydrazine [101], were deposited on thermally grown SiO_2 by drop casting. More discussion on the structure of GO and reduced graphene oxide is given in annexure A of this thesis. Uniform distribution of rMLG sheets over a large area of SiO_2 was confirmed by scanning electron microscopy (SEM) and atomic force microscopy (AFM) as shown in figure 4.2(a) and 4.2 (b) respectively. The thickness of rMLG sheets was around 2- 3 nm (AFM image 4.2(b)) and this indicates that the rMLG consist of 6 - 7 layers.

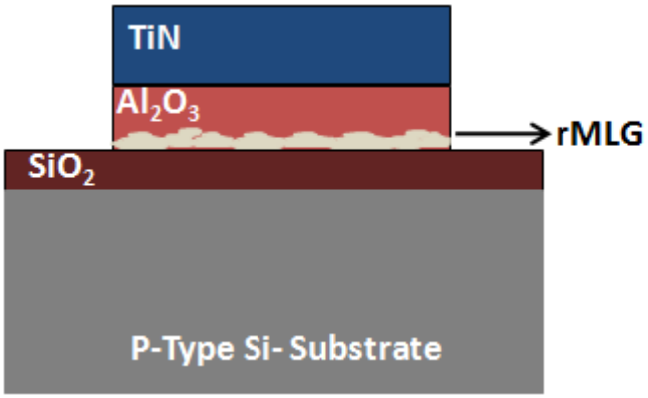


Figure 4.1: Schematic of the cross-section of rMLG CSL flash MOS capacitor.

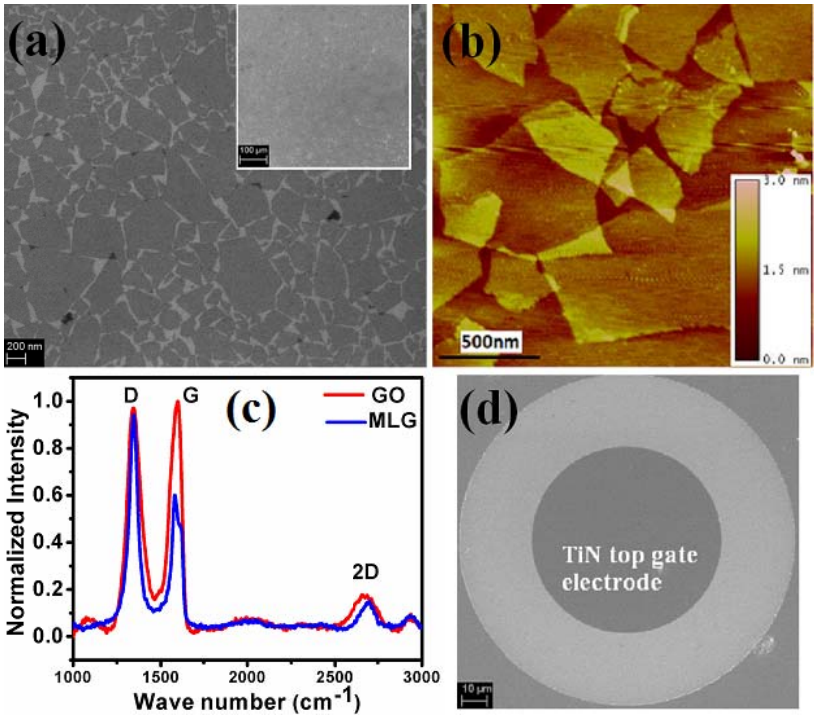


Figure 4.2: (a) SEM image of rMLG sheets on tunnel oxide. Inset shows the large area uniform distribution of rMLG sheets. (b) AFM image of rMLG sheets on tunnel oxide. Thickness of rMLG is 2-3 nm. (c) Raman spectra for multilayer graphene obtained after reduction of graphene oxide and that of pure graphene oxide (GO). (d) SEM image (top view) of the as fabricated flash MOS-Capacitor. The circle in dark shade is the TiN gate. TiN, Al_2O_3 and rMLG are etched away from the surrounding ring, seen in the lighter shade.

Raman spectrum of multilayer graphene used in the devices is shown in figure 4.2(c). For comparison, Raman spectrum of GO is also shown. Higher ratio of D peak to G peak intensities (1.59 for rMLG and 0.987 for GO) signifies the reduction of GO into more graphene like structure [101]. The positions of rMLG sheets on the substrate were carefully noted using RAITH 150Two system in the SEM mode for further patterning the top gate electrode at these locations. This procedure affirmed the presence of rMLG sheets in memory devices. On the same wafer, a region without graphene sheets was also identified for control devices using the same SEM system. For blocking dielectric and top gate electrode, 15 nm of high quality Al_2O_3 and 80 nm TiN were deposited by physical vapor deposition in Applied Materials ENDURA PVD system. PMMA was spin coated on the sample and circular TiN dots of 75 μm diameter were patterned at the already noted coordinates by electron beam lithography using RAITH 150Two direct write setup. Finally, dry etching of the TiN and Al_2O_3 was performed in Cl_2/BCl_3 plasma and rMLG sheets were etched by O_2 plasma in Applied Materials Etch Centura system. Top-view SEM image of a typical, as fabricated device is shown in figure 4.2(d). The circle in dark shade in figure 4.2(d) is the TiN gate and TiN, Al_2O_3 and rMLG were etched away from the surrounding ring, seen in the lighter shade. Control samples were fabricated using an identical procedure on the same wafer in locations where no MLGs are seen using SEM.

4.3.1.2 Results and Discussion

Charge storage capability of rMLG memory MOS-capacitors were tested by performing CV, and program/erase (P/E) transients using Agilent 4284 LCR meter and Agilent 4156C semiconductor parameter analyzer respectively. Effective oxide thickness of the gate stack of the device as calculated from CV plot, shown in figure 4.3(a), is 13.5 nm. CV plots after different P/E voltages (8V to 18V) are shown in figure 4.3(a). CV curves for the control sample (without MLG sheets) programmed and erased at ±10V are shown in figure 4.3(b).

Programming was achieved by applying a positive program voltage for a predetermined time on the gate and subsequently measuring the CV to find the shift in flat band voltage (V_{FB}). The programming intervals were varied exponentially as is the practice [26]. Since the devices were made on p- type silicon wafers, a positive gate voltage would result in inversion of the silicon

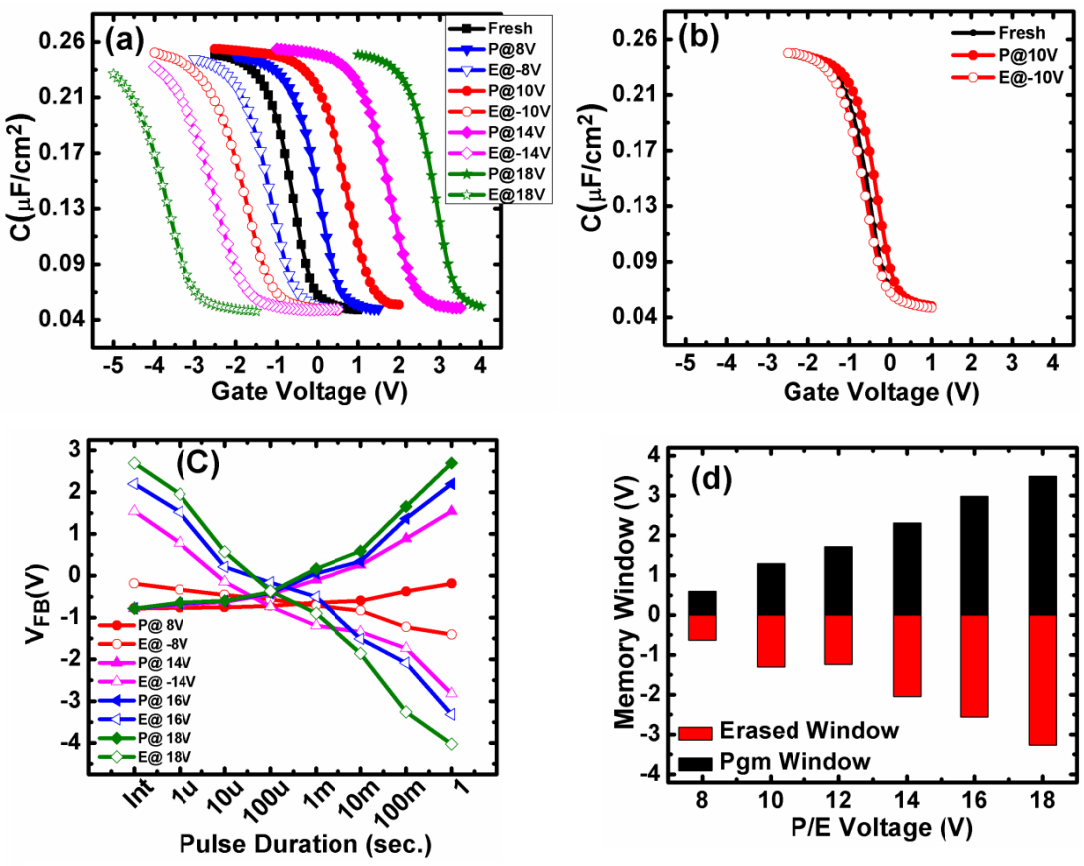


Figure 4.3: (a) CV plots of the flash memory capacitors after successive programming and erasing. P and E stand for the program and erase respectively. "Fresh" indicates the CV of the pristine device before application of P/E pulses. (b) CV plots for control sample (without MLG sheets) programmed and erased at ±10 V for 1 sec. (c) Program/Erase transients of the MLG flash devices. (d) Program and erase memory window at each P/E voltage used.

surface. Under high positive gate voltage the inversion carriers (electrons) would undergo Fowler-Nordheim (F-N) tunneling through tunnel oxide and these carriers would fall into the potential well formed by the graphene sheets. Right shift in the CV after programming confirms the electron storage in the rMLG sheets. During the erase operation, a high negative potential is applied to the gate. The silicon surface would be accumulated with holes, which tunnel through the tunnel oxide and recombine with the electrons stored in the floating gate. The procedure used for erasing, is similar to that for programming, except for the polarity of the voltage applied to the gate. The devices were programmed and erased at each voltage for 1sec. The Program/erase

transients for the devices with rMLG are shown in figure 4.3(c). Significant change in the V_{FB} is observed for all program/erase voltages except ±8 V. The difference in the slope of programming and erasing curves in figure 4.3(c) is due to the different kind of carriers (electrons for programming and holes for erasing) involved in the mechanism. In the MOS capacitor structure, inversion carriers, i.e. electrons for a p-type silicon substrate, have to be generated by thermal generation processes and this may take several 10's of micro seconds depending on the quality of the silicon wafer [102] and hence results in delayed programming [103].

Difference between the V_{FB} after 1sec. programming and the V_{FB} after 1sec. erasing is taken as the memory window. Program window is defined as the difference between the V_{FB} after 1sec. programming and the V_{FB} of the device prior to programming. Similarly, erase window is defined as the difference between the V_{FB} after 1sec. erasing and the V_{FB} of the device prior to programming. The memory window for the control sample after 10 V program/erase (P/E) is only 0.16 V while for rMLG memory capacitors the corresponding memory window is 2.61 V. This confirms that significant charge storage is taking place in the rMLG sheets. Maximum memory window was observed after 18 V P/E and is 6.8 V for 1 sec. programming. Smaller memory window reported in [37] for reduced SLG may be because of the lower DOS in SLG. A memory window of 6.8 V is sufficient for reliable multi level data storage applications. The program and erase windows of the device for various P/E voltages are shown in figure 4.3(d).

The total number of electrons per unit area stored in the rMLG sheets after programming was calculated using the equation [42]

$$N_{Gr} = \frac{\Delta V_{MG}\, \varepsilon_0 \varepsilon_{Al2O3}}{q\, t_{Al2O3}} \quad \ldots\ldots\ldots \quad (4.1)$$

where V_{MG} is the midgap voltage corresponding to the midgap capacitance, ΔV_{MG} is the difference in the midgap voltage of the device before and after programming, ε_0 is the permittivity of free space, ε_{Al2O3} is relative permittivity of Al_2O_3, t_{Al2O3} is the physical thickness of Al_2O_3, and q is the electronic charge.

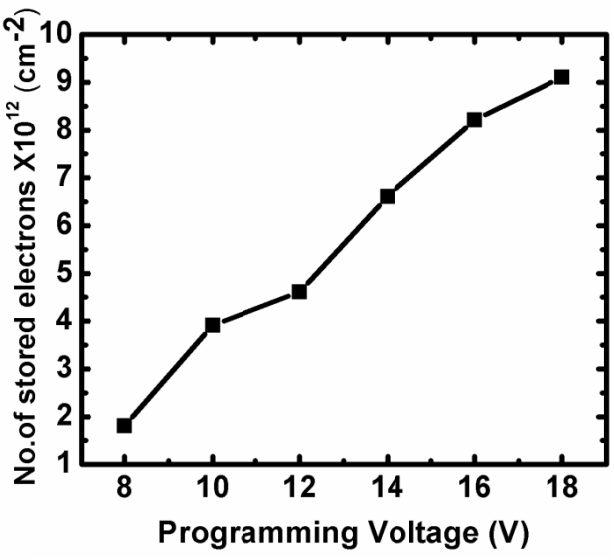


Figure 4.4: Number of stored electrons per cm^2 in the rMLG sheets of the programmed device.

Using the thickness of the tunnel and blocking dielectric measured using ellipsometry, ε_{Al2O3} and t_{Al2O3} were calculated from the CV of the control device. Number of electrons stored in the rMLG sheets is shown in figure 4.4. Density of stored electrons at 18 V programming voltage is 9.1 x 10^{12} cm^{-2} while the DOS for ideal SLG, based on the model given in [98] is 8 x 10^{12} $cm^{-2}eV^{-1}$. This observation further supports the multilayer nature of graphene sheets used in the present study. Comparison of this data with the work reported in [37], shows that higher DOS of MLG sheets leads to more charge storage and hence larger memory window.

Slow programming with graphene flash MOSCAP structure observed in the figure 4.3(c) can be avoided if programming carriers are supplied without any time delay. This can be achieved with a MOSCAP structure having heavily doped n-type regions surrounding the gate stack. This structure, we call it as pseudo source/drain structure. This structure has the implanted regions surrounding the memory gate stack. These implanted regions serve as a minority carrier supplier during programming of the memory devices. This structure is very commonly used in the flash memory studies [8]. In the next section, device fabrication of this pseudo S/D structure and memory performance is discussed.

4.3.2 Graphene floating gate flash MOSCAP with implanted surroundings

4.3.2.1 Device Fabrication

Graphene CSL flash MOSCAPs with phosphorous implanted around the gate stack were fabricated on p-type silicon substrate with 8 nm SiO_2 as tunnel oxide, 22 nm Al_2O_3 as blocking dielectric and TiN as the top gate electrode. The devices were fabricated with gate last scheme, i.e. n-type implantation and activation anneal were performed before fabricating actual memory gate stack by protecting the active region with a dummy gate. In this experiment, rMLG is obtained after thermal reduction of GO. More discussion on the thermal reduction of GO sheets is given in the annexure A of this thesis. Other device fabrication details *viz*. dielectric deposition and metal etching etc. were similar to the device fabrication flow of section (4.3.1). More details of the device fabrication are given in the annexure C of this thesis.

A schematic of the process flow for device fabrication is depicted in figure 4.5(a-e). Schematic cross-section image of the as fabricated devices is shown in Figure 4.6(a) and the AFM image of the rMLG used in this experiment is shown in Figure 4.6(b). Diameter of the MOSCAPs was 80 µm. Control samples with no rMLG sheets were also fabricated using identical procedure.

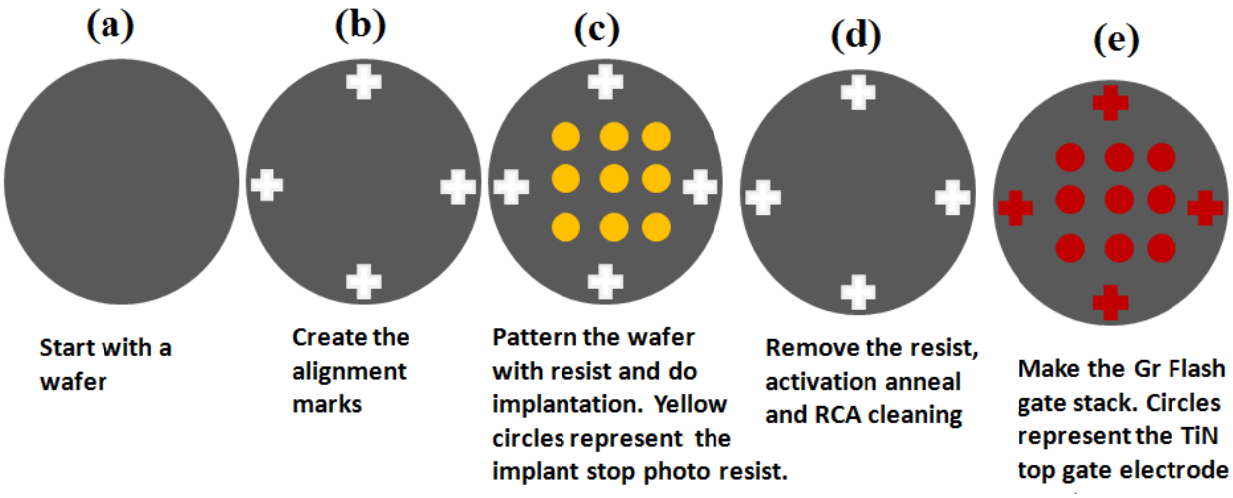


Figure 4.5: Process flow for fabricating pseudo source/drain rMLG CSL flash devices.

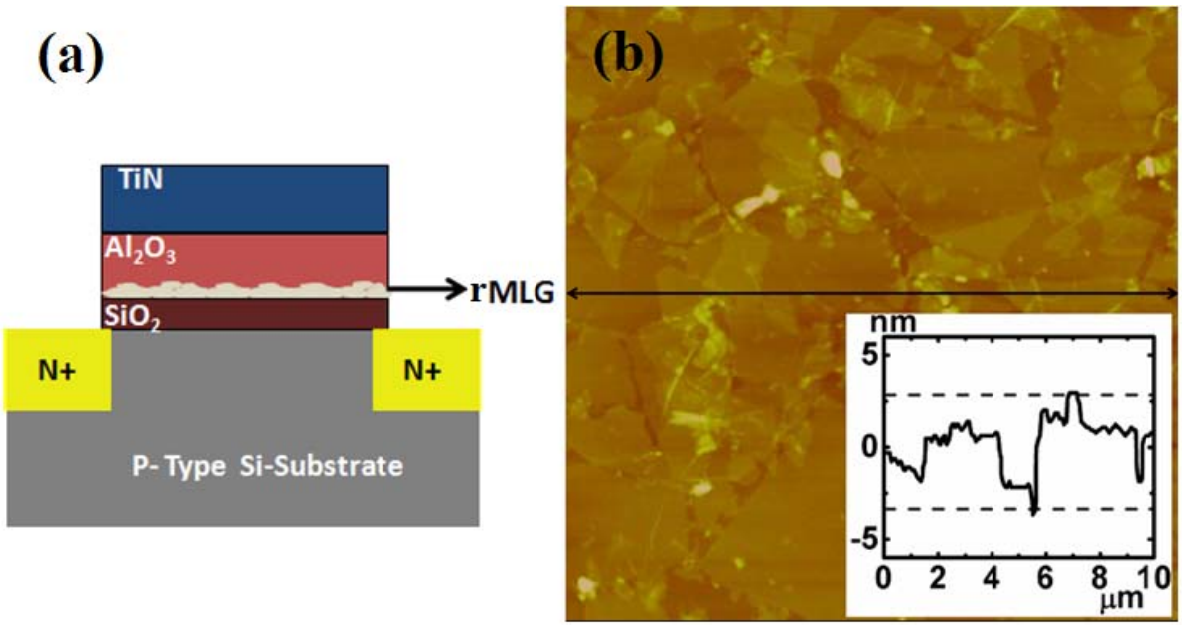


Figure 4.6: (a) Schematic of the cross-section of rMLG CSL flash MOS capacitor, (b) AFM image of the graphene sheets used in this work. Inset shows the height profile. Difference between the minimum and maximum level is about 5 nm.

4.3.2.2 Results and Discussion

CV and P/E transients of rMLG FG flash MOSCAPs were obtained using Agilent 4284 LCR meter and Agilent 4156C semiconductor parameter analyzer respectively. CV plots of the memory devices after different P/E operations at different voltages for 1 sec. pulse duration are shown in figure 4.7(a). Effective oxide thickness of the gate stack is 19 nm. P/E transients for different P/E voltages are shown in figure 4.7(c). The program transients shown in figure 4.3(c) were having a slow rise, e.g. a significant shift in the programmed V_{FB} was observed only after 1 msec. while in the present study significant change in V_{FB} is observed just after 1µsec. programming pulse. Hence, the use of n+ source around the MOSCAPs allows us to explore the P/E transients with fast programming and clear saturation. Fast and clear saturation of the program transients also indicates the reduced ballistic transport through rMLG FGs.

Total memory window (MW) due to electron and hole storage obtained for rMLG FG devices at 20 V P/E is 9.4 V while for control sample, the difference in the flatband voltage (ΔV_{FB})

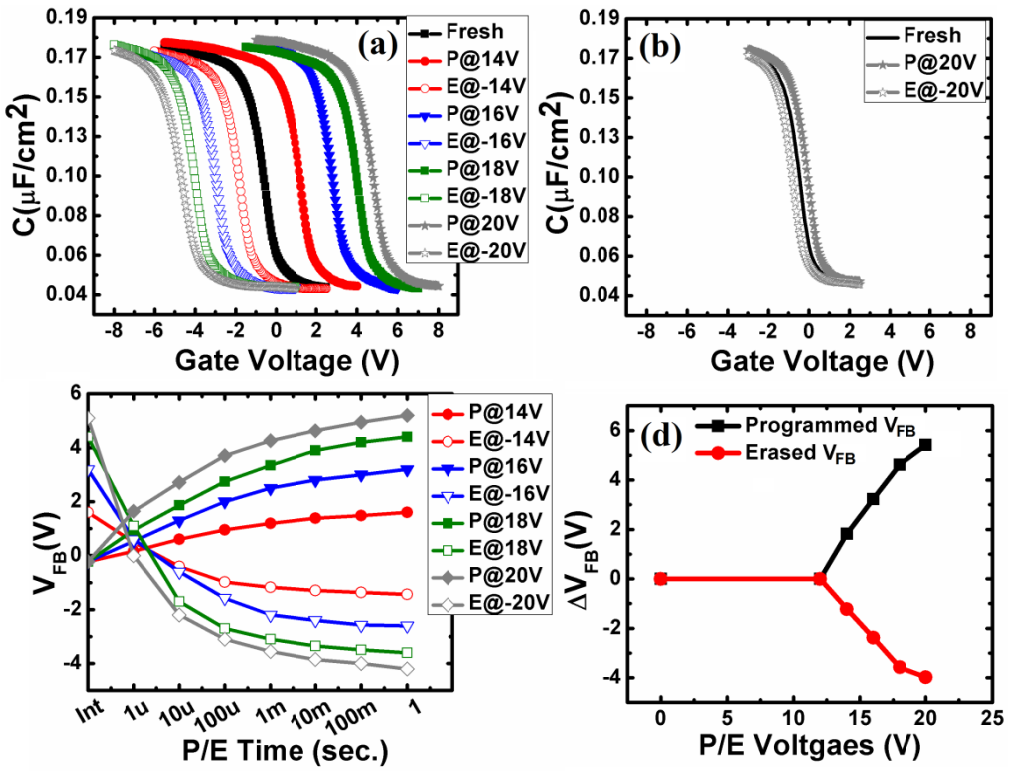


Figure 4.7: (a) CV plots of the rMLG flash devices after successive program (P) and erase (E) operation for 1 sec. "Fresh" indicates the CV of the device before P/E operation. (b) CV plots for control sample P/E at ±20V for 1sec. (c) P/E transients of rMLG flash devices. (d) ΔV_{FB} vs. P/E voltages.

between programmed and erased states under identical P/E conditions is only 0.78 V, figure 4.7(b). This confirms that significant charge storage is taking place in the rMLG sheets. The MW of 9.4 V is significantly higher than those reported in [36, 37]. A significant component of this memory window is the over-erase i.e. hole storage observed in these memory devices. In [37], Hong et al. have shown insignificant over-erase and hence low hole storage for multilayer graphene. This could arise from the nature of the graphene and the defects in it. Further, ambipolar nature of the graphene also suggests the possibility of hole storage in Gr FG devices. This over-erase is beneficial as it would enhance the overall MW.

An important point to discuss here is the slope of the ΔV_{FB} with P/E voltages plotted in figure 4.7(d). For a good quality high-k as a blocking dielectric, if the current through the blocking

dielectric is small, the slope of these curves should be close to 1 before saturation starts [28]. Here, the slope for the programmed ΔV_{FB} is 0.7, which is slightly less than 1. Dielectric constant (K) of the Al_2O_3 used in this work is 7.8 as calculated from the CV characteristics. This slope value can be improved by improving the quality and K value of the blocking dielectric.

Density of stored charge carriers (electrons and holes) (N_{Gr}) in rMLG sheets is calculated using equation (4.1) [42]. Density of stored electrons after 20 V, 1 sec. programming is 10^{13} cm^{-2} which is close to the DOS for MLG i.e. 4.4 x 10^{13} cm^{-2} eV^{-1} already calculated in section 4.2 of this chapter. Similar number of stored charge density in MLG sheets is recently reported with MOS_2 channel in memory devices [38]. Density of stored holes is less (7.8 x 10^{12} cm^{-2} eV^{-1}) than the density of stored electrons after similar erase conditions. This difference can be due to the small asymmetry in the DOS around the Fermi level for MLG sheets [104].

Retention of stored charges in rMLG CSL flash devices were tested at room temperature as well as at high temperatures as shown in figure 4.8(a). Extrapolation of the P/E states retention curves at room temperature and at 150°C to 10 years yields a MW of 6.9 V (74% charge retained) and 2.8 V (30% charge retained) respectively. At room temperature, charge loss for electrons is 21% while for holes is 24%. The slightly higher electrons loss in the present study compared to 8% reported in [36] can be due to the reduced thickness of the blocking dielectric used in this work. This retention figure can be improved further by improving the quality of the blocking dielectric.

Activation energies (E_a) for P/E states retention loss (at 20% V_{FB} decay), obtained from the slope of the linear fit of the Arrhenius equation [8], are 1.05 eV and 1.11 eV respectively (figure 4.8(b)). Activation energy depends on the type of CSL, on the quality of the surrounding dielectric and on the type of the charge loss mechanism. Activation energies reported for nitride based memory in [8] range from 1.01 eV to 1.88 eV depending on the nitride composition. Slightly low value of E_a (0.95 eV) is reported for nitride based memory in [105]. In a very recent work, activation energies were categorized based on different charge loss mechanisms [106]. The activation energies obtained in the present study agrees well with that reported in [106] for charge detrapping at high temperatures. However it is not clear if the fundamental process of

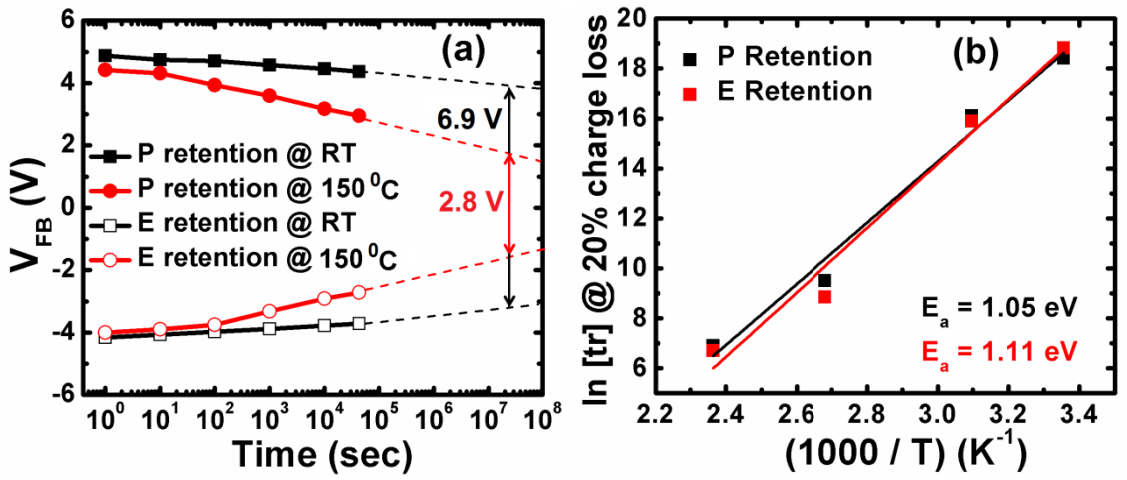


Figure 4.8: (a) Retention curves for rMLG FG flash capacitors obtained at room temperature and at 150°C. A minimum MW of about 2.8V after extrapolating the retention data at 150°C to 10 years is retained. (b) Arrhenius plot of programmed and erased state retention loss (taken at 20% V_{FB} decay).

charge loss in rMLG floating gate is the same as for the other floating gates. Cycling endurance, an important figure of merit to qualify the flash devices, is also performed and the devices could be programmed and erased for ~1000 cycles at 20 V, 1 msec. pulse duration. Endurance of these memory devices can be enhanced by improving the quality of the blocking dielectric. Some variability in the MW from device-to-device was also observed that is attributed to the variation in the rMLG thickness. A minimum of 4 V and a maximum of 15 V MW were also noticed for a few devices, however, the majority of them yielded a MW of ~ 9.4 V, which is reported in this thesis.

One important point to discuss here is how the quality of the Gr affects the memory performance. Graphene obtained by reduction of GO may have higher defect density compared to chemical vapor deposited (CVD) graphene, which may affect the density of stored charge carriers. Even though the density of stored electrons in the present study is same as reported for CVD graphene in the literature [37, 38], significant over-erase (i.e. hole storage) observed in our devices can be due to different nature of the graphene used in this work. However, this hypothesis needs further investigation and a conclusion can be drawn only after comparing the

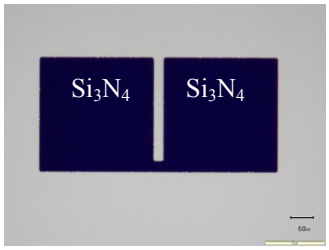
memory performance of different kinds of graphene under identical device processing conditions.

4.3.3 Graphene floating gate flash Transistor

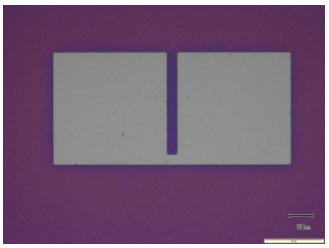
Though the issue of slow programming with rMLG flash MOS structure, observed in the figure 4.3(c), could be resolved with modified MOSCAP structure having implanted surrounding around the memory gate stack discussed in section 4.3.2 of this thesis. However, the device structures of the section 4.3.1 and 4.3.2 were the MOSCAPs with large area diameter (~ 80 μm). Some of the scaling issues discussed in chapter 2 of this thesis appear as the device dimensions scales down. With the extremely small gate area MOSCAPs, electrical characterizations are difficult because of the low capacitance values offered by these structures. These issues can be studied in rMLG FG flash transistor structure. Also, the CHE programming mechanism can only be demonstrated in a transistor structure. In order to study the performance of rMLG CSL flash devices in small area dimensions, flash transistor structures with different channel length were fabricated.

4.3.3.1 Device Fabrication

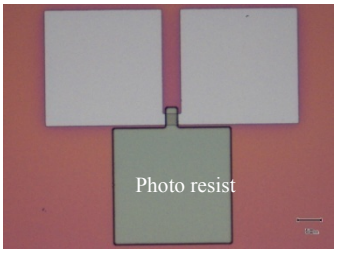
Long channel graphene flash transistors were fabricated using standard CMOS process flow with gate last step. After active area pattering, channel region was protected by thick photo resist and phosphorous implantation was done for creating n-type source/drain regions. After removing the implant stop photo resist, dopant activation anneal was performed at 950°C in Ar + O_2 ambient for 20 sec. After this, wafers were cleaned by standard RCA cleaning and same gate stack with rMLG as charge storage layers, as used in the MOS-capacitor structure of section 4.3.2, was grown and patterned by dry etching. Images of flash transistor structure at various stages of the fabrication process are shown in figure 4.9(a-f). More details of the device fabrication are given in the annexure B of this thesis.



(a): After active area pattering and Si_3N_4 at the active region.

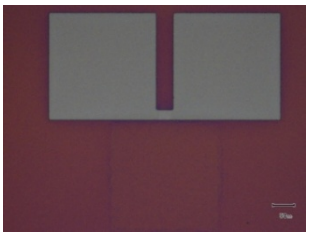


(b): After field oxide growth and removal of Si_3N_4 from the active region.

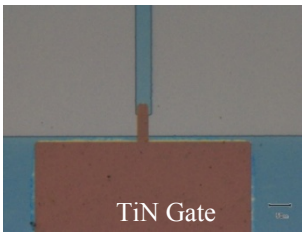


(c): After protecting the channel region from thick photo resist before implantation.

(d): After implantation.



(e): After removal of resist and dopant activation anneal.

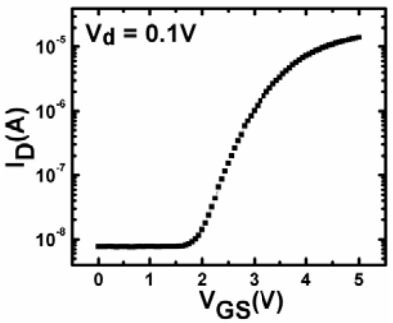


(f): After final device fabrication.

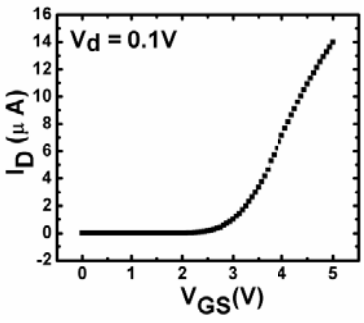
Figure 4.9: Optical micrograms after different steps in the fabrication of the flash transistor.

4.3.3.2 Results and Discussion

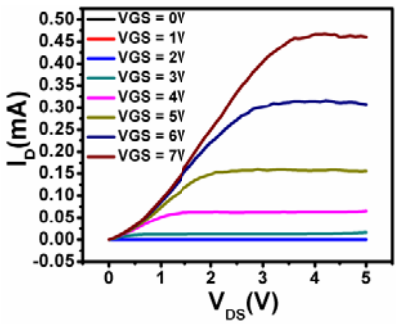
I_D-V_{GS} and I_D-V_{DS} characteristics of 3 µm channel length transistors; where I_D is the drain current, V_{GS} is the gate source voltage and V_{DS} is the drain source voltage, are shown in figure 4.10(a-c). From input characteristics (linear plot) threshold voltage of the devices is about 1.1 V. Figure 4. 10(d) and figure 4.10(e) show the I_D-V_{GS} characteristics of the 3 µm channel length devices after application of 14 V program/erase pulse for different P/E times starting from 1 µsec. to 100 msec. Figure 4.10(d) demonstrates that devices could be programmed successfully with the application of program pulse for different times. A program window of 3.43 V is obtained after 14 V, 100 µsec. programming pulse. However, these devices could not be erased fully as shown in figure 4.10(e). A very negligible erase is observed even after 100 ms of -14 V erase pulse. This behavior is observed with all the devices tested for program/erase mechanism. A possible reason for this slow erase could be the high WF of the rMLG sheets used in these



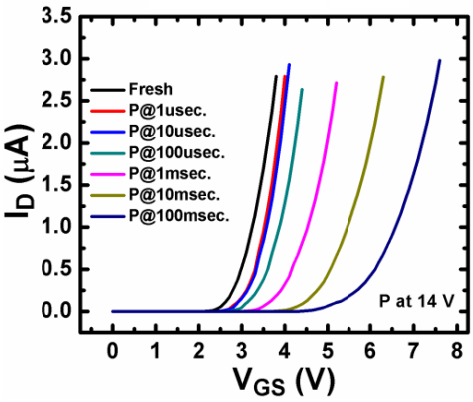
(a): I_D-V_{GS} (log plot) curve for 3 µm channel length device.



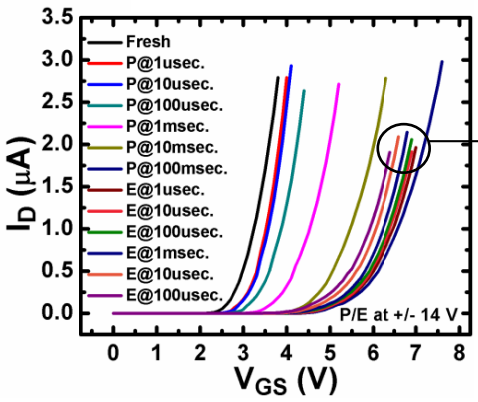
(b): I_D-V_{GS} (linear plot) curve for 3 µm channel length device.



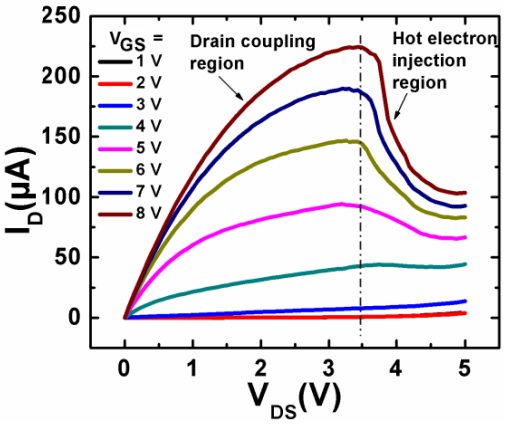
(c): I_D-V_{DS} curve for 3 µm channel length device.



(d): I_D-V_{GS} curve for 3 µm channel length device after 14V program pulse for different time.



(e): I_D-V_{GS} curve for 3 µm channel length device after 14V program / Erase pulse for different time.



(f): I_D-V_{DS} curve for 3 µm channel length device showing the CHE programming effect observed in some devices

Figure 4.10: Electrical characteristics of graphene charge storage flash transistor.

devices, (WF of the rMLG sheets is calculated in the chapter 5 of this thesis). A FG layer with high WF forms the deep potential well which can cause the erase saturation. To avoid the erase saturation in flash devices, hybrid FG concept is proposed as discussed in the chapter 2 of this thesis [28-30]. Further, charge tunneling from the gate electrode during erase operation due to the low work function of the TiN metal (4.5 eV) could also be the reason of slow or negligible erase in these particular memory devices. However, these hypotheses need further investigations.

An important point to discuss here is the observation of CHE programming in some of the memory devices. Figure 4.10(f) shows the I_D-V_{DS} characteristics of 3 μm channel length rMLG CSL memory devices with a signature of the CHE programming. In a FG memory transistor, I_D-V_{DS} characteristics of the devices are modified by the capacitive coupling between the drain and the FG of the devices. This effect is particular to the FG flash transistor only and is not seen in the conventional MOS transistors [42]. At higher drain voltages, hot electron injection to the FG causes the drain current to reduce. In these particular memory devices, drain current starts to reduce at about 3.5 V of V_{DS}.

4.4 Summary of the chapter

In this chapter, reduced graphene is successfully demonstrated as a charge storage layer with three different kinds of test structures viz. MOSCAP, MOSCAP with implanted surroundings and flash transistors. A large memory window of 9.4 V is obtained with pseudo S/D test structure. Data retention at room temperature as well as at elevated temperatures is demonstrated. A 6.8 V remnant MW at room temperatures and 2.8 V remnant MW at 150°C after 10 years is extrapolated. Endurance of the devices was tested, however, devices showed poor endurance and large device to device variability.

Transistor structures were fabricated with the aim to study some of the scaling issues at reduced device dimensions. These memory transistors could be programmed successfully to a 3.4 V program window at 14 V programming voltages, however, these devices could not be erased successfully. High WF of the rMLG sheets or tunneling of the charge carriers from the top gate

electrode could be the reason for erase saturation. A signature of the hot electron programming in the I_D-V_{DS} characteristics of these memory devices is demonstrated.

Though the desired aim, i.e. the study of the rMLG FG memory performance at small device dimensions using flash transistor structure could not be achieved due to issues in erasure and low throughput of these devices. Nevertheless, the process flow for working FG flash transistor fabrication is established and some of the interesting observations (like CHE programming) are reported.

It is worth discussing here that quality of the graphene would affect the memory performance. In the present study, multilayer graphene used for charge storage purpose is obtained by thermal reduction of graphene oxide whose properties (e.g. work function) would depend on several parameters like extent of reduction, thickness etc. Therefore, more careful experiments should be performed to study the effect of different parameters on the memory performance. Large device to device variability in the memory window, observed in these devices, can be controlled by controlling the quality (WF, thickness, defects) of the graphene sheets.

Chapter 5

Experimental determination of work function of reduced graphene oxide: reduced graphene oxide as a gate electrode in MOS devices

For any material intended to be used as charge storage layer, knowledge of its work function (WF) is very important. Memory performance strongly depends on the well depth formed by charge storage layer which in turn depends on its WF. In this chapter, WF of the graphene layers used as a CSL in the chapter 4 of this thesis is experimentally determined using it as a gate electrode in a MOS structure. This experiment not only yields the WF of reduced graphene oxide (rGO) layers but also provides us the opportunity to explore rGO as a gate electrode for complementary metal oxide semiconductor (CMOS) devices, as metal gate electrodes are proposed for future CMOS technology [3]. However, the incorporation of metals in the front end of device processing is a concern because of the metallic contaminations introduced by metals into the surrounding dielectrics [107, 108]. The interfacial reaction between metal electrodes and the dielectric could also lead to fermi level pinning [109]. Graphene (Gr), because of its two dimensional sheet like structure, extraordinary high conductivity and high thermal stability, can be an interesting candidate as a gate material in MOS devices. In this chapter, first a short literature review on the different methods used to extract the graphene work function with a discussion on available literature on the graphene as a gate electrode in MOS structures is presented. Then, the experimental results obtained with rGO as a gate electrode under different contact metals are discussed.

The results presented in this chapter were reported in Applied Physics Letters vol. 100, p. 233506, 2012, ACS Applied Materials and Interfaces (DOI: 10.1021/am404649a) and at International Conference on Emerging Electronics (ICEE) 2012 held in IIT Bombay, India.

5.1 Graphene Work function: A brief literature review

WF of graphene with different number of layers is studied by a number of techniques like Kelvin probe force microscopy (KPM) [110-113], photoelectron emission microscopy [98] and by measuring the open circuit voltage in graphene based solar cells [114]. WF of graphene is also a subject matter of intense modeling efforts [115-116]. Theoretically, the doping of graphene by metals modeled by Giovannetti et al. [115] predicted that the WF of graphene would increase (decrease) when it is doped with metals with higher (lower) work function than that of graphene due to electron transfer for Fermi level alignment. However Pi et al. [117] experimentally demonstrated that the graphene would be n-doped when in contact with Pt, contradicting the predictions in [115]. In other theoretical studies as well [118], charge transfer from high work function metal to monolayer graphene (i.e. n- type doping) is discussed, contradicting the observed experimental results [119]. However, most of the experimental studies performed so far converge to a monotonous increase in the Gr WF as the number of layers increase and saturate to some value for more than 4-6 layers. WF of Gr on SiC increases from 4.2 eV for single layer graphene (SLG) to 4.6 eV for more than 4 layers as measured by KPM method [98] while for Gr on Si, same trend of increasing WF from a value of about 4.35 eV for monolayer to 4.61 eV for more than 4 layers is reported [114]. WF of Gr on SiO_2 as measured by KPM method also increases from 4.57 eV for SLG to 4.69 eV for bi-layer graphene (BLG) [113]. Theoretical studies of Ziegler et al. [116] predicted that graphene WF depends on the carrier concentration and doping type for 1 to 3 layers of graphene and WF is insensitive to doping for thicker layers. It is clear that the WF of graphene for any application should be determined in the material environment in which it is being used.

Recently, Park et al. [120] proposed the use of monolayer graphene synthesized by CVD as the gate metal in charge trap memory devices. Since graphene is extremely thin, a metal for making contacts for measurements is required. Park et al. had deposited Nickel on a small area on top of Gr for this purpose. They have compared the performance of such devices with devices containing TaN as gate metal. Significant improvement in the memory performance is demonstrated for graphene gated devices compared to TaN gated devices [120]. The

performance improvement is attributed to the higher work function of the monolayer graphene on Al_2O_3 compared to the WF of TaN and reduced mechanical stress in the underlying gate dielectric. In another work by the same group [121], monolayer graphene as a gate electrode is integrated with high dielectric constant gate materials and improvement in the device reliability is attributed to the reduced mechanical stress in the high-K dielectrics. In [122], again by same group, monolayer graphene work function is evaluated with Cr/Au, Ni and Pd top metal contacts. It is reported that With Cr/Au and Ni contact, the work function of graphene is pinned to that of the contacted metal, whereas with Pd or Au contact the work function assumes a value of 4.62 eV irrespective of the work function of the contact metal [122]. Authors have no explanation for these different behaviors of graphene work function under different metals.

In summary, even though Gr can be an interesting candidate for gate applications in MOS devices, no study of the WF of Gr as a function of the number of layers using MOS test structures is reported in available literature. In this chapter, we systematically study the electrical performance of reduced multi layer graphene (rMLG) gate electrodes with varying number of rGO sheets under different contact metals *viz*. Pt, Ir, Al and TiN. We find that WF of the graphene gate electrode devices increases as the number of graphene layers increase and saturates for thick graphene sheets.

5.2 Work function determination of reduced graphene oxide as a Gate Electrode in MOS with different contact metals

5.2.1 MOSCAP Fabrication and electrical characterization with rGO/TiN contacts metal

5.2.1.1 Device Fabrication

To estimate the WF of graphene of different thicknesses, MOS capacitor structures with and without rGO sheets were fabricated on p-type Si substrate (resistivity of 1-5 Ω.cm) with 7.6 nm (± 0.2 nm) thermally grown SiO_2 as the gate dielectric. Complete device fabrication procedure is

depicted in figure 5.1(a-e). Schematic of the cross section and the top view SEM of as fabricated MOS capacitors are shown in figure 5.2(a) and figure 5.2(b) respectively. Thickness of the SiO_2 films was measured using a spectroscopic ellipsometer (Sentech, SE 800). After the growth of gate oxide, graphene oxide (GO) obtained by Hummer's method was deposited on gate dielectric with drop cast technique and it is thermally reduced to Gr at 500°C in Ar ambient for 1hr. The same procedure of thermal reduction of GO into Gr at the device locations was demonstrated earlier [123]. Raman spectra of GO and thermally reduced graphene sheets are shown in figure 5.2(c). Thermal reduction results in an enhancement in the 2D band at ~ 2700 cm^{-1} and a suppression of the defect related D band at ~ 1350 cm^{-1}. Further, ratio of D band intensity to G band intensity decreases, from 0.97 for GO to 0.40 for reduced graphene. The G band is seen to shift towards lower wave number from 1590 cm^{-1} for GO to 1580 cm^{-1} for reduced graphene.

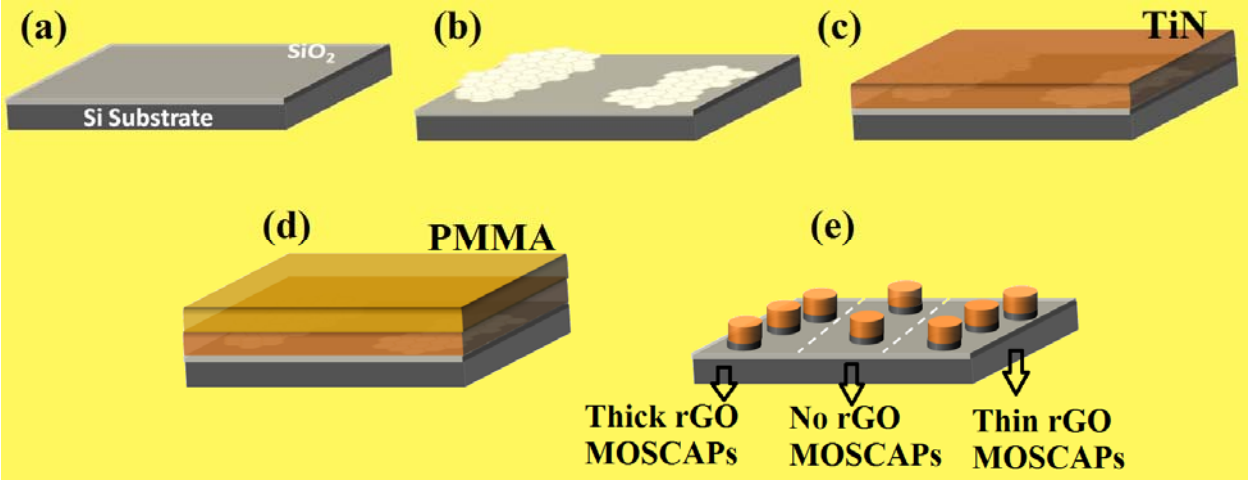


Figure 5.1: Complete procedure of MOSCAPs fabrication with different rGO thicknesses under TiN contact metal, (a) starting Si substrate with thermally grown SiO_2. (b) Si/SiO_2 substrate with different number of rGO layers after thermal reduction of GO. A region on the same sample is left without rGO for control sample. Location of the different thicknesses of the rGO is identified in SEM equipped with lithography technique and the co-ordinates are noted for further MOSCAP fabrication. (c) deposition of 80 nm TiN as a top capping metal. (d) PMMA is spin coated and the EBL is performed at the already noted positions in step (b). (e) Final top view of as fabricated MOSCAPs. Regions of the MOSCAPs with different rGO thicknesses are demarked for better understanding of the procedure.

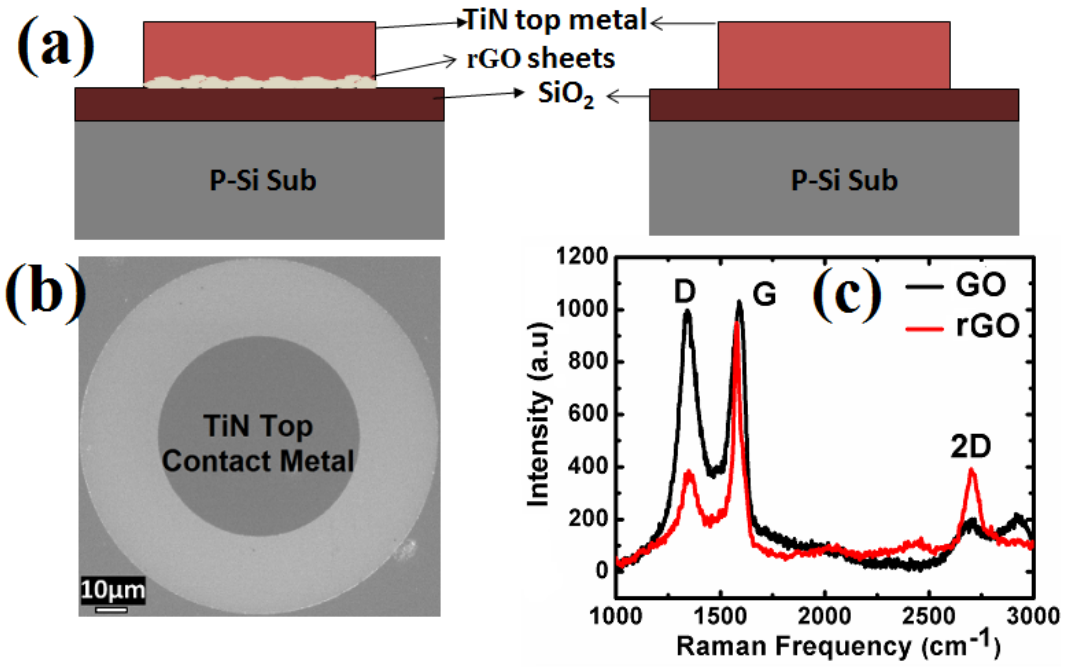


Figure 5.2: (a) Schematic of the devices with and without rGO sheets sandwiched between TiN gate electrode and SiO_2 gate dielectric. (b) Top view of as fabricated MOS-capacitor. (c) Raman Spectra of graphene oxide (GO) and graphene obtained after thermal reduction of GO.

These observations attest to the reduction of GO into graphene like character [124, 125]. More discussion on the I_D/I_G ration in Raman spectra is given in the annexure A of this thesis. Different ranges of rGO thicknesses were obtained by varying the number of drop casts i.e. one drop cast of liquid at some locations and multiple drop casts at other locations. To make the devices with varying number of rGO layers under the contact metal, extensive scanning electron microscopy (SEM) was performed in an electron beam lithography system (RAITH 150TWO) to identify the location of various thicknesses of graphene (very thin, moderate thin, and thick) on the gate dielectric. The co-ordinates of these locations were stored for further top gate patterning. SEM and atomic force microscopy (AFM) (Vecco Digital Instruments, Nanoscope IV) images of very thin, moderate thin and thick layer rGO sheets are shown in figure 5.3(a-f). It is seen that the thinnest films obtained are not continuous and as the thickness increases, closed films of multi layer rGO are obtained. Figure 5.3(g - i) shows the corresponding AFM sectional analysis of the height profile of rGO on SiO_2. Figure 5.3(g) shows the sectional analysis for the

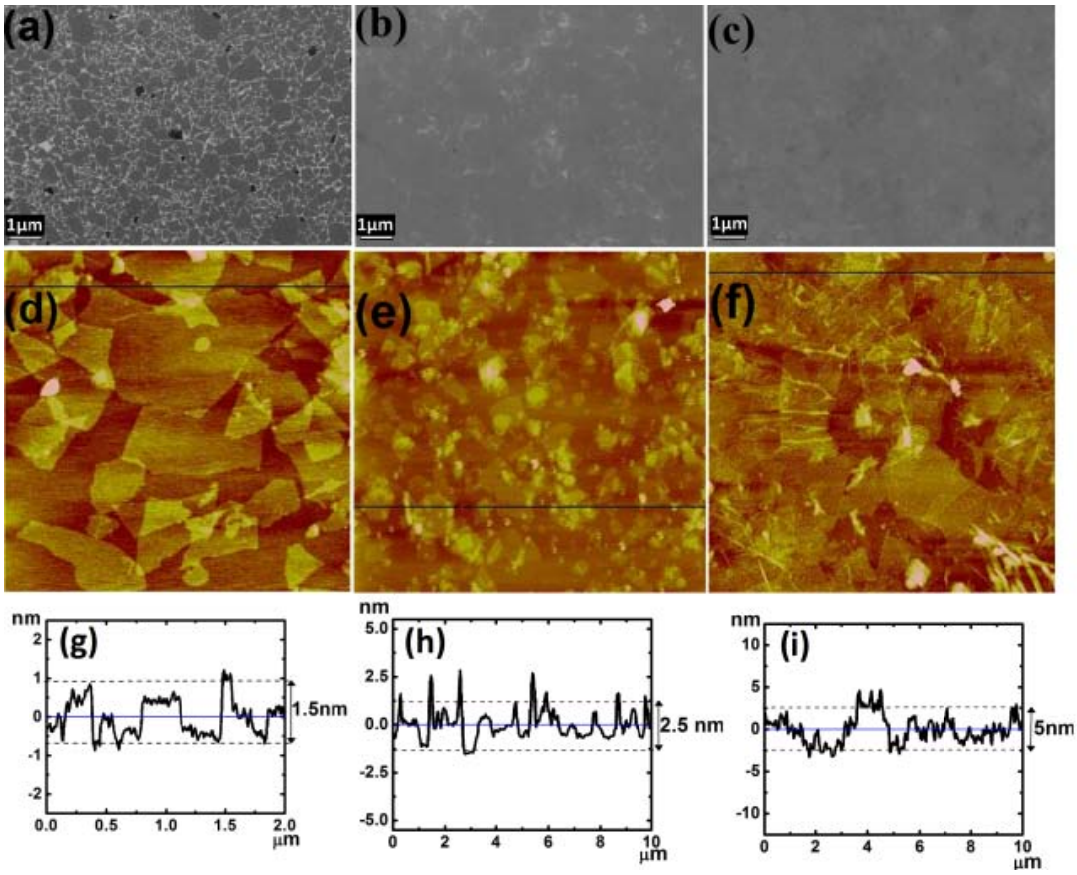


Figure 5.3: SEM images of different layers of rGO: (a) 1-3 layers of rGO sheets, (b) 3-5 layers of rGO sheets, (c) more than 5 layers of rGO sheets. (d-f) AFM images of rGO sheets corresponding to figure (a-c) respectively. (g-i) AFM sectional analysis showing the thickness of rGO sheets corresponding to figure (d-f) respectively.

thinnest films. The discrete topographical levels indicate different layers of graphene. The minimum height differences between discrete levels in the figure are in the range of 0.5 nm, which is consistent with literature data for rGO obtained by reduction of GO [125]. From the height profile of AFM images, thickness of very thin graphene sheets is about 0.5 - 1.5 nm, i.e. 1-3 layers (figure 5.3(g)), thickness of moderate thin graphene sheets is about 1.5 - 2.5 nm, i.e. 3-5 layers (figure 5.3(h)) while thickness of thick graphene sheets is more than 2.5 nm, i.e. more than 5 layers (figure 5.2 (i)). In figure 5.3(g – i), dashed lines show the low and high levels in the line scans. The spikes beyond the dashed lines observed at few sites correspond to the local variation in the rGO thickness or the number of rGO layers. The width of these spikes is about

50 nm. The area occupied by these spikes would be less than 5% of the overall gate dimensions of the MOS-capacitors and hence will not affect the characteristics of the capacitors significantly. TiN as top contact electrode was deposited by sputtering process (Applied Materials ENDURA). PMMA was spin coated on the sample and circular patterns of TiN dots of 80 μm diameter are patterned at the already noted coordinates by electron beam lithography system (RAITH 150Two). This procedure affirmed the presence of graphene sheets of different known thickness under the TiN contact metal. Finally, dry etching of the TiN was performed in Cl_2 + BCl_3 plasma and graphene sheets were etched by O_2 plasma in a reactive ion etching system (Applied Materials Etch Centura). On the same wafer, a region without rGO sheets was also identified using the same SEM system for control devices ($Si/SiO_2/TiN$).

5.2.1.2 Results and Discussion

Both kinds of MOS-capacitors ($Si/SiO_2/TiN$ and $Si/SiO_2/rGO/TiN$) are electrically characterized using Agilent 4284 LCR meter and Agilent 4156C semiconductor parameter analyzer. CV and conductance - voltage (GV) plots of these devices are shown in figure 5.4(a) and figure 5.4(b) respectively. As the number of rGO layers increases, CV curves shift rightwards. This is due to the increasing WF of the graphene with increasing number of layers. Variation of the flatband voltage with increasing number of rGO layers in rGO/TiN electrode is shown in figure 5.3(c). Work function for the gate electrode was calculated using the equation [126]

$$\boldsymbol{V_{FB} = \Phi_{ms} - \frac{Q_{OX}}{C_{OX}}} \text{ …….. (5.1)}$$

$$\text{Where, } \boldsymbol{\Phi_{ms} = \Phi_m - \Phi_s} \text{ …….. (5.2)}$$

where V_{FB} is the flatband voltage, Φ_{ms} is the WF difference between the gate electrode and the semiconductor substrate, Q_{ox} is the oxide charge and C_{OX} is the oxide capacitance. WF for TiN electrode as calculated from the given equation is 4.45 eV while for rGO/TiN stack, it has a value of 4.6 eV, 4.74 eV and 4.91 eV for thin rGO, moderate thick rGO and thick rGO respectively. In these calculations a substrate doping of 3 x 10^{15} cm^{-3} is used (calculated from

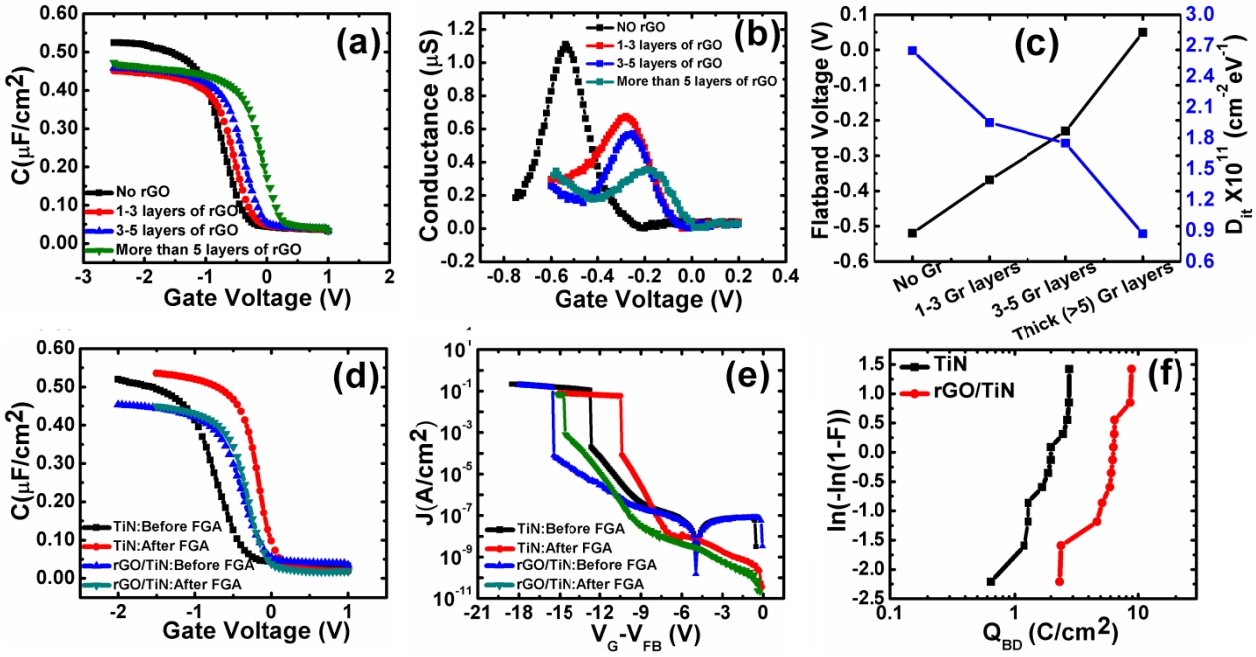


Figure 5.4: Comparison of CV and GV plots respectively for TiN gate electrode and rGO/TiN gate electrode devices with different numbers of rGO layers. (c) Flatband voltage and the interface state density with increasing number of rGO layers in rGO/TiN electrode devices. (d, e) CV and current density-voltage (JV) plots of SiO_2/TiN and SiO_2/rGO/TiN devices before and after FGA at 420°C for 20 minute with 3-5 layers of rGO, (f) Charge to breakdown behavior for only TiN and rGO/TIN gate electrode devices.

C_{min} value of high frequency CV curve) and since all the MOS capacitors are made on same wafer under identical process conditions, it is assumed that the effect of oxide charges on the flatband voltage is same for all devices [126]. We have corrected the work function values after including the contribution made by oxide charges in a separate experiment and found the error in work function values with and without correcting for the charges are negligible. The results are presented and explained in the section 3.3 of this chapter. The WF of TiN obtained in present study is within the range of the WF values reported in the literature for PVD deposited TiN on SiO_2 [127]. The values of Gr WF on SiO_2 reported in the literature vary between 4.45 eV for SLG to about 5 eV for MLG depending on the type of Gr, measurement technique and contact metal used [110, 113, 116]. The work function values obtained in this work, for different number of rGO layers on SiO_2 are in close agreement with the WF values reported in the literature for

few layer (4.57 eV for SLG, 4.69 eV for bi-layer graphene) [113] and multilayer (4.93 eV to 4.95eV for MLG) exfoliated graphene on SiO_2, all obtained by KPM method [110]. In [120], work function of monolayer graphene is reported to be about 5.2 eV-5.3 eV on Al_2O_3 which has a higher dielectric constant than that of SiO_2. These results suggest that the role of TiN in a rGO/TiN stack is to provide contact whereas the rGO layers decide the WF. This may be an important technique for the WF tuning of the gate electrode in MOS devices provided that a methodology to deposit Gr films with control on the number of layers is developed.

Another observation in CV plots is that the accumulation capacitance is higher for SiO_2/TiN stack compared to the SiO_2/rGO/TiN, irrespective of the thickness of the rGO. Effective oxide thickness (EOT) calculated from the accumulation capacitance for TiN electrode is 6.6 nm while the EOT for rGO/TiN electrode is 7.5 nm to 7.3 nm. The EOT of the gate dielectric for TiN electrode is about 1 nm less than the value obtained from ellipsometry while EOT for rGO/TiN electrodes is less by only about 0.1 to 0.3 nm. This indicates the consumption of the dielectric by some chemical reaction at the SiO_2/TiN interface. Any chemical reaction at dielectric/metal interface is undesirable as it could lead to Fermi level pinning [109] and causes contamination in the thin gate dielectrics and thus degrade the device reliability [107, 108]. The reduction of EOT with metal gate electrode via chemical reaction at dielectric/metal interface is also reported in literature [128]. In our study, the chemical reaction at the dielectric/metal interface ceases to happen even when 1-3 layers of rGO are sandwiched between metal and the dielectric.

Inclusion of rGO between metal and dielectric also improves the interface quality as clearly seen in GV plots shown in figure 5.4(b). Conductance peak in the GV plot indicates the amount of interface states at the Si/SiO_2 interface and the peak position occurs around the flatband voltage of the devices. With the increasing number of rGO layers in the device, GV peak height reduces. The peak position also shifts towards right and this is in accordance with the increasing WF of the rGO with increasing number of layers. Interface states calculated as per the equation given in [129] are plotted in figure 5.4(c). Values of the interface states reduce from 2.7×10^{11} $cm^{-2}eV^{-1}$ for only-TiN electrode to a value of 8×10^{10} $cm^{-2}eV^{-1}$ for the devices with more than 5 layers of rGO under TiN. In the present study, TiN is deposited by sputtering process which invariably causes some damage to the Si/SiO_2 interface. An improvement in the Si/SiO_2 interface quality

suggests that graphene, because of its superior mechanical properties, shields the Si/SiO_2 interface from sputtering damage.

Robustness of the gate stacks to thermal annealing is also investigated for these devices. CV characteristics of the $Si/SiO_2/TiN$ and $Si/SiO_2/rGO/TiN$ before and after forming gas anneal (FGA) at 420°C for 20 minutes are shown in figure 5.4(d). Improved slope of the CV curves after FGA signifies the reduced interface state density [126]. The CV curve of the $Si/SiO_2/TiN$ stack has shifted to the right after FGA, indicating an increase in negative fixed charges in the oxide or an increase in the WF or both [127]. This is likely due to an increased interaction between SiO_2 and TiN at the temperature used for the FGA. However the $Si/SiO_2/rGO/TiN$ is seen to be robust against FGA, except for the (beneficial) improvement in interface state density.

The rGO gate electrode also results in much improved breakdown characteristics of the devices. Figure 5.4(e) compares the current density versus gate voltage plots for TiN electrode and rGO/TiN electrode before and after FGA. The rGO/TiN electrode devices have lower leakage compared to TiN electrode devices. This improvement in the breakdown characteristics of the devices is independent of the number of rGO layers in the rGO/TiN electrode. Average breakdown field for TiN and rGO/TiN devices are 17 MV/cm and 20.5 MV/cm respectively before FGA, while after FGA, breakdown fields for TiN and rGO/TiN electrode devices are 16.1 MV/cm and 19.7 MV/cm respectively. Further, charge to breakdown (Q_{BD}) is an important figure of merit for qualifying the dielectric reliability [130]. We performed Q_{BD} measurements on TiN and rGO/TiN electrode devices under identical constant current stress. Figure 5.4(f) shows the Q_{BD} characteristics for both kinds of devices. Average value of Q_{BD} for $Si/SiO_2/rGO/TiN$ devices is about 6 C/cm^2 while Q_{BD} value for $Si/SiO_2/TiN$ devices is only 2 C/cm^2.

5.2.2 Exact determination of rGO work function after correcting for the charges in the gate stack

In the previous experiment (section 5.1.1), we have demonstrated the shifts of upto 0.5 V in flatband voltage of MOS capacitors with $SiO_2/rGO/TiN$ gate stack compared to SiO_2/TiN gate

stack, under the assumption that the charges in the dielectric are negligible and that the work function of the Si substrate is a constant and also the flatband voltage shift is equal to the work function shift. However it is of particular importance to justify this assumption in the present study as the processing of graphene, especially its integration in the MOS system could potentially introduce charges in the gate dielectric. A thickness series method is widely used to eliminate the effect of charges in the determination of work function from MOS devices [126]. Equation 5.1 can be re written as

$$\boldsymbol{V_{FB} = \Phi_{ms} - \frac{Q_{OX}}{C_{OX}} = \Phi_{ms} - \frac{(Q_{OX})t_{OX}}{\varepsilon_0 \varepsilon_{OX}}} \quad \ldots\ldots\ldots (5.3)$$

where ε_0 is the absolute permittivity of free space, ε_{OX} is the relative permittivity of the SiO_2 and t_{OX} is the thickness of SiO_2 calculated from the CV data. Other terms are already explained in equation 5.1. In this technique, MOS devices are fabricated with at least three different dielectric thicknesses. Assuming that the charges in the dielectric are independent of thickness and that the charges are at or near the interface, the work function difference between the gate and the semiconductor can be determined by plotting V_{FB} versus t_{OX}. The intercept of the linear fit of the data on the V_{FB} axis gives the work function difference. The slope of the plot can be used for estimating the total charge. The goodness of the fit is a test of the assumptions stated above.

5.2.2.1 Device Fabrication

MOSCAPs were fabricated with similar experimental procedure as discussed in the previous section but with three SiO_2 thicknesses i.e. 7.3 nm, 10 nm and 14 nm and only reduced multilayer GO under TiN contact metal as a gate electrode. Multilayer rGO was chosen for this experiment to avoid any flat band voltage variation due to different number of graphene layers.

5.2.2.2 Results and Discussion

Figure 5.5(a) shows the high frequency CV curves of the MOS capacitors with and without rGO sheets and three different SiO_2 thicknesses in each case. The flatband voltage and oxide thickness were determined from the CV and these are plotted in figure 5.5(b). It is seen that the V_{FB} versus

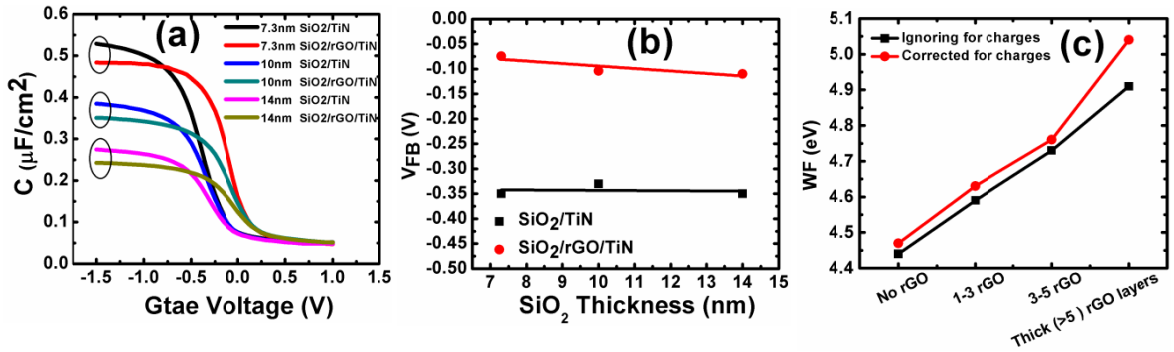


Figure. 5.5: (a) CV curves for different thickness of gate oxide (7.3 nm, 10 nm and 14 nm: all measured by ellipsometry) with TiN and rGO/TiN as a gate electrode material, (b) variation in the flat band voltage of the devices with different oxide thickness (calculated from CV). Intercept and slope for SiO_2/TiN devices (black line) are -0.34 V and -4.3 x 10^{-4} V/nm respectively while intercept and slope for SiO_2/rGO/TiN devices (red line) are -0.048 V and 0.0047 V/nm respectively. (c) Work function with increasing number of rGO layers in rGO/TiN electrode devices after ignoring and correcting for charges.

oxide thickness data in either case can be fitted fairly well using a straight line. Intercepts of the straight line on the Y axis for SiO_2/TiN stack and SiO_2/rGO/TiN stack are -0.34V (with a root mean square error of 0.036 V) and -0.048 V (with a root mean square error of 0.028 V) respectively. From equation (5.3), these intercepts are related with the work function difference between gate electrode and the semiconductor. From the minimum of capacitance value (C_{min}) in CV curves, semiconductor doping is calculated as 2.8 x 10^{16} cm^{-3} which in turn gives the semiconductor work function (Φ_s) value of 4.98eV. Using this value of Φ_s and equation (5.2), gate electrode work function for SiO_2/TiN stack and SiO_2/rGO/TiN stack are 4.64 eV and 4.93 eV respectively. The difference in these values of the work function suggests that multilayer rGO/TiN stack has higher work function than the TiN metal. Slope of the straight lines for SiO_2/TiN stack is -4.3x 10^{-4} V/nm (with a root mean square error of 0.0038 V/nm) and -0.0047 V/nm (with a root mean square error of 0.0026 V/nm) for SiO_2/rGO/TiN stack. Total oxide charges calculated using equation (1) are 1.48 x 10^{-9} Coulomb. cm^{-2} (9.2 x 10^9 cm^{-2}) and 1.6 x 10^{-8} Coulomb. cm^{-2} (10^{11} cm^{-2}) for SiO_2/TiN stack and SiO_2/rGO/TiN stack respectively. The flat band voltage shift due to these values of the oxide charges will be only 0.0028 V for SiO_2/TiN

stack and 0.03 V for SiO_2/rGO/TiN stack for the 7.3 nm thick oxide work function of multilayer rGO sheets after neglecting the contribution of fixed oxide charges is 4.90 eV, 4.88 eV and 4.87 eV respectively for the three thicknesses (i.e. 7.3 nm, 10 nm and 14 nm) of SiO_2, while it is 4.93 eV after correcting for the contribution of the charges. Neglecting the second term (charge) in equation (5.1) causes an error of 30 mV in the work function calculation for the smallest oxide thickness used. In the previous experiment (section 5.1.1), the flat band voltage shift of the rGO gate electrode devices with varying number of rGO layers was calculated. In that calculation, contribution of charges to the flat band voltage was neglected. Now, the work function of different rGO layers is recalculated by taking into account the flat band voltage shift caused by charges as obtained in the present experiment. The dielectric and interface charges are assumed to be identical as in the present case since the devices were fabricated in the same batch using identical processing. However the EOTs values obtained for the experiment reported in section 5.1.1 were slightly different but close to the lowest value shown in figure 5.5. The work functions extracted from CV by ignoring the contribution of the charges and by correcting for the charges are shown in figure 5.5(c). Table 5.1 summarizes the difference in work function values for different number of rGO layers when the contribution of charges is included and when this contribution is corrected for. The error in work function estimation by ignoring the contribution of charges is also listed.

Table 5.1: Work function (WF) values for the different layers of rGO with and without ignoring the contribution of charges and the error values in the WF estimate when charge is ignored.

No. of rGO layers	WF (eV) (corrected for charges)	WF (eV) (ignoring charges)	Error when the charges are ignored (meV)
No graphene	4.47	4.44	30
1-3 layers of rGO	4.63	4.59	40
3-5 layers of rGO	4.76	4.73	30
> 5 layers of rGO	5.04	4.91	130

It is evident from the two values obtained that the error when the charge contribution is ignored is not significant. The work function tuning obtained by different number of layers reported in the previous experiment is confirmed. However correction for charges results in a slightly higher work function tuning range than reported in previous experiment.

5.2.3 Device Fabrication and electrical characterization of rGO gate electrode devices with Pt/TiN, Ir/TiN and Al/TiN metal contacts

5.2.3.1 Device fabrication

The p-type <100> wafer with resistivity of 1-5 Ω cm was RCA cleaned, and SiO_2 of different thicknesses (6.3 nm, 8.5 nm, 10 nm and 15 nm) were thermally grown in high purity oxygen ambient at 850°C and subsequently annealed at 900°C for 15 minutes in high purity nitrogen ambient. Type of the rGO and deposition method is same as reported in the previous section. Top contact metals (Pt, Ir, Al) were deposited by sputtering at sufficiently reduced power to prevent any plasma damage in the rGO sheets as well as in the gate dielectric. The thickness of Pt and Ir metals was kept at 20 nm and 80 nm thick and TiN was deposited on top. Such a structure would be easier to etch for device fabrication than a structure with thick Pt or Ir. Other device fabrication details are same as adopted in section 5.2.1. Complete device fabrication procedure is depicted in figure 5.6(a-f). Extensive use of the SEM combined with EBL system facilitates the making of the devices on the different thicknesses of rGO sheets. Thickness of the thick, moderate thick and very thin rGO sheets is about 7 nm, 2.5 nm and 1.7 nm respectively as obtained by cross section high resolution transmission electron microscopy (HRTEM) images (figure 5.7(a-c)). Different layers of the MOS gate stack are clearly visible in figure 5.7. Top metal layers (Pt/TiN) are appearing as dark contrast in the TEM image and rGO sheets as layered sheets. For TEM sample preparation, thickness of the Pt metal was reduced to 5 nm. The samples were prepared by mechanical polishing followed by ion milling. The stack with 20 nm Pt did not withstand the sample preparation procedure due to the poor adhesion of thick Pt films on the SiO_2 dielectric. Thin metal films also have the poor adhesion, however when capped with thick TiN, these could sustain the sample preparation procedure. The thicknesses of the rGO sheets

obtained by HRTEM analysis are consistent with the thicknesses obtained by the height profile of the AFM images for thick, moderate thick and very thin rGO sheets respectively. In this work, very thin, moderate thick and thick rGO means 1-3 layers, 3-5 layers and more than 5 layers respectively, unless otherwise stated.

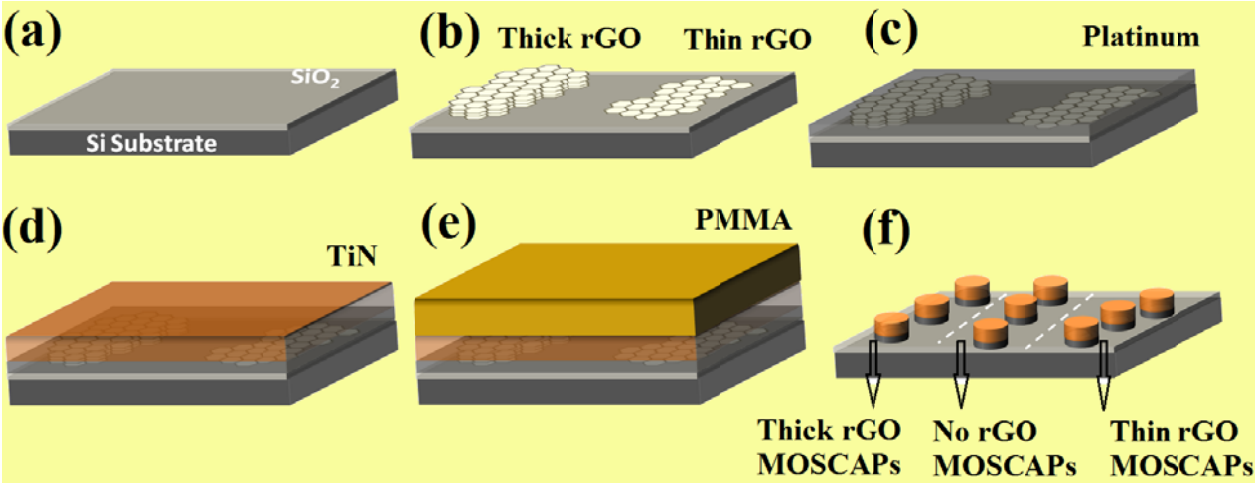


Figure 5.6: Complete procedure of MOSCAPs fabrication with different rGO thicknesses under Pt/TiN contact metal (a) starting Si substrate with thermally grown SiO_2. (b) Si/SiO_2 substrate with different number of rGO layers after thermal reduction of GO. A region on the same sample is left without rGO for control sample. Location of the different thicknesses of the rGO is identified in SEM equipped with lithography technique and the co-ordinates are noted for further MOSCAP fabrication. (c), (d) deposition of 20 nm Pt and 80 nm TiN as a top capping metal respectively. (e) PMMA is spin coated and the EBL is performed at the already noted positions in step (b). (f) Final top view of as fabricated MOSCAPs. Regions of the MOSCAPs with different rGO thicknesses are demarked for better understanding of the procedure.

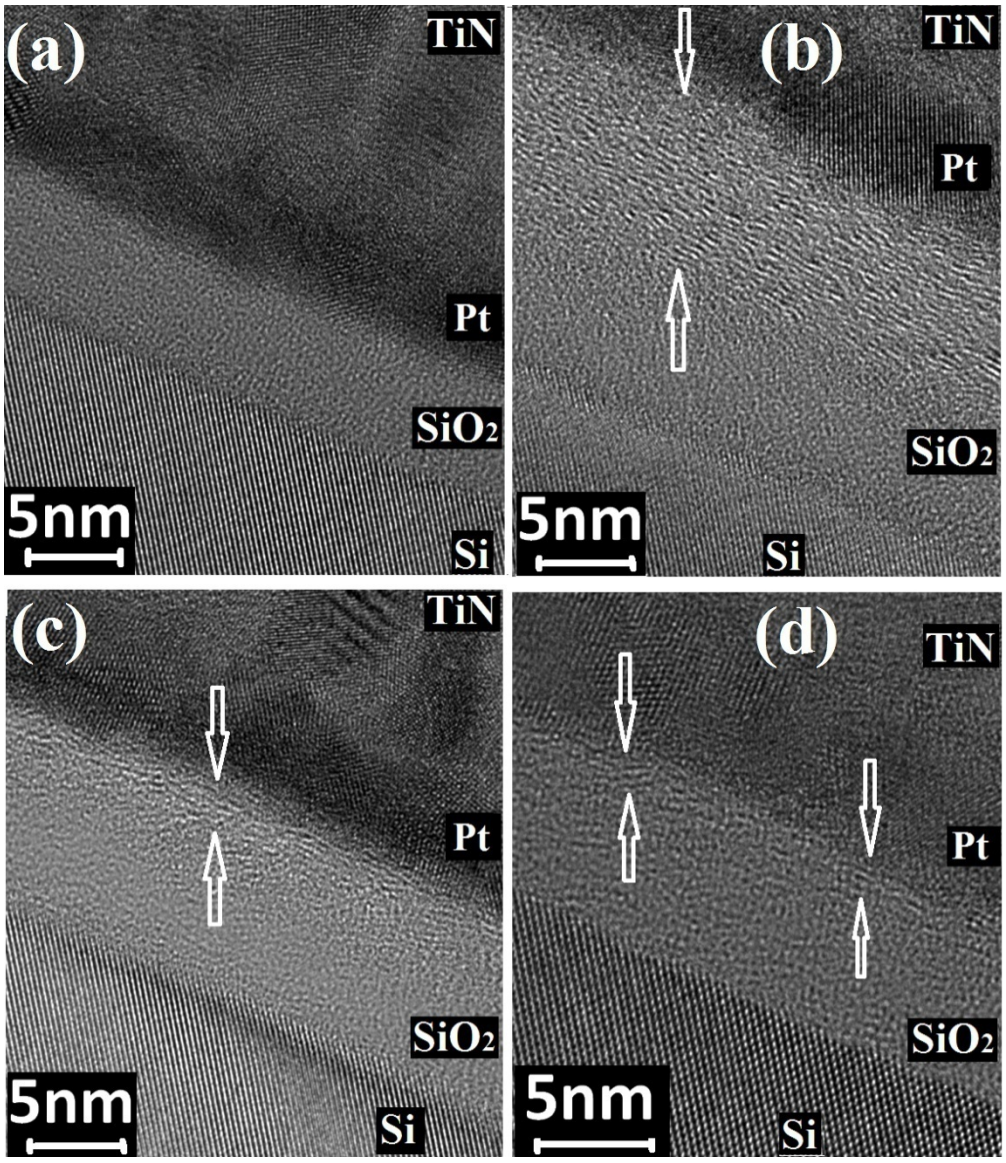


Figure 5.7: Cross section HRTEM images of fabricated MOSCAPs. (a) Without any rGO under Pt/TiN contact metal (b) Thick rGO (~20 layers) under Pt/TiN contact metal. (c) Moderate thick rGO (3-5layers) under Pt/TiN contact metal (d) Thin rGO (1-3 layers) under Pt/TiN contact metal.

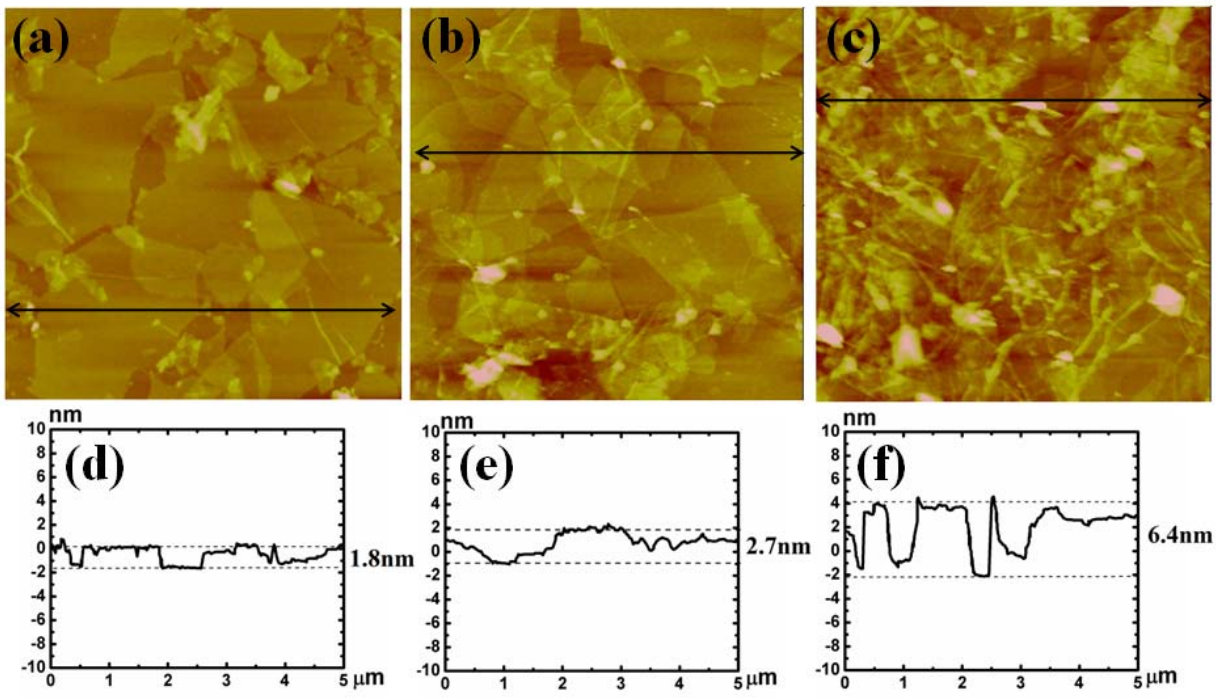


Figure 5.8: AFM images of typical (a) thin, (b) moderate thick and (c) thick rGO sheets on SiO_2 used to fabricate the MOSCAPs, (d-f) height profile of the AFM images shown in (a-c) respectively.

5.2.3.2 Results and Discussion

MOS capacitors (with and without rGO sheets) were electrically characterized using Agilent 4284 LCR meter and Agilent 4156C semiconductor parameter analyzer. CV plots for rGO gate electrode devices with different capping metals (Pt/TiN, Ir/TiN and Al/TiN) are shown in figure 5.9. CV curves for the corresponding control sample (without rGO sheets) are also shown in the same figures. From the minimum capacitance value (C_{min}) of the CV plot, doping concentration of the semiconductor substrate is calculated as 1.5 x 10^{16} cm^{-3} which in turn gives semiconductor WF as 4.96 eV. Flat band voltage for $Si/SiO_2/Pt/TiN$ stack is 0.05 V while it decreases to -0.6 V and -0.2 V for very thin and moderate thin rGO sheets respectively and again increases to 0.29 V for thick rGO layers below Pt/TiN contact metal. Equation (1) implies that the shift in flatband voltage can have contributions from charges in the oxide and the interface. To accurately

determine the WF of a gate electrode in MOS system, the contributions of the charge should be properly taken into account. Estimation of the oxide charges becomes even more important in the present study as the processing of rGO, especially its integration in the gate stack of a MOS structure could potentially introduce charges in the gate dielectric. In order to extract the value of WF of various layers of rGO under contact metal, thickness series experiment with four SiO_2 thicknesses (6.3 nm, 8.5 nm, 10 nm and 15 nm) has been performed. The flatband voltage vs. t_{OX} data obtained from these experiments is plotted in figure 5.9(c). WF of Pt/TiN, very thin rGO/Pt/TiN and thick rGO/Pt/TiN gate electrodes, extracted using equation (5.3), are 5.04 eV, 4.35 eV and 5.28 eV respectively. This shows that the WF of the gate electrode can be varied from 4.35 eV to 5.28 eV by varying the number of rGO layers under Pt/TiN contact metal as shown in CV plot of figure 5.9(a). Total oxide charge densities, calculated using the slope of V_{FB} vs. t_{OX} plot and equation (5.3), are 9 x 10^{10} cm^{-2}, 8.4 x 10^{10} cm^{-2} and 1 x 10^{11} cm^{-2} for Pt/TiN, thin rGO/Pt/TiN and thick rGO/Pt/TiN gate electrodes respectively. In conductance vs. gate voltage (GV) plot of a MOS structure, conductance peak (G_{Peak}) occurs close to flatband voltage and hence the variation in the flatband voltage with different number of rGO layers under Pt/TiN contact metal can also be verified by noting the position of conductance peak as shown in figure 5.8 (b). G_{Peak} position shifts towards positive gate voltage from a value of -0.61 V for very thin rGO to 0.3 V for thick rGO, which is consistent with the flatband voltage values, obtained from the CV curves. Decreasing height of the conductance peak in GV plot signifies the reduced interface state density at the gate dielectric/semiconductor interface. Graphene, because of its superior mechanical properties, protects the gate dielectric from any plasma damage during top metal deposition (by sputtering in the present study), which results in an improved gate dielectric/semiconductor interface for graphene gate electrode devices.

The results of bi-directional (+ve as well –ve) modulation of V_{FB}, (and hence the gate electrode WF) are also obtained with Si/SiO_2/rGO/Ir/TiN stack. For calculating the WF of different number of rGO layers under Ir/TiN contact metal, V_{FB} shift is directly equated to the Φ_{ms}. Here it is assumed that the amount of oxide charges for Ir/TiN contact metal devices would be same as for Pt/TiN contact metal as all the devices were processed under identical process conditions. These oxide charges would shift the V_{FB} by only ~ 0.04 V for Ir/TiN and rGO/Ir/TiN gate

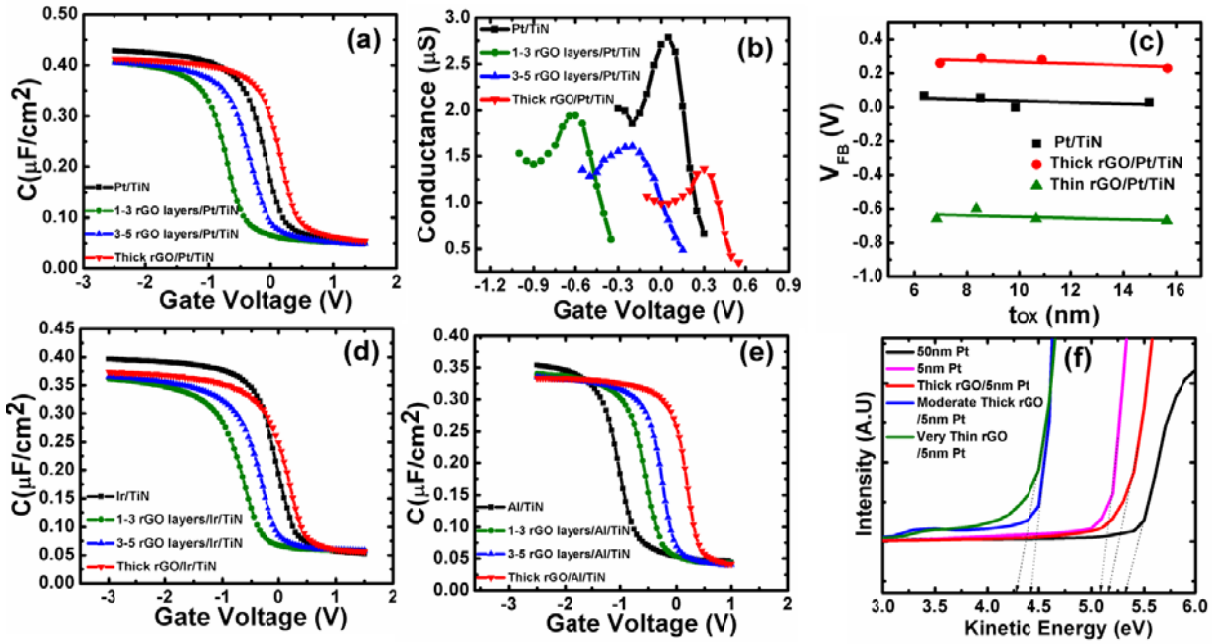


Figure 5.9: (a), (b), (c) CV, GV and V_{FB} vs t_{OX} plots of MOSCAPS with different thicknesses of rGO under Pt/TiN contact metal respectively. (d) and (e) are CV curves of MOSCAPs with different thickness of rGO under Ir/TiN and Al/TiN contact metals respectively. (f) UPS spectra for 50 nm thick Pt film on SiO_2, 5 nm thick Pt film on SiO_2 and rGO with different layers on SiO_2 with 5 nm Pt deposited on it. Intersection of the dotted line with arrow on the X-axis gives the WF for different materials.

electrode devices for 10 nm SiO_2 thickness. Hence the assumption of neglecting the contribution of oxide charges to the flatband voltage is justified. WF values for different rGO layers under Ir/TiN contact metal are 5.06 eV for no rGO and 4.4 eV, 4.7 eV and 5.21 eV for very thin, moderate thick and thick rGO layers respectively. We have also carried out the same CV analysis with Al, which has a low WF value. With Al/TiN contact metal, a minimum WF of 4.46 eV with very thin rGO layers and a maximum WF of 5.16 eV for very thick rGO layers is obtained.

The difference in the WF values for thin rGO layers under different capping metals (Pt, Ir, Al) lies only within 0.12 eV. The spread in the data is similar for thick rGO sheets. This observation suggests that the flatband voltage (and hence the WF of gate electrode) is determined mainly by

the number of rGO layers, and the capping metals used in this work do not play any significant role in it. A possible explanation for this is discussed later in this chapter.

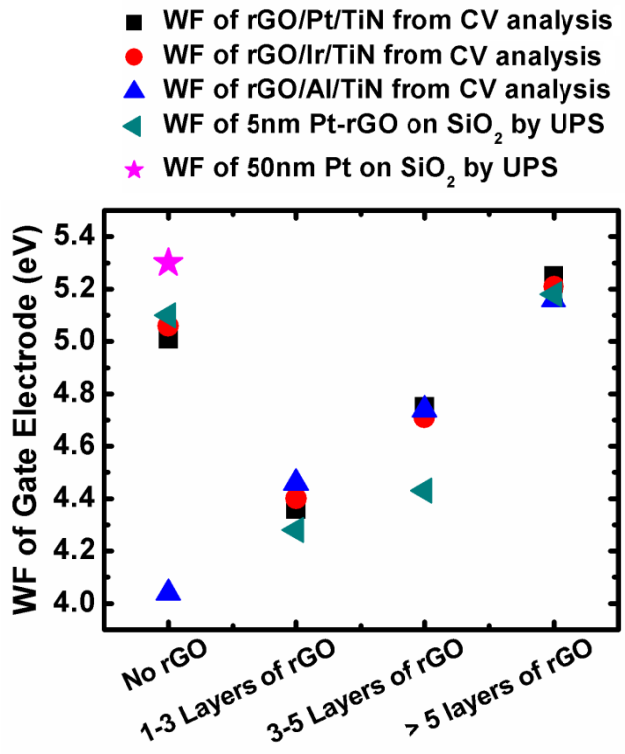


Figure 5.10: Modulation of the gate electrode WF with different number of rGO sheets under Pt/TiN, Ir/TiN, and Al/TiN capping metals. For comparison, WF values obtained from UPS analysis for 5 nm Pt-rGO on SiO_2 and 50 nm Pt on SiO_2 are also plotted.

It is worth discussing here that in the present case maximum WF difference (low value to high value) for thin and thick rGO is about 0.9 eV with Pt/TiN metal while the corresponding difference was 0.5 eV when TiN contact metal was used in section 5.2.1. This difference can be due to the fact that in the present work, the metals Pt, Ir, Al are pure metals and are deposited by sputtering in pure Ar plasma while the TiN metal was deposited in a reactive environment in the presence of N_2 and Ar plasma. Metals like Pt, Ir and Al physisorb on Gr and thus may not affect the Gr properties. However, the interaction of TiN with Gr, especially for TiN deposited by

reactive sputtering, is not clear. We anticipate that the metal deposition conditions or the different interactions of TiN with graphene play some role in this.

Table 5.2: Modulation of WF of top gate electrode with varying number of rGO layers with different top contact metals. Values in the bracket (bold) in the Pt/TiN column are the WF values obtained for rGO/5nm Pt system using UPS.

No. of rGO layers	WF (eV) with different top contact metals		
	Pt/TiN	Ir/TiN	Al/TiN
No rGO	5.04 (**5.08)**	5.06	4.04
1-3 layers of rGO	4.35 **(4.28)**	4.4	4.46
3-5 layers of rGO	4.75 **(4.43)**	4.71	4.74
Thick (> 5) layers of rGO	5.28 **(5.16)**	5.21	5.16

The trend of WF variation of rGO layers under Pt metal was also verified by ultraviolet photoelectron spectroscopy (UPS). For UPS analysis, rGO with different thicknesses were deposited on 4 nm SiO_2 and a 5 nm Pt layer was deposited by sputtering on top of them. Pt thickness has been deliberately limited to 5 nm for UPS analysis as the energy of the photoelectrons is very low and usually very thin films are deposited for WF measurements by UPS [131]. A thick Pt (50 nm) sample on SiO_2 is also prepared for reference. UPS spectra are collected using a Thermo VG Scientific Multilab 2000 Photoelectron Spectrometer, which is equipped with a high photon flux He gas discharge source (modes: He I 21.2 eV and He II at 40.8 eV). UPS spectra supports the thickness dependent WF of the rGO/Pt system (figure 5.9 (f)). WF values obtained by CV measurements for different rGO thickness with different contact metals and by UPS measurements for rGO/Pt system are plotted in figure 5.10. In the present study, WF values by UPS measurements are lower than that obtained by CV measurements. UPS technique is known to detect the lower limit of the WF of a system [132]. Hence, WF values obtained by UPS measurements are in general lower than the standard WF values [131- 133]. Even though the WF values obtained from the two methods differ marginally, the trend of WF shift with varying number of rGO layers under Pt metal is same in both the measurement

techniques. WF of different rGO sheets under different contact metals is tabulated in table 5.2 of this chapter.

A possible explanation for the experimental observations of the WF tuning reported above is as follows. WF of the graphitic structure increases when the graphene layers increase from monolayer to multilayers as measured by different techniques like photoelectron spectroscopy and Kelvin probe force microscopy (KFM) methods [98, 113, 116]. Further, in a very recent theoretical study [134], WF of the rGO is calculated using molecular dynamics simulations and density functional theory. This theoretical study reports that the WF of rGO increases with increasing oxygen concentration in the different groups like carbonyl, hydroxyl and epoxy attached to it. Structure of graphene oxide is discussed in annexure A of this thesis. A range of WF values of GO with different oxygen concentrations i.e. 4.4 eV - 6.8 eV for carbonyl group, 4.35 eV - 5.6 eV for epoxy group and 4.25 eV - 4.95 eV for hydroxyl group is reported. WF values obtained in our experiments (minimum 4.35 eV for very thin layer rGO and maximum 5.28 eV for thick layer rGO) are well within the range of these theoretical WF values.

When the thermal reduction of GO is performed, thin layer GO can be expected to reduce more readily and loose its oxygen faster as compared to that of the thick GO layers. In multilayer GO, the layers below the top layers can be expected to reduce less effectively. Hence the oxygen concentration would increase from thin rGO to thick rGO. To confirm this hypothesis, Fourier Transform Infrared Spectroscopy (FTIR) and XPS analysis is performed on different thicknesses of GO sheets before and after thermal reduction.

Fourier Transform Infrared Spectroscopy (FTIR) Analysis

FTIR analysis is performed with a Bruker 3000 Hyperion Microscope with Vertex 80 FTIR system, spectral resolution 0.2 cm^{-1}) and is shown in figure 5.11(a-b). For FTIR analysis, GO sheets of different thicknesses were drop casted on Si/SiO_2 substrate and annealed at 550°C for 1 hr in Ar ambient. Peaks corresponding to different functional groups *viz*. hydroxyl (C-OH), epoxide (C-O-C), carboxyl (COOH) and ketonic (C=O) attached to the GO are marked in FTIR spectra shown in figure 5.11(a-b). The peaks are identified based on [135]. For very thin GO,

after thermal reduction, the peak heights corresponding to different functional groups are almost negligible compared to that for moderate thick and very thick rGO. Slightly slow reduction of peak height for ketonic (C=O) and carboxyal groups (COOH) for moderate thick and very thick rGO is due to the high binding energy of oxygen with these functional groups [134]. This FTIR analysis clearly indicates that under identical reduction conditions, removal of different functional groups depends on the GO thickness. As a result, oxygen concentration would be low in very thin rGO sheets (due to efficient removal of different functional groups) and would increase with rGO thickness.

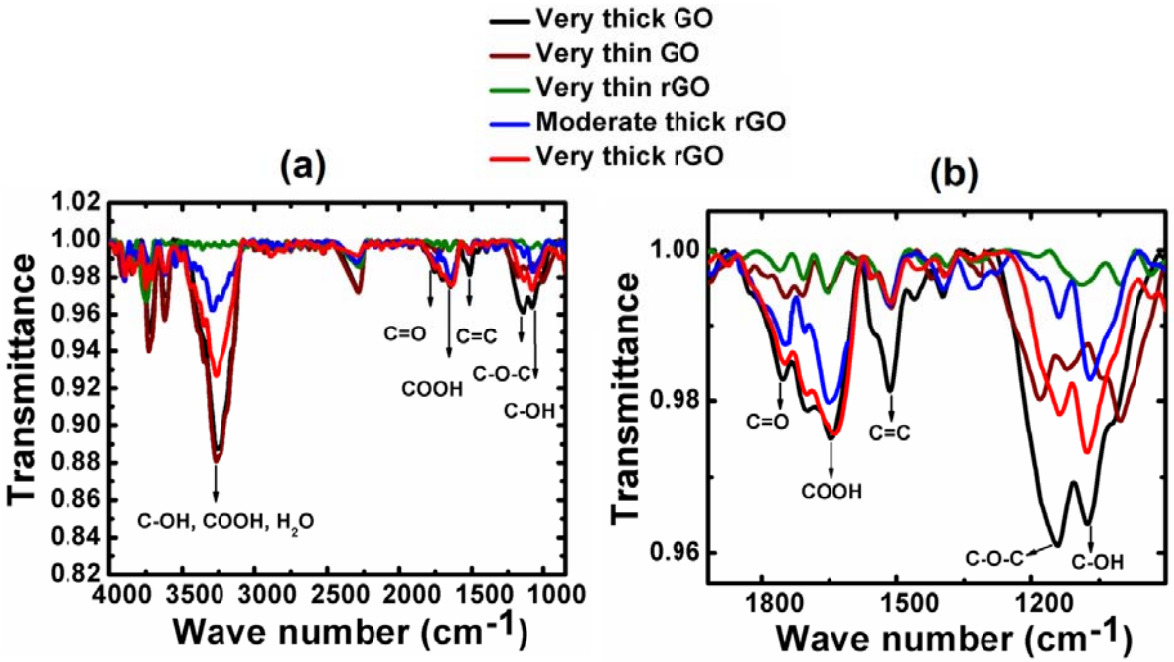


Figure 5.11: FTIR spectra, (a) full range, and (b) lower range, of GO and rGO obtained after thermal reduction of GO at 550°C for 1hr in Ar ambient. Peaks corresponding to different functional groups attached to GO are assigned as per reference [135].

X-Ray photo electron spectroscopy analysis

To quantify the oxygen concentration in different thickness of rGO layers, XPS analysis is performed for very thin and thick GO sheets on Si/SiO_2 substrate before and after thermal reduction. The XPS spectra of C1s peak is shown in figure 5.12(a-d). Peaks corresponding to

different functional groups are marked in the figure 5.12(a-d). The peak assignment is as per [136- 138]. For GO, figure 5.12(a, c), peak at binding energy of about 284.8 - 285.1 eV is assigned to sp^2 carbon – carbon (C=C or C-C) bonds while the peaks at 285.76 eV, 286.78 eV, 287.55 eV and 288.81 eV are assigned to C-OH, C-O-C, C=O and COOH groups respectively. After the thermal reduction, peak positions slightly decrease to lower binding energies which is consistent with other reported studies [136]. After the thermal reduction, peaks corresponding to C-OH and C-O-C for very thin rGO cannot be deconvoluted and are assigned as a single peak at 285.70 eV while for very thick rGO sheets, both peaks can be assigned separately. However, the intensity of these peaks decreases significantly after thermal reduction.

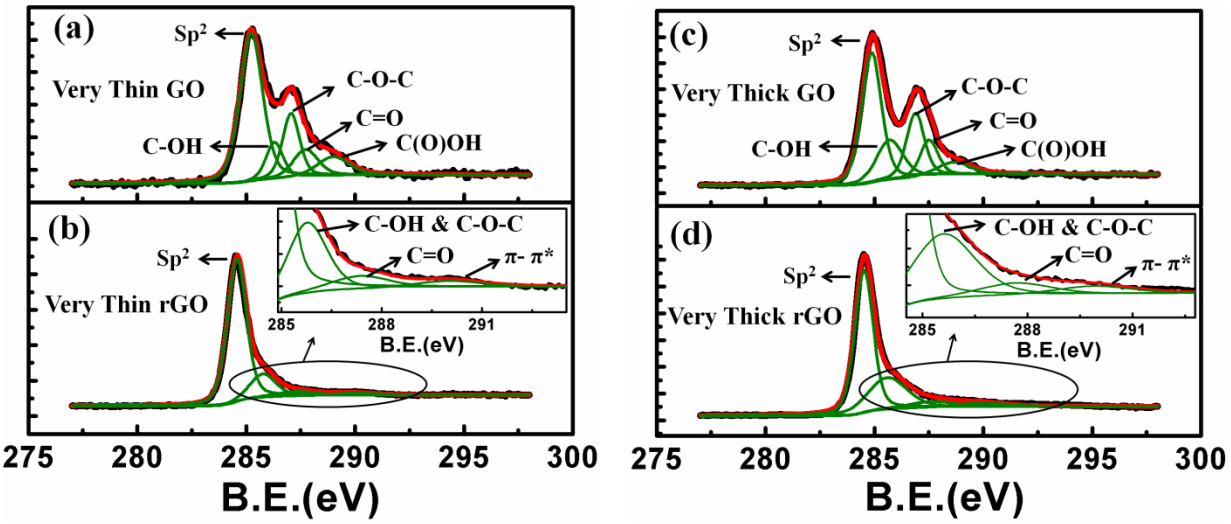


Figure 5.12: C1s peaks of XPS spectra for very thin and very thick GO before annealing (a and c) and after annealing (b and d).

Figure 5.12(a-d) clearly demonstrates that peaks corresponding to different oxygen containing functional groups decreases faster for very thin rGO sheets as compared to those in very thick rGO sheets. Oxygen concentration, obtained after dividing the area of all peaks corresponding to oxygen containing functional groups to the total area of C1s peak, in very thin and very thick GO (figure 5.12(a and c)) is about 46% and 55% while it reduces to 21% and 34% for very thin and very thick GO sheets after thermal reduction (figure 5.12(b and d)). This variation in oxygen concentration, according to [134], would cause the WF of the rGO layers to vary from low value

to high value. The observed WF modulation in the present study can be attributed to the combined effect of the two factors, namely a thickness dependent WF and the oxygen concentration dependent WF.

5.2.4 Oxygen concentration dependent work function calculation

To further confirm that change in oxygen concentration would change the WF of rGO sheets, thick GO layers were reduced at different temperatures ranging from 450°C to 750°C and the MOSCAPs were fabricated with rGO under Al gate electrode. Other device fabrication details were same as mentioned previously in section 5.2.3.1 of this chapter. Thick GO layers were chosen to exclude the layer dependent WF. XPS spectra of thick GO layers reduced at different temperatures are shown in figure 5.13(a-d). Remnant oxygen concentration in the rGO sheets and its WF under Al contact metal is plotted in figure 5.14(a). Change in GO WF with oxygen concentration is also plotted in figure 5.14(b).

As the reduction temperature increases, peak heights corresponding to different oxygen containing functional groups and hence oxygen concentration decreases. For GO, oxygen concentration is about 55% while it decreases to 37%, 34%, 28% and 20% after thermal treatment at 450°C, 550°C, 650°C and 750°C temperatures respectively. WF of the rGO sheets also depends on the reduction temperature as determined from the CV measurements of Si/SiO2/rGO/Al MOSCAPs. As the reduction temperature increases (from 550°C to 750°C), WF of the rGO/Al stack decreases. WF of the rGO/Al stack reduces from 5.14 eV for 35% O_2 concentration (reduced at 550°C) to 4.42 eV for 20% O_2 concentration (reduced at 750°C) as shown in figure 5.14(b). This experiment clearly demonstrates that oxygen concentration plays a significant role in determining the reduced graphene oxide work function.

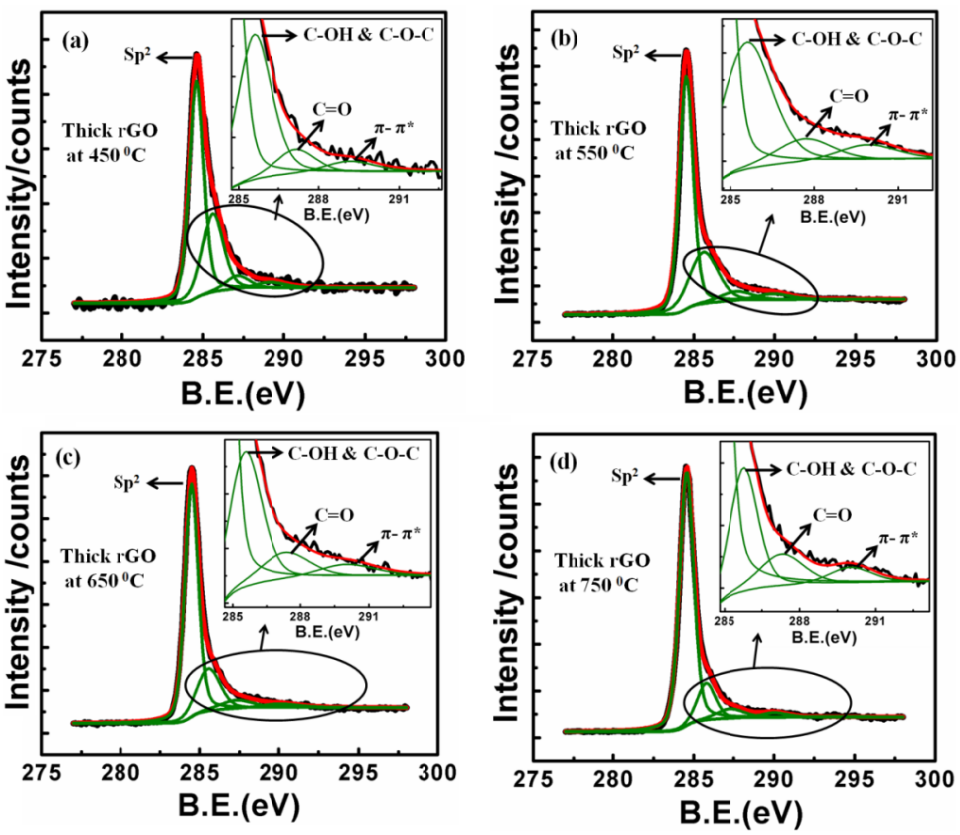


Figure 5.13: C1s peak in XPS spectra for thick GO sheets reduced at different temperatures, (a) at 450°C, (b) 550°C, (c) at 650°C and (d) at 750°C.

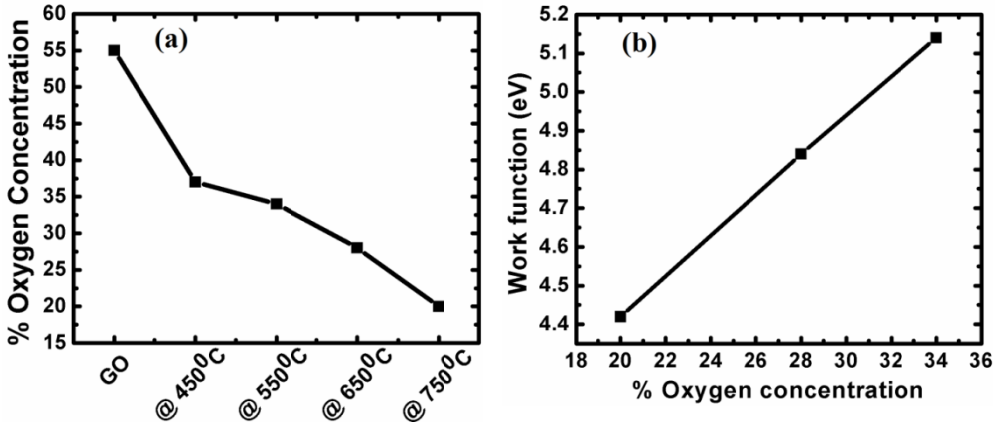


Figure 5.14: (a) Percent oxygen concentration of the rGO sheets with thermal reduction at different temperatures, (b) WF of the rGO sheets with oxygen concentration.

It is important to isolate the contribution of the two factors in the rGO WF i.e. thickness dependent WF and the oxygen defect dependent WF. However, with GO as a starting material, it may be misleading as any number of rGO layers (thin or thick) would have some amount of oxygen which would contribute to the rGO WF. This can only be achieved by using the pristine graphene as a starting material and then introducing some known amount of oxygen in this before performing the electrical measurements.

5.3 Summary of the Chapter

Work function of the rGO sheets is experimentally determined under different contact metals. It is demonstrated that the WF of the rGO sheets strongly depends on the oxygen concentration attached to the rGO. This oxygen concentration can be controlled either by controlling the thickness of the rGO sheets or by performing the reduction at different temperatures. By varying the thickness of the rGO sheets, a minimum of about 4.35 eV and a maximum of about 5.28 eV WF values are obtained with very thin and very thick rGO sheets under Pt/TiN contact metal. This variation in the WF is attributed to the two factors, namely thickness dependent WF and the oxygen concentration dependent WF. Thick rGO is supposed to have higher oxygen concentration which results in the large values of the WF.

Further, thermal reduction of the thick GO is carried out at different temperatures to control the overall oxygen concentration in the rGO sheets. WF of these rGO sheets, reduced at different temperatures, depends on the reduction temperatures. As the reduction temperature increases, oxygen concentration and hence the WF of the rGO sheets decreases. The obtained electrical data are well supported by the physical characterizations like XPS, FTIR and UPS.

Chapter 6

Investigation of metal/dielectric interaction and thermal stability of reduced graphene oxide gate electrode devices

In the metal gate electrode technology, it is well known that the metal reacts with the top of the gate dielectric and hence reduces the effective oxide thickness [107]. In the previous chapter, in section [5.1.3], it is noticed that effective oxide thickness (EOT), as obtained from capacitance in accumulation, of the rGO and non rGO gate electrode devices differ. EOT of the rGO gate electrode devices is close to the physical thickness, as obtained from the ellipsometer, of the gate dielectric while EOT of the non rGO gate electrode devices is about 1 nm less than the physical thickness. In this chapter, we performed high resolution transmission electron microscopy (HRTEM) analysis to analyze this difference in EOT. Further, the thermal stability of the different techniques adopted for WF modulation in CMOS technology is an important consideration [139, 140]. Hence, we also examined the thermal stability of the WF values as obtained with the technique proposed in the chapter 5 of this thesis.

The results presented in this chapter are published in ACS Applied Materials and Interfaces (DOI: 10.1021/am404649a) and are under review in IEEE Transactions on Nanotechnology.

6.1 Suppression of the metal/dielectric interaction with rGO incorporation between metal and dielectric

To study how the incorporation of rGO sheets affects the metal dielectric interaction, MOSCAPs were fabricated with three different gate dielectric (SiO_2) thicknesses with rGO/TiN gate

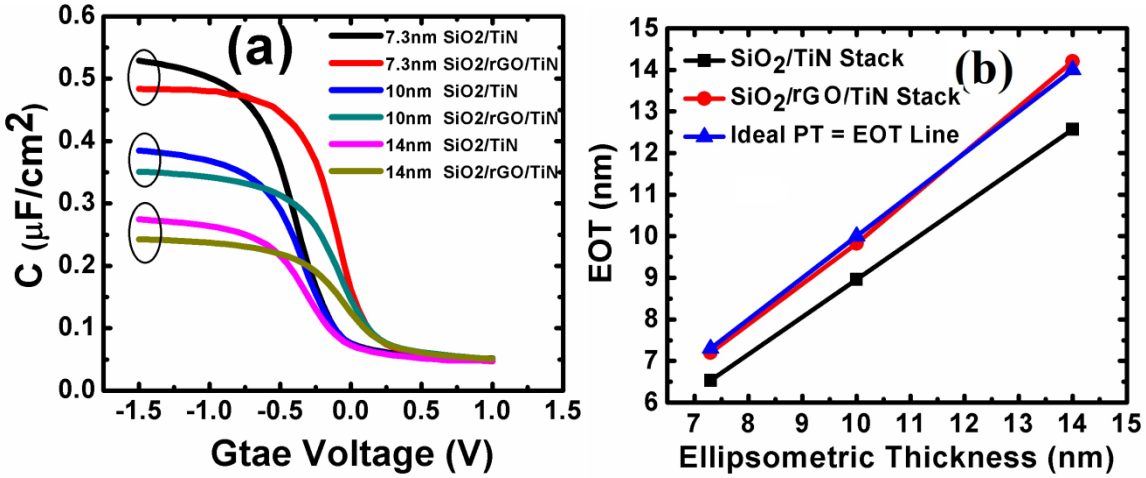


Figure 6.1: (a) CV curves for different thickness of gate oxide (7.3 nm, 10 nm and 14 nm: all measured by ellipsometry) with TiN and rGO/TiN as a gate electrode material, (b) Comparison of the EOT values as obtained from the accumulation CV plot and physical thickness obtained from the ellipsometer.

electrode. For all the 3 sets of devices, physical thickness and effective oxide thickness is compared. CV plots for 3 different dielectric thicknesses with rGO/TiN and only TiN gate electrode devices are shown in figure 6.1(a). For all the thicknesses, it is clear that rGO/TiN devices results in the lower accumulation capacitance and hence larger EOT. In fact, the EOT values obtained for the rGO/TiN devices are close to the physical thickness (PT) of the gate dielectric obtained from the ellipsometer while EOT for the only TiN gate electrode devices are about 1 nm less than the physical thickness as plotted in figure 6.1(b). This indicates the consumption of the dielectric by some chemical reaction at the TiN/SiO_2 interface which ceases to happen when rGO is incorporated between the metal and the dielectric.

We proposed that rGO is impermeable to TiN. The impermeability of graphene for TiN can be explained by considering the area of the graphene hexagon and the size of TiN molecule. The distance between the opposite edges of Gr hexagon is about 2.46 Å [141] which suggests that the maximum diameter of a sphere which can permeate through the Gr hexagon is about 2.46 Å. Ti-N has a bond length of about 1.57 Å [142] and Ti and N have atomic diameters of 2.84 nm and 1.36 nm respectively [143]. This implies that TiN, being larger than the Gr hexagon, would not

permeate through Gr film as demonstrated in figure 6.2. As a result, even a single defect free Gr layer would prevent the metal diffusion into the dielectric and hence the metal/dielectric reaction

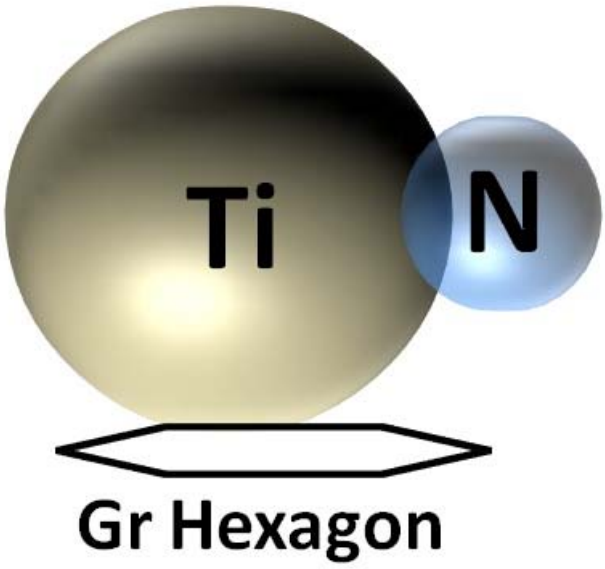


Figure 6.2: Schematic showing the impermeability of TiN through graphene hexagon.

would be suppressed. Possibility of diffusion of metals like Co, Ni and Au in Gr is also ruled out in the literature [144]. To confirm our hypothesis, elemental analysis along the cross section of the gate stack in HRTEM equipment is performed as shown in figure 6.3(a,b). A detectable signal of Ti is found in SiO_2 in non rGO gate electrode device while in Si/SiO_2/rGO/TiN devices, Ti signal dies very fast in the rGO. This result indicates that the TiN metal cannot diffuse through the graphene sheets and hence graphene prevents any metal/dielectric interaction. As a result, the EOT values and physical thickness of the gate dielectric for rGO gate electrode devices are almost same. This elemental mapping data supports our hypothesis based on the geometrical considerations that graphene prevents the metal (TiN) / dielectric interaction.

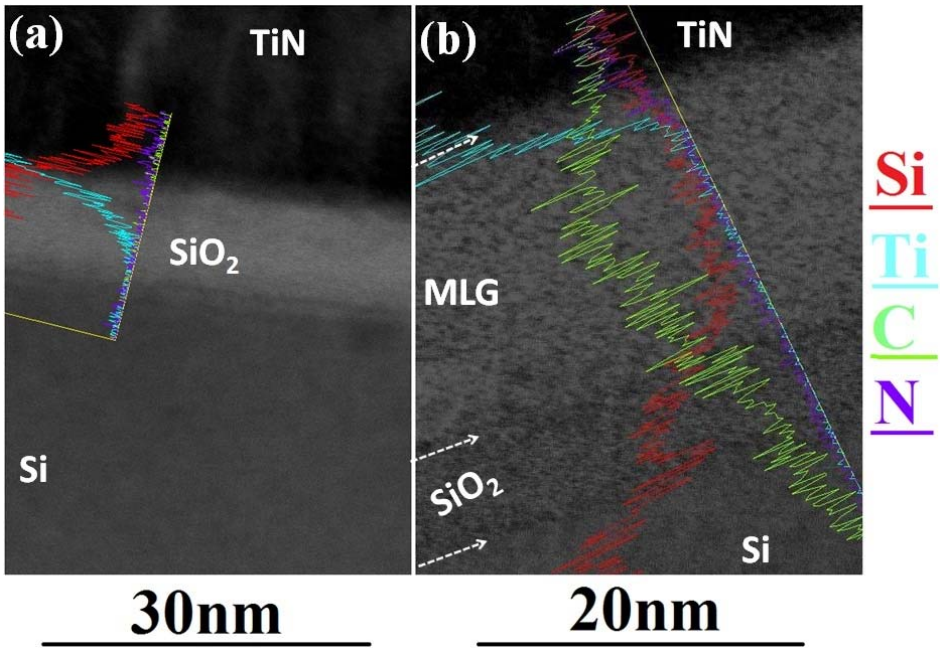


Figure 6.3: Dark filed cross section HRTEM image of (a) Si/SiO_2/TiN, (b) Si/SiO_2/rGO/TiN MOSCAP devices. A detectable Ti signal can be seen in the SiO_2 of figure 6.3(a) while in figure 6.3(b), Ti signal dies quickly in the top few nanometers of rGO layers. This shows that rGO sheets can be used as a diffusion barrier for metal atoms.

6.2 Thermal stability of the WF values obtained for rGO gate electrode devices

In order to test the thermal stability of the rGO/Pt/TiN gate electrode, MOSCAPs (same devices fabricated in chapter 5 of this thesis) were subjected to rapid thermal annealing in nitrogen ambient for temperatures ranging from 400°C to 950°C for 5 sec. each. The same sample from each split in the rGO thickness was subjected to an anneal–electrical measurement–anneal–electrical measurement...sequence.

6.2.1 Results and analysis of thermally treated rGO/Pt/TiN MOSCAPs

The CV plots after thermal treatment of Pt/TiN and rGO/Pt/TiN gate electrode devices are shown in figure 6.4(a-d). CV curves with rGO gate electrode devices remain steep even after 900°C thermal annealing while a significant stretch out is observed in the CV curve for Pt/TiN gate electrode devices. This indicates the degradation of the gate dielectric in non-rGO gate electrode devices as a result of increasing anneal temperatures. This observation suggests that rGO gate

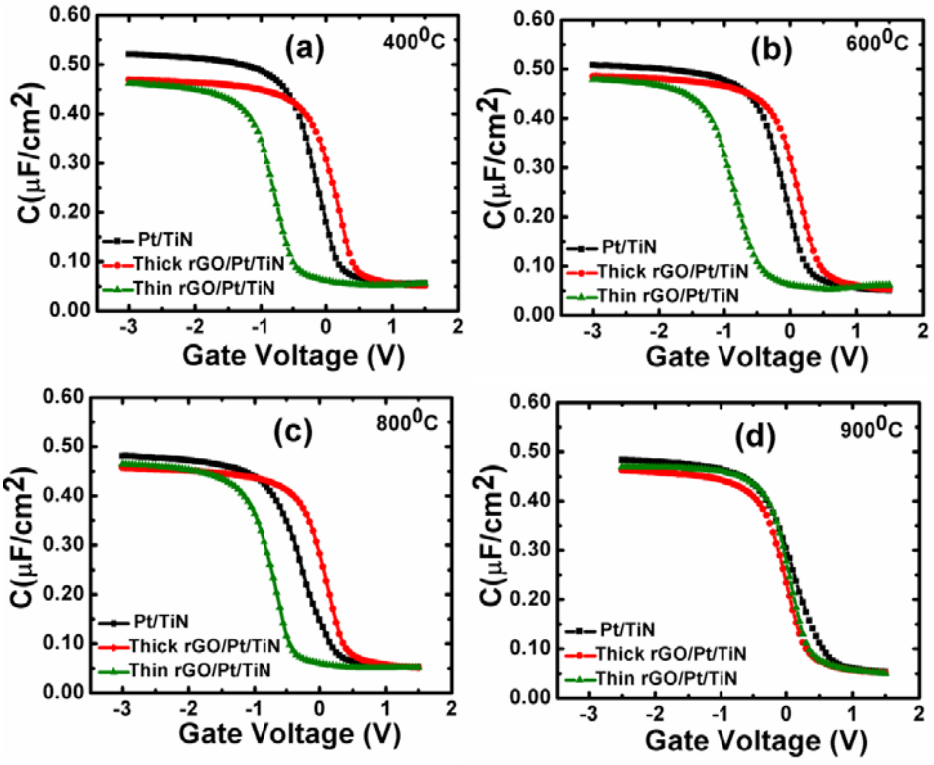


Figure 6.4: CV curves for MOSCAPs with and without rGO under Pt/TiN contact metal after rapid thermal treatment at different temperatures. (a) after annealing at 400°C (b) after annealing at 600°C (c) after annealing at 800°C (d) after annealing at 900°C. Increasing stretch out in the CV of Pt/TiN gate electrode devices is observed after annealing at different temperatures. Stretch out becomes more prominent after 900°C while the CVs for rGO under Pt/TiN are steep even after 900°C anneal.

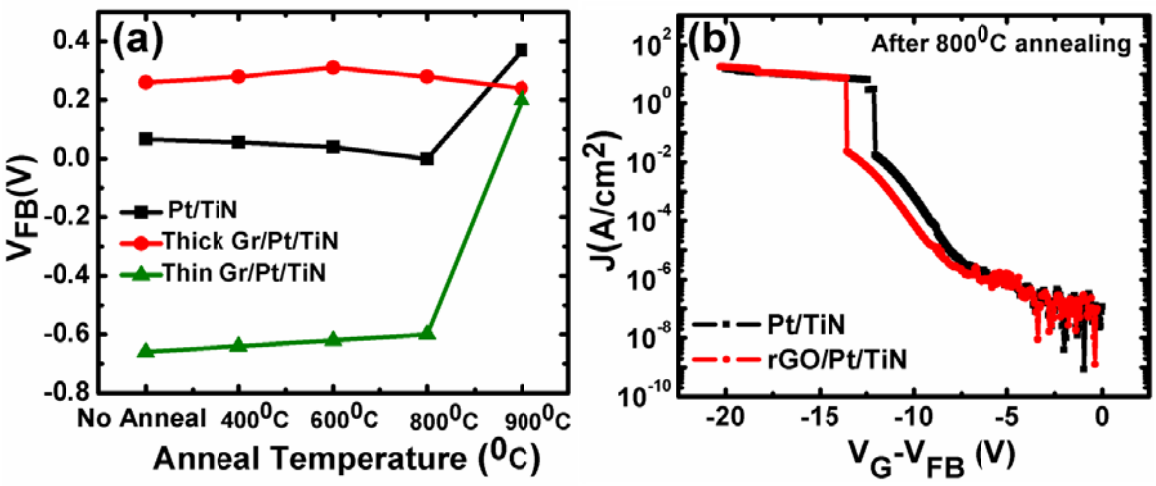


Figure 6.5: (a) Variation in the flat band voltage with annealing temperature for thick (> 5 layers) and thin rGO (1-3 layers) under Pt/TiN contact metal. (b) Comparison of breakdown characteristics for Si/SiO_2/PtTiN and Si/SiO_2/thick rGO/Pt TiN after 800°C annealing step.

electrode devices are more robust against any thermal treatment induced damage to the gate stack of the MOS devices. Flat band voltage values of the Pt/TiN and rGO/Pt/TiN gate electrode devices remain stable till 800°C annealing temperature as shown in figure 6.5(a). After 800°C anneal, flatband voltage changes only by 0.06 V and by 0.02 V for thin and thick rGO/Pt/TiN gate electrode devices respectively. However, after 900°C thermal annealing, flatband voltage of thin rGO/Pt/TiN devices increases significantly and approaches to that of Pt/TiN electrode devices while flatband voltage of thick rGO/Pt/TiN MOSCAPs remains almost stable to its initial value of 0.26 V. Figure 6.5(b) compares the breakdown behavior of Pt/TiN gate electrode devices with and without inclusion of rGO beneath it, after 800°C thermal annealing step. The rGO/Pt/TiN gate electrode devices result in less leakage and higher breakdown voltage values compared to only Pt/TiN electrode devices. This again indicates that the rGO gate electrode devices not only allow the tuning of WF but the quality of the gate dielectric also improves substantially.

6.2.1.1 HRTEM investigation of rGO/Pt/TiN gate electrode devices after thermal treatment

To investigate the cause of this V_{FB} change after high temperature processing, cross-section HRTEM analysis of the sample annealed at 900°C is performed. High temperature (900°C) thermal anneal causes the Pt to diffuse through the rGO sheets as shown in figure 6.6. For thick and moderate thick rGO (figure 6.6(a) and 6.6(b)), Pt could not pass through the rGO sheets completely while for very thin rGO (figure 6.6(c)), Pt diffuses entirely through the rGO sheets and reacts with the SiO_2. As the Pt could not reach to the SiO_2 dielectric after annealing for the devices having thick rGO under Pt metal, the V_{FB} for these devices remains stable even after 900°C anneal. V_{FB} for thin rGO devices approaches to the V_{FB} of SiO_2/Pt/TiN stack as the Pt is now in direct contact with SiO_2 after passing through the rGO sheets as clear from the HRTEM image of figure 6.6(c).

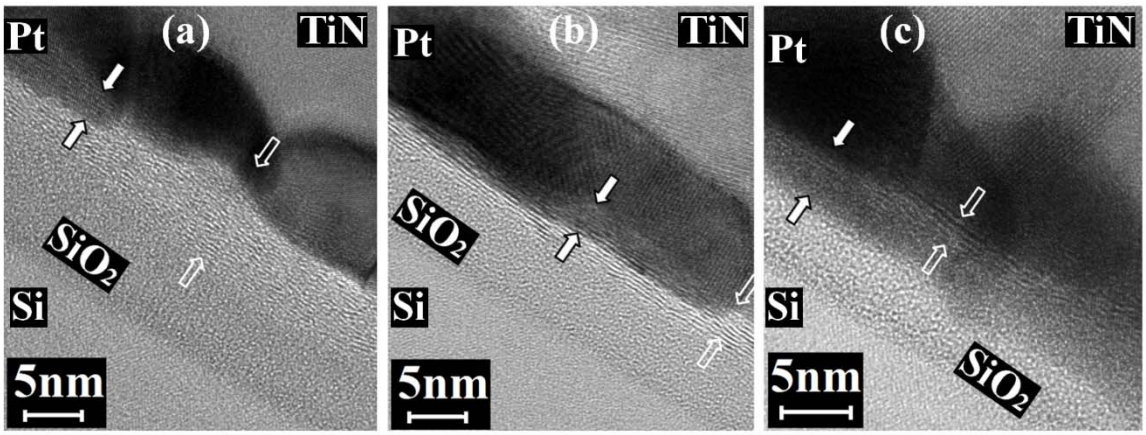


Figure 6.6: Cross section HRTEM images of Si/SiO_2/rGO/Pt/TiN stack after thermal annealing at 900°C for 5 sec. in N_2 ambient, (a) For thick rGO (b) for moderate thick rGO and (c) for very thin rGO. Hollow arrows demarcate the rGO while solid arrows indicate the region of Pt diffusion in rGO/SiO_2. For thick and moderate thick rGO (a and b), Pt could not cross the full rGO thickness while for very thin rGO, Pt pass through the rGO and reacts with SiO_2.

6.3 Summary of the chapter

In this chapter, effect of incorporation of rGO sheets between metal and gate dielectric (SiO_2) is discussed. It is demonstrated that rGO sheets in between the top metal and the gate dielectric not only allow to modulate the WF but also prevents any dielectric contamination by the top metal. Elemental mapping along the cross section of the gate stack clearly demonstrates that signal of top TiN metal dies out very fast in the top few nanometers of the rGO sheets and hence could not react to the SiO_2 dielectric. Further, high temperature stability of the WF values obtained with rGO as a gate electrode is also investigated and it is found that obtained WF values are thermally stable till 800°C annealing conditions. HRTEM analysis reveals that diffusion of metal through rGO sheets, as a result of high temperature processing, is the main cause of WF instability.

Chapter 7

Summary of the thesis and future directions

In the initial part of the thesis, NC formation statistics and Pt nanocrystal memory effects are discussed. It is demonstrated that the NC distribution (size, density and area coverage) strongly depends on the parameters like initial metal thickness, anneal time, anneal temperature and underlying dielectric. Memory results of Pt NC memory devices are also demonstrated. A memory window of about 3.6 V at ± 10 V P/E voltages is obtained. Further, based on the recent advancements in the flash memory research, it is concluded that the metal NC memory technology is not a viable option for technology node 20 nm and below. In this wake, novel material like graphene is proposed and experimentally investigated as a charge storage layer in conventional floating gate flash memory structure. Further, as it is necessary to know the WF of the CSL, WF of reduced graphene oxide sheets (which are used as a CSL in this thesis) was experimentally determined using it as a gate electrode in MOS devices.

Height reduction of the flash memory gate stack is one of the key solutions to reduce the parasitic coupling among the neighboring FGs of flash memory devices. We demonstrated that the graphene (reduced graphene oxide is used in this thesis) can be used as charge storage layer in conventional FG flash memory structure. Graphene as a charge storage layer would offer the ultimate scalability to the vertical dimensions memory gate stack. A large memory window of 9.4 V is demonstrated with reduced graphene oxide as a charge storage layer. Retention of the stored charges is tested at room temperature as well as at elevated temperatures. A 6.9 V remnant memory window at room temperature and 2.8 V remnant memory window at 150°C is demonstrated. Activation energies are calculated for the first time for graphene charge storage devices. One issue that is encountered with these devices is the device to device variability and poor endurance. The devices could be programmed/erased for about 1000 cycles at 20V, 1ms P/E operation. Endurance of the devices can be improved by improving the quality of the

blocking dielectric. However, to minimize the device to device variability, precise control on the graphene thickness and reduction conditions is needed.

As the memory performance strongly depends on the WF of the CSL, the WF of the rGO sheets is also experimentally determined by integrating them as a gate electrode in MOS structure. It is demonstrated that WF of the rGO sheets can be varied by varying the number of rGO sheets as well as by controlling the amount of oxygen concentration in the rGO sheets. A wide range of work function modulation (from n-type to p-type) is demonstrated with varying number of rGO layers under the contact metals. WF of the gate electrode in a MOS structure could be modulated from 4.35 eV (n-type metal) to 5.28 eV (p-type metal) by sandwiching different numbers of rGO layers between top contact metals and gate dielectric SiO_2. WF of the gate electrode shows strong dependence on the number of rGO sheets and is seen to be nearly independent of the contact metals used. The dependence of the WF on the number of rGO sheets is verified by ultra violet photoelectron spectroscopy. The observed WF modulation is attributed to the different amounts of oxygen concentration in different thicknesses of rGO layers. This is experimentally verified by X- ray photoelectron spectroscopy (XPS) analysis and by Fourier Transform Infrared Spectroscopy (FTIR) analysis. XPS analysis yields different amounts of oxygen concentration in different thicknesses of rGO sheets annealed under identical conditions. The difference between the WF values for various thicknesses of rGO layers under different capping metals (Pt, Ir, Al) lie within 0.12 eV. This suggests that the WF of gate electrode is predominantly determined by the thickness of the rGO layers and the amount of oxygen present in the different functional groups attached to it.

It is also demonstrated that for very thick rGO sheets where layer dependent WF no longer comes in to picture, WF can be modulated by reducing rGO sheets at different temperatures. Reduction at different temperatures causes different amounts of remnant oxygen (high oxygen concentration at low reduction temperature and vice versa) in the rGO sheets which in turn changes the WF from low to high value. The obtained WF values are thermally stable up to 800°C. Post annealing, X-HRTEM analysis of the samples reveals that the diffusion of metal through rGO layers at higher temperature is the main cause of WF instability.

This information on the WF of rGO sheets is very important to optimize the overall performance of rGO CSL flash memory devices. The study, in a very distinct manner, suggests that by controlling the oxygen concentration in the rGO sheets, depth of the potential well formed by rGO sheets and hence the memory performance of the rGO CSL flash devices can be tailored.

7.1 Future Directions

- Density of states (DOS) in the graphene is very low ($1 \times 10^{12} - 1 \times 10^{13}$ cm^{-2}), low DOS is undesirable for a material intended to used as a charge storage layer. One possible way to enhance the DOS of graphene is combining it with some metal. Hence a combination of graphene and thin metal layers can be further examined for floating gate application.

- The rGO sheets can also be investigated as a hybrid floating gate i.e. low WF of the bottom rGO sheets which are in the direct contact with the tunnel oxide and the high WF of the top rGO sheets which are in contacts with the blocking dielectric. Such combination of the FG layer would allow to have the efficient program/erase.

- Endurance characteristics of the graphene charge storage devices are to be studied rigorously.

- Chemical vapor deposited (CVD) graphene is expected to have less defect density compared to that in graphene obtained after reduction of the graphene oxide, therefore, it would be interesting to compare the memory performance of two kinds of graphene, i.e. CVD grown graphene and rGO.

- As demonstrated in the thesis, oxygen concentration plays a significant role in determining the WF of the graphene sheets, hence, controlled oxidation of the CVD grown graphene and study of its memory performance would be interesting to pursue. This will again allow us to modulate the WF and hence the P/E characteristics of the graphene FG devices.

- Graphene can also be examined as a diffusion barrier for different metals (e.g. copper) at high temperatures.

Annexure A

Structure of Graphene Oxide (GO) and Reduced Graphene Oxide (rGO)

Graphene oxide is obtained by oxidative exfoliation of graphite films. In fact the graphite oxide also has layers of carbon atoms networks similar to that of graphite, however the interlayer distance among the planes is large compared to the graphene interlayer distance (0.7 nm to 1 nm compared to 0.34 nm in graphene) and the carbon atoms are decorated by the various oxygen containing functional groups. Structure of the GO can be thought of as individual graphene sheets decorated with oxygen functional groups on both basal planes and edges [136, 145] as shown in figure A1 (a, b). These oxygen functional groups are mostly hydroxyl (C-OH) and epoxy (C-O-C) groups attached on the basal plan and carboxy (COOH), carbonyl (C=O) and phenol, attached to the sheet edges. These functional groups can be easily identified by physical characterization techniques like XPS and FTIR. These oxidized carbon layers are hydrophilic and can be easily exfoliated in water that produces the mono or few layered graphene oxide flakes. These graphene oxide flakes can be further reduced to graphene like flakes by the removal of the oxygen containing groups. Binding energies of the different oxygen containing functional groups attached to the GO, calculated from ab initio simulations, are tabulated in table A1 [135]. Some of these binding energies are relatively low and hence the corresponding functional groups can be very easily removed by mild chemical or thermal treatment. The so achieved (chemically or thermally derived) graphene flakes are most commonly known as reduced graphene oxide (r-GO) sheets.

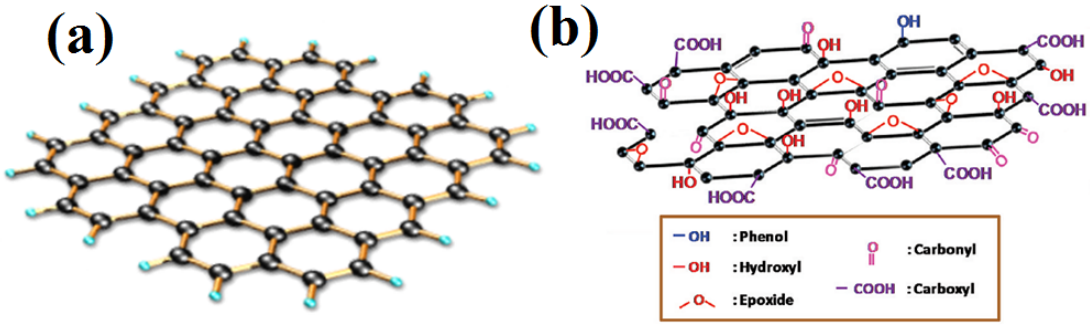


Figure A1: structure of (a) Gr [146] and (b) GO [136]. Different functional groups attached to GO are shown in (b).

Table A1: Binding energies of different oxygen containing functional groups attached in GO [135].

Chemical species	Binding energy (eV/functionality)
Hydroxyl (C-OH)	1.5 eV/OH
Epoxide (C-O-C)	3.1 eV/O
Carboxyl (COOH)	5.8 eV/COOH
Ketonic species (C=O)	8.0 eV/O
Ether species (C-O) (non aggregated)	4.9 eV/O
Aggregated cyclic edge ether (-O-)	9.1 eV/O

Electronic, optical and mechanical properties of the graphene oxide can be tuned by controlling the reduction extent. For example, experimentally, it is demonstrated that the band gap of the GO sheets can be tuned from 2 eV to 0.02 eV by different level of reduction [147]. Theoretically also, by using density functional theory, Huang et al. demonstrated that total density of states and band gap of GO can be varied by varying the oxygen to carbon (O/C) ratio. Band gap of the rGO with different O/C ratio is shown in figure A2 [148].

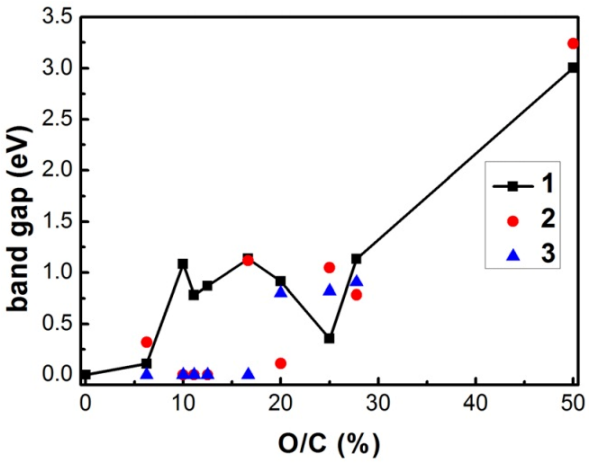


Figure A2: The band gap of the rGO with different O/C ratio [148].

GO is considered as an important precursor to obtain the graphene sheets by chemical or thermal reduction of the GO. In chemical reduction process, hydrazine treatment of the GO is the most

common method to reduce the GO to more graphene like character [136, 145, and 149]. In this method reduced graphene oxide (rGO) agglomerates are obtained in the aqueous or in some polar solvents polar solvents like N-Methylpyrrolidone (NMP) or dimethylformamide (DMF) [149].

GO can also be reduced to more graphene like nature by thermal treatment. GO starts losing its different oxygen functional groups at temperatures starting from 200°C [136]. One way to achieve this is to deposit the GO sheets (either by drop casted or by spin coating) on some suitable substrate (Si or Si/SiO_2) and place it in some heated chamber. A range of temperatures starting from 200°C to 1000°C under different ambient conditions like Argon (Ar), Nitrogen (N_2) and vacuum are proposed for efficient reduction of GO sheets [120, 136, 137]. Figure A3 shows the evolution of C1s spectra of GO thermally reduced at different temperatures.

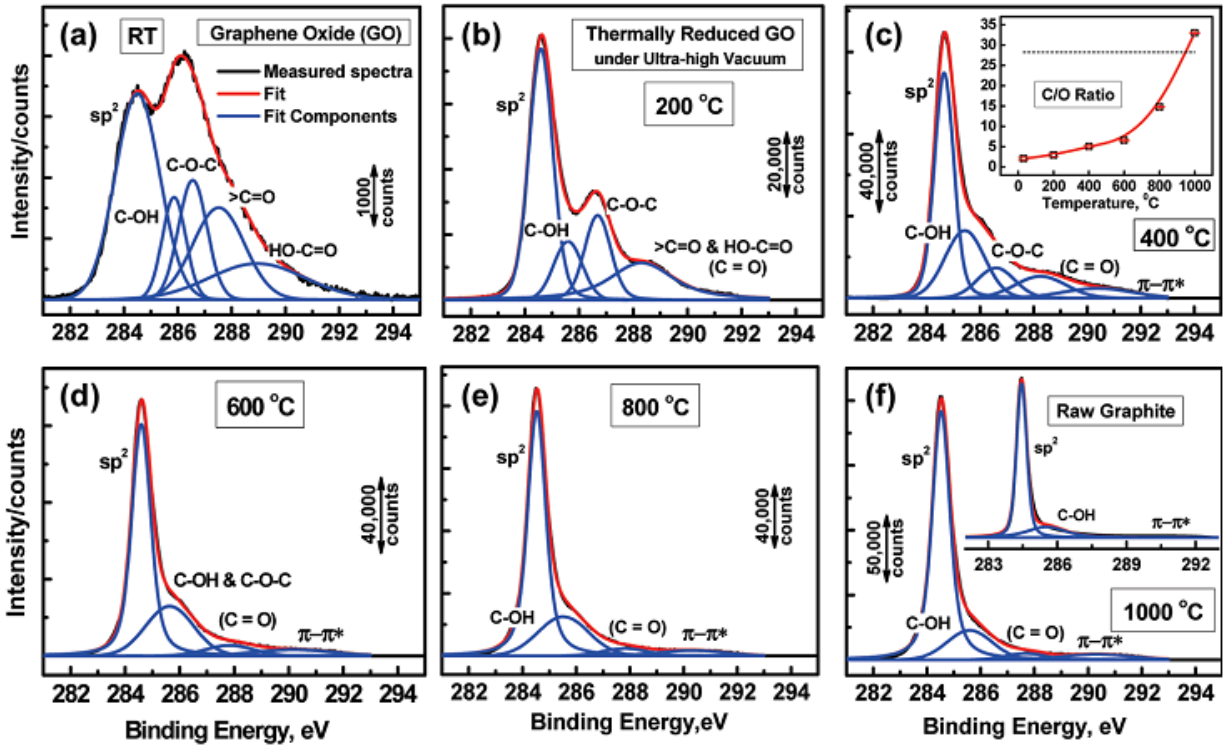


Figure A3: High-resolution C1s XPS spectra of thermally reduced GO sheets at different temperatures [136].

In the present study graphene oxide was thermally reduced to graphene (rGO) at 550^0C in Ar ambient for 1hr. Different ranges of reduced graphene oxide (rGO) thicknesses were obtained by using pre-optimized solution of different concentration of GO in DI water. A solution of 1mg of GO in 20 ml of DI water for very thin rGO sheets, 1mg of GO in 10ml of DI water for moderate thick rGO sheets and 1mg of GO in 5ml of DI water for very thick rGO sheets was prepared. Different number of drops of these liquids combined with the SEM and AFM techniques was used to obtain the different thicknesses of rGO sheets.

This procedure (thermal reduction of the GO sheets at the device location) is suited for MOS technology as GO is easily dispersed in de-ionized (DI) water. Because of the presence of the negative charge on GO sheets, it remains well separated for long time in DI water. Reduced graphene oxide and graphene agglomerates in DI water and polar solvents like N-Methylpyrrolidone (NMP) or dimethylformamide (DMF) are used as solvents to avoid agglomeration [149]. However, these chemicals can potentially contaminate the gate dielectric while on the other hand DI water is regularly used in CMOS device processing without any detrimental effect on the gate dielectric quality.

It is worth discussing here that I_D/I_G ratio decreases for thermally reduced GO sheets figure 5.2(c), while I_D/I_G ratio increases for chemically reduced GO sheets as shown in figure 4.2(c) of this thesis. There are several reports where an increase or decrease or even no change in the I_D/I_G ratio after reduction is observed. In [136, 137], a reduction in the I_D/I_G ratios is observed for chemically or thermally reduced GO while a negligible change in the I_D/I_G ratio has been observed by D. Zhan et al. [150]. In [136], it is mentioned that I_D/I_G ratio is a measure of the distance between defects (L_D) in graphene. This ratio may have a high value as well as a low value depending on the distance between the defects.

Annexure B

Process recipes for pseudo source drain rGO flash transistors fabrication

Step No.	Process step	Tool used	Process recipe	Any other comment
1	RCA Clean	Wet process bench	• 2% HF dip -- 1152ml DI water + 48ml (49%HF) for 30 sec. • RCA1-- NH4OH : H_2O_2 : DI water in the ratio 125ml : 250ml : 875ml @750°C, duration: 1200 sec.; • 2% HF dip -- 1152ml DI water + 48ml (49%HF) for 30 sec. • RCA2-- HCL : H_2O_2 : DI water in the ratio 125ml : 250ml : 875ml @750°C, duration: 1200 sec.; • 2% HF dip -- 1152ml DI water + 48ml(49%HF) for 30 sec.	This can be escaped as the second step is directly for alignment mark patterning
2	Alignment mark patterning (To expose the alignment marks region and protect the rest of the wafer)	Double side aligner (EVG 620)	• Dehydration—300°C on hot plate for 300 sec. • Resist used -- S1813; Spin recipe — Step 1 = 500 rpm, 30 sec. step 2 = 6000 rpm, 45 sec. step 3 = 300 rpm, 10 sec. • Expected resist thickness ~ 1.2 μm • Prebake: 90°C, 120 sec. • Exposure : 50mJ/cm^2, soft contact, 1um separation • Development : MF 319(vertical), 25 sec. • Post development bake :	

			90°C, 60 sec.	
3	Si etching from the alignment marks	STS RIE	• Power = 400W, SF6 = 40sccm, pressure = 100mTorr, Time 7min	
4	Ashing for Resist removal from rest of the wafer	AMAT Ash chamber	• Power 1400W, O_2 = 300sccm, Time 2Min	
5	Pirahna	Micro 1 wet bench	• H_2SO_4 : H_2O_2 in the ratio 910ml : 390ml for 3600 sec.	
6	RCA	Micro 1 wet bench	• Same as in step 1	
7	Pad oxide growth	Ultech furnace, stack 1, tube 1 - Dry oxidation	• Same as in step 2; only growth time is 1500 sec. • Target thickness = 8nm	Thickness needs to recheck before doing any fresh run. (This May not be required as the implantation in the next step is to be done on non active region. In that case step 6 can also be escaped)
8	Active area patterning (To protect the actual MOSCAP region with resist))	Double side aligner (EVG 620)	• Dehydration—300°C on hot plate for 300sec. • Resist used -- S1813; Spin recipe — Step 1 = 500 rpm, 30 sec. step 2 = 6000 rpm, 45 sec. step 3 = 300 rpm, 10 sec. • Prebake: 90°C, 120 sec.	To increase the resist thickness, main spinning speed can be reduced to 2000 rpm

			• Exposure : 50 mJ/cm^2, soft contact, 1um separation • Development : MF 319(vertical), 25 sec. • Post development bake : 90°C, 60 sec.	for 45 sec.
9	Implantation	PIII (Micro 2)	• Power = 1000W, pressure 0.12mbar, frequency = 5KHz, duity cycle = 10%, On time = 20µs, implant duration = 30 sec. bias = -2KV. Distance between plasma and wafer = 9cm.	Power can be reduced to 950W or even 900W
10	Ashing for Resist removal	AMAT Ash chamber	• Power 1400W, O_2 = 300sccm, Time 2Min	
11	Pirahna	Micro 1 wet bench	• H_2SO_4 : H_2O_2 in the ratio 910ml : 390ml for 3600 sec.	
12	RCA	Micro 1 wet bench	• Same as in step 1	
13	Tunnel oxide growth	Ultech furnace, stack 1, tube 1 - Dry oxidation	• Same as in step 2; only growth time is 1500 sec. • Target thickness = 8nm	Thickness needs to recheck before doing any fresh run
14	GO drop cast	Hot plate at 120°C		Can be tried at other temperature like 70°C to 100°C
15	GO thermal reduction	in high-K chamber	• At 550°C, Ar flow 2000sccm, pressure 100torr	
16	Blocking dielectric Al_2O_3 deposition	AMAT ENDURA,	• Power = 200W, Ar:O_2 =	

		Chamber 1	10:14, time 20min	
17	Top metal deposition	AMAT ENDURA, Chamber D	• Power = 350W, Ar:N_2 = 20:20, time 15min	
18	Final MOSCAP patterning	Double side aligner (EVG 620)	• Same as in step 4,	
19	Metal + gate stack etching	AMAT etch tool	• RF/bias = 1000/130, Cl_2:BCL_3 = 100:40, pressure = 9 mtorr, helium pressure = 4000mtorr	
20	Ashing	AMAT etch tool	• Same as in step 6	
21	Back side etching	Micro 1 wet bench	• 2% HF	
22	Back side metallization	Aluminum evaporator		

Annexure C

Process recipes for rGO flash transistors fabrication

Step No.	Process step	Tool used	Process recipe	Any other comment
1	RCA Clean	Wet process bench	• 2% HF dip -- 1152ml DI water + 48ml (49%HF) for 30 sec. • RCA1-- NH4OH : H_2O_2 : DI water in the ratio 125ml : 250ml : 875ml @750°C, duration: 1200 sec.; • 2% HF dip -- 1152ml DI water + 48ml(49%HF) for 30 sec. • RCA2-- HCL : H_2O_2 : DI water in the ratio 125ml : 250ml : 875ml @750°C, duration: 1200 sec.; • 2% HF dip -- 1152ml DI water + 48ml(49%HF) for 30 sec.	
2	Thin dry oxide (Pad oxide)	Ultech furnace, stack 1, tube 1 - Dry oxidation	• Pre-growth step – 850°C, O_2 gas flow = 50 sccm, N_2 gas = ON, duration = 120 sec. • Growth step-- 850°C, O_2 gas flow = 5000 sccm, N_2 gas = OFF, duration = 1800 sec. • Post growth anneal = 900°C, O_2 gas flow = 0seem, N_2 gas ON, duration 300 sec. • Expected thickness ~ 10 nm	Thickness needs to recheck before doing any fresh run
3	Silicon Nitride	Ultech furnace, stack 2, tube 3 -	• Ramp up step-- 300 sec., temperature	3 to 4 dummy

	deposition	LPCVD.	Stabilization step: 120 sec., pressure Stabilization step : 300 sec. • Deposition step -- SiH_4 gas flow: 80 sccm, NH_3 gas flow: 100 sccm, N_2 gas flow: 1000 sccm, temperature: 780°C, pressure: 0.3 torr, duration: 3900 sec. • Expected thickness = 60 nm, RI = 1.99	runs require to get the RI of 1.99
4	Active area patterning	Double side aligner (EVG 620)	• Dehydration—300°C on hot plate for 300 sec. • Resist used -- S1813; Spin recipe — Step 1 = 500 rpm, 30 sec. step 2 = 6000 rpm, 45 sec. step 3 = 300 rpm, 10 sec. • Prebake: 90°C, 120 sec. • Exposure : 50mJ/cm^2, soft contact, 1um separation • Development : MF 319(vertical), 25 sec. • Post development bake : 90°C, 60 sec.	
5	Silicon Nitride etch	STS RIE	• CF_4 gas flow : 40 sccm, O_2 gas flow : 4sccm, RF Power for Selective Etch (SE) = 50W, Chamber pressure(SE) = 110 mTorr	RI should be exactly 1.99 for this to happen nicely
6	Ashing for Resist removal	AMAT Ash chamber	• Power 1400W, O_2 = 300sccm, Time 2Min	
7	Pirahna	Micro 1 wet bench	• H_2SO_4 : H_2O_2 in the ratio 910ml : 390ml for 3600 sec.	

8	Filed oxide growth	Ultech furnace, stack 2, tube 1 - Pyrogenic oxidation	• Growth step: H_2 gas flow = 8000 sccm, O_2 gas flow = 6000sccm, temperature = 1000°C,torch temperature: 735°C, Time: 2400 sec.	
9	Silicon Nitride etch	Micro 1 wet bench	• Hot phosphoric acid, 80°C, time ~ 45minute	RI should be exactly 1.99 for this
10	RCA	Micro 1 wet bench	• Same as in step 1	
11	Gate area patterning (GATE level lithography for Dummy gate for implant stop layer)	Double side aligner (EVG 620)	• Same as in step 4,	To increase the resist thickness, main spinning speed can be reduced to 2000 rpm for 45 sec.
12	Implantation	PIII (Micro 2)	• Power = 1000W, pressure 0.12mbar, frequency = 5KHz, duity cycle = 10%, On time = 20μs, implant duration = 30 sec. bias = -2KV. Distance between plasma and wafer = 9cm.	Power can be reduced to 950W or even 900W
12	Ashing for Resist removal	AMAT Ash chamber	• Same as in step 6	
13	Pirahna	Micro 1 wet bench	• Same as in step 7	
14	Activation anneal	Annealsys RTP (Nano lab)	• Temperature = 950°C, time = 5 sec., N_2 gas flow = 1000sccm	

15	RCA	Micro 1 wet bench	• Same as in step 1	
16	Tunnel oxide growth	Ultech furnace, stack 1, tube 1 - Dry oxidation	• Same as in step 2; only growth time is 1500 sec.	Thickness needs to recheck before doing any fresh run
17	GO drop cast	Hot plate at 120°C		Can be tried at other temperature like 70°C to 100°C
18	GO thermal reduction	in high-K chamber	• At 550°C, Ar flow 2000sccm, pressure 100torr	
19	Blocking dielectric Al_2O_3 deposition	AMAT ENDURA, Chamber 1	• Power = 200W, Ar:O_2 = 10:14, time 20min	
20	Top metal deposition	AMAT ENDURA, Chamber D	• Power = 350W, Ar:N_2 = 20:20, time 15min	
21	Gate area patterning (Final gate patterning)	Double side aligner (EVG 620)	• Same as in step 4,	
22	Metal + gate stack etching	AMAT etch tool	• RF/bias = 1000/130, Cl_2:BCL_3 = 100:40, pressure = 9 mtorr, helium pressure = 4000mtorr	
23	Ashing	AMAT etch tool	• Same as in step 6	
24	Back side etching	Micro 1 wet bench	• 2% HF	
25	Back side metallization	Aluminum evaporator		

References:

[1] Available online at http://agigatech.com/blog/page/2/

[2] Available online at http://agigatech.com/blog/system-uses-for-nand-flash/

[3] International Technology Roadmap for Semiconductors. 2010, update, available on-line at http://www.itrs.net/reports.html

[4] M. F. Beug, N. Chan, T. Hoehr, L. M. Meskamp and M. Specht, "Investigation of Program Saturation in Scaled Interpoly Dielectric Floating-Gate Memory Devices", *IEEE Transactions on Electron Devices*, vol. 56, no. 8, pp. 1698-1704, 2009.

[5] 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, 2002.

[6] B. De Salvo, C. Gerardi, R. van Schaijk, S. A. Lombardo, D. Corso, C. Plantamura, S. Serafino, G. Ammendola, M. van Duuren and P. Goarin, et al., "Performance and Reliability Features of Advanced Nonvolatile Memories Based on Discrete Traps (Silicon Nanocrystals, SONOS)", *IEEE Transactions on Devices and Materials Reliability*, vol. 4, no. 3, pp. 377-389, 2004.

[7] C. H. Lee, S. H. Hur, Y. C. Shin, J. H. Choi, D. G. Park and K. Kim, "Charge-Trapping Device Structure of SiO_2 / SiN / High-k Dielectric Al_2O_3 for High-Density Flash Memory", *Applied Physics Letters*, vol. 86, p. 152908, 2005.

[8] C. Sandhya, A. B. Oak, N. Chattar, A. S. Joshi, U. Ganguly, C. Olsen, S. M. Seutter, L. Date, R. Hung and Juzer Vasi, et al., "Impact of SiN Composition Variation on SANOS Memory Performance and Reliability under NAND(FN/FN) Operation", *IEEE Transactions on Electron Devices*, vol. 56, no. 12, pp. 3123 – 3132, 2009.

[9] S. Fujii, J. Fujiki, N. Yasuda, R. Fujitsuka and K. Sekine, "Transition of Erase Mechanism for MONOS memory depending on SiN Composition and its Impact on Cycling Degradation", *in the proc. of International Reliability of Physics Symposium*, 2010, pp. 956-959.

[10] E. Vianello, F. Driussi, L. Perniola, G. Molas, J. P. Colonna, B. De Salvo and L. Selmi, "Explanation of the Charge-Trapping Properties of Silicon Nitride Storage Layers for NVM Devices Part I: Experimental Evidences From Physical and Electrical Characterizations", *IEEE Transactions on Electron Devices*, vol. 58, no. 8, pp. 2483-2489, 2011.

[11] H. C. You, T. H. Hsu, F. H. Ko, J. W. Huang, W. L. Yang and T. F. Lei, "SONOS-Type

Flash Memory Using an HfO_2 as a Charge Trapping Layer Deposited by the Sol–Gel Spin-Coating Method", *IEEE Electron Device Letters*, vol. 27, no. 8, pp. 653-655, 2006.

[12] H. W. You and W. J. Cho, "Charge Trapping Properties of the HfO_2 Layer with Various Thicknesses for Charge Trap Flash Memory Applications", *Applied Physics Letters*, vol. 96, p. 093506, 2010.

[13] Y. N. Tan, W. K. Chim, W. K. Choi, M. S. Joo, T. H. Ng and B. J. Cho, "High-K HfAlO Charge Trapping Layer in SONOS-type Nonvolatile Memory Device for High Speed Operation", *In the Proc. of International Electron Device Meeting*, 2004, pp. 889-892.

[14] M. B. Gonzalez, J. M. Rafí, O. Beldarrain, M. Zabala and F. Campabadal "Charge Trapping and Electrical Degradation in Atomic Layer Deposited Al_2O_3 Films", *Microelectronic Engineering*, vol. 109, pp. 57-59, 2013.

[15] M. She, T. J. King, "Impact of Crystal Size and Tunnel Dielectric on Semiconductor Nanocrystal Memory Performance", *IEEE Transactions on Electron Devices*, vol. 50, no. 9, pp. 1934-1940, 2003.

[16] Y. Shi, H. G. Yang, J. Lv, L. Pu, R. Zhang, B. Shen and Y. D. Zheng, "Silicon Nanocrystal Memories", *In the proc. of Solid State and Integrated Circuit and Technology* 2004, pp. 881-884.

[17] J. D. Blauwe, "Nanocrystal Nonvolatile Memory Devices", *IEEE Transactions on Nanotechnology*, vol. 1, no. 1, pp. 72-77, 2002.

[18] Z. Liu, C. Lee, V. Narayanan, G. Pei and Edwin C. Kan, "Metal Nanocrystal Memories—Part I: Device Design and Fabrication" *IEEE Transactions on Electron Devices*, vol. 49, No. 9, pp. 1606-1613, 2002.

[19] P. H. Yeh, J. J. Chen, P. T. Liu, D. Y. Wang and T. C. Chang, et al., "Metal Nanocrystals as Charge Storage Nodes for Nonvolatile Memory Devices", *Electrochimica Acta*, vol. 52, pp. 2920-2926, 2007.

[20] T. H. Hou. C. Lee, V. Narayanan, U. Ganguly and E . C. Kan, "Design Optimization of Metal Nanocrystal Memory-Part I: Nanocrystal Array Engineering", *IEEE Transactions on Electron Devices*, vol. 53, no. 12, pp. 3095-3102, 2006.

[21] A. Nainani, S. Palit, P. K. Singh, U. Ganguly, N. Krishna, J. Vasi and S. Mahapatra, "Development of a 3D Simulator for Metal Nanocrystal (NC) Flash Memories under NAND Operation", *In the Proc. of International Electron Device Meeting*, 2007, pp. 947-950.

[22] R. F. Steimle, R. Muralidhar, R. Rao, M. Sadd, C. T. Swift, J. Yater, B. Hradsky, S. Straub, H. Gasquet and L. Vishnubhotla, "Silicon Nanocrystal Non-Volatile Memory for

Embedded Memory Scaling", *Microelectronics Reliability*, vol. 47, pp. 585-592, 2007.

[23] T. C. Chang, F. Yen, J. b, S. C. Chen and Yu-Ting Tsai, "Development in Nanocrystal Memory", *Materials Today*, vol. 14, no. 12, pp. 608-615, 2011.

[24] N. Ramaswamy, T. Graettinger, G. Puzzilli, H. Liu, K. Prall, S. Gowda, A. Furnemont, C. Kim and K. Parat, "Engineering a Planar NAND Cell Scalable to 20 nm and Beyond", *In the Proc. of International Memory Workshop*, 2013.

[25] S. Raghunathan, T. Krishnamohan, K. Parat and K. Saraswat, "Investigation of Ballistic Current in Scaled Floating-Gate NAND Flash and a Solution", *In the Proc. of International Electron Device Meeting*, 2009, pp. 819-822.

[26] S. Jayanti, X. Yang, R. Suri and V. Misra, "Ultimate Scalability of TaN Metal Floating Gate with Incorporation of High-K Blocking Dielectrics for Flash Memory Applications", *In the Proc. of International Electron Device Meeting*, 2010, pp. 106-109.

[27] P. Blomme, M. Rosmeulen, A. Cacciato, M. Kostermans and C. Vrancken, "Novel Dual Layer Floating Gate Structure as Enabler of Fully Planar Flash Memory", *In the Digest of Technical Papers on the Symposium on VLSI Technology*, 2011, pp. 129-130.

[28] P. Blomme, A. Cacciato, D. Wellekens, L. Breuil, M. Rosmeulen, G. S. Kar, S. Locorotondo, C. Vrancken, O. Richard and I. Debusschere, "Hybrid Floating Gate Cell for Sub-20-nm NAND Flash Memory Technology", *IEEE Electron Device Letters*, vol. 33, no. 3, pp 333-335, 2012.

[29] D. Wellekens, P. Blomme, M. Rosmeulen, T. Schram, A. Cacciato, I. Debusschere, J. V. Houdt and S. V. Aerde, "An Ultra-thin Hybrid Floating Gate Concept for Sub-20nm NAND Flash Technologies", *In the Proc. of International Memory Workshop*, 2011.

[30] G. S. Kar, L. Breuil, P. Blomme, H. Hody, S. Locorotondo, N. Jossart, O. Richard, H. Bender, G. Van den Bosch, I. Debusschere and J. Van Houdt, "Ultrathin Hybrid Floating Gate and High-K Dielectric as IGD Enabler of Highly Scaled Planer NAND Flash Technology", *In the Proc. of International Electron Device Meeting*, 2012, pp. 17-20.

[31] P. K. Singh, K. K. Singh, R. Hofman, K. Armstrong, N. Krishna, S. Mahapatra,"Au Nanocrystal Flash Memory Reliability and Failure Analysis", *In the proc. of IEEE Intl. Symposium on the Physical and Failure Analysis of Integrated Circuits*, (IPFA) 2008.

[32] S. Jayanti, X. Yang and V. Misra, "Investigation of Thermal Stability of High-K Interpoly Dielectrics in TaN Metal Floating Gate Memory Structures", *In the Proc. of International Memory Workshop*, 2011.

[33] K. S. Novoselov, V. I. Falko, L. Colombo, P. R. Gellert, M. G. Schwab and K. Kim, "A

Roadmap for Graphene" *Nature*, vol. 490, pp. 192-200, 2012.

[34] Y. M. Lin, K. A. Jenkins, A. V. Garcia, J. P. Small, D. B. Farmer and P. Avouris, "Operation of Graphene Transistors at Gigahertz Frequencies", *Nano Letters*, vol. 9, no. 1, pp. 422-426, 2009.

[35] L. Liao, J. Bai, R. Cheng, Y.C. Lin, S. Jiang, Y. Qu, Y. Huang and X. Duan, "Sub-100 nm Channel Length Graphene Transistors", *Nano Letters*, vol. 10, pp. 3952-3956, 2010.

[36] S. Wang J. Pu, D. S. H. Chan, B. J. Cho and K. P. Loh, "Wide Memory Window in Graphene Oxide Charge Storage Nodes", *Applied Physics Letters,* vol. 96, p.143109, 2010.

[37] A. J. Hong, E. B. Song, H. S. Yu, M. J. Allen, J. Kim, J. D. Fowler, J. K. Wassei, Y. J. Park, Y. Wang and J. Zou, "Graphene Flash Memory", *ACS Nano*, vol. 5, no. 10, pp. 7812-7817, 2011.

[38] S. Bertolazzi, D. Krasnozhon and A. Kis, "Nonvolatile Memory Cells Based on MoS_2/Graphene Heterostructures", *ACS Nano*, vol. 7, no. 4, pp. 3246-3252, 2013.

[39] A. Rani, J. M. Song, M. J. Lee and J. S. Lee, "Reduced Graphene Oxide Based Flexible Organic Charge Trap Memory Devices", *Applied Physics Letters*, vol. 101, p. 233308, 2012.

[40] D. J. Baek, M. L. Seol, S. J. Choi, D. Il Moon and Y. K. Choi, "Nonvolatile Memory with Graphene Oxide as a Charge Storage Node in Nanowire Field-Effect Transistors", *Applied Physics Letters*, vol. 100, p. 093106, 2012.

[41] S. M. Kim, E. B. Song, S. Lee, J. Zhu, D. H. Seo, M. Mecklenburg, S. Seo and Kang L. Wang, "Transparent and Flexible Graphene Charge-Trap Memory", *ACS Nano*, vol. 6 no. 9, pp. 7879-7884, 2012.

[42] P. Cappelletti, C. Golla, P. Olivo and E. Zanoni, "Flash memories", Springer, 2004.

[43] D. F. Bentchkowsky, "Memory Behavior in a Floating Gate Avalanche Injection MOS (FAMOS) Structure", *Applied Physics Letters*, vol. 18, no. 8, pp. 332-334, 1970.

[44] R. Bez, E. Camerlenghi, A. Modelli and A. Visconti, "Introduction to Flash Memory", *Proc.of the IEEE*, vol. 91, no. 4, pp 489-503, 2003.

[45] S. K. Lai, "Flash Memories: Success and Challenges*", IBM Journal of Research & Devices*, vol. 52 no. 4.5, pp 529-534, 2008.

[46] P. Pavan, R. Bez, P. Olivo and E. Zanoni, "Flash Memory Cells-an Overview", *Proc. of the IEEE*, vol. 85, no. 8, pp 1248-1274, 1997.

[47] G. Atwood, "Future Directions and Challenges for ETox Flash Memory Scaling", *IEEE*

Transactions on Device and Materials Reliability, vol. 4, no. 3, pp. 301-306, 2004.

[48] M. Lenzlinzer and E.H. Snow, "Flower-Nordheim tunneling into thermally grown SiO_2", *Journal of Applied Physics*, pp. 278-283, vol. 40, no 1, 1969.

[49] S. Lee, "Scaling Challenges in NAND Flash Device toward 10 nm Technology", *In the Proc. of International Memory Workshop*, 2012.

[50] K. Woo Lee, S. K. Choi, S. J. Chung, H. L. Lee, S. M. Yi, B. Il Han, B. I. Lee, D. H. Lee, J. H. Seo and N. Y. Park et al., "A Highly Manufacturable Integration Technology of 20 nm Generation 64 Gb Multi-Level NAND Flash Memory", *In the Digest of Technical Papers on the Symposium on VLSI Technology*, 2011, pp. 70-71.

[51] C. Y. Lu, "Future Prospects of NAND Flash Memory Technology-The Evolution from Floating Gate to Charge Trapping to 3D Stacking", *Journal of Nanoscience and Nanotechnology*, vol. 12, no. 10, pp. 7604-7618, 2012.

[52] D. Wellekens and J. Von. Houdt, "The Future of Flash Memory: is Floating Gate Technology Doomed to Lose The Race?", *Integrated Circuit Design and Technology and Tutorial* 2008, pp. 189-194.

[53] C. Y. Lu, K. Y. Hsieh and R. Liu, "Future Challenges of Flash Memory Technologies", *Microelectronic Engineering*, vol. 86, pp 283-286, 2009.

[54] L. Breuil, J. Lisoni, P. Blomme, G. Van den Bosch and J. Van Houdt, "A Novel Multilayer Inter-Gate Dielectric Enabling up to 18V Program / Erase Window for Planar NAND Flash", *In the Proc. of International Memory Workshop*, 2013

[55] Y. S. Kim, D. J. Lee, C. K. Lee, H. K. Choi, S. S. Kim, J. H. Song, D. H. Song, J. H. Choi, K. D. Suh and C. Chung, "New Scaling Limitation of the Floating Gate Cell in NAND Flash Memory", *In the Proc. of IEEE International Reliability Physics Symposium*, 2010, pp. 599-603.

[56] Y. Li and K. N. Quader, "NAND Flash Memory: Challenges and Opportunities", *Computer*, vol. 46, no. 8, pp. 23-29, 2013.

[57] J. H. Yoon and G. A. Tressler, "Advanced Flash Technology Status, Scaling Trends & Implications to Enterprise SSD Technology Enablement", presentation *Flash Memory Summit* 2012. Available online at

http://www.flashmemorysummit.com/English/Collaterals/Proceedings/2012/20120821_TA12_Yoon_Tressler.pdf

[58] C. Sandhya, U. Ganguly, N. Chattar, C. Olsen, S. M. Seutter, L. Date, R. Hung, J. M. Vasi and S. Mahapatra, "Effect of SiN on Performance and Reliability of Charge Trap Flash (CTF) Under Fowler-Nordheim Tunneling Program/Erase Operation", *IEEE Electron Device Letters*, vol. 30, no. 2, 2009.

[59] S. Tiwari, F. Rana, H. Hanafi, A. Hartstein, E. F. Crabbe and K. Chan, "A Silicon Nanocrystals Based Memory", *Applied Physics Letters*, vol. 68, no. 10, pp. 1377-1379, 1996.

[60] C. Gerardi, G. Molas , G. Albini , E. Tripiciano , M.Gely , A. Emmi, O. Fiore, E. Nowak, D. Mello and M. Vecchio, "Performance and reliability of a 4Mb Si nanocrystal NOR Flash memory with optimized 1T memory cells", *In the Proc. of International Electron Device Meeting*, 2008.

[61] J. H. Chen, Y. Q. Wang, W. J. Yoo, Y. C. Yeo, G. Samudra, D. S. Chan, A. Y. Du and D. L. Kwong, "Nonvolatile Flash Memory Device Using Ge Nanocrystals Embedded in HfAlO High-K Tunneling and Control Oxides: Device Fabrication and Electrical Performance", *IEEE Transactions on Electron Devices*, vol. 51, no. 11, pp. 1840-1848, 2004.

[62] W. K. Choi, W. K. Chim, C. L. Heng, L. W. Teo, V. Ho, V. Ng, D. A. Antoniadis and E. A. Fitzgerald, "Observation of Memory Effect in Germanium Nanocrystals Embedded in an Amorphous Silicon Oxide Matrix of a Metal–Insulator-Semiconductor Structure", *Applied Physics Letters*, vol. 80, no. 11, pp. 2014-2016, 2002.

[63] W. Guan, S. Long, M. Liu, Q. Liu, Y. Hu, Z. Li and R. Jia, " Modeling of Retention Characteristic for Metal and Semiconductor Nanocrystal Memories", *Solid State Electronics*, vol. 51, pp. 806-811, 2007.

[64] C. Lee , U. Ganguly , V. Narayanan , H. Hou , J. Kim and E. C. Kan, "Asymmetric Electric Field Enhancement in Nanocrystal Memories", *IEEE Electron Device Letters*, vol. 26, no. 12, pp.879-881, 2005.

[65] Z. Liu, C. Lee, V. Narayanan, G. Pei and Edwin C. Kan, "Metal Nanocrystal Memories-Part II: Electrical Characteristics", *IEEE Transactions on Electron Devices*, vol. 49, no. 9, pp. 1614-1622, 2002.

[66] C. Lee, J. Meteer, V. Narayanan and E. D. Kan, "Investigation on Process Dependence of Self-Assembled Metal Nanocrystals", *In the Proc. Materials Research Society Symposium*, vol. 737, pp f8.18.1-f8.18.6.

[67] C. LEE, J. Meteer, V. Narayanan and E. D. Kan, "Self-Assembly of Metal Nanocrystals on Ultrathin Oxide for Nonvolatile Memory Applications", *Journal of Electronic Materials*, vol. 34, no. 1, pp 1-11, 2005.

[68] L. I. Maissel and R. Glang, *Handbook of Thin Film Technology*, McGraw-Hill, p. 8-32, 1970.

[69] P. K. Singh, R. Hofmann, K. K. Singh, N. Krishna and S. Mahapatra, "Performance and Reliability of Au and Pt Single-Layer Metal Nanocrystal Flash Memory under NAND(FN/FN) Operation", *IEEE Transactions on Electron Devices*, vol. 56, no. 9, pp.

2065-2072, 2009.

[70] Y. N. Chen, K. E. J. Goh, X. Wu, Z. Z. Lwin, P. K. Singh, S. Mahapatra, and K. L. Pey, "Temperature-dependent Relaxation Current on Single and Dual Layer Pt Metal Nanocrystal-based Al_2O_3/SiO_2 Gate Stack", *Journal of Applied Physics*, vol. 112, no. 10, p.104503, 2012.

[71] P. K . Singh, G. Bisht, K. Auluck, M. Sivatheja, R. Hofmann, K. K. Singh and S. Mahapatra, "Performance and Reliability Study of Single-Layer and Dual-Layer Platinum Nanocrystal Flash Memory Devices under NAND Operation", *IEEE Transactions on Electron Devices*, vol. 57, no. 8, pp. 1829-1837, 2010.

[72] P. K. Singh, G. Bisht, R. Hofmann, K. Singh, N. Krishna and S. Mahapatra, "Metal Nanocrystal Memory With Pt Single and Dual-Layer NC with Low-Leakage Al_2O_3 Blocking Dielectric", *IEEE Electron Device Letters*, vol. 29, no. 12, pp. 1389-1391, 2008.

[73] F. M. Yang, T. C. Chang, P. T. Liu, U. S. Chen, P. H. Yeh, Y. C. Yu, J. Y. Lin, S. M. Sze and J. C. Lou, "Nickel Nanocrystals with HfO_2 Blocking Oxide for Nonvolatile Memory Application", *Applied Physics Letters*, vol. 90, p. 222104, 2007.

[74] D. Shahrjerdi, D. I. G. Gutierrez and S. K. Banerjee, "Fabrication of Ni Nanocrystal Flash Memories Using a Polymeric Self-Assembly Approach", *IEEE Electron Device Letters*, vol. 28, no. 9, pp. 793-795, 2007.

[75] J. Lee and D. L. Kwong, "Metal Nanocrystal Memory With High-K Tunneling Barrier for Improved Data Retention", *IEEE Transactions on Electron Devices*, vol. 52, no. 4, pp. 507-509, 2005.

[76] C. W. Hu, T. C. Chang, C. H. Tu, P. K. Shueh, C. C. Lin, S. M. Sze, T. Y. Tseng and M. C. Chen, "Cobalt Nanodots Formed by Annealing the CoSiO Layer for the Application of the Nonvolatile Memory", *Applied Physics Letters* vol. 94, p. 102106, 2009.

[77] Y. Zhu, D. Zhao, R. Li and J. Liu, "Self-aligned TiSi2/ Si Hetero Nanocrystal Nonvolatile Memory", *Applied Physics Letters*, vol. 88, p.103507, 2006.

[78] S. W. Ryu, J. W. Lee, J. W. Han, S. Kim and Y. K. Choi, "Designed Work function Engineering of Double-Stacked Metal Nanocrystals for Nonvolatile Memory Application", *IEEE Transactions on Electron Devices*, vol. 56, no. 3, pp.377-382, 2009.

[79] J. Yater, M. Suhail, S. T. Kang, J. Shen, C. Hong, T. Merchant, R. Rao, H. Gasquet, K. Loiko, B. Winstead, S. Williams, M. Rossow, W. Malloch, R. Syzdek and G. Chindalore, "16Mb Split Gate Flash Memory with Improved Process Window" *In the Proc. of International Memory Workshop*, 2009.

[80] W. K. Choia, W. K. Chim, C. L. Heng, L. W. Teo, Vincent Ho, V. Ng, D. A. Antoniadis and E. A. Fitzgerald, "Observation of Memory Effect in Germanium Nanocrystals Embedded in an Amorphous Silicon Oxide Matrix of a Metal-Insulator-Semiconductor

Structure", *Applied Physics Letters*, vol. 80, no. 11, pp. 2014-2016, pp. 2002.

[81] C. W. Hu, T. C. Chang, C. H. Tu, Y. H. Huang, C. C. Lin, M. C. Chen, F.S. Huang, S. M. Sze and T. Y. Tseng, "High Density Ni Nanocrystals Formed by Co evaporating Ni and SiO_2 Pellets for the Nonvolatile Memory Device Application", *Electrochemical and Solid-State Letters*, vol. 13, no. 3, pp. H49-H51, 2010.

[82] D. U. Lee, H. J. Lee, E. K. Kim, H. W. You and W. J. Cho, "Low Operation Voltage and High Thermal Stability of a WSi_2 Nanocrystal Memory Device using an $Al_2O_3/HfO_2/Al_2O_3$ Tunnel Layer", *Applied Physics Letters*, vol. 100, p. 072901, 2012.

[83] C. C. Lin, T. C. Chang, C. Hao Tu, W. Ren Chen, C. W. Hu, S. M. Sze, T. Y. Tseng, S. C. Chen and J. Y. Lin, "Charge Storage Characteristics of Mo Nanocrystal Dependence on Mo Oxide Reduction", *Applied Physics Letters*, vol. 93, p.222101, 2008.

[84] W. Banerjee, S. Maikap, T.-C. Tien, W.-C. Li and J.-R. Yang, "Impact of Metal Nano Layer Thickness on Tunneling Oxide and Memory Performance of Core-Shell Iridium-Oxide Nanocrystals" *Journal of Applied Physics*, vol. 110, p. 074309, 2011.

[85] D. B. Farmer and R. G. Gordon, "High Density Ru Nanocrystal Deposition For Nonvolatile Memory Applications", *Journal of Applied Physics*, vol. 101, p. 124503, 2007.

[86] K. S. Seol, S. J. Choi, J. Y. Choi, E. J. Jang, B. K. Kim, S. J. Park, D. G. Cha, I. Y. Song, J. G. B. Park, Y. Park and S. Ho Choi, "Pd-Nanocrystal-Based Nonvolatile Memory Structures With Asymmetric SiO_2/ HfO_2 Tunnel Barrier", *Applied Physics Letters*, vol. 89, p.083109, 2006.

[87] Y. H. Lin, C. H. Chien, C. T. Lin, C. Y. Chang and Tan-Fu Lei, "High-Performance Nonvolatile HfO2 Nanocrystal Memory", *IEEE Electron Device Letters*, vol. 26, no. 3, pp. 154-156, 2005.

[88] H. C. You, T. H. Hsu, F. H. Ko, J. W. Huangand and T. F. Lei, "Hafnium Silicate Nanocrystal Memory using Sol–Gel-Spin-Coating Method", *IEEE Electron Device Letters*, vol. 27, no. 8, pp. 644-646, 2006.

[89] S.Sadana, "Metal nanaocrystals for flash memory and contact resistance application", M.Tech Thesis, Department of electrical engineering, IIT Bombay, 2011.

[90] M. Bhaisare, A. Misra, M. Waikar and Anil Kottantharayil, "High quality Al2O3 dielectric films deposited by pulsed- DC reactive sputtering technique for high-k applications", Nanoscience and Nanotechnology Letters, vol. 4 no.6, pp. 645 – 650, 2012.

[91] M.Waikar, "Fabrication and Electrical Characterization of Floating Gate NAND Flash Transistor", M.Tech Thesis, Department of electrical engineering, IIT Bombay, 2012.

[92] V. Satya Suresh, "Optimization of Silicon Nitride For SONOS Flash Memories by Trap Characterization", M.Tech Thesis, Department of electrical engineering, IIT Bombay, 2011.

[93] S. K. Samanta, W. J. Yoo, G. Samudra, E. S. Tok, L. K. Bera and N. Balasubramanian, "Tungsten Nanocrystals Embedded in High-K Materials for Memory Application", *Applied Physics Letters* 87, 113110, 2005.

[94] J. C. D. Delgado, Y. A. Kim, T. Hayashi, A. M. Gomez, M. Hofmannn, H. Muramatsu, M. Endo, H. Terrones, R. D. Shull, M. S. Dresselhaus and M. Terrones,"Thermal Stability Studies of CVD-grown Graphene Nanoribbons: Defect Annealing and Loop Formation", *Chemical Physics Letters*, vol. 469, pp. 177-182, 2009.

[95] B. Standley, W. Bao, H. Zhang, J. Bruck, C. N. Lau and M. Bockrathet, "Graphene-based Atomic-Scale Switches", *Nano Letters*, vol. 8, no. 10, pp. 3345-3349, 2008.

[96] T. J. Echtermeyer, M.C. Lemme, M. Baus, B. N. Szafranek, A. K. Geim and H. Kurz, "Nonvolatile Switching in Graphene Field-Effect Devices", *IEEE Electron Device Letters*. vol. 29, no. 8, pp. 952-954, 2008.

[97] E. Song, B. Lian, S. M. Kim, S. Lee, T. K. Chung, M. Wang, C. Zeng, G. Xu, K. Wong, Y. Zhou, et al., "Robust Bi-stable Memory Operation in Single-layer Graphene Ferroelectric Memory", *Applied Physics Letters*, vol. 99, p. 042109, 2011.

[98] H. Hibino, H. Kageshima, M. Kotsugi, F. Maeda, F.-Z. Guo, and Y. Watanabe, "Dependence of Electronic Properties of Epitaxial Few-layer Graphene on the Number of Layers Investigated by Photoelectron Emission Microscopy", *Physical Review B*, vol. 79, p. 125437, 2009.

[99] W. Zhu, V. Perebeinos, M. Freitag, and P. Avouris, "Carrier Scattering, Mobilities and Electrostatic Potential in Monolayer, Bilayer and Trilayer Graphene", *Physical Review B*. vol. 80, p. 235402, 2009.

[100] C. Ramirez, L. Garzón, P. Miranzo, M. I. Osendi and C. Ocal, "Electrical Conductivity Maps in Graphene Nanoplatelet/Silicon Nitride Composites using Conducting Scanning Force Microscopy", *Carbon*, vol. 49, pp. 3873-380, 2011

[101] S. Stankovich, D. A. Dikin, R. D. Piner, K. A. Kohlhaas, A. Kleinhammes, Y. J. Y. Wu, S. T. Nguyen and R. S. Ruoff, "Synthesis of Graphene-based Nanosheets via Chemical Reduction of Exfoliated Graphite Oxide", *Carbon*, vol.45, pp.1558-1565, 2007.

[102] S. M. Sze, Physics of Semiconductor Devices (2nd Edition), John Wiley and Sons, 1981.

[103] T. Lee and S. K. Banerjee, "Device characteristics of HfON Charge-trap Layer Nonvolatile Memory", J*ournal of Vacuum Science & Technology B* . vol. 28, no. 5, pp.

1005-1010, 2010.

[104] C. L. Lu, C. P. Chang, Y. C. Huang, J, M. Lu, C. C. Hwang and M. F. Lin, "Low-Energy Electronic Properties of the AB-stacked Few-layer Graphites", *J. Physical Condens. Matter*, vol. 18, pp. 5849-5859, 2006.

[105] R. van Schaijk, M. van Duuren, N. Akil, A. Huerta, S. Beckx, F. Neuilly, Z. Rittersma, M. Slotboom, S. Van Elshocht and J. Wouters, "A Novel SONOS Memory with HfSiON/Si_3N_4/HfSiON Stack for Improved Retention", *in Proc. of IEEE Non-Volatile Semiconductor Memory Workshop*, 2006, pp. 50-51.

[106] K. Lee, M. Kang, S. Seo, D. Kang, S. Kim, D. H. Li and H. Shin, "Activation Energies (Ea) of Failure Mechanisms in Advanced NAND Flash Cells for Different Generations and Cycling", *IEEE Transactions on Electron Devices*, vol. 60, no. 3, pp. 1099-1107, 2013.

[107] T. Ushiki, K. Kawai, I. Ohshima and T. Ohmi, "Chemical Reaction Concerns of Gate Metal with Gate Dielectric in Ta Gate MOS Devices: An Effect of Self-Sealing Barrier Configuration Interposed between Ta and SiO_2", *IEEE Transactions on Electron Devices* vol. 47, no. 11, pp. 2201-2207, 2000.

[108] V. Misra, G. Lucovsky and G. Parsons, "Issues in High-K Gate Stack Interfaces", *MRS Bulletin*, pp. 212-216, 2003.

[109] C. Ren, H. Y. Yu, Y. C. Yeo, J. F. Kang, X. P. Wang, H. H. H. Ma, M. F. Li, D. S. H. Chan and D. L. Kwong, "A Dual-Metal Gate Integration Process for CMOS With Sub-1-nm EOT HfO_2 by Using HfN Replacement Gate", *IEEE Electron Device Letters*, vol. 25, no. 8, pp. 580-582, 2004.

[110] N. J. Lee, J. W. Yoo, Y. J. Choi, C. J. Kang and D. Y. Jeon, "The Interlayer Screening Effect of Graphene Sheets Investigated by Kelvin Probe Force Microscopy", *Applied Physics Letters*, vol. 95, p. 222107, 2009.

[111] S. S. Datta, D. R. Strachan, E. J. Mele and A. T. Charlie Johnson, "Surface Potentials and Layer Charge Distributions in Few-layer Graphene Films", *Nano Letters*, vol. 9, no. 1, pp. 7-11, 2009.

[112] T. Filleter, K. V. Emtsev, T. Seyller and R. Bennewitz, "Local Work Function Measurements of Epitaxial Graphene", *Applied Physics Letters*, vol. 93, p. 133117, 2008.

[113] Y. J. Yu, Y. Zhao, S. Ryu, L. E. Brus, K. S. Kim and P. Kim, "Tuning the Graphene Work Function by Electric Field Effect", *Nano Letters*, vol. 9, no. 10, pp. 3430-3434, 2009.

[114] K. Ihm, J. Tim, K. J. Lee, J. W. Kwon, T. H. Kang, S. Chung, S. Bae, J. H. Kim, B. H. Hong and G. Y. Yeom, "Number of Graphene Layers as a Modulator of the Open-circuit Voltage of Graphene-based Solar Cell", *Applied Physics Letters*, vol. 97, p. 032113, 2010.

[115] G. Giovannetti, P. A. Khomyakov, G. Brocks, V. M. Karpan, J. van den Brink and P. J. Kelly, "Doping Graphene with Metal Contacts", *Physical Review Letters*, vol. 101, p. 026803, 2008.

[116] D. Ziegler, P. Gava, J. G.uttinger, F. Molitor, L. Wirtz, M. Lazzeri, A. M. Saitta, A. Stemmer, F. Mauri and C. Stampfer, "Variations in the Work Function of Doped Single- and Few-layer Graphene Assessed by Kelvin Probe Force Microscopy And Density Functional Theory", *Physical Review B*., vol. 83, p. 235434, 2011.

[117] K. Pi, K. M. McCreary, W. Bao, W. Han, Y. F. Chiang, Y. Li, S. W. Tsai, C. N. Lau, and R. K. Kawakami, "Electronic Doping and Scattering by Transition Metals on Graphene", *Physical Review B*., vol. 80, p. 075406, 2009.

[118] Y. Tang, Z. Yang and X. Dai, "Trapping of Metal Atoms in the Defects on Graphene", *Journal of Chemical Physics Letters*, vol.135, no. 22, p. 224704, 2011.

[119] E. J. H. Lee, K. N. Balasubramanian, R. T. Weitz, M. Burghard and K. Kern, "Contact and edge effects in graphene devices", *Nature Nanotechnology*, vol.3, pp. 486-490, 2008.

[120] J. K. Park, S. M. Song, J. H. Mun and B. J. Cho, "Graphene Gate Electrode for MOS Structure-Based Electronic Devices", *Nano Letters*, vol. 11, no. 12, pp. 5383-5386, 2011.

[121] J. K. Park, S. M. Song, J. H. Mun and B. J. Cho, "Dramatic Improvement of high-K Gate Dielectric Reliability by Using Mono-layer Graphene Gate Electrode", *In the proc. of Symposium on VLSI Technology* 2012, p.31.

[122] S. M. Song, J. K. Park, O. J. Sul and B. J. Cho, "Determination of Work Function of Graphene under a Metal Electrode and its Role in Contact Resistance", *Nano Letters*, vol. 12, no.8, pp. 3887-3892, 2012.

[123] X. Wang and L. Zhi, K. Mullen, "Transparent Conductive Graphene Electrodes for Dye-Sensitized Solar Cells", *Nano Letters,* vol. 8, no. 1, pp. 323-327, 2008.

[124] L. Liu, S. Ryu, M. R. Tomasik, E. Stolyarova, N. Jung, M. S. Hybertsen, M. L. Teigerwald, L. E. Brus and G. W. Flynn, "Graphene Oxidation:Thickness-dependent Etching and Strong Chemical Doping", *Nano Letters*. vol. 8, no. 7, pp. 1965-1970, 2008.

[125] M. J. Allen, V. C. Tung and R. B. Kaner, "Honeycomb Carbon: A Review of Graphene", *Chemical Review*, vol. 110, pp. 132-145, 2010.

[126] E. H. Nicollian and J. R. Brews, Wiley, New York, p. 468, 1982

[127] J. Westlinder, G. Sjoblom and J. Olsson, "Variable Work Function in MOS Capacitors Utilizing Nitrogen-controlled TiNx Gate Electrodes", *Microelectronics Engineering*, vol.75, pp. 389-396, 2004.

[128] G. S. Lujan, T. Schram, L. Pantisano, J. C. Hooker, S. Kubicek, E. Rohr, J. Schuhmacher, O. Kilpelä, H. Sprey, S. De Gendt and K. De Meyer, "Impact of ALCVD and PVD Titanium Nitride Deposition on Metal Gate Capacitors", *In Proc. European Solid State Device Research Conference*, 2002, pp. 583-586.

[129] W. A. Hill and C. C. Coleman, "A Single-Frequency Approximation for Interface-State Density Determination", *Solid State Electronics*. vol. 23, pp. 987-993, 1980.

[130] E. Rosenbaum, J. C. King and C. Hu, "Accelerated Testing of Sio_2 Reliability", *IEEE Transaction on Electron Dev*ices. vol. 43, no. 1, pp. 70-80, 1996.

[131] Y. Park, V. Choong, Y. Gao, B. R. Hsieh and C. W. Tang, "Workfunction of Indium Tin Oxide Transparant Conductor Measured by Photo Electon Spectroscopy", *Applied Physics Letters*, vol. 68, p. 2699, 1996.

[132] J. S. Kim, B. Lagel, E. Moons, N. Johansson, I. D. Baikie, W. R. Salaneck, R. H. Friend and F. Cacialli, "Kelvin Probe and Ultraviolet Photoemission Measurements of Indium Tin Oxide Work Function: A Comparison", *Synthetic Metals* vol. 111-112, pp. 311-314, 2000.

[133] S. Gutmann, M. Conrad, M. A. Wolak, M. M. Beerbom and R. Schlaf, "Workfunction Measurements on Nano-Crystalline Zinc Oxide Surfaces", *Journal of Applied Physics*, vol. 111, p. 123710, 2012.

[134] P. V. Kumar, M. Bernardi and J. C. Grossman, "The Impact of Functionalization on the Stability, Work Function and Photoluminescence of Reduced Graphene Oxide", *ACS Nano*, vol.7, no. 2, pp. 1638-1645, 2013.

[135] M. Acik, G. Lee, C. Mattevi, M. Chhowalla, K. Cho and Y. J. Chabal, "Unusual Infrared Mechanism in Thermally Reduced Graphene Oxide", *Nature Materials*. vol. 9, pp. 840.845, 2010.

[136] A. Ganguly, S. Sharma, P. Papakonstantinou and J. Hamilton, "Probing the Thermal Deoxygenation of Graphene Oxide using High-resolution in situ X-ray-based Spectroscopies", *Journal of Phyical Chemistry*, vol. 115, pp. 17009-17019, 2011.

[137] D. Yang, A. Velamakanni, G. L. Bozoklu, S. Park, M. Stoller, R. D. Piner, S. Stankovich,

I. Jung, D. A. Field, C. A. Ventrice Jr. D and R. S. Ruoff, "Chemical Analysis of Graphene oxide films after Heatand Chemical Treatments by X-ray Photoelectron and Micro-Raman Spectroscopy", *Carbon*, vol. 47, pp. 145–152, 2009.

[138] S. W. Lee, C. Mattevi, M. Chhowalla and R. M. Sankaran, "Plasma-Assisted Reduction of Graphene Oxide at Low Temperature and Atmospheric Pressure for Flexible Conductor Applications", *Journal of Physical Chemistry Letters*, vol. 3, pp. 772-777, 2012.

[139] B. Y. Tsui and C. F. Huang, "Wide Range Work Function Modulation of Binary Alloys for MOSFET application", *IEEE Electron Device Letters*, vol 24, no. 3, pp 153-155, 2003.

[140] H.Y. Yu, C. Ren, Y. C. Yeo, J. F. Kang, X. P.Wang, H. H. H. Ma, M. F. Li, D. S. H. Chan and D. L. Kwong, "Fermi Pinning Induced Thermal Instability of Metal Gate Work Functions", *IEEE Electron Device Letters*, vol. 25, no. 5, pp. 337-839, 2004.

[141] F. Schwierz, "Graphene Transistors", *Nature Nanotechnology*, vol. 5, pp. 487-496, 2010.

[142] Y. C. Bae, H. Osanai, K. Ohno, M. Sluiter and Y. Kawazoe, "All Electron Mixed-basis Calculation of Structurally Optimized Titanium Nitride Clutters", *Materials Transactions*, vol. 43, no. 3, pp. 482-484, 2002.

[143] C. Ma, S. Ishihara, H. Soejima, N. Nishiyamam and A. Inoue, "Formation of New Ti-based Metallic Glassy", *Materials Transactions* vol. 45, no. 5, pp. 1802-1806, 2004.

[144] S. Entani, S. Sakai, Y. Matsumoto, H. Naramoto, T. Hao and Y. Maeda, "Interface Properties of Metal/Graphene Heterostructures Studied by Micro-Raman Spectroscopy", *Journal of Physical Chemistry C*. vol. 114, pp. 20042-20048, 2010.

[145] D. Chen, H. Feng and J. Leng, "Graphene Oxide: Preparation, Functionalization, and Electrochemical Applications", *Chemical Review*, vol. 112, pp. 6027-6053, 2012.

[146] Available onlne at http: //www.usa-graphite.com/basic-page/graphene

[147] Y. Shen, S. Yang, P. Zhou, Q. Sun, P. Wang, L. Wan, J. Li, L. Chen, X. Wang, S. Ding, D. W. Zhang, "Evolution of the band-gap and optical properties of graphene oxide with controllable reduction level" *Carbon*, vol. 62, pp. 157-164, 2013

[148] H. Huang, Z. Li, J. She, W. Wang, "Oxygen density dependent band gap of reduced graphene oxide" *Journal of Applied Physics*, vol. 111, p. 054317, 2012.

[149] S. Park and S.Ruoff, "chemical methods for the production of graphenes", *Nature Nanotechnology*, vol. 4, pp. 217-224, 2009.

[150] D. Zhan, Z. Ni, W. Chen, L. Sun, Z. Luo, L. Lai, T. Yu, A. T. S. Wee, Z. Shen, "Electronic structure of graphite oxide and thermally reduced graphite oxide" *Carbon*, vol. 49, pp. 1362-1366, 2011

List of Publications

(1) Abhishek Mishra, Amritha Janardanan, Manali Khare, Hemen Kalita, and Anil Kottantharayil, "Reduced Multilayer Graphene Oxide Floating Gate Flash Memory With Large Memory Window and Robust Retention Characteristics" *IEEE Electron Device Letters*, vol. 34, no. 9, pp. 1136- 1138, 2013. **DOI**: http://dx.doi.org/10.1109/LED.2013.

(2) Abhishek Misra, Mayur Waikar, Amit Gour, Hemen Kalita, Manali Khare, Mohammed Aslam, and Anil Kottantharayil, "Work function tuning and improved gate dielectric reliability with multilayer graphene as a gate electrode for metal oxide semiconductor field effect device applications" *Applied Physics Letters*, vol. 100, p. 233506, 2012. **DOI**: http://dx.doi.org/10.1063/1.4726284

(3) Abhishek Misra, Hemen Kalita, Anil Kottantharayil "Workfunction Modulation and Thermal Stability of Reduced Graphene Oxide Gate Electrodes in MOS Devices" *ACS Applied Materials and Interfaces*, **DOI**: 10.1021/am404649a

(4) Abhishek Misra, Hemen Kalita, Mayur Waikar, Amit Gour, Meenakshi Bhaisare, Manali Khare, Mohammed Aslam and Anil Kottantharayil, "Multilayer Graphene as Charge Storage Layer in Floating Gate Flash Memory" *In the Proc. of International Memory Workshop*, 2012. **DOI**: 10.1109/IMW.2012.6213626

(5) Abhishek Misra, Sunny Sadana, Satya Suresh, Meenakshi Bhaisare, Senthil Srinivasan, Mayur Waikar, Amit Gaur and Anil Kottantharayil, "Effect of Different High-K Dielectrics on the Pt Nanocrystal Formation Statistics (size, density and area coverage) for Flash Memory Application" *In the Proceedings of the MRS Fall Meeting* 2010, vol. 1288. **DOI**: dx.doi.org/10.1557/opl.2011.208

(6) Abhishek Mishra, Manali Khare, Hemen Kalita, M. Aslam, Anil Kottantharayil, "Extraction of Graphene/TiN Work Function Using Metal Oxide Semiconductor (MOS) Test Structure", presented at the *International Conference on Emerging Electronics (ICEE 2012)*, IIT Bombay, Dec. 15 - 17, 2012. **DOI**: 10.1109/ICEmElec.2012.6636272

(7) Abhishek Misra, Mayur Wakikar, Amit Gour, Meenakshi Bhaisare and Anil Kottantharayil "SiO_2/Al_2O_3 dielectric stack with low power pulsed-DC reactive sputtered High-K Al_2O_3 as blocking dielectric for NAND flash application" Presented at the International Workshop on Physics of Semiconductor Devices (*IWPSD)* 2011 IIT Kanpur, India.

Manuscript under review

(1) Abhishek Misra, Hemen Kalitha, Anil Kottantharayil "On the Investigation of Metal / dielectric interaction and oxide charges with graphene/TiN gate electrode devices" a revised version is to be submitted to *IEEE Transaction on Nanotechnology*.