Detailed investigation of the charge storage in size-controlled Si nanocrystals

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum naturalium

(Dr.rer. nat)

genehmigt durch die

Mathematisch-Naturwissenschaftlich-Technische Fakultät
(mathematisch-naturwissenschaftlicher Bereich)
der Martin-Luther-University Halle-Wittenberg

von Tiezheng Lu
geboren am 30.01.1977 in Hunan, China

Gutachter:

1. Prof. Dr. Margit Zacharias
2. Prof. Dr. Ulrich Goesele

Halle (Saale), den 17 November 2006
verteidigt am 02.07.2007

urn:nbn:de:gbv:3-000012040
[http://nbn-resolving.de/urn/resolver.pl?urn=nbn%3Ade%3Agbv%3A3-000012040]

ACKNOWLEDGEMENTS

This work would not have been possible without the help of many people. This thesis is the result of three years of work whereby I have been accompanied and supported by a lot of people at Max Planck Institute of Microstructure Physics. Now it is the time that I can express my gratitude to all of them.

First, I would like to thank my research advisor, **Prof.Dr. Margit. Zacharias** gave me tremendous support and invaluable advice for me to finish the graduate study. I have benefited a lot from her continuous encouragement. I am also very impressed by her diligence and enthusiasm for the scientific pursuit. I feel very fortunate for the opportunity to have her as my research advisor.

I wish to express my gratitude to **Prof.Dr. Ulrich. Goesele** who always kept eyes on the progress of my work and was available when I needed his advises. His insightful advices and kind guidance accompanied me for the past three years. I enjoyed his broad scientific background and his focussed nature. There were many useful lessons to learn. His enthusiasm and integral view on research has made a deep impression on me.

I am also grateful to **Dr. Marin. Alexe**, who guided me into the community of electrical characterization. Many thanks for his technical guidance throughout my graduate research. He has been supportive of my work as well and I also enjoyed discussing any kind of questions with him.

Special thanks go to **Dr. Roland. Scholz**. He did all TEM images for my research work and took great effort with valuable comments on further analysis.

I wish to acknowledge Dr. V. Talalaev and Mr. K. P. Meyer for technical supporting me in my research.

The "0D and 1D Nanomaterias" group at MPI-Halle substantially contributed to the development of this work. The strict and extensive comments and the many discussions within the group had a direct impact on the final form and quality of this thesis. I am indebted to Dr. J. Heitmann, Dr. F. Kolb (now they are at Infineon company in Dresden) and Dr. S. Richer(now at Zeiss company in Jena), the previous members of this group, for not only sharing their experience for research, but also patiently supervisor me to do experiments at my beginning time. I thank all current members of this group from the bottom of my heart.

I feel a deep sense of gratitude for Dr. Hofmeister, Dr. Breitenstein, Dr. Erfurth, and all colleagues in our Experimental Department II. I would therefore express my faithful thanks to the researchers in our institute, whom I had pleasure to work with.

I deeply thank all Chinese friends in Halle for their support and friendship. During the last 3 years, I was continuously encouraged and supported by them. I would like to thank Professor Jun Shen at Tongji University, China, Professor K.N.Tu at University of California at Los Angeles, US for the discussion, help and collaboration. Their supports made my research life enjoyable.

Last and most important,this thesis is dedicated to my wife Lili Zhao and my parents. Without their endless love and sacrifice, nothing would have been accomplished.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Introduction

1.1 State of the art for non-volatile memory

1.1.1 Basics of non-volatile memory devices

In the last twenty years, microelectronics has been strongly developed, concerning higher integration density and lower cost. Research on microelectronics has been focused on the miniaturization of devices with the reliability constraint on the device. For instance, a computer CPU contains more than 100 millions of transistors, and microelectronics researchers concentrate to scale devices down while keeping stability constrains. In this frame with the memory devices are paramount importance. Memory chips with low power consumption and low cost have attracted more and more attention due to the booming market of portable electronic devices and are indispensable components of modern life as shown in Fig. 1.1. Memory chips are used in PC, mobile phones, digital cameras, smart-media, networks, automotive systems, and global positioning systems. All semiconductor memories can be divided into two main types, both based on CMOS technology, volatile and non-volatile memory. Volatile memory is fast but loses its contents when power is removed. Non-volatile memory is slower but retains the information without power supply. Typically volatile memories are Static Random Access Memory (SRAM) and Dynamic Random Access Memory (DRAM). The most important device for semiconductor industry is the Metal-Oxide-Semiconductor Field Effect Transistor (MOSFET), which was first reported by Kahng and Atalla in 1960 [1]. MOSFET and related integrated circuits now constitute about 90% of the semiconductor device market. A nonvolatile memory (NVM) device based on semiconductors is a typical MOS transistor that has a source, a drain, and a gate. Tunnelling is the process by which a NVM can be either erased or programmed and is usually dominant in thin insulating layer. Storage of the charge on the floating gate allows the threshold voltage (V_T) to be electrically altered between a low and a high value to represent logic 0 and 1, respectively. Typical requirement for a non-volatile memory is a ten years data retention time without power supply.

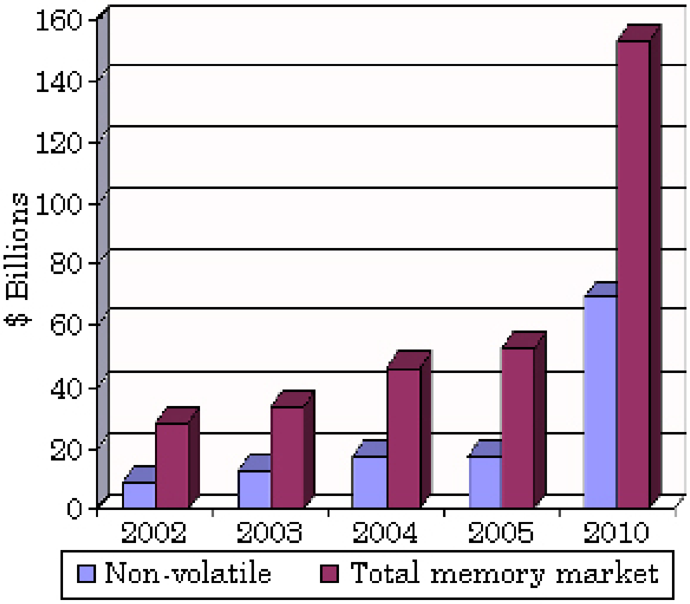


Figure 1.1: Worldwide Non-Volatile and Total Memory Markets Prediction, 2002-2010

1.1.2 Various concepts of non-volatile memories and their realization

Most of non-volatile memories used today consist of four types: flash memory, FeRAM (Ferro-electric RAM), MRAM (Magnetic RAM) and phase change memory. Flash memory is the most suitable structure of non-volatile memory, since one cell consists of only one transistor. Charge can be stored in this floating node and determines the state of the memory by changing the threshold voltage of the transistor. A FeRAM memory cell generally consists of one transistor and one capacitor. A MRAM cell needs a transistor and a magnetic tunnel junction. This results in a certain incompatibility of FeRAM and MRAM with standard silicon technology. Phase change storage seems to be a new concept. Phase change materials can be switched rapidly back and forth between amorphous and crystalline phases by applying appropriate heat pulses. They are widely used in optical information technologies (DVD, CD-ROM and so on). Recently they have also been considered for non-volatile memory applications. IBM researchers show erasable thermal phase-change recording at a storage density of 3.3 Tb inch^{-2}[2]. However, the requirement of a high current to switch the phase is the main limitation of the phase change memory. Although the last three types of non-volatile memory have better features and show better programming performance, they could have to become key memory components of chips in the near future. However, as for further application, the physics and material properties of these kinds of devices mean that scaling down to nanometer cell sizes will be a big challenge.

1.1.3 Flash memory development

A flash memory chip actually is an array of floating gate transistors. It is structurally different from a standard MOSFET in its floating gate, which is electrically isolated. Electrons are transferred from the floating gate to the substrate by tunnelling through a thin insulating layer. It has two gates instead of just one. One gate is the control gate (CG) like in MOS transistors, but the second is a floating gate (FG) that is insulated all around by an oxide layer. The FG is located between the CG and the substrate as shown in Fig.1.2. Because the FG is isolated by its insulating oxide layer, any carriers placed on it get trapped there and thus store the information. The cross-section scheme of the continuously developing design of flash memory is presented in figure 1.3 with the polysilicon continuous floating gate (a) and the discrete storage nodes (b,c). The polysilicon floating gate device is the first model for flash memory, and was invented by Kahng and Sze in 1967 at Bell Labs. The individual storage node device, making use of traps in Si_3N_4, was proposed by Wegener *et al.* in 1967. A widely exploited non-volatile memory is using a continuous layer of polycrystalline silicon as a trap site for carriers. A poly-silicon floating gate or Si_3N_4 charge trap layer is sandwiched between a tunnel oxide and an inter-poly oxide as a control oxide to form a charge storage layer. In 1995, S. Tiwari *et al.* from IBM studied the properties of a MOSFET with embedded Si nanocrystals in the gate prepared by CVD [3]. It was a creative work and opened up a new research direction of nanocrystal-based memory.

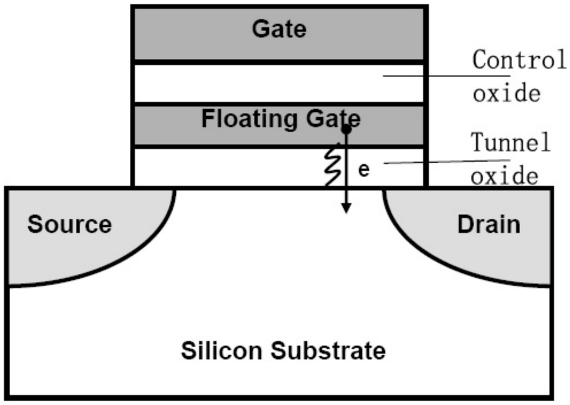


Figure 1.2: A typical floating gate memory structure. Other than traditional MOSFET, it has two gates: floating gate and control gate. The trapping layer is sandwiched between tunnel oxide and control oxide.

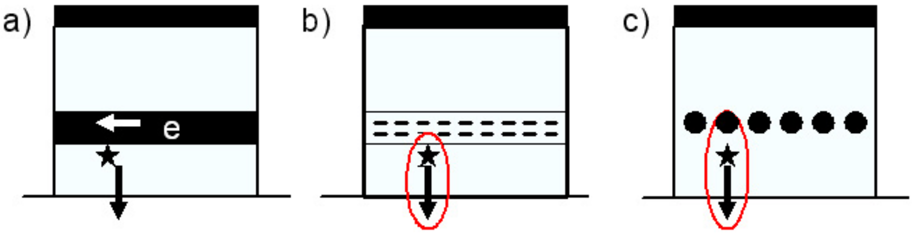


Figure 1.3: (a) Continuous floating gate, and discrete trap floating gate, based on (b) electronic traps naturally existing in nitride, and (c) nanocrystal induced trap

In 2003, the first functional 6 V, 4 Mb silicon nanocrystal based nonvolatile memory arrays using conventional 90 nm process technologies were produced at Freescale's Austin Technology and Manufacturing Center [4]. In 2005, Freescale manufactured the world's first 24 Mb memory array based on silicon nanocrystals, a major step toward replacing conventional floating gate-based flash memories. The production of a working 24-Mbit memory device requires that silicon nanocrystals be deposited with excellent uniformity and integration approaches that keep the nanocrystal properties intact during subsequent processing. Meanwhile, as the industry begins manufacturing at smaller geometries with 90-nm and smaller, embedding floating gate-based flash becomes difficult to produce cost-effectively. At such dimensions, the chip need at high-voltage of 9-12 V and the transistors required to write and erase the conventional flash module cannot be scaled down. Furthermore, engineers cannot reduce the high working voltage in floating gate-based flash memory without compromising reliability or risking memory failures and loss of data. Nanocrystal memories are part of an advanced class of next generation of memory technologies. They show better scalability than conventional floating gate-based flash technology, as their tunnel oxide thickness can be reduced without seriously impacting data retention. The ultimate limit in scaling down the floating gate memory is to use one electron to represent a bit, the so-called single electron transistor (SET). To make such practical memory requires a proper design of the device structure and that the voltage for charging a single electron is discrete and well separated, as is the shift in threshold voltage caused by the storage of a single electron [5].

The data read/write operation can be programmed and erased using conventional techniques in floating gate memories and can substantially reduce the cost of embedded flash at the 90 nm node or beyond. It is easier and more reliable to integrate flash memory with logic and analog devices in order to achieve better chip performance and integration density. The non-volatility and high density of flash memory give it a wide window of opportunities from code storage to mass data storage. As a result, it accounts already for more than half of the DRAM market and is currently the fastest growing memory segment. For flash memory possessing the multi-bit/cell property, several distinct threshold voltage states can be achieved in the flash memory cell by changing the amount of charge stored in its floating gate which further improves the performance of chip density.

Although flash memory has become the mainstream of current non-volatile memory, it also exhibits a number of disadvantages: the operation speed of flash memory still is slow, compared to volatile memory. The fastest programming times are in the range of μs (also depend on what kind of tunnel process is used) and the erase times are in the *ms* range. Usually, regarding different architectures of flash memory technology (NOR and NAND flash technology), they have different operation speed. NOR flash memory is much like address-mapped memory with small capacity and is faster than NAND flash memory which can store huge amounts of data. Additionally, its working voltage and endurance also needs to be optimized. For the progress of the nanocrystal memory development, retention and narrow program window has remained a main issue for practical applications, because nanocrystals are usually formed by CVD or ion implantation with lower than 10^{12}cm^{-2} density. From the point of view of integration, the mismatching of the chips also is unavoidable, because uncontrollability of the exact number of Si NCs in each transistor cell could cause similar, but not identical threshold voltage shift for individual cells and degrade the performance of chips which contains millions of transistors [6].

1.2 Non-volatile memory scaling

The scaling demands very thin gate insulators in order to keep short channel effects under control and to maximize performance. But leakage current through these thin insulators has become a major concern for many applications. The minimum feature size of an individual MOSFET has shrunk to 15 nm with an equivalent gate oxide thickness (EOT) of 0.8 nm in 2001. However, semiconductor flash memory scaling is far behind CMOS logic device scaling and still requires a working bias of more than 10 V, which is far behind the operation voltage of CMOS logic. According to the prediction from the International Technology Roadmap for Semiconductor (ITRS) in Fig. 1.4, silicon MOSFETs are already in the nano-scale. Hence new physics, such as quantum confinement needs to be considered, in order to further improve the properties of silicon devices.

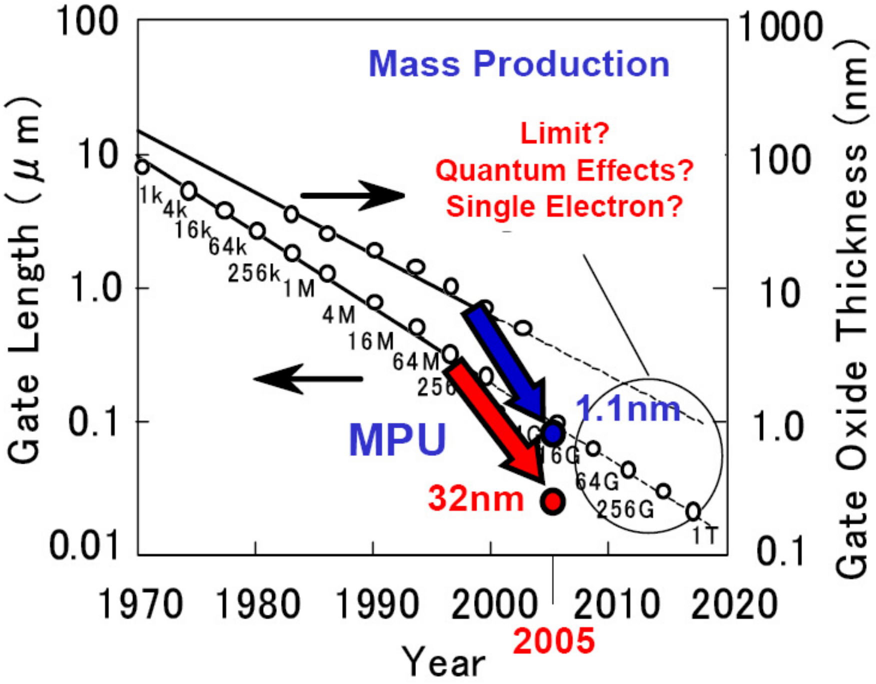


Figure 1.4: The trend of MOSFET scaling from ITRS. Data from ITRS Corp (www.itrs.com).

Floating gate device became the prevailing non-volatile memory devices when individual storage nodes devices based on storage in traps in Si_3N_4 or in semiconductor or metal nano-dots attracting more attention as shrinking devices. The injection and ejection of careers is accomplished by tunnelling across a gate insulator barrier, and the state of the memory is read by sensing the current between source and drain at a gate voltage between the two threshold current values.

The scaling of the gate stack and operation voltages are often dependent on each other. A tunnel oxide thickness larger than 8 nm is required in the commercial flash memory chip to satisfy the ten years data retention time. If the tunnel oxide is to be decreased below 2 nm, the operation voltage could be reduced from more than 10 V to less than 4 V. If the equivalent gate oxide thickness is below 2 nm, carriers can directly tunnel through the insulator barrier which shortens the retention period rapidly from 10 years to several seconds. However, a high reliability is associated with the discrete trap sites in floating gate. In fact, a single leakage path can only discharge a single storage node. In contrast, in a conventional continuous floating gate, a single defect could destroy completely the behavior of the transistor. In the discrete natural traps in silicon nitride, the electron wave is spread out and the charge associated with a discrete trap could harm the performance of whole floating gate. If individual dots are sufficiently isolated from each other and identical, the electrons in the floating gate are localized and stabilized. The concept of a NC based memory in Fig. 1.5 demonstrates the scheme of a Si NC-based transistor.

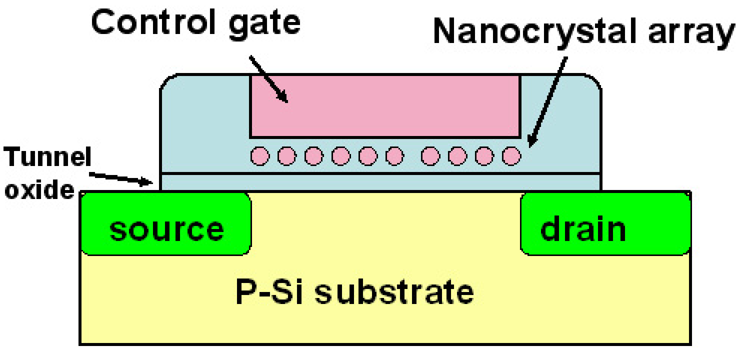


Figure 1.5: Scheme of Si nanocrystal transistor

1.3 Fabrication of silicon quantum dots

Dimensionality plays an important role in determining the properties of nano-structured matter. Quantum confinement is a result of the nanocrystal being smaller than the bulk semiconductor Bohr exciton diameter. The term quantum confinement, when applied to low-dimensional semiconductors, describes the confinement of the exciton within the physical boundaries of the semiconductor. This is a quantum phenomenon - hence the names, ”quantum well”, ”quantum wire”, and ”quantum dot”, which describe confinement in 1, 2 and 3 dimensions, respectively, as shown in Fig. 1.6. The exciton Bohr radius is often used as a yard-stick to judge the extent of confinement in a low-dimensional structure. Meanwhile, as the device size shrinks for higher speed and lower power, certain properties cannot be scaled due to the wave nature of electrons. In nanostructures, whenever the electron mean free path exceeds the appropriate dimensions of the device structure, quantum behavior may dictate the physical properties of devices.

The capability of charge storage is an important property of Si nano-dots, which act as trap sites influenced by quantum confinement effects [7]. This storage ability

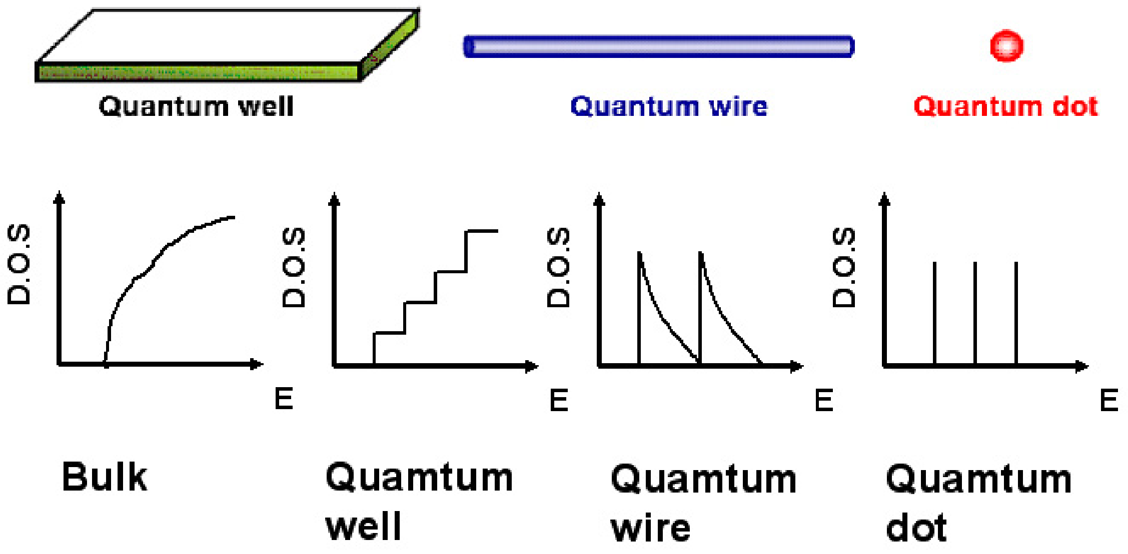


Figure 1.6: Nanostructures with quantum confinement in one or more directions. Geometry and density of states (DOS) as a function of electron energy.

also is size-dependent for nano-scale Si dots [8].

1.3.1 Important quantum features for nanoscale Si dots

Bulk silicon has an intrinsic indirect band gap property as showing in Fig. 1.7 which was described in the book of Yu and Cardona [9]. Its valence band maximum and conduction band minimum are not located at same momentum wavector k. Carrier band gap transition have to be assisted by phonon absorption or emission [10]. For nano-scale silicon dots, firstly, the spatial localization of the electron and hole wavefunctions within a dot increases their spread in momentum space, hence enhance their overlap probability in the Brillouin zone. As a result, the radiative recombination rate can increase by orders of magnitude. Secondly, nano-dots contain a very small number of atoms. Therefore, imperfections, such as point defects and dislocations which cause nonradiative recombination are unlikely to exist. Because nonradiative recombination pathways are suppressed and the radiative recombination rate is enhanced, carrier confinement in nanocrystals causes the efficiency of luminescent recombination to dramatically improve as well. For silicon nanocrystals, excitonic luminescence can easily be observed at room temperature in the near-IR (approximately 700 to 800 nm) that is hundreds of times brighter than the approximately 1100-nm bulk intrinsic luminescence at cryogenic temperatures.

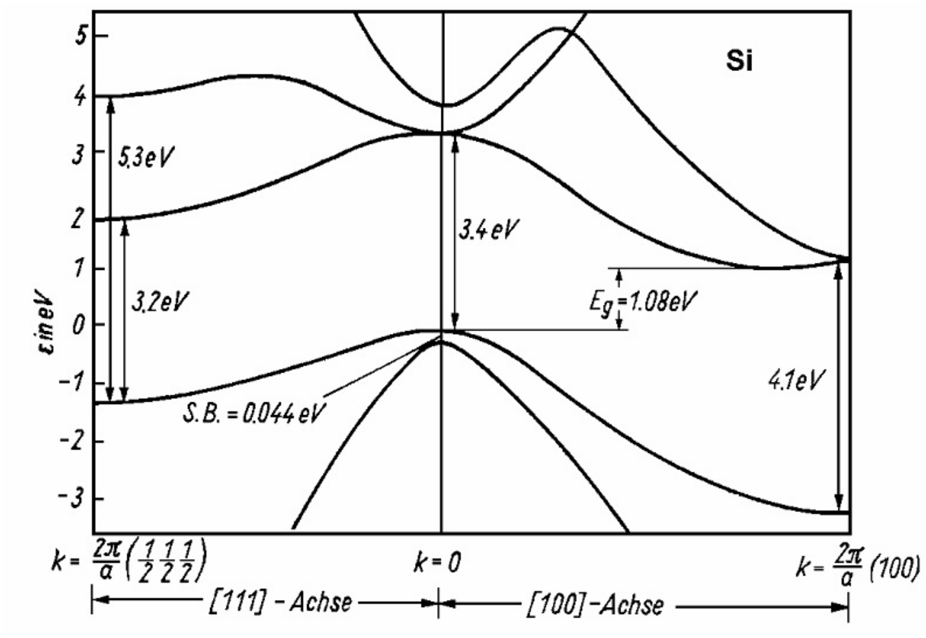


Figure 1.7: Simple energy diagram and most frequent transitions. Band gap E_g is about 1.1 eV. Figure from Yu and Cardona reference.

Except light emission, quantum confinement effects also lead to a change of capacitance of a nanoscale sphere and a reduction of dielectric constant. We can calculated the quantum-confinement capacitance of a small sphere by the following equation:

$$E_2 - E_1 = e^2/2C_{eff} \tag{1}$$

where E_1, E_2 are the one, two-electron ground state energies of a silicon sphere embedded in an amorphous SiO_2 matrix [11]. For simplicity, in order to bring in a second electron to the silicon particle, the required bias is determined by the ground states, and the Coulomb energy involving the capacitance (which is owning the so-called Coulomb blockade). Due to quantum confinement, an electron has an energy in addition to the electrostatic energy and this energy is inversely proportional to the square of the dimension of nano-dots. The quantum confinement capacitance C_{eff} decreases when the ground-state energies dominate over the electrostatic energy, and the appreciable reduction can be observed only for dot sizes under 10 nm. Table 1 summarizes a number of characteristics for silicon dots in the gate stack. The capacitance (C_{eff}) is the self-capacitance of the silicon dot [8]. These vary linear for capacitance and inverse linearly for charging energy E_c. The quantization effect of confinement of energy E_0 varies as the inverse square of the dimension.

Table 1: Parameters of a spherical Si dot charge. Values from the review of S. Tiwari [8].

Size/nm	C_{eff}/aF	E_c/eV	E_0 /eV	Shift V_T if single electron transistor/V
20	4.45	0.018	0.007	0.062
10	2.23	0.036	0.03	0.225
5	1.11	0.072	0.104	0.84
3	0.68	0.118	0.29	2.31
2	0.45	0.178	0.65	>5
1	0.22	0.364	2.6	>10

The reduction of the static dielectric constant also becomes significant as the size is reduced down to quantum-confined systems, such as nanocrystals and nanowires. A reduced static permittivity increases the Coulomb interaction energy between electrons, holes and impurities, and thus can significantly modify the optical and electrical properties of these quantum-confined structures [11], such as their refractive index or capacitance.

1.3.2 State of the art of Si NC preparation

Silicon NCs embedded in an insulating matrix has attracted a lot of scientific attention. A number of techniques are used to produce Si NCs, such as ion implantation of Si in SiO_2 films, Pulsed Laser Deposition (PLD), Plasma Enhanced Chemical Vapor Deposition (PECVD), Low Pressure Chemical Vapor Deposition (LPCVD), thermal evaporation of silicon rich oxide, and so on. Usually, a high temperature annealing process is needed to induce the formation of the NCs. Techniques based on CVD methods have the advantage of being more stable as well as typical in industrial applications, but inevitably bring in hydrogen from the silane source which deteriorates the properties of Si NCs. Obviously, ion implantation can not guarantee the uniformity of Si particles, because of the depth distribution of the implanted ions. PLD or magnetron sputtering are suffering by the problem of size and position control and their incompatibility with traditional microelectronics technology.

The SiO/SiO$_2$ superlattice method was developed at MPI-Halle for Si NC growth which overcomes intrinsic problems of other methods, and further optimizes some important parameters for Si NCs, such as size and size distribution, as well as the location and density of the dots. In this dissertation, I will show how these properties will influence the performance of devices based on Si NCs.

1.4 Optical properties of Si NC

First experimental results more than a decade ago demonstrated room temperature luminescence of nanocrystalline silicon and showed its potential application for silicon based photonics [12]. In addition to room-temperature photoluminescence, electroluminescence (EL) measurements involving various types of silicon NCs showed efficiencies higher than bulk Si. This could make electrically pumped light emission from silicon NC attractive for future photonics applications. Because of quantum confinement and the elimination of bulk or surface defects, these dots can emit light from the near infrared throughout the visible with quantum efficiencies in excess of 0.6% [13]. Meanwhile, for optoelectronic applications, the key device is a suitable light source: the laser. Pavesi *et al.* demonstrated that light amplification is possible using silicon itself in the form of quantum dots dispersed in a silicon dioxide matrix [14]. This invention could be considered as a milestone in silicon photonics research and would open a route to design a siliocn laser although the results are still controversial. A number of papers reported the observation of optical gain in these systems by following this idea and gain has been reported in a number of different experiments in Si NC formed by different techniques [15]. However, since the origin of optical gain is still unclear, these results are under debate [16].

1.5 Further device application of Si NC

1.5.1 Photonics based on Si NC

Various photonics devices using Si NC have been investigated: STMicroelectronics announced a breakthrough based on silicon nanocrystal technology involving light-emitting diodes (LED). Zhang demonstrated whispering-gallery modes in microdisk arrays based on size-controlled Si NCs [17]. If the nanocrystal structures are implanted with erbium ions, very effective energy transfer occurs from the nanocrystals to the Er^{3+} ions, and the luminescence shifts into the technologically useful wavelength range of 1.54 μm [18].

1.5.2 Silicon optical nanocrystal memory

An optical analogue to the electronic nanocrystal memory was demonstrated recently by the group of Atwater at Caltech [19].

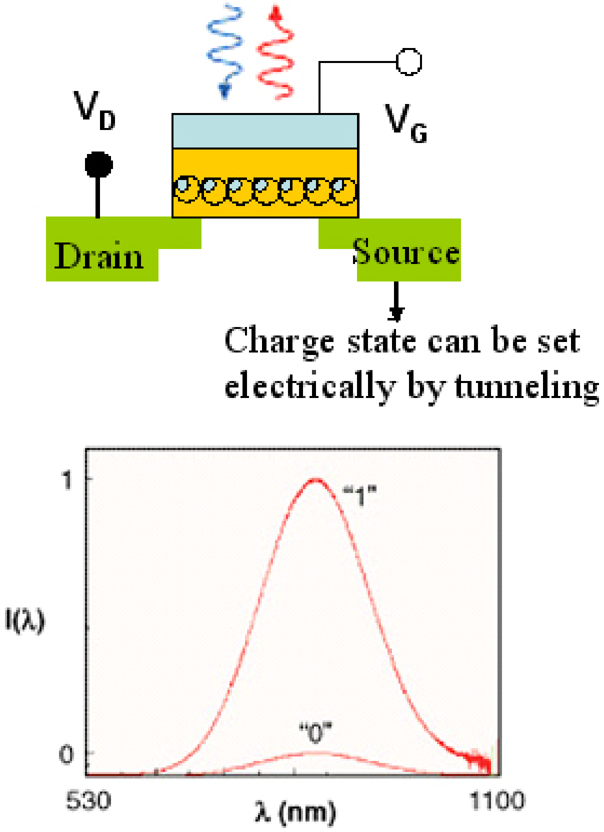


Figure 1.8: A two-state memory which can be set electronically and read optically through a transparent top gate. Figure from R. Walter [19].

When the optical nanocrystal is uncharged, incident light excites hole-electron pair as photoluminescence, but the photoluminescence process is quenched in a charged nanocrystal, ascribed to fast nonradiative Auger recombination processes in which relaxation of optically generated excitons occurs by energy transfer to a nearby excess charge carrier. Thus results in two states of photoluminescence which can be used for a memory system as showing in Fig. 1.8. The hope is to potentially replace electrical data buffers in optical communication systems and allow for the elimination of the accompanying optical-to-electrical conversion hardware for future all-optical communications systems. However, the relatively long emission lifetime of silicon NCs which may critically limit the program and erase speeds. A millisecond lifetime would make the device speed at about a MHz level, which is much slow than realized for concentrated memory devices [20].

1.6 Organization of the thesis

In this thesis, I investigate Si nanocrystal fabrication and their application for memory devices, based on artificial discrete-trap storage node, as one alternative for *flash* memory scaling. This dissertation addresses the aforementioned issue of gate stack scaling for future generations of semiconductor flash memory and proposes solutions based on new memory structures and new materials that are compatible with the current CMOS process technology. In Chapter 1, one new candidate for future non-volatile memory devices, based on Si NCs, is introduced, and discussed in comparison to previous concepts. Diverse methods of fabricating Si NCs and the properties of Si NCs will be summarized. Chapter 2 describes the fabrication of silicon NC by phase separation and this method is highlighted by its regular NC size control and well-layered location. A set of Si NCs embedded in SiO_2 samples are fabricated and investigated by Transmission Electron Microscopy (TEM). These samples are used for various purposes in the following chapters. In Chapter 3, the typical MOS test architecture was applied to various samples. The conduction mechanisms for insulators, the principle of operation for MOSFETs working with Si NCs, and the MOS high frequency capacitance-voltage behavior are introduced. In order to understand its conductive mechanism and capacitive behavior, the electrical properties of MOS structure containing a single individual silicon NC trap layer are investigated by static I-V, high frequency C-V, G-V and the conductance method. For reliability, IR Thermography is employed for detection of breakdown below the metallization layer after many times tunneling operation. In Chapter 4, a new concept of the multibit/cell is introduced. The relevant MOS prototype with multi Si NC layers is characterized by a family of capacitance-voltage curves. Charging states due to different number of Si NC charging layers show the feasibility of multibit/cell design. A theoretical charging model and further retention investigations strengthen the idea for multi-bit storage in one cell operation. Chapter 5 demonstrates the tuning of the Si NC density of our samples by changing the stoichiometry of the SiO_x in alternative SiO_x/SiO_2 deposition process. TEM, photoluminescence and C-V curves show the viability of this method. This adjustable parameter might optimize the performance of non-volatile memory devices based on the discrete nodes. Since the ability of the Si NCs to trap carriers, Deep Level Transient Spectrscopy (DLTS) allows to study the trapping behavior of Si NCs. The parameters of the trap center are determined, Chapter 6 shows this measurement and related results. A summary of the work and future perspectives for semiconductor non-volatile memory devices are given in Chapter 7.

CHAPTER II

Functional Nanocrystals and Their Fabrication

2.1 Devices based on silicon nanocrystals

Nowadays, nonvolatile memories are mainly used in portable electronic devices. Portable electronics devices, such as mobile phones and mp3 players, have significant economic impact on the semiconductor industry. Most commercially available nonvolatile memories are produced with a polycrystalline silicon or Si_3N_4 floating gate as the charge storage layer. However, the polycrystalline silicon can not be used with very thin tunnel oxide, which is a scaling problem of these devices, and the silicon-oxide-nitride-oxide-silicon (SONOS) memory is limited by its data endurance due to shallow trap sites. From the previous chapter we already know that a single leakage path in a polycrystalline Si based memory can discharge the memory with a total loss of the information as a consequence. Because of NCs as discrete-trap storage nodes in floating gate devices has been suggested as one of the key items for increasing flash memory stability and decreasing the node size for higher information density and further scaling down of chips. In such NC based memories the Si NCs should be sufficiently isolated from each other by an oxide. Hence, the electrons are localized in the NCs and a single defect would not discharge all of the NCs, which means such a NC-based memory should have a longer retention time than its SONOS counterparts.

In contrast to other metal or III-V, II-VI semiconductor NCs [21], one main advantage of silicon NCs is their compatibility with standard silicon CMOS technology. Therefore, silicon NCs embedded in dielectric layers like silicon dioxide, or nitride have attracted an increasing attention over the past decade. However, NC memories have a rather low charge density which will cause a relatively narrow threshold voltage shift and a high power consumption which might be a drawback of this approach. In addition to the above discussed non-volatile memories based on Si NCs, other applications of Si NCs have been proposed, too. Such applications are based on the tunnel effect and Coulomb blockade in multiple tunnel junctions (MTJ) [22] and single electron transistor (SET) [5], respectively. These devices require to subtly tune and keep NC identical. Recently, researchers from Cavendish Laboratory at the University of Cambridge realized isolated double Si quantum dots as charge qubit which might offer a new way for quantum computation [23].

2.1.1 Introduction of SiO/SiO_2 superlattice

Presently, Si NC research is focused on the preparation of Si NCs embedded in an insulator host, and uses the simple surface passivation of the Si nano-dots by oxidation or hydrogenation. Si NCs as small as a few nms in diameter have been synthesized by a variety of techniques. Synthesis of silicon NCs can be realized by electrochemically etching of single crystalline silicon in hydrofluoric acid resulting in so-called porous silicon [24], by ion implantation of silicon into a high quality oxide matrix followed by thermally induced Ostwald-ripening of silicon clusters and their crystallization [25], or by deposition of sub-stoichiometric oxide films from CVD methods [26–28], magnetron sputtering [29] or reactive evaporation [30] and a thermally induced phase separation and crystallization. In all these methods the size control is realized by changing the chemical stoichiometry of the films and by the control of the annealing process, such as annealing time and steps. However, with these methods the NC size can only be decreased by simultaneously decreasing the NC density. Other parameters, such as the position of NCs, in-plane separation distances of neighboring Si NCs are hard to control. For device applications, accurate engineering of spatial position, size and density of the silicon quantum dots including the realization of single layers or layer control is essential.

The SiO/SiO_2 superlattice process that allows the preparation of size-controlled silicon NCs was developed [31, 32] at MPI-Halle in 2001. Reactive evaporation of SiO powder in oxygen atmosphere is used mainly for preparation of bulk amorphous SiO_x films with variation of x in the range between 1 and 2 and frequently used in coating of optical materials. Therefore, a lot of experimental data in thermal evaporation, such as evaporation temperature and pressure dependence of SiO_x, are available for the SiO_x deposition. The alternate depositions of SiO_x/SiO_2 ($x<2$) pairs realize that the SiO_x layer is sandwiched by adjacent SiO_2 layers. Amorphous SiO is not a stable chemical compound and will easily decompose into Si and SiO_2 during high temperature annealing. The high T annealing (between 900-1100 ℃) of such initially amorphous SiO_x films results in phase separation described by

$$2SiO_x \rightarrow xSiO_2 + (2-x)Si \tag{2}$$

and causes Si clusters which are surrounded by a SiO_2 matrix. The high T phase separation of the SiO_x automatically leads to a Si-O bond breaking. The Si atoms assemble to Si clusters, further nucleate to Si NCs and will be separated by a SiO_2 shell. The surrounded SiO_2 merges with the upper and lower SiO_2 layers, but the Si nuclei are confined in the previous SiO layer with a certain size close to that of the SiO_x layer thickness. Based on the above equation the distance of the oxide between the adjacent Si nanoclusters depends on the stoichiometry of the SiO_x as well. The NCs have an average size depending on the original SiO_x layer thickness and are randomly arranged within the plane of the former layer. Usually we used a constant stoichiometry of x around 1.0 for preparation of the thin SiO layers under base vacuum conditions. The NC sizes are controlled independently by using a layer thickness approximately equal to the desired crystal sizes. This method produces well-passivated Si NCs and can control not only the size but also the location of the NCs as will be demonstrated in following chapters.

2.2 Sample fabrications

The samples were prepared by evaporation of SiO powder using a conventional evaporation chamber which enables the preparation of SiO layers of uniform thickness on 4 inch wafers by two symmetrically arranged resistance-heated Mo boats. The evaporation temperature of the boats is about 1200-1600 ℃, and the silicon substrate temperature is 100 ℃. The material used for evaporation was commercial SiO powder (Chempur, 99.9%) with very tiny grain size about 0.5 mm which was either evaporated under vacuum 1×10^{-6} mbar or oxygen gas atmosphere 2×10^{-4} mbar for the SiO_x or SiO_2 deposition, respectively. Usually SiO can be evaporated from boats of Mo, W or Ta because of their heating efficiencies and matching resistances. In addition, their high melting points and low vapor pressures have to be considered, too. For instance, the vapor pressure of W is only 1 Pa at 3477 K, and much less at 1500-1900 K. Thermal evaporation could bring a heavy metal contamination, so the heating current is always kept at a moderate level to keep the amount of heavy metal under the tolerance limit.

Table 2: Growth parameters and experimental purpose

sample	type	resistivity (Ω.cm)	configuration (SiO_x/SiO_2+SiO_2 nm)	annealing N_2 flow	purpose
A1	p	1-30	1×4/4+4	1100°C(1h)	Charge and tunnel
A2	p	1-30	1×4/4+4	1100°C(1h)	Control sample
B1	n	1-5	1×3/4+26=33	all have	Multilevel
B2	n	1-5	2×3/4+19=33	the same condition:	charge
B3	n	1-5	3×3/4+12=33	1100°C	
P	p	1-5	3×4/4+12=36	0.5h	permittivity estimation
C1	n	1-5	all identical	same conditions	Tune
C2	n	1-5	structures as follows, but	as follows:	Si NC
C3	n	1-5	variable O_2 pressure:	900°C(0.5h)+	density
C4	n	1-5	3×4/4+12=36	1100°C(0.5h)	
D1	n	0.05-0.1	1×4/4+24	1100 °C(0.5h)	DLTS

Rotation of the wafer holder enables a thickness homogeneity better than 10% over the whole wafer. By alternatively depositing SiO_x and SiO_2 layers a superlattice structure can be easily prepared. In general, the requirements for cleaning elemental group-IV substrates are more stringent than those for III-V compound semiconductor substrates. Improper surface preparation prior to growth could cause reoxidation or contamination, thus degrade the insulating. Conventional Si substrate cleaning is using a combination of standard RCA (Radio Corporation of America) and HF dip. RCA consists of 2 steps which are a 10 min treatment in a 70-80 ^{o}C heated solution of 1 H_2O_2: 1 NH_4OH : 5 H_2O (volume ratio) plus 1 min immersion in 1H_2O_2: 1HCI: 5H_2O (volume ratio) solution. After 1 min 5% HF dip, rinse in deionized water and dry, the wafer is immediately loaded into the chamber for deposition.

A series of samples were grown for various purposes of the experiments in the following chapters by choosing different kinds of wafers, based on doping type, resistivity and orientation. In order to summarize, all parameters of the grown samples are listed in table 2. Sample A1 and A2 were evaporated on p-type, 1-30 Ω.cm, (100) Si substrates. The single layer of silicon NCs in sample A1 was realized by preparing a 4 nm thick SiO_x layer embedded in 4 nm thick SiO_2 layers on both sides. The difference to experiments reported before is the nearly equal and thin thickness of the bottom and the top oxide layer [33], because I wanted to observe the tunneling behavior through the whole oxide layer and to determine the conduction mechanism. For comparison, a control sample, i.e, a SiO_2 film sample A2 with the same total thickness as sample A1, was prepared on the same type of substrate using the same deposition conditions. Subsequently thermal annealing for 1 h was performed in a quartz tube furnace under N_2 flow environment at 1100 ^{o}C to realize the Si NC formation. They are used for the investigation of the conduction mechanism as well as the charge storage behavior in chapter 3.

Similarly, three samples B1, B2, B3 were prepared on n-type (100) silicon substrates (1-5 Ωcm) after cleaning. First, a 3 nm SiO_2 film was deposited as a tunnel oxide in each case, then 1, 2 or 3 layers of SiO were prepared and separated by a thin layer of SiO_2 by alternative thermal deposition of 4 nm of SiO and 3 nm of SiO_2. On top of this structure, an additional SiO_2 capping layer was evaporated as the upper control oxide with a respective layer thickness of 26, 19 or 12 nm for the sample B1, B2, B3, respectively. Thus the total thicknesses of these 3 samples are identical. In order to form the Si NCs, thermal annealing was performed in a quartz tube furnace under N_2 ambient (1100 ℃, 0.5 h) for phase separation and crystallization. These samples are used as the prototypes of multi-level charge devices and will be discussed in chapter 4.

Sample C1-C4 are a series of samples which are based on n-type, 1-5 Ωcm, (100) silicon substrates, but are prepared having a different stoichiometry during the deposition. The sample structures contain three pairs of SiO_x/SiO_2 stacks where the stoichiometry of the SiO_x layers was fixed for each of the samples to a different value. First a 4 nm SiO_2 film was deposited as the tunnel oxide, then the three layers of the SiO_x (4 nm) alternated by thin SiO_2 layers (3 nm) were prepared by alternative thermal evaporation of SiO powder under selected oxygen pressure. An additional SiO_2 layer was evaporated on top as the upper control oxide. For the SiO_2 layers an oxygen pressure of $2x10^{-4}$ mbar was always used. For the above samples, SiO_x layers were prepared with less than $1x10^{-6}$ mbar (sample C1), which is the base pressure of our vacuum system, $5x10^{-6}$ mbar (sample C2), $1x10^{-5}$ mbar (sample C3), and $5x10^{-5}$ mbar (sample C4). Sample C1 represents the control sample with pure SiO deposition and properly represents the highest NC layer density possible by using SiO powder. A summary of these used parameters and results can be found in the table 4 in the corresponding chapter. Rutherford Back Scattering (RBS) measurements were reported for estimation of the oxygen content, i.e. stoichiometry of SiO_x. The measurements were performed on bulk SiO_x thin films. These sample will be used to demonstrate the tuning of Si NCs density in chapter 5. Last, the D1 sample based on a highly doped substrate was specially designed to have a single layer Si NC with asymmetric tunnel oxide and control oxide for the observation of deep level transient spectroscopy (DLTS) in chapter 6.

2.3 Physical characterization of the samples

2.3.1 RBS measurement of stoichiometry in SiO_x and TEM introduction

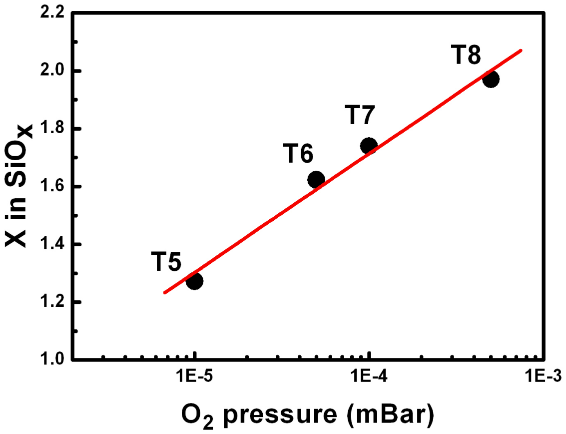


Figure 2.1: Stoichiometry of SiO_x as function of oxygen pressure, measurement based on the SiO_x thin film grown in the same chamber.

RBS is an effective composition analysis tool for thin films. The sample is bombarded with very high energy helium ions, the yield and energy of backscattered helium are measured. It is an energy spectrometry of MeV He^+ ions, elastically scattered by nuclei of the analytical sample, both composition and depth distribution of elements in the sample can be deduced non-destructively from its spectrum. This makes RBS an easy tool to determine stoichiometric abundances of elements in materials. Before growing the SiO_x/SiO_2 sample, we measured the RBS on bulk SiO_x films (200 nm) as T5 to T8 showing in Fig. 2.1. The obtained RBS spectra confirms the dependence of the stoichiometry of the SiO_x layer on the growing oxygen pressure. Under the allowed experimental oxygen pressures, the x in SiO_x can be changed approximately between 1 and 2.

TEM is a valuable characterization tool for structure analysis. A variety of signal generated by the electron beam passing through the specimen can be used to extract information on the structure of the specimen. TEM works sometimes similar to optical microscopy, but with an electron beam instead of light. If the specimen is thin enough, the interference between directly passing electron waves and diffracted electron waves can be used to obtain images.

2.3.2 Structural characterization of samples A1

Cross section TEM was used to characterize the sample structure using a Philips CM20T operating at 200 kV. Because of the low contrast between Si nano-dots and surrounding SiO_2, TEM investigations are normally taken at under-focus. The typical sample A1 structure is schematically shown in Fig. 2.2 (a) and the cross section image is represented in Fig. 2.2(b). The cross section specimens were prepared in the usual way glueing two wafer face to face, sawing, grinding, dimpling, etching and final Ar ion milling to reduce the thickness of specimens until few tens nm. Ar is used because it is inert and not naturally present in the samples.

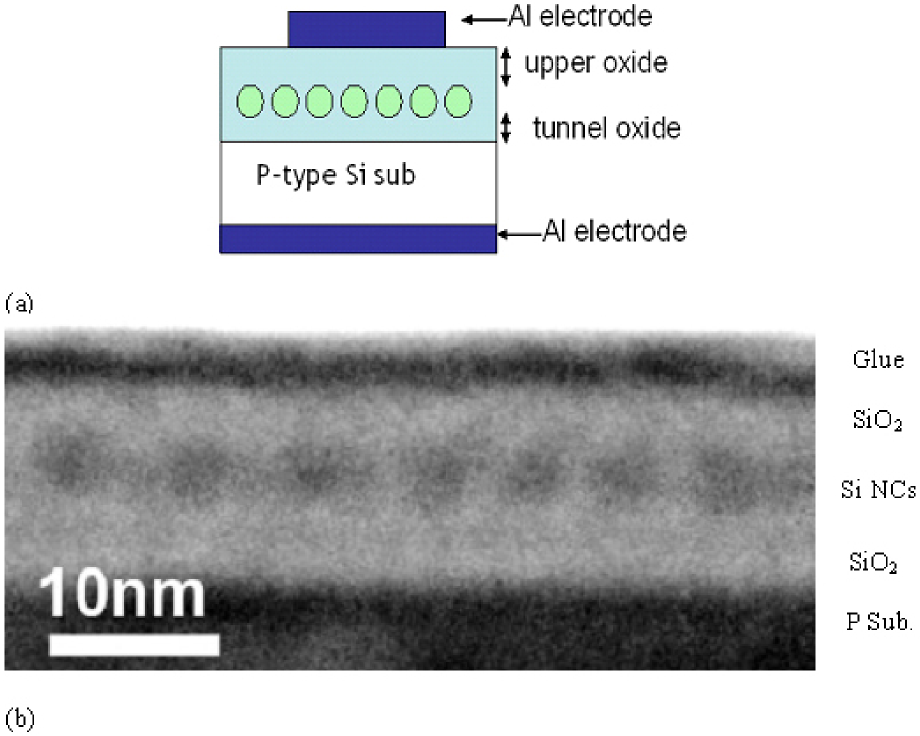


Figure 2.2: (a) A schematic model of the sample structure. (b) TEM cross section image of the sample A1.

In the bright field Fig. 2.3, many tiny spots are observed and have random distribution in plane. These tiny spots are the Si dots and have consistent size, but random localize. As seen in the plan-view TEM image Fig. 2.3 of sample A1, the silicon NCs

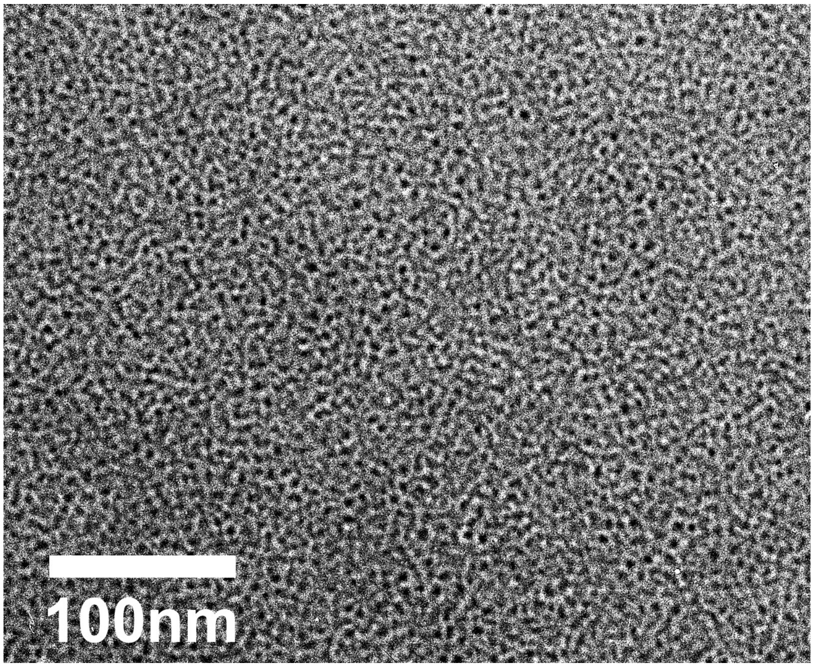


Figure 2.3: In plane TEM image of sample A1. This image was taken by a Philips CM20T at 200kV

have a diameter of approximately 4 nm. The silicon NCs on average have a distance of about 3-4 nm, hence the estimated area density of silicon NCs is approximated to be around $10^{12} cm^{-2}$. Electron diffraction pattern provides important information on the atomic or molecular crystal lattice of the specimen, such as spacing between lattice plane reflections. The measured spacing can be compared with the uniform value from XRD database, thus the lattice plane, morphology can be determined. Atomic-resolution plane-view TEM images were investigated using a JEOL JEM4010 with 400 KV. Because the electron beam is vertically transmitted through the sample, the texture of (111) lattice plane is easier to visualize for diamond cubic crystal of Si. Several Si NCs are marked in the Fig. 2.4. We roughly estimate the interplanar spacing (d spacing) to be about 3.14 Å. The observable Si crystal (111) planes occur. Zooming into this in-plane image, another texture of the crystal plane (220) will appear with interplanar spacing of 1.9 Å. Only NCs having the right orientation to the incident electron beam can be seen. These interplanar spacing values are identical to the bulk values from literature and demonstrate that dots are unstrained by the surrounding SiO_2. This evidence could help us to understand the mechanic property of the Si dots in the matrix. Strain could influence the carrier mobility, hence, the message that the Si NCs obviously are not under strain is important for the property

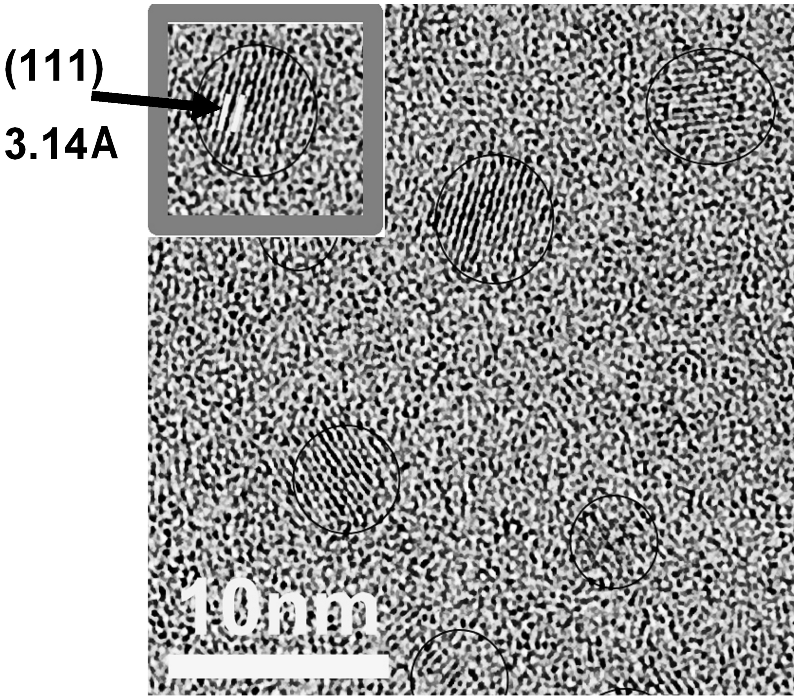


Figure 2.4: Plane view HRTEM image of sample A1. This high resolution image was taken at JEOL JEM4010.

of Si NC MOSFET.

2.3.3 Structural characterization of sample B1-B3

The layer structure of the annealed samples B1, B2, B3 was checked by TEM in cross section by Philips CM20T, too. Fig. 2.5 shows the cross section TEM images of the 3, 2 and 1 layers samples, respectively. Images are taken under bright field and underfocus. The silicon NCs are represented by the dark dots in the images, having a diameter of 3.9 nm (±0.4 nm), and being separated in plane by 2-3 nm of SiO_2. The distance between adjacent Si NC layers is about 3 nm. All 3 samples have nearly the same total thickness. The estimated total area density of silicon NCs in sample B3 is about 9 x $10^{12}/cm^2$. From former experiments, it was known that for thicker layers, i.e. larger crystals, the size distribution slightly increased. The resulting NCs are nearly spherical in shape, thus have roughly identical charging capacity and a good isolation from each other. Due to the presence of several layers, the samples offer the possibility to charge step by step the respective Si NC layers. The SiO_2 layer between the Si NC layers offers a constant barrier for carrier trapping. This is in contrast to other Si NC fabrication methods. In-plane TEM investigations were not performed for the multilayer structures, because different layers may contribute to the plan-view image and make analysis of such images difficult.

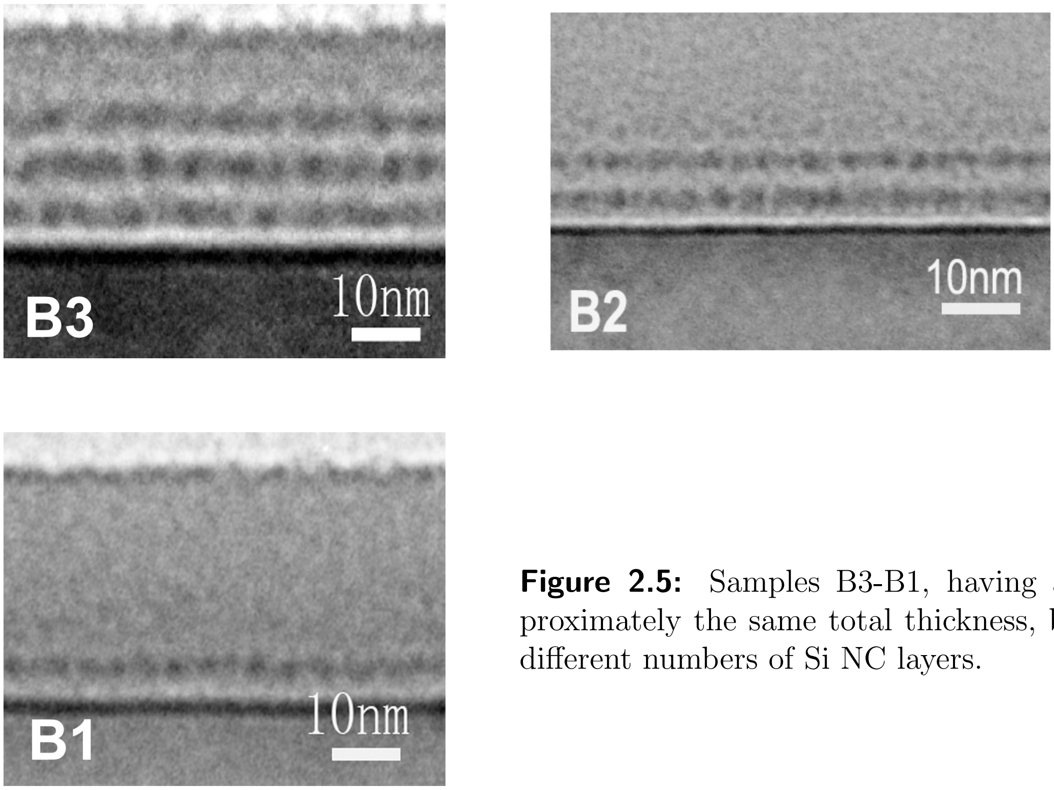


Figure 2.5: Samples B3-B1, having approximately the same total thickness, but different numbers of Si NC layers.

2.3.4 The composition and structural characterization of sample C1-C4

The selected cross section TEM images of the samples C1-C4 are shown in the Fig. 2.6. Again, the resulting NCs are spherical in shape and have a good isolation from each other. In addition, the trend of decreasing number of Si NCs in the oxide shows up from (a) to (d). Please note, that the samples prepared in cross section have still a thickness of around 10 nm which means that more than one row of crystals contribute to the TEM image in depth. The samples have silicon NCs with a diameter of 3.8 ± 0.5 nm, as was estimated from the TEM images. The diameter is nearly identical in all samples and controlled by the former SiO layer thickness. The very small deviation in the total thickness of the three layers in the samples demonstrates the process stability of our growth and the good reproducibility.

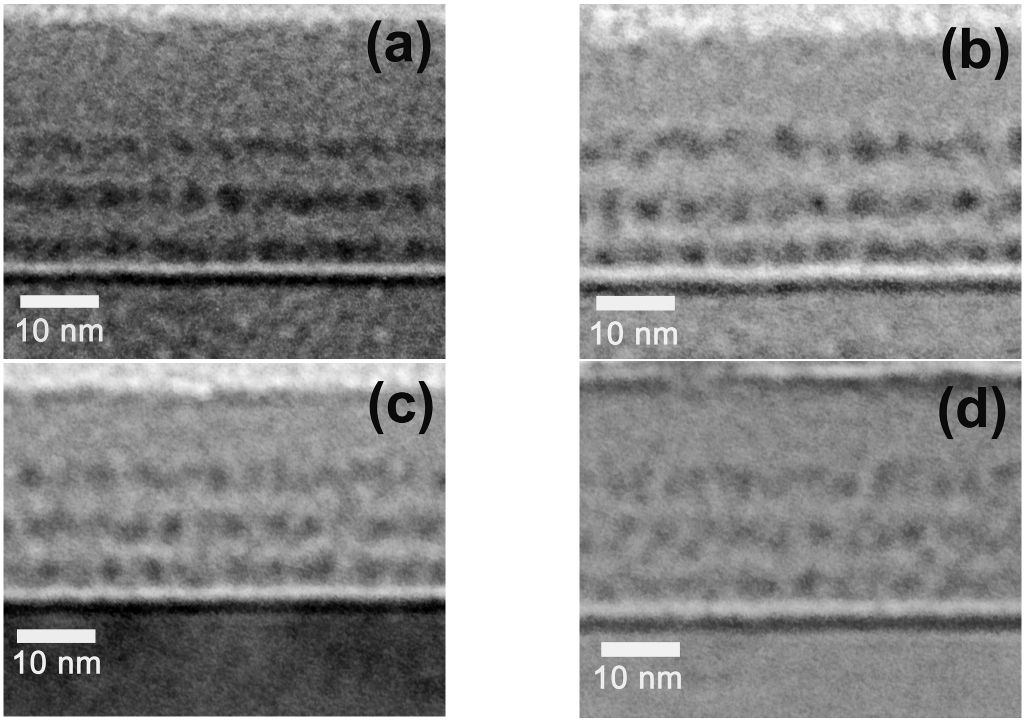


Figure 2.6: Cross-section TEM images of the selected samples, (a)-(d) correspond to samples C1-C4.

2.4 Estimation of Si NC density in stacked sample by model

Overall, the TEM pictures prove the existence of Si NCs in the oxide matrix. Unlike the formation of Si NCs by other methods, such as ion implantation, our method realizes Si NCs homogenously located within the previous SiO layer with a rather high in-plane density. In the following, an estimation of the maximum layer density of the NC is performed. If we assume that Si NCs have identical size d and each one is regularly in one quadratic cell, neighboring dots have fixed distance a, adjacent Si NCs layers are separated by h as demonstrated in Fig. 2.7, then the area density can be estimated by $(1/(a+d))^2$, in case of sample A1, a=3-4 nm and d=4 nm, then the area density is about $1.56 - 2.04 \times 10^{12}/cm^2$. The distance between two adjacent layer is typically chose to h=4 nm, the volume density $(1/(a+d))^2/(d+\mathrm{h})$ can evaluate further to be about $1.95 - 2.55 \times 10^{18}/cm^3$. These values are confirmed by the estimation from TEM pictures. I will demonstrate its consistency in the next chapter with the electrical measurements.

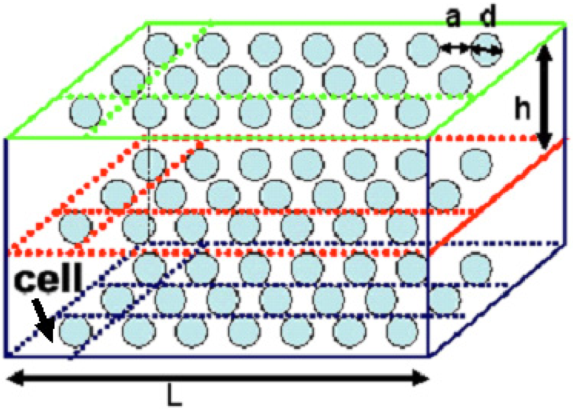


Figure 2.7: Regularly distributed Si NCs in sample structure. Each color layer represents the respective Si NC layer, assuming each cell only contains one Si dot and dots have same separation distance

This estimation gives the density range in my sample structure by ideally positioning each dot in a quadratic lattice. Compare with other methods [34, 35], the density of my sample reaches the $10^{12}/cm^2$ level without degradation of the separation of the Si NCs. The width of threshold voltage shift representing the memory states is proportional to the amount of the contributed Si NCs, so we have to increase the density of Si NCs in the matrix, but keep the balance for the insulating matrix to improve the reliability of floating gate memory device.

2.5 Conclusions of the chapter

A size controlled SiO/SiO_2 approach was used for the fabrication of samples with single layers of silicon NCs in oxide, samples containing 1, 2, 3 layers of silicon NCs with same total thickness, respectively. Samples with gradually adjusted Si NC density in their structures were also prepared. Structural characterizations were performed and demonstrated the suitability of this approach for the preparation of a few silicon NCs layers in thin oxides. TEM, as a main structural investigation tool, was successfully applied for all samples. Based on the TEM pictures, I demonstrated the availability of this SiO/SiO_2 approach for fabrication of the thin layer MOS device.

CHAPTER III

Electrical Properties of MOS Containing a Single Layer of Silicon NCs

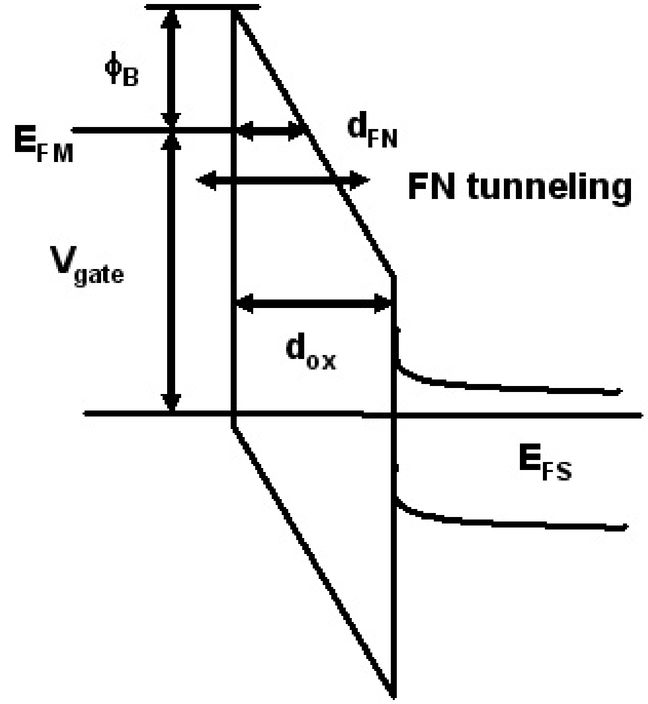
3.1 Metal-insulator-semiconductor structures

Our insulating layer containing Si NCs represents the floating gate in the MOSFET device. Metal-Oxide-Silicon (MOS) is the basic structure for FET transistors in silicon, and could be used to control the conductive channel by the gate bias. Therefore, after fabrication the Si NCs in SiO_2, we have to complete our sample to fabricate a MOS structure which is essential for further electrical characterization. For electrical characterization an aluminum top electrode (area of 0.275 mm^2) and a backside ohmic contact electrode were evaporated after removing the native SiO_2 on the wafer backside. Our sample thus represents a typical Metal-Oxide-Semiconductor (MOS) structure.

3.1.1 Conduction mechanism of insulator

The current mechanisms through the insulating materials, which do not contain free carriers, can be distinctly different from those in doped semiconductors or metals. T. Hori and E. H. Nicollian *et al.* described the conduction mechanism in details in their books [36, 37], especially focusing on the conduction mechanism valid in SiO_2. Most introductions on conduction in SiO_2 in this chapter is based on these books. A thermal SiO_2 film could be considered basically an ideal insulator under moderate bias conditions. Although the oxide resistivity is very high, on the order of $10^{15}\Omega$cm, it is not infinite. Hence currents flow through an oxide for any given gate voltage. Most amorphous insulator at an E fields in excess of 10^4V/cm show a range of nonlinear current-voltage dependence, and can be interpreted based on certain conduction mechanisms. In the following part we discuss Fowler-Nordheim (FN) tunneling, Poole-Frenkel (PF) emission, as well as direct tunneling transport, Schottky emission and ohmic behaviour. FN and direct tunneling are substantially independent of the temperature unlike other mechanisms, because of the dependence of tunneling phenomena on the quantum state in the insulator.

FN tunneling has been studied extensively in MOS structures where it has been shown to be the dominant current mechanism, especially for thick oxides ($> 40\mathring{A}$), such as in Fig. 3.1. The basic idea is that quantum tunneling of carriers occurs through a triangular potential barrier in the presence of a high electric field. The barrier of the insulator is pulled down by the E field so far that electron tunneling from the metal Fermi level into the oxide conduction band becomes possible. Once the carriers have tunnelled into the insulator they are free to move within the valence or conduction band of the insulator. To check for this current mechanism, experimental I-V characteristics are typically plotted as $ln(J_{FN}/E_{ox}^2)$ - $1/E_{ox}$, a so-called Fowler-Nordheim plot. Provided the effective mass of the insulator is known (for SiO_2, $m_{ox}* = 0.42m_0$), one can fit the experimental data to a straight line yielding a value for the barrier height under the valid E field.



ϕ_B...barrier height
E_{FM} ... metal Fermi level
E_{FS}... semiconductor Fermi level
d_{Ox}...oxide thickness
d_{FN}...tunneling distance V
V_{Gate}....gate voltage

Figure 3.1: Schematic of Fowler-Nordheim tunneling

FN tunneling implies that carriers are injected into the conduction band of the insulator and free to move through the insulator. However, in deposited insulators, which contain a high density of structural defects, this is not the case, Si_3N_4 is an example of such a material. The structural defects cause additional energy states close to the band edges and restrict the current flow by capture and emission processes, thereby becoming the dominant PF emission mechanism. Since the existence of a large density of shallow traps in CVD deposited films, Si_3N_4 makes PF emission a well-characterized mechanism and field-enhanced thermal excitations of trapped electrons into the conduction band are frequently observed in Si_3N_4 based non-volatile memory systems. Other possible processes include ohmic conduction as well as trap assisted tunneling, can be found in the literature [37]. In the simple case for $d_{Ox} \leq 2$ nm, direct tunneling dominates, the electrons pass through the full oxide thickness and the gate current is due to direct tunneling. Typically, Fowler-Nordheim tunneling and Pool-Frenkel emission with the corresponding J-E functions are described follows:

$$J = \frac{A}{4\phi_B} E^2 \exp \frac{-2B\phi_B^{3/2}}{E} \tag{3}$$

and

$$J \propto E \exp(-\frac{\phi_B - 2\sqrt{E/C}}{\phi_t}) \tag{4}$$

respectively, where J, E, and ϕ_t are current density, electrical field, and thermal energy of about 26 meV at room temperature. ϕ_B is the potential barrier height at the nanocrystal-insulator interface, and A, B and C are constants. For a substrate in strongly accumulation, where $E_F > \frac{4\phi_B^2}{3qBt_{ox}}$, the direct tunneling lead to a similar formula as Fowler-Nordheim tunneling and can be simplified to following with additional constant terms B1 and B2:

$$J_{dir} = \frac{q^2 m_{eff} E^2}{8\pi\phi_B m B_1} \exp[\frac{-4\sqrt{2m(q\phi_B)^3} B_2}{hqE}] \tag{5}$$

3.1.2 Basic programming/erasing mechanism for MOSFET

In floating gate memories, the charge used for programming the device has to be injected into the floating gate. In order to change the charge or data content of NVM, two major tunnel mechanisms are viable: FN tunneling through thin oxides and channel hot-electron (CHE) injection.

We already discussed the FN tunneling, as a very important injection mechanism valid in NVM. In the operation of NVM, it can also be programmed by hot-carrier injection. The method of programming is by hot-electron injection for n-type NVM built on p-substrates or by hot hole injection for p-type NVM built on n-substrates. Hot-hole injection is relatively slow due to the hole mass (0.56 m_e) as well as the Si-SiO_2 energy barrier of 4.7 eV for holes, which is why most NVMs manufactured today are n-type MOSFET on p-substrates. The memory cell is programmed by charging the floating gate via the injection of hot-electrons from the drains pinch-off region. The hot-electrons get their energy from the voltage applied to the drain of the memory cell. They are accelerated by the lateral electric field along the channel into even higher fields surrounding the drain depletion region. Once these electrons gain sufficient energy they surmount the energy barrier of 3.2 eV between the silicon substrate and the dielectric layer or gate oxide and arrive at the floating gate. Hot-electron injection also is commonly used in actual flash memory cells, but is not the operation mechanism for the erase procedure. However frequent CHE operations normally degrade the tunnel oxide integrity of flash cells, thus the program/erase cycling endurance is worsened. The basic strategy for minimizing hot-carrier effects in MOSFETs is to introduce a lighter ion implanted region adjacent to heavily drain doped region or lightly nitried gate oxide.

3.1.3 High frequency capacitance-voltage analyze

Grove (the founder of Intel) *et al.* used the MOS capacitor as the main tool in the study of the MOS system [38]. Here the differential capacitance is the most essential property, because small-signal measurements determine the changing rate of the charge with voltage. To understand capacitance-voltage measurements one must first be familiar with the frequency dependence of the measurement. This frequency dependence occurs primarily in inversion since a certain time is needed to generate the minority carriers in the inversion layer. High- and low-frequency C-V measurements are often useful among various methods to evaluate the MOS characteristics. Most capacitance measurements are performed with admittance bridges or capacitance meters. Using Gauss' law, the small-signal equivalent circuit of the MOS capacitor was derived as follows [36]:

$$\frac{1}{C} = \frac{1}{C_{si}(\psi_s)} + \frac{1}{C_{ox}} \tag{6}$$

Eq.6 gives the total capacitance of the MOS devices as the series sum of the silicon capacitance and the oxide capacitance, per unit area. The majority and minority carrier response times to ac gate voltages are very different. The minority carrier response is typically as long as 0.01-1s, much slower than the frequency of bias at high frequency and certainly not instantaneous over the frequency range of interest.

The high frequency capacitance is used to evaluate the fixed-charge density and doping density. The low frequency capacitance is conventionally used to evaluate the interface-state density and surface potential [39]. At high frequency, typically 0.1-1MHz, majority carriers respond instantaneously. It is instructive to consider the band diagrams for the MOS structures discussed in this chapter. Fig. 3.2 shows the energy-band diagrams for ideal MIS diode structure using (a) n-type and (b) p-type semiconductor substrates. For these ideal structures, at zero applied voltage on the metal gate, it is a state of flat band. However, because of the difference between the gate metal workfunction, ϕ_M, and the semiconductor workfunction, ϕ_s, many dielectrics exhibit a charge (Q_f), resulting in a required applied voltage $V_{FB} \neq 0$ to achieve a flat band condition.

The simplest and most widely used method for measuring oxide charge density is to infer this density from the voltage shift of C-V curves. Because of the existence of oxide charge, as shown in the right part in Fig. 3.2, the entire C-V curve is shifted along the voltage axis with respect to the ideal high frequency C-V curve. For n-type substrate, positive charge in oxide causes the C-V curve to shift to more negative value of the bias with respect to the ideal C-V curve, and negative charge cause the C-V curve to shift to more positive value of the bias with respect to the ideal C-V curve. For p-type substrate, the opposite is valid.

The bias shift of the C-V curve caused by oxide charge Q can be explained by image charges. Using a n-type substrate as an example, for a certain gate bias without Q, at depletion region, the ideal depletion layer width is such that negative charge on the gate is balanced by the positive dopant ions in the depletion layer. If positive Q is joined in the oxide as shown in the upper-right side of Fig. 3.2, the above charge balance is interrupted, its image charge (actually electrons for this case) is introduced in the silicon substrate. These additional electrons are located at the depletion layer, partly neutralize and reduce the depletion layer width, because the capacitance of Si (C_{si}) is inverse to the width of the depletion layer (C_{si} is linearly dependent on its flat-band capacitance, flat-band capacitance is inversely proportional to extrinsic Debye length which is related to depletion layer width) and series connection of C_{si} and C_{ox}, thus from eq.6, the actual capacitance with oxide charge Q becomes larger than for the ideal capacitance without Q. At strong accumulation, this influence of image charge is omitted because of the accumulation of carriers from the substrate, capacitances are always equal to C_{ox} and at strong inversion, capacitances reach the same saturation value and the image charge is omitted too. Therefore, for MOS structures containing Si NCs charge in the oxide, its high frequency C-V looks like a hysteresis loop. It has the same values at two extreme sides, but different value in the remaining middle part, such as in the depletion region and weak inversion region.

3.1.4 Interface trap properties from conductance method

After certain assumptions are made, a general equivalent circuit is derived for the MOS capacitor, on the basis of capture and emission by interface trap levels distributed throughout the silicon bandgap. We are noting that there are many prevailing mechanisms to explain the origin of trap phenomena of nano-dots in the dielectric matrix. The interface traps between dots and the surrounding SiO_2 matrix have also been suggested to explain the charge storage, because of the large surface volume ratio of the dots.

Among the various methods of interface trap investigations, the conductance method is the simplest but effective one to use. That is, in the depletion region interface trap occupancy changes by capture and emission of majority carriers, since the minority carrier density is very low in depletion. For simplicity, we directly obtain the definition of the interface trap admittance over the silicon bandgap Y_{it} from Nicollian [36], and its real part is

$$\frac{\langle G_p \rangle}{\omega} = \frac{\omega C_{ox}^2 G_m}{G_m^2 + \omega^2 (C_{ox} - C_m)^2} \tag{7}$$

Here G is conductance. Only $\frac{\langle G_p \rangle}{\omega}$ will peak as a function of frequency f in the curve. The procedure is to measure the amplitude change of this curve either between the points f_p and f_p/n or between the points f_p and nf_p where f_p is the frequency corresponding to the peak value of $\frac{\langle G_p \rangle}{\omega}$. Here we choose n=5. The ratio of the amplitudes are single-valued function of σ_s (standard deviation of band bending), using the experimental ratios, the corresponding σ_s is obtained from the available figure in the book of Nicollian [36]. The defined universal function $f_D(\sigma_s)$ is dependent on the σ_s [36], therefore, we will obtain the interface trap density as follows, where q is electric charge:

$$D_{it} = (\frac{\langle G_p \rangle}{\omega})_{f_p} [f_D(\sigma_s) q]^{-1} \tag{8}$$

In summary, we need only two points on a $\frac{\langle G_p \rangle}{\omega}$ versus log frequency, plus additional data from empirical curve to estimate D_{it} with unit of $10^9 cm^2 eV^{-1}$-$10^{10} cm^2 eV^{-1}$.

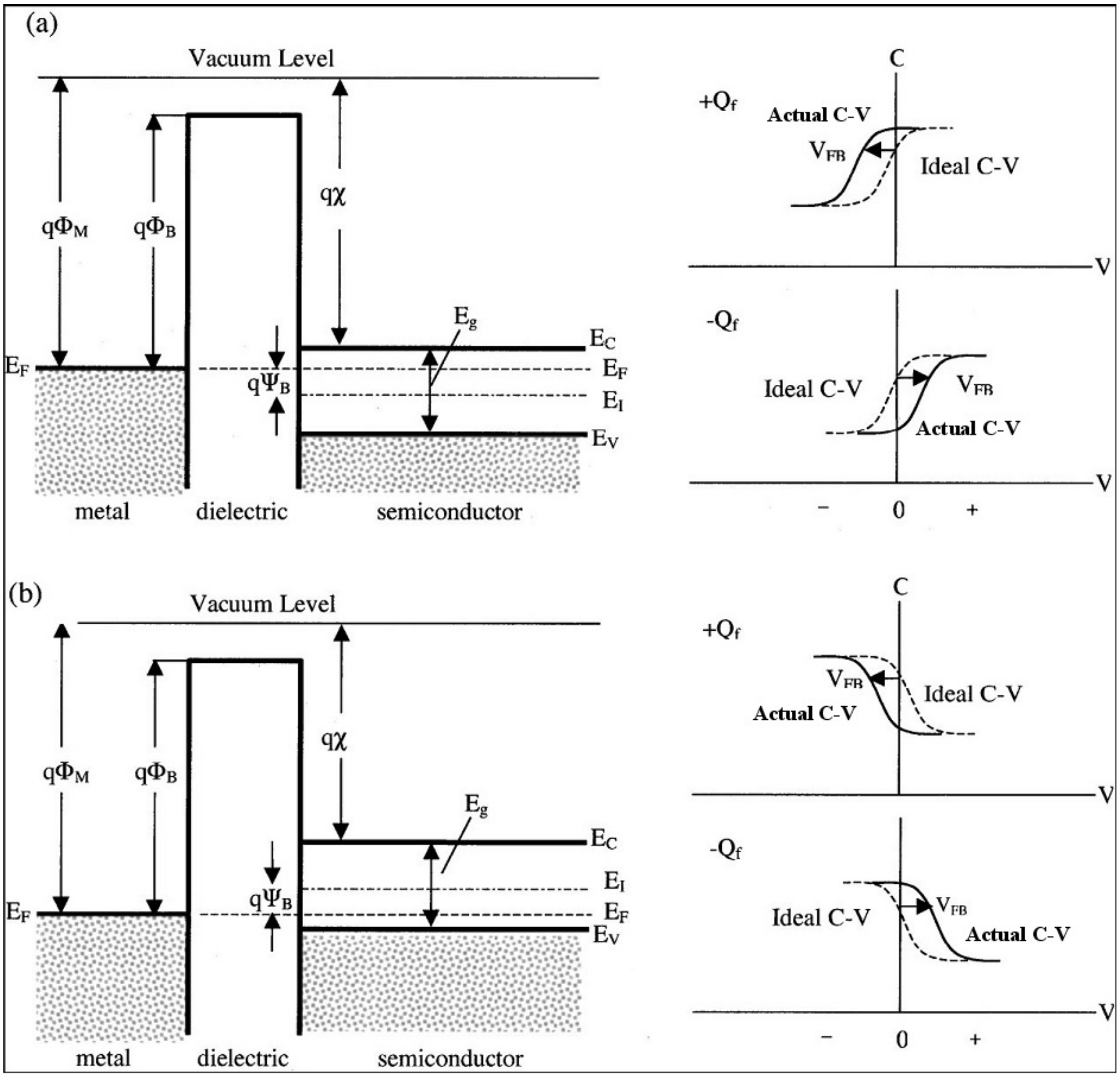


Figure 3.2: Energy-band diagrams and associated high frequency C-V curves for ideal MIS diodes for (a) n-type and (b) p-type semiconductor substrates. For these ideal diodes, V=0 corresponds to a flatband condition. For dielectrics with positive (Q_f) or negative ($-Q_f$) fixed charge, an applied voltage (V_{FB}) is required to obtain a flatband condition and the corresponding C-V curve shifts in proportion to the charge in dielectrics [40].

3.2 Conduction mechanism of the sample

Usually, current-voltage (I-V) measurements can indicate the conduction meachanism. Quasi-static I-V measurements were performed using a Keithley 6517 electrometer. A staircase voltage with a fixed voltage step of 0.1 V and a step delay time of 15 s were applied to the device under investigation. To measure correctly, errors due to displacement currents should be avoided. Hence, the I-V characteristic has to be realized which means the applied voltage should be swept as slowly as possible and a time delay should be used to measure the steady state characteristic. The voltage was swept to negative values from 0 V to -10 V for the sample A1 and the control sample A2, hence, into the accumulation region. Usually, the leakage current is required to be lower than 10^{-8} A/cm^2 in a dielectric film used for a memory storage capacitor. The behavior of the characteristics has two regions for common dielectric material. In the low field region where field E $<$ 10^5 V/cm, the current is due to thermally generated carriers. While in the higher field region, in order to analyze the conduction mechanism of our sample, we fitted our experimental J-E curve by available models at these electrical field.

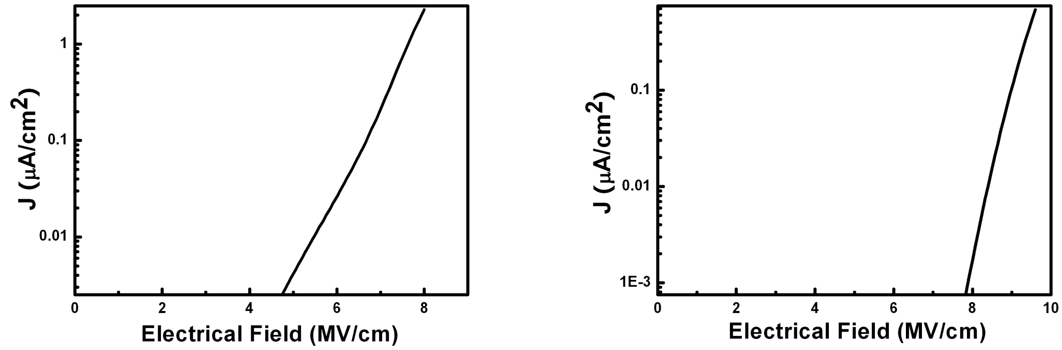


Figure 3.3: Typical J-E curve for the sample A1 (left) and control sample A2 (right). Under same E field, the current density of control sample A2 is obviously less than for sample A1

In Fig. 3.3 we plot the J-E curves for the sample A1 and control sample A2. It is obvious that Si NCs inside the matrix enhances the conductivity of sample A1, so the current density of control sample A2 is much less than sample A1 only about 1/1000 under the same E field. Above the threshold E field, the current densities of samples are increasing extremely. The threshold E field of sample A1 is less than the value of sample A2 because of the embedded Si NCs in sample A1. In fig 3.4

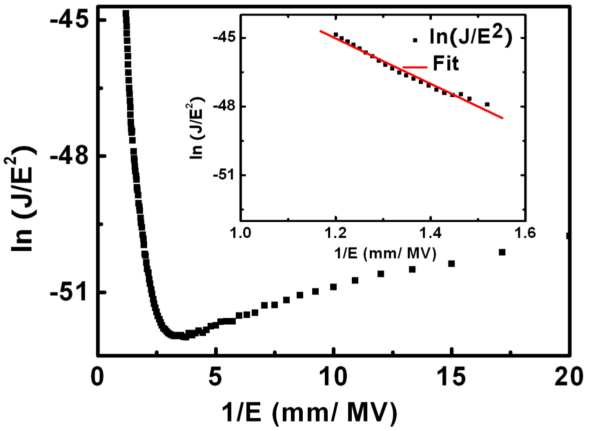


Figure 3.4: Fowler-Nordheim plot of the silicon nanocrystal MOS structure in sample A1. The inset shows the linear fitting of the Fowler-Nordheim plot for a limited range of 1/E values.

we plotted $ln(J/E^2)$ as a function of 1/E, which is known as a Fowler-Nordheim plot. Linear fitting prove the FN tunneling is the main conduction mechanism for the sample A1 under certain E field. Using the linear part of the Fowler-Nordheim plot and choosing typical values of the effective mass of the electron $0.26m_0$ to fit eq.3, we interpret the potential barrier to be about 1.6 eV. This parameter only indicates the effective barrier height and is the height between the isolator barrier SiO_2 and the quantized states of the silicon nanocrystals, the latter being shifted up by quantum confinement effects. Carrier injection from the gate electrode into the oxide of an MOS structure containing Si NCs is schematically indicated in Fig. 3.5. The electric field on the insulator is represented by the sloped potential curve, steeper slopes mean higher E field. When the local field is high, carriers tunnel through the triangular barrier and enter the Si NC conduction band and sequently tunnel to the conduction band of the Si substrate via the triangular barrier on the right side. Please note, that the sizes of the silicon nanocrystals are not completely identical, but have a certain size distribution which means a certain distribution of the effective potential barriers. In addition, we got some indications from excitation of photoluminescence measurements by Heitmann *et al.* [41] that the barrier height might be reduced due to a very thin non-stoichiometric oxide layer on the surface of the nanocrystals. Using photon-in photon-out soft X-ray spectroscopy, the electronic structure of Si NC embedded in a SiO_2 host matrix was revealed by a cooperation with the BESSY's group [42]. They confirm the Si NC to be of a core-shell structure with a crystalline Si core and thin layer of suboxide. The transition between the Si core and surrounding SiO_2 has to occur gradually. Because this transition layer varies with stoichiometry,

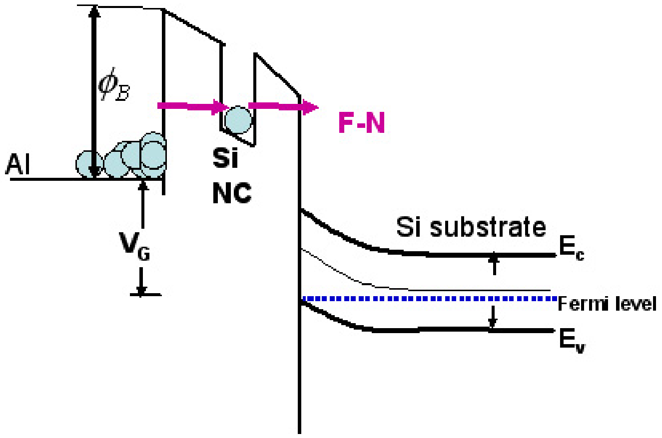


Figure 3.5: Band diagram of MOS structure under reverse bias and its tunnelling process

the electronic structure varies as well. Such an effect might also be of influence for the tunneling behavior.

3.3 Detection of the oxide breakdown in MOS sample

It was shown FN tunneling as a prevailing conduction mechanism in the test MOS sample. However, FN tunneling is worked under high E field (MV/cm level) operation, from the point of view of dielectric breakdown of the insulating oxide layer, reduplicatively using the FN tunnelings to force the carrier for charge/discharge performance gradually make defect or carrier accumulation, inevitably deteriorate the quality of oxide and finally cause percolation of this thin insulating oxide, thus breakdown happens.

Gate insulator breakdown effects the conducting behavior due to dielectric film became conductive, leakage current from gate to substrate is rising distinctly. But in the I-V measurement, it is hard to distinguish the breakdown existence only by the current jump. The fluctuation of current density can be explained by a lot of other reasons, such as thermal excitation, mechanic-stress.

Since breakdown path gets carrier through and have heat generation in the transport process, we get the hint from solar cell technology [43]. Using thermography method, several kinds of shunts in solar cell can be detected by mapping the thermal contrast. We can confirm whether has the breakdown below the metallization by this method as well. If so, detection and localization of breakdown by thermal contrast in thermography image can be done.

Lock-in IR thermography measurements were done at TDL 384 M lock-in system

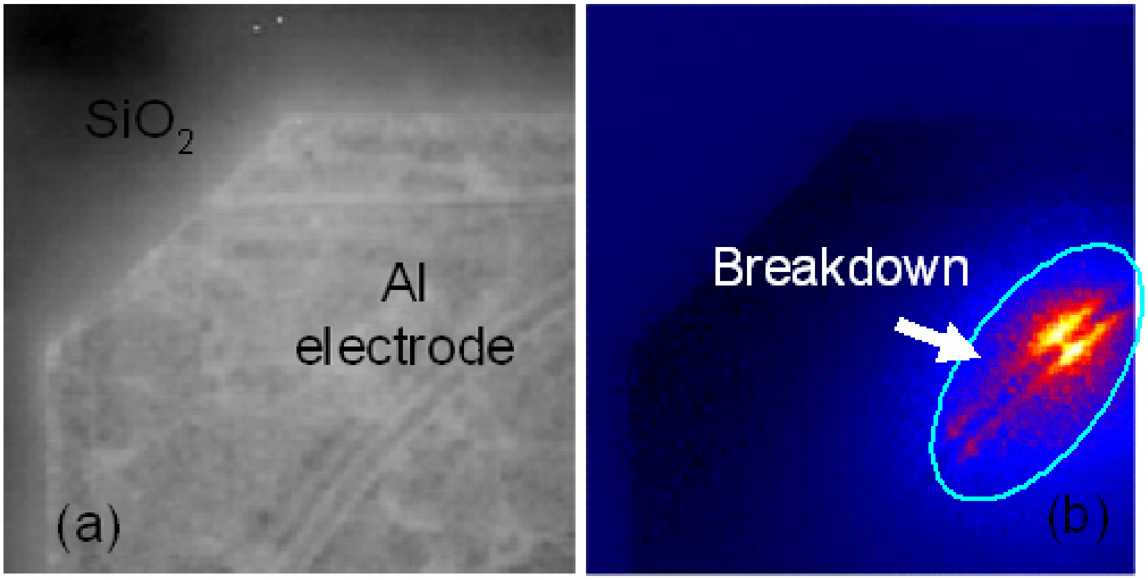


Figure 3.6: (a) Topography image of the breakdown sample. (b) Thermography image.

at 10Hz for breakdown samples under 0.5 V forward bias condition, IR modulated pulses penetrate the samples and detect by MCT detector. Their images are shown as in the Fig. 3.6.

Images are 138x138 pixel with pixel resolution 5 μm. In the topography image (Fig. 3.6(a)), SiO_2 and hexagonal Al electrode are visible, but Al electrode was scratched in the operation. In the thermography image (Fig. 3.6(b)), breakdown is exactly localized underneath the scraping Al metallization layer, probably this scratch attenuated the oxide thickness and breakdown ultimately.

3.4 The storage and charge capacity of the Si NCs

To investigate the trapping characteristics of the structure we performed high frequency (100 kHz) multiple up-down C-V and conductance-voltage (G-V) sweeps beteween inversion and accumulation region for our MOS test structure. The measurements were carried out using a HP4194A impedance analyzer in a dark box for avoiding photocurrent. Due to the ability of storage and trap-detrap processes of carriers in Si NCs embedded in the dielectric matrix [44], we can describe the properties by high frequency C-V curves. The C-V and G-V measurements of sample A1 are presented in Fig. 3.7 and Fig. 3.8, respectively. For the C-V characteristics the voltage was swept forth and back between -4 and 2 V, which means for a p-type MOS structure, bias swept from the accumulation region to the inversion region, and a hysteresis was observed only for the sample containing the nanocrystals. Comparison with the schematic MOS C-V case in Fig. 3.2, the observed hysteresis was clockwise indicating a net negative charge. If there is no charge present in the oxide or at the oxide-semiconductor interface, the flat-band voltage simply equals the difference between the gate metal work-function ϕ_M, and the silicon work-function ϕ_S. The actual value of the work-function of a metal deposited onto silicon dioxide is well known from the literature, such as about 4.1 eV for Al. The work-function of silicon ϕ_S requires some more thought since the Fermi energy varies with the doping type as well as with the doping concentration. This work-function equals the sum of the electron affinity in the semiconductor χ, the difference between the conduction band energy and the intrinsic energy divided by the electronic charge in addition to the bulk potential. This is expressed by the following equation:

$$V_{FB} = \phi_M - \phi_S = \phi_M - \chi - \frac{E_g}{2q} - V_t \ln \frac{N_a}{n_i} \tag{9}$$

where N$_a$ is doping density and dependent on sample, all other value are constant and available from literature. One also notices in Fig. 3.8 the presence of a conductance peak with a peak position close to the flat-band voltage. The magnitude of the hysteresis in C-V and the shift in the peak position in G-V are both about 0.35 V. In contrast, no hysteresis and no conductance peak shift in the C-V and G-V characteristics, respectively, were observed for the control pure MOS sample A2. Therefore, these hysteresis and peak shifts could be attributed to electrons trapped

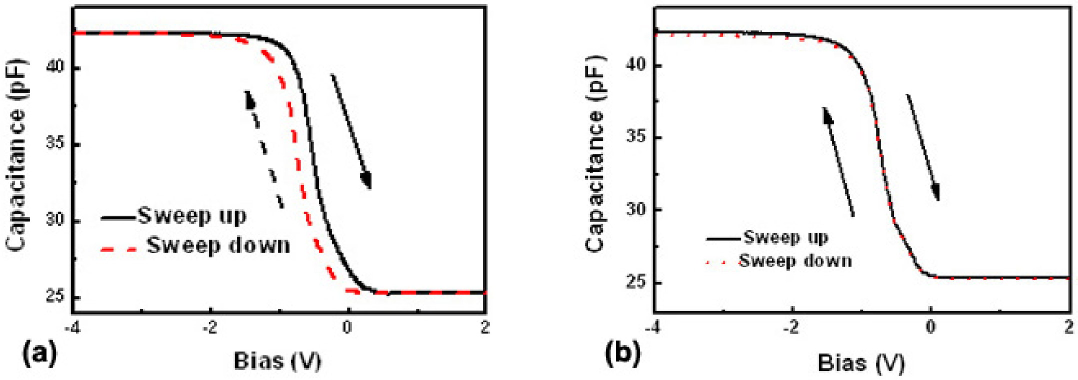


Figure 3.7: Typical C-V curve of sample A1 in (a) and A2 in (b) obtained by sweeping gate voltage forth and back between -4 and 2 V.

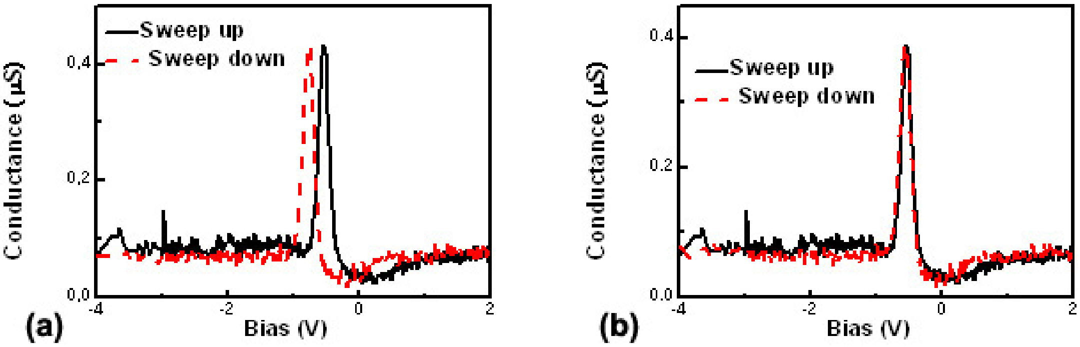


Figure 3.8: Typical G-V curve of sample A1 in (a) and A2 in (b) obtained by sweeping gate voltage forth and back between -4 and 2 V.

in the sandwiched silicon nanocrystals or at the interface of the nanocrystals. Defect states in the oxide or at the silicon substrate-oxide interface in control sample can not be responsible for such a behavior. We estimated the charge density by using a model of charging a single layer of silicon nanocrystals of density n$_{dot}$, which was suggested by S. Tiwari [3] using the following equation:

$$\triangle V_{FB}^{T} = \frac{e n_{dot}}{\epsilon_{ox}}(t_{upper} + t_{dot}\frac{\epsilon_{ox}}{2\epsilon_{si}}) \tag{10}$$

where $\triangle V_{FB}^{T}$ is the flatband shift, t_{upper} is the upper gate oxide thickness, t_{dot} is the silicon nanocrystal diameter, ϵ_{ox} and ϵ_{si} are the permittivities of the oxide and silicon, respectively. Using the given parameters, the nanocrystal density of $1.6 \times 10^{12}/\text{cm}^2$ can be estimated from the above equation for a single layer of Si NCs if one electron per nanocrystal is assumed. In addition, we can estimate the silicon nanocrystal density from our TEM images. However this is a rather crude estimation and represents the upper limit of the nanocrystal density because the silicon nanocrystals shown in the TEM image are from a sample thickness of some tens of nanometers. On the other hand, from an in-plane view of a single nanocrystal layer we know that the nanocrystals are nearly homogeneously distributed within the plane of the layer which supports our estimation. Comparing the estimated nanocrystal density with the above value derived from C-V measurements we can conclude that the nanocrystal density is in the same range as the charge density.

In our sample, the Coulomb charging energy was estimated to be about 93 meV by using $q^2/2C_{self}$ in table 1 of chapter 1, where C$_{self}$ is the self-capacitance of a Si nanocrystal embedded in SiO$_2$. This charging energy is larger than the thermal energy and properly can block additional electrons to be stored in the conduction band of the Si NCs. The observed clockwise hysteresis is attributed to charge trapping from the gate. When the gate electrode is in writing operation at -4 V the charges are stored in the Si nanocrystal by Fowler-Nordheim tunneling through the thin oxide, and if the gate is swept into erase operation, a number of charges will be able to escape from the Si nanocrystals, and the neutral charge state is achieved again. The whole sweep process results in a shift of the capacitance. In our case, the sample structure was realized by a symmetric SiO_2/Si nanocrystals/ SiO_2 structure which means that if we apply a bias, this forces a tunneling of the electrons of the metal gate or the holes of the p-substrate into the silicon nanocrystals simultaneously. However, it is known that the electrons more easily tunnel into the nanocrystals than holes which is related to the larger barrier height (4.7 eV) for the holes tunneling in the valence band.

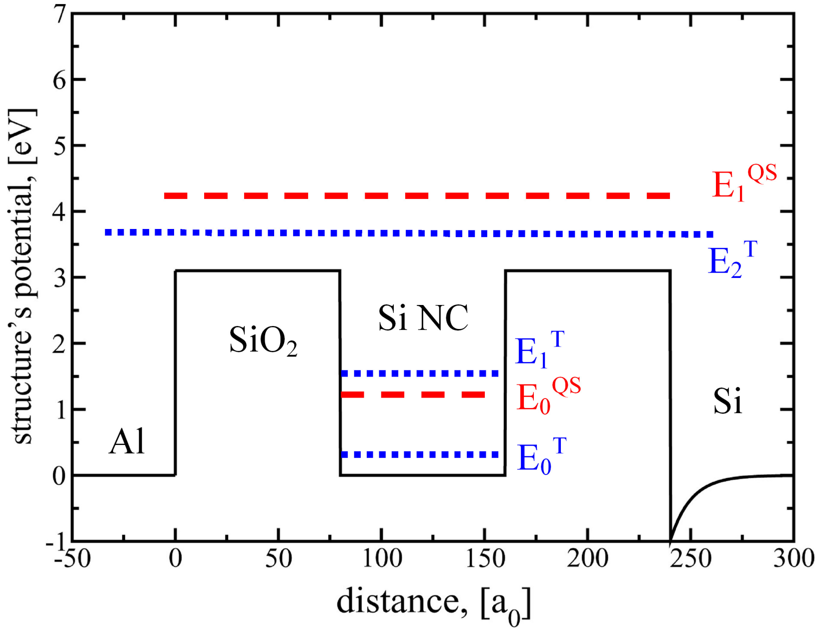


Figure 3.9: Energy spectrum of the MOS structure model. One QS state is presented within the well and its energy is lower than the barrier maxima.

A model with a one-dimensional potential is used for the description of the electron tunneling current through the MOS system as described in Fig. 3.9. This spectrum was calculated by J. Berakdar (theory department of MPI-Halle). Considering all the parameters given in the current sample, there are two kinds of energy states associated with Si NCs. They are conventionally referred as quasi-stationary (QS) states and transparent states. The term "quasi" means that the electron deposited within the well at some time t_0 will tunnel through the barriers. Such initially well-prepared state is decaying in time and has a finite and large lifetime. Moreover, if the energy of the electron coincides with the energy of the transparent state, the particle behaves, as if it would be free. For given characteristics (heights and widths of the barriers, material dielectric constants and respective effective masses, a_0 is atomic unit with about equal distance 0.45 Å) there is a single QS state within the well (red dashed line) to assist the tunnel process. Its energy level is lower than the barrier maxima and the barrier height is about 1.7 eV. This value is consistent with our electrical measurement.

3.5 Interface charge estimation from conductance method

Since defects in MOS structures cause instabilities, we can figure out that the silicon nanocrystal-SiO_2 interface quality will remain a key factor for future silicon nanocrystal memory technology. The conductance method is a simple but powerful tool to estimate the interface state parameters. For each bias point in the depletion region, we measured and plotted the G/ω value as a function of the frequency f using the HP4194A semiconductor analyzer. The results are presented in Fig. 3.10, where G is the conductance, ω is the angular frequency, and f is the frequency and the f_P is the frequency corresponding to the peak of the G/ω curve. The standard deviation of band bending and the universal function can be obtained sequentially from the G/ω ratio between the frequency, either low $f_p/5$ or high $5f_p$ and the frequency f_p corresponding to the peak of the measured G/ω.

The interface trap density calculated using these two parameters is about $10^{10}/cm^2eV^{-1}$, which includes the interface states between the Si substrate and the SiO_2. Except the interface trap density, the intrinsic trap time constant $\tau_p(\tau_n)$ can be calculated from $\tau_p = \frac{\xi_p}{\omega_p}$. It is at the μs level, where ξ_p is an empirical value and proportional to the standard deviation of band bending. It is inverse to the carrier tunnel rate and significantly influences the operation speed of memory devices. The capture cross section is related to the capture probability by $\sigma_n = c_n/\arg\nu$, where $\arg\nu$ is the average thermal velocity of the carriers. The capture probability c_n is calculated from the interface trap time $\tau_n = \frac{1}{c_n N_D}\exp(-\langle v_s\rangle)$, therefore the capture cross section can

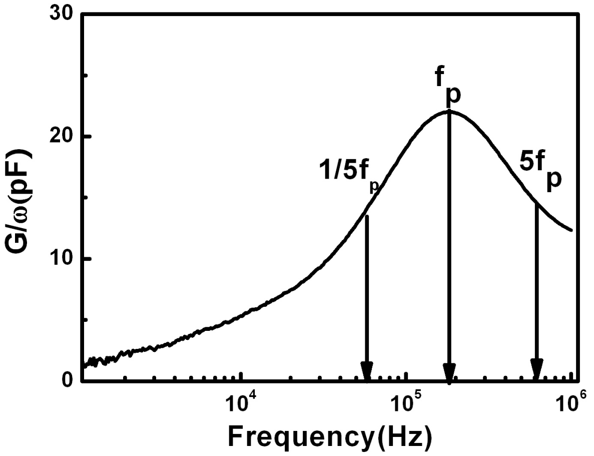


Figure 3.10: G/ω vs. the frequency f for sample A1, the value at 5f$_P$ and f$_P$/5 define the width of the curve on the high and low frequency side of f$_P$, respectively.

be obtained. Typically for the sample A1, based on its resistivity, the doping density is about $1 \times 10^{16}/cm^3$, τ_p is about 1.9 μS, thermal velocity of the carriers is approximately corresponding to a thermal energy of 3k$_B$T/2 which leads to 10^7cm/s, therefore the capture cross section is at a level of 10^{-16}-10^{-17} cm^2.

It is well known that hydrogen passivates silicon dangling bonds at the Si-SiO$_2$ interface in MOS devices thereby reducing the interface trap density. In our case, phase separation requires long time annealing of the former SiO layers which could reduce the trap density at the silicon nanocrystal/SiO$_2$ interface too. For floating-gate non-volatile memories, repeated write/erase cycling will reduce the isolating capability of the tunnel oxide, thus enhancing the floating-gate charge loss by stress induced leakage current (SILC) after many periods of cycling. A quantitative model for steady-state SILC for flash EPROM was developed before [45], which shows a linear relation between interface trap density and steady state SILC. The SILC we estimated via the interface trap density is far below the Fowler-Nordheim tunneling current density measured in the way described above. Hence, we conclude that our silicon nanocrystal MOS structure demonstrates a good retention capability.

We therefore conclude that on average all silicon nanocrystals inside the oxide matrix are charged. The interface trap density estimated from the conductance method is much lower than the sample charge density. This is a further indication that the dominant charges are confined in the silicon nanocrystals.

3.6 Conclusions of the chapter

A size controlled SiO/SiO_2 approach was used for the fabrication of single layers of silicon nanocrystals in a MOS structure. Electrical characterizations are performed and demonstrate the suitability of this approach for silicon nanocrystal based memory devices. The charge tunneling behavior of samples which contain a single layer of Si NCs was studied. Static I-V measurement show that under certain bias, carrier tunnel the tunnel oxide by FN tunneling behavior. Borrowed IR-Thermography technology from solar cell, we confirm the breakdown existence and localize it in the thin oxide after long term tunneling with ultrahigh E field. Based on the C-V and the G-V measurements and the conductance methods, the phenomena observed could be explained by the charge storage effect of electrons within Si NCs. Because of the fairly uniform lateral distribution of Si NC, we observed nearly complete charging for all Si NCs. It will help us to understand the capacitance of the gate oxide containing Si NCs and optimize the properties of the transistor. The approach is well matched to the standard silicon technology and hence will be of benefit for the discussed application in NVM.

CHAPTER IV

Multilevel Charge Storage in Silicon nanocrystal Multilayers

4.1 The concept of multi-bit/cell

Recent performance improvements in products such as mobile phones, laser printers, and GPS systems have brought significant increases in cell number and data size requirements. This results in an increasing market for flash memory with faster speed, greater density and higher integration. Many companies are developing SONOS trapping site storage devices. Such SONOS flash has the advantage of being able to storage two physical bits as multiple storage levels. The multi-bit cell technology usually stores two bits of data per cell, or per transistor, reducing the area of the memory array by as much as 50% compared to single-bit cell technology, thus reducing the cost per bit significantly. However, the multi-bit cell technology needs a distinct distribution of several threshold voltages (V_T), because of the need of sufficient V_T windows between multi-states. Liu *et al.* proposed a source/drain flash memory device for two-bit per cell storage in his recent IEEE paper [46]. Fig. 4.1 shows how this two bit/cell flash memory concept works: flatly localized storage charges at the floating

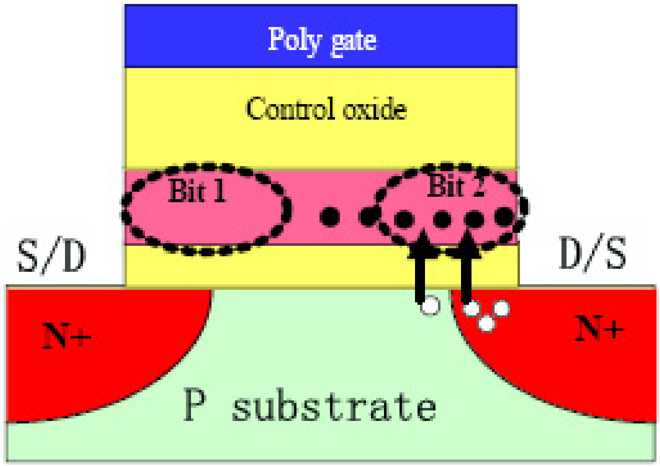


Figure 4.1: MirrorBit technology symmetric design which doubles the density of a flash memory array. The two bit/cell layout greatly simplifies the manufacturing process.

gate were obtained through channel hot electron programming by reverse reading of the polarities between source and drain allows a single cell with the two different discrete traps, one near the source and another near the drain as show in Fig. 4.1. Fig. 4.2 schematically shows the corresponding four states of the memory. When a large negative voltage is applied to the gate electrode, the electrons in the traps are ejected and the memory is 'Erased'. When a positive drain bias and a large positive gate voltage are applied, the electrons are injected into the traps near the drain edge due to local hot carriers, which represents the 'Write (D)' state. On the other hand, when source and drain are switched i.e. the positive bias is applied to the source and positive gate voltage is applied, the electrons are injected into the trap near the source edge which is the 'Write (S)' state. When electrons are injected into traps near both drain and source sides under sufficient lateral E-field, thus this represents the 'Write (S+D)'state. This operation has been successfully demonstrated in NAND flash devices. A good improved performance was shown for memory applications [47]. A novel vertical channel NVM memory cell with ONONO (dual nitride trapping layers) dielectrics stack was proposed recently at ESSDERC2006 [48] for the multi-bit per cell operation. According to the prediction of the Roadmap, 8-bit/cell operation could be realized by 2010. Although the system of Si NCs embedded in SiO_2 has a much steeper potential wall for hole and electrons compared with Si_3N_4 traps sites, the realization of two-bit/cell device by using this S/D nanocrystal structure has not

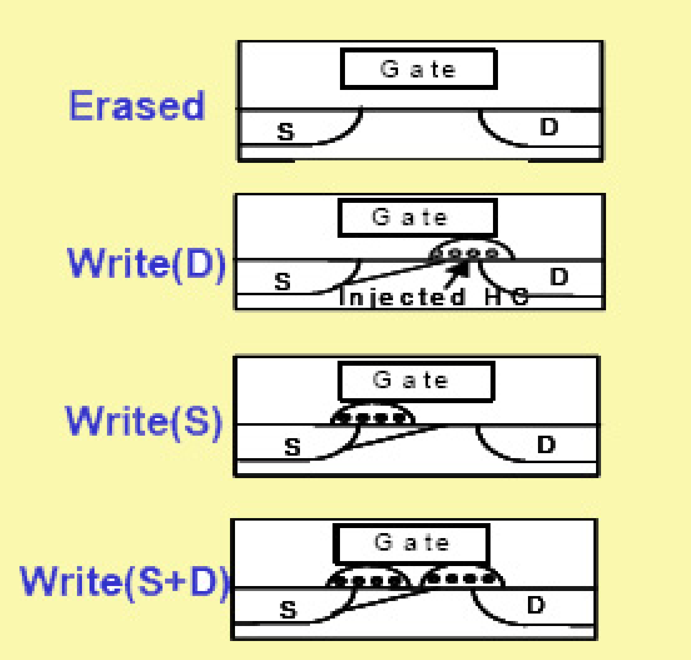


Figure 4.2: The scheme and working principle of two-bit operation. It has four working state sequentially as described in text.

been realized so far.

4.2 Si NC/SiO$_2$ multilayer introduction

The concept of multilevel charge storage using a quasi-superlattice amorphous Si and Si_3N_4 structure was suggested recently [49]. In this kind of structure, the electrons are captured at the charge-trapping sites of trap states of the Si_3N_4 layers and interface states between the Si_3N_4 and a-Si layers. Considering the naturally discrete charge storage of the Si_3N_4, such a stack structure might be charged randomly, thus influencing the multi-bit/cell performance.

I will concentrate on the charging process of multiple, regular localized Si NCs layers in a dielectric matrix. Such Si NC based layered structures are quite similar to the above nitride based structures but have the advantage of artificial discrete charge nodes. Here I typically chose a set of samples (B1-B3), which have the same total oxide thickness, but have 1,2 and 3 layers of Si NC embedded in the matrix. Afterwards, I proceeded samples to fabricate the MOS structures and performed high frequency (100 kHz) multiple up-down C-V and G-V sweeps as discussed in the previous chapter. During the C-V and G-V measurements, the programming biases gradually rise. Comparing a family of C-V curves, the multilevel step charge storage was proved by sequential charging of up to three Si NC layers embedded in the oxide for multi-bit/cell performance. A ultrahigh charge density was realized too. This concept might overcome the limitation of previously introduced multi-bit design, and simplify the circuit design as well as increase the integration level.

4.3 The permittivity of Si NCs in a SiO$_2$ matrix

We already know that in the design of MOSFET devices, the gate oxide capacitance is an effective parameter to influence the drain current. In the literature [6], the drain current in the saturation region is given as follows:

$$I_D = (W/L)C_{ox}(V_G - V_T)V_D\mu \tag{11}$$

where W and L are width and length of the transistor, μ is the electron mobility, V_G, V_T, V_D are gate bias, threshold value and drain bias. C_{ox} is the gate capacitance.

When the Si NCs are embedded in the SiO_2 film, the dielectric properties of the film will be different from that of the pure dioxide film. Therefore, for flash memory devices with the Si NCs embedded in the gate oxide, the existence of the Si NC will definitely affect the gate capacitance. On the other hand, as a Si NC is an isolated structure with a size of about 4 nm embedded in the SiO_2 matrix, its physical properties should be different from that of bulk crystalline Si. Therefore understanding the dielectric properties of the Si NCs is inevitably important for understanding the device behavior. In addition, if the dielectric properties of the Si NCs are known, the dielectric properties of a SiO_2 thin film containing the Si NCs are clear too.

The fabrication processes of sample P are similar to those introduced in the previous chapter. Typically the 3 periods of the SiO/SiO_2 superlattice film were grown by thermal evaporation with 12 nm capping SiO_2 and a total thickness of 36 nm. High temperature annealing was subsequently carried out at 1100 ^oC in N_2 ambient for 1 h in order to induce the Si NCs formation. The profile of Si NCs in the SiO_2 matrix can be easily obtained from the phase separation reaction. The Si NCs are distributed layer by layer throughout the oxide. As such, the dielectric constant of the oxide film is no longer equal to 3.9 of pure SiO_2. The MOS capacitance of the structure with Si NCs distributed in the oxide will be different from the capacitance of a MOS with a pure SiO_2 thin film of the same thickness.

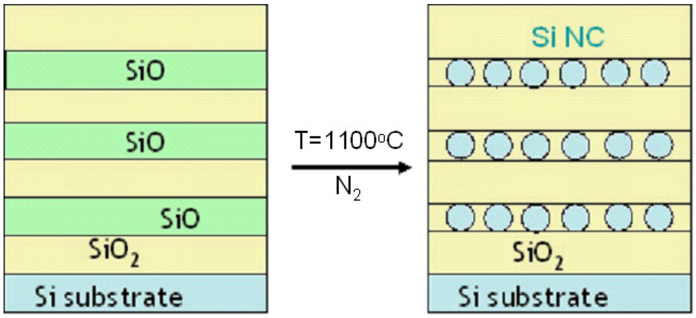


Figure 4.3: The scheme of process flow for Si NC formation (not to scale).

For simplicity, we assume that the volume fraction of Si NCs v_i in the previous SiO layers is constant and related to the stoichiometry of the SiO layer. The whole oxide layer is virtually divided into N pairs of sublayers with an equal $SiO+SiO_2$ thickness plus the top oxide thickness as shown in Fig. 4.3 where N is 3. Each sublayer can be schematically described as an effective medium, which is composed of the SiO_2 layer with embedded Si NCs and the pure SiO_2 layer. The effective permittivity ϵ_e of the SiO_2 containing Si NCs is represented by the Maxwell-Garnett effective medium approximation (EMA) [50]:

$$\frac{\epsilon_e - \epsilon_{SiO_2}}{\epsilon_e + 2\epsilon_{SiO_2}} = v_i \frac{\epsilon_{NC} - \epsilon_{SiO_2}}{\epsilon_{NC} + 2\epsilon_{SiO_2}} \tag{12}$$

where ϵ_{SiO_2}=3.9 is the permittivity of pure SiO_2 and ϵ_{NC} is the permittivity of the Si NCs, v_i is the volume fraction of Si NC in layer. Based on the above serial sublayer capacitances, the total MOS capacitance C per unit area in the accumulation region can be expressed as:

$$\frac{1}{C} = \sum [\frac{\epsilon_e \epsilon_0}{d_0} + \frac{\epsilon_{SiO_2}\epsilon_0}{d_1}]^{-1} + [\frac{\epsilon_{SiO_2}\epsilon_0}{d_{top}}]^{-1} \tag{13}$$

where ϵ_0 is the permittivity in vacuum, d_0, d_1 and d_{top} are the size of Si NC, thickness of neighbor layer of SiO_2 and the top oxide thickness, respectively.

Using molar, atom weight and density listed in table 3, we can deduce from eq.14 the volume fraction v_i of Si NC for each sublayer from these parameters which is about 33.6% . Eq.12 is used to calculate ϵ_e and replace it in eq.13. The thicknesses

Table 3: Parameters of SiO, Si and SiO_2

	SiO	Si	SiO_2
atom weight	44	28	60
density (g/cm^3)	2.15	2.33	2.53

$$\begin{aligned} 2SiO &\rightarrow SiO_2 + Si \\ 88 &\rightarrow 28 + 60 \end{aligned} \tag{14}$$

d_0, d_1 and d_{top} are estimated from the respective TEM images (Fig. 4.4). The 291 pF oxide capacitance at accumulation region is obtained from C-V curves in Fig. 4.5. Finally, using the above values, the ϵ_{NC} value obtained from Eq.13 is 10. Obviously, the permittivity of the Si NC is much larger than that of pure SiO_2 but smaller than that of bulk Si. This value can also be confirmed from theoretical estimates [51], which express the static dielectric constant of the Si NC as a function of the NC size as follows:

$$\epsilon_{static}(D) = 1 + \frac{\epsilon_{Si} - 1}{1 + \sqrt[1.37]{\frac{1.38}{D}}} \tag{15}$$

where ϵ_{Si} is the dielectric constant of bulk Si (11.9), D is the size in the unit of nm. We chose D=4 nm, ϵ_{Si}=11.9 and calculated ϵ_{static} is about 9.8.

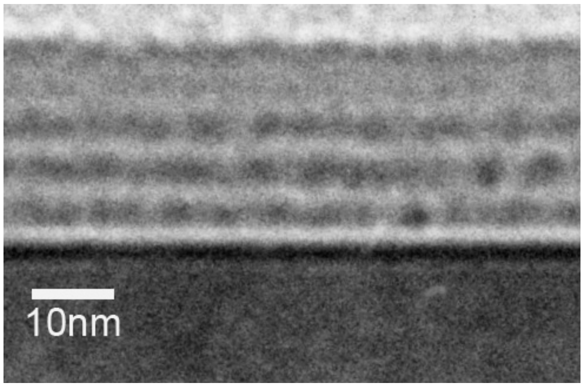


Figure 4.4: TEM image of the 3 layer sample P. In order to calculate the capacitance, this image is used to estimate the effective thicknesses

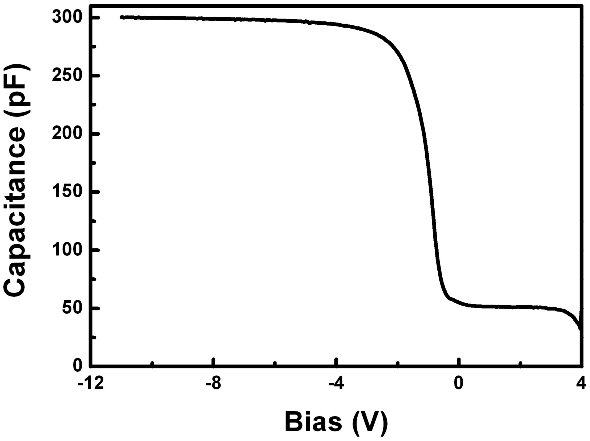


Figure 4.5: C-V curve of the sample P. The capacitance in the accumulation region is obtained from measurement

The estimation process is similar to that used in the work of Ng *et al.* [52]. Because of the advantage of phase separation, we can estimate the volume fraction of silica rather precisely and avoid the problem of inhomogeneity caused by ion implantation. The resulting Si NC dielectric constant is not only useful for fundamental physics but also useful to design the Si NC based transistor, because the gate capacitance influences the channel conductance and will determine the behavior of the transistor.

4.4 The prototype of multi-level charge memory based on the multi-layers of Si NCs

The samples B1-B3 are the prototypes showing multi-level charge behavior. For electrical characterization, a top ohmic aluminum contact electrode and a backside electrode were thermally evaporated. The backside electrode was evaporated after removing the native SiO_2 on the wafer backside, using diluted HF solution. To investigate the trapping characteristics of the sample B1-B3 we performed multiple up-down capacitance-voltage (C-V) sweeps between inversion and accumulation regions. The measurements were performed using commercially available admittance bridges, such as an HP4194A impedance analyzer at 100 kHz and room temperature.

Fig. 4.6 (a) and (b) illustrate the characteristic multilevel charge storage for our MOS structures B2 and B3 containing 2 (a) and 3 (b) layers of Si NCs. A family of C-V hysteresis was observed on the upper part of Fig. 4.6, which is caused by the successive charge trapping and de-trapping processes in the Si NC layers. As can be seen here, a higher programming bias applied at the gate results in a wider corresponding hysteresis. Hence, the narrowest C-V loop corresponds to the charging of only 1 layer and the widest one to the charging of 2 (or 3) layers corresponding to the 2 (or 3) layer sample. The conductance peak has the proportional shift under different programming bias as shown on the lower part of Fig. 4.6. (a) and (b) correspond to our samples B2 and B3. The sweep is from inversion to accumulation region for the MOS devices based on n-type substrates. For positive program bias, more and more electrons are trapped in the NCs through the thin tunnel oxide which is evident by the clockwise hysteresis. After each sweep, electrons charged in the NCs were erased by negative bias to bring the device back in the neutral stage. We calculated the flatband voltage as performed in literature [36]. The flatband capacitance can be easily calculated to be about 0.3 C_{ox} when the doping density n_i is known as $2 \times 10^{14}/cm^3$ and uniform, because $C_{FB} = \epsilon_S/\lambda$, where $\lambda = \sqrt{\frac{\epsilon_S kT}{2q^2 n_i}}$ is the Debye length. In the left C-V curves of Fig. 4.6, the left (sweep-up) part of the hysteresis curve represents the discharge stages, because of its around zero volts flat band position, and the right (sweep-down) part of the hysteresis the charge stages. It explains why choosing the n-type Si substrate in this measurement was sample. In the C-V measurement, all the bias sweep is from - to +, and is sweeped from inversion to accumulation region for the n-type substrate. Therefore carrier accumulation and tunnelling only happens at the positive sweep. Using varied programming voltage and fixed erasing voltage, the relationship between flatband voltage shift and programming voltage can be evaluated.

Fig. 4.7 demonstrates the gate voltage dependence of the memory window, i.e. the flatband shift. Here, two (three) flat stairs are observed for the 2 (3) layer sample,

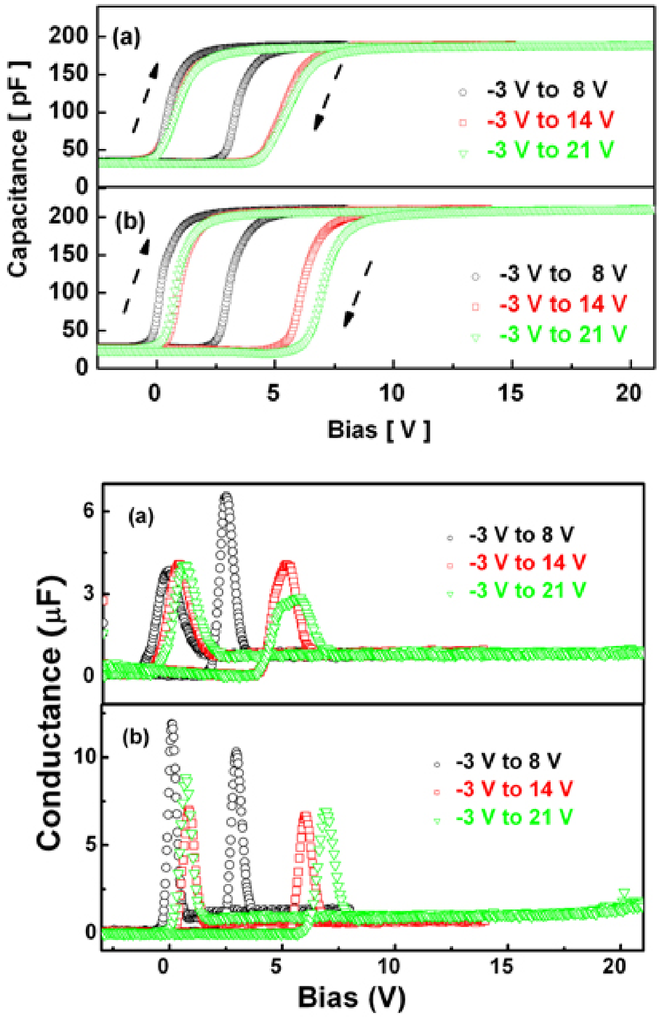


Figure 4.6: High frequency C-V (left part) and G-V (right part) characteristics of the sample (a) B2 with 2 layers, and (b) B3 with 3 layers. The widths of hysteresis and distance between conductance peaks gradually widen with increased the programming bias.

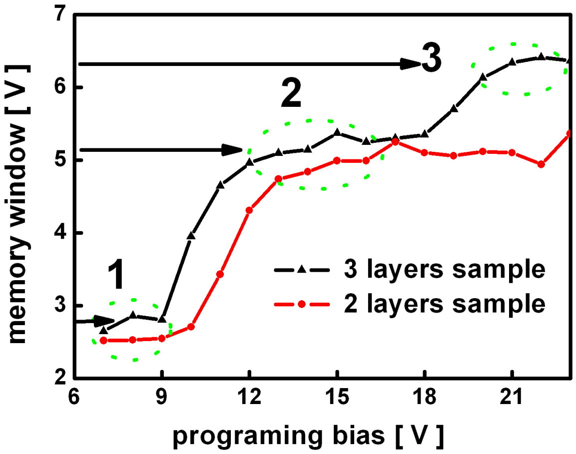


Figure 4.7: The dependence of the memory window on programming voltage. There are three apparent stages observed for the 3 layers sample B3, and two stages for the 2 layer sample B2. Each stage is spaced by about 7 V.

which are spaced by about 7 V. However, the increase of the flatband shift is less for the third step, which might be either related to a reduced number of crystals in the third layer or to a smaller size of the nanocrystals in that layer. The latter explanation seems to be more probable due to the possibility of some oxidation of the upper nanocrystals layer during the crystallization procedure.

Usually, in addition to charges trapped in these discrete Si NCs, there are other general types of charges associated with the SiO_2-Si system. There are fixed oxide charges, mobile oxide charges, oxide trapped charges, and interface trapped charges. The origin of these traps is related to the type of substrate, oxidation, such as the oxidation ambient and temperature, even on wafer orientation, or is caused by ionic impurities. Although some traps, for instance, oxide trapped charges, can be removed by low temperature annealing, most other traps still influence the behavior of the C-V curve of MOS devices. Therefore we can not obtain the perfect flat stairs. We believe that after minimizing interface trapped charges using forming gas anneal, the performance could be improved.

The remarkable good properties of the SiO_2 make it work as the gate oxide under certain medium field. However, the demands of a high working bias (>15 V) for such additional memory states in this prototype could degrade the MOS transistor. Compared with other prototypes of the multi-bit/cell designs, it has a simple structure, but needs a high working bias. The programming bias for the third level is very closed to the breakdown value of the SiO_2 dielectrics. Once breakdown happens, the charge storage ability will vanish and the whole cell will fail.

4.5 Theoretical derivation of charging equation for multi-layer Si NCs

It is well-known that, for a simple MOS structure with only one layer of Si NCs and if only one electron per NC is assumed, the flatband shift of the voltage is correlated to the NC density and the control oxide as suggested by Tiwari in eq.10. He considered the integration of the Poisson's equation in the oxide. For our multilayer Si NCs sample, its integration process is similar, but more complex than the case with only one layer Si NCs in the oxide. To get further charging information of these multilayer structure, we must reconsider the integration of Poisson's equation in the oxide. Poisson's equation leads as follows:

$$\frac{d^2\psi}{dx^2} = \frac{-q\eta_0(x)}{\epsilon_{ox}} \tag{16}$$

where x is the distance measured from the metal-SiO$_2$ interface, $\eta_0(x)$ is the volume density of oxide charge, and ψ is the band bending in the oxide. Integration of equation 16 leads to the following expression:

$$\frac{d\psi}{dx} = \frac{-q}{\epsilon_{ox}} \int_x^{x_0} dx^{'} \eta_0(x^{'}). \tag{17}$$

The integration is performed from a point x in the oxide to the silicon surface, x=x$_0$. Integrating a second time, from the gate x=0 to the silicon surface, then:

$$\psi(0) = \frac{-q}{\epsilon_{ox}} \int_0^{x_0} dx \int_x^{x_0} dx^{'} \eta_0(x^{'}). \tag{18}$$

On the other hand, at the gate, $\psi(0)$ depends on the gate bias at flat-band V$_{FB}$ and W$_{ms}$, where W$_{ms}$ is the work function difference between metal and semiconductor substrate.

$$\psi(0) = V_{FB} - W_{ms} \tag{19}$$

Mathematically, the double integral in eq.18 can be simplified to a single integral by interchanging the order of integration.

$$\int_0^{x_0} dx \int_x^{x_0} dx^{'} \eta_0(x^{'}) = \int_0^{x_0} dx^{'} \eta_0(x^{'}) \int_0^{x^{'}} dx = \int_0^{x_0} dx^{'} x^{'} \eta_0(x^{'}) \tag{20}$$

Finally, it can be written as:

$$V_{FB} - W_{ms} = \frac{-q}{\epsilon_{ox}} \int_0^{x_0} dx^{'} x^{'} \eta_0(x^{'}) \tag{21}$$

For the case of individual node charges in the oxide, the charge in the oxide causes a different V$_{FB}$ to the case without charge in the oxide. Therefore, for the same W$_{ms}$, the shift of V$_{FB}$ is

$$\Delta V_{FB} = \frac{-q}{\epsilon_{ox}} \int_0^{x_0} dx^{'} x^{'} \eta_0(x^{'}) - 0 \tag{22}$$

In our case, for the sake of simplicity, each Si NC layer is considered as one Si continuous layer, the charges are only located here where the previous SiO layers were before. Each integration step can be divided to SiO$_2$ layer segment integrations plus these Si layer segments (previous SiO layer segment) integrations, as schematically shown in Fig. 4.8. Integration of the SiO$_2$ layer segment leads to 0, because we assume no charge in these segments. Due to the mismatch of the permittivity of the Si with its adjacent SiO_2, we must choose an "effective" integration length for the Si layer which is sandwiched by SiO_2 layers. Usually the size of Si nanocrystal t_n is equal to the thickness of this layer, considering the influence of mismatch of permittivities, the 'effective' length of integration is chosen to $\frac{\epsilon_{ox}}{\epsilon_{si}}t_n$. For convenience, we convert area density d_{well} to volume charging density as following,

$$\eta_0(x) = \frac{d_{well}}{\frac{\epsilon_{ox}}{\epsilon_{si}}t_n} \tag{23}$$

so in our case, the eq.22 becomes:

$$\begin{aligned}\Delta V_{FB} &= \frac{-q}{\epsilon_{ox}}[\int_0^{t_1} x^{'}\eta_0(x^{'})dx^{'} + \int_{t_1}^{t_2} x^{'}\eta_0(x^{'})dx^{'} + \int_{t_2}^{t_3} x^{'}\eta_0(x^{'})dx^{'} .... + \int_{t_n}^{d} x^{'}\eta_0(x^{'})dx^{'}] \\ &= \frac{-q}{\epsilon_{ox}}[0 + \int_{t_1}^{t_2} x^{'}\eta_0(x^{'})dx^{'} + 0 + \int_{t_3}^{t_4} x^{'}\eta_0(x^{'})dx^{'} + ... + 0] \\ &= \frac{-qd_{well}}{\epsilon_{ox}}[\int_{t_1}^{t_1+\frac{\epsilon_{ox}}{\epsilon_{si}}t_n} \frac{x^{'}dx^{'}}{\frac{\epsilon_{ox}}{\epsilon_{si}}t_n} + \int_{t_3}^{t_3+\frac{\epsilon_{ox}}{\epsilon_{si}}t_n} \frac{x^{'}dx^{'}}{\frac{\epsilon_{ox}}{\epsilon_{si}}t_n} + ...]\end{aligned} \tag{24}$$

After replacement of the volume density with eq.23 and integration for total N Si NC layers, eq.24 becomes

$$\Delta V_{FB} = \frac{qd_{well}}{\varepsilon_{ox}}(\frac{\varepsilon_{ox}t_n N}{2\varepsilon_{si}} + t_1 + t_3 ... + t_N) \tag{25}$$

where $t_1, t_3 ... t_N$ are the upper control thicknesses for each Si NC layer as shown in Fig. 4.8. We assume that all layers have the same nanocrystal density. From this modified equation, we derive that the flatband shift of the samples with a varied number of charging layers is mainly dependent on the amount of upper control oxides for each layer plus the additional term $\frac{\epsilon_{ox}d_{well}}{2\epsilon_{si}} \times N$ as described in the above equation.

In Fig. 4.7, we compared the flatband shift of the 3 layer sample and the 2 layer sample as a function of the program bias. As mentioned above, having a sample with three layers of NCs inside, each of these layers is "seeing" a different upper control oxide thickness. That means, for the first, second and third layer the control oxide has to be chosen as 26, 19, 12 nm, respectively. Using eq.25, we estimate that the respective flatband shifts are proportional to 26.7 : 46.3 : 59.0 =1: 1.74: 2.21. This ratio is nearly the same as observed in our measurement in Fig. 4.7 where the 3 steps show ratios of 2.8 V : 5.2 V : 6.4 V=1 : 1.85 : 2.28. Measured values are always larger than the theoretical estimates, because of the contribution of other impurity charge. As can be seen in Fig. 4.7, the 2 layer sample shows a similar behavior with steps at

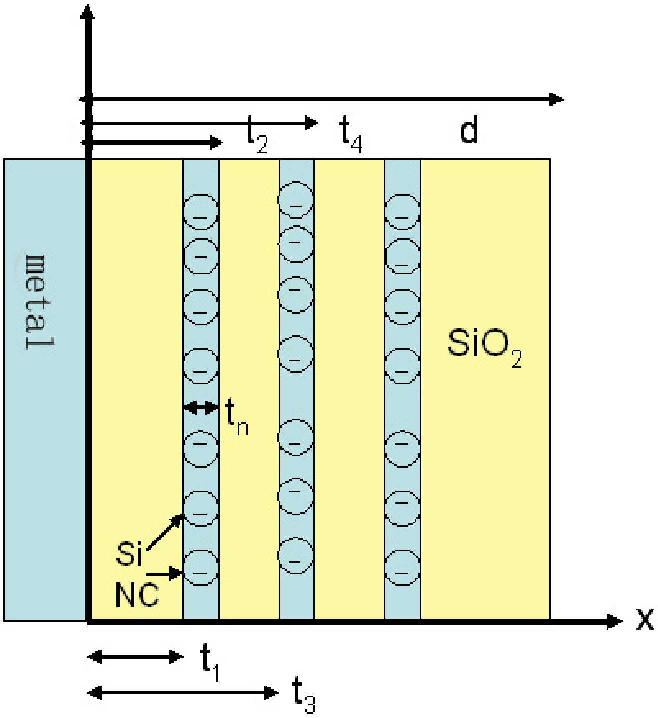


Figure 4.8: Schematic diagram of integration for multilayer charge equation

around 2.4 V and 5.2 V, however, with no third step, as one would expect. Hence, these charge steps are from the charges of the respective number of layers which are driven by the programming bias in both samples. Actually, the frequency behaviors of devices with from different number of layers in the oxide are different because of different carrier tunnel rates, thus represent a drawback of this prototype for future multi-bit memory devices.

4.6 Retention characteristics of the sample

4.6.1 Introduction of retention test

Silicon technology owes much to the remarkable properties of SiO_2 as a good gate oxide. However, the gate oxide simultaneously is also the source of several crucial degradation effects in MOS transistors. NVM cells have some important functional characteristics for practical applications, such as retention and endurance. In order to evaluate the performance of the cell, these characteristics must be tested. The primary failure mechanism of the gate oxide pertains to leakage current after high injection electric field stressing during hot-electron injection or FN tunneling, dielectric breakdown, and some modes of electrostatic discharge [53]. All these failure phenomena are intimately associated with charge transport in SiO_2 thin films. Therefore they degrade the charging capability of the SiO_2 matrix containing the Si NCs. Retention is a measurement of the time that a nonvolatile memory cell can retain the charge whether it is powered or un-powered. In floating gate memories, the stored charge can leak away from the floating gate through the gate oxide. This leakage which is caused by contamination, such as mobile ions, oxide defects or thermionic emission results in a shift of the threshold voltage of the memory cell. To improve the retention characteristics of the memory cell, various improvements to the quality of the gate oxide become very important.

4.6.2 Retention test of Si NC multilayer stack

In order to examine the reliability of our stack structure, the retention characteristics was investigated. Because this sample is not a MOSFET, we can't measure the retention window in the I_d-V_g curve, we can only measure the retention indirectly. For the sake of simplicity, we observe the capacitance change to characterize the reliability of MOS-capacitors based on the multilayer stack. First we measured the C-V curve of the uncharged sample, then continuously measure the capacitance after a fixed writing voltage and certain time, thus we get the C-t dependent, and subsequently estimate the flat band shift.

It is known by the eq.6 of the previous chapter that the total capacitance at flat band condition is composed of the silicon capacitance at flat band condition and the invariable oxide capacitance. Leakage currents loose the charge in floating gate and cause the reduced image charge to influence the depletion layer width. They further cause the deviation of the silicon capacitance at flat band condition. Therefore, we can realize an indirect retention measurement from the total measured sample capacitance as a function of time.

In Fig. 4.9, the retention measurements were performed for the three layer sample B3 for different stages. As can be seen here, the flatband voltage shift of these programming stages was stable over nearly 25 h of operation for each programming bias. Such an improved charge retention time was demonstrated as well in the doubly-stacked Si NC memory structures by Nassiopoulou [54]. Coulomb repulsion between the adjacent Si NC layers which blocks the leakage could be a reason for the improvement retention properties. The lateral diffusion of the charge, necessary to warrant

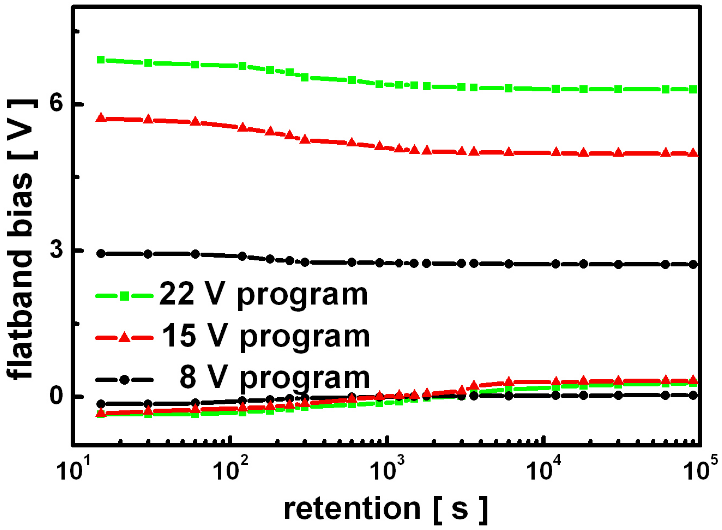


Figure 4.9: Retention characteristics of the 3 layer sample after charging at different writing biases.

for the multi-bit operation, is limited and negligible, even for our case of high density (about $10^{12}/cm^2$). If adding the influence of tuning of areal density of Si NC, which will be discussed in the next chapter, diluted Si NC distribution on such multilayer stacks could further reduce the hopping process among the Si NCs and optimize its ability of retention. The reliability is also significantly improved. Therefore shrinking these devices even further seems feasible.

4.7 Further discussion and conclusion

The estimated permittivity of Si NC from the C-V curve is less than the bulk value and demonstrates the reduction due to quantum confinement. It will influence the saturated value of the transistor memory. Si NC memory performance also depends on the shape of the NCs as well as on their crystalline orientation, because the orientation influences the carrier distribution (for example heavy and light holes) and the transmission efficiency. Even if we don't consider these variations, the feasibility of multilevel charges in layered arranged Si NCs in an MOS structure was shown. Experiments and modified charge equation demonstrate the validity of a family of C-V loops and the corresponding staircase in the flatband shift. Our results not only show a good memory capability of such multi-layer Si NC devices, but also the feasibility of memory devices having a two-bit/cell storage behavior, which means if we apply this 3 Si NC layer as a floating gate in a transistor, four distinct threshold voltage states could be easily found out, under different gate biases. We think that this can be used to design the related transistor in the future.

CHAPTER V

Tunability of the NC Density for Memory Application

5.1 The motivation of changing the NCs density

We already know that nanocrystals permit scaling of tunnel oxide and operating voltages for memory applications. However, there are some remaining device related concerns for this structure. The impact of the nanocrystal parameters such as size and density on memory bit performance has not been elucidated. Since these nanocrystals are typically formed by random nucleation and growth processes, the impact of fluctuations on memory arrays is not well understood.

When using Si nanocrystals as a charge storage device, the size, size distribution and density of the nanocrystals will certainly influence the properties of the devices. For instance, the threshold voltage shift due to electron storage in nanocrystal memory devices is proportional to the number density of nanocrystals. Smaller sized Si nanocrystals will offer a better control of the numbers of stored carriers but would also result in an increased spacing energy and Coulomb charging energy, as was demonstrated [55]. The spacing of the nanocrystals should be large enough to prevent a significant dot to dot hopping at the working temperature of the device. Different methods for preparing Si nanocrystals were reported, such as using CVD, the implantation of silicon into SiO_2, or the preparation of silicon rich silica. The reported sizes are in the range of 1.7-6 nm with a normally rather broad size distribution of some nanometers. The control of the size and size distribution in single and multilayers was demonstrated previously by using our approach. An independent control of the interdot spacing was not demonstrated so far. Also, a controlled variation of the nanocrystal area density by maintaining the same size and a small size distribution was not possible up to now. The simplest way for size and size distribution is to prepare a layered arrangement of Si nanocrystals as can be accomplished by the deposition of SiO_x/SiO_2 multilayers followed by a high temperature annealing.

The feasibility of multilevel charges in layered arranged Si nanocrystals in an MOS structure was demonstrated in the previous chapter. With our method we have a very simple way to control the area density within the layer by maintaining nearly equally sized nanocrystals. That is with a variation of the stoichiometry parameter x from 0.9 to 1.63. The stoichiometry of the SiO_x layers is controlled by adjusting the oxygen pressure during the growth which influences the resulting area density of the Si nanocrystals. The method gives an unique possibilities to study the influence of the nanocrystals density (without having an additional influence on the NC size) on the electrical properties. The properties of the resulting MOS devices as a function of the nanocrystal area density will be discussed in this chapter.

5.2 Experimental procedure

The deposition of the SiO_x/SiO_2 were done in the same way, but using a variable stoichiometry (i.e variable oxygen pressure as described in chapter 2). The variation of the stoichiometry corresponds to the x value in SiO_x from 0.9 to 1.63 for samples C1 to C4. After deposition, the samples were annealed in two-steps sequently under N_2 atmosphere in a quartz tube furnace at 900 ^oC for 0.5 h to form amorphous Si clusters firstly, then the temperature was raised to 1100 ^oC for 0.5 h to form the Si nanocrystals in the SiO_2 matrix. The samples were investigated by all measurements, including PL, C-V and cross-section TEM. A HeCd laser was focused on a sample area of about 1 mm^2 as an excitation source, and a power density of 1.5 W/cm^2 without any filter. Photoluminescence measurements were performed using an Acton Research 500L monochrometer with an attached liquid nitrogen cooled CCD camera. The slit of the monochrometer was 200 μm and accumulation time of CCD was 0.2 s. The spectra were corrected for the spectral response of the measurement system. For electrical characterization an aluminum top gate electrode and an Ohmic contact as backside electrode were evaporated after removing the native SiO_2 on the wafer backside by diluted HF. To investigate the trapping characteristics of the structure we performed again multiple up-down C-V sweeps between the inversion and the accumulation region of these MOS structures. The measurements were performed at 100 kHz using an HP4194A semiconductor impedance analyzer at room temperature. The layer structures of the annealed samples C1 to C4 were checked by TEM in cross section geometry.

5.2.1 Optical characterizations of diluted Si NC layers

Silicon is a excellent material for electronics, however, its intrinsic indirect band gap makes Si an inefficient material for light emission. Because of quantum confinement effect of Si nanostructure and the elimination of bulk or surface defects, nanometer-sized Si objects (nanocrystals or nano-dots) can emit light at room temperature from the near infrared throughout the visible [12]. The normalized photoluminescence spectra of the sample series are presented in Fig. 5.1(a). We choose 800 nm as a central wavelength and the width of the spectra is about 350 nm. The photoluminescence (PL) peak positions of the four samples are nearly the same, around 770 nm but the intensity gradually decreases from sample C1 to C4. In these spectra, the nanocrystal density ratio among the samples might be extracted from the integrated PL intensity under the assumption that the samples are excited by the same power density, using an identical sample area, and samples with a nearly identical diameter of the nanocrystals. As can be seen here, the nanocrystal density gradually decrease from sample C1 to C4 with increasing O_2 pressure. The nearly constant PL peak position observed for the samples prepared with the same layer thickness but variable oxygen content is a clear hint that the average NC size is in the same range within the limit of our preparation. Under the assumption of equally sized NCs, the radiative life time τ_{rad} and the absorption cross-section σ of such NCs are the same. Using the same area flux density of exciting photons I_{ex}, the PL intensity I_{PL} depends mainly on the total number of excited NCs N according to the equation suggested by Kovalev

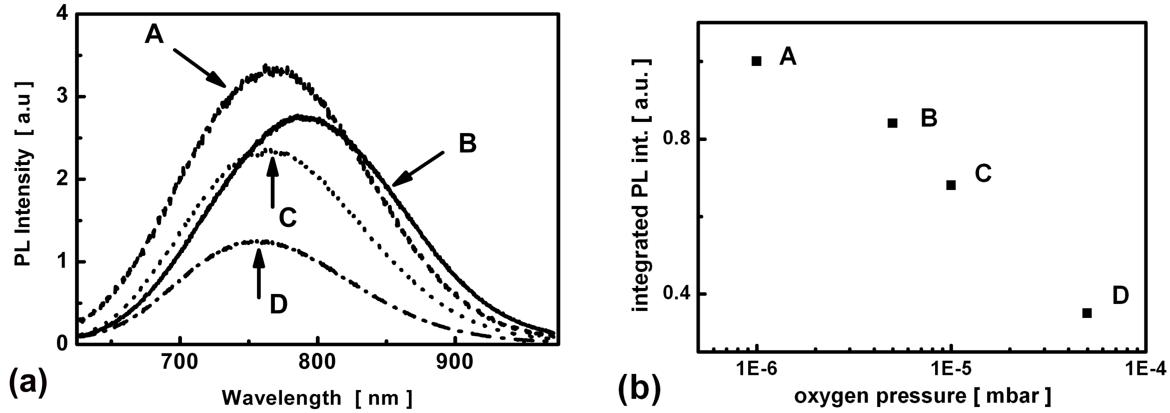


Figure 5.1: (a) Photoluminescence (PL) spectra of our four samples. A, B, C and D represent sample C1, C2, C3 and C4, respectively. (b) PL intensity as function of oxygen pressure.

Table 4: Norminal growth oxygen pressure and parameters extracted from photoluminescence and C-V experiment

sample	O_2pressure	stoichiometry	norm.inte.	flatband shift			average NC	
	$\times10^{-6}$	of the SiO_x	PL int	for bias of			density	
no.	(mBar)	x	A.U	7V	14V	21V	$10^{12}cm^{-2}$	unity
ratio								
C1	1	0.9	1	3.15	5.37	6.46	2.79	1
C2	5	1.13	0.84	2.78	4.46	5.2	2.25	0.81
C3	10	1.27	0.68	2.22	3.29	4.00	1.73	0.62
C4	50	1.63	0.35	0.98	1.67	2.16	0.93	0.33

et al. [56]:

$$I_{PL} \sim \frac{N\sigma(E_{ex} - E_{gap})I_{ex}}{1 + \sigma I_{ex}(E_{ex} - E_{gap})\tau_{rad}(gap)} \tag{26}$$

where E_{ex}, E_{gap} are the excitation and band gap energy, respectively, N is the NC density and will be estimated from this equation, if the same excitation I_{ex} is used. In this PL measurement, the measured spectra is spread from 627 to 970 nm with a range of about 340 nm. In order to count all contributing Si NCs, we have to integrate the intensity within this certain wavelength range, hence, the ratio of the integrated PL intensities of our fours samples scales directly to the NC densities of the samples which is $1 : 0.84 : 0.68 : 0.35$ (normalized to the sample with SiO layers), respectively. The results are summarized in table 4.

5.2.2 Electrical characterizations of diluted Si NC layers

We selected sample C1 and C4 for demonstration of the typical influence of the NC density on the charge behavior in Fig. 5.2. A family of C-V hysteresis curves was observed which are presented here for typical programming biases. The C-V hysteresis is caused by the trapping, storing, and de-trapping of carriers to and from the silicon NC. Applying higher bias at the gate results in a wider corresponding hysteresis because of the sequential charging of the three NC layers. The narrowest C-V loop corresponds to the charging of the first layer only, the middle one to the charging of the first and second layers and the widest one to the charging of all three layers. Comparing the behavior of samples C1 and C4 having different densities of NC we observe a continuous decrease of the flat band shift (i.e. the width of C-V loop) with decreasing the NC density in the layers under the same programming bias as is shown in left part of Fig. 5.2 a and b. The NC density can be estimated using the flat band shift and the respective charge equation developed in last chapter. We also observed that the memory windows as function of programming bias on the right part of Fig. 5.2 show 3 stairs due to sequential charge of 3 Si NC layers.

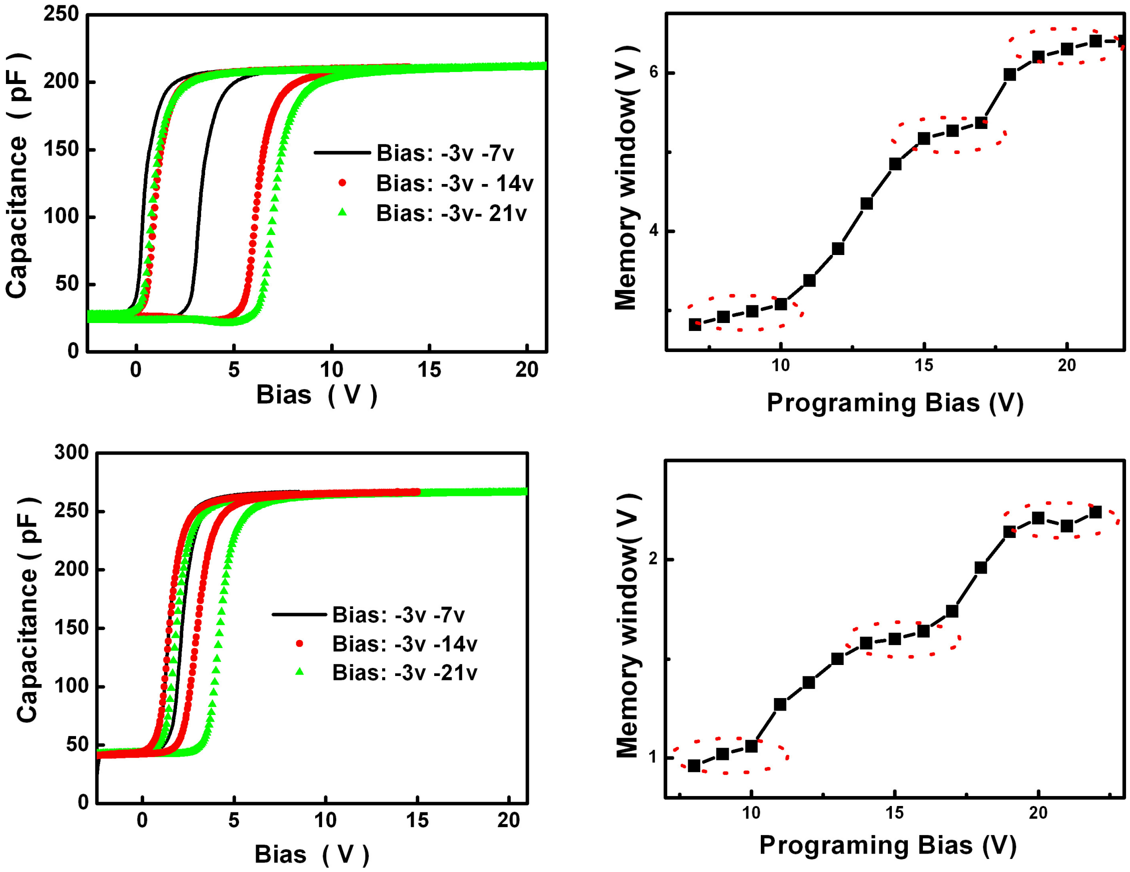


Figure 5.2: Upper: Typical high frequency C-V curves of sample C1 (left) and their memory window as function of programming bias (right). Lower: The typical high frequency C-V curves of sample C4 (left) and their memory window as function of programming bias (right). The widths of the loops are proportional to the charging density of Si NCs.

As a result, one of the essential growth parameter, the oxygen pressure, can be correlated to the resulting area nanocrystal density as summarized in table 4 and graphically presented in Fig. 5.3. The PL intensity and the flatband shift in C-V are both directly depending on the NC density, hence, both measurements represent an independent way for an estimate of the NC density. As can be seen in Fig. 5.4, the NC density estimated from C-V measurements scales linearly with the integrated PL intensity, the implications of this relation will be discussed below. Error bars in the figure indicate that due to the unavoidable deviation in estimation of average Si NCs density, these data are only valid within a range and have uncertainty. Comparing now the samples having an increased NC density, we observe an increased flatband shift. For each sample, we sequentially charge the layers by increasing the programming bias. So comparing the samples, we find that the memory window linearly scales with the PL intensity and hence the NC density.

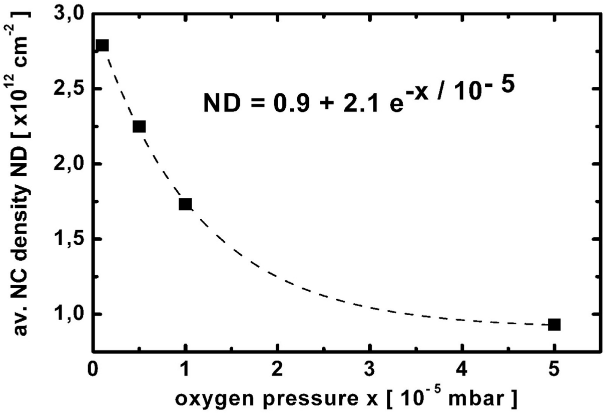


Figure 5.3: The dependence of the calculated Si NC density ND on the oxygen pressure shows an approximately exponential behavior.

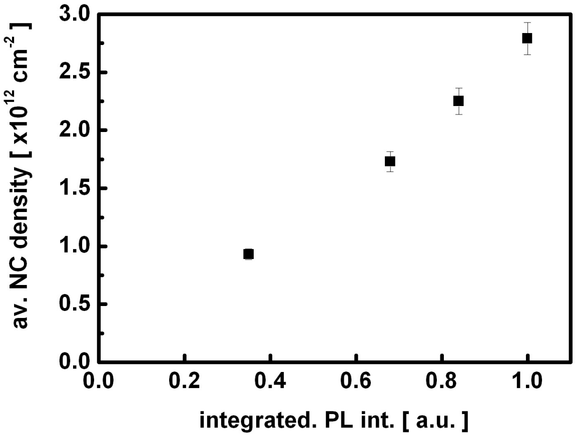


Figure 5.4: The relationship between calculated Si NCs density from memory window and the integrated photoluminescence intensity.

5.3 Discussion of the chapter

5.3.1 Improvement of annealing process

One crucial step in sample preparation is the crystal formation because only samples with well established and isolated nanocrystals show charge storage behavior. High temperature annealing is an important step for the crystal formation, the thermal diffusion of Si atoms make them gradually nuclei. For our samples, the highly oxygen composition induce the Si clusters to be separated further more each other than in our previous samples. Therefore we must optimize the annealing process for these samples. We tested pieces of the same samples after an 900 ^{o}C anneal. No charge storage and no luminescence were observed. As suggested by Yi *et al.* [57] in her three stage model of the phase separation, annealing at 900 ^{o}C results in a complete phase separation of the former SiO layers but with amorphous Si nanocluster surrounded by amorphous SiO_2. Above 900 ℃, the crystallization of amorphous clusters into Si nanocrystals takes place. In our samples, the additional oxygen during the growth results in an increased SiO_2 and a decreased Si percentage. Hence, the additional SiO_2 forces the Si clusters to be more separated from each other within the layers compared to the samples previously produced in our group. A decrease in size can be excluded because the deviation in the PL peak position observed for the samples is in the order of the fabrication tolerance. In contrast to the annealing of 1 h at 1100 ^{o}C normally used for multilayers, we chose here a variation of the annealing process to reduce the thermal budget and the danger of destroying the layers by oxidation. We annealed the samples for 30 min at 900^{o}C to form the amorphous Si clusters. At this temperature, oxidation of the clusters by diffusion of residual oxygen into the films is drastically reduced and hence the delicate layers are conserved. An additional annealing at 1100^{o}C for 30 min then transformed the amorphous Si nuclei into nanocrystals.

5.3.2 Determination of density ratios

The change in the average density of the crystals is hard to prove by TEM cross sectional investigation because each of the images still represents an overlay of a different number of NC rows. The final thinning of the TEM samples is different in each case and the image conditions also depend on the focused spot under each investigation. Using electrical characterization, we can obtain the Si NC area density ratios as we did in the photoluminescence measurement.

The NC density can be obtained from C-V measurements. From our TEM investigations, we found that the total thickness of sample C1 is slightly larger than in the other samples. Therefore in the C-V curves, its capacitance in accumulation C_{oxide} is smaller than that of the other samples, which is consistent with the expression given in literature [58]. It is known from chapter 2 that the various volume fractions of Si

NCs in the oxide matrix from sample C1 to C4 slightly affects the permittivities of their oxide layers. Thus this influence in a change of capacitances in accumulation C_{oxide} for these four samples. The upper oxide thicknesses and size of the dots for each layer were measured from the TEM images. According to the suggested fixed oxide charge model, the theoretical value of the flat-band shift of multilayer charges can be estimated as eq.25 in the previous chapter. For simplicity, we assume that within one sample all layers have the same nanocrystal density. Assuming now only one trapped carrier per NC, the values of flatband shifts of the sequently charged layers are obtained from the C-V curves. We calculated the average NCs densities assuming a charging of all three layers. The results are also presented in table 4. As can be seen, the calculated nanocrystal density shows a similar ratio dependence as the integrated (normalized) PL intensity.

The above Fig. 5.3 shows the Si NC density calculated from the C-V measurements as a function of the oxygen pressure. From the fitting of the curve we obtained the expression: $D = 0.9 + 2.1exp(-P/10^{-5})$, where D is the nanocrystals area density (unit of $10^{12}cm^{-2}$) and P is the oxygen pressure (unit of mbar). The empirical formula can be used for a rough prediction of the nanocrystal area density within a certain O_2 pressure range. The upper limit of the area density under the condition of using pure SiO powder, an ultimate small base pressure of zero mbar and a layer thickness of around 4 nm is around 3 x $10^{12}/cm^2$. The lower limit, however, which would be a density of 0.9 x $10^{12}/cm^2$ for a very high oxygen pressure is not correct, because according to the Fig. 2.1, for a pressure of larger than 1x10^{-4} mbar the resulting layer is already consisting of SiO_2 which does not contain any nanocrystals after annealing. A further larger density could be possible if the size of the nanocrystals is decreased, which could be realized with thinner SiO_x layers. Based on the estimated NC density and under the assumption of equally sized spherical nanocrystals, we can estimate the average distance between the neighboring nanocrystals by using a simple geometric model of putting equally sized spheres with a certain density on a plane in a regular way which was already shown as Fig. 2.7 in chapter 2. Then the medium distance between the nano-spheres can be estimated which tunes from 2.2 nm to 6.7 nm for changing the stoichiometry from around 0.9 to 1.63. Hence, we can systematically adjust the separation of the NCs into the range needed for avoiding in-plane carrier leakage between adjacent nanocrystals. The nanocrystal density calculated from the flatband behavior is linearly related to the observed PL intensity relation (see Fig. 5.4) as one would expect. Both properties can equally be used for device characterization.

In Fig. 5.5, the relationship between the memory window under different biases and the integrated photoluminescence intensity is shown. Error bar indicates that for each programming bias, the measured value of memory window is variable within certain range, but tolerant. The observed linear trend under variable programming biases is in agreement with our discussion above. All the linear fits overlap at the zero point, which means zero nanocrystal density results in zero PL intensity, as one would expect. In addition, with a proper desired nanocrystal density the memory window of each of the layers can be designed by a properly chosen oxygen content. This offers new possibilities for device design and optimization, specially enough isolating inter-distance between dots is helpful to prevent hopping for long retention. A highly density of dots contributes to the threshold value shift to cause more pronounced memory state. This approach is universal not only for group IV nanocrystals, but might also be used for other materials which are derived from the decomposition of their oxides, such as Pt nanocrystal from the PtO_x decomposition in an insulating matrix [59]. In addition, this approach might allow to design single electron transistors with only one nanocrystal in the channel.

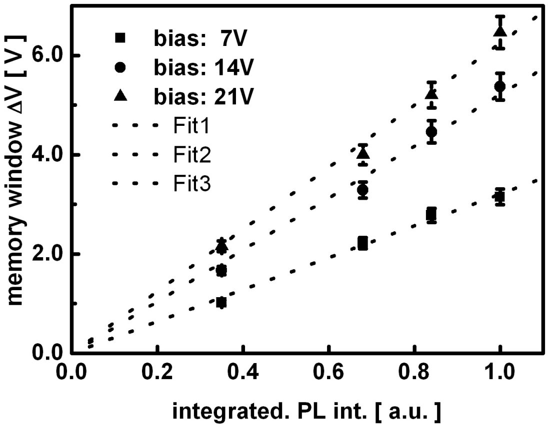


Figure 5.5: Relationship between the memory windows under different biases and the integrated photoluminescence intensity.

Tuning of the NC density in plane not only benefits memory operation, but also influences the lateral carrier transport of the sample. Lateral conductivity generally increases as Si concentration increases, probably because of an increase in carrier transport as the average distance between neighboring Si NCs decreases. It benefits for the lateral hopping process of Si NCs in plane and could be used to optimize the current output for third generation of tandem solar cell [60].

5.4 Conclusion of the chapter

A simple approach to control the area density of the Si nanocrystals within each layer was demonstrated by varying the stoichiometry of the SiO_x layers. The varied stoichiometry results in a varied area density of the nanocrystals after high temperature annealing of the multilayer stacks as was demonstrated by TEM, photoluminescence and by the results of the electrical characterization. Based on the results reported here, an adaptive design of Si NC based memories can be accomplished with an independent variation of the nanocrystal density from 0.9×10^{12} up to $2.8 \times 10^{12}/cm^2$ within each of the layers, which represents an average distance between the NCs from 2.2 nm up to 6.7 nm. This approach allow to limit the in plane hopping process among the Si NCs which will degrade the reliability of devices by tuning the spacing between NCs. Since the nanodots embedded in the dielectric matrix will enhance the conductivity of the dielectrics remarkably, therefore this system may be developed to a conductivity-tunable building block in future applications.

CHAPTER VI

The Silicon NC trap center studied by DLTS

6.1 Introduction

Memory-cell structures employing discrete traps as the charge storage media have received much attention as promising candidates to replace conventional polycrystalline silicon or silicon nitride non-volatile memories for future high capacity and low power consuming memory devices. NC memory devices employing distributed nano-dots as storage elements have shown great potential in device applications [3]. A floating gate composed of individual nanocrystals reduces the problem of charge loss encountered in conventional commercial floating-gate, such as short retention, endurance and SILC (Stress Induced Leakage current), and allow further scaling down of tunnel oxides, thereby lowering working bias, and faster write/erase speeds.

The real origin of the trap/detrap phenomena in the silicon nano-floating gates is still under debate, even including the generation of the luminescence of the nanocrystalline silicon. Several models of charge storage were proposed such as three-dimensional quantum confinement effects [61], interface and defects states [62]. The existence of defects due to lattice damage of ion implantation and hydrogen effects from CVD process [63] in the NC-Si based MOS structure might generate parasitic traps and influence the charge performance. Among these methods, considering the good adhesion of SiO with SiO_2 and simplicity of the sources, the SiO/SiO_2 superlattice method shows robust charging capability due to its low interface and defect density.

6.2 Transient capacitance spectroscopy

In contrast to the steady-state high frequency C-V methods described so far, transient capacitance spectroscopy gives information by measuring how the nonsteady-state high-frequency capacitance changes with time t. Deep Level Transient Spectroscopy (DLTS) invented by Lang [64] is now widely used to detect traps of so-called ”deep levels” in the band gap. Initially, the method utilized measurements of transient capacitance following the pulsed bias in a p-n junction or Schottky barrier diode to monitor changes in the charge state of defect centres. Schulz and Johnson [65] extended applying DLTS to study the charge emission from interface states in MOS structures. Therefore, for Si NCs embedded MOS structure, if the charge emission from interface states of Si NC is detectable, it is worthwhile to use DLTS for characterization. The DLTS signal is the difference of capacitances at two different times after a filling pulse. It shows peaks for different trap levels in the sample at the respective temperatures T. If traps are filled by a filling pulse and the reverse bias is switched on again, the sample is in thermal non-equilibrium and relaxes into equilibrium by detrapping the charges back. This relaxation process is related with a capacitance transient. Its time constant is governed by the thermal emission rate e$_{n;p}$, which depends on the trap energy E$_t$ and the temperature T:

$$e_{n,p} = \frac{1}{\tau_e} = \frac{N_{n,p}\sigma_{n,p}v_{n,p}}{g} exp(\frac{-E_t}{kT}) \tag{27}$$

Here N$_{n,p}$ is the effective density of states in the conduction band or valence band, $\sigma_{n;p}$ is the capture cross section of the trap, $v_{n,p}$ is the carrier velocity, g is the degeneracy factor. We can see here, the emission rate is exponential dependent on 1/T, so the thermal energy (or activation energy E$_t$) determining its slope and the capture cross section σ determining its intercept, are the main features of the curve. Because of

$$DLTS = a(C(t_2) - C(t_1)) \tag{28}$$

the DLTS signal is scaled in units of capacitance (usually pF). In principle, a DLTS measurement starts at a low temperature. The signal is recorded, and temperature is ramped up for measurement. During the raising of T, according to $exp(\frac{-E_t}{kT})$ dependence of the thermal emission rate, as long as T is too low for significant thermal emission until t$_2$, the difference in eq.28 is zero. If T is so high that the thermal emission is already over at t$_1$, the difference eq.28 is also zero. Only if the emission time constant (or its inverse, the emission rate for convenience) of one level falls into the so-called "rate window" given by the definition of t$_1$ and t$_2$, a DLTS peak appears. We obtain the following condition for the DLTS peak to appear:

$$e_{n,p} = \frac{\ln \frac{t_2}{t_1}}{t_2 - t_1} \tag{29}$$

The "rate window", is the inverse of τ_e, according the formula of the relaxation time constant τ_e, and has the unit of s^{-1}. Because on eq.27, for each trap, the trap energy and the capture cross section are determined by its slope and intercept, respectively, may be determined from an Arrhenius plot $\ln(e_{n,p}) \sim 1/T$.

Few researchers have used DLTS to observe the Si NCs charging process [63, 66], however. Because of the drawback of their fabrications methods, only limited information about Si nano-dots were obtained and were usually hidden by artificial parasitic traps. Here we present our DLTS measurements of Si NCs samples based on the SiO/SiO_2 superlattice. Several results are explained and are consistent with previous reports. These results might help us to understand the trap process of the Si nano-floating gates and optimize the future operation of non-volatile memory devices based on Si nanocrystals. The charge-pumping method is a kind of nonsteady-state measurement, but it is widely used to evaluate the interface states in MOS transistors which has a more complex 4 terminals structure, so we don't concentrate on this method in this chapter.

6.3 DLTS experimental details

The non-symmetrical sandwich structure samples were prepared as usually on highly doped n-type (100) silicon substrates (0.05-0.1Ω cm). First, a 4 nm SiO_2 film was deposited as a tunnel oxide, then a 4 nm layer of SiO was deposited by SiO powder evaporation. On top of this structure, an additional SiO_2 layer was evaporated as the upper control oxide with the layer thickness of 24 nm. In order to form the Si NCs, thermal annealing was performed in a quartz tube furnace under N_2 ambient (1100 ^oC, 0.5 h) for phase separation and crystallization. For comparison, a control sample, i.e. a pure SiO_2 film with the same total dioxide thickness as the sample with Si nanocrystals, was prepared on the same type of substrate using the same deposition conditions. The electrical measurements were performed at 100 kHz using an HP4194A impedance analyzer at room temperature, as we did before. For more information about the traps of the samples, DLTS were performed in the 100 - 310 K temperature range with variable pulse bias and rate windows. The heating rate was 0.4 K/min. The cryostat containing the sample was attached to the capacitance-meter with preamplifier, and the amplified transient capacitance was processed by the electronics and a PC. The capacitance of the sample is measured under depletion region conditions. The "rate windows" were chosen as 34.7, 17.3, and 8.7 s^{-1} for the measurements.

6.4 C-V and DLTS analysis

Fig. 6.1 presents typical C-V curves of the sample. The loop of the forward and reverse sweep C-V characteristics indicates the electron charging and discharging effects of Si NCs embedded in SiO$_2$. The details of charging behavior of Si nanocrystals were discussed previously. The DLTS study gives more detailed information, in addition to C-V characterization in Fig. 6.1. The process of measurement is the following: electron injection to dots appears by positive reverse bias on the gate, whereby, near the interface of Si-SiO$_2$, the potential slope is so steep that electrons are populated and confined by oxide and the steep potential. The electron energy perpendicular to the interface is now quantized into discrete states, this phenomenon is called surface quantization and it helps the electron injection [67]. Electron escape emission is from the interface or quantum levels of the dots by repeating bias filling pulses. The emission of electrons is recorded by measuring the capacitance transient. Fig. 6.2 shows the DLTS signal of the control SiO$_2$ sample in (a) and a sample with Si nanocrystals in (b). Negative peaks mean that the electrons (majority carriers) are in the space charge region. In Fig. 6.2a, only one peak around 190 K is obtained in the spectra, but the spectra in Fig. 6.2b present two peaks around 190 K and 250 K, respectively. The signal noise in its left part is due to problems associated with the metal contact

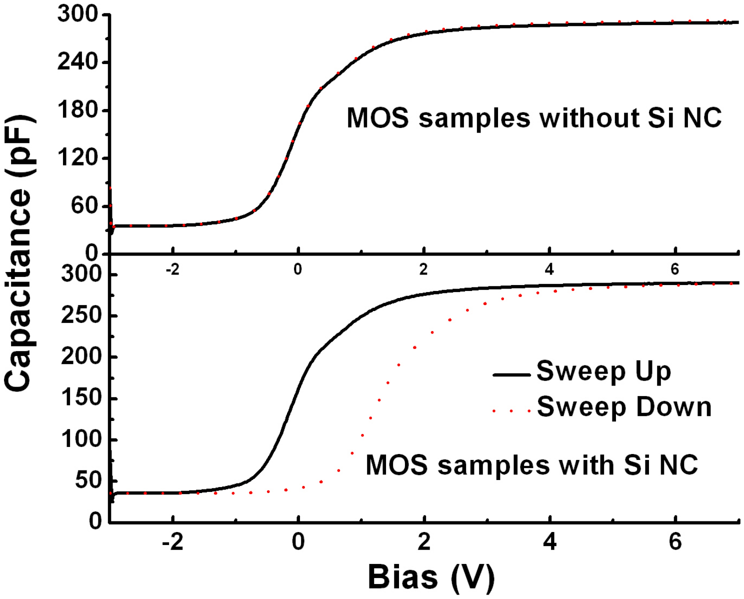


Figure 6.1: C-V curve of the control sample and sample D1

at low temperatures. For the simple MOS control sample, the signal consists of the interface charge states between the oxide and Si substrate, and for the MOS sample with Si NC, except the above peak, a second and dominant peak appears. It should be related to the charge within the Si NCs. Furthermore, to distinguish whether the charges are stored in the quantum confined conduction band or in deep levels of the band gap in the Si NCs, we tuned the rate window of DLTS to obtain the thermal activation energy and capture cross sections to explain this peak. In Fig. 6.2, we measure the DLTS spectra under three rate window 34.7, 17.3, and 8.7 s^{-1} by choosing certain t_2 and t_1, and corresponding peak shifts are shown. This emission rate shift as function of temperature will be presented in the Arrhenius plot in Fig. 6.4.

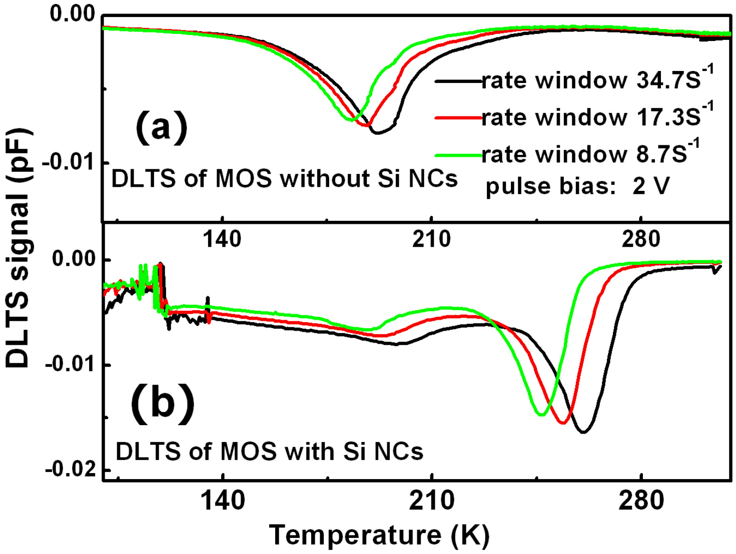


Figure 6.2: DLTS spectra of control sample (a) and samples D1 (b) dependent on rate window: samples have same thickness, but upper one is MOS control sample without Si NCs, lower one is MOS sample with Si NCs

Yamasaki theoretically analyzed and measured the bulk traps and interface states in Si MOS diodes [68] and demonstrated the distinction between interface charge and bulk traps. Because the emission rate of bulk traps at a temperature is constant regardless of the pulse bias, the peak temperature and shape of $\triangle(C)$ should not change with pulse bias. We keep the same rate window, constant reverse bias 1 V, but variable pulse bias 0.5, 1, and 2 V on the samples. The obtained spectra are presented in Fig. 6.3. The same peak temperature at around 250 K and shape

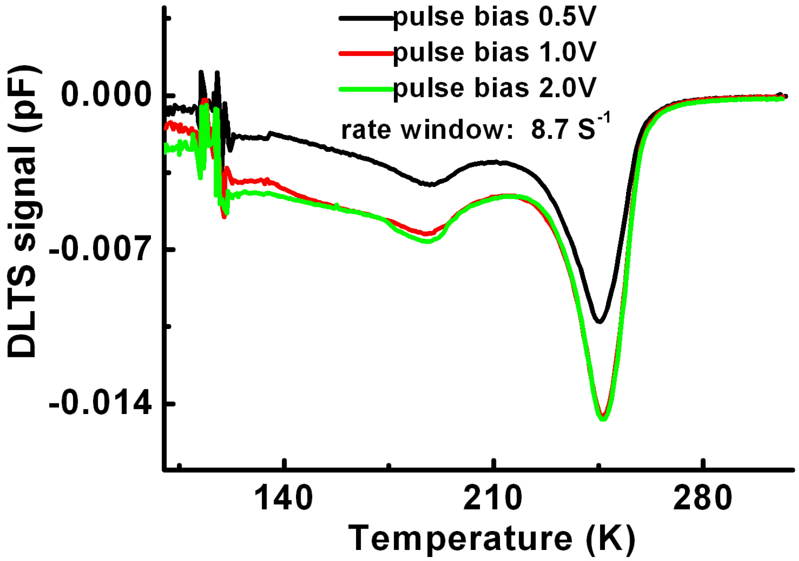


Figure 6.3: DLTS spectra of MOS sample containing Si NCs dependent on fitting pulse biases: as the pulse bias rises, the transient signal rise 'too' until saturation.

convince us that this trap is probably a bulk trap, unrelated to the interface states. Without applying a pulse bias, the MOS sample is in the accumulation region and electrons are injected from the substrate and stored in the Si NCs. It is well-known that these stored electrons alter the electrostatic potential and cause the positive flat-band shift. Therefore, the sample actually is in the inversion region under the pulse bias. Electron emission from Si NCs occurs and contributes to transient capacitance. Similar phenomena were observed by Souifi *et al.* [66] in their CVD samples, However, the results from Souifi show other noisy peaks in the DLTS spectra, which means there exist trap sites which are different from the Si quantum dots and interface in their oxide, such as hydrogen-related traps in CVD process or defects. Their sample and only a 2 nm tunnel oxide, whereas my sample had a 4 nm tunnel oxide. We believe it is because of high density ($>10^{12}$cm^{-2}) of Si NCs in our sample, since the tunnel probability is proportional to the total electron emission number. Normally, because of the Si-O bond breaking in the phase separation process, the sample used here might contain some dissociating dangling bond in the matrix as a deep level, but sufficient heat treatment nucleate them for the crystallization of Si dots, so their amount appears to be below the detection limit of DLTS.

In the case of pulse bias 1 V and 2 V, the DLTS signals at the peak around

250 K are nearly the same, because the stored electrons in Si NCs are fully released under these pulse biases and contribute to the same transient capacitances. From the shape of the DLTS peaks, we consider these deep traps as consistent with the expected discrete behavior of the nanocrystal. We observe the shift of DLTS peak related to Si NCs for different "rate windows". The different DLTS peaks have their own corresponding temperature T, so the Arrhenius plot is used to determine the parameters of the trap center in Fig. 6.4. We can find the value of carrier velocity $v_{n,p} = 10^7 cm/s$, and $N_c = 8.4 \times 10^{18} cm^{-3}$ from the literature and use them in eq.27. For measuring E_f and σ, we fitted the Arrhenius plot in Fig. 6.4 according to eq.27. The calculated activation energy E_t of the main peak is about 0.56 eV. Using this value and the above value of m^*, $v_{n,p}$ and N_c in eq.27, the capture cross section is found to be about $1 - 7\times\ 10^{-13}$ cm^2.

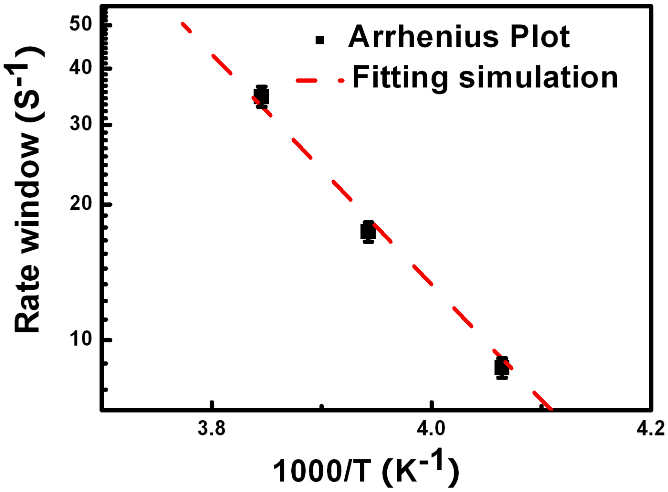


Figure 6.4: The Arrhenius Plot of the NC sample D1. Together with parameter simulation.

These estimations are much larger than the counterparts of trap impurities associated with the deep levels of silicon band gap, but are adjacent to the values of Si NCs in the SiO_2 matrix, such as the obtained potential barrier of Si NCs by our method to SiO_2 insulator is 1.6 eV in previous I-V characterization. Biteen [69] and Kovalev [70]showed the large cross section of silicon NCs is at the $10^{-13}cm^2$ level and much larger than the cross section of usual deep level trap centers. Silva confirmed the large capture cross section of Si NCs [71] in a dielectrics matrix as well. In contrast, interfacial traps or single dangling bands are mainly located at the gap center with a capture cross section of 10^{-15} cm^2-10^{-17} cm^2 [72]. This was demonstrated by the conductance method measurements in previous chapter. The capture cross section is an important parameter for memory operation, a reduced capture cross section and Coulomb blockade will result in lower programming speed and saturation V_T. The larger capture cross sections of Si NCs not only benefit NVM operation, but also play a key role as a sensitizer in the Er doped silicon dioxide containing Si NC system already discussed in chapter 1,

6.5 Further discussion and conclusion

Using DLTS, we were able to observe thermal emission from Si NC around 250 K for certain rate windows. From our estimates for some intrinsic parameters of the DLTS curve, such as capture cross section σ at 10^{-13} cm^2 level and activation energy E_T about 0.6 eV, we concluded that these data are consistent with the properties of Si NCs which were measured by optical methods, such as photoluminescence Auger saturation [70]. These results could be a clue that the trapping mechanism in MOS systems containing Si NCs is related to the quantum levels of the Si quantum dots at around 300 K. DLTS supplied information other than obtained as C-V measurement on other electrical characterizations.

CHAPTER VII

Conclusions and Outlook

7.1 Conclusions

The advantages of Si NC based NVM are low cost and compatibility with current CMOS technology. The driving force for flash memory scaling is cost reduction. In a flash memory chip, both the core memory cell array and the peripheral circuitry need to be scaled.

Although Si nanocrystals in SiO_2 represent a promising material system which has been studied over last two decades, tuning the properties of Si nanocrystals for device applications still needed to be addressed. The fabrication of appropriate samples for different purposes were introduced in chapter 2. TEM characterizations for these samples were also included in this chapter. The SiO/SiO_2 superlattice method has the advantages of a controled NCs size, density and location. Thus, a much better performance in floating gate MOSFET devices can be expected. A nearly fully charged single Si NCs layer embedded in the insulating matrix under medium bias was realized by a uniform distribution of Si NCs and will benefit the stability of threshold values in memory operation. The full charges of Si NCs can not be accomplished by other methods, such as implantation, sputtering. The feasibility of the sequential layer by layer charging in our multilayer samples was shown. The mass point of this sequential charge was recognized in its memory window dependence of the programming bias as stable stairs. A comparison of the ratio of voltage stairs in memory windows from experimental results and theoretical charge model revealed that the clear stairs were caused by successive charging of a varied number of Si NC layers in the floating gate. Retention tests of all three programming states show their stability in retention process, and multi-layer structures could be designed as a prototype of multi-bit/cell MOS transistors. The stoichiometry of the SiO_x layers was controlled by adjusting the oxygen pressure, which influenced the resulting area density of the Si NC. The tuning of the Si nanocrystal area density in the layers was demonstrated by TEM as well as comparison of capacitance-voltage and photoluminescence measurements. The influence of the nanocrystal density on the charge behavior was discussed. The SiO_x/SiO_2 method realizes a simple but effective way to control the area density by maintaining other parameters of NCs and gives an unique possibility to study the influence of the nanocrystal density on the electrical properties as well as optical properties. The trap center study of Si NCs based on the DLTS treated the individual Si nano-dots as a single point defect deep level defects in the oxide and revealed essences of Si NC, such as a large capture cross section and potential barrier. The properties we observed are consistent with those obtaining by optical methods and helpful to optimize the performance of real Si NC device. These properties were also carried out with electrical characterizations in previous chapters.

7.2 Outlook

7.2.1 Optimized scaling down for Si NC device

In order to integrate Si NCs into existing functional devices as components for future devices, a better control of the quality of this MOS stack containing Si NCs is needed. Especially the following tasks have to be investigated:

- Thermal oxide has been used as a tunnel dielectric for a long time. It is a high quality material with very small defect density. However, The high barrier of the tunnel oxide results in small hot electron injection efficiency or FN tunneling injection efficiency. Its operation voltage cannot be scaled if a certain programming speed is to be achieved. High quality Si_3N_4 and some high κ materials, such as H_fO_2, Al_2O_3 are good replacements for amorphous thermally grown SiO_2 [73]. The lower injection barrier can enhance the electron injection efficiency during programming. For the same EOT (Effective Oxide Thickness), silicon nitride is physically thicker than thermal oxide, and high κ materials are even thicker than silicon nitride. Therefore, better retention might be achieved if the tunnel insulating layer is some other material than SiO_2.
- It has become clear that the shrinking transistors dimensions will reach less than 10 nm in the next ten year. Quantum and single electron charging effects are bringing new functionalities into transistors as shown in Fig. 7.1. The size-controlled Si NC method realizes a rather high Si NC density in the plane of

the gate and adjustable distance between neighbor Si dots, and could be very convenient to fabricate single electron transistors.

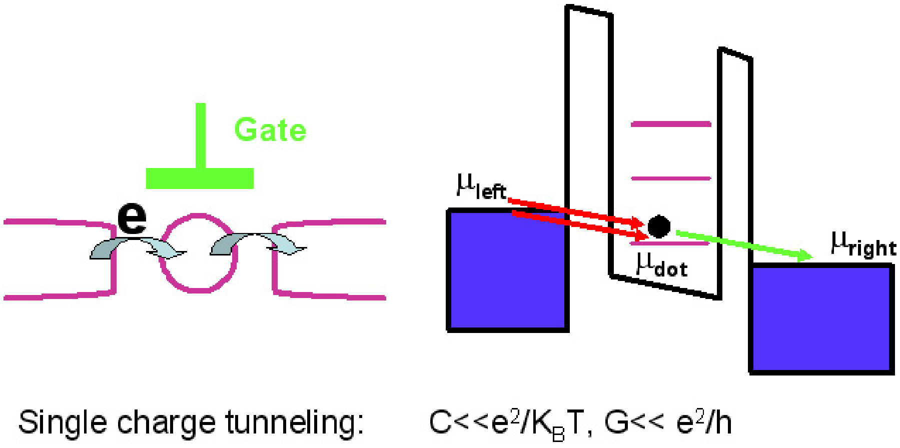


Figure 7.1: Single electron transistor and its principle. Coulomb blockade of quantum dot in trap will reject the tunnelling of second electron

New materials and novel device structures are always indispensable for flash memory scaling and device performance improvements. A trap-based flash memory is more scalable than other floating gate flash memory types. Good trap materials/species with large areal trap density and deep trap energy level are desirable for enhancing the programming efficiency and retention time. More work is required to investigate a good charge trap material that is compatible with current CMOS technology.

Novel device structures represent another approach to achieve small memory cell size and make flash memory scalable. For instance, a backside trapping structure memory has been proposed by Silva *et al.* [74]. Backside storage memories shown in Fig. 7.2 present an alternative to the conventional front-floating gate geometries by storing charge in discrete sites on the back of a thin depleted silicon channel. The devices are fabricated using a modified 'smart-cut' substrate preparation process followed by standard CMOS processing. Because the charge trapping layer is buried in ultra thin single crystal silicon, this upper ultrathin silicon becomes a channel for scaled backside trapping silicon non-volatile memories and eliminates the possible parasitic junctions between Source/Drain and substrate. In this back storage design, the choice of the charge trapping layer is optional. Not only an ONO stack, but also semiconductor or metal nanocrystals can be considered, and the 'smart-cut' process must be adjusted for these charge trapping layers.

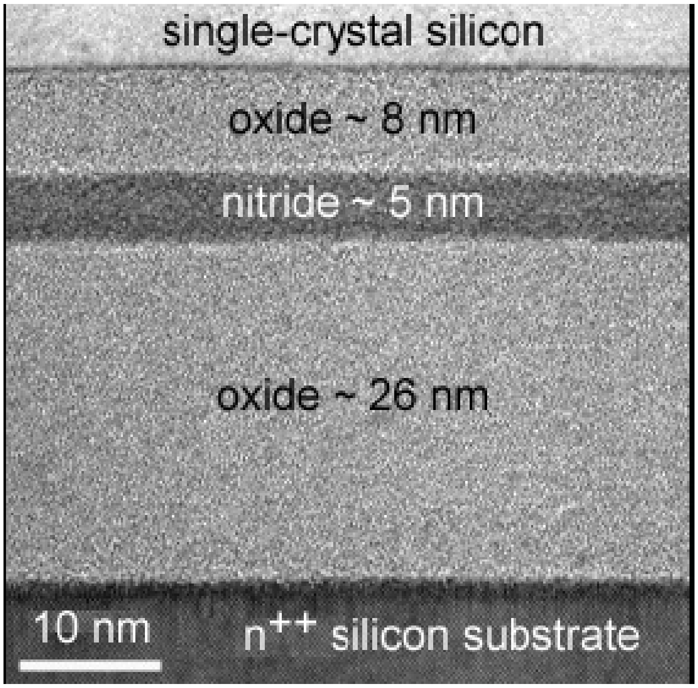


Figure 7.2: TEM images of a prepared substrate (single crystal silicon above a ONO stack) after smart-cut wafer bonding for backside storage. Image was obtained at Cornell Univ [74].

7.2.2 Nanoscale approach for memory devices

The developments in nanoscale silicon electronics require new tools to locally characterize the electrical properties of the Si NCs. In this direction, Scanning Capacitance Microscopy and Conductive Atomic Force Microscopy (C-AFM) are alternative techniques to perform an electrical characterization at the nanoscale. C-AFM has also been used to estimate the amount of charge stored in few Si NCs from its electrical measurements due to high lateral resolution of the technique, which allows to reveal details that would be masked when standard wafer level electrical tests had been carried on [75]. Nanoscale approaches, such as C-AFM, STM or point contact will play an important role in further single Si NC research.

7.2.3 Approach for optical modulation

Integrated optics in silicon is very attractive for a combination of technology and cost reasons. Besides lasing and waveguiding, one of the requirements of photonics is the optical modulation, especially by electrical field. Most efficient way of implementing optical modulation in silicon via the E-field is to use carrier injection or depletion. It is clear that the concentration of free carries in silicon contributes to the loss via absorption, therefore changing the concentration of free charges can influence the refractive index of the material. Liu *et al.* from Intel produced high-speed optical phase modulation based on a MOS capacitor structure embedded in a silicon waveguide [76]. They used a thin accumulation charge layer on both lateral sides of the gate oxide in accumulation operation and carriers change the refractive index and optical absorption through Kramers-Kronig analysis. At $\lambda = 1.55\mu$m, Kramers-Kronig relation is:

$$\begin{gathered}\triangle\eta = \triangle\eta_e + \triangle\eta_h = -[8.8 \times 10^{-22}\triangle N_e + 8.5 \times 10^{-18}(\triangle N_h)^{0.8}] \\ \triangle\alpha = \triangle\alpha_e + \triangle\alpha_h = 8.5 \times 10^{-18}\triangle N_e + 6.0 \times 10^{-18}\triangle N_h\end{gathered} \tag{30}$$

where $\triangle\eta_{e,h}$ is the change in refractive index resulting from change of electron or, hole carrier concentration, $\triangle\alpha_{e,h}$ is the change in absorption resulting from change of electron, hole carrier concentration, $\triangle N_{e,h}$ is the carrier concentration change [50].

Considering Si NCs in an SiO_2 matrix, it is well known that suitable bias can adjust the amount of the charges of the Si NCs in SiO_2, typical the area charging Si NC density is the level of $10^{12}/cm^2$ in few tens of nm thick SiO_2 matrix, approximate value charging density is level of $10^{18}/cm^3$. Furthermore we estimate the refractive index change by eq. 30 using these charging levels, higher refractive index change ($> 10^{-4}$) can be obtained. However, eq. 30 is only valid for silicon medium, since it derived from the analysis of experimental silicon electro-absorption spectra and impurity-doping spectra [77]. Because its working speed is proportional to the charge tunneling process, one of the key limitations for using silicon as a photonic material is that it only works at relatively low speed of silicon optical modulators compared to those fabricated from III-V semiconductor compounds and electro-optic materials such as lithium niobate.

7.2.4 Magnetic ordering in Mn-ion implanted Si NCs

It is known that the silicon device is approaching its limitations. New device physics and models are required for next generation data process. Spintronics which combine magnetism and solid state electronics, might open into a new route for device physics. Dilute magnetic semiconductors, a combination of transition metals such as Mn, Ni with group IV materials will take advantage of prevailing Si technology and magnetism from transition metal to realize Si-based spintronics devices. There only a few works on the magnetic property of metal doped silicon NCs. Scientists at Sandia National Lab speculate on the existence of a ferromagnetic ordering in Mn-ion doped silicon NCs based on the temperature dependence of PL with and without a magnetic field [78]. We might realize magnetic structure by either thermal diffusion or ion implantation of transition metal into an SiO_2 matrix containing Si NCs.

7.2.5 Tandem solar cell based on Si quantum dots

Solar cell is a rather old topic. Tandem solar cells are currently discussed in "third generation" approaches because of its higher convert efficiency than the first two generations of solar cells. The concept of tandem solar cells is using stack structure and vary the band gap of each stack from top surface to bottom as Fig. 7.3. These different band gaps serve as desired filters and convert the different wavelength of solar energy. The performance increases as the number of cells in the tandem increases [79].

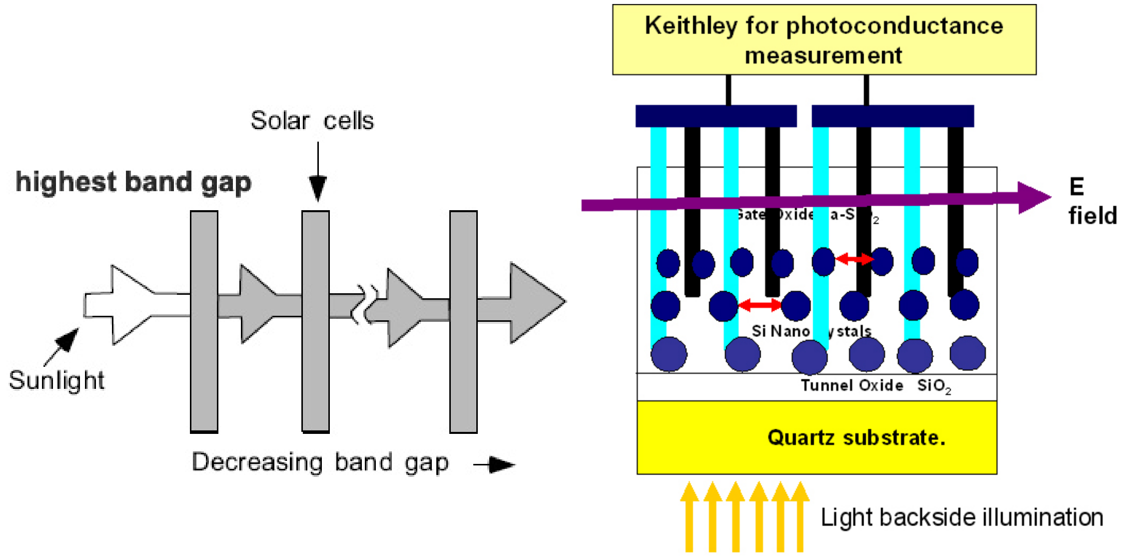


Figure 7.3: Tandem cell structure from the group of M. Green at Univ of New South Wales, Australia

Si NCs having excellent properties, such as controlled band gap by adjusting the dot size, abundant and nontoxic properties, silicon NCs have became a promising candidate for all-silicon tandem solar cell. However, Si NCs are surrounded by an insulating matrix, for instance SiO_2, Si_3N_4, which deteriorates the current output. Green model featured values for solar cell, such as bandwidth, mobility dependence on the properties of Si quantum dots and propose size-controlled Si NCs in a matrix which could potentially serve as the stack in all-silicon tandem solar cells [60].

Bibliography

[1] D. Kahng and M. M. Atalla. *IRE Device Research Conference*, 1960.

[2] R. Lankhorst, S. Ketelaars, and M. Wolter. *Nature Materials.*, 4:347, 2005.

[3] S. Tiwari, F. Rana, H. Hanafi, A. Hartstein, E. F. Crabbé, and K. Chan. *Appl. Phys. Lett.*, 68:1377, 1996.

[4] R. Muralidhar, R. F. Steimle, M. Sadd, and R. Rao. *IEDM Technical Digest*, 2003.

[5] L. J. Guo, E. Leobandung, and Stephen Y. Chou. *Science*, 275:649, 1997.

[6] S. M. Sze. Modern semiconductor device physics. *John Wiley & Sons, INC*, 1998.

[7] Y. Shi, K. Saito, H. Ishikuro, and T. Hiramotob. *J. Appl. Phys.*, 84:2358, 1998.

[8] S. Tiwari, J. A. Wahl, H. Silva, F. Rana, and J. J. Welser. *Appl. Phys. A*, 71:403, 2000.

[9] Y. Yu and M. Cardona. Fundamentals of semiconductors : physics and materials properties. *Spring Press*, Berlin, 1996.

[10] L. Pavesi. *Mat. Today*, 1:18, 2005.

[11] R. Tsu. *Appl. Phys. A*, 71:391, 2000.

[12] A. G. Cullis and L. T. Canham. *Nature*, 353:335, 1991.

[13] A. Irrera. Light emission devices based on the silicon nanostructures. *PhD Thesis*, 2003.

[14] L. Pavesi, L. Dal Negro, C. Mazzoleni, G. Franzo, and F. Priolo. *Nature*, 408:440, 2000.

[15] J. Ruan, P. M. Fauchet, L. Dal Negro, M. Cazzanelli, and L. Pavesi. *Appl. Phys. Lett.*, 83:5479, 2003.

[16] L. Dal Negro, P. Bettotti, M. Cazzanelli, D. Pacifici, and L. Pavesi. *Optics Communications*, 229:337–348, 2004.

[17] R. J. Zhang, S. Y. Seo, A. Milenin, M. Zacharias, and U. Goesele. *App. Phys. Lett.*, 88:153120, 2006.

[18] A. Polman and F. C. J. M. van Veggel. *J. Opt. Soc. Am. B*, 21:871, 2004.

[19] R. Walters, P. Kik, J. Casperson, H. Atwater, R. Lindstedt, M. Giorgi, and G. Bourianoff. *Appl. Phys. Lett.*, 85:2622, 2005.

[20] R. J. Walters, G. I. Bourianoff, and H. A. Atwater. *Nature Materials.*, 4:143, 2005.

[21] Liu. H. T and A. P. Alivisatos. *Abstracts of Papers of the American Chemical Society*, 230:u2249, 2005.

[22] S. Decossas, F. Mazen, T. Baron, G. Bremond, and A. Souifi. *Nanotechnology*, 14:1272, 2003.

[23] J. Groman, D.G. Hasko, and D. A. Williams. *Phys. Rev. Lett*, 95:090502, 2005.

[24] A. G. Cullis, L. T. Canham, and P. D. J. Calcott. *J. Appl. Phys.*, 82:909, 1997.

[25] T. Shimizu-Iwayama, K. Fujita, S. Nakao, K. Saitoh, T. Fujita, and N. Itoh. *J. Appl. Phys.*, 75:7779, 1994.

[26] N. M. Hwang W. S. Cheong and D. Y. Yoon. *J. of Cryst. Growth*, 204:52, 1999.

[27] L. He, T. Inokuma, Y. Kurata, and S. Hasegawa. *J. Non-cryst. Solids*, 185:249, 1995.

[28] M. Fujii, A. Mimura, S. Hayashi, D. Kovalev, and F. Koch. *Mat. Res. Soc. Symp. Proc.*, 638:F9.2.1, 2001.

[29] H. Seifarth, R. Grötzschel, A. Markwitz, W. Matz, P. Nitzsche, and L. Rebohle. *Thin Solid Films*, 330:202, 1998.

[30] U. Kahler and H. Hofmeister. *Opt. Mat.*, 17:83, 2001.

[31] M. Zacharias, J. Heitmann, R. Scholz, U. Kahler, M. Schmidt, and J. Bläsing. *Appl. Phys. Lett.*, 80:661, 2002.

[32] J. Heitmann, F. Mueller, M. Zacharias, and U. Goesele. *Adv. Mat.*, 17:795, 2005.

[33] S. Huang, S. Banerjee, R. T. Tung, and S. Oda. *J. Appl. Phys.*

[34] J. S. Biteen, N. S. Lewis, H. A. Atwater, and A. Polman. *Appl. Phys. Lett.*, 84:5389, 2004.

[35] A. Irrera, D. Pacifici, M. Miritello, G. Franzò, F. Priolo, F. Iacona, D. Sanfilippo, G. Di Stefano, and P. G. Fallica. *Appl. Phys. Lett.*, 81:1866, 2002.

[36] J. R. Brews E. H. Nicollian. **MOS** (metal oxide semiconductor) physics and technology. *John Wiley & Sons, Inc. New York*, 1982.

[37] T. Hori. Gate dielectrics and **MOS** ulsis principles, technologies and applications. *Springer, Inc*, 1997.

[38] A. S. Grove and E. H. Snow. *Proceedings IEEE,*, 54:894, 1966.

[39] D. K. Schroder. Semiconductor materials and device characterization. *John Wiley & Sons, Inc. New York*, 1998.

[40] G. D. Wilk, R. M. Wallace, and J. M. Anthony. *J. Appl. Phys.*, 89:5243, 2001.

[41] J. Heitmann, F. Müller, L. Yi, M. Zacharias, D. Kovalev, and F. Eichhorn. *Phys. Rev. B*, 69:195309, 2004.

[42] A. Zimina, S. Eisebitt, W. Eberhardt, J. Heitmann, and M. Zacharias. *Appl. Phys. Lett.*, 88:163103, 2006.

[43] O. Breitenstein and M. Langenkamp. Lock-in thermography : basics and use for functional diagnostics of electronic components). *Springer, Berlin*, 2003.

[44] S. Huang, S. Banerjee, R. T. Tung, and S. Oda. *J. Appl. Phys.*, 93:576, 2003.

[45] J. De Blauwe and G. Groeseneken. *IEEE Trans. Electr. Dev.*, 45:1745, 1998.

[46] Liu. Z, Lee. C, Narayanan. V, Pei. G, and Kan. E.C. *IEEE Transactions on Nanotechnology*, 2:345, 2003.

[47] I. Kim, K. Yanagidaira, and T. Hiramoto. *IEEE Electron Devices Lett.*, 25:265, 2004.

[48] F. L. Zhou, Y. Cai, R. Huang, X. Zhang, and Y. Y. Wang. *Proceedings of the 36th European Solid-State Device Research Conference*, page 226, 2006.

[49] T. C. Chang, S. T. Yan, P. T. Liu, C. W. Chen, H. H. Wu, and S. M. Sze. *Appl. Phys. Lett.*, 85:248, 2004.

[50] G. T. Reed and A. P. Knights. Silicon photonics: an introduction. *Willey*, 2004.

[51] L.W. Wang and A. Zunger. *Phys. Rev. Lett.*, 73:1039, 1994.

[52] C. Y. Ng, T. P. Chen, A. L. Ding, Y. Liu, M. S. Tse, D. Fung, and Z. L. Dong. *Appl. Phys. Lett*, 88:063103, 2006.

[53] M. Ohring. Reliability and failure of electronic materials and devices. *Academic Press*, 1998.

[54] A. Salonidou, A. G. Nassiopoulou, and A. Salonidou. *Program of EMRS2006 Spring meeting, Syposium C 0702*, 2006.

[55] A. Kumar, J. J. Welser, S. Tiwari, F. Rana, and K. K. Chan. *Mat. Res. Soc. Symp. Proc.*, 638:F2.4.1, 2001.

[56] D. Kovalev, H. Heckler, G. Polisski, and F. Koch. *phys. stat. sol. (b)*, 215:871, 1999.

[57] L. X. Yi, J. Heitmann, R. Scholz, and M. Zacharias. *Appl. Phys. Lett*, 81:4248, 2002.

[58] S. M. Sze. Physics of semiconductor devices. *John Wiley & Sons, Inc. New York*, 1981.

[59] J. Tseng, C. W. Cheng, S. Y. Wang, T. B. Wu, K.Y. Hsieh, and R. Liu. *Appl. Phys. Lett.*, 85:2595, 2004.

[60] C. W. Jiang and M. A. Green. *J. Appl. Phys*, 99:114902, 2006.

[61] Y. Shi, K. Saito, H. Ishikuro, and T. Hiramoto. *J. Appl. Phys.*, 84:2358, 1998.

[62] G. G. Qin. *MRS Reserach Bulletin*, 33:1857, 1998.

[63] Y. H. Kwon, C. J. Park, W. C. Lee, D. J. Fu, Y. Shon, T. W. Kang, C. Y. Hong, H. Y. Cho, and K. L. Wang. *Appl. Phys. Lett.*, 80:2502, 2002.

[64] D. V. Lang. *J. Appl. Phys.*, 45:3023, 1974.

[65] M. Schulz and N. M. Johnson. *Solid. State. Commun.*, 25:481, 1978.

[66] A. Souifi, P. Brounkov, S. Bernardini, C. Busserret, L. Militaru, G. Guillot, and T. Baron. *Material. Scien. Eng. B.*, 102:99, 2003.

[67] H. Oda and D. K. Ferry. Silicon nanoelectronics. 2005.

[68] K. Yamasaki, M. Yoshida, and T. Sugano. *Jpn. J. Appl. Phys*, 18:113, 1979.

[69] J. S. Biteen, D. Pacifici, N. S. Lewis, and H. A. Atwater. *Nano Lett.*, 5:1768, 2005.

[70] D. Kovalev, J. Diener, H. Heckler, G. Polisski, N. Künzner, and F. Koch. *Phys. Rev. B*, 61, 2000.

[71] H. Silva, M. K. Kim, U. Avci, A. Kumar, and S. Tiwari. *MRS Bulletin*, 3:845, 2004.

[72] L. Militaru and A. Souifi. *Appl. Phys. Lett*, 83:2456, 2003.

[73] Wan. Q, Wang. T. H, Zhu. M, and Lin. C. L. *Appl. Phys. Lett.*, 81:538, 2002.

[74] H. Silva and S. Tiwari. *IEEE Transactions on Nanotechnology.*, 3:264, 2004.

[75] M. Porti, M. Avidano, M. Nafria, X. Aymerich, J. Carreras, and B. Garrido. *Micro. Engineering*, 80:268, 2005.

[76] A. S. Liu, R. Jones, L. Liao, D. Samara-Rubio, D. Rubin, O. Cohen, R. Nicolaescu, and M. Paniccia. *Nature*, 427:615, 2004.

[77] R. A. Soref and B. R. Bennett. *J. Quant. Electron*, 23:123, 1987.

[78] W. Pan, R. G. Dunn, M. S. Carroll, and Y. Q. Wang. *Mater. Res. Soc. Proc.*, 910:A08, 2006.

[79] M. A. Green. *Physica E*, 14:65, 2002.

APPENDIX A

Publications and Presentation

- **T.Z. Lu**,J. Shen, B. Mereu, M. Alexe, R. Scholz, V. Talalaev and M. Zacharias, “Electrical behavior of size-controlled Si nanocrystals arranged as single layers”, *Appl. Phys. A: Materials Sciences and Process* 80, 1631 (2005).
- **T.Z. Lu**, M. Alexe, R. Scholz, V. Talalaev and M. Zacharias, “Multilevel charge storage in Silicon nanocrystal multilayers”,*Appl. Phys. Lett.* 87, 202110 (2005).
- Zhao LL, **Lu TZ**, Zacharias M, Yu J, Shen J, Hofmeister H, Steinhart M, Goesele U, “Integration of erbium-dooped lithium niobate microtubes into ordered macroporous silicon”, *ADVANCED MATERIALS* 227, 332(2005).
- **T.Z. Lu**,M. Alexe, R. Scholz, V. Talalaev and M. Zacharias, “Si nanocrystal based memories: Effect of the nanocrystals density”,*J. Appl. Phys.* 100, 014310 (2006).
- **T.Z. Lu**, M. Alexe, R. Scholz, and M. Zacharias, ”Multilevel charge storage in Silicon nanocrystal multilayers”, European Material Research Society Spring Meeting 2006, May 29-Jun 2, Nice, France, talk.
- **T.Z. Lu**, M. Alexe, R. Scholz, V. Talalaev and M. Zacharias, ”Si nanocrystal memory: Effect of the nanocrystal density”, European Material Research Society Spring Meeting 2006, May 29-Jun 2, Nice, France, poster.
- **T.Z. Lu**, M. Alexe, R. Scholz, and M. Zacharias, ”Multilevel charge storage in Silicon nanocrystal multilayers”, European Solid State Device Conference 2006, Sep 16-22, Montreux, Switzerland.

APPENDIX B

Curriculum vitae

First Name	Tiezheng
Family Name	Lu
Date of Birth	January 30, 1977 (Hunan, P. R. China)
Nationality	Chinese
E-mail	*adanlu@mpi-halle.de* or *lutiezheng@yahoo.com*

EDUCATION

Nov. 2003 - present	Max Planck Institute of Microstructure Physics and Martin-Luther-University Halle Wittenberg (Halle, Germany) Ph.D. candidate (Supervisors: Prof. Dr. Margit.Zacharias and Prof. Dr. Ulrich Goesele)
Sep. 1998 - Jun. 2001	Institute of Physics, Chinese Academy of Sciences (Beijing, China) Master Degree in Optics
Sep. 1994 - Jun. 1998	Peking University (Beijing, China) Bachelor's Degree in Physics

---

PROFESSIONAL EXPERIENCE

May. 2003 - Nov. 2003	Max-Born-Institute and Free University at Berlin research assistant Inert gas ionization by HHG from femtosecond laser
Jan. 2002 - May. 2003	Hong Kong University of Science and Technology Research Assistant The modulation of Femtosecond pulse by Liquid crystal

Award

- Young Scientist Award of European Material Research Society Spring Meeting 2006, Nice, France

APPENDIX C

Eidesstattliche Erklaerung

Ich erklaere hiermit, dass ich keine anderen als die von mir angegebenen Quellen und Hilfsmittel zur Erstellung meiner Dissertation verwendet habe. Den benutzten Werken woertlich oder inhaltlich entnommene Stellen sind als solche gekennzeichnet.

Tiezheng Lu

Halle (Saale), im Jan 2007