Characterization of SONOS nonvolatile semiconductor memory (NVSM) devices for space and military applications.

Characterization of SONOS Nonvolatile Semiconductor Memory (NVSM) Devices for Space and Military Applications

by

Stephen J. Wrazien

A Dissertation

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Doctor of Philosophy

in

Electrical Engineering

Lehigh University

May 2005

This dissertation is approved and recommended for acceptance as a Dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering.

April 25th, 2005
Date

Marvin H. White
Dissertation Advisor

April 25th, 2005
Accepted Date

Committee Members:

Marvin H. White

Dennis A. Adams

Ralph J. Jaccodine

Svetlana Tatic-Lucic

To my parents, Barbara and Carl,

my brother, Dan,

my girlfriend, Lorie,

and my friends,

for all of their love and encouragement.

And to the 2004 Boston Red Sox

Acknowledgements

Thinking back on all of the experiences and events over the past few years of my graduate studies, it's amazing how many people have contributed to this degree. Whether it was working with me for a summer, collaborating on a paper, teaching me device physics, demonstrating processing techniques, or just encouraging me to keep going because I was so close to finishing and what a great achievement it will be once I'm finished, without this support and encouragement, this dissertation would not exist.

First, I would like to thank my advisor and mentor, Dr. Marvin White, for the opportunity to study at the Sherman Fairchild Lab as both an undergraduate and a graduate student. His financial support, technical expertise, and passion for innovative device research have fueled this research. The members of my Ph.D. committee, Dennis Adams, Dr. Ralph Jaccodine, and Dr. Svetlana Tatic-Lucic, have provided invaluable insight into device research and helped shape this dissertation. I would especially like to thank Dennis Adams, who has collaborated with our research group over the course of my studies, provided myself an internship at the Northrop Grumman Corporation, and has always supported my graduate and professional careers.

The faculty, staff, and students of the Sherman Fairchild Center have made the lab an innovative and stimulating research environment. I would never have been able to fabricate working devices without the expertise and patience of Raymond Filozof. His dedication to excellent processing techniques and to helping his students have made semiconductor processing at the lab an enjoyable and rewarding experience. Dr. Floyd

Miller has provided insights into fabrication techniques as well as introduced me to the art of SEM imaging. Mrs. Linda Dreisbach has provided constant support, ensuring that the administrative side of graduate research is always under control.

I would like to thank my fellow graduate students; Themis Afentakis, Jiankang Bu, Ta-Ko Chuang, James Fuerherm, Abbas Jamshidi-Roudbari, Gaoshan Jing, Joel Krayer, Santosh Pandey, Huiling Shang, Matias Troccoli, Yu "Richard" Wang, Xiaonan "Stephen" Wang, Yi Yao, Yu "Anne" Zeng, Wenyue Zhang, Yanli "Esther" Zhang, and Yijie "Sandy" Zhao, for their technical expertise and friendship. I'd especially like to thank; Matias, who has been a good friend and put up with living with me for the past 2 years, Gaoshan, a fellow member of Red Sox Nation, Joel, with whom I have enjoyed many music-related and random discussions, and Santosh, Anne, and Sandy, who have provided me with mentorship and friendship since day one at the lab. I'd also like to thank all of the summer students I have worked with on the SONOS project over the years (in chronological order); Jon Faul, Joel Krayer, Bilal Khan, Rajiv Mehrotra, and Wayne Miller. In addition, Andy Bronczyk, who made the summer of 2004 one of the most enjoyable in memory.

Huge thanks to my family and all of my friends who have supported me all along. Special thanks to my Mom, Dad, Dan, and my girlfriend Lorie, who have given me overwhelming love and encouragement. I'd also like to thank Eric, Adam, and Steve for providing me with many more than memorable non-research moments.

Thanks to the Northrop Grumman Corporation, the Army Space and Missile Defense Command, the Sherman Fairchild Foundation, and the National Science Foundation for their financial support.

Contents

List of Tables

List of Figures

Abstract

The increasing miniaturization of portable electronics, application of electronic systems for deep-space exploration, and the advent of nuclear terrorism have created the need for a high-speed, low-power, high-density non-volatile semiconductor memory (NVSM) capable of long-term data retention under heavy radiation doses. This dissertation characterizes polysilicon-oxide-nitride-oxide-silicon (SONOS) transistors, and proposes, characterizes, and models a device derived from the SONOS technology, the substrate injection nitride based (SNROM) device, to meet the future requirements of space and military systems.

This research characterizes novel short-channel substrate-injection nitride-based (SNROM) transistors (10 x 0.22 μm) fabricated with a 36Å-thick tunnel oxide, 30Å silicon nitride, and a 36Å blocking oxide as the gate dielectric. The thick tunnel oxide prevents tunneling of electrons and holes through the gate dielectric and allows the transistor to be programmed using hot electron and hole injection. This research programs SNROM transistors to produce a 2-volt memory window. The device is written using channel initiated secondary electron injection (CHISEL) by applying 3V to the gate, 2.5V to the drain, and -4V to the bulk for 5ms. The device is erased using hot hole injection by applying -7V to the gate and 4V to the drain for 5ms. The programmed transistor surpasses the 10-year retention requirement with a 0.5V memory window at the 10-year mark. The CHISEL injection mechanism is characterized and

modeled by extrapolating the threshold voltage shift, trapped charge, and injected current, from the reduction in channel current during programming. In addition, the impact ionization current is measured at the drain terminal and correlated with the current injected into the ONO gate insulator.

Additionally, electrical characterization is performed on conventional SONOS transistors, with an emphasis on programming efficiency, endurance, and data retention. An acceleration factor is presented to linearly extrapolate 10-year retention from high temperature measurements. This dissertation demonstrates the ability of current conventional SONOS transistors to meet space and military system requirements and presents an innovative SNROM structure as a low-power, high-density NVSM for future applications.

Chapter 1

Introduction to Non-Volatile Semiconductor Memory (NVSM)

Semiconductor memory devices have evolved greatly over the past 10 years. Semiconductor memory is used in a wide variety of applications from personal computer RAM, used to store information required to run software programs and operating systems, to complex electronic systems embedded in satellites and most recently, portable personal electronics. Traditionally, there have been two types of semiconductor memory, volatile and non-volatile memory. Volatile semiconductor memory (DRAM, SRAM) stores data in dense arrays and has quick access times and can endure up to 10^{12} program/read cycles, but requires a power supply in order to retain data. Non-volatile semiconductor memory (NVSM) stores data in dense arrays with comparable access and programming times, but does not require power to retain data. However, the majority of NVSM's currently commercially available cannot withstand elevated programming ($>10^6$ write/erase cycles) and therefore cannot replace volatile semiconductor memory, such as DRAM personal computers [1].

Data is stored in semiconductor memory devices as a difference in turn-on (threshold) voltage, electrical resistance, or polarization, typically manifesting itself in a difference in read-out current when specific voltages are applied to the device. Writing the transistor will be referred to as a logic 'one', and erasing the device will be a logic

'zero'. The data must be retained by the NVSM for a minimum of 10 years.

The major NVSM's are floating gate (FG), discrete-trap (SONOS, NROM™), and novel concept (FeRAM, MRAM, PRAM). These technologies differ in the method in which they sense a difference in memory state, see Table 1.1. This dissertation focuses mainly on the SONOS/NROM™ technologies, which store positive and negative charge (holes and electrons, respectively) in discrete traps in a silicon nitride film in the gate dielectric. The ability of NVSM's to retain information without a power

NVSM	Technology	Memory Effect
Floating Gate	FG	Conductive Gate
Discrete Trap	SONOS	Si_3N_4 film
	NROM	Si_3N_4 film
Novel Concept	MRAM	Magnetic orientation
	FeRAM	Polarization
	PRAM	Physical state

Table 1.1: Summary of current generation NVSM technologies and method employed to produce memory effect.

supply makes them ideal for use in portable electronics and space systems running off finite battery supplies. In addition, NVSM's which store charge in discrete traps are inherently radiation-hard and can withstand vast doses of radiation required for use in outer space where they are exposed to cosmic radiation and military systems where radiation from nuclear detonations is a concern..

1.1 Space and Military Electronic Systems

Nonvolatile semiconductor memory is used in a variety of space and military applications. The NVSM technologies utilizing discrete charge trapping (SONOS) are

inherently radiation-hardened, due to their thin tunneling oxides. During radiation exposure, ions trap in the tunneling oxide of the transistor and shift the threshold voltage, causing leakage currents to flow while the device is in the off state. Thinner tunneling oxides trap fewer radiated ions and therefore do not suffer from excessive performance degradation due to radiation exposure. Floating gate memories cannot be utilized in radiation applications due to the nature of the charge storage mechanism and the thick tunneling oxides. Charge is stored in a conductive gate isolated from the silicon substrate by a tunnel oxide, typically thicker than 7 nm. This thick oxide accumulates 2-3 times more charge during radiation exposure than SONOS or nitride

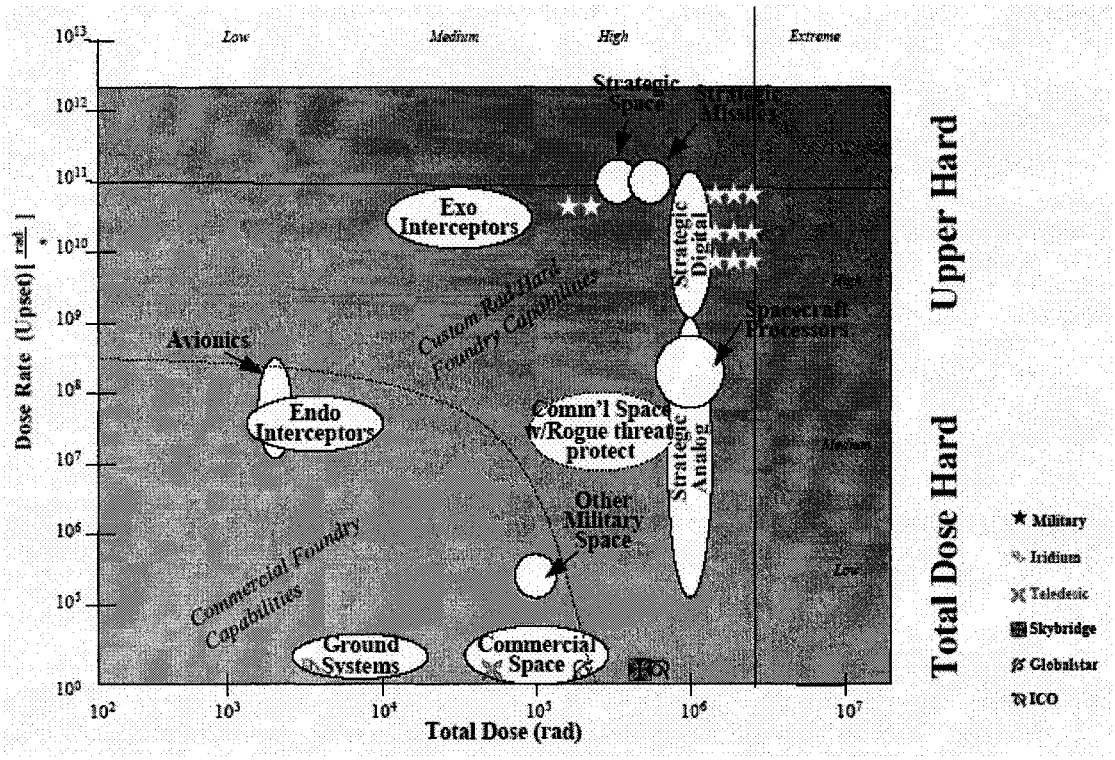


Figure 1.1: The unique radiation requirements of space and military applications with regards to lifetime and single event doses. The area under the dotted line represents the capability of commercial off-the-shelf (COTS) electronics [2].

based (NROM™) NVSM's. The radiated ions accumulating in the tunnel oxide create defects, which the charge in the conductive gate uses to escape the gate and the charge (memory state) is lost. In a SONOS structure, the charge must first escape the discrete trap it is stored in before tunneling to the trap and subsequently tunneling to the silicon bulk.

The Northrop Grumman Corporation has utilized SONOS NVSM's in a number of satellites and military applications such as the F-22 advanced fighter's electronically steered radar [1]. To this day, the SONOS technology is the only space-flown NVSM. SONOS devices are used to store module addressing and voltage biasing configurations for high-power Ga-As transistors in these applications [3].

Voltage	7V
Current	Low
Speed	< 10 ms
Retention	10 Years
Commercial	85 °C
Military	125 °C
Cycling	10^4
Radiation Dose	300 krad (Si)
Denisty	1 Mb

Table 1.2: Systems requirements for NVSM's utilized in space and military applications [4].

Typical NVSM's used in space and military applications must meet a number of specifications. Included are requirements for programming voltages and currents to ensure low power operation. Fast programming and read times are required given that

the longer it takes for a device to access and program, the more power it uses. NVSM's must retain data for 10-years within a range of specified temperatures. The device must be able to be programmed a number of times, but will generally not be repeatedly re-programmed since it will be used as a data storage device. A lifetime radiation dose requirement and density of the memory array must be met in order to integrate the part into an electronic system.

Electronics operating in orbit of Earth no longer face radiation hazards solely from the natural environment, but also nuclear terrorism. The availability of nuclear materials due to the collapse of the Soviet Union in the late 1980's, coupled with the determination of hard-line regimes, such as North Korea, which has publicized its development of nuclear weapons and nuclear-capable long-range missiles [5], has made this threat increasingly relevant to NVMS system design.

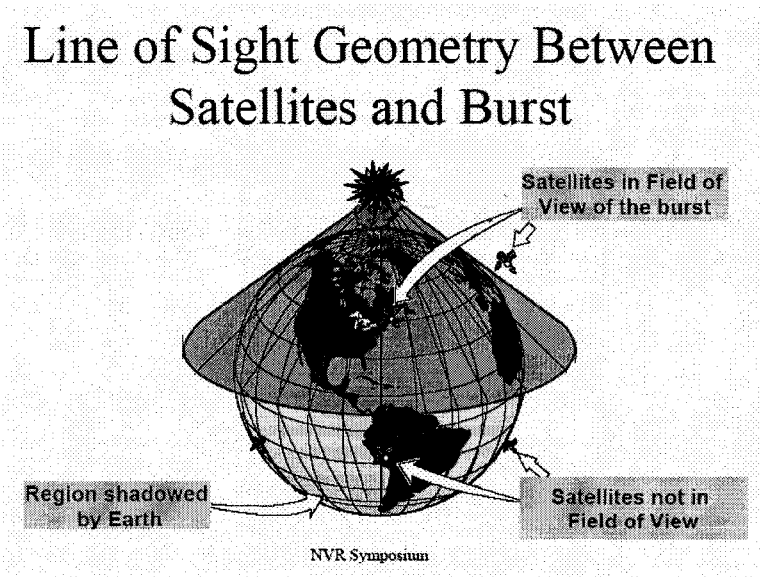


Figure 1.2: Effect of a nuclear detonation on commercial and military satellites in Earth orbit [2].

Researchers have run computer simulations to examine the effects of a single 50 kT nuclear burst at a height of 120 km over the Earth. The radiation from the nuclear detonation immediately destroys all satellites and space systems within line-of-sight of the explosion. The radiation from the detonation is then collected by the Van Allen belts and trapped within the Earth's atmosphere. This trapped radiation will travel around the Earth in the Van Allen belts and destroy the remaining satellites which were shielded by the Earth from the initial blast. Computer simulations have determined that within 50 days (< 2 months), all of the remaining commercial and un-protected satellites will be destroyed by this trapped radiation [2].

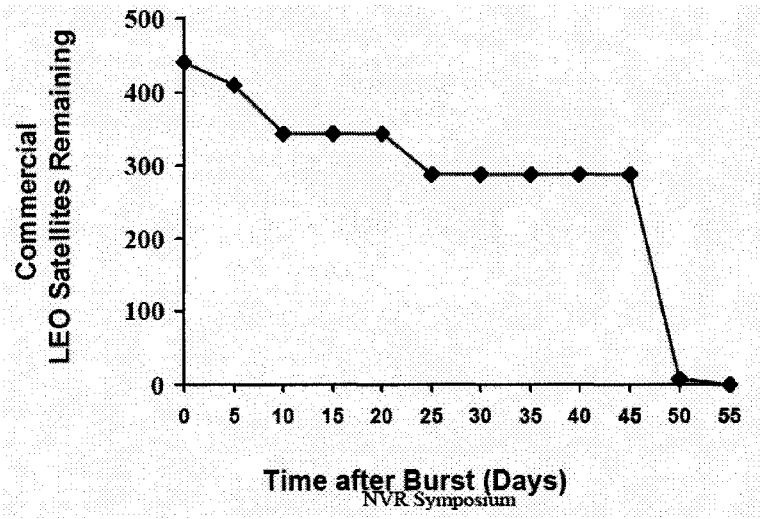


Figure 1.3: Number of satellites remaining after simulated 50 kT nuclear detonation in the Earth's atmosphere [2].

Historically, the defense industry has been 2 – 4 generations behind commercial electronics corporations in semiconductor device scaling, Fig. 1.4. A good metric for device scaling is CMOS transistor gate length. Aggressive NVSM device scaling has not been a priority of the defense industry, since NVSM are mostly used in fighter jet

radars or satellite systems which require small volumes of non-volatile memory and have sufficient chip area for the required parts. However, the increasing complexity of electronic systems due to the integration of diverse microelectronics on a single chip (system-on-a-chip), have pushed scaling to new heights for both commercial and defense electronics. The benefits of SONOS transistors programmed with CHISEL/HHI mechanisms (speed, retention, power) could encourage increased government funding of the defense industry to scale its microelectronics and remain competitive with commercial semiconductor foundries.

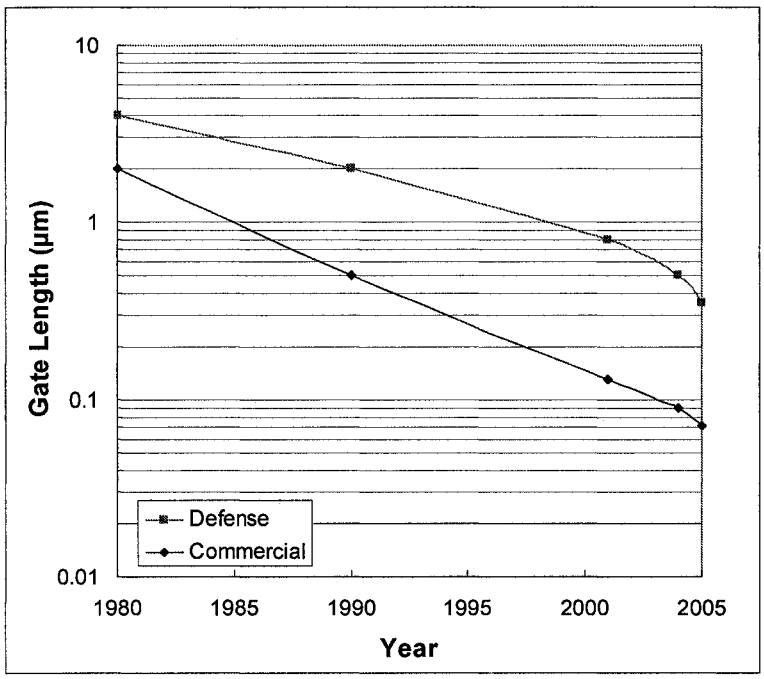


Figure 1.4: Gate length dimension scaling of commercial and defense electronics industries over the past 25 years [6].

1.2 NVSM Technologies for Space and Military Applications

Conventional non-volatile semiconductor memory (NVSM) devices store holes or electrons in a charge-storage layer in the gate dielectric. This charge can be stored in a variety of manners. The charge can either be stored in a conductive layer (floating gate) or in traps in the charge-storage layer (floating trap). NVSM devices are programmed by applying either a positive or negative voltage to the gate to promote the tunneling of electrons or holes, respectively, into the charge-storage medium.

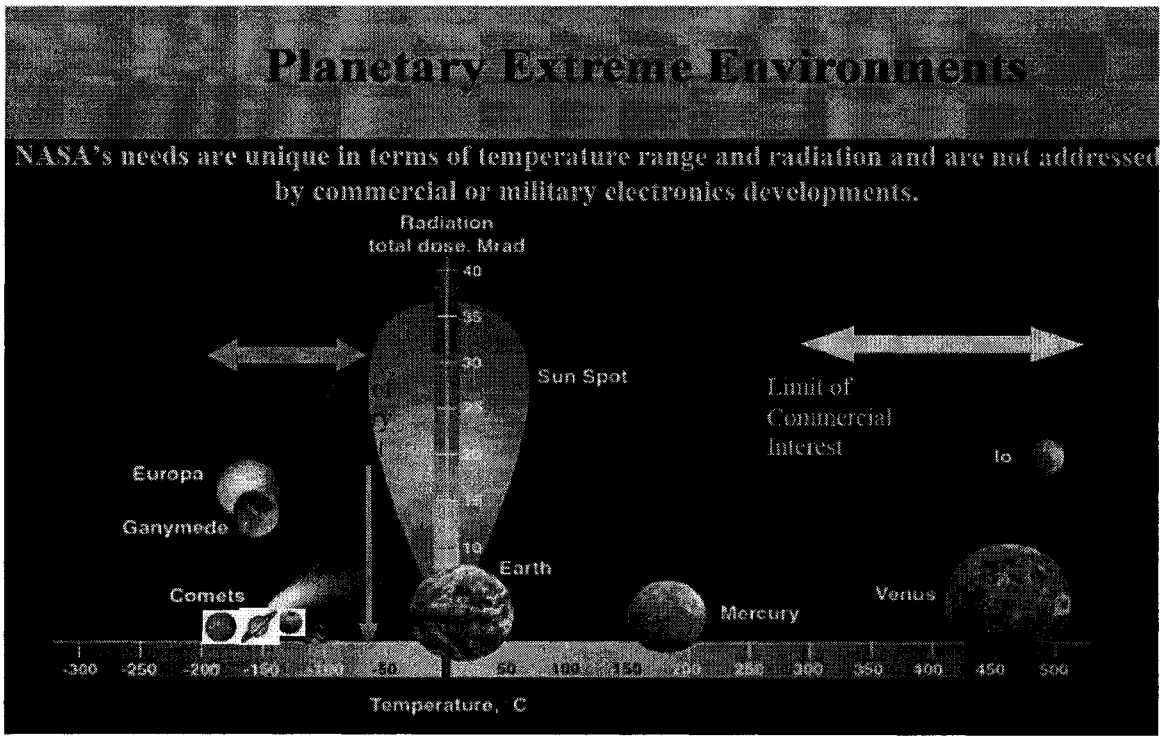


Figure 1.5: Unique temperature and radiation requirements for inter-planetary exploration [7].

In order to qualify an NVSM for space and military applications, we must examine the specific characteristics of that technology. Several key electrical characteristics of EEPROM devices include retention, endurance, and programming

speed. Retention refers to the ability of the device to store and sense charge in the nitride layer after a number of write and erase cycles at a specific temperature. Endurance refers to the ability of a device to withstand repeated write and erase cycles and still be able to retain and recover charge. Programming speed refers to the time over which the programming voltage is applied to the gate in order to switch the device from the write to the erase state.

1.2.1 Floating Gate

Floating gate transistors are written using channel hot electron (CHE) injection and erased using Fowler-Norheim (FN) tunneling. To program the device with CHE, a high lateral electric field is applied along the channel of the device, causing electrons to accelerate through the channel until they are scattered by the lattice or another carrier. The appropriate vertical field will result in "lucky" electrons, which will accumulate enough energy to cross the Si-SiO_2 barrier and inject themselves into the gate dielectric where they are stored in traps. The Carrier Hot Electron (CHE) current is defined using the equation [8]

$$I_G = \left(\frac{\Delta L}{\lambda_R}\right)\left(\frac{\lambda_S \varepsilon_y}{4\phi_B}\right) I_S e^{-\phi B/\lambda_S \varepsilon_y} e^{-X_B/\lambda_{OX}} e^{-\bar{x}/\lambda_S} \tag{1.1}$$

where ΔL is the length of the charge injection region, λ_R is the redirection mean free path, λ_{OX} is the mean free path in silicon-dioxide, λ_S is the electron mean free path in silicon, ϕ_B is the SiO_2 barrier height measured from the substrate, I_S is the source

current, E_y is the maximum vertical field in the vicinity of the drain, expressed as $E_y = \frac{qN_A X_d}{\varepsilon_S}$ where N_A is the substrate doping, X_D is the drain depletion width, and ε_S is the permittivity of silicon.

Floating gate devices offer rapid programming using channel hot electron injection and high density, since two devices may share a single source line [9]. Floating gate transistors are currently the most widely used NVSM devices for applications such as memory cards used in digital cameras and cellular phones, smart cards, and automotive applications. These applications all require low voltage and low power operation, rapid programming, extreme high and low temperature tolerance, long-term data retention, and endurance up to one-million write/erase cycles [7].

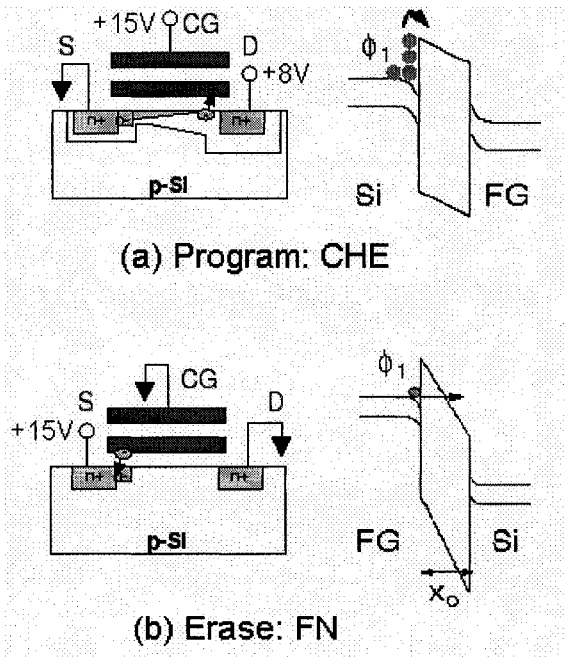


Figure 1.6: Write/erase mechanisms of mainstream flash technologies. Programming voltages and energy-band diagrams for a) CHE injection write and b) FN erase operation [10].

The conductive charge storage medium in a floating gate Flash memory transistor allows a single defect in the oxide to remove all of the stored charge. The charge will tunnel from the conductive gate to the interface defect then to the silicon substrate. Due to this vulnerability to single-defect charge loss, floating gate devices suffer from reliability issues stemming from oxide hot carrier degradation, punch-through avalanche effects, and tunnel oxide scaling limits [11, 12]. Floating gate devices also suffer from higher programming voltages and cell scaling limits [13].

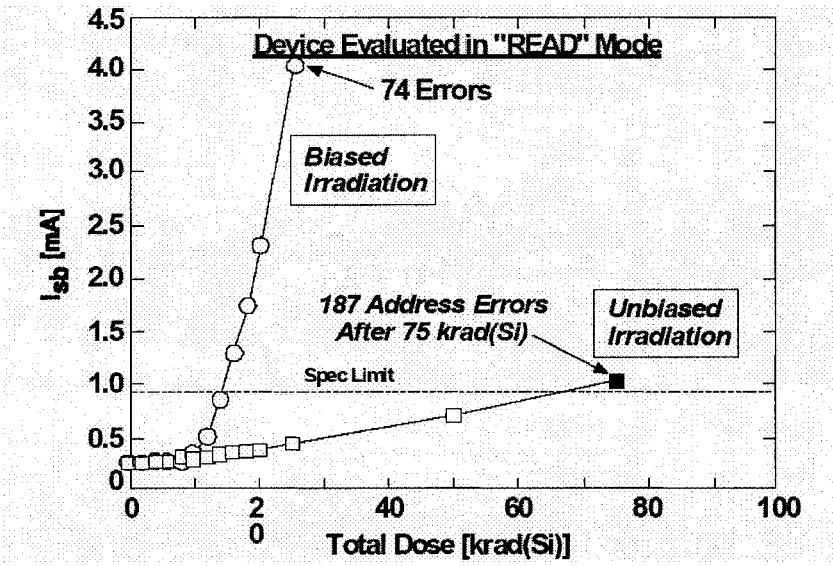


Figure 1.7: Experimental data showing floating gate failure, depicted as an increase in leakage current versus total dose radiation [14].

Floating gate memory, the traditional choice for many memory-intensive applications, has several major limitations with respect to cell size scaling and programming voltage [15]. Due to the complicated nature of the gate dielectric, scaling the device below the 100 nm node is difficult. Recent advances show the ability of the device to be scaled below the 120 nm node [16]. However, the complex capacitances

involved in the gate dielectric hamper vertical gate scaling, requiring higher programming voltages. Floating gate memories also suffer from dielectric hot carrier degradation which damages the bottom insulating oxide during the write function, avalanche effects at the drain junction, and high-voltage breakdown of the oxides due to the high electric fields required to program these devices [13]. However, the major drawback of the floating gate technology is its inability to function in a radiation environment, with peripheral charge pumping circuits failing at 10krad, sensing amplifiers failing at 50krad, and floating gate NVSM erasure at 100krad [14]. This performance is eclipsed by the SONOS technology which can operate without incident at radiation levels up to 500krad [1].

1.2.2 SONOS

Floating trap devices, such as polysilicon-oxide-nitride-oxide-silicon (SONOS), transistors offer several advantages over floating gate transistors. SONOS devices store charge in spatially isolated deep level traps in the charge-storage layer, see Fig. 1.8. The SONOS device is written by grounding the source, drain, and bulk, and applying a positive voltage to the gate, enabling Fowler-Nordheim tunneling of electrons to the charge-storage medium where they are stored in traps [17]. Erasing the SONOS device is accomplished by grounding the source, drain, and bulk, and applying a negative voltage to the gate, causing holes to tunnel from the silicon substrate and trap in the nitride layer [13]. The charge remains trapped in the nitride after all voltages are removed.

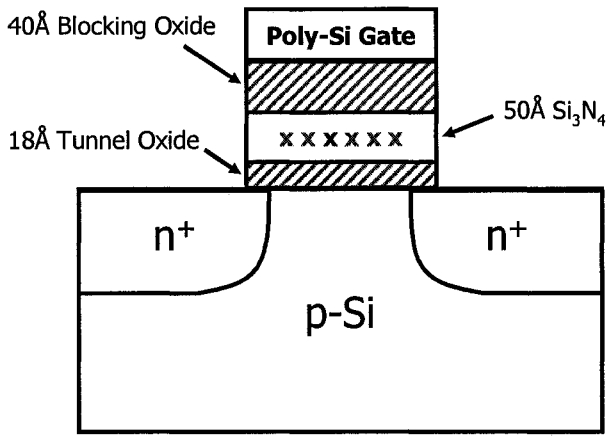


Figure 1.8: The vertical cross-section of a SONOS device, X's represent charge storage traps in the silicon nitride film.

Storing charge in traps in the nitride layer prevents charge from tunneling through the oxide layers to either the gate or the substrate when the transistor is in the retention mode. Eventually, holes and electrons will escape from their traps and tunnel back to the silicon bulk. This occurs over an extremely long period of time, typically, 10 to 100 years. There are several mechanisms responsible for charge loss: trap to band tunneling, trap to trap tunneling, and thermal excitation of electrons [18]. In an N-channel SONOS device, storing negative charge, electrons, in the nitride will cause the threshold voltage to increase, and storing positive charge, holes, in the nitride will cause the threshold voltage to decrease.

SONOS devices offer several advantages over floating gate memory transistors. SONOS devices have a higher tolerance to radiation [19], use lower operating voltages and less current, have greater endurance to extended write/erase cycling, and are compatible with scaled 'standard' CMOS technology with fewer processing steps [15]. SONOS transistors are more easily scaled than floating gate memory devices. The

	SONOS	Floating Gate
Carriers	Write: Electrons Erase: Holes	Electrons
Structure	One Gate One Dielectric	Two Gates Two Dielectrics
Gate Control	Direct	Capacitively coupled to control gate
Scaleable	Linear with program voltage	Not easily scaled
Reliability	Immune to single defect memory loss	Susceptible to single defect memory loss in tunneling oxide
Programming Voltages	Lower voltages (5-7 V)	High Voltages (~15 V)
Electric Fields in Tunnel Oxide	Reduced Fields (Modified Fowler Nordheim Tunneling)	High Fields (Fowler Nordheim Tunneling and Channel Hot Electrons)

Table 1.3: Summary of the major differences between SONOS and floating gate technologies [20].

nature of the floating gate makes it difficult to discern the exact voltage and field applied to the conductive gate. Bias is not directly applied to the gate, which is surrounded by four oxide and nitride films. A SONOS device is biased by applying a voltage directly to the gate, making it is easier to scale the programming voltages with the gate dielectric thickness [21].

1.2.3 NROM™

Several novel methods for programming SONOS devices can be utilized to provide performance enhancements over the floating gate technology. One such innovation is the concept of storing charge in two distinct localized regions in the

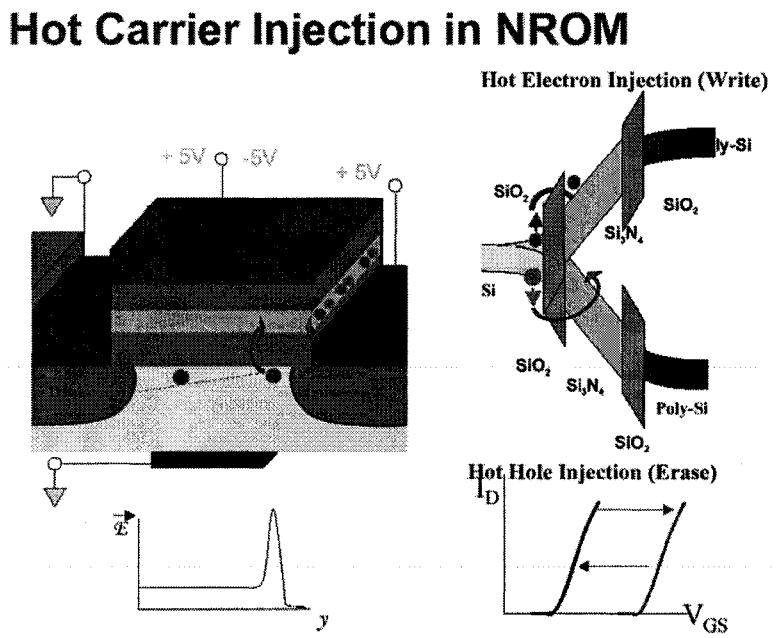


Figure 1.9: Programming a short-channel NROM™ transistor using channel hot electron (CHE) and hot hole injection (HHI) at the drain side of the device [22].

nitride [23]. B. Eitan, et al., pioneered the research and development of the so-called 2-Bit NROM™. This device utilizes channel hot electron (CHE) injection at the source and drain junctions to store a small amount of charge locally in the nitride. A positive voltage is applied to the gate, a separate positive voltage is applied to either the source or the drain to select that side of the gate dielectric for charge injection, and the remaining terminals are grounded [24]. The tunneling oxide for NROM™ devices must be at least 3 nm thick in order to significantly block direct tunneling during the write and erase operations [23] and reduce so-called 'back-tunneling' in order to enhance the charge storage or memory retention.

Erasure of the NROM™ device is accomplished using hot hole injection (HHI). A negative voltage is applied to the gate, a positive voltage is applied to either the

source or drain, selecting the side of the transistor where the holes will be injected, and the remaining terminals are grounded [24]. The blocking oxide must be thick enough (> 4 nm) in order to prevent electron injection from the gate during the erase operation. The device is read by performing a "reverse read" operation where the transistor is read in the direction opposite the write or erase [25]. For example, to read the bit at the drain side, a small voltage of roughly 1.5 volts will be applied to the source, the drain and bulk will be grounded, and the gate will be swept from 0 to 3 volts [24]. The exact

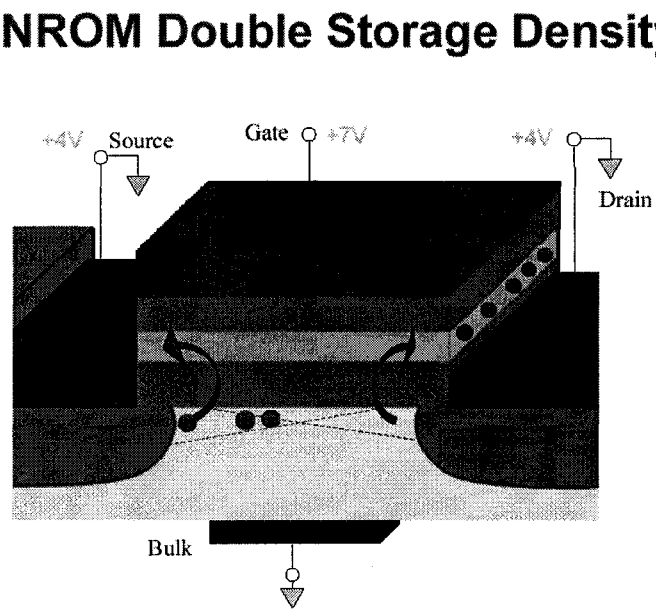


Figure 1.10: Multi-bit data storage in a single NROM™ transistor achieved via independently injecting charge at both the source and drain junctions [22].

'read' voltage will depend on the device design, since too low a voltage will not provide sufficient drive current and too high a voltage will cause excessive drain induced barrier lowering, thereby degrading the programming window. The programming and read operations are detailed in Table 1.4.

The NROM™ device is a hybrid of the floating gate and floating trap devices. It possesses the quick programming time and source sharing of the floating gate technology along with the retention endurance, durability, and low operating power and voltages of the SONOS technology. Research has shown NROM™ devices can be easily scaled with CMOS technology, possess short write (10μs) and erase times (10ms) [26], use low programming voltages (5-7 V), share source lines to achieve cell sizes of 2.5- $5F^2$, exhibits write/erase durability (up to 10^5 cycles), and long term retention (>10

	Bit 1			Bit 2		
	WL (V)	**BL1 (V)**	**BL2 (V)**	**WL (V)**	**BL1 (V)**	**BL2 (V)**
Write	7	4	0	7	0	4
Erase	-7	4	0	-7	0	4
Reverse Read	0-3	0	1.5	0-3	1.5	0

Table 1.4: Applied voltages for the write, erase, and read states for NROM™ 2-bit transistor operation with 90Å equivalent oxide gate thickness. Applied voltages are for the Word Line (WL), Bit Line 1 (BL1), and Bit Line 2 (BL2). The substrate remains grounded in all states. Voltages are for a device with a 36Å tunnel oxide, 30Å silicon-rich nitride, and a 36Å blocking oxide. All terminals are grounded for a device in the idle/retention mode.

years) due to the increased tunnel oxide thickness as compared to their SONOS counterparts [24]. This concept has been demonstrated on NSONOS transistors with gate lengths as small as 0.13 μm [27].

Reliability issues involving damage to the gate dielectric and residual charge buildup in the charge-storage layer, due to mismatch of the hole and electron injection

areas, have led researchers to develop new methods for programming and erasing 2-bit NVSM's. The first alternative method uses hot electron injection to write a device and uniform tunneling to erase the device [26]. Hot electron injection provides fast write times and storage of electrons in two distinct, localized areas. However, residual charge buildup occurs over time due to small lateral differences along the channel of the location of electrons injected during the write operation and the location of holes injected during device erasure. This residual charge causes the threshold voltage window to degrade after as few as 100 write/erase cycles [28].

A uniform channel erase can also be used to cycle the NROM™ structure. During a channel erase, holes tunnel from the entire length of the silicon substrate, ensuring that the nitride is completely erased [26]. Using a uniform channel erase will erase both data bits, limiting the flexibility of this device, but allowing for Flash operation. The drawbacks to uniform channel erasure include increased erase gate voltage and a reduced tunnel oxide thickness to achieve hole tunneling.

Another alternative programming method utilizes hot hole injection to *write* and Fowler-Nordheim tunneling of electrons to *erase* the device [29]. Typically programming the device with electrons is referred to as *writing* the device and programming with holes is referred to as *erasing* the device. However, C.C. Yeh, et al, use the opposite convention. Their arrangement uses hot hole injection for fast programming times and 2-bit memory operation. These researchers contend Fowler-Nordheim (FN) tunneled electrons occupy traps deep in the nitride and electrons in shallow traps will be drawn out by the high electric fields. Programming with channel

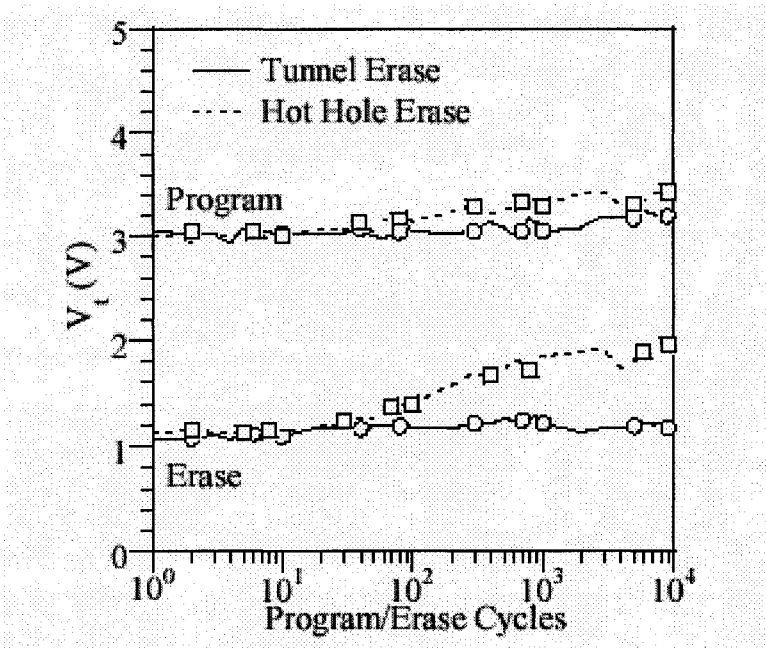


Figure: 1.11 Write/erase endurance with CHE/tunneling and CHE/HHI in an NROM™ VNSM. Hot hole injection results in electron build up over the channel [26].

hot electron (CHE) injection randomly captures the electrons in deep and shallow traps. The electrons in shallow traps will easily escape during charge retention [29]. However, hot hole injection causes greater damage to the gate dielectric during programming than channel hot electron injection, resulting in degraded retention performance after write/erase cycling [26]. These varied programming systems represent the versatility of the NROM™ device and its suitability for use in space and military applications.

1.2.3 Novel Concept NVSM's

In addition to storing charge in a film in the gate dielectric of a transistor, there are several competitors to the SONOS technology for use in space and military

applications. These technologies utilize the change in polarity, magnetic direction, or physical composition of a film in the gate dielectric of a transistor. Ferroelectric memory (FRAM) uses a ferroelectric material $PbZr_XTi_{1-X}O_3$ (PZT) or $SrBi_2Ta_2O_9$ (SBT) film. Applying a positive electric field to the film causes the polarization of the film to shift positively. Conversely, a negative voltage will cause a negative shift in polarization [30].

These films are used in capacitors as replacements for the oxide gate dielectric. The polarity of these films can be read by applying a voltage to the gate and sensing a high (logic 1) and low (logic 0) read current [31]. This technology features extremely fast read (55 ns) and write (100ns) times, high endurance (10^8-10^{13} write/erase cycles) and 10-year retention at 150C [32-34]. However, FeRAM devices require a large amount of chip area and are not compatible with standard CMOS processing [33].

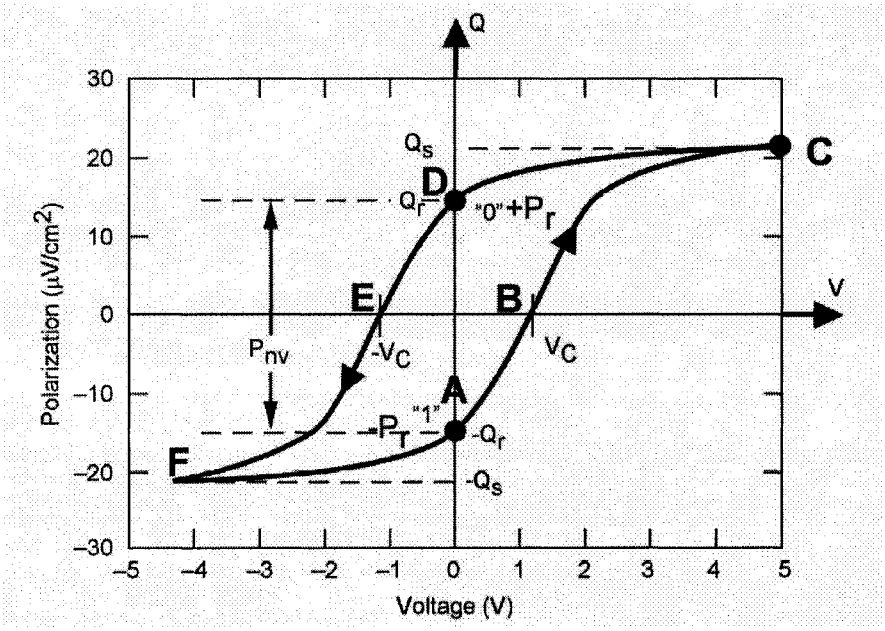


Figure 1.12: Hysteresis loop of a ferroelectric capacitor [35].

Another novel NVSM is magnetic memory (MRAM). This technology utilizes a ferromagnetic film which can switch magnetic direction via an applied electric field. Pairing this film with another ferromagnetic film with a fixed magnetic direction in a gate stack of a transistor or capacitor can create a magnetic memory device. If the magnetic orientations of the free and fixed films are in the same direction, the junction will have a low resistance. A high resistance will be obtained if the magnetic orientations are opposite [36, 37]. Applying a read voltage will produce a high read current if the resistance is low and a low read current for a high resistance, effectively reading out the memory sate [38].

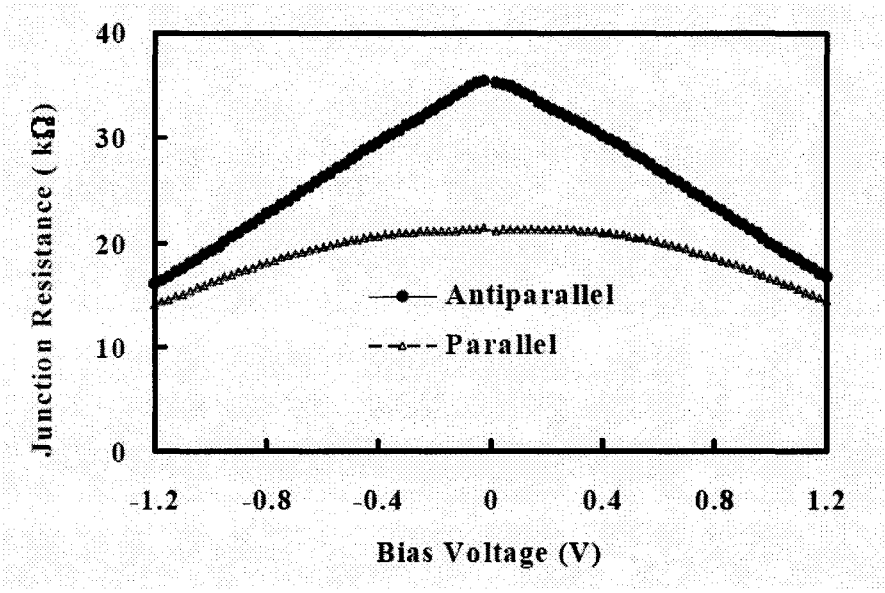


Figure 1.13: Bias voltage dependence of the anti-parallel and parallel resistances for a magnetic tunneling junction [39].

The MRAM technology features extremely low programming voltages (1-3.3V), nanosecond read and programming times, 10 year data retention, infinite cycling, and high density. However, this technology is plagued by high currents required to switch

the magnetic orientation of the free film, thermal relaxation where the film spontaneously switches its magnetic orientation, and incompatibility with standard CMOS processing [40]. In addition, process control has been a major obstacle to manufacturing cost-effective MRAM devices.

Finally, the phase change memory (PRAM) alters the physical state of a film in the gate dielectric. PRAM uses 'Chalcogenide' films which are composed of group VI elements, such as GeSbTe (GST) alloys. The application of heat changes the film from an amorphous state, characterized by a dull appearance and high electrical resistance

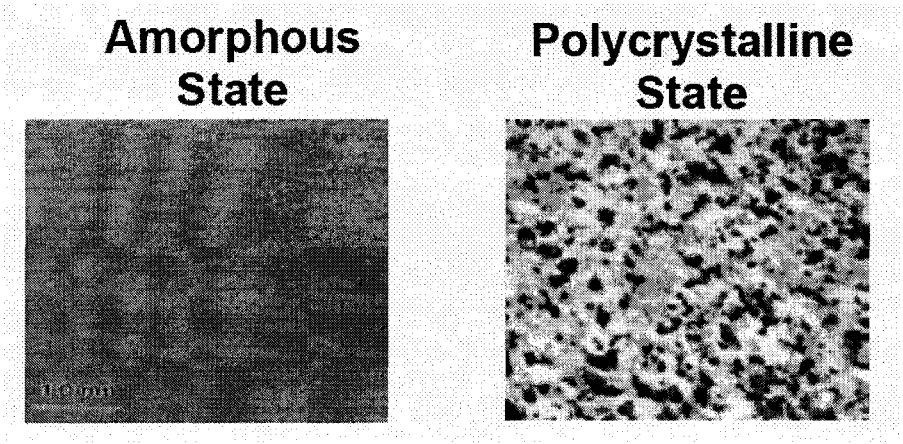


Figure 1.14: Transmission Electron Microscope images of the two phases of a GeSbTe alloy [41].

(mega-ohms), to a polycrystalline state, characterized by a highly reflective appearance and low electrical resistance (kilo-ohms) [41]. This technology is currently in use in CD/DVD-RW media where difference in reflectivity between the two states is used to determine the memory state.

PRAM memory devices use a Chalcogenide film in the gate of a transistor or capacitor. The Chalcogenide film is locally heated and, depending on the heating time,

becomes either amorphous or crystalline, altering the resistance of the junction [42]. The memory state is sensed by applying a read voltage and measuring the resulting current. This technology features nanosecond programming times, greater than 10 year retention, 10^{12} programming cycles with low current, high density, moderate scaling, and radiation hardness up to 2Mrad [40, 43, 44]. However, PRAM films often become frozen in a specific state when high current is applied, thermal cross-talk heats neighboring bits during programming, and high programming currents (mA) are required [45, 46].

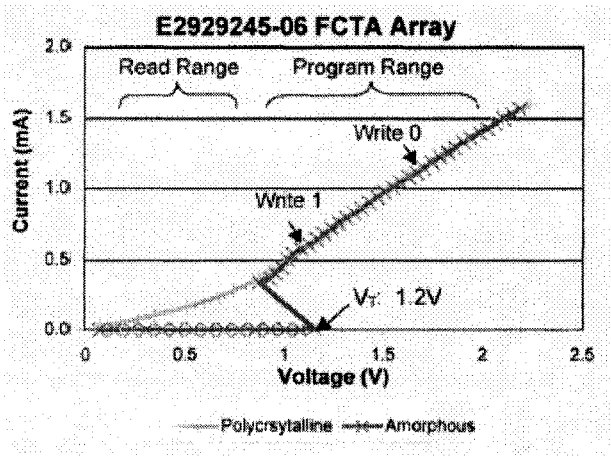


Figure 1.15: Memory effect of a PRAM [46].

1.3 Scope of the Dissertation

Nonvolatile semiconductor memory devices operating off of fixed battery supplies in space and military applications require low-power programming and long-term data retention. The memory device must not draw more than a milliamp of current. Scaling programming voltages to below 5V aids in reducing the power dissipated by the

device. However, rapid programming of a memory is useless if the device cannot retain the data for up to 10 years without a power supply. These two requirements are central to the design of a radiation-hardened memory device suitable for space and military applications. This dissertation focuses on scaling the programming voltages and times through utilizing a write mechanism adapted from the floating gate technology and developing a method to forecast 10-year retention of a SONOS memory transistor. This includes characterizing the write and erase processes of scaled SNROM devices using channel initiated secondary electron (CHISEL) injection, NROM™ programmed with channel hot electron (CHE) injection, characterizing damage to the interface of the silicon bulk and silicon-dioxide layers in various NVSM devices, and measuring and extrapolating data retention of SONOS devices at elevated temperatures.

A low power substrate injection nitride based (SNROM) device is presented, for the first time, and programming using channel initiated secondary electron (CHISEL) injection is analyzed and modeled in Chapter 2. The SNROM device is a short-channel transistor with an oxide-nitride-oxide charge storage gate dielectric. The CHISEL concept is demonstrated by programming SNROM devices with a negative bulk bias and positive drain and gate biases and measuring the obtained memory window. The threshold voltage of the transistor is modeled and extracted from the measured channel current during CHISEL programming. Impact ionization current at the drain side of the device is measured and correlated to the injected current and threshold voltage shift in the SNROM device.

Chapter 3 presents retention measurements performed on SONOS devices at

temperatures ranging from room to 250°C. Retention models based on the density of traps in the nitride are employed to model the decay rate of the threshold voltage for a specific temperature and time. Data retention is characterized using charge decay rate modeling. Scaled SONOS transistor fabrication techniques and design concerns are outlined in Chapter 4. Techniques for fabricating sub-micron SONOS transistors using a source-drain drive-in anneal to create a 0.5 μm channel length are presented and simulated.

Chapter 5 concerns the electrical characterization of SONOS devices. Oxide degradation during charge injection is studied using charge pumping measurements. Charge pumping is performed with both square and triangular voltage pulses and the two methods are compared and contrasted. Linear voltage ramp, capacitance, and variable frequency charge pumping measurements are utilized to characterize SONOS gate dielectrics. Chapter 6 presents results of electrical measurements, performed in the course of this research, on SONOS, NROM™, and SNROM devices for qualification for space and military applications. Conventional NVSM devices are programmed at different voltages, interface quality and initial memory window are measured up to 10^6 write/erase cycles, and data retention is measured at temperatures up to 250°C. Threshold voltage decay trends are studied and a standard procedure for forecasting 10-year charge retention is formed. Chapter 7 presents conclusions based on the analysis, measurement, and modeling of NVSM devices in this dissertation. Recommendations for future research are presented.

Chapter 2

Channel Initiated Secondary Electron Injection (CHISEL) Programming in SNROM Transistors

Programming SONOS devices for space and military applications requires low currents, rapid charge transport, and minimal power dissipation. Conventional programming methods such as Modified Fowler-Nordheim (MFN) and direct tunneling (DT) have been introduced in Chapter 1. Tunneling mechanisms require very little current but necessitate relatively high programming voltages, which have reached the scaling limit. The gate voltage determines the electric field in the gate stack and the speed of the programming operation. SONOS transistors fabricated with conventional SiO_2 and Si_3N_4 films cannot be further vertically scaled due to the thickness requirements of the oxides, necessary for insulating during the retention mode, and the nitride, requiring a certain dimension in order to have enough traps to store a sufficiently large volume of charge. In addition, programming using tunneling requires several milliseconds to obtain an adequate memory window.

2.1 CHISEL Injection in Floating Gate Transistors

Floating gate devices are programmed by applying high vertical and lateral electric fields in order to impart enough energy to excite electrons flowing in the channel over the bottom oxide barrier. Channel hot electron injection has recently been

adapted for use in SONOS transistors, with the pseudonym NROM™, meaning nitride based read only memory [23]. The NROM™ structure is a short-channel SONOS device which is programmed using hot electron and hot hole injection. Recently, researchers working on floating gate devices have begun pioneering a newer programming mechanism known as CHannel Initiated Secondary ELectron (CHISEL) injection [47]. This programming mechanism is not only more efficient and faster than CHE injection, but it requires smaller channel currents and programming voltages [48]. Programming with CHISEL injection has not been widely studied in published literature for use in SONOS NVSM's for use in space and military applications.

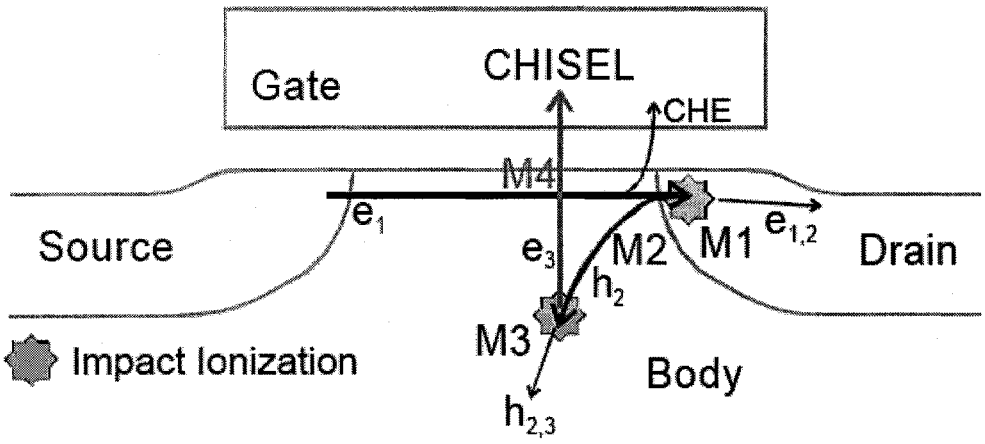


Figure 2.1: Programming a floating gate transistor with CHISEL injection. Impact ionizations create hole currents ($h_{2,3}$) and electron currents ($e_{2,3}$) [49].

CHISEL programming of a floating gate transistor is similar to channel hot electron injection, outlined in Chapter 1. Positive biases are applied to the gate and the drain junctions and a negative bias is applied to the substrate. The drain voltage is typically equal to or greater than the gate voltage. An inversion layer is formed in the channel due to the gate voltage, and pinches off towards the drain junction. Electrons

flow in the channel from the source to the drain (e_1). A small portion of the electrons gain enough energy to become 'hot' (CHE) and inject into the gate dielectric due to the high lateral electric field in the channel. At the drain junction, it is believed that the electrons traveling through the channel undergo an impact ionization at the region of the inversion layer pinch-off [49, 50].

This impact ionization creates an electron-hole pair (e_2, h_2), where the electron (e_2) is either accelerated to the drain or into the channel to propagate further impact ionizations. The hole (h_2) is accelerated towards the substrate due to the negative bulk bias, where the hole may trigger a secondary impact ionization, creating another electron-hole pair (e_3, h_3). The hole (h_3) is accelerated to the bulk by the vertical electric field and a portion of the ionized electrons (e_3) are accelerated into the channel and injected into the gate dielectric [49].

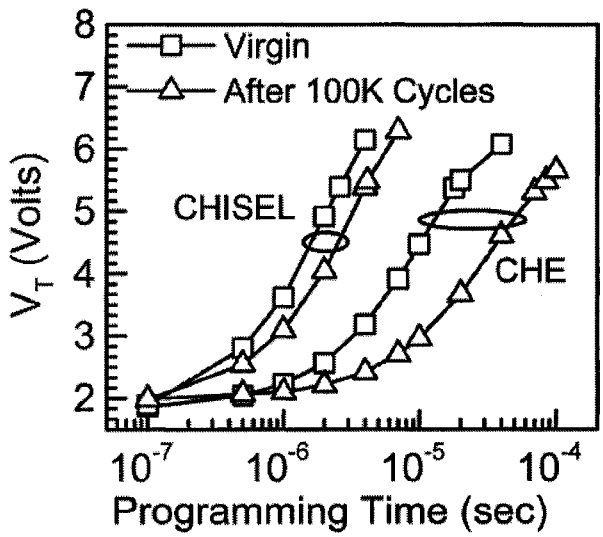


Figure 2.2: Floating gate memory programmed with CHISEL and CHE injection. CHISEL injection is one decade faster than CHE injection [51].

Therefore, the drain current may be expressed as a combination of the channel current (e_S) and electron currents from the primary (e_1) and secondary (e_2) impact ionizations.

There are several advantages to programming a floating gate transistor with CHISEL versus programming with CHE [52]:

1) Programming times are one decade faster
2) Programming voltages are comparable or smaller
3) Channel current required for charge injection is smaller
4) Damage to oxide films in gate dielectric is comparable or less

The current injected into the gate of a floating gate transistor has been extensively modeled and may be expressed as a function of the drain and bulk currents

$$\frac{I_G}{I_D} = C(E_{ox}) \cdot \left(\frac{I_B}{I_D}\right)^{\alpha} \tag{2.1}$$

where $\alpha = \Phi_B / \Phi_{II} \approx 2.5$, where Φ_B is the Si-SiO_2 oxide barrier height and Φ_{II} is the effective threshold for impact ionization [47].

2.2 CHISEL Injection in SNROM Transistors

This dissertation presents unique research on a novel short-channel (10 x 0.22μm) NVSM device programmed with CHISEL injection. The gate dielectric is composed of a 36Å tunnel oxide, a 30Å oxygen-rich oxynitride, and a 36Å blocking oxide. These devices will be referred to as substrate injection nitride based (SNROM)

devices. The tunneling oxides of these devices are thicker than tunneling oxides in conventional SONOS transistors, which are typically 15 to 20 Å thick. The relative dielectric constant for the oxynitride film is 5.07, with an equivalent oxide thickness of 23.5Å. These transistors are intended to be programmed with a channel hot electron (CHE) write and a hot hole injection (HHI) erase. It should be noted that these transistors are not optimized for CHISEL injection. Future SNROM structures, film thickness, and ion implants can be optimized for CHISEL programming in order to provide enhanced results.

Modeling short-channel SNROM transistors programming with CHISEL and CHE injection begins with demonstrating memory operation. SNROM devices (10 x 0.22 μm) were written with CHISEL injection by applying +3V to the gate, +2.5V to

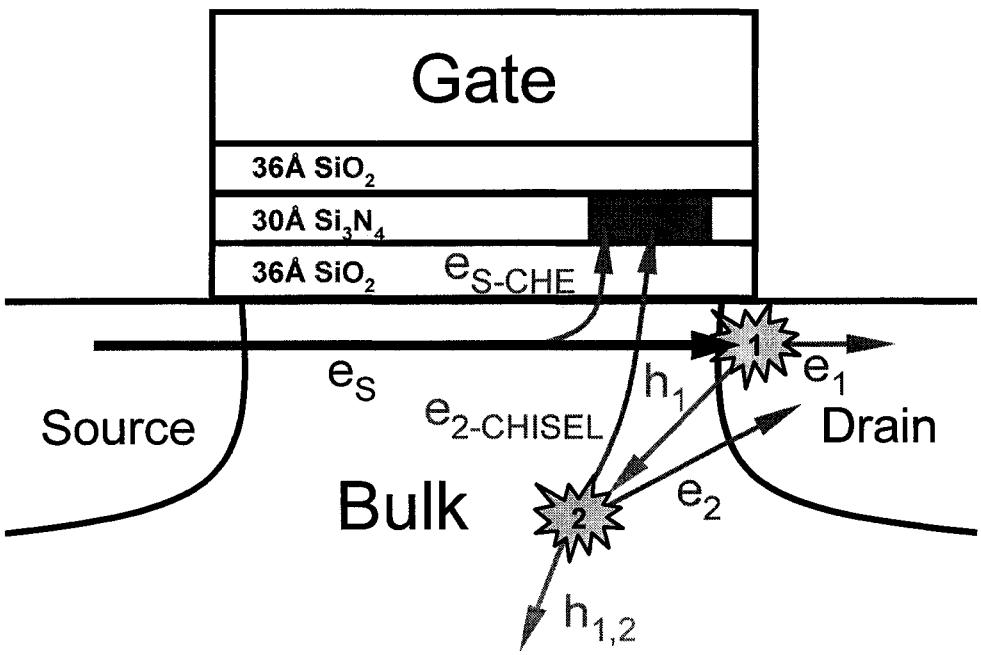


Figure 2.3: Vertical cross-section of a short-channel SNROM transistor during CHISEL injection. The drain current is composed of electron currents ($e_{S,1,2}$).

the drain, -4V to the substrate, and grounding the source. These transistors were erased using HHI by applying -7V to the gate, +4V to the drain, and grounding the source and substrate. The programming times for both write and erase were 5 ms, each. SNROM devices programmed under these conditions produced a 2 – 2.7 volt memory window. Programming, read, and retention voltages and timing were controlled using the NVSM characterization setup described in section 5.2.

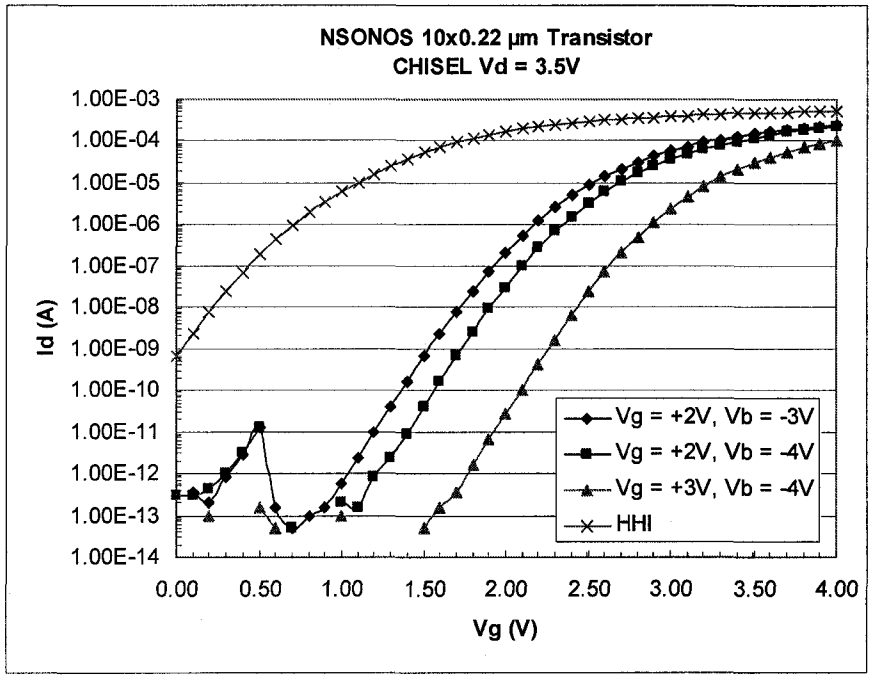


Figure 2.4: Drain current versus gate voltage measurements depicting SNROM programming with CHISEL injection at various gate and substrate biases. Drain bias during I-V sweeps is 50mV. The device is preconditioned with a hot hole injection (HHI) erase at the drain (red). The device is programmed with 2-3V applied to the gate, -3 to -4V applied to the substrate, and a drain bias of +3.5V (blue curves).

CHISEL programming of an SNROM transistor occurs in the same manner as CHISEL injection in a floating gate transistor, described in section 2.1, with the exception of the electron current from the secondary impact ionization. Electrical

characterization, discussed later in this section, has shown the drain current is approximately equal to the channel current (e_S) and electron currents from the primary (e_1) and secondary (e_2) impact ionizations. Therefore, we model the electron current from the secondary impact ionization as $e_{2\text{-CHISEL}}$, which is the portion of the ionized electrons which are injected into the nitride, and e_2, which is the remainder of the electrons which are captured by the drain.

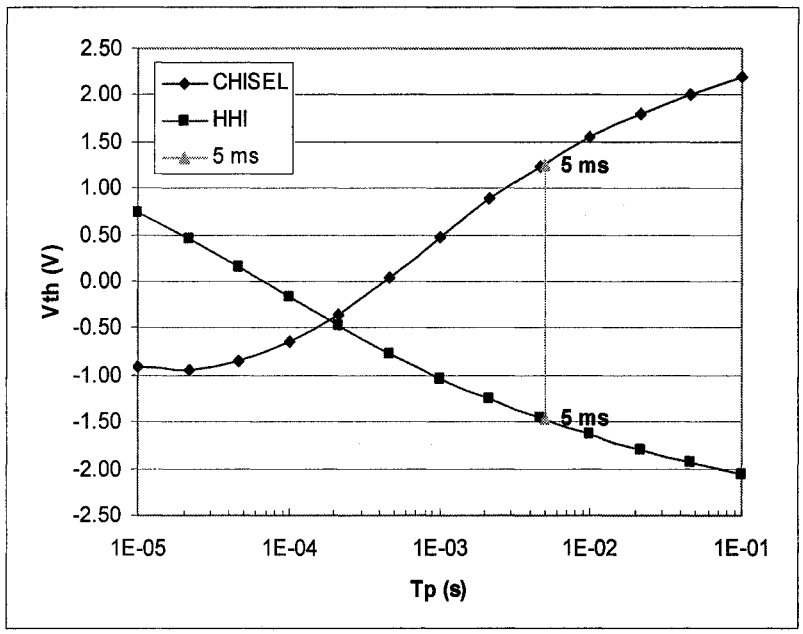


Figure 2.5: Threshold voltage versus programming time for a short-channel 10 x 0.21 μm SNROM transistor programming with CHISEL and HHI mechanisms. CHISEL write is performed by applying +3V to the gate, +3.5V to the drain, -4V to the substrate, and grounding the source for 5ms. HHI erase is performed by applying -7V to the gate, +4V to the drain, and grounding the source and substrate for 5ms.

The threshold voltage of the transistor may be modeled from the drain to source current of an SNROM transistor programmed with CHISEL injection, giving insight into the current-multiplication and impact ionization occurring within the channel and

substrate of the device. The SNROM device is programmed using an HP4145 parameter analyzer. The gate voltage is fixed at +3 V, -4 V is applied to the substrate, and the source is grounded. The channel current, measured at the drain, is monitored as the drain voltage is swept from zero to +3.5 V. The measurement is more accurate when the source current is measured, since this current does not include the electrons escaping to the drain from the electron-hole pairs generated by impact ionizations, a critical component for CHISEL injection. As the drain voltage is swept, the channel current increases, as is expected for a conventional MOS transistor. However, when the lateral electric fields, created by the drain bias, are high enough, impact ionizations begin to occur at the drain junction and in the bulk, and CHISEL injection occurs.

2.2.1 Modeling SNROM Threshold Voltage Shift during CHISEL Injection

During CHISEL injection, electrons are excited over the silicon-dioxide barrier and trap in the nitride, creating a positive shift in the threshold voltage, where the threshold voltage is expressed as [21]

$$V_{th} = \Phi_{GS} + 2\Phi_F + \frac{\sqrt{4\varepsilon_{si} q N_B \Phi_F}}{C_{eff}} - Q_N\left[\frac{x_{OB}}{\varepsilon_{OX}} + \frac{x_n}{2\varepsilon_n}\right] - \frac{Q_{SS} + Q_{it}}{C_{eff}} \qquad (2.2)$$

where Φ_{GS} is the gate to semiconductor work function, Φ_F is the Fermi potential, ε_{SI} is the permittivity of the silicon substrate, q is the charge of an electron, N_B is the channel doping density, C_{eff} is the effective capacitance, Q_N is the charge trapped in the nitride, Q_{SS} is the charge at the Si-SiO_2 interface, and Q_{it} is the charge stored in interface traps

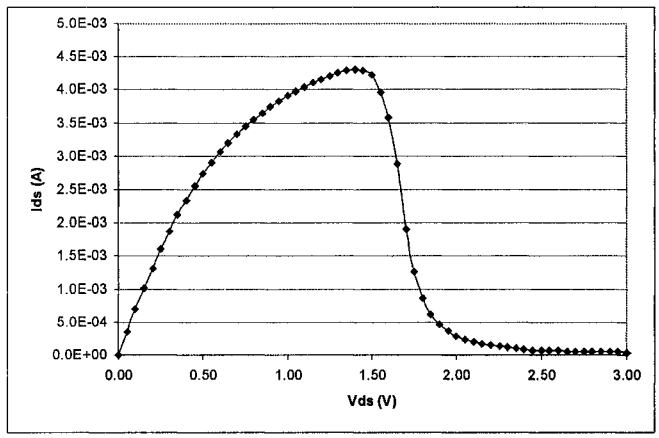


Figure 2.6: Drain current of an SNROM transistor (W/L = 10 x 0.22 μm) measured during CHISEL programming with 3V applied to the gate and -4V applied to the bulk.

in the bottom oxide [21].

The threshold voltage increases until it approaches the value of the fixed gate voltage, causing the channel to turn off, evident in the decrease in channel current, see Fig. 2.6. This effect allows us to extract and monitor the change in the threshold voltage of the device as the drain voltage is swept. The channel current may be expressed as

$$I_{DS} = \frac{1}{2}\beta_0 (V_{GS} - V_{TH})^2 (1 + \lambda V_{DS}) \tag{2.3}$$

where V_{GS} is the gate voltage, V_{TH} is the threshold voltage, V_{DS} is the drain bias, λ is the inverse of the Early voltage, and $\beta_0 = \mu_n C_{OX} W/L$ where μ_n is the effective mobility for holes, and W and L are the width and length of the transistor, respectively. Solving Eq. 2.3 for the threshold voltage yields

$$V_{TH} = V_{GS} - \sqrt{\frac{2I_{DS}}{(1 + \lambda V_{DS})\beta_0}} \tag{2.4}$$

The extracted SNROM threshold voltage shows the CHISEL injection occurs when the lateral electric field is high enough to create hot carriers and the subsequent impact ionizations. The current from the impact ionizations is measured during the CHISEL injection. This current is the sum of the electrons which are not injected from the channel current, primary impact ionization, and electrons not injected following the secondary impact ionization.

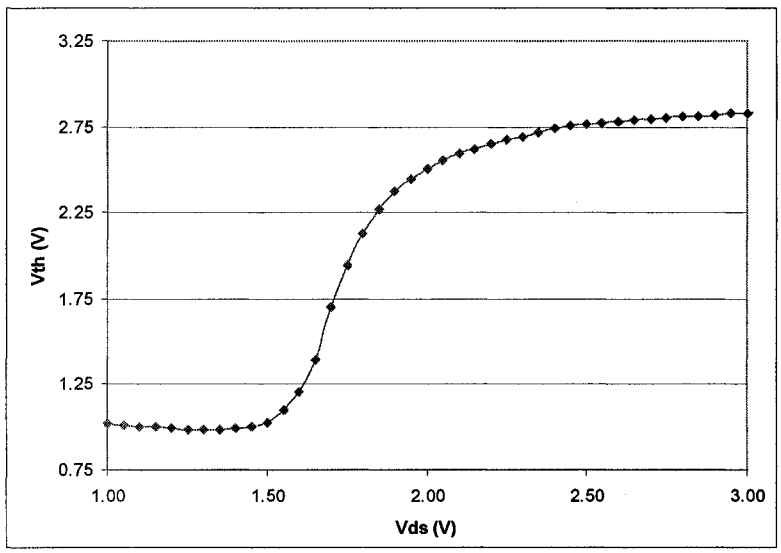


Figure 2.7: Extrapolated threshold voltage shift of the SNROM transistor during CHISEL programming.

The charge per unit area injected and trapped in the nitride may be calculated from the threshold voltage shift. The threshold voltage shift as a function of injected charge may be expressed as

$$\Delta V_{TH} = \Delta Q'_N \left(\frac{X_{OB}}{\varepsilon_{OX}} + \frac{X_N}{2\varepsilon_N} \right) \tag{2.5}$$

where Q_N is the charge density in the nitride, X_{OB}, ε_{OX}, X_N, and ε_N are the thickness and permittivity of the blocking oxide and nitride layers, respectively. Since the charge is not injected over the entire channel length, the charge density is expressed as $Q'_N = Q_N\left(\frac{d}{L}\right)$ where L is the channel length and d is the length of the region where the charge is injected. This scaling factor represents the equivalent effect on the threshold voltage of a smaller volume of charge injected over a localized region of the channel. The charge is then calculated by multiplying the charge density by the area of the SNROM device.

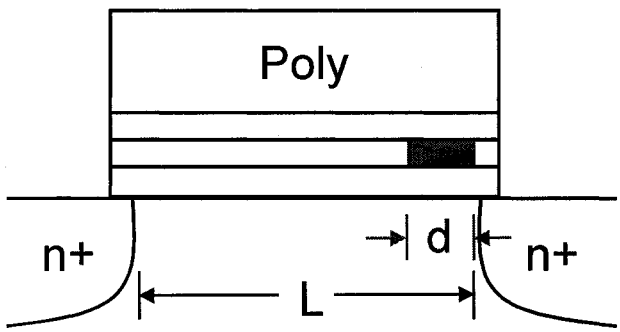


Figure 2.8: Lateral cross-section of a SNROM transistor depicting the CHISEL charge injection region (red).

The threshold voltage shift has been calculated previously from the channel current during programming. The change in the threshold voltage is used to numerically compute the amount of charge injected into the nitride, Fig. 2.9. The current injected into the nitride may be expressed as change in charge per unit time

$$I_{inj} = \frac{dQ_N}{dt} = \frac{dQ_N}{dV_{DS}} \cdot \frac{dV_{DS}}{dt} \tag{2.6}$$

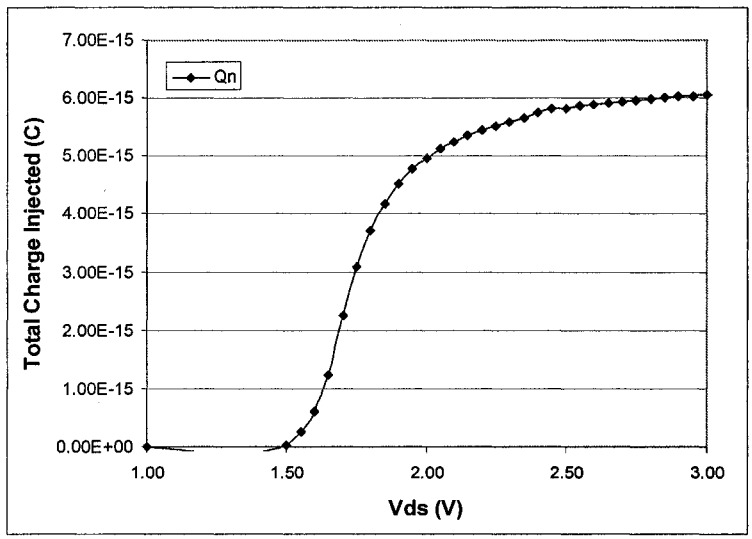


Figure 2.9: Trapped charge required to produce observed V_{TH} shift.

During the SNROM CHISEL injection measurement, the drain voltage is swept from 0 to 3.5 volts in steps of 0.1V. The drain voltage step function is approximated as a ramped voltage with a slope expressed as alpha, $\alpha = \frac{dV_{DS}}{dt}$. The ramp rate (slope) used for CHISEL injection was 500 mV/second. Substituting the ramp rate into Eq. 2.6, we may express the injected current as

$$I_{inj} = \alpha \frac{dQ_N}{dV_{DS}} \tag{2.7}$$

The change in the threshold voltage, extrapolated from the channel current during charge injection, is substituted into Eq. 2.5 to calculate the total charge injected. The channel length is 0.22 μm and the length of the charge injection region is approximate from literature as 450 Å [53]. The slope of the charge with respect to the

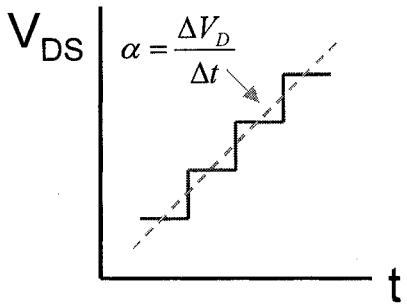


Figure 2.10: Drain voltage stepped (solid) during SNROM CHISEL programming approximated as a ramped voltage (dashed).

drain voltage is substituted into Eq. 2.7, with the ramp rate (500 mV/second), to calculate the current trapping in the nitride which will shift the threshold voltage as measured. We will refer to this required injection current as the 'calculated injection current'. The total injected charge, 5.83×10^{-15} C, may be divided by the charge per electron, 1.6×10^{-19} C/e$^-$, to calculate that 36,000 electrons are injected into the nitride

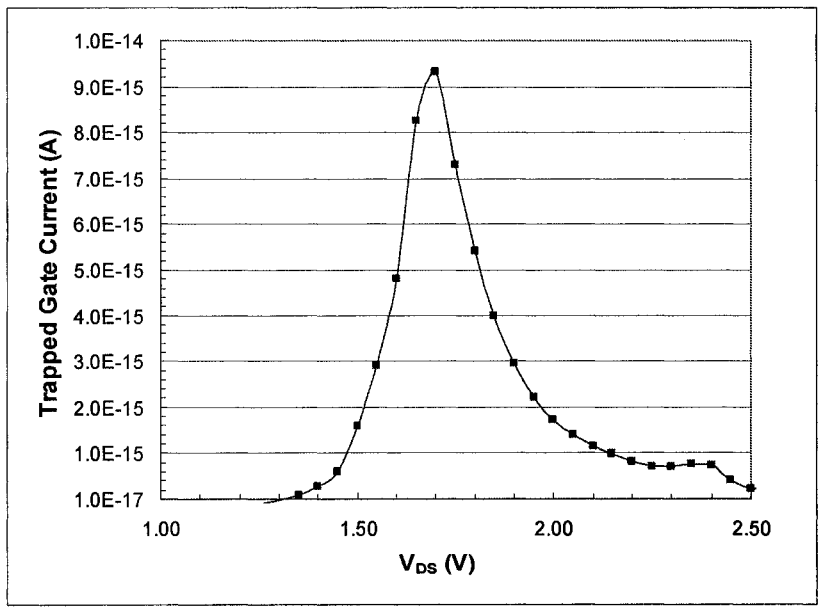


Figure 2.11: Current injected into the gate dielectric and trapped in the nitride extrapolated from the threshold voltage shift.

during the CHISEL programming. In addition, the total current injected into the nitride, 58 fA, is very low. These results demonstrate the low-power operation of the SNROM.

2.2.2 SNROM CHISEL Impact Ionization Current

The calculated current injected into the nitride may be correlated to the ionization current measured at the drain terminal. The ionization current, may be expressed as the difference between the source and drain currents during DC

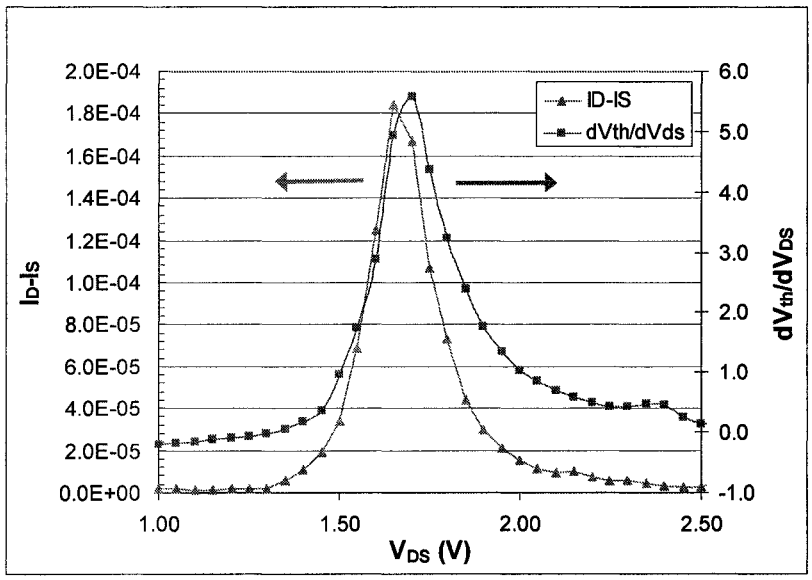


Figure 2.12: Measured impact ionization current (triangles) and slope of threshold voltage shift (squares) during SNROM CHISEL injection.

programming of the SNROM, since the majority of the electrons from the impact ionizations are not injected but are captured by the drain terminal.

$$I_{Ion} = |I_D| - |I_S| \tag{2.8}$$

Plotting the measured impact ionization current versus the change in the threshold

voltage with respect to drain voltage, we see that the impact ionization current spikes when the threshold voltage changes, Fig. 2.12. The threshold voltage of the SNROM transistor is shifted when the electrons generated by the impact ionizations are injected into the nitride.

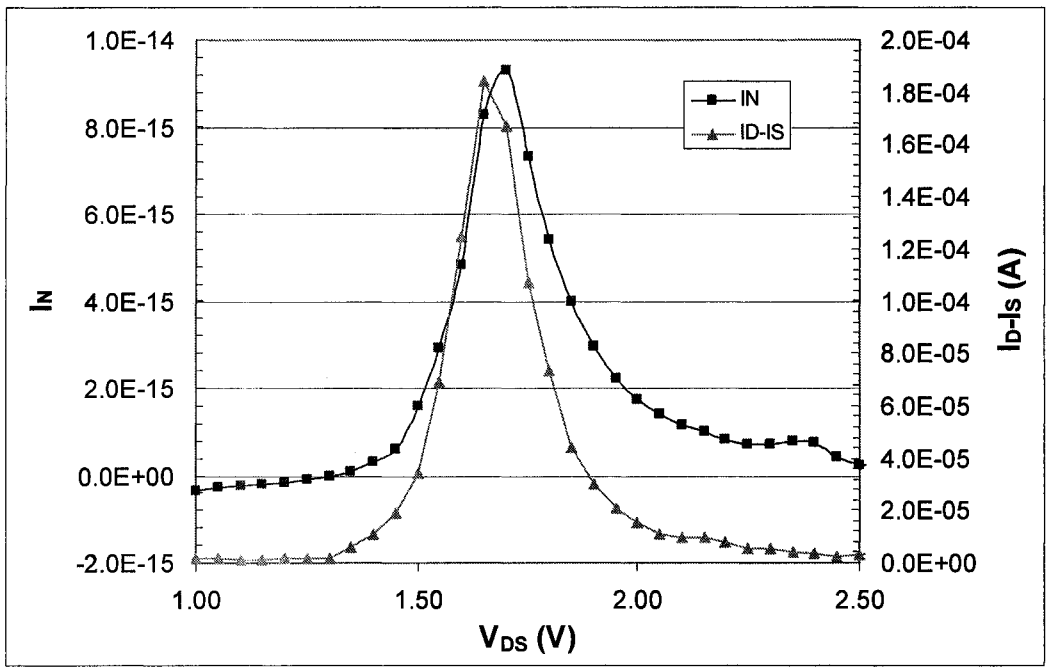


Figure 2.13: Calculated trapped injection current (dark squares) and measured impact ionization current (light triangles) in the SNROM device.

As already discussed, the electron current from the impact ionizations is proportional to the current injected into the nitride. This relationship is demonstrated by plotting the calculated injection current with the measured impact ionization current while the drain bias is varied, Fig. 2.13. Both currents peak at a drain bias of 1.7V, where the CHISEL injection and threshold voltage change occur. This provides a visual correlation between the ionization current, the current injected and trapped in the nitride, and the shift in the threshold voltage.

Chapter 3

Modeling Density of Traps in Oxynitride Charge Storage Films in Scaled SONOS Transistors

In order to achieve improved SONOS charge retention, oxygen-rich silicon nitride films called oxynitrides (SiO_xN_y) have been used in the fabrication of the devices tested in this dissertation. Kapoor et al investigated oxynitrides as storage media, replacing silicon-rich nitride layers. Oxynitride films have lower trap densities compared to silicon-rich nitrides [54], which is verified in this dissertation. Oxynitride films demonstrate reduced Coulombic repulsion between trapped charges. Coulombic repulsion contributes to these charges coming free from their bonds and creating interface traps, which are used as stepping-stones for charge stored in the nitride to escape to the silicon substrate.

A decrease in the interface trap density has been correlated with an increase in oxygen content in the oxynitride film [54]. Thin oxynitride films provide an improved gate dielectric compared to silicon nitride films. The addition of oxygen impurities in the oxynitride reduces the trap densities by compensating the silicon-dangling bonds at the oxide-oxynitride interface and results in a decreased charge decay rate and current conduction and improves the SONOS memory properties [55]. Interface states must be minimized in SONOS memory devices to achieve superior device performance, such as lower programming voltages, and shorter write and erase times [56].

3.1 Historical Perspective

Since the invention of the nitride based memory structure in 1967 [57], there has been a great deal of experimental and analytical work focusing on understanding data retention characteristics and charge loss mechanisms. The following is a summary of charge decay mechanisms in MNOS and SONOS devices proposed in literature [10]:

1) Lundkvist, *et al,* [58] have modeled the charge decay behavior of MNOS transistors at room temperature as back tunneling of trapped charge stored in the nitride. The charge in the nitride is assumed to be stored in an area extending from the oxide-nitride interface to about 10 nm into the nitride. This trapped charge tunnels through the tunnel oxide to the silicon substrate due to the internal field present in the retention mode.
2) In order to describe the increased rate of loss of charge at elevated temperatures, Lundkvist, *et al.*, [59] proposed that the thermal emission of trapped charges from the nitride is another path of charge decay. It is assumed that the nitride traps are continuously distributed in energy from depth ϕ_{min} to ϕ_{max} in this model.
3) Lehovec, *et al.*, [60] assumed that the retention mode loss is limited by the nitride conduction through Poole-Frenkel detrapping. Using this assumption, a simple analytical expression for charge retention in MNOS devices was derived. A mono-energetic trap model was used in their work. Using a numerical data fit for the results, at a trap depth of 1.5 eV, a Poole-Frenkel coefficient of about 6 x

10^{-4} $cm^{1/2}V^{-1/2}eV$ and an effective escape attempt rate factor of 1.2 x 10^8 sec^{-1} were calculated.

4) Trap-assisted charge injection was revealed as a major factor in MNOS [61] and SONOS [62] device operations, particularly for thin tunnel oxides at low fields (1-4 MV/cm). Charge tunnels through part of the nitride and the entire thickness of the oxide using a nitride trap as a "stepping stone". It has been revealed that these traps are shallow and are not the same as the traps responsible for charge storage in the nitride, which are much deeper in the nitride.

5) White, et al., [18] have explained the charge retention for thin-oxide MNOS memory transistors in terms of the direct tunneling of charges from traps in the nitride to the Si-SiO_2 interface states. A mono-energetic donor and acceptor trap level, localized spatially at the nitride-oxide interface, was considered. They observed a slower rate of charge loss to acceptor Si-SiO_2 interface states as compared with donor Si-SiO_2 interface states. This is interpreted as a result of a larger increase in acceptor interface states near the conduction band edge as compared with donor interface states near the valence band edge.

6) Wang and White [63] have expounded on the Yang and White model by examining the contribution of trap-to-band tunneling to the charge decay rate.

3.2 Trap Density Modeling

The decay of the charge stored in the nitride layer at room temperature has been

modeled by a number of investigators, such as White and Cricchi [64], Lundkvist, Lundstrom, Svensson [58], Roy and White [65], Kamagaki and Minami [66, 67], and Hu and White [68]. These models invoke back tunneling of charge from the nitride to the semiconductor substrate. An internal field due to trapped charges enhances the process of tunneling.

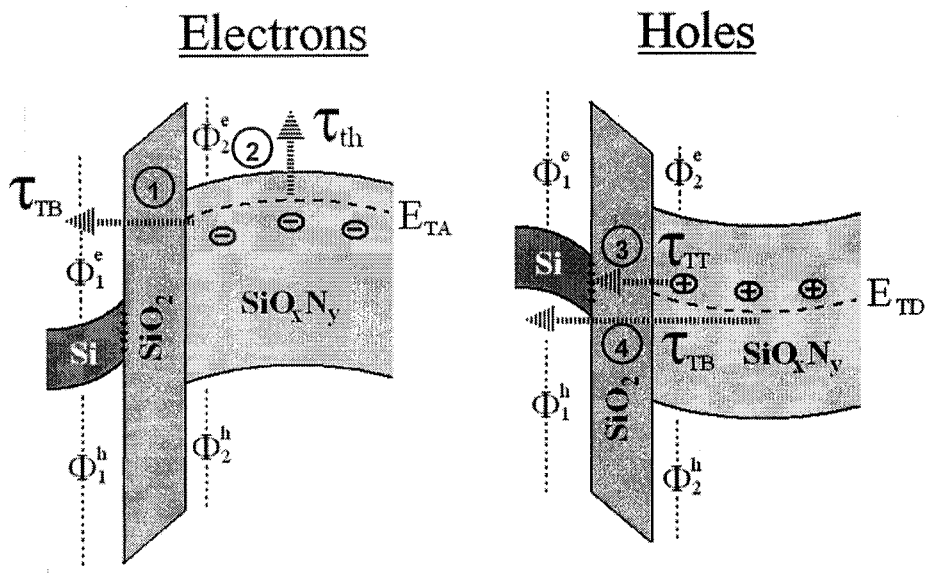


Figure 3.1: Energy-band diagrams depicting charge loss mechanisms of a conventional SONOS transistor in the write (electrons) and erase (holes) state [69].

The characterization of charge trapped in a nitride dielectric at elevated temperatures has been investigated for SNOS devices [59, 70, 71]. Recently, with the advent of scaled SONOS devices, the temperature dependence has been investigated with an amphoteric trap model [18], which attributes the electron and hole charge storage to a silicon-dangling bond. Their results indicate the trapped 'electrons' in the nitride layer are thermally excited at elevated temperatures and 'back tunnel' through an ultra-thin tunnel oxide to the silicon. In contrast, the distribution of trapped 'holes' is

influenced very little with increasing temperature. This result has been explained [18] by suggesting the electron trap activation energy lies closer to the conduction band edge in the silicon nitride than the activation energy for hole traps to edge of the valence band in the silicon nitride as shown in Fig. 3.1.

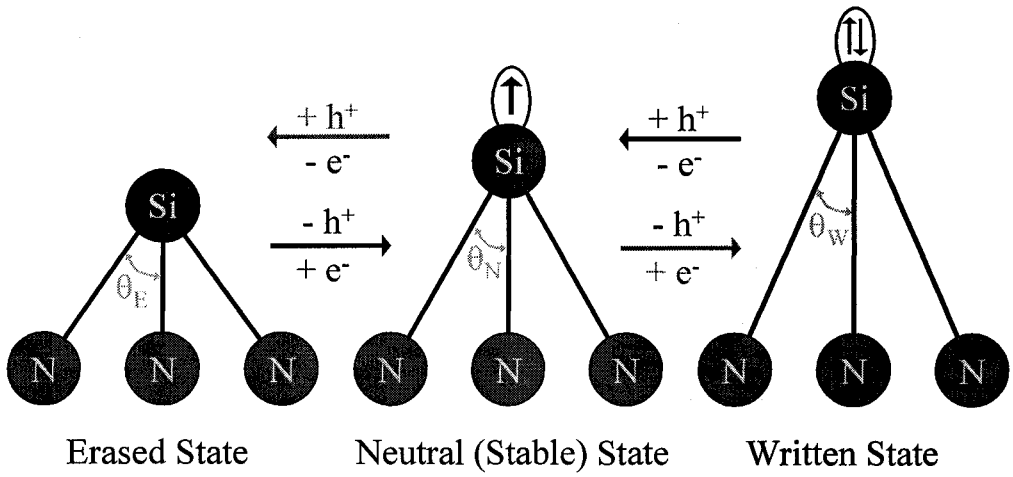


Figure 3.2: Amphoteric trap model applied to Si-N bonds in the silicon nitride film. The bond angle θ decreases and bond length increases as electrons are trapped due to the attractive and repulsive forces.

Using an amphoteric trap model, an expression has been derived [72] for the charge trapped in the nitride

$$\rho_n(x, E_{TA}, t) = -qg(x, E_{TA})f^- \tag{3.1}$$

where E_{TA} is the energy level of the trap, f^- is the trap occupancy function for electrons, and $g(x,E_{TA})$ is the density of traps in the nitride (traps/cm^{3}eV) at a distance 'x' from the tunnel oxide-nitride interface into the nitride. The charge stored in the nitride causes a shift in the threshold voltage of the device, ΔV_{TH}, which can be written as [18]

$$\Delta V_{TH}(t) = -\int_0^{X_N} \int_{E_{GN}} \left(\frac{X_N - x}{\varepsilon_n} + \frac{X_{OB}}{\varepsilon_{ox}} \right) \rho_N(x, E_{TA}, t) dx dE_{TA} \tag{3.2}$$

Assuming a spatially uniform trap density throughout the nitride, $g(x,E_{TA}) = g(E_{TA})$, using Eq. 3.1, and differentiating Eq. 3.2 with respect to t, we can write ΔV_{TH} as [18],

$$\frac{\partial \Delta V_{TH}}{\partial \log(t)} = -2.3 k_B T X_N \left(\frac{X_N}{2\varepsilon_N} + \frac{X_{OB}}{\varepsilon_{OX}} \right) g(E_{TA}) \tag{3.3}$$

where X_N and X_{OB} are the thicknesses of the nitride and blocking oxide, respectively, while ε_N and ε_{OX} are the dielectric constants of the nitride and blocking oxide, respectively. Eq. 3.3 assumes a uniform distribution of nitride traps and the activation energy responsible for the decay rate is [18]

$$E_{TA} = k_B T \ln(AT^2 t) \tag{3.4}$$

where A is a constant given as

$$A = 2\sigma_n \sqrt{\frac{3k_B}{m^*}} \left[\frac{2\pi m^* k_B}{h^2} \right]^{3/2} \tag{3.5}$$

where σ_n is the trap capture cross-section, m^* the effective electron mass in the nitride and 'h' is Planck's constant. These relationships are used to interpret measurements of retention at elevated temperatures for retention times out to 10^5 seconds for the write state of a scaled SONOS device, see Fig. 3.3. Since the thermal activation energy of

electron traps is near the conduction band edge, the electrons are thermally excited from these traps and back-tunnel through the tunneling oxide. Eq. 3.4 shows the trap energy may be probed at a given temperature by measuring the slope of the decay characteristics as a function of retention time. The capture cross-section, σ_n, may have temperature dependence.

The threshold voltage of the device in the write and erase states at elevated temperatures is shown in Fig. 3.3. The results demonstrate the decay rate of the write state is affected by increasing the temperature. The write state threshold voltage decay rate increases as the temperature increases. The erase state decay rate is virtually unchanged. These trends are the same for both the 'oxynitride' and 'silicon-rich' nitride

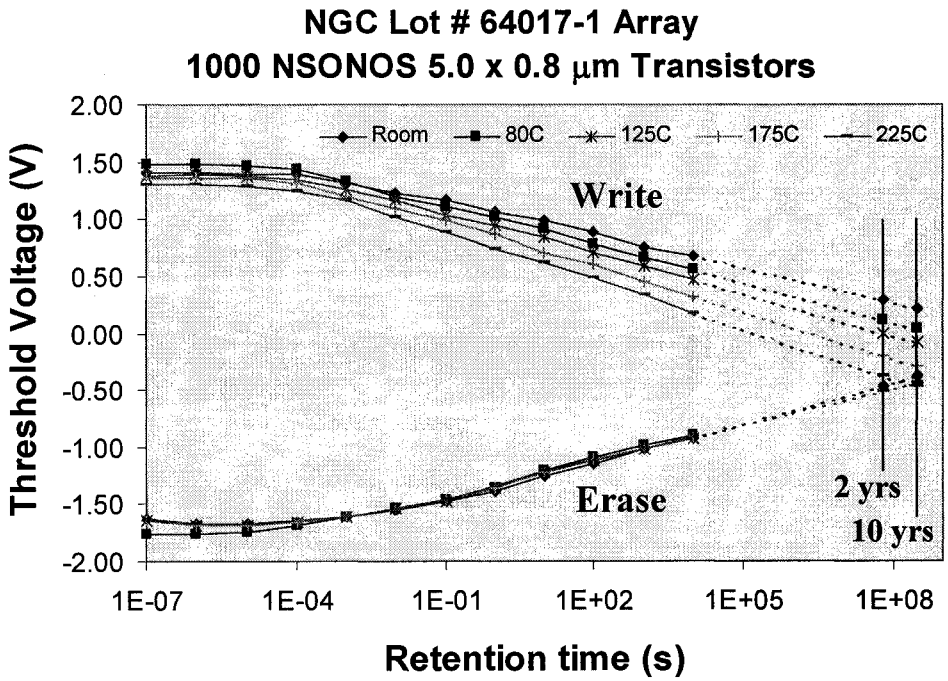


Figure 3.3: Charge retention at elevated temperatures for a SONOS NVSM memory device with ONO dielectric thicknesses (18Å/80Å/40Å) [69].

films. In both films we observe, for write and erase states, a small initial shift of the threshold voltage with increasing temperature [69].

Retention data has been extrapolated as shown in Fig. 3.3. At $3x10^8$ sec (10 years) and room temperature (22C) we have a 1.2V window. At 80C the window has decreased to 0.8V at 10 years and at 150C the window decreases to 0.3V. The data indicates these SONOS devices should be acceptable for 7 V programming and 10 year memory retention applications for operating temperatures below +125 C [69].

The trap density in the oxynitride layer of the SONOS device is determined as a function of trap energy using Eqs. 3.1-3.4. Eq. 3.3 is used to calculate the energies of the traps in the nitride layer. At 250 C the trap energy ranges from 0.6 eV to 1.43 eV. Eq. 3.2 is used to calculate the density at points along the threshold voltage curve. The

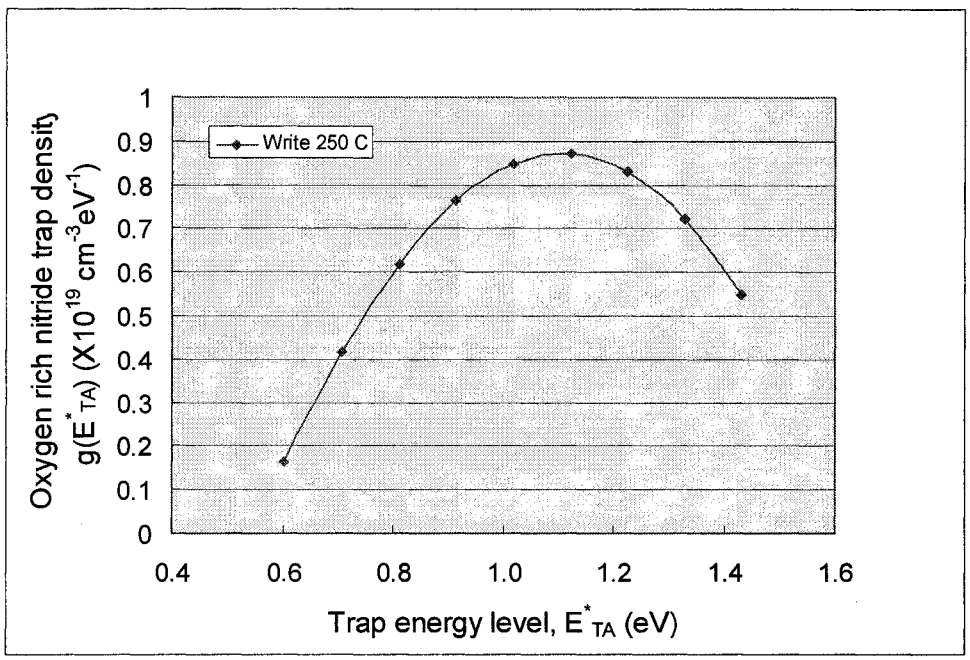


Figure 3.4: SONOS oxynitride film trap density, $g(E_{TA})$, versus trap energy, E_{TA}, for the write (excess electron) state at 250C [69].

change in threshold voltage divided by the logarithm of the change in time is calculated at each decade of time. The electron trap density, $g(E_{TA})$, is plotted versus trap energy, E_{TA}, as shown in Fig. 3.4. The electron trap density for the oxynitride film peaks at 1.1 eV below the edge of the nitride conduction band, similar to the silicon-rich nitride film, shown in Fig. 3.5. The trap density of the oxynitride film is less than the trap density of the silicon-rich nitride film [69]. This is expected as the presence of oxygen will tie-up silicon dangling bonds, which are the memory traps in the nitride.

The temperature effect on the threshold voltage of scaled (7V programming) SONOS oxynitride NVSM devices has been investigated. At elevated temperatures, the write state threshold voltage decay rate, $\partial V_{TH}/\partial \log(t)$, of SONOS devices with oxynitride films increases with an increase in temperature, while the erase state

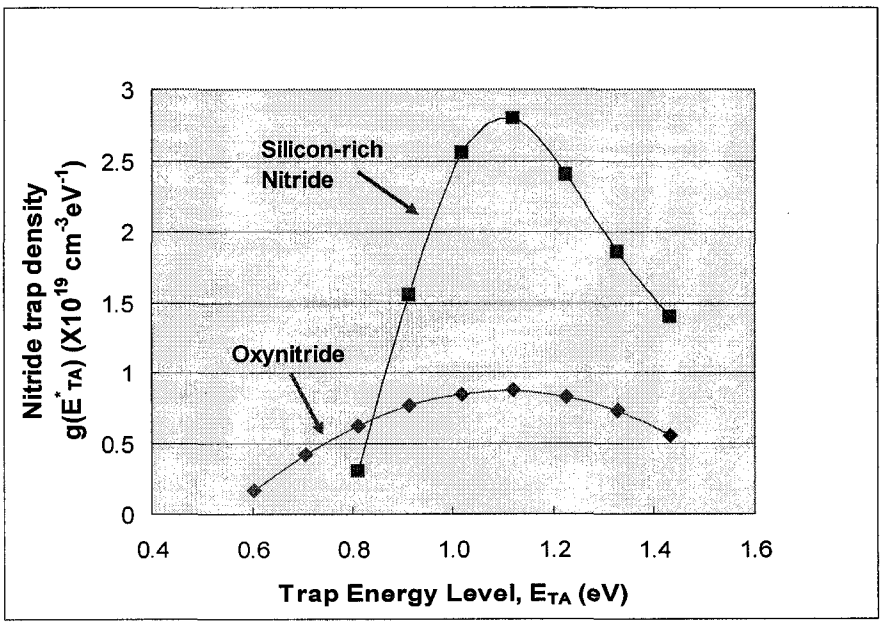


Figure 3.5: A comparison of trap densities in 'silicon-rich' and 'oxynitride' charge storage layers in SONOS NVSM gate dielectrics [69].

threshold voltage decay rate remains unchanged. The nitride trap density is extrapolated from the decay rate of the write state at 250 C. This allows us to model the trap distribution within the oxynitride band gap.

In previous studies of elevated temperature affects, Yang and White [18] have extrapolated the electron trap density for a SONOS device with a 'silicon-rich' nitride layer at 175 C. Research for this dissertation demonstrates the trap density for a silicon-rich nitride film is roughly three times greater than the trap density of an oxynitride layer, as shown in Fig. 3.5. The oxynitride and silicon-rich nitride trap densities peak at the same activation energy, 1.1 eV.

The extrapolated memory retention window at 3 x 10^8 sec (10 years) and 22C is 1.2V, a 0.8V window at 80C, and a 0.3V window at 150C. Oxynitride layer trap densities are three times smaller than trap densities for silicon rich nitride layers. A smaller nitride trap density may correlate to improved retention results due to the unavailability of free traps which act as "stepping stones" for stored charge tunneling out of the nitride and back to the silicon substrate. A smaller trap density indicates decreased internal electric fields in the nitride, and therefore less repulsion of charge from the nitride, resulting in lower charge decay rates. These measurements on scaled SONOS devices demonstrate the retention performance of SONOS devices with oxynitride charge storage layers up to 10 years at temperatures up to 150C.

Chapter 4

SONOS Device Design and Fabrication

The step-by-step process of SONOS device fabrication at the Sherman Fairchild Center Micro-Electronics Lab is outlined in Appendix A. This chapter is an overview of several design optimizations that must be taken into consideration in order to fabricate reliable conventional and short-channel SONOS transistors. SONOS transistors are fabricated in the Sherman Fairchild Laboratory on 15-30 Ω-cm, boron-doped, (100) 3" Si wafers. The devices will be defined using standard photolithography, etching, and ion implant with the TP-400 mask set. This mask set contains SONOS and standard CMOS transistors with channel lengths ranging from 1.0 μm to 50 μm, in various increments. A wide variety of channel lengths is necessary for device characterization, as channel hot electron and channel initiated secondary electron injection are used to examine write/erase characteristics in sub-micron dimension devices. Others devices, with larger areas, are better suited for charge pumping and linear voltage ramp measurements, which are used to examine the properties of the gate dielectric.

4.1 Threshold Voltage Adjust Implant

The threshold voltage of a SONOS transistor is determined by a number of factors. Most important is the substrate doping in the channel beneath the gate of the

transistor. An ion implant into the channel area can shift the threshold voltage by altering the doping of the channel. The SONOS transistors fabricated for this dissertation are designed with a blanket implant so the NMOS and PMOS threshold voltages are of the same magnitude but opposite polarities, i.e. +1V (NMOS) and -1V (PMOS). Another consideration is the memory window positioning. It may be preferable to have NMOS transistors with write and erase state threshold voltages which are positive, to avoid accessing more than one transistor in an array when the gate is grounded to read out the memory state [10]. Another factor is the end of memory life position. The natural threshold voltage is the threshold voltage the transistor will revert to after the charge stored in the nitride has escaped over years of storage. If the memory position is negative, grounding the gates of an array of NMOS transistors will cause current to flow in all the bits that are not in the erase state, i.e. either written or un-programmed [3].

The threshold voltage of a traditional NMOS transistor is given by the sum of the charge in the device [21]

$$V_{th} = \Phi_{GS} + 2\Phi_F + \frac{\sqrt{4\varepsilon_{si} q N_B \Phi_F}}{C_{eff}} - Q_N\left[\frac{x_{OB}}{\varepsilon_{OX}} + \frac{x_n}{2\varepsilon_n}\right] - \frac{Q_{SS} + Q_{it}}{C_{eff}} \qquad (4.1)$$

Where ϕ_F is the Fermi potential, ϕ_{GS} is the gate to semiconductor work function, QS is the charge in the substrate, Q_{SS} is the charge at the Si-SiO_2 interface, Q_{it} is the charge in interface traps, and C_{eff} is the effective capacitance of the gate dielectric. The Fermi potential is given by

$$\phi_F = \frac{k_B T}{q} \ln\left(\frac{N_B}{n_i}\right) \tag{4.2}$$

Where k_B is Boltzmann's constant, T is the absolute temperature, q is the charge of a single electron, N_B is the doping density in the silicon bulk, and n_i is the intrinsic doping of the semiconductor. Analyzing the doping of the polysilicon gate and typical values for the semiconductor doping, the shift in threshold voltage due to ion implant can be written as [10]

$$\Delta V_{TH} = \frac{q}{C_{eff}} \Phi_{ION} \tag{4.3}$$

where Φ_{ION} is the ion implant concentration. The threshold voltage will increase if a positive ion (boron) is implanted, and decrease with a negative ion (phosphorus) implant.

Assuming a bulk doping of $7x10^{14}$ cm^{-3} and an effective gate thickness of 140Å (thicker due to the thickened tunnel oxide for hot carrier programming), the natural threshold voltage for an NMOS transistor is -0.260 V and -1.126 V for a PMOS transistor. Therefore, a boron implant of 10^{12} cm^{-2} at 50 keV has been incorporated in order to shift the threshold voltages to 0.35 and -0.5 for NMOS and PMOS transistors, respectively. Calculations of the effect on this ion implant on the threshold voltage have been calculated using MathCAD.

4.2 SONOS Gate Stack

The gate dielectric formation is the most critical and delicate stage of the SONOS transistor processing. First the tunnel oxide is grown using a triple wall oxidation (TWO) furnace. The triple wall oxidation furnace is a unique tool for creating ultra-thin oxide films with low defect densities. Three walls surround the furnace and gas is flowed between the walls to create defect-free oxides with low alkali, water-related, and heavy metal ion contamination. Extremely thin oxides can be grown and oxide thickness can be finely controlled [10]. Graduate research at the Sherman Fairchild Center has demonstrated improved performance and reliability from tunnel oxides grown at high temperatures in the triple wall oxidation furnace [73]. Tunnel oxide thicknesses have been varied for this wafer processing run. Thinner tunnel oxides of 2 nm have been deposited for programming using Modified Fowler-Nordheim (MFN) tunneling. In addition, thicker tunnel oxides of 5 nm have been deposited for use in studying programming using channel hot electron, channel initiated secondary electron injection, and hot hole injection. The short-channel devices will have thicker tunnel oxides in order to prevent direct and MFN tunneling.

The silicon nitride film is deposited using low pressure chemical vapor deposition (LPCVD). The silicon nitride film is roughly 5 nm thick. Yang, *et al.*, [74] conducted studies on nitride films, which were deposited at different temperatures. These studies concluded the optimum temperature for Low-Pressure-Chemical-Vapor-Deposition (LPCVD) nitride deposition for a minimum surface roughness is 680 C.

Minami, *et al.*, [75] demonstrated that using LPCVD to form blocking oxides improved memory retention dramatically. During LPCVD nitride deposition, Silicon dangling bonds create deep level traps in the nitride layer. The number of traps can be reduced using oxygen and hydrogen atoms as terminators for dangling bonds [15]. Trap-rich nitrides facilitate fast programming speeds [76].

Finally, a 5.0 nm blocking oxide is deposited using LPCVD. The blocking oxide must be thick enough to prevent charge from tunneling to the gate. In addition, the lever arm of the charge stored in the nitride is greater if the blocking oxide is thicker. The threshold voltage can be written as

$$\Delta V_{TH} = \Delta Q_N \left[\frac{X_N}{2\varepsilon_N} + \frac{X_{OB}}{\varepsilon_{OX}} \right] \tag{4.4}$$

demonstrating the need for a sufficiently thick blocking oxide. Conventional SONOS transistors fabricated for electrical characterization must be designed with the specific requirements of space and military applications in mind. For example, the tunnel oxide thickness must be exact. A 2Å difference will change the programming time by an order of magnitude and the oxide must be a high quality film with minimal interface traps.

4.3 Polysilicon Gate Etch

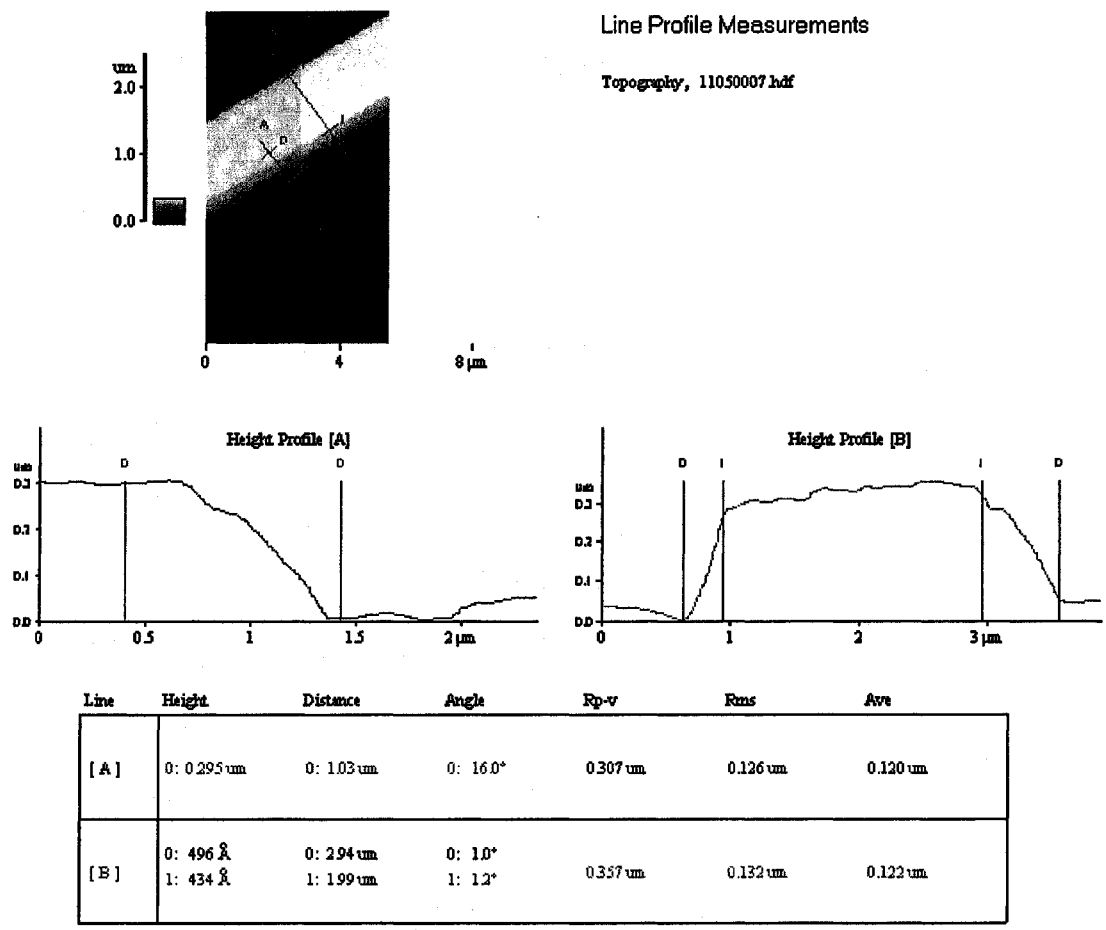
The polysilicon gate etch is the most important step in the gate dielectric formation. Etching the gate for too long will cause the etchant to either cut into the

silicon surface, damaging the source and drain regions, or etch the tunneling oxide laterally until the gate is destroyed. The most common gate etch is the plasma etch. The plasma etch in theory is an anisotropic etch, meaning the etch process is completely vertical. In practice, there is some degree of undercutting in the gate stack. Another drawback to the plasma etch is the lack of selectivity, meaning it will etch thousands of angstroms of polysilicon in minutes and then, in a matter of seconds, etch through the relatively thin oxide-nitride-oxide dielectric and continue to etch into the source and drain regions.

A solution to this problem is improving etch selectivity. Selectivity is the ratio of the etching rates of the two films. For example, an etchant that etches polysilicon at 2kÅ/minute and etches oxide at 50Å/minute has a polysilicon/oxide selectivity of 40:1. An etchant with a high selectivity will etch the polysilicon quickly and stop when it reaches the delicate triple dielectric. A variety of wet chemical etchants are available for use with polysilicon films. The most common etchant is a mixture of acetic acid ($HC_2H_3O_2$), nitric acid (HNO_3), and hydrofluoric acid (HF). Combining this solution in a 250:200:4 ratio will yield a polysilicon etch rate of about 2kÅ/minute [10].

A 'TRW' chemical etch is used to pattern the polysilicon gate isotropically during the device fabrication for this dissertation. The etch is composed of $HF:HNO_3:H_20$ in a ratio of 1:60:60 [77]. Experiments on patterning polysilicon films have shown etch rates of approximately 1.5-1.8kÅ/min. The etch rate slows for very long times and large features due to re-deposition of etched materials. The TRW etch is not anisotropic as there is a 0.5 μm region on either side of the gate where the sidewall

is sloped. The advantage of using this etch is that it has a selectivity of 30 from polysilicon to silicon dioxide. The TRW etch is used to etch the polysilicon gate material and the blocking oxide, stopping on the silicon nitride film. The selectivity of



Line	Height	Distance	Angle	Rp-v	Rms	Ave
[A]	0: 0.295 um	0: 1.03 um	0: 16.0°	0.307 um	0.126 um	0.120 um
[B]	0: 496 Å 1: 434 Å	0: 2.94 um 1: 1.99 um	0: 1.0° 1: 1.2°	0.357 um	0.132 um	0.122 um

Figure 4.1: Atomic force microscope (AFM) cross-section of a polysilicon film etched using the 'TRW' solution. Etch rate was approximately 1.5kÅ/minute. Undercutting is 0.5 µm on either side of the gate.

TRW from polysilicon to silicon nitride is 150:1. The nitride is then etched using heated H_3PO_4 and a short BHF dip etches the tunnel oxide and removes any residual oxides from the etched regions.

4.4 Short Channel Transistor Fabrication

There are several considerations one must evaluate when fabricating short-channel transistors. First, a punch-through protection implant must be applied to the channel of the device to prevent the depletion regions from expanding and shorting the source and drain together when a bias is applied to the drain [44]. For an NMOS short-channel transistor, the source and drain are formed with phosphorus, therefore, a positive ion, boron, must be implanted into the channel as a 'channel-stop'. Silvaco simulations have been run to determine the ion implant dosage and energy. The punch-through protection implant is performed at a dosage of $1x10^{12}$ ions/cm^2 boron at 180 keV and $2x10^{12}$ ions/cm^2 phosphorus at 180 keV for NMOS and PMOS transistors, respectively. The implant is performed at a much higher energy, 180 keV, than the threshold voltage adjust implant, 50 keV, in order to drive the ions deeper into the substrate to prevent a depletion region from forming in the channel. Performing the implant at a low energy level would cause the ions to accumulate at the surface and merely shift the threshold voltage of the transistor as discussed in section 4.1.

In addition to the punch-through protection implant, we must examine the definition of the source and drain areas. Source and drain implants are performed at the Sherman Fairchild Center using a self-aligned gate technology. After the gate dielectric is patterned, the source and drain implants are applied using the gate dielectric as a mask. This is an extremely effective processing technique. However, the channel length is limited by the photolithography capabilities of the facilities. The photolithography

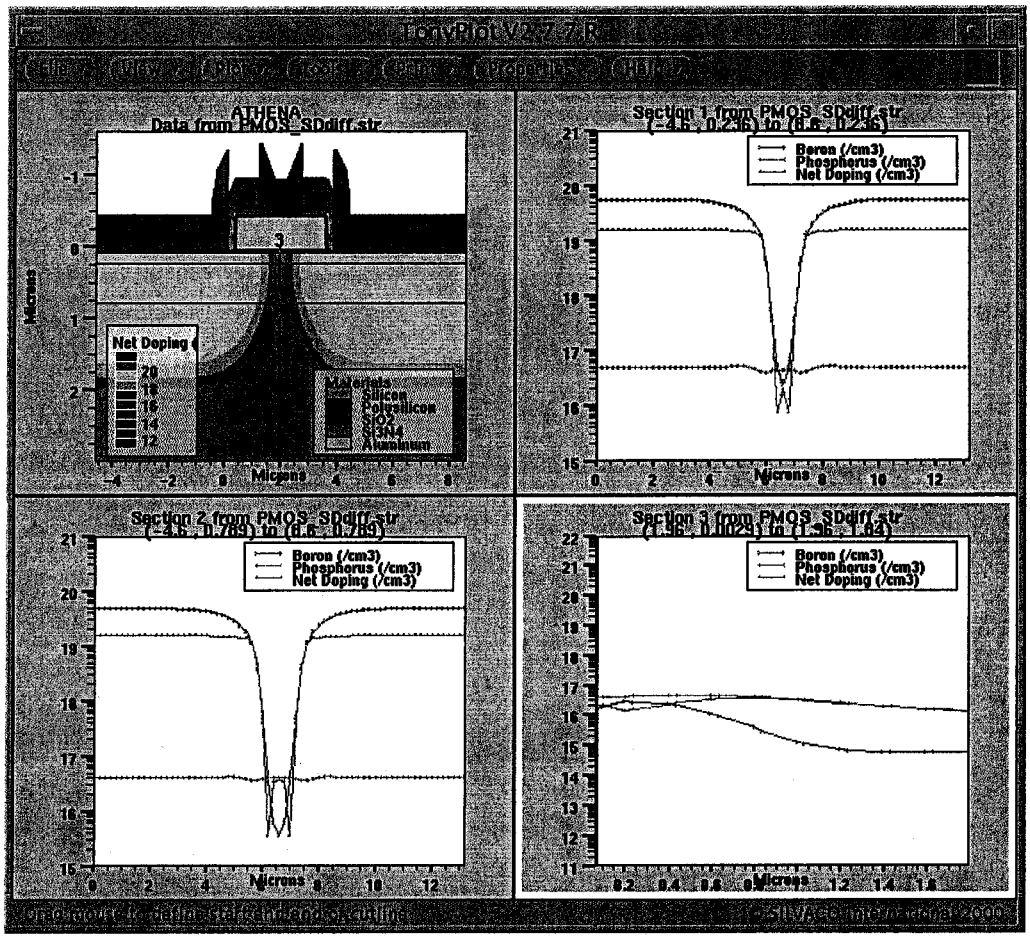


Figure 4.2: Athena simulation of a PMOS SONOS transistor with a 0.35 μm channel length. The channel length was 3 μm before a 360-min. anneal at 1100 C. Net doping is shown in the device cross-section (top-left), impurity/net doping in the lateral cross-section at 0.2 μm (top-right) and 0.8 μm (bottom-left) into bulk, and doping from the center of the channel into substrate (bottom-right) are shown.

equipment in use at Lehigh University has a minimum feature size of approximately 3 μm.

After etching the oxides and nitride to pattern the gate dielectric, the source and drain will be implanted using boron ($5x10^{15}$ ions/cm^2 at 32 keV) and phosphorus ($5x10^{15}$ ions/cm^2 at 40 keV) implants. Conventional SONOS transistors require an

implant anneal performed at 950 C for roughly 2 hours to diffuse the implanted ions deeper into the device [10]. Conventional SONOS transistors were fabricated for this dissertation in addition to short-channel transistors. A long-term anneal is performed to drive the source and drain together to produce short-channel transistors.

Computer simulations have been run with the Silvaco software package, Fig. 4.2, to calculate the time and temperatures required to perform this anneal to produce devices with channel lengths on the order of 0.5 – 0.8 μm. Considering the change in physical and electrical gate lengths from the drawn mask gate length, a drive-in anneal of 300-330 minutes at 1100°C is required to produce short channel transistors using the TP400 mask-set. This drive-in anneal is 100°C higher than the typical anneal and is twice the time duration.

Chapter 5

Electrical Characterization Techniques

Electrical measurements can be performed on SONOS transistors in order to test almost any parameter of the device, such as gate dielectric composition, data retention, and even trap density in the interfaces between gate films or in a specific gate film. These measurements include write/erase, threshold voltage, linear voltage ramp (LVR), programming stress, and charge pumping measurements. These measurements test the quality of the device as a transistor, and can quantify the durability of the device over an extended period of use. In addition, the memory properties of the device must be tested using the write/erase and retention measurements for virgin and stressed devices. These tests determine how well the device retains charge under extreme use and over time.

The measurements presented in this dissertation are performed in a LabVIEW™ environment. LabVIEW™ is a software platform developed by National Instruments which utilizes a graphical programming language to control test equipment, such as oscilloscopes, electrometers, and function generators. The programs are written similar to a flow-chart where commands are given inside of specific blocks and constants and variables are referenced by connecting lines or wires. Communication between the computer and the measurement devices is accomplished through a GPIB cable connected to an IEEE-488 port on the electronic instrument and an identical port on a GPIB card which is plugged into a PCI slot on a personal computer. LabVIEW™ can

simultaneously set up several devices for a single measurement since each device has its own addressing number and will only respond to commands sent to its individual address. LabVIEW™ controls the instrument during the electrical measurement and sends data to the computer through the GPIB interface, where it can be formatted and displayed on the computer screen and written to a text file. Screen shots of the LabVIEW™ program panels will be provided for the measurements discussed in the following sections.

The programming and retention measurements are performed utilizing a threshold voltage sensing circuit connected to a National Instruments data acquisition card. The sensing circuit utilizes an op-amp to read the threshold voltage from the gate of the device without disturbing the memory state. A reference voltage and resistance are selected in order to produce a reference current, typically 1-100μA depending on the

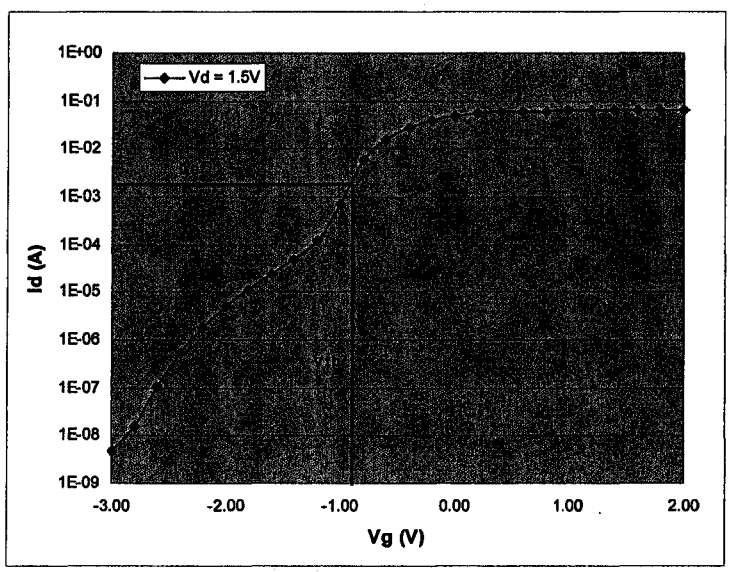


Figure 5.1: Demonstration of reference current and biasing of a transistor to read the threshold voltage at the gate terminal with a 1.5V drain bias.

device dimensions. This current is forced through the memory transistor while the source and bulk terminals are grounded. The voltage measured at the gate will be equal to the threshold voltage of the device [44].

The NI data acquisition card sends 0V and 5V signals to the circuit to switch the circuit between the read, write, erase, or idle states. Programs written in LabVIEW™ control the timing and output of the instrumentation, giving the user control over write and erase times, retention time duration, and number of programming cycles.

5.1 Programming

Applying a voltage to the gate can alter the threshold voltage of a SONOS device. The transient behavior of a SONOS device is a function of the polarity, amplitude, and duration of the gate voltage pulse. The write/erase measurement determines the programming speed and initial memory window of a device. The greater the amplitude of the applied voltage pulse, the quicker the device will erase or write. The typical voltage pulse duration is in the range of 1 ms to 10 ms. The erase pulse will sometimes be longer if a block or page erase system is utilized.

The shift in the threshold voltage during a write/erase operation is due to charge tunneling from the silicon bulk and trapping in the nitride film. In the operation of SONOS devices, charge injection is governed by tunneling processes, which are strong functions of the electric fields in a device. This makes the amplitude of the applied programming voltage and the amount of the charge initially stored in the nitride

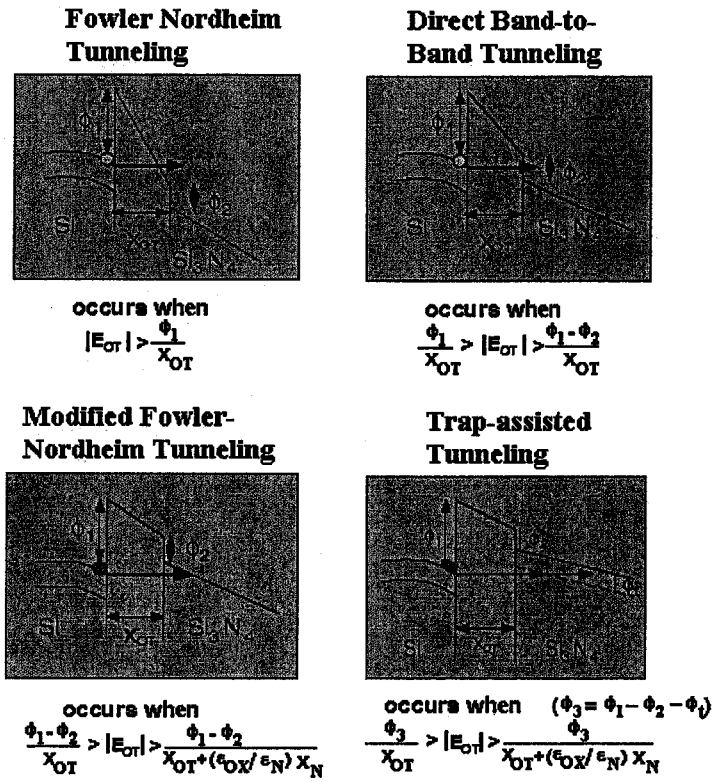


Figure 5.2: Energy band diagrams and tunnel oxide electric field requirements for FN, DT, MFN and TAT tunneling mechanisms [21].

important in determining the programming speed of a SONOS memory transistor. Assuming Modified Fowler-Nordheim (MFN) tunneling is the dominant charge injection mechanism, French, *et al.*, [62], derived an expression to describe the dependence of the SONOS threshold voltage change on several essential parameters

$$\Delta V_{TH}(V_P, t_p) = \frac{x_{eff}}{E_T}\left[\frac{V_p + \left(\frac{x_{OB}}{\varepsilon_{ox}} + \frac{x_N - \bar{x}}{\varepsilon_N}\right) Q_N(0) - \phi_{GS} - \phi_S}{Xeff}\right]^2 \ln\left(1 + \frac{t_p}{\tau}\right) \qquad (5.1)$$

where $x_{eff} = x_{OT} + x_N\left(\frac{\varepsilon_{ox}}{\varepsilon_n}\right) + x_{OB}$ is the effective thickness of the stacked dielectrics, $\bar{x}$ is the location of the charge centroid in the nitride film, E_T, V_p, t_p are the magnitude and duration of the programming pulse, $Q_N(0)$ is the initial charge stored in the nitride, ϕ_{GS} is the gate-semiconductor work function difference, ϕ_S is the surface potential, and τ is the turn-on time [10].

The programming time, t_p, is specified by customers or applications, typically in the range of 1 ms to 10 ms for writing for writing, and 10 ms to 100 ms for erasing. For a fixed t_p, the dependence of the threshold change on the programming voltage can be described by the differentiation of Eq. 5.1 with respect to Vp

$$\frac{\partial \Delta V_{TH}}{\partial V_p} = 1 - \frac{\tau}{t_p} + \frac{2}{E_T}\left[\frac{V_p + \left(\frac{x_{OB}}{\varepsilon_{ox}} + \frac{x_N - \bar{x}}{\varepsilon_N}\right)Q_N(0) - \phi_{GS} - \phi_S}{Xeff}\right]\ln\left(\frac{t_p}{\tau}\right) \approx 1 \qquad (5.2)$$

where $t_p >> \tau$ is assumed. There is a 1:1 correspondence in the change of the threshold voltage with respect to the programming voltage [10].

The write/erase characteristic for an n-channel SONOS transistor using erase and write voltages of 5, 6, and 7 volts is shown in Fig. 5.3. The transistor has a virgin threshold of 0.2 V. During the measurement, the drain, source, and substrate of the device are grounded. To precondition the device for a write, the circuit will write and erase the transistor. The erase measurement is preconditioned with an erase and a write.

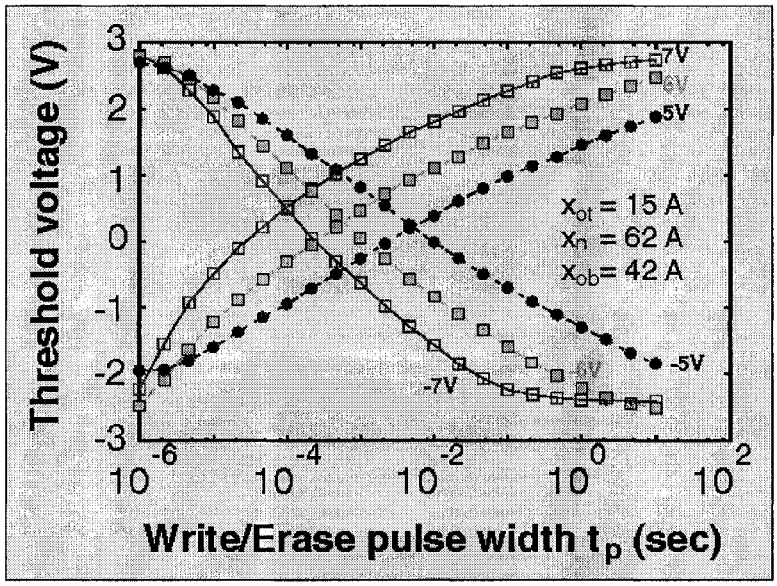


Figure 5.3: Write/erase curves for a SONOS transistor programming with gate voltages of 7V (blue), 6V (teal), and 5V (red) [13].

Next, a write pulse, +5 V, +6 V, or +7 V, is applied to the gate of the device for the time duration, t_p. The erase operation is performed in the same manner, except the gate voltages used are -5 V, -6 V, and -7 V.

Fig. 5.3 represents a typical result of the write/erase measurement. The initial threshold voltages are -2.0 V t0 -2.5 V for the erased device, and 2.7 V to 2.8 V for the written device. The crossover point where the write and erase curves meet is used to characterize the programming speed of the device at that particular gate voltage. As the programming voltage is increased from 5 to 7 volts, crossover point decreases from 10 ms to 0.1 ms, a rate of 1 decade/V [10]. The crossover time will increase 1 decade of time for every 2Å increase in tunneling oxide thickness. The electric field during programming may be calculated by dividing the gate voltage by the effective oxide thickness of the gate dielectric.

5.2 Retention

Retention measurements are used to characterize the ability of a SONOS device to store data over time. Retention is the most important characteristic for a non-volatile memory device. It is difficult to make a device with fast programming speed, good data retention, high endurance, and low voltage operation. As the tunnel oxide of the device is scaled, the device will have faster programming times and require lower programming voltages, but will have degraded data retention and endurance.

Data retention measurements are performed at room temperature, 22 C, and 85 C. To perform the retention measurement, the device is written or erased. The read

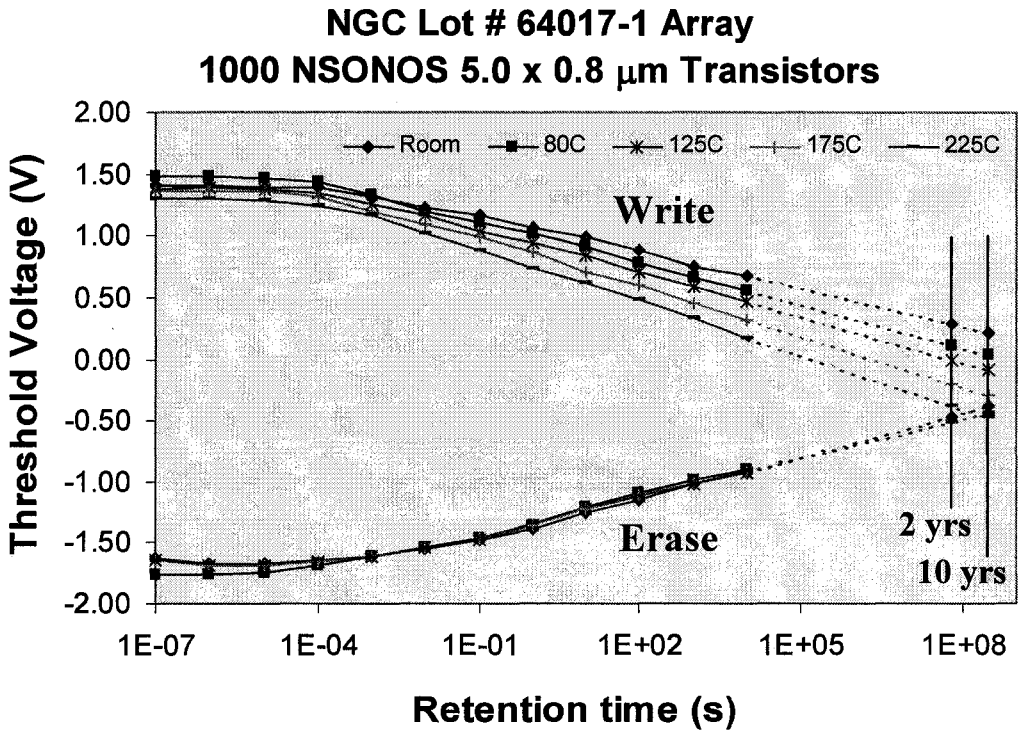


Figure 5.4: Threshold voltage of a SONOS device versus read delay time extrapolated to 10 years for temperatures ranging from 22°C to 175°C [20].

measurement is performed after a specified delay time. The read operation is repeated, and the delay time is increased by an order of magnitude. This process is repeated for delay times typically between 10^{-6} seconds to 10^{5} seconds. After the read operation is performed, the memory state is refreshed in order to prevent read disturb effects. During the delay, when the device is in the idle mode, all terminals are grounded.

The threshold voltage decay rate is small for short read delay times, up to roughly 1 second. At 1 second, the decay rate increases for both the erased and writing states. The retention measurement can be performed at elevated temperatures. There are several effects of temperature on the write state, but not the erase state. Device operation, retention, and modeling at elevated temperatures are discussed in Chapter 3. Retention data can be roughly extrapolated using the last 3 or 4 data points and a straight line fit in order to characterize retention at times up to 10 years. The standard used by Northrop Grumman for data retention is a 50-mV window at 10 years, at 85 C.

5.2.1 High Temperature Electrical Measurement Setup

High temperature measurements require the hot chuck to heat the wafer and hold it at temperatures of 22 C, 80 C, 150C, and 250 C. The hot chuck used to collect this data has been operated routinely at temperatures up to 400 C. Due to this wear on the hot chuck, the surface where the wafer is placed is no longer planar. This creates difficulty in making good contacts between the probes and the contact pads of the device terminals. When the device is heated and cooled, there is a degree of vibration associated with the operation of the hot chuck. Increasing the temperature to 125C

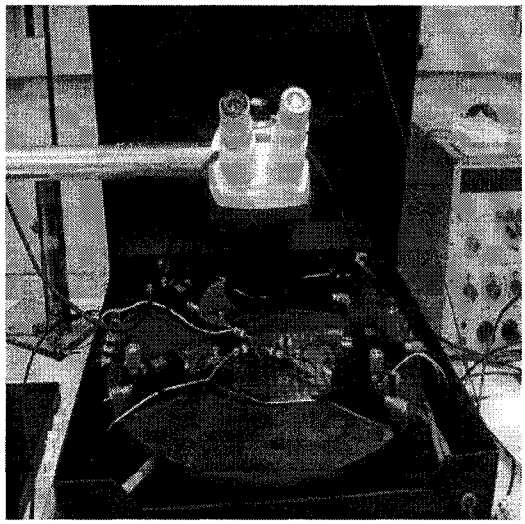


Figure 5.5: SONOS wafer under test, probes, and hot chuck [19].

causes the probes to expand and slide off of the contact windows, scratching the device, and losing contact with the appropriate device terminals, especially for long-term retention measurements take out to 10^4 seconds.

The research presented in this dissertation compensates for these effects with special expansion probes and using smaller sections of the wafers. The smaller pieces are not as affected by the non-planar surface of the hot chuck or by vibrations in the

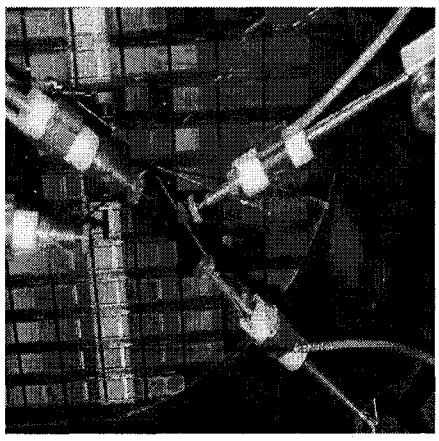


Figure 5.6: Close-up view of SONOS wafer on hot-chuck with probes [19].

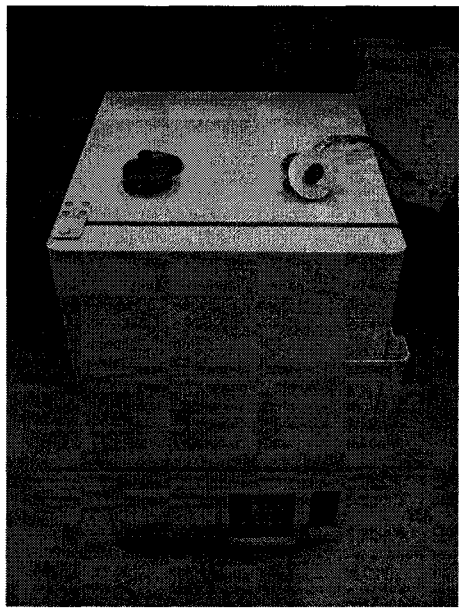


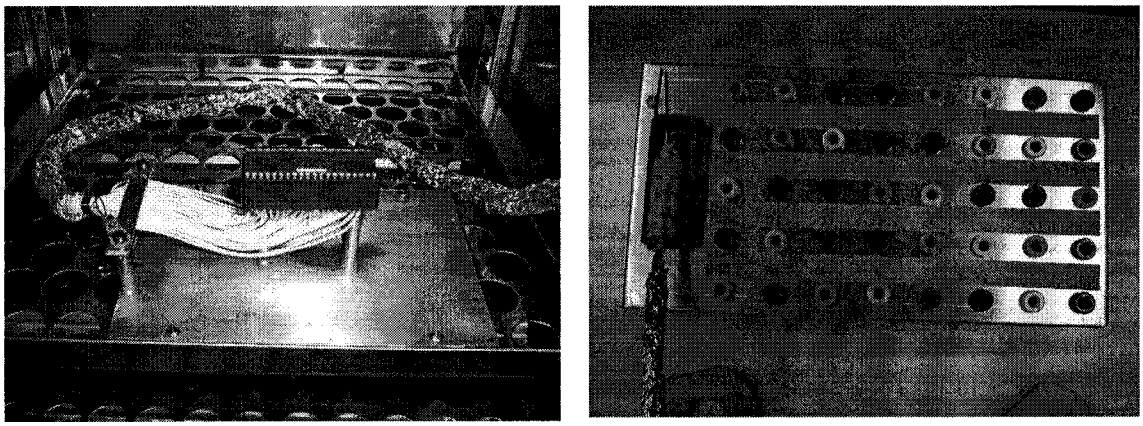
Figure 5.7: Yamato DX 300 high temperature oven capable of heating to 300°C.

laboratory, allowing the retention measurement to be performed at extended times and temperatures up to 225C. Probes designed specifically for high temperature measurements are thin and flexible, allowing them to bend in order to compensate for the contact pads on a device moving due to expansion or contraction of the wafer. The flexibility of these probes also allows the probes to maintain contact with the wafer while experiencing vibrations which would cause conventional probes to slide off of the contact windows.

Extremely high temperatures (> 100 C) can cause the probes to stray from the contact pads even with specially designed probes. Therefore, Northrop Grumman has packaged SONOS transistors in a 40-pin integrated circuit (IC) chip. Each chip contains 7 SONOS transistors of various dimensions. The transistor terminals are accessed using the pins of the IC chip. In addition, the IC chip is connected to a zero force insertion

Figure 5.8: Front panel of the DX 300 oven where the temperature is set and monitored. Temperature is rigidly controlled and does not fluctuate when oven has reached set point. Over-temp control prevents overheating in case of microcontroller malfunction.

socket where the pin connections are fed out through a wire connected to a metal box, where connections can be made with conventional 'banana-plug' cables to each individual IC pin. The IC socket and connectors are rated for use in temperatures up to 300C. A Yamato DX300 high temperature oven is used to heat the IC chip up to a



(a) (b)

Figure 5.9: IC socket fixture for high temperature measurements (a) and IC connector for wiring to electrical characterization equipment (b).

maximum temperature of 300C where the cable connecting the socket to the metal output box is fed out through a vent in the top of the oven. This flexible measurement setup allows for easy high temperature characterization of SONOS devices without the risk of probes disconnecting from contact pads due to thermal expansion of the wafer, thermal expansion of the probes, vibrations generated by the occupants of the research lab, and vibrations generated by electronic equipment (specifically cooling fans in nearby computers and characterization equipment).

5.3 Charge Pumping

The charge pumping measurement is a widely accepted and utilized technique used to measure the interface trap density in MOS devices [10]. This technique is utilized on SONOS devices to measure the interface trap density before and after the device has been erased and written (pre/post-stress). Interface states are used by charge in the nitride as 'stepping-stones' for back-tunneling to the silicon bulk when the device is in the retention mode. Measuring the density of interface states after a set number of write/erase cycles at specific voltages gives the user an indication of the negative effects of programming using a specific mechanism and the quality of the gate oxide.

5.3.1 Square Waveform

The charge pumping measurement excites electrons stored in traps at the Si-SiO_2 interface by applying a voltage waveform to the gate of the device. The source and

drain are connected together. A trapezoidal wave function is applied to the gate. The base voltage of the gate function is swept from negative to positive values while the amplitude is held constant. The gate potential of the transistor is altered so the underlying surface potential moves from strong accumulation to strong inversion. Electrons flow from the source and drain into the channel where a fraction of the electrons are captured by interface states and near oxide traps in the tunneling oxide and the nitride film. The surface is then driven back to accumulation, causing the mobile

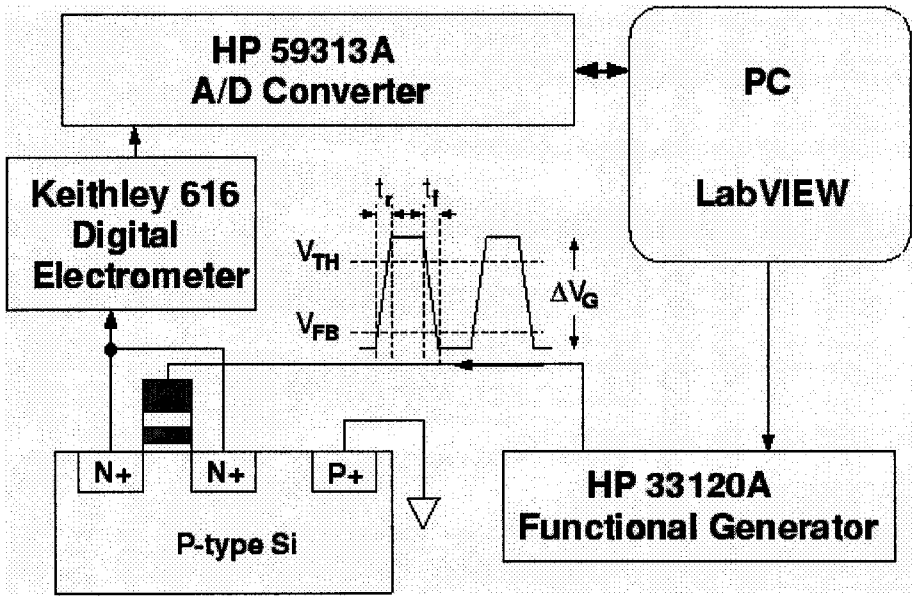


Figure 5.10: Conventional charge pumping setup with the source and drain shorted [10].

charge to drift back to the source and drain. The charge trapped in the interface states recombines with majority carriers from the substrate, causing a substrate current to flow. This substrate current is proportional to the density of Si-SiO_2 interface traps, D_{it}. The charge pumping measurement can also be performed by applying the same voltage to the gate, but grounding the bulk, connecting the source and drain together, and

measuring the source-drain current. The magnitude of the charge pumping current will remain the same, however, it will be of opposite polarity to the bulk current in the conventional setup. The substrate current, I_{cp}, is proportional to the mean density of interface traps, D_{it}, and the frequency of the pumping pulse, f, shown in the following equation [78]

$$I_{cp} = 2qD_{it}fA_g kT \ln\left[v_{th} n_i \left| \frac{V_{FB} - V_{TH}}{\Delta V_G} \right| \sqrt{\sigma_n \sigma_p t_r t_f} \right] \tag{5.3}$$

where A_G is the area of the gate, v_{TH} is the thermal velocity of the carriers, σ_n and σ_p are the capture cross sections for electrons and holes, V_{FB} and V_{TH} are the flatband and threshold voltages of the device, ΔV_G, t_r, and t_f are the amplitude, rise time, and fall time of the pulse applied to the gate, respectively. Using this technique, the charge pumping current I_{cp} is measured and the interface trap density D_{it} can be extracted.

In order to determine the quality of the SONOS device and its long-term operation, a SONOS transistor is often stressed by repeatedly writing and erasing the device. The write and erase times are determined by the user, typically a 2.5 ms +7V write and 7.5 ms at -7V erase are used with Northrop Grumman devices. Charge pumping and retention measurements will be performed on a virgin unstressed device and repeated after every decade of stress. Stressing the devices is used to determine how the device will hold up after it being used. However, NVSM's are typically not programmed more than 10^4 times due to their nature as a memory where data is permanently written, stored, and accessed.

When the device is programmed and erased, holes and electrons tunnel through the oxide of the device, often disrupting the covalent bonds between silicon and oxygen atoms in the oxide layer, forming traps know as interface states. As more electrons and holes tunnel, more traps, dangling bonds, will be created in the oxide of the device. When the device is in written or erased, and is in the retention mode (all terminals connected to ground), charge from the nitride will leak out through the oxide with assistance from interface traps. Electrons and holes will tunnel through the oxide barrier to these traps at the interface and then they will continue into the silicon. As the device is stressed, the charge will leave at an increased rate, which can be seen during the retention measurement. The density of traps in the interface can be measured as D_{it} by a charge pumping measurement, and is a good test of the oxide quality and severity of damage from a particular programming mechanism [15].

5.3.2 Triangular Waveform

If a triangular (sawtooth) waveform is applied to the gate electrode, the interface trap density may be extracted from the charge pumping current as [78]

$$I_{cp} = 2q\overline{D}_{it} f A_G kT \left[\ln(v_{th} n_i \sqrt{\sigma_n \sigma_p}) + \ln\left\{ \frac{|V_{FB} - V_{TH}|}{|\Delta V|_G} \frac{1}{f} \sqrt{\alpha(1-\alpha)} \right\} \right] \tag{5.4}$$

where A_G is the area of the gate, v_{TH} is the thermal velocity of the carriers, σ_n and σ_p are the capture cross sections for electrons and holes, V_{FB} and V_{TH} are the flatband and

threshold voltages of the device, respectively. In addition, ΔV_G and f are the amplitude and frequency of the triangular waveform applied to the gate, respectively, while α is defined as the fraction of the period when the gate voltage is rising and $(1-\alpha)$ is the fraction of the period when the gate voltage is falling.

The triangular wave pulse does not suffer from as much distortion from the cable used during wafer-level testing in contrast with the use of trapezoidal pulses. A triangular pulse will retain its basic shape and not distort as a square or trapezoidal wave, since rise and fall times are affected by noise and stray capacitances with the use of the latter waveform. In a triangular pulse, the rise and fall times are defined by the frequency of the signal. The average interface trap density, $\overline{D_{it}}$, may be extracted by differentiating Eq. 5.4 as

$$\frac{dQ_{CP}}{d\log f} = \frac{2qkT\overline{D_{it}}}{\log e} \cdot A_G \tag{5.5}$$

where $Q_{cp} = I_{cp}/f$ is the charge pumped per cycle. The charge pumping measurement is performed at several frequencies, between 100 kHz – 300 kHz, and the maximum charge pumped is plotted versus the logarithm of the frequency [78]. This plot yields a straight line, where the slope is used to extract $\overline{D_{it}}$ from Eq. 5.5. This technique does not rely on the threshold or flatband voltage, capture cross-sections, or gate voltage applied during the measurement and is an accurate method for determining the interface trap density. The charge pumping measurement is repeated at intervals of 10^4, 10^5, 10^6

write/erase cycles. These measurements yield a comparison of various programming techniques and the damage to the tunneling oxide induced per write/erase cycle. This analysis allows us to determine the most efficient and damage-free mechanisms for programming NVSM devices.

5.3.3 Variable Frequency

The charge pumping measurement utilizes an applied voltage waveform with a frequency which is commonly in the range of 10 kHz to 1 MHz. At these frequencies, the sole contributors to the charge pump current are the interface traps within 10 Å of the Si-SiO_2 interface. However, if the frequency of the applied pulse decreases, the trapped carriers at the interface will have enough time to tunnel to Near-Interface-Oxide-Traps (NIOT's). The tunneling time constant must be less than half the period of

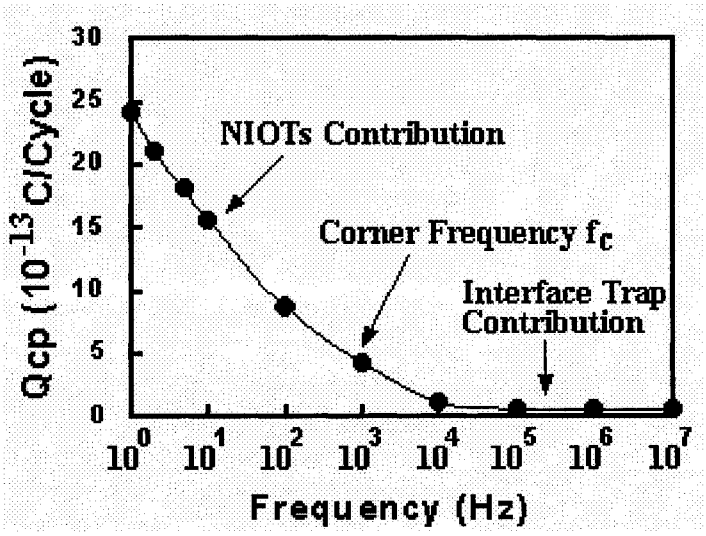


Figure 5.11: The recombined charge per cycle Q_{CP} as a function of frequency for a thin tunnel oxide SONOS device (W=100μm L=8μm). The increase in Q_{CP} at low frequency is attributed to an additional charge pumping current associated with nitride traps, which are only 15Å away from the Si-SiO_2 interface [79].

the applied waveform [79]. In Fig. 5.11, the charge pumped remains constant at high frequencies, since the interface traps are the only contributors to the recombination current.

Variable frequency charge pumping involves monitoring the charge pumping current over a wide range of gate pulse frequencies. As the frequency is decreased, Q_{cp} will increase due to the contribution of the NIOT's. The contribution of mono-energetically distributed NIOT's to the charge pumping per cycle, Q_{cp} (NIOT), is given by [80]

$$Q_{cp}(NIOT) = qA_g \int_{x_{\min}}^{x_m(f)} dx N_T(x)[1 - e^{-\frac{1}{2\tau_{TT}(E_{TO},x)}}] \tag{5.6}$$

where

$$\tau_{TT}(E_{TO},x) = \frac{(m^*_{ox,e})^2 x(1+\frac{t}{2\kappa_1 x})}{2\pi^2 \kappa_2 \hbar^3 D_{it}} e^{2\kappa_1 x} \approx \tau_0 e^{2\kappa_1 x} \tag{5.7}$$

is the trap-to-trap tunneling time constant used to describe the carrier tunneling phenomenon between interface traps and NIOT's, τ_0 is inversely proportional to D_{it} and is weakly dependent on x, $N_T(x)$ is the spatial trap density of NIOT's, x_{min} is the minimum distance at which a NIOT is distinguishable from an interface trap, x_m is the maximum distance into the oxide at which a gate pulse with a frequency f can probe, $\hbar$ is the reduced Planck's constant, D_{it} is an averaged density of interface traps, and the parameters κ_1 and κ_2 are the attenuation coefficients in the oxide and semiconductor,

$m_{ox,e}$ and $m_{si,e}$ are the effective masses of an electron in the oxide and semiconductor, respectively. The slope of the experimental curve in Fig. 5.11 yields the density of NIOT's at the maximum tunneling distance x_m as

$$N_T(x_m) \approx \frac{2\kappa}{2.3qA_G}\left(-\frac{\partial Q_{cp}}{\partial \log f}\right) \tag{5.8}$$

Variations of the gate pulse frequency, f, allow for the determination of the spatial distribution of NIOT's [10].

The tunnel oxide thickness may be obtained using the variable frequency charge pumping measurement using the corner frequency f_C and the interface trap density D_{it}.

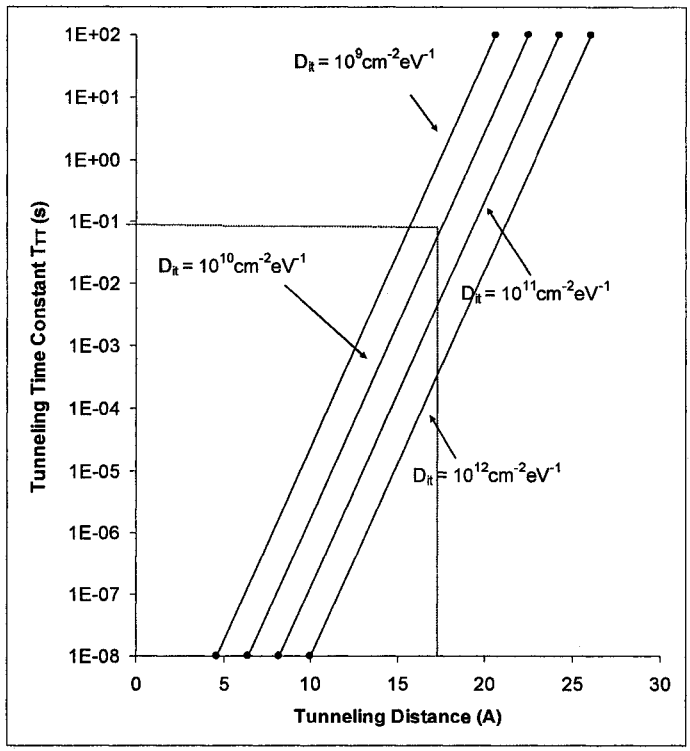


Figure 5.12: Tunneling distance plotted versus tunneling time constant for determining tunnel oxide thickness [17]. A SONOS transistor with f_C = 10Hz has $\tau_{TT} = 10^{-1}$ s. Assuming $D_{it} = 10^{10}$ $cm^{-2}eV^{-1}$, the extracted tunnel oxide thickness is 17.5Å.

First, the tunneling time constant, τ_{TT}, is calculated by taking the reciprocal of the corner frequency, where the charge pumped begins to increase drastically. The interface trap density is determined from the charge pumping measurement at high frequency, typically 100kHz. Cross-referencing these values on the plot in Fig. 5.12 will yield the tunneling distance, or the tunnel oxide thickness. Since the charge pumping current increases at low frequencies when the charge has enough time to tunnel from the nitride to recombine in the substrate, we know the tunneling distance is equal to roughly the tunnel oxide thickness [17].

5.4 Quasi-Static Capacitance (Linear Voltage Ramp)

The Linear Voltage Ramp measurement is a quasi-static C-V measurement used to determine the effective thickness of the ONO dielectric in a SONOS device. The standard LVR setup, Fig. 5.13, involves connecting the source and drain together and tying them to the bulk. A function generator, controlled by a PC using LabVIEW™, sends a ramping voltage to the bulk of the device. The gate current is measured using an electrometer, which is controlled by a PC using LabVIEW™. The ramping voltage can be expressed as

$$Vgb = V_O \pm \alpha t \tag{5.9}$$

where V_O is a DC voltage level. The measured gate current is given by

$$I_g = \frac{\partial Q_g}{\partial t} A_G = \frac{\partial Q_g}{\partial V_{gb}} \frac{dV_{gb}}{\partial t} A_G = \alpha C_{eff} A_G = \alpha \frac{\varepsilon_{ox}}{X_{eff}} A_G \tag{5.10}$$

where $x_{eff} = x_{OT} + x_{OB} + \frac{\varepsilon_{OX}}{\varepsilon_N} x_N$ is the effective thickness of the ONO dielectric, ε_{OX} is the oxide permittivity, and A_G is the area of the gate.

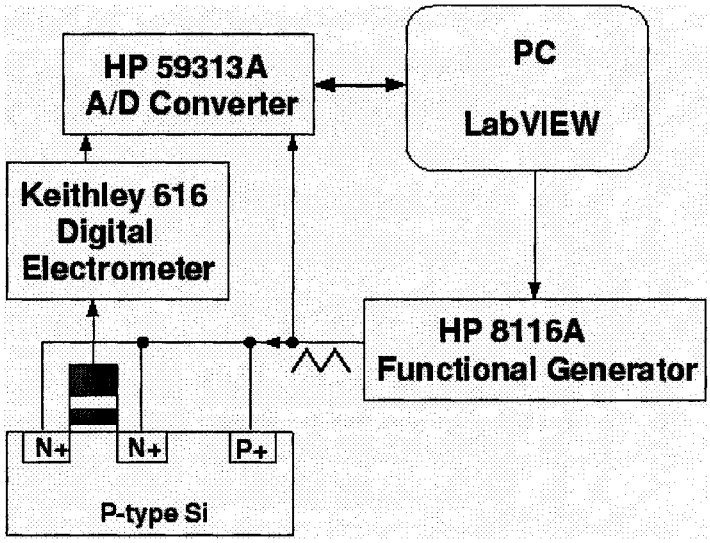


Figure 5.13: Linear Voltage Ramp Setup [10].

The LVR measurement can be used to examine the static memory window of a SONOS device. The voltage ramp starts at a positive value where the semiconductor surface is in accumulation. The nitride traps are filled with holes. Sweeping the bulk voltage from a positive to a negative value causes the surface of the semiconductor to go from accumulation to inversion. As the surface goes to inversion, electrons are injected and trapped in the nitride. This shifts the flatband voltage negatively. Fig. 5.14 shows the result of an LVR measurement on a SONOS capacitor on an N+ gridded p-type wafer, where V_{GB} is the horizontal axis. This measurement was performed at a ramp rate of 200mV/s. The effective oxide thickness is extrapolated as 81Å. Ellipsometry has shown the device oxide thickness as 79Å, which agrees with this experimental result. The device static memory window is extrapolated as 2.5V for a

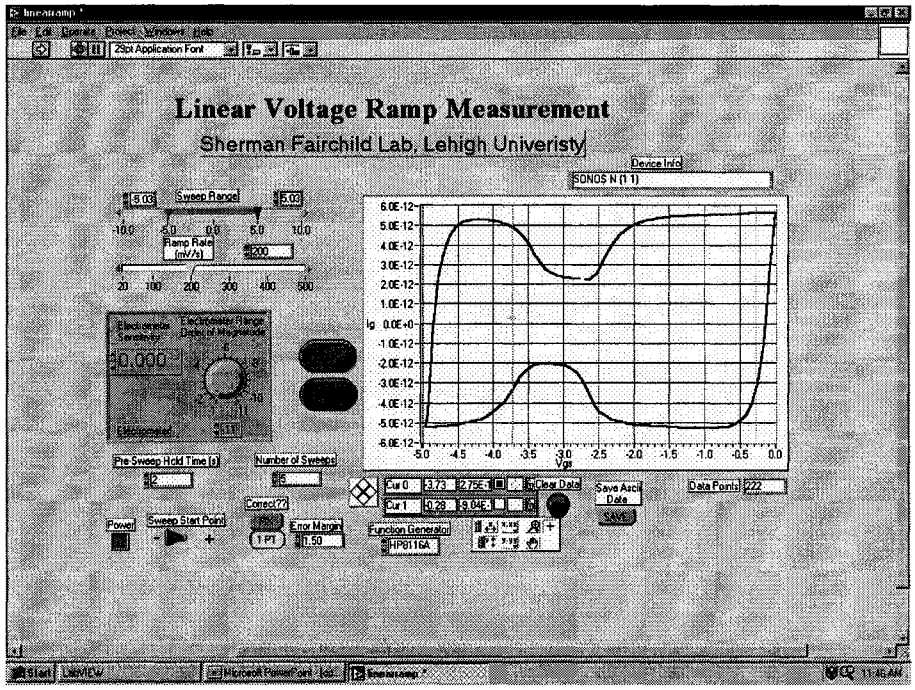


Figure 5.14: Front panel of LabVIEW™ LVR measurement program [21].

programming voltage of 7V [10].

The nitride charge storage film thickness in a SONOS device can also be determined from LVR measurements. The LVR measurement is repeated sweeping the bulk to gate voltage from +5V to -5V. The data is recorded and the measurement is repeated for +/- 6 and 7V. We assume all of the charge injected into the gate is trapped in the nitride as we sweep the gate voltage. Assuming the charge centroid in the nitride is located at half the nitride thickness, we can express the flatband voltage shift as [81]

$$\Delta V_{FB} = \Delta Q_n^{inj}\left[\frac{X_N}{2\varepsilon_N} + \frac{X_{OB}}{\varepsilon_{OX}}\right] \tag{5.11}$$

The flatband voltage shift is calculated for each LVR measurement. The injected charge

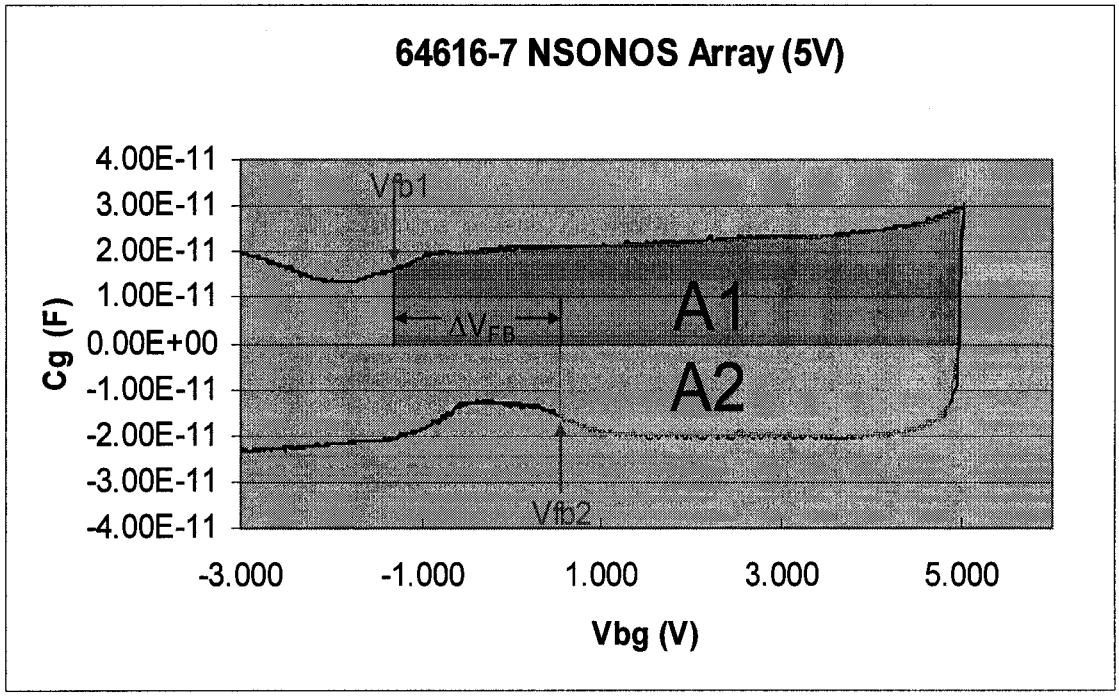


Figure 5.15: Linear voltage ramp measurement showing the flatband voltage shift (ΔV_{FB}). The charge is calculated from subtracting the injected holes (A1) and electrons (A2) [82].

is calculated for each measurement by subtracting the hole current from the electron current. This calculation is performed by subtracting the two areas shown in Fig. 5.15. Plotting the flatband voltage shift versus injected charge yields a straight line with slope

$$\lambda = \frac{X_N}{2\varepsilon_N} + \frac{X_{OB}}{\varepsilon_{OX}} \tag{5.12}$$

This allows the nitride thickness to be calculated assuming the blocking oxide thickness is known, or calculated using separate capacitance measurements and solving for the nitride thickness, as demonstrated in Chapter 6.

Chapter 6

Characterization of Scaled SONOS/NROM™/SNROM Devices for Space and Military Applications

The primary requirement for using an NVSM in space and military applications is radiation hardness. A memory device which has nanosecond programming, milliwatt power dissipation, and 10-year data retention is useless if it cannot withstand a 100krad lifetime radiation dose. The SONOS technology is inherently radiation-hard. The process of hardening a SONOS device for space and military applications is classified and will not be dealt with in detail in this dissertation. Exposing a transistor to radiation causes the NMOS off-state (leakage) current to increase. SONOS devices have been designed by the Northrop Grumman Corporation possessing low leakage currents at radiation doses as high as 1Mrad. The radiation tolerance of the NROM™ and SNROM

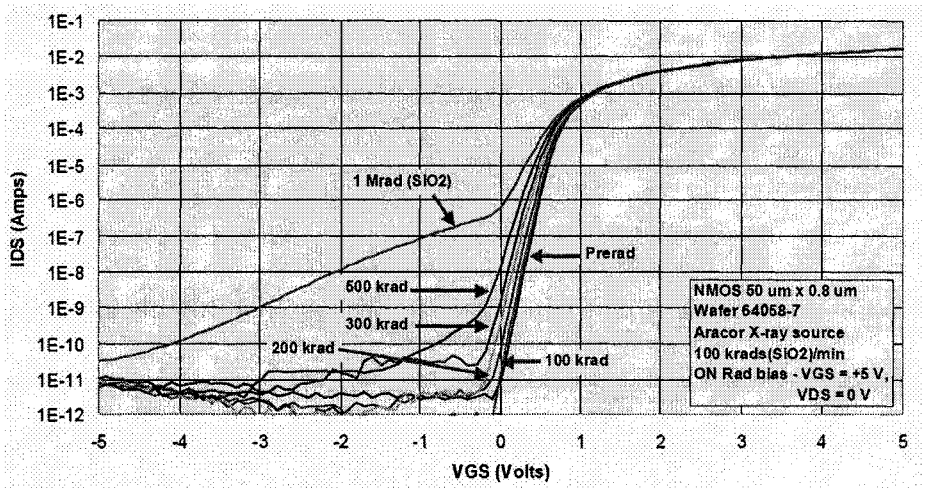


Figure 6.1: Off-state leakage in a radiated 50 x 0.8 μm SONOS transistor [1].

technologies has not been tested. However, these device structures are derivative of the SONOS technology and it is assumed that the radiation hard traits of the SONOS technology will carry over to these novel structures. In addition to radiation tolerance, programming time, programming voltages, power dissipation, data retention, and programming endurance must be taken into account when classifying an NVSM for space and military applications. This chapter investigates these aspects of programming in conventional and short channel multi-bit NROM™ and SNROM transistors using a variety of mechanisms, including Modified Fowler-Nordheim (MFN) tunneling, Channel Hot Electron (CHE) injection, Hot Hole Injection (HHI), and CHannel Initiated Secondary ELectron (CHISEL) injection.

Measurements are performed on floating trap SONOS/NROM™/SNROM nonvolatile memory devices with gate dielectrics consisting of a 2-4 nm tunneling oxide, 4-7 nm silicon-nitride charge-storage layer, and a 3.5-4 nm 'blocking' oxide underneath a phosphorus-doped polysilicon gate. The gate dielectric is programmed with tunneling by applying either a +7V or -7V pulse to the gate terminal. The applied gate voltage attracts electrons or holes to the surface of the silicon depending on the polarity of the gate voltage. These charges either tunnel through an ultra-thin "tunneling oxide" (SONOS) or are electrically excited to jump over the oxide barrier by a positive voltage applied to the drain (NROM™) or a positive bias applied to the drain and a negative bias applied to the bulk (SNROM), and store in traps within the nitride layer.

However, these innovative techniques for programming and storing charge have several drawbacks related to the nature of the charge injection mechanisms. NROM™

reliability issues have been reported involving damage to the gate dielectric due to hot hole and hot electron injection [51]. This is a consequence of the high electric fields required for these programming mechanisms to be successful. Charge pumping measurements are performed on these devices programmed using various mechanisms up to one-million write/erase cycles. The interface trap density is compared for various levels of stressing with the different programming mechanisms to determine the best method for writing and erasing the transistor quickly, with low voltages, to obtain two distinct memory states.

6.1 Programming Efficiency

The programming efficiency of an NVSM may be quantified by determining the voltages applied to the device during programming, the time duration of the programming, the current required for programming, and the disparity between the to memory states produced by the programming action. First, the memory window produced is examined. The maximum memory window producible is a function of the nitride thickness and blocking oxide thickness. The threshold voltage shift may be expressed as

$$\Delta V_{TH} = \Delta Q_N \left(\frac{X_{OB}}{\varepsilon_{OX}} + \frac{X_N}{2\varepsilon_N} \right) \tag{6.1}$$

where Q_N is the charge trapped in the nitride film, X_{OB}, ε_{OX}, X_N, and ε_N are the blocking oxide thickness and permittivity and nitride thickness and permittivity, respectively.

The centroid of the charge in the nitride is assumed to be located at the middle of the nitride film, $X_N/2$. The term $\left(\frac{X_{OB}}{\varepsilon_{OX}} + \frac{X_N}{2\varepsilon_N}\right)$ represents the 'lever-arm' of the charge stored in the nitride. The greater the film thicknesses, the greater a volume of charge stored in the nitride will shift the threshold voltage.

Therefore, since all three mechanisms are used to program SONOS, NROM™, and SNROM devices with similar nitride and blocking oxide thicknesses, the maximum memory window will be the same for each programming mechanism. The tunnel oxide thickness will not affect the maximum memory window achievable. However, a thin blocking oxide will allow electrons to be injected to the nitride from an N+ doped poly-silicon gate during erasure. The charge injected from the gate will eventually equal the

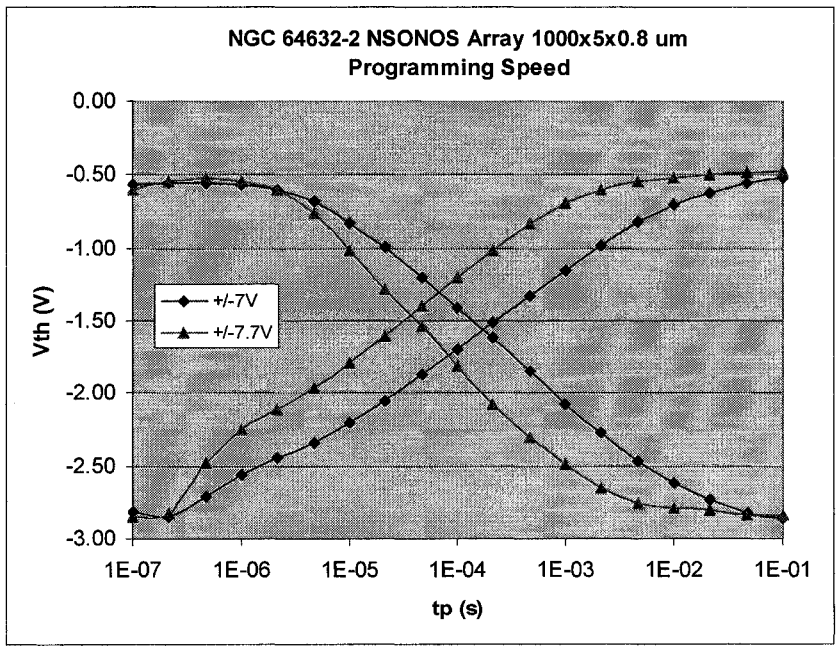


Figure 6.2: Programming an NSONOS Array of transistors (1000 x 5.0 x 0.8 μm) with MFN tunneling at gate voltages of 7 and 7.7 volts.

charge injected from the substrate, causing the negative edge of the memory window to be fixed after a specific programming time.

The initial programming window desired for a SONOS transistor is typically > 1.5 V. The charge decay rate is determined by the tunneling oxide thickness and is generally 100 mV/decade. Charge decay begins at around 1.0 seconds for a tunnel oxide thickness of ~18Å. To retain charge for 10 years, the device must have a 50 mV memory window after 8 decades of time, 3.15×10^8 seconds. Therefore, over approximately 8 decades of time, the memory window will close by 1.6 V, making the minimum memory window approximately 1.65 V. The initial memory window required will change with tunnel oxide thickness since the charge decay rates will also change.

The programming time will be determined by the voltages applied to the device and the thickness of the oxide-nitride-oxide (ONO) stack. A thinner bottom oxide will be easier for charge to tunnel through, allowing the device to be programmed quicker and with lower applied voltages. The programming times and voltages studied have been selected in order to produce a 2V initial memory window in order to ensure 10-year data retention for a variety of tunnel oxide thicknesses and programming voltages. Programming a SONOS transistor with a fixed gate dielectric thickness with MFN tunneling, the time where the write and erase curves intersect (cross-over) will decrease by one decade for every one-volt increase in gate voltage or 2Å decrease in tunnel oxide thickness [13].

Due to the fluctuations in effective gate dielectric thickness, tunnel oxide thickness, and physical film composition, it is difficult to make a comparison between

	Voltage (V)			
Mechanism	**Gate**	**Drain**	**Source**	**Bulk**
MFN	7	0	0	0
	-7	0	0	0
CHE	7	4	0	0
HHI	-7	4	0	0
CHISEL	3	1.5-2.5	0	-4

Table 6.1: Voltages required for programming a nitride-based NVSM using a variety of charge transport mechanisms.

devices designed and programmed with MFN tunneling and short-channel devices which use injection of hot electrons and holes. The Northrop Grumman Corporation uses MFN to program SONOS transistors with a 2V memory window by applying a 2.5 ms write and a 7.5 ms erase. Therefore, we will program NROM™ devices with CHE and SNROM devices with CHISEL using 5 ms write and erase to produce a 2V initial memory window. Enhancements to the programming speeds are possible once the device design is optimized for programming with channel or substrate hot electrons.

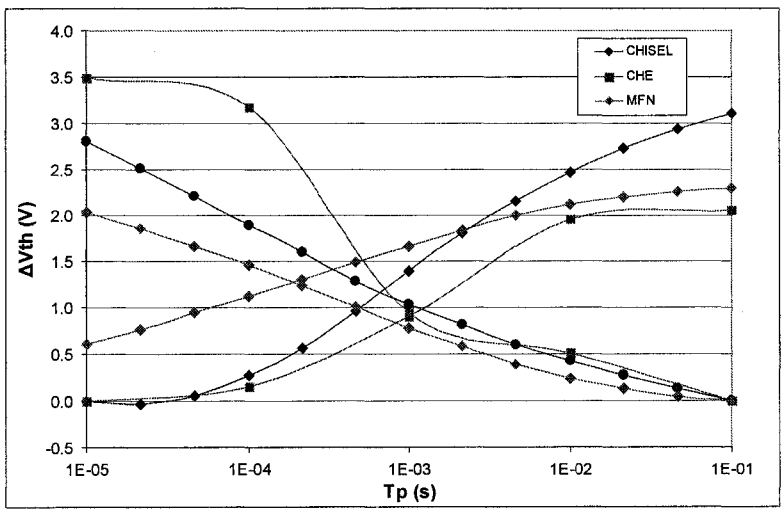


Figure 6.3: Programming nitride-based transistors with CHISEL, CHE, and MFN tunneling (voltages outlined in Table 6.1).

The power dissipated by an NVSM device during programming is a function of the applied voltages and the current required for the charge transport mechanisms to occur. The programming voltages used for all mechanisms described in this dissertation are approximately 7V gate-to-bulk and 0-4 V drain-to-source. The currents required for effective charge injection differ greatly. MFN tunneling requires very little current

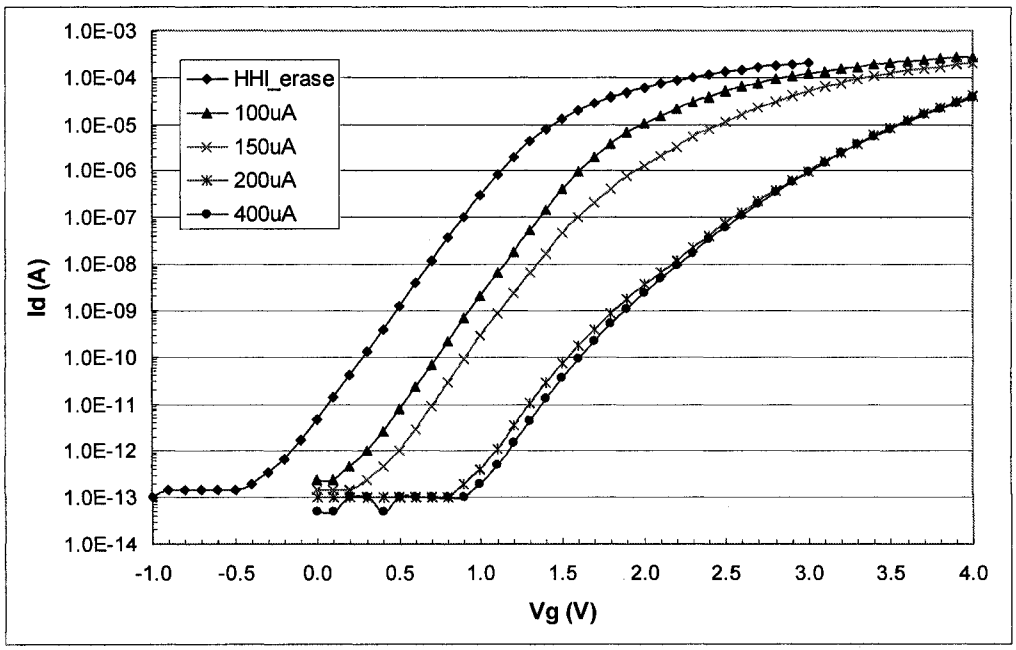


Figure 6.4: Threshold voltage after programming an SNROM (10 x 0.22 μm) transistor with hot hole injection (diamonds – V_G = -7V, V_D = +4V), channel initiated secondary electron injection (– V_G = +3V, V_D = +1.5V, V_B = -4V) while limiting channel current.

compared to CHE and HHI often require milliamps of current. CHISEL injection may be performed using hundreds of micro-amps of current for reduced power dissipation compared to CHE/HHI. Currents during CHE and CHISEL injection are characterized by programming a 10 x 0.22 μm NROM™ and SNROM transistors with a 4145 parameter analyzer while fixing the channel current and floating the source terminal.

The SNROM transistor programmed with CHISEL produced a 1.0V memory window at 100μA and a 2.0V window with 200 μA of current.

Electrical characterization of the channel current during CHISEL programming demonstrates the influence of the bulk bias over the CHISEL mechanism. A significant reduction in the channel current indicates an increase in the threshold voltage of the device, a result of the CHISEL injection. Electrical measurements demonstrate that CHISEL injection occurs at a lower drain bias for a more negative bulk bias. In addition, the drain current drops more rapidly under high negative bulk biases, indicating the CHISEL injection is more efficient with more electrons ionizing, injecting, and trapping in the nitride.

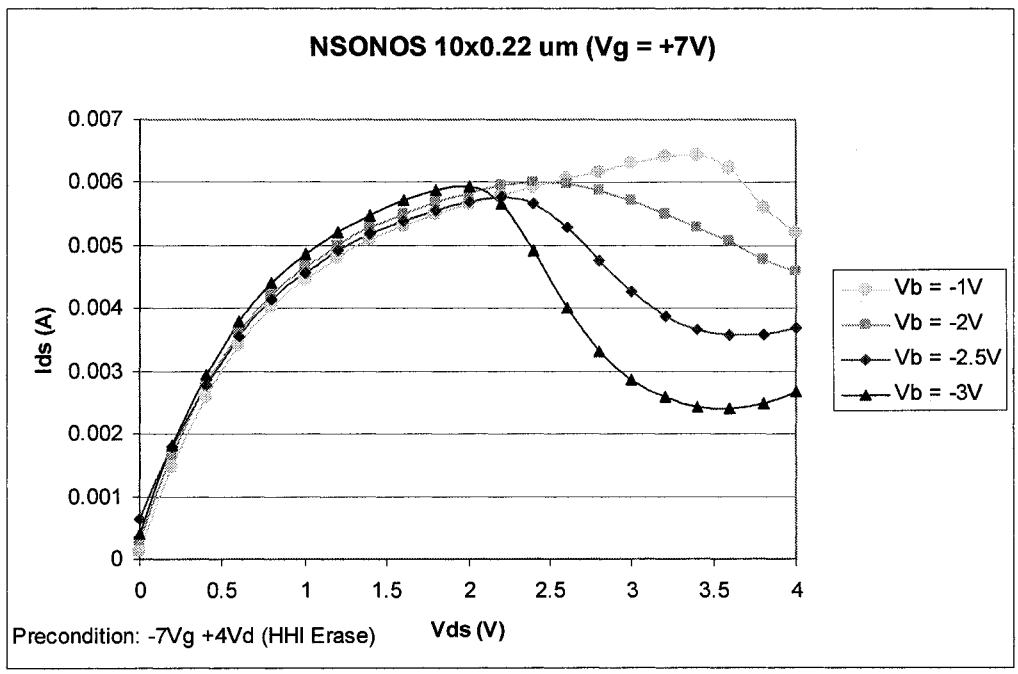


Figure 6.5: SNROM Drain current measurements demonstrating the decrease in drain bias required for CHISEL injection as the negative bulk bias is increased.

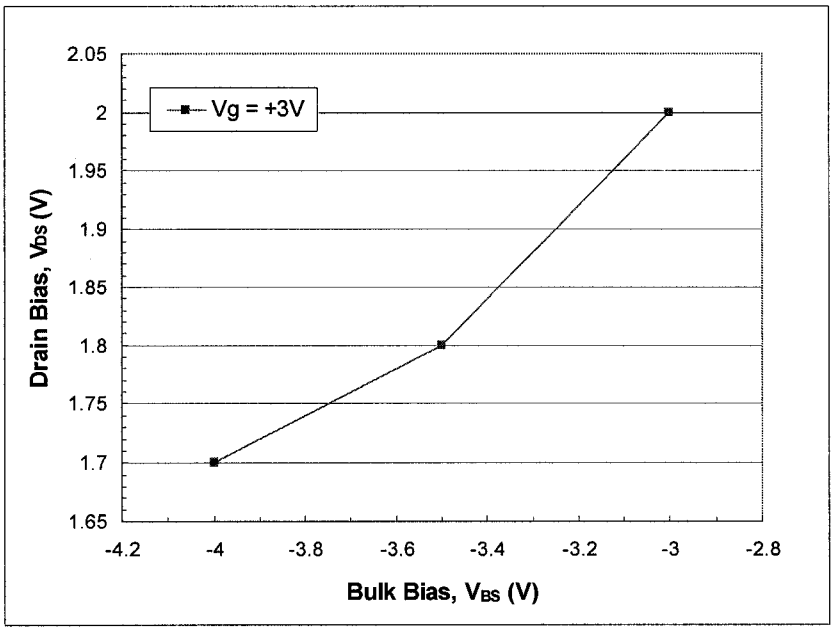


Figure 6.6: SNROM Drain bias required for CHISEL injection for various bulk biases. Gate voltage is fixed at +3 V for all measurements.

6.2 Extrapolating Memory Retention

Nonvolatile memory devices are designed to retain data for up to 10 years. This data, a logic one or zero, is stored as positive or negative charge trapped in the nitride in order to create a threshold voltage differential. To test for data retention, we program the device and measure the threshold voltage repeatedly for every decade of time that passes. The measurement can typically be carried out to 10^3 seconds (16 minutes 40 seconds) for short-term measurements to obtain the initial memory window magnitude and position, 10^4 seconds (2 hours 40 minutes) for most measurements, and 10^5 seconds (28 hours) for long-term measurements. The goal is determining whether enough of the charge stored will remain after 10 years (3.15×10^8 seconds). Therefore, modeling the

charge loss mechanisms and charge storage properties of the transistor provides a number of methods for extracting future retention performance.

Lehigh University has characterized the data retention characteristics of SONOS transistors fabricated by the Northrop Grumman Corporation. Data retention

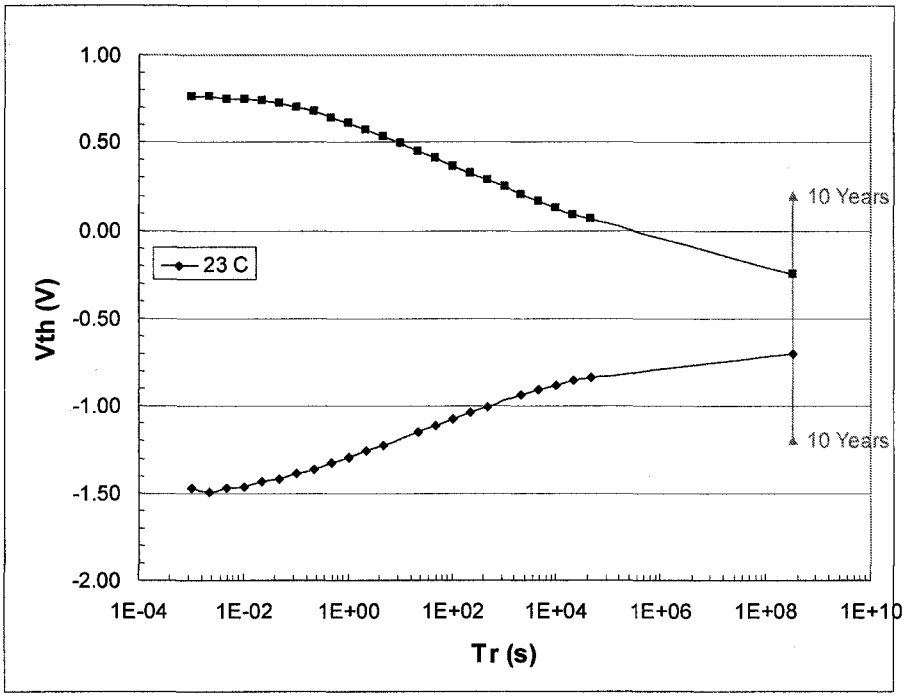


Figure 6.7: Extrapolated retention of a 50 x 0.8 μm SONOS transistor.

measurements are performed on arrays of 1000 NSONOS 5.0 x 0.8 μm transistors connected in parallel. The transistor arrays are written by applying +7V to the gate for 2.5ms while grounding the source, drain and bulk. The array is erased by grounding all terminals and applying -7V to the gate for 7.5ms. Programming voltages and programming times are controlled by a LabView™ environment running on a PC connected to a dynamic electrical characterization circuit developed at Lehigh University [44]. Data retention at 10 years is predicted using a linear trendline from the

last 3 or 4 points from the retention measurement.

The critical element for long-term data retention at room temperature is the charge decay rate. The charge decay rate is determined by the tunnel oxide, nitride, and blocking oxide thicknesses. The tunnel oxide thickness determines the ability of charge

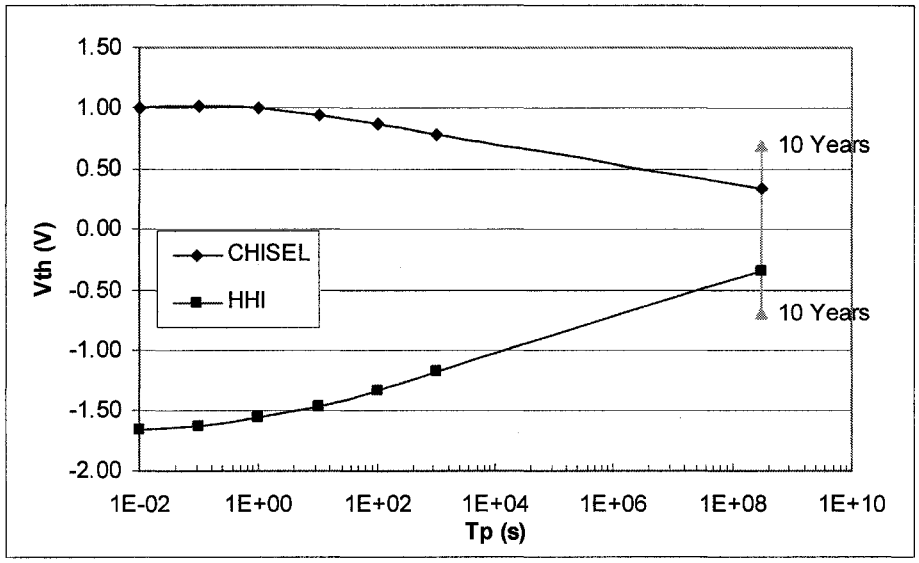


Figure 6.8: Memory retention of a 10 x 0.22 μm SNROM transistor extrapolated to 10 years. The memory window at 10 years is ~0.8 V with standard CHISEL/HHI programming conditions.

to tunnel to and from the nitride, and the nitride and blocking oxide thicknesses determine the 'lever arm' of the trapped charge, giving a loss or gain of charge greater influence on the threshold voltage with thicker films. Retention measurements are performed on conventional SONOS transistors written and erased with MFN tunneling, short-channel NROM™ transistors written with CHE and erased with HHI injection, and short-channel SNROM transistors written with CHISEL and erased with HHI injection. The initial memory window for all of these devices was 1.5 – 2.5 V. The average memory window extrapolated at 10 years was approximately 0.5 – 0.8 V for an

NROM™ or SNROM device programmed under the conditions outlined in section 6.1. The short-channel NROM™/SNROM transistors have tunnel oxide thicknesses of 36.1 Å and the conventional SONOS transistors have tunnel oxide thicknesses of 18 Å. This corresponds to a lower charge decay rate, about 80 mV/decade, for the NROM™ and SNROM transistors.

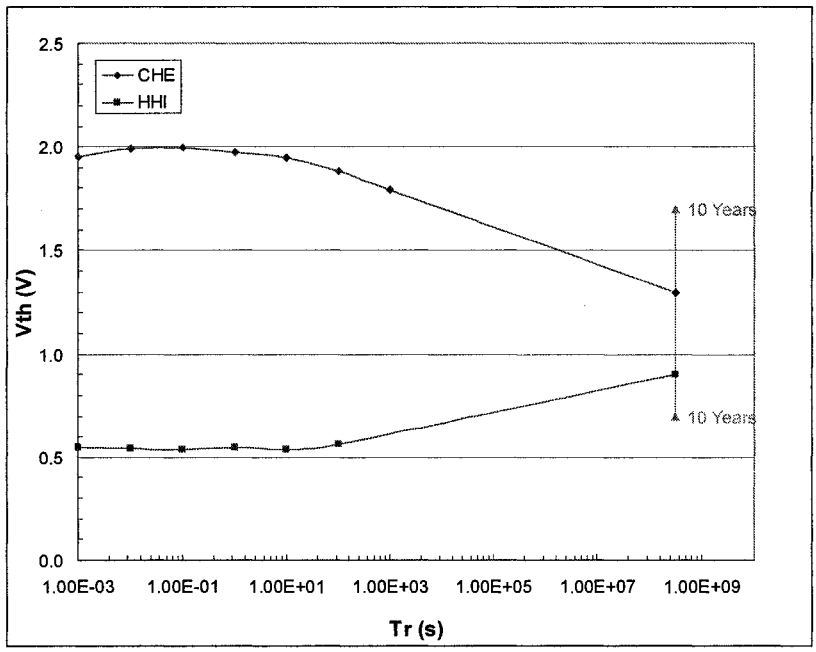


Figure 6.9: An NROM™ 10 x 0.22 μm transistor programmed with CHE and HHI. The 10-year memory window is approximately 0.5V.

Ideally, we would like to identify an 'acceleration factor' for data retention at elevated temperatures. At elevated temperatures the electrons are thermally excited and escape from traps in the nitride and back-tunnel to the substrate. A greater number of electrons will be excited as temperatures increases, resulting in additional loss of charge. High temperatures accelerate the loss of charge and allow us to perform a screening measurement on the SONOS devices. For example, if an acceleration factor

of 1000 (3 for log [tr]) is observed over 100 degrees, a SONOS device retaining charge at 150 C for 10^5 seconds is equivalent to a SONOS device retaining the same quantity of charge at 50 C for 10^8 seconds, roughly 10 years. The acceleration factor is calculated by finding the point where the write and erase states converge for a given high and low temperature. The change in the cross-over point between the two temperatures is the acceleration factor.

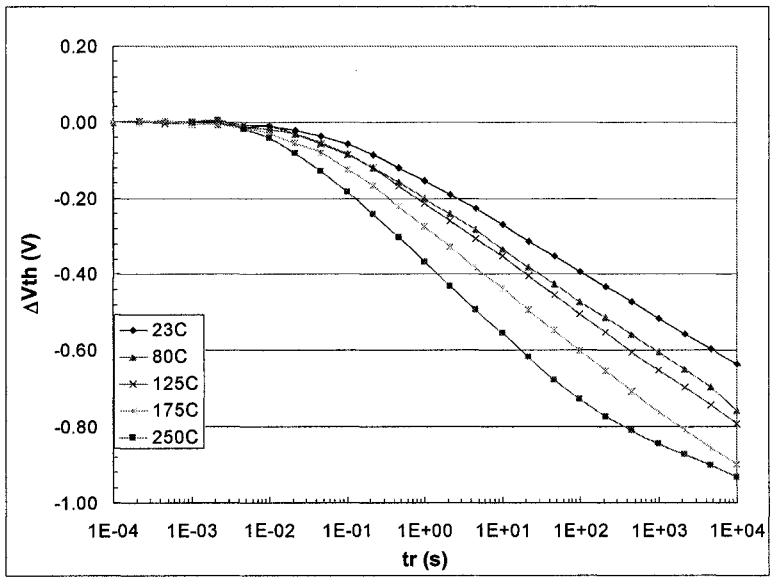


Figure 6.10: Decay from the initial write state threshold voltage of an NSONOS Array (1000 x 5.0 x 0.8 μm) at various temperatures between 23 (room) and 250 °C.

First, we must assume that the device may only lose charge while in the retention mode. Therefore, the threshold voltage may not cross the mid-point of the memory window. The memory state is created by storing equal amounts of holes or electrons in the nitride. The neutral state of the device will occur at the mid-point of the initial memory window. The memory state reaching the memory window mid-point will be referred to as 'memory window closure'.

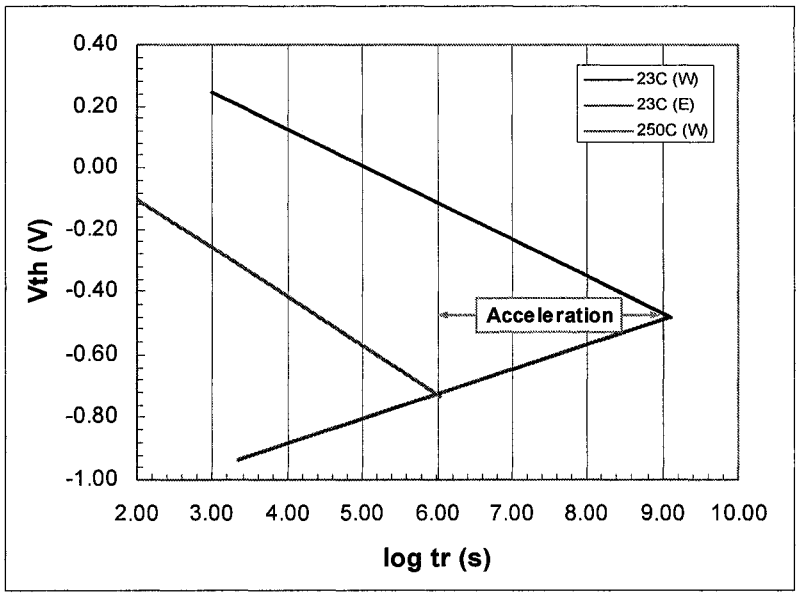


Figure 6.11: Extrapolated retention at 23C (dark) and 250C (light) out to intersection with extrapolated erase at 23C. Acceleration marker notes 3 decade shift in window closure point, identifying the acceleration factor of 2.95, varying by 0.013/°C.

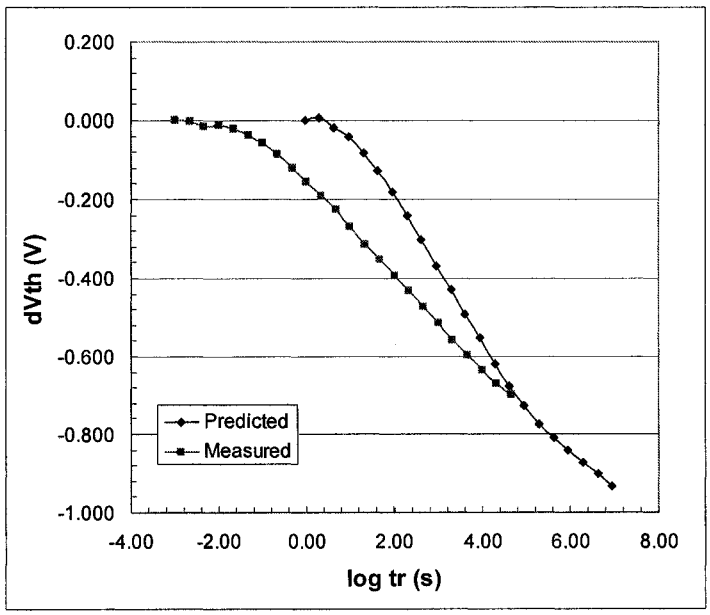


Figure 6.12: Threshold voltage decay at room temperature (light) and predicted room temperature retention (dark) from retention data at 250C. A log-time acceleration factor of 2.95 is used, meaning the time axis is shifted roughly 3 orders of magnitude. This data assumes threshold voltage decay rate is a linear function of temperature.

Examining how the threshold voltage decay rate changes at elevated temperatures gives us some insight into the charge decay. The decay rate is affected by a number of factors. The tunnel oxide thickness determines the ability of charge to back-tunnel to the substrate. The location, energetically and spatially, in the nitride is important as well. If the charge is located in deep energy traps, it will require higher temperature to excite the charge from its traps. Traps which are physically closer to the tunnel oxide will be more likely to loose their charge due to tunneling through the thin oxide barrier. These factors are greatly affected by the thicknesses of the tunnel oxide and nitride, the physical composition of the nitride, and the quality of all films involved. Therefore, it is difficult to quantify many of these factors and determine a simple acceleration factor (i.e. 3X charge loss rate) for retention measurements performed at various temperatures.

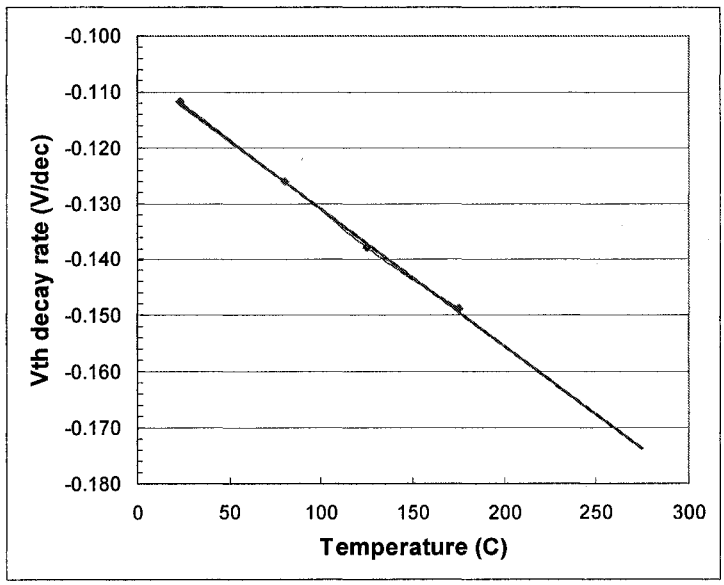


Figure 6.13: NSONOS array write state threshold voltage decay rate vs. temperature.

The threshold voltage decay rate remains constant after 1 second for all temperatures tested except for 250 C. The change in the decay rate versus temperature for this region is calculated as 2.4 mV/C, where the decay rate increases 2.4 mV/dec for every degree C increase in temperature. Depending on the tunnel oxide and nitride thicknesses, trap density, location of the injected traps, and interface trap density, the threshold voltage decay rate changes between 2 – 5 mV/dec for every degree change in temperature. Increasing the temperature will increase the magnitude of the decay rate.

Retention measurements at elevated temperatures and long-term room temperature measurements show that the decay rate decreases over time, typically after 10^3 to 10^4 seconds at room temperature. Examining the slope of the threshold voltage decay versus time shows a parabolic trend in the charge decay rate. This is most likely a result of the parabolic nature of the trap distribution within the silicon nitride layer.. However, there is a discrepancy in the shape of the trap distribution in the nitride film.

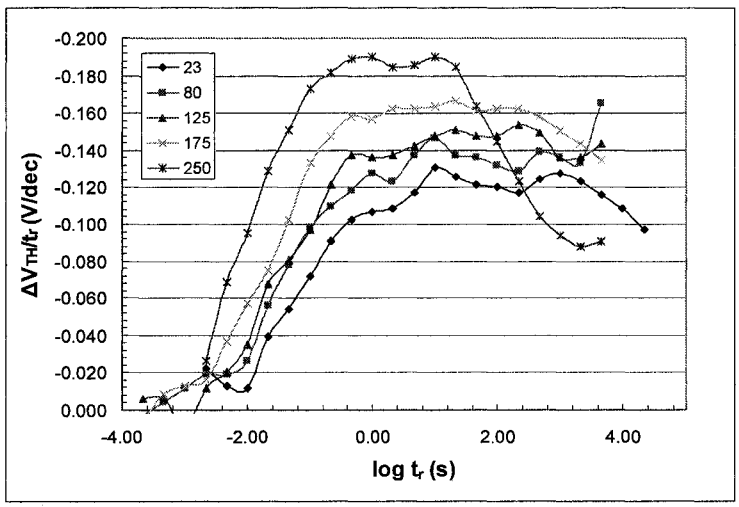


Figure 6.14: Threshold voltage decay rate of the write state versus the logarithm of retention time for various temperatures.

Wang and White [63] have proposed a model which uses a fourth-order fit to the trap distribution. However, this fit does not center the trap distribution at 1.1 eV, the accepted activation energy for electron traps in a silicon-nitride film [18, 68, 71]. Future research must examine the analysis of the trap densities within the silicon nitride film to provide further accurate analysis of the predicted charge decay.

The charge decay rate drops and levels off, producing a linear threshold voltage past this point. The room temperature data behaves similarly, but the decay rate change is not as drastic and occurs over a longer period. It is difficult to determine how the change in the slope of the threshold voltage decay will translate to room temperature since retention measurements taken out past one week (10^5) seconds are not practical.

The variation of charge loss rates from device to device heavily depends on the physical composition, quality, chemical composition, spatial and energetic trap distribution of the nitride layer. Therefore, we suggest the 10-year retention should be

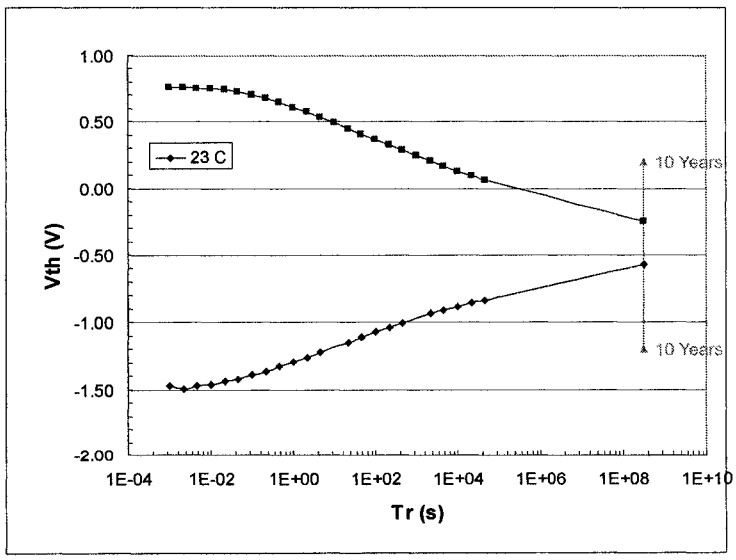


Figure 6.15: NSONOS (50 x 0.8 μm) retention programmed with MFN tunneling.

extracted from a straight line after the retention curve levels partially. This will provide a somewhat pessimistic yet realistic estimation of 10 year data retention. The primary difficulty of forecasting retention is the fact that it is difficult to know what the actual retention characteristics will look like after months and years of data retention. The most accurate method for forecasting data retention at up to 10 years is extracting the energetic trap distribution in the nitride from elevated temperature measurements and apply these results to room temperature retention factoring in film thicknesses and interface trap contributions to charge loss.

6.3 Endurance

The ability of an NVSM device to operate after being repeatedly written and erased is known as endurance. After a small number of write/erase cycles, the electric field at the injection interface is enhanced by the hole trapping in the tunnel oxide, leading to an enlarged memory window. During each cycle, electrons tunnel through the bottom oxide, with a portion of the electrons trapping in defects in the oxide. At the same time, under high local electric fields, additional electron traps are generated in the oxide by dislodging electrons from covalent Si-SiO_2 bonds. As the number of write/erase cycles increases, more electrons are trapped in the oxide until the electron trapping outweighs the hole trapping. The net charge in the oxide causes a reduction in the injection field and inhibits further tunneling and CHE injection [83].

SONOS, NROM™, and SNROM devices are written and erased using combinations of MFN tunneling, CHE, CHISEL, and hot hole injection. The first

programming mechanism examined is modified Fowler-Nordheim (MFN) tunneling in an NROM™ transistor with X_{OT} = 40Å, X_N = 42Å, X_{OB} = 44Å. NROM™ devices are characterized by thicker bottom oxides compared to their SONOS counterparts. The NROM™ device is written by applying +12V to the gate and erased by applying -12V to the gate. The source, drain, and bulk are grounded. After erasing, the device has a threshold voltage of 2.5V, and after writing, V_{TH} is 5.7V. The memory window, the difference between the write and erase state threshold voltages, is 3.2V. A large gate voltage (12V) must be used to generate high electric fields in the device, since the tunneling oxide is rather thick (40Å), which prevents MFN tunneling at voltages <10V in the NROM™ transistor. The device is stressed at 10^4, 10^5, and 10^6 cycles and D_{it} is measured at each stage. The interface trap density increases by a factor of 10.8 when the device is stressed to 10^5 cycles and increases by a factor of 20.5 at 10^6 cycles.

Modified Fowler-Nordheim tunneling is utilized to program a SONOS transistor with X_{OT} = 20Å, X_N = 70Å, X_{OB} = 40Å. This transistor has a thinner bottom oxide than the NROM™ transistor to enable tunneling. The use of tunneling permits the device to be programmed with 7V on the gate to obtain a memory window of 3V. The device is stressed and the charge pumping measurements show the D_{it} increases by a factor of 6 after 10^5 cycles, and increases by a factor of 14 after 10^6 cycles.

Finally, hot electron and hot hole injection are used to program an NROM™ transistor with X_{OT} = 4 nm, X_N = 4.2 nm, X_{OB} = 4.4 nm. The device is written by applying +7V to the gate and +4V to the drain while grounding the source and substrate. The device is erased by applying -7V to the gate and +4V to the drain while grounding

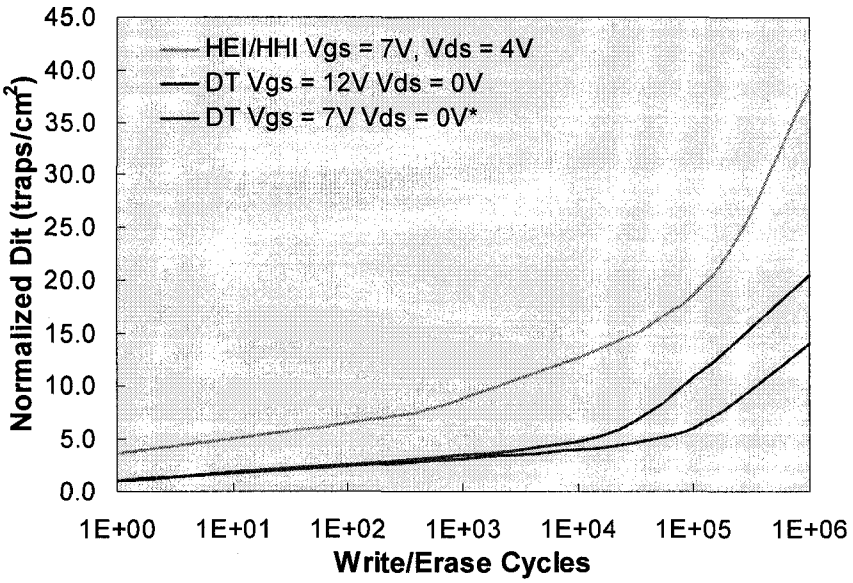


Figure 6.16: Normalized interface trap density versus cycling for 20 x 0.25 μm NROM™ transistors. The interface trap density doubles when programming with CHE/HHI versus DT/HHI. *Conventional SONOS transistor [84].

the source and substrate. The threshold voltage of the erase state is 3V and the write state threshold voltage is 6V, yielding a memory window of 3V. This device is optimized for CHE/HHI injection allowing for low programming voltages (≤7V). The device is stressed up to 10^6 cycles and charge pumping measurements determine the interface trap density increases by a factor of 18.6 at 10^5 cycles, and 38.3 at 10^6 cycles. The charge pumping measurements on the SONOS/NROM™ transistors confirm the results of recent research demonstrating that hot electron injection (CHE) and hot hole injection (HHI) cause greater damage to the gate dielectric during programming than Modified Fowler-Nordheim (MFN) tunneling [50, 85].

Conventional SONOS transistors are programmed with modified Fowler-Nordheim (MFN) tunneling under the conditions outlined in section 6.1. The memory

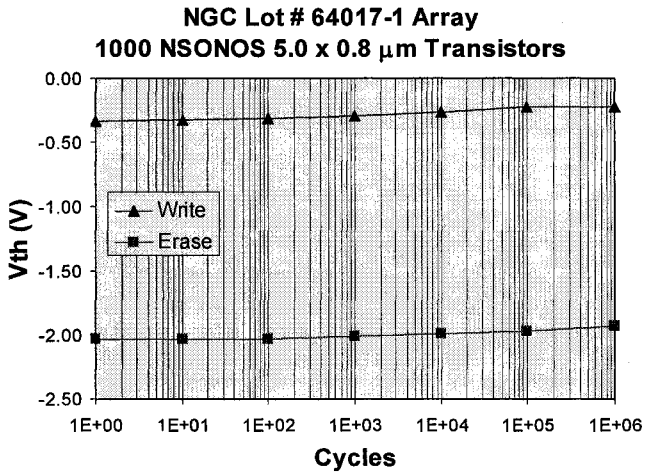


Figure 6.17: Initial memory window versus cycling for an array of 1000 NSONOS 5.0 x 0.8 μm transistors [1].

window shifts positively as the device is stressed due to the buildup of negative charge in interface traps which are generated by extended programming cycles, Fig 6.17. The initial memory window of the NSONOS array shifts positively less than 200 mV as the device is stressed 10^6 cycles. This memory window shift is not large enough to degrade long-term device performance.

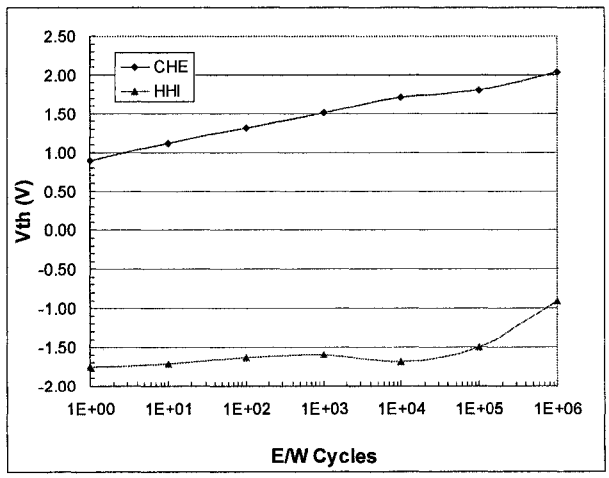


Figure 6.18: Cycling a 10 x0 0.22 μm NROM™ transistor with CHE/HHI.

The memory window shift was also monitored for short-channel NROM™ devices written using channel hot electron (CHE) injection and SNROM devices programmed with channel initiated secondary electron (CHISEL) injection. Both the NROM™ and SNROM devices are erased using hot hole injection (HHI). The memory window for the NROM™ device shifted positively one volt. The SNROM device shifted a mere 250 – 300 mV, which is comparable to the SONOS device programmed with modified Fowler-Nordheim (MFN) tunneling.

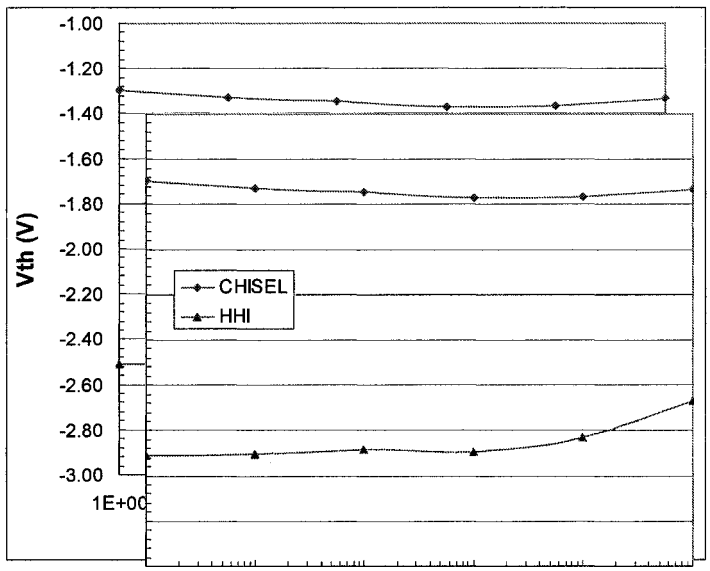


Figure 6.19: Cycling a 10 x 0.22 μm SNROM transistor CHISEL/HHI.

Device retention is characterized for a variety of cycling stress levels. Fig. 6.20 shows data retention of an array of 1000 NSONOS 5.0 x 0.8 μm transistors at room temperature for various degrees of write/erase cycling (+/-7V). The device is written and erased for 10^4, 10^5, and 10^6 cycles. The charge decay rate increases significantly for the erase state due to the build-up of interface traps which promote back-tunneling of

holes from the nitride. The virgin device has a 10-year memory window of 250mV. When this device is stressed, the 10-year memory window remains 250mV at 10^4 and 10^5 cycles, and reduces to 120mV at 10^6 cycles, exceeding the requirement of a 50 mV window at 10 years after 10^4 cycles.

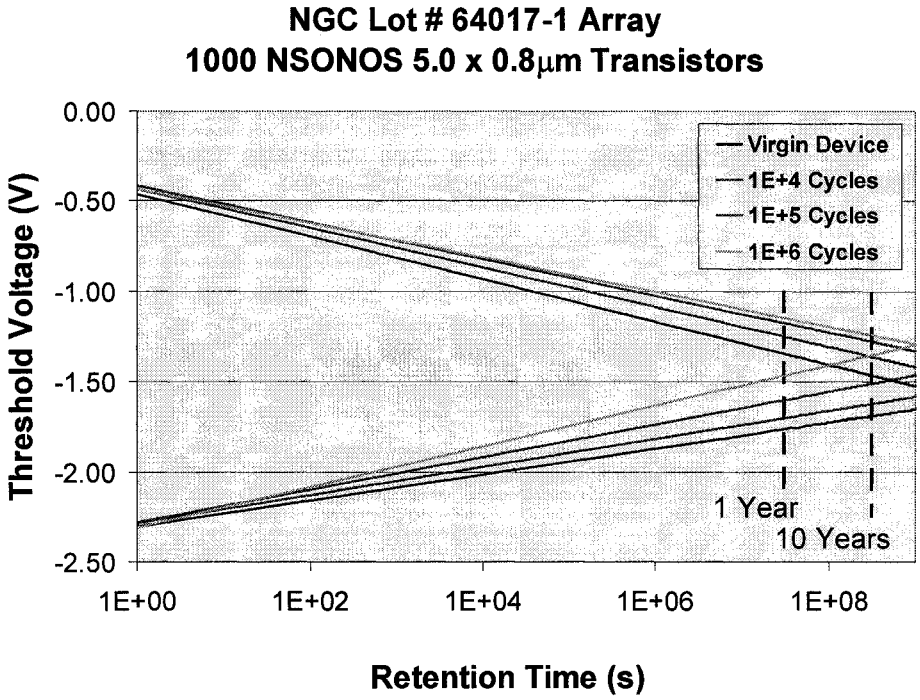


Figure 6.20: Data retention for various levels of cycling of an array of 1000 NSONOS 5.0 x 0.8 μm transistors. The 10-year memory window reduces to an acceptable 120 mV after 10^6 write/erase cycles [1].

6.4 Ideal NVSM for Space and Military Applications

In order to determine which nitride-based NVSM discussed in this dissertation is most suitable for space and military applications, we must consider the whole of the device characterization presented in this chapter. Electrical characterization is

performed on conventional SONOS transistors with gate dielectric films thicknesses as follows; 18Å tunnel oxide (X_{OT}), 45Å oxynitride (X_N), and a 35Å blocking oxide (X_{OB}). These transistors are written by applying +7V to the gate for 2.5ms and erased by applying -7V to the gate for 7.5ms, all other terminals were grounded. Short-channel NROM™ and SNROM transistors had a gate dielectric consisting of; 36.6Å X_{OT}, 30.3Å X_N, and 36.1Å X_{OB} determined from transmission electron microscopy (TEM) and capacitive measurements. The NROM™ transistors are written and erased in two different configurations; direct tunneling write and erase (DT) by applying +/- 12V to the gate for 5 ms each and channel hot electron injection write by applying +7V to the gate and +4V to the drain for 5 ms and a hot hole injection erase by applying -7V to the gate and +4V to the drain for 5 ms (CHE/HHI). The SNORM devices are programmed with channel initiated secondary electron injection by applying +3V to the gate, -4V to the substrate, and 1.5 – 2.5 V to the drain (depending on transistor channel length) for 5ms each (CHISEL/HHI).

First we must examine programming voltages, currents, and timing. The programming voltages are similar for all technologies, including a gate-to-bulk combined bias of 7 volts. The MFN programming has an advantage over CHE-CHISEL/HHI since no drain voltage is required. The programming current is lowest in SONOS devices using MFN tunneling. SNROM CHISEL injection uses less current than NROM™ CHE injection, and coupling this decreased programming current with the ability to scale voltages as the device geometry is scaled will yield future power consumption improvements. The programming time is equivalent for all mechanisms

and devices since the combined write and erase times are a total of 10ms.

The programming conditions for each of the mechanisms have been tailored to produce a 2-volt initial memory window. In addition, 10-year data retention is obtained with each of the programming mechanisms on reasonably designed transistors. The CHE/HHI and CHISEL/HHI programmed devices produce 0.5-0.8V memory windows at 10 years, due to the thickened tunneling oxide and reduced threshold voltage decay rates. Conventional SONOS devices produced acceptable 10-year memory windows, which could be increased by extending the write and erase times.

The conventional SONOS device has the best endurance performance. The electric fields are lowest in this device as is the damage to the Si-SiO_2 interface. The short-channel SNROM device programmed with CHISEL injection did not degrade as fast as the NROM™ device programmed with CHE injection. Future research monitoring the interface trap density will provide further analysis and characterization of the CHISEL mechanism.

Transistor density is the final parameter scrutinized. All of the charge transport mechanisms can be used to program a sub-micron scaled SONOS transistor. However, MFN tunneling only stores one bit per transistor, whereas the CHE programmed NROM™ and CHISEL programmed SNROM devices store two bits per transistor. This is the major advantage of these novel programming mechanisms. Charge is injected at the source or drain where a positive bias is applied during programming allowing one bit to be stored at either side of the gate dielectric. The improved density and ability to store two bits simultaneously give the SNROM technology the edge for space and

military applications. This technology utilizes programming voltages, times, power, and initial memory window comparable to conventional SONOS devices. In addition, the The write state decay rate of the SNROM is lower than that of single-bit SONOS devices because of the thicker tunneling oxide. Programming endurance is better than NROM™ programming and is not a major concern of the NVSM designer, as these

	Space & Miltary	Floating Gate	SONOS	NROM™	SNROM	FeRAM	MRAM	Phase Change
Area	$5F^2$	$5F^2$ - 70 nm	$6F^2$ - 0.13 µm	***2.5-5F² - 0.13 µm***	***2.5-5F² - 0.13 µm***	$15F^2$ - 0.25 m	$8F^2$ - 0.18 µm	Large
Voltage	5 - 7	> 10	3 - 7	3 - 10	***-3 - +2***	***1 - 3.3***	***1 - 3.3***	***1 - 3.3***
Current	Low	High	Low	High	Moderate	***Very Low***	Very High	Very High
Endurance (Cycles)	$> 10^4$	10^5	10^6	$> 10^5$	$> 10^5$	$10^8 - 10^{13}$	∞	$> 10^{12}$
Retention (Years)	> 10	> 10	> 10	> 10	> 10	10	~10	∞
Radiation Dose	***300 krad***	10 - 100 krad	***1 Mrad***	1 Mrad	1 Mrad	--	--	***2 Mrad***
Memory Density	1 Mb	***1 Gb***	1 Mb	***4 Mb***	***4 Mb***	***32 Mb***	***4 Mb***	***4 Mb***
Programming Speed	< 10 ms	***µs***	ms	***µs***	***µs***	***ns***	***ns***	***ns***
CMOS	Y	Y	Y	Y	Y	N	N	Y
Extra Masks	+ 2	+ 12	+ 2	+ 5	+ 5	+ 3-4	+ 3-4	+2-4

Table 6.2: Summary of the capabilities of NVSM's for space and military applications. Parameters which are excellent are shown in bold-italics, trouble areas are bold.

devices are not intended to be re-programmed repeatedly. The programming voltages, current, and timing may be improved with future SNROM device optimizations for CHISEL injection. The floating gate technology has been optimized for CHISEL injection and research has shown CHISEL to be three times more efficient than CHE injection and uses micro-amperes of current during programming [52]. These

advantages over CHE injection should carry over to the nitride-based SNROM, since CHE injection performance was demonstrated in nitride-based NROM™ devices after being pioneered in floating gate devices.

Chapter 7

Conclusions

This dissertation concerns the electrical characterization, modeling, and classification of SONOS NVSM transistors for space and military applications. In this dissertation, SONOS, NROM™, and SNROM transistors are programmed using a variety of charge transport mechanisms, and the resulting programming efficiency, memory effect, and reliability of the devices are electrically observed and modeled.

7.1 Characterization and Modeling of CHISEL Injection in SNROM NVSM Transistors

For the first time, short-channel SNROM transistors are programmed with channel initiated secondary electron (CHISEL) injection. Charge injection is monitored for varying electric fields applied to the device. The threshold voltage shift is modeled from the change in the channel current during programming. Charge injected and trapped in the nitride as well as injection current required to produce this threshold voltage shift are extracted. The number of electrons injected into the nitride is calculated as 36,000, demonstrating the low-current and low-power operation of the SNORM. In addition, for the first time, the current due to impact ionizations occurring in the channel is recorded and correlated to the current injected and trapped in the

nitride. SNROM devices have been programmed with CHISEL injection at low currents with acceptable programming times and voltages. Ten-year data retention has been predicted using linear extrapolation.

7.2 Predicting 10-Year SONOS Data Retention

Retention measurements are performed on a variety of conventional and short-channel SONOS devices programmed using a variety of charge injection mechanisms. Retention characteristics of these devices are measured for a variety of temperatures and stress levels. As expected, elevated temperature increases the charge loss mechanisms during retention. A leveling off of the charge decay is identified and indicates the linear extrapolation method commonly used is a pessimistic measure of 10-year data retention. Using a parabolic fit to the change in the slope of the threshold voltage during the retention measurement, 10-year retention is predicted taking into account internal field relaxation due to back-tunneling. In addition, retention measurements at elevated temperatures are performed to identify an acceleration factor to aid in accurately predicting 10-year data retention.

7.3 Short-Channel SNROM Transistor Fabrication

Short channel SNROM/NROM™ transistors are designed and fabricated at the Sherman Fairchild Center Microelectronics Laboratory. The gate dielectric of these transistors is composed of a 40 Å tunneling oxide, 50 Å silicon nitride, and a 40 Å

blocking oxide for operation with CHISEL, CHE, and HHI. Simulations are performed to provide custom processing conditions including a long-term source and drain drive-in anneal which yields sub-micron memory devices. Simulations also provide conditions for a channel implant to prevent the source and drain depletion regions from connecting during hot carrier injection.

7.4 Characterizing the Ideal NVSM for Space and Military Applications

Measurements and analysis of short-channel SNROM transistors programmed using channel initiated secondary electron (CHISEL) injection and hot hole injection (HHI) have been presented for use in space and military applications. Electrical characterization demonstrates the radiation hardness, programming characteristics, superior 10-year data retention, and endurance of these devices. Coupling these qualities with the ability to store two bits simultaneously in the gate dielectric makes short-channel SNROM transistors programmed with CHISEL and HHI an ideal candidate for space and military applications. Future characterization must address residual charge build-up in the nitride due to CHISEL/HHI injection region mismatch, trap location in the nitride (i.e. shallow/deep), lateral diffusion of charge at elevated temperatures, and radiation effects.

7.5 Future Research Recommendations

- Etch the gate dielectric of SNROM/NROM™ transistors with a reactive ion etch (RIE). The RIE is an anisotropic etch, meaning the walls of the etched space will be vertical. This etch will produce well defined sub-micron transistors.
- In order to produce a mask with sub-micron features, electron-beam (E-beam) lithography must be used to pattern the gate geometries in a photoresist film.
- In addition to promoting CHISEL injection through sub-micron fabrication techniques, SNROM devices may be designed with substrate injection in mind in order to improve injection efficiency. Buried channel implants may be used to provide additional carriers for secondary impact ionizations and a greater number of electrons for injection into the nitride.
- To further the understanding of the CHISEL mechanism, a measurement setup may be constructed which monitors the currents at all terminals of a transistor with respect to time when the device is programmed using fixed voltages. The programming time will allow for greater characterization of the threshold voltage shift as currents are produced and monitored in the device.

References

[1] D. Adams, J. T. Smith, J. Murray, M. H. White, and S. J. Wrazien, "Design, Fabrication and Characterization of High Density Radiation Hardened SONOS/CMOS EEPROMs for Space and Military Systems," NVMTS, Orlando, FL, November 2004.

[2] L. Cohn, "Non-Volatile Memory Technology for Space and Missile System Application," NVMTS, San Diego, CA, 2001.

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

[4] A. Niebel, Invited Speaker, Non-Volatile Semiconductor Memory Workshop, Monterey, CA, 2001.

[5] "The C.I.A. World Factbook," http://www.cia.gov/cia/publications/factbook/geos/kn.html#Military, April 2005.

[6] M. Bohr, "Intel's 65 nm Process Technology," http://www.intel.com/research/silicon, April 2005.

[7] L. Alkalai, "Advanced Avionics Systems for Future Space Exploration," Non-Volatile Memory Technology Symposium (NVMTS), San Diego, CA., 2001.

[8] M. H. White, "ECE 483 Lecture Notes," Lehigh University, 2001.

[9] L. Lancaster, "Why SONOS?," Non-Volatile Semconductor Memory Workshop (NVSMW), Monterey, CA, 2001.

[10] Y. Yang, "Retention Reliability of Scaled SONOS Devices," Doctoral Dissertation, Lehigh University, May 1999.

[11] C. Hu, *Nonvolatile Semiconductor Memories*. New York: IEEE Press, 1991.

[12] J. Choi, "Highly manufacturable 1 Gb NAND flash using 0.12 μm process technology," IEDM, Washington D.C., 2003.

[13] M. H. White, D. A. Adams, and J. Bu, "On the go with SONOS," *Circuits and Devices Magazine, IEEE*, vol. 16, pp. 22-31, 2000.

[14] S. Kayali, "Radiation Factors of Reliability on Non-Volatile Electronics,"

NVMTS, 2000.

[15] J. Bu and M. H. White, "Design Considerations in Scaled SONOS Nonvolatile Memory Devices," *Solid-State Electronics*, vol. 45, 2001.

[16] J. Lee, H.-S. Im, and D.-S. Byeon, "High-Performance 1-Gb NAND Flash Memory with 0.12 μm Technology," *Journal of Solid-State Circuits*, vol. 37, 2002.

[17] J. Bu and M. H. White, "Electrical characterization of ONO triple dielectric in SONOS nonvolatile memory devices," *Solid-State Electronics*, vol. 45, pp. 47-51, 2001.

[18] Y. Yang and M. H. White, "Charge Retention of Scaled SONOS Nonvolatile Memory Devices at Elevated Temperatures," *Solid-State Electronics*, vol. 44, 2000.

[19] S. J. Wrazien, J. Faul, M. H. White, D. A. Adams, and J. Murray, "Characterization of Scaled SONOS NVSM Devices for Space and Military Applications," NVMTS, San Diego, CA, November 2001.

[20] J. D. Krayer, Sherman Fairchild Summer Lecture Series, Lehigh University, 2002.

[21] J. Bu, "Polysilicon-Oxide-Nitride-Oxide-Silicon (SONOS) Nonvolatile Semiconductor Memory Devices Design, Fabrication, and Characterization," Doctoral Dissertation, Lehigh University, April 2001.

[22] B. Khan, Sherman Fairchild Summer Lecture Series, Lehigh University, 2003.

[23] B. Eitan, P. Pavan, I. Bloom, E. Aloni, A. Frommer, and D. Finzi, "NROM: A novel localized trapping, 2-bit nonvolatile memory cell," *IEEE Elect Dev Ltrs*, vol. 21, pp. 543-545, 2000.

[24] I. Bloom, P. Pavan, and B. Eitan, "NROM - a new non-volatile memory technology: from device to products," *Microelect Engr*, vol. 45, pp. 213-223, 2001.

[25] M. H. White, D. Adams, J. Murray, S. J. Wrazien, Y. S. Zhao, Y. R. Wang, B. Khan, W. Miller, and R. Mehrotra, "Characterization of Scaled SONOS EEPROM Memory Devices for Space and Military Systems," NVMTS, Orlando, FL, November 2004.

[26] C. T. Swift and G. L. Chindalore, "An embedded 90nm SONOS nonvolatile memory utilizing hot electron programming and uniform tunnel erase," IEDM,

San Francisco, CA, December 8-11, 2002.

[27] Y. Kamigaki and S. Minami, "MNOS Nonvolatile Semiconductor Memory Technology: Present and Future," *IEICE Trans Electron*, vol. E84-C, pp. 713-723, 2001.

[28] M. H. White, Y. Zhao, Y. Wang, S. J. Wrazien, J. D. Krayer, D. Adams, and J. Murray, "Scaled SONOS NVSM Devices for Space and Military Systems," NVMTS, San Diego, CA, November 2003.

[29] C. C. Yeh and W. J. Tsai, "PHINES: a novel low power program/erase, small pitch, 2-bit per cell flash memory," IEDM., San Francisco, CA, December 8-11, 2002.

[30] H. Ishiwara, "Recent progress in FET-type ferroelectric memories," IEDM, Washington D.C., 2003.

[31] S. Y. Lee and K. Kim, "Future 1T1C FRAM technologies for highly reliable, high density FRAM," IEDM, San Fransisco, CA, 2002.

[32] Hayashi, "A novel stack capacitor cell for high density FeRAM compatible with CMOS logic," IEDM, San Fransisco, CA, 2002.

[33] G. Derbenwick, D. A. Camp, S. C. Philipy, and A. F. Isaacson, "Advances in FeRAM Technologies," NVMTS, 2000.

[34] G. Derbenwick, A. D. DeVilbiss, F. Gnadinger, and G. Heubner, "Single Transistor Ferroelectric Memory Device," Non-Volatile Semiconductor Memory Workshop (NVSMW), Monterey, CA., 2001.

[35] A. Sheikholeslami and P. G. Gulak, "A survey of circuit innovations in Ferroelectric random-access memories," *Proceedings of the IEEE*, vol. 88, pp. 667-689, 2000.

[36] J. Zhu, "Vertical Magnetoresistive Random Access Memory," NVMTS, San Diego, CA, 2001.

[37] K.-M. Lenssen, "MRAM and its Prospects," NVMTS, San Diego, CA, 2001.

[38] D. Wang, "70% TMR at Room Temperature for SDT Sandwich.," 9th Joint MMM/Intermag Conference, 2004.

[39] J. S. Moodera, "Large Magnetoresistance at Room Temperature in Ferromagnetic Thin Film Tunnel Junctions," *Phys Rev Ltrs*, vol. 74, 1995.

[40] M. Durlam, "A 0.18 μm 4Mb toggling MRAM," IEDM, Washington D.C.,

2003.

[41] S. Lai, "Current status of the phase change memory and its future," IEDM, Washington D.C., 2003.

[42] J. D. Maimon and E. Spall, "Chalcogenide-Based Non-Volatile Memory Technology," NVMTS, 2000.

[43] J. H. Park, "An 8F2 MRAM technology using modified metal lines," IEDM, Washington D.C., 2003.

[44] Y. Wang, Sherman Fairchild Summer Lecture Series, Lehigh University, 2003.

[45] J. D. Maimon, K. K. Hunt, L. Burcin, and J. Rodgers, "Chalcogenide memory arrays: characterization and radiation effects," *Nuclear Science, IEEE Transactions on*, vol. 50, pp. 1878-1884, 2003.

[46] A. Pirovano, "Scaling analysis of phase-change memory technology," IEDM, Washington D.C., 2003.

[47] D. Esseni, L. Selmi, A. Ghetti, and E. Sangiorgi, "Injection Efficiency of CHISEL Gate Currents in Short MOS Devices: Physical Mechanisms, Device Implications, and Sensitivity to Technological Parameters," *IEEE Trans on Elect Dev*, vol. 47, 2000.

[48] S. Mahapatra, S. Shukuri, and J. Bude, "CHISEL Flash EEPROM -- Part I: Performance and Scaling," *IEEE Trans on Elect Dev*, vol. 49, 2002.

[49] L. Larcher and P. Pavan, "A new Analytical Model of Channel Hot Electron (CHE) and Channel Initiated Secondary Electron (CHISEL) Current Suitable for Compact Modeling," Workshop on Compact Modeling, 2002.

[50] M. Mastrapasqua, "Low voltage flash memory by use of a substrate bias," *Microelect Engr*, vol. 48, pp. 389-394, 1999.

[51] D. R. Nair, N. R. Mohapatra, S. Mahapatra, S. Shukuri, and J. Bude, "Effect of P/E Cycling on Drain Disturb in Flash EEPROMs Under CHE and CHISEL Operation," *IEEE Trans on Dev and Mat Rel*, vol. 4, 2004.

[52] S. Chung, "Low Voltage/Power and High Speed Flash Memory Technology for High Performance and Reliability," The 3rd WIMNACT, Singapore, 2003.

[53] E. Lusky, Y. Shacham-Diamand, I. Bloom, and B. Eitan, "Characterization of channel hot electron injection by the subthreshold slope of NROM™ device," *IEEE Elect Dev Ltrs*, vol. 22, pp. 556-558, 2001.

[54] D. Xu and V. J. Kapoor, "Effects of oxygen content and oxide layer thickness on interface state densities for metal-oxynitride-oxide-silicon devices," *Journal of Applied Physics*, vol. 70, pp. 1570-1574, 1991.

[55] V. J. Kapoor, R. S. Bailey, and R. A. Turi, "Chemical Composition, Charge Trapping, and Memory Properties of Oxynitride Films for MNOS Devices," *Journ. of the Electrochem. Soc.*, vol. 137, pp. 3589, 1990.

[56] V. J. Kapoor, R. A. Turi, D. Xu, and R. S. Bailey, "MNOS Memory Technology with Oxynitride Thin Films," *IEEE Trans. on Components, Packaging, and Manufacturing Tech. - Part A*, vol. 17, 1994.

[57] H. A. R. Wegener, A. J. Lincoln, H. C. Pao, M. R. O'Connel, and R. E.Oleksiok, "Metal Insulator Semiconductor Transistors as a Nonvolatile Storage," International Electron Devices Meeting (IEDM), Washington, DC, 1967.

[58] L. Lundkvist, I. Lundstrom, and C. Svensson, "Discharge of MNOS structures," *Solid-State Electronics*, vol. 16, pp. 811-818, 1973.

[59] L. Lundkvist, C. Svensson, and B. Hansson, "Discharge of MNOS structures at elevated temperatures," *Solid-State Electronics*, vol. 19, pp. 221-227, 1976.

[60] K. Lehovec and A. Fedotowsky, "Charge retention of MNOS devices limited by Frenkel-Poole detrapping," *Applied Physics Letters*, vol. 32, pp. 335-338, 1978.

[61] C. Svensson and I. Lundstrom, "Trap-assisted charge injection in MNOS structures," *Journal of Applied Physics*, vol. 44, pp. 4657-4663, 1973.

[62] M. L. French and M. H. White, "Scaling of multidielectric nonvolatile SONOS memory structures," *Solid-State Electronics*, vol. 37, pp. 1913-1923, 1994.

[63] Y. Wang and M. H. White, "An analytical retention model for SONOS nonvolatile memory devices in the excess electron state," *Solid-State Electronics*, vol. 49, pp. 97-107, 2005.

[64] M. H. White and J. R. Cricchi, "Characterization of Thin-Oxide MNOS Memory Transistors," *IEEE Trans on Elect Dev*, vol. ED19, pp. 1280, 1972.

[65] A. Roy and M. H. White, "Determination of the trapped charge distribution in scaled silicon nitride MONOS nonvolatile memory devices by tunneling spectroscopy," *Solid-State Electronics*, vol. 34, pp. 1083-1089, 1991.

[66] Y. Kamigaki, S.-i. Minami, and H. Kato, "A new portrayal of electron and hole traps in amorphous silicon nitride," *Journal of Applied Physics*, vol. 68, pp.

2211-2215, 1990.

[67] Y. Kamigaki, S. Minami, T. Hagiwara, K. Furusawa, T. Furuno, K. Uchida, M. Terasawa, and K. Yamazaki, "Yield and reliability of MNOS EEPROM products," *IEEE J. Solid-State Circuits*, vol. 24, pp. 1714, 1989.

[68] Y. Hu and M. H. White, "Charge Retention in Scaled SONOS Nonvolatile Semiconductor Memory Devices - Modeling and Characterization," *Solid-State Electronics*, vol. 36, pp. 1401, 1993.

[69] S. J. Wrazien, Y. Zhao, J. D. Krayer, and M. H. White, "Characterization of SONOS oxynitride nonvolatile semiconductor memory devices," *Solid-State Electronics*, vol. 47, pp. 885-891, 2003.

[70] S. L. Miller, "Effect of Temperature on Data Retention of SNOS Transistors," 8th IEEE NVSM Workshop, Vail, CO, August 1986.

[71] P. J. McWhorter, S. L. Miller, and T. A. Dellin, "Modeling the memory retention characteristics of silicon-nitride-oxide-silicon nonvolatile transistors in a varying thermal environment," *Journal of Applied Physics*, vol. 68, pp. 1902-1909, 1990.

[72] A. Roy, "Characterization and Modeling of Charge Trapping and Retention in Novel Multi-Dielectric Nonvolatile Semiconductor Memory Devices," Doctoral Dissertation, Lehigh University, May 1989.

[73] A. B. Joshi, G. Q. Lo, and D. K. Kwong, "Improved Performance and Reliability of MOSFETs with Thin Gate Oxides Grown at High Temperature," IRPS, 1991.

[74] Y. Yang, A. Purwar, and M. H. White, "Reliability Considerations in Scaled SONOS Nonvolatile Memory Devices," *Solid-State Electronics*, vol. 43, pp. 2025-2032, 1999.

[75] S. Minami and Y. Kamigaki, "A Novel MONOS Nonvolatile Memory Device Ensuring 10-Year Data Retention After 10^7 Write/Erase Cycles," *IEEE Trans. Elect. Dev.*, vol. 40, 1993.

[76] S. Fujita, H. Toyoshima, M. Nishihara, and A. Sasaki, "Variation of Trap States and Dangling Bonds in CVD Si_3N_4 Layer on Si Substrate by NH_3/SiH_4 Ratio," *Journ. of Electron. Mater.*, vol. 11, pp. 795, 1982.

[77] Y. Zeng, "A Study of Electron Transport in the Inversion Layer of Advanced Silicon Carbide (SiC) Power MOSFETs," Doctoral Dissertation, Lehigh

University, 2004.

[78] G. Groeseneken, H. E. Maes, N. Beltran, and R. F. D. Keersmaecker, "A Reliable Approach to Charge Pumping Measurements in MOS Transistors," *IEEE Trans. Electron Devices*, vol. ED-31, 1984.

[79] R. E. Paulsen and M. H. White, "Theory and Application of Charge Pumping for the Characterization of Si-SiO2 Interface and Near-Interface Oxide Traps," *IEEE Trans. Electron Devices*, vol. 41, 1994.

[80] N. E. Cohen, R. E. Paulsen, and M. H. White, "Observation and Characterization of Near-Interface Oxide Traps with C-V Techniques," *IEEE Trans. Electron Devices*, vol. 42, pp. 2004-2009, 1995.

[81] M. H. White and A. Roy, "A new approach to study electron and hole charge separation at the semiconductor-insulator interface," *IEEE Trans on Elect Dev*, vol. 37, 1990.

[82] B. Khan and R. Mehrotra, "Electrical Characterization of SONOS NVSM Devices," Sherman Fairchild Summer Lecture Series, Lehigh University, 2004.

[83] S. Yoon and M. H. White, "Study of Thin Gate Oxide Grown in an Ultra-Dry/Clean Triple-Wall Oxidation Furnace System," *J Electro Mater*, vol. 19, pp. 487-493, 1990.

[84] S. J. Wrazien, Y. Wang, B. M. Khan, and M. H. White, "Characterizing damage to ONO dielectrics induced during programming SONOS/NROMTM non-volatile semiconductor memory (NVSM) devices," *Solid-State Electronics*, vol. 48, pp. 2035-2039, 2004.

[85] S. Mahapatra, S. Shukuri, and J. Bude, "CHISEL Flash EEPROM -- Part II: Reliability," *IEEE Trans on Elect Dev*, vol. 49, 2002.

[86] M. E. Castagna, M. M. S. Coffa, A. M. L. Caristia, R. Mangano, and C. Bongiorno, "Si-based Materials and Devices for Light Emission in Silicon," *Physica E: Low-Dimensional Systems and Nanostructures*, vol. In Press, Corrected Proof Available Online, 2002.

[87] M. E. Castagna, S. Coffa, L. Caristia, and A. Messina, "Quantum dot materials and devices for light emission in silicon," ESSDER, 2002.

[88] M. Kanellos, "ST Micro sees silicon light." CNETNews.com: http://news.com.com/2100-1001-963620.html, 2002.

[89] J. C. Pivin, M. J. d. Castro, H. Hofmeister, and M. Sendova-Vassileva, "Exiton-

erbium coupling in SiOx suboxide films prepared by combining sol-gel chemistry and ion implantation," *Materials Science and Engineering*, vol. B97, pp. 13-19, 2003.

[90] H. S. Han, S. Y. Seo, and J. H. Shin, "Optical gain at 1.54 μm in erbium-doped silicon nanocluster sensitized waveguide," *Appl. Phys. Lett.*, vol. 79, 2001.

Appendix A

SONOS Device Fabrication Sequence

A. 1 Starting Material

1. 3-inch p-type Si Wafers, boron-doped 13-15 Ω-cm

A. 2 N-Well Formation

1. RCA clean with HF dip
2. Oxidation, 1000 Å (dry, 1000 C, 160 min)
3. Photolithography, mask NW (N-Well)
4. BHF etch to obtain 200 Å pad oxide (14 min)
5. N-well implantation, phosphorus, 100 KeV $4.8x10^{12}cm^{-2}$
6. Plasma photo resistor (PR) strip
7. Chemical PR strip
8. RCA clean
9. Anneal, 1100 C, 60 min, N_2, 2 slpm
10. Oxidation, dry, 1100 C, 15 min
11. Anneal (N-well drive in), 1100 C for 24 hours, N_2, 2 slpm

A. 3 LOCOS Isolation

1. RCA Clean
2. LPCVD nitride 1000 Å (725 C, 0.3 Torr, 20 sccm dichlorosilane (DCS), 100 sccm ammonia, 65 min)
3. Photolithography, mask AD (Active Device)
4. Plasma etch nitride (0.3 Torr, 300 Watts, 60 sccm CF_4)
5. Chemical PR strip
6. Photolithography, mask FI (Field Implant)
7. Field implantation, boron, 40 KeV, $5x10^{14}cm^{-2}$
8. Plasma PR strip
9. Chemical PR strip
10. RCA clean
11. Field oxidation/implant activation (wet, 1000 , 180 min)
12. Field oxidation, 8000 Å (wet, 1000 C, 140 min, anneal 30 min in dry N_2)
13. Etch oxynitride over the nitride (10:1 BHF, 1 min)
14. Etch nitride, 1000 Å (hot phosphoric acid, 180 C, 60 min)
15. Etch buffer oxide, 1000 Å (10:1 BHF, 100 sec)
16. RCA clean
17. Oxidation – remove residual nitride (wet, 900 C, 20 min)
18. Etch oxide (10:1 BHF, 100 sec)
19. RCA clean (no HF)

20. Oxidation, 250 Å (wet, 900 C, 15 min)

A. 4 Channel Doping

1. Threshold adjust implant, boron 50 KeV, $1x10^{12}cm^{-2}$
2. RCA clean
3. Implant anneal (900 C, N_2, 30 min)

A. 5 Gate Formation and Polysilicon Deposition

1. Etch pad oxide (100:1 BHF, 12 min)
2. RCA clean (after HF dip to hydrophobic, do not rinse)
3. Tunnel oxide growth, 20 Å (TWO, 800 C, 30 sccm O_2 in 3000 sccm Ar, 40 min)
4. LPCVD silicon-rich nitride, 45 Å (680 C, 90 sccm DCS, 30 sccm NH_3, 7 min)
5. LPCVD nitride, 45 Å (680 C, 10 sccm DCS, 100 sccm NH_3, 13 min)
6. Optional – LPCVD oxynitride, 45 Å (680 C, 30:40:10 N_2O:NH_3:DCS, 4 Å /min)
7. LPCVD oxide, 55 Å (725 C, 10 sccm DCS, 100 sccm N_2O, 34 min)
8. Oxide steam densification (wet, 900 C, 30 min)
9. Polysilicon gate deposition, 5 K Å (625 C, 800 mTorr, 284 sccm Silane, 30 min)
10. RCA clean
11. $POCl_3$ diffusion (900 C, N_2 and O_2 flow with $POCl_3$ bubbler at 19 C, 25 min, then N_2 only for 25 min)
12. Etch P-glass (100:1 BHF, 4 min)

13. Photolithography, mask PY (Polysilicon)
14. Plasma etch polysilicon (300 Watts, 150 mTorr, 45 sccm SF_6, 60 sec)
15. Chemical PR strip

A. 6 Source and Drain Implants

1. Etch blocking oxide (100:1 BHF, 3 min)
2. Etch nitride (H_3PO_4,180 C, 3 min)
3. Etch tunnel oxide (100:1 BHF, 3 min)
4. RCA clean (with HF dip)
5. Pad oxidation, 200 Å (wet, 800 C, 28 min)
6. Photolithography, mask PP (Poly-diffusion)
7. P-Source/Drain implant, boron, 32 KeV, $5x10^{15}cm^{-2}$
8. Plasma PR strip
9. Chemical PR strip
10. Photolithography, mask NN (N-diffusion)
11. N-Source/Drain implant, phosphorus, 40 KeV, $5x10^{15}cm^{-2}$
12. Plasma PR strip
13. Chemical PR strip
14. RCA clean
15. Implant drive-in anneal (1000 C in 2 slpm N_2 for 160 min)
16. Boron/Phosphorus-silicate glass strip (10:1 BHF, 1 min)

A. 7 Contact Window

1. RCA clean (with HF)
2. Intermediate oxide, 1100 Å (wet, 950 C, 30 min, anneal 30 min)
3. Photolithography, mask CW (Contact Window)
4. Etch windows (10:1 BHF), rinse thoroughly
5. Chemical photo strip

A. 8 Pre-metal High Temperature deuterium Anneal

1. RCA clean (with HF dip to hydrophobic)
2. Anneal (700 C, 10% D_2/N_2, 2 slpm, 4 hours, PMA furnace)

A. 9 Metalization

1. DC sputter (2 hours, cool down 10 min)
2. Photolithography, mask MET (Metalization)
3. Etch aluminum (PAN etch at 43 C)
4. Chemical photo-strip

A. 10 Post-Metal Anneal (PMA)

1. Organic clean (Acetone & Methanol)
2. PMA (400 C, 10% D_2/N_2, 30 min)

Appendix B

Measurement Procedures

B. 1 Current Versus Voltage Measurement

Goal – To plot the drain current versus applied gate voltage on a single SONOS device and to become familiar with the measurement setup using the HP- 4145, multi-meter, and LabVIEW™ interface. The I-V curve is useful in determining the threshold voltage of the device in the fresh, programmed, and erased states.

1) Setup

A) Verify that the computer is connected to the 4145 scope (bottom right) and the Keithley 7001 Switch System (top right) using the GPIB cables.

B) Place the wafer onto the probing station.

C) Locate a single SONOS device.

D) Make contact with the device at pads indicated in the chart.

Table B.1 – Current versus Voltage Measurement Pin Setup

	Source	Substrate	Gate	Drain
Pad #	9	32	33	8
Pad #	7	34	35	6
Probe #				

E) Close the lid on the probing station and watch so the lid does not close on the wafer.

F) Connect the pins on the side of the lid to the wires from the Keithley meter in this order from top to bottom: 1, 2, 4, 5.

G) On the PC, load the program "matrix.vi" in the directory \\Calvin\pcdir\labview \programs\subvis. Run the program by clicking on the white arrow in the top left of the screen.

H) On the PC, load the I-V measurement program "isvgfam" in the directory \\Calvin\pcdir\labview\programs\iv. The program will take a few seconds to load.

2) Measurement

A) Choose a title for your data measurement.

B) Set the bulk voltage to zero volts.

C) Set the Gate Voltage Range from –3.0 to 1.0 volts.

D) Set the data points to 51 for a higher resolution curve.

E) Set the Drain Voltage to 50 mV.

F) Set the Bulk Steps to 1. Setting this to a value greater than one will repeat the measurement that many times.

G) Set the Channel Definitions. (refer to chart above)

H) Run the program by clicking on the white arrow in the top left of the screen. Do

not bump the table during the measurement.

I) Save your data to a file in a directory you set up for yourself.

J) Print the LabVIEW™ screen (scale window to 85% in printer settings).

K) Write the device

1) Change the Gate Voltage Range to +7V to +8V.

2) Run the program for roughly 10 seconds then click on the stop icon next to the arrow.

L) Change the title of your measurement (part A) to indicate the device has been programmed.

M) Repeat steps 2B through 2J.

N) Erase the device

1) Change the Gate Voltage Range to –7V to –8V.

2) Run the program for roughly 10 seconds then click on the stop icon next to the arrow.

O) Change the title of your measurement (part A) to indicate the device has been erased.

P) Repeat steps 2B through 2J.

Q) Open Microsoft Excel and import your data and plot it as best you can.

B. 2 Charge Pumping Measurement

Goal – To sweep the gate of the device with a square wave with specific amplitude in

order to excite electrons out of traps in the Si-SiO2 interface. This will result in a current, which will be measured. This current can be related to the charge stored in the device by dividing the current by the frequency. LabVIEW™ will extract the interface trap density, Dit, from the charge stored in the traps.

1) Setup

A) Verify that the computer is connected through the GPIB cables with the HP 33120A function generator and the Keithley 617 programmable electrometer.

B) Follow steps B – E from the I-V measurement.

C) Connect the HP 33120A to the gate of the device.

D) Connect the source and drain terminals together using the "T" connector and to Keithley 617 electrometer.

E) Place the short circuit connector onto the substrate pin to ground this terminal.

F) Load the program "sonosBASE_swp.vi" in the directory "C:\steve\SONOS LabVIEW".

2) Measurement

A) Set the amplitude to 2 volts (default).

B) Set the frequency to 1.00E+5 Hz (default).

C) Leave the rise and fall times at 1E-7 seconds (default).

D) Set samples per point to 3 or 5 if you want a more accurate measurement.

E) Set the type of device, NMOS (default) for the Northrop Grumman lot # 63850D

wafers.

F) Set termination to High Z (default).

G) Set pulse generator to HP-33120A (default).

H) Set the base level range to values appropriate for the device. Start with a range from –3 to +1 volts in the forward direction. You will have to change these values in order to view the entire charge pumping curve.

I) Enter a name including the device lot #, wafer #, and any other information that characterizes the device.

J) Set the width and length of the device. For NGC lot # 63850D, W =10 um and L = 0.8 um.

NOTE: When measuring an array of devices, width becomes # of devices x width of 1 device, in order to account for the large number of devices in the array, and correctly extract I_{CP} and D_{it}.

K) Run the program by clicking on the white arrow in the top left of the screen. Do not bump the table during the measurement.

B. 3 Retention Measurement

1. ***Goal*** – To determine the ability of a SONOS device to retain charge over a long period of time, to investigate the retention characteristics at elevated temperature, and to determine the effect of stressing the device on the retention characteristics.

1) Setup

A) Turn on all 4-power supplies.

B) Download FPGA through the hardware debug cable.

a. Northrp1- to erase device in 7.5 ms

b. Northrop- to write device in 2.5 ms

c. In Xilinx Foundation software, choose Northrop.bit or Northrp1.bit → Download → Download design → OK

C) Verify that the computer is connected through the GPIB cables with the Tek TDS 460 oscilloscope.

D) Connect the cables from the box to the exact pad.

E) Check the connection of V_w and V_e.

a. When writing the device, V_w should be positive and V_e should be negative.

b. When erasing the device, V_w should be negative and V_e should be positive.

F) Load the program "FPGAretention_readonly.vi" in the directory "C:\steve\SONOS LabVIEW".

G) Reset SW-1 and press the middle square button to start.

2) Measurement

A) Set up windows

a. read window 1: 2.49E-4 read window 2: 3.47E-4

b. ground window 1: -3.18E-4 ground window 2: -2.41E-4

c. "read window" goes on the left side of the transition on the oscilloscope,

which is Vth. "ground window" is on the right side on the oscilloscope, which establishes ground for Vth.

B) Measurements start from 1E-7, 1E-6... to 1E+5 s.

B. 4 Write/Erase/ Cycling Measurement

2. ***Goal*** – To simulate prolonged use of the device, and determine how the device will hold up under periods of extreme use.

1) Setup

A) Connect the HP-33120A to the gate of the device.

B) Short the source, drain and bulk to ground.

C) Verify that the computer is connected through the GPIB cable with HP-33120A.

D) Load the program "stress1.vi" in the directory "C:\steve\SONOS LabVIEW".

2) Measurement

A) Set Input Stress Cycles from 1E4 to 1E6.

B) Set Write Width as 2.5 ms and Erase Width as 7.5 ms.

C) Set Write Amplitude and Erase Amplitude as 7 V.

Appendix C

Fabrication Simulations

'SONOS 1/01' Fabrication Run
Varied channel lengths for conventional and 2-bit operation

20 wafers

*** No Anneal for Vth or Punch-through Implants – Done with S/D Anneal

Vth Implant – Vth PMOS = 0.35 V, NMOS = -0.5 V
10 – $1x10^{12}$ ions/cm^2 (Boron) – Wafers # 0 - 9
5 – $1.3x10^{12}$ ions/cm^2 (Boron) – Wafers # A - E
5 – $1.7x10^{12}$ ions/cm^2 (Boron) – Wafers # F - J

Channel Punch-through Protection Implant
NMOS - Use Mask "NN" (TP-400) - 200 Å Pad Oxide
Implant: $1x10^{12}$ ions/cm^2 (Boron), 180 KeV *** Specify: LOW BEAM CURRENT

PMOS - Use Mask "PP" (TP-400) - 200 Å Pad Oxide
Implant: $2x10^{12}$ ions/cm^2 (Phosphorus), 180 KeV *** Specify: LOW BEAM CURRENT

Wafers # 0-4 Are test wafers for Punch-through Implant

Table C.1: Implant Drive-In Anneal

Wafer #	Type	Vth Implant	Punch Implant	Anneal Time
0	*NMOS*	$1.0x10^{12}$	$1x10^{12}$	90
ω	***NMOS***	***$1.3x10^{12}$***	**$1x10^{12}$**	**120**
F	***NMOS***	***$1.7x10^{12}$***	**$1x10^{12}$**	**120**
1	*NMOS*	$1.0x10^{12}$	$1x10^{12}$	160
G	*NMOS*	***$1.7x10^{12}$***	$1x10^{12}$	240
2	*NMOS*	$1.0x10^{12}$	$1x10^{12}$	270
H	***NMOS***	***$1.7x10^{12}$***	**$1x10^{12}$**	**300**
3	***NMOS***	$1.0x10^{12}$	**$1x10^{12}$**	**300**
B	*NMOS*	*$1.3x10^{12}$*	$1x10^{12}$	330
4	*NMOS*	$1.0x10^{12}$	$1x10^{12}$	330
5	*PMOS*	$1.0x10^{12}$	$2x10^{12}$	90
C	***PMOS***	***$1.3x10^{12}$***	**$2x10^{12}$**	**120**
I	***PMOS***	***$1.7x10^{12}$***	**$2x10^{12}$**	**120**
6	*PMOS*	$1.0x10^{12}$	$2x10^{12}$	160
J	*PMOS*	*$1.7x10^{12}$*	$2x10^{12}$	240
7	*PMOS*	$1.0x10^{12}$	$2x10^{12}$	270
D	***PMOS***	***$1.3x10^{12}$***	**$2x10^{12}$**	**300**
8	***PMOS***	**$1.0x10^{12}$**	**$2x10^{12}$**	**300**
E	*PMOS*	*$1.3x10^{12}$*	$2x10^{12}$	330
9	*PMOS*	$1.0x10^{12}$	$2x10^{12}$	330

Silvaco Source/Drain Anneal Simulations

NMOS Simulation

```
go athena
# NMOS
# Simulate Source/Drain implants and anneal for short-channel SONOS transistor

# set up lateral 1/2 device
line x loc=-2.6 spac=0.1
line x loc=2.0 spac=0.1

# set up vertical device
line y loc=0.0 spac=0.1
line y loc=3.0 spac=0.1

# 1.0 set up bulk conditions
init silicon boron resistivity=23 orientation=100 space.mult=2.0

# 3.0.29 Oxidation
diffus time=15 temp=900 weto2 press=1.00 hcl.pc=0

# 4.1 Threshold voltage adjust implant (blanket implant)
implant boron dose=1.0e12 energy=50 tilt=0 rotation=0 crystal lat.ratio1=1.0 \
    lat.ratio2=1.0
# Channel punchthrough protection implant
implant boron dose=1.0e12 energy=180 tilt=0 rotation=0 crystal lat.ratio1=1.0 \
    lat.ratio2=1.0

# 5.0 ONO Formation
# 5.0.45 Etch pad oxide (BHF)
etch oxide all

# 5.0.46.3 Tunnel oxide (TWO)
deposit oxide thick=0.002

# 5.0.46.5 LPCVD Silicon-rich-nitride
deposit nitride thick=0.0045

# 5.0.46.7 LPCVD oxide
deposit oxide thick=0.0055

# 5.0.56.8 Oxide steam densification
diffus time=30 temp=900 weto2 press=1.00 hcl.pc=0

# 5.0.46.9 Polysilicon gate deposition
deposit poly thick=0.5

# 5.0.59 POCL3 Diffusion
diffus time=40 temp=900 nitro press=1.00 c.phosphor=1.0e19

# 5.0.46.14 Plasma etch poly-Si/define ONO
etch poly left p1.x=0.4
```

```
# Etch blocking oxide
etch oxide left p1.x=0.4
# Etch Nitride
etch nitride left p1.x=0.4
# Etch tunnel oxide
etch oxide left p1.x=0.4

# 6.0.64 Pad oxidation - 200A
# DEPOSIT TO PROTECT QUALITY OF ONO
deposit oxide thick=0.02

# 6.0.70 N-Source/Drain implant
implant phosphor dose=5e15 energy=40 tilt=0 rotation=0 crystal \
      lat.ratio1=1.0 lat.ratio2=1.0

# Follow flow appendix A - J. Bu dissertation
# 6.0.A6.15 Implant Drive-In anneal
# ************************************************************************
diffus time=360 temp=1000 nitro press=1.0

# 6.0.A7.2 Intermediate Oxide/Anneal
deposit oxide thick=0.5
# deposit pad and intermediate oxides instead of thermally grow??? *****
#diffus time=30 temp=950 weto2 press=1.00 hcl.pc=0
diffus time=30 temp=950 nitro press=1.00

# 7.0.A7.4 Etch windows
etch oxide right p1.x=1.25 p1.y=-0.15 p1.x=1.25 p2.y=-1.0
etch oxide left p1.x=0.1

# 7.0.A8.2 Pre-metal high temperature deuterium anneal (used N2 anneal)
diffus time=160 temp=700 nitro press=1.00

# 8.0.82 Metallization
deposit alumin thick=0.50

# 8.0.84 Etch Aluminum
etch aluminum start x=0.1 y=0.00
etch cont x=0.1 y=-2.0
etch cont x=1.25 y=-2.0
etch done x=1.25 y=0.0

# Mirror
struct mirror right

structure outfile=NMOS_SDdiff.str
# plot structure
tonyplot NMOS_SDdiff.str
```

File for PMOS Simulation

```
go athena
# PMOS
# Simulate Source/Drain implants and anneal for short-channel SONOS transistor

# set up lateral 1/2 device
line x loc=-2.6 spac=0.1
line x loc=2.0 spac=0.1

# set up vertical device
line y loc=0.0 spac=0.1
line y loc=3.0 spac=0.1

# 1.0 set up bulk conditions
init silicon boron resistivity=23 orientation=100 space.mult=2.0

# 2.0 N-Well Formation
# 2.2 1000A dry oxidation - pad oxide
diffus time=145 temp=1000 dryo2 press=1.00 hcl.pc=0

# 2.3 (photo) - pattern pad oxide for N-Well implant
# 2.4 300A pad oxide desired
etch oxide dry thick=0.07

# 2.5 N-Well implant
implant phosphor dose=4.8e12 energy=100 tilt=0 rotation=0 amorph \
      lat.ratio1=1.0 lat.ratio2=1.0

# 2.9 Implant Anneal
diffus time=60 temp=1100 nitro press=1.00
# Oxidation
diffus time=15 temp=1100 dryo2 press=1.00 hcl.pc=0
#  N-Well drive-in/Nitrogen Anneal
diffus time=1065 temp=1100 nitro press=1.00

# 3.0.22-24 Remove oxides
etch oxide all

# 3.0.29 Oxidation
diffus time=15 temp=900 weto2 press=1.00 hcl.pc=0

# 4.1 Threshold voltage adjust implant
implant boron dose=1.0e12 energy=50 tilt=0 rotation=0 crystal lat.ratio1=1.0 \
      lat.ratio2=1.0
```

```
# Channel punchthrough protection implant
implant phosphor dose=2.0e12 energy=180 tilt=0 rotation=0 crystal \
      lat.ratio1=1.0 lat.ratio2=1.0

# 4.3 vth adjust anneal
#diffus time=30 temp=900 nitro press=1.00

# 5.0 ONO Formation
# 5.0.45 Etch pad oxide (BHF)
etch oxide all

# 5.0.46.3 Tunnel oxide (TWO)
deposit oxide thick=0.002

# 5.0.46.5 LPCVD Silicon-rich-nitride
deposit nitride thick=0.0045

# 5.0.46.7 LPCVD oxide
deposit oxide thick=0.0055

# 5.0.56.8 Oxide steam densification
diffus time=30 temp=900 weto2 press=1.00 hcl.pc=0

# 5.0.46.9 Polysilicon gate deposition
deposit poly thick=0.5

# 5.0.59 POCL3 Diffusion
diffus time=40 temp=900 nitro press=1.00 c.phosphor=1.0e19

# 5.0.46.14 Plasma etch poly-Si/define ONO
etch poly left p1.x=0.4
# Etch blocking oxide
etch oxide left p1.x=0.4
# Etch Nitride
etch nitride left p1.x=0.4
# Etch tunnel oxide
etch oxide left p1.x=0.4

# 6.0.64 Pad oxidation - 200A
# DEPOSIT TO PROTECT QUALITY OF ONO
deposit oxide thick=0.02

# 6.0.70 P-Source/Drain implant
implant boron dose=5.0e15 energy=32 tilt=0 rotation=0 crystal \
```

```
    lat.ratio1=1.0 lat.ratio2=1.0

# Follow flow appendix A - J. Bu dissertation
# 6.0.A6.15 Implant Drive-In anneal
#
**********************************************************************
***
diffus time=360 temp=1000 nitro press=1.00

# 6.0.A7.2 Intermediate Oxide/Anneal
deposit oxide thick=0.5
# deposit pad and intermediate oxides instead of thermally grow??? *****
#diffus time=30 temp=950 weto2 press=1.00 hcl.pc=0
diffus time=30 temp=950 nitro press=1.00

# 7.0.A7.4 Etch windows
etch oxide right p1.x=1.25 p1.y=-0.15 p1.x=1.25 p2.y=-1.0
etch oxide left p1.x=0.1

# 7.0.A8.2 Pre-metal high temperature deuterium anneal (used N2 anneal)
diffus time=160 temp=700 nitro press=1.00

# 8.0.82 Metallization
deposit alumin thick=0.50

# 8.0.84 Etch Aluminum
etch aluminum start x=0.1 y=0.00
etch cont x=0.1 y=-2.0
etch cont x=1.25 y=-2.0
etch done x=1.25 y=0.0

# Mirror
struct mirror right

structure outfile=PMOS_SDdiff.str
# plot structure
tonyplot PMOS_SDdiff.str
```

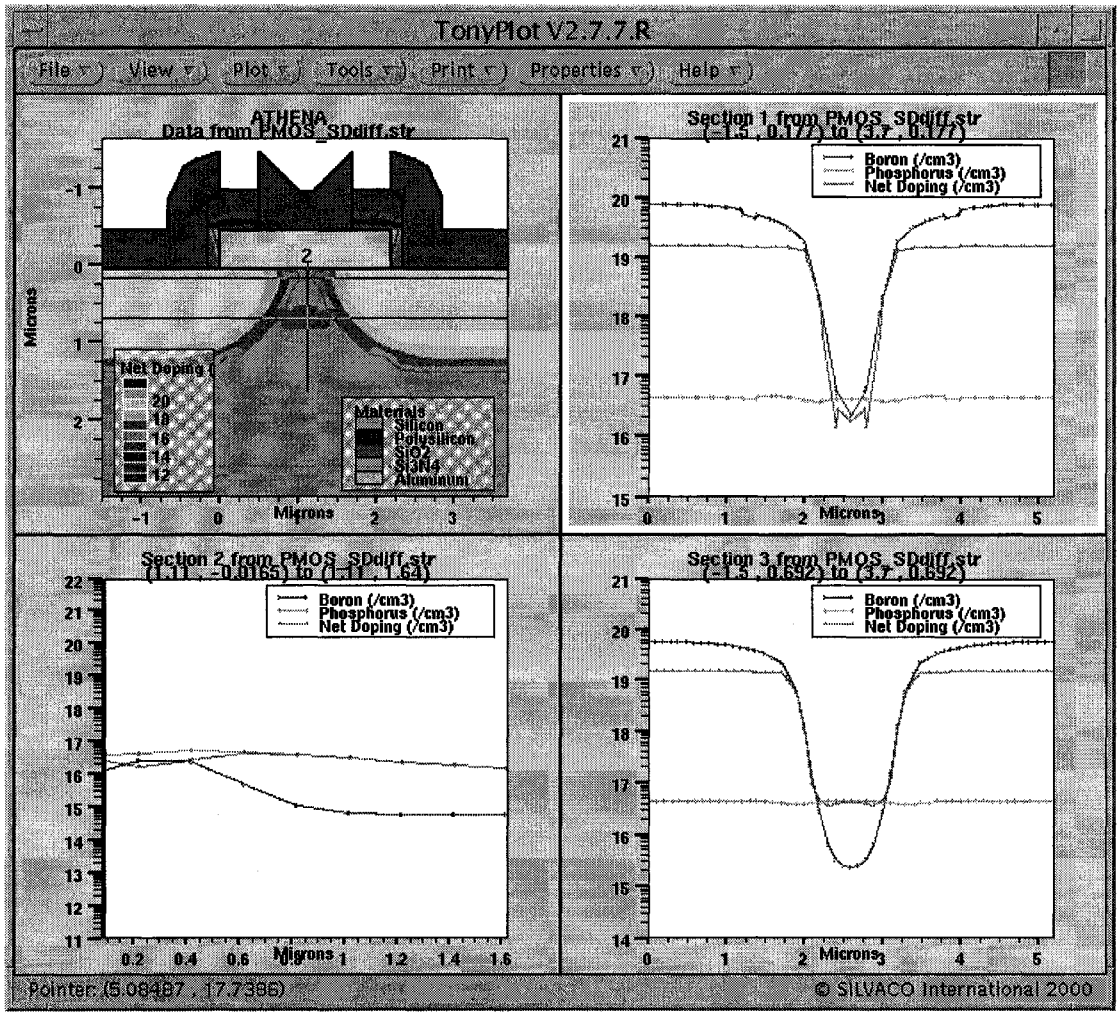


Figure C.1: PMOS SONOS Transistor after 120 minute anneal at 1000 C (conditions for SONOS 2000). Original channel length was 2.2 μm, post-anneal channel length is approximately 0.25 μm.

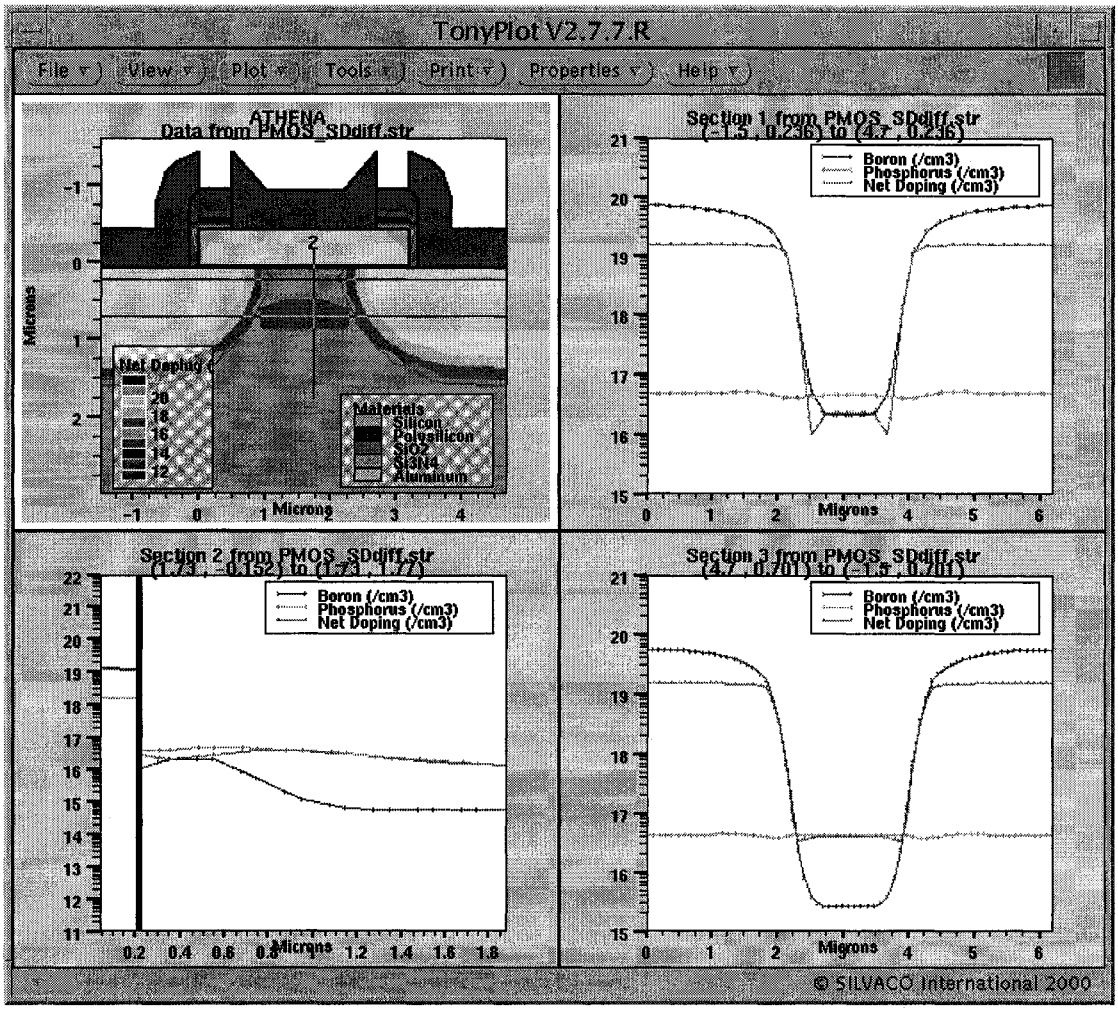


Figure C.2: PMOS SONOS Transistor after 160 minute anneal at 1000 C (conditions for SONOS 2000). Original channel length was 3.2 μm, post-anneal channel length is approximately 1.75 μm.

Appendix D

Erbium-Doped CMOS Capacitors (Proposal)

CMOS Capacitors with Er-Doped Silicon Rich Oxide Gate Dielectrics for Ultra-Efficient Si-based Light Emission

Purpose

We will study the light emission properties CMOS capacitors with Erbium (Er) doped gate oxides. CMOS capacitors will be fabricated at the Sherman Fairchild Laboratory for Solid-State Studies. CMOS devices of all sizes with high-quality films are routinely fabricated at the Sherman Fairchild Laboratory. Fabrication will include computer simulations for tuning the Erbium implant conditions. The capacitors will be sent to an outside company, which has the facilities required to properly perform the Erbium implant. Fabrication will conclude at Sherman Fairchild Laboratory and the devices will be ready for test and evaluation. Depending on observations from preliminary experiments, custom mask sets, circuits, simulations, and layer depositions can be designed in order to produce further advances in the field of Si-based light emission. Previous research using capacitors with Erbium-implanted gate oxides has produced favorable results. ST Microelectronics reported 1.54 μm electroluminescent devices at 300 K with a 10% external quantum efficiency, comparable to standard light emitting diodes using III-V semiconductors.

Introduction

Silicon is a semiconductor utilized in the fabrication of almost all advanced modern electronics. Silicon allows high-yield, low-cost fabrication to occur on a large scale. Implementing silicon as an efficient optical emitter would allow for even greater achievements in microelectronic fabrication and design [86]. However, the physical properties of silicon make it undesirable for optical applications. Silicon is commonly used as a substrate for electronic devices and is often integrated with optical electronics fabricated on a separate substrate with a direct band gap [87]. Therefore, it would be convenient to identify a material or composite of materials which could be utilized as a substrate for both optical and power devices. Such a material would allow both optical and power devices to be fabricated on the same substrate, reducing system area, heat dissipation, and cost [88].

Implanting a rare earth ion known as Erbium (Er) into the gate oxide of capacitors formed on a silicon substrate may be the solution. Research performed at ST Microelectronics has produced 1.54 μm electroluminescent (EL) devices operating at 300 K with a 10% external quantum efficiency [86, 87]. Erbium ions are excited by collisions with hot electrons. The lifetime of the device is governed by oxide breakdown due to hot electron injection. Therefore, a silicon rich oxide (SRO) is used

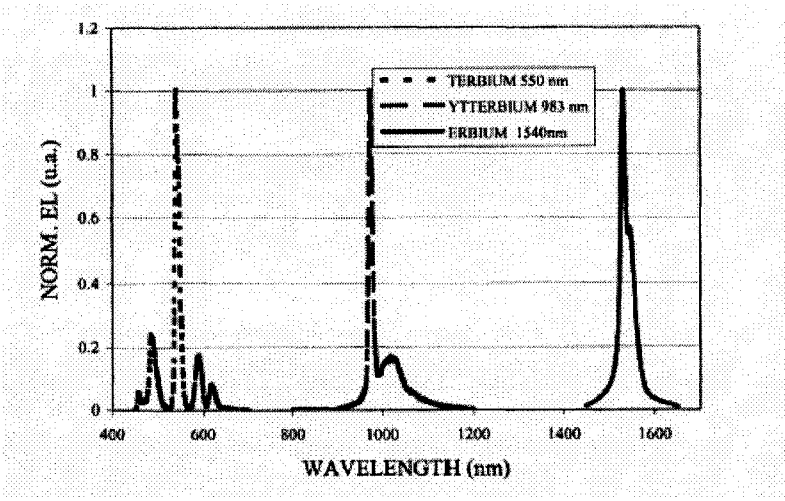


Figure D.1: Normalized room temperature EL spectrum from MOS devices with a rare-earth-doped SiO_2 gate dielectric: Er-doped (–), Yb-doped (- -), and Tb-doped (· ·) [86].

as the gate dielectric in place of a conventional SiO_2 layer [87]. Rare-earth ions other than Erbium, such as Terbium and Ytterbium, may be implanted with similar effect. The rare-earth ions are coupled to nanocrystals in the SRO. Recombination of electron-hole pairs in the SRO causes the excitation of the rare-earth ions [89]. The frequency of the emitted light depends on the identity of the implanted ion, see Fig. D.1 [86].

Device Fabrication

The fabrication of the CMOS capacitors would be similar to the previous research with Er-doped capacitors, see Fig. D.2. ST Microelectronics utilized an epitaxial p^- layer grown on (100) Si boron-doped p^+ substrates. The active areas of the device were formed using a LOCOS structure by growing a 500A oxide layer. A 700Å-thick nitride layer was then formed to protect the oxide and the active area. After the LOCOS formation, the protective nitride and oxide are removed. Next, either a 620Å-

thick oxide is thermally grown, or a 1000Å-thick SRO is deposited by PECVD. Next, the structures are implanted with Er at a dosage of $1x10^{15}$ ions/cm^2 at 50 keV. An implant energy of 50 keV was selected so the ion concentration will be roughly centered at the middle of the gate oxide. After the implantation, annealing at 1000 C for 30 minutes with nitrogen flowing was performed in order to remove any damage from the ion implant and to obtain the growth and crystallization of the nanostructures in the gate oxide. A 3000 A-thick n^+ polysilicon layer is deposited and doped in-situ using a CVD reactor. The poly Si layer is defined using lithography and reactive ion etching to

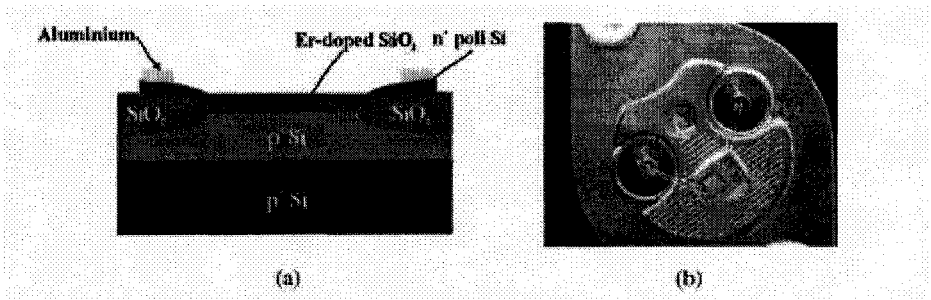


Figure D.2: Schematic cross section (a) and picture (b) of the ST Microelectronics rare-earth light –emitting MOS device [86].

provide electrical contact to the device while being thin enough for the 1.54 μm light generated to emit from the device. Finally, an Al-Si-Cu layer, 3 μm thick is deposited in a metal ring around the device structure to allow bonding to the packaging. Active areas for the MOS devices are 1500 x 1500, 350 x 350 and 100 x 100 μm^2 [87].

The capacitors fabricated at the Sherman Fairchild Laboratory will be fabricated on $7x10^{14}$ ions/cm^2 boron-doped (100) Si substrates. The devices will be defined using standard photolithography, etching, and ion implant with an existing mask set. The mask set includes annular and square capacitors with dimensions of 100 x 100, 200 x

200, 500 x 500, and 1000 x 1000 μm^2. LOCOS isolation will be utilized to define the device structures. After the nitride and oxide are removed and LOCOS formation is complete, a silicon-rich oxide (SRO) will be thermally grown on the substrate. The thickness of the oxide will be varied on individual wafers, as will the composition of the SRO. Thicknesses ranging from 500Å – 1000Å will be thermally grown.

The Erbium will then be implanted into the SRO. Prior to implant, a series of computer simulations will determine the correct ion density and energy to achieve an ion distribution centered at the middle of the gate oxide. The simulations will be run for the various oxide thicknesses used. The wafers will then be sent to an outside company to perform the Erbium ion implant. Sherman Fairchild has been utilizing Implant Sciences Corporation in Wakefield, MA for the majority of its ion implants. This vendor would be able to implant the Erbium ion. After the Er-implant is performed, a thick (~3000Å) layer of polysilicon will be deposited using low pressure chemical vapor deposition (LPCVD). A special mask will be designed using the L-edit software program in order to pattern an aluminum gate which leaves an area open for the generated light to emit from the device. The aluminum will be deposited on sections of the poly in order to create a gate contact for the capacitors.

Conclusions from Previous Research

The results research performed by ST Microelectronics on Er-doped CMOS devices are rather promising. They have reported strong emission from a 2.5 mm^2 device at 300 K at a current of 100 μA. The shape of the spectra observed are typical

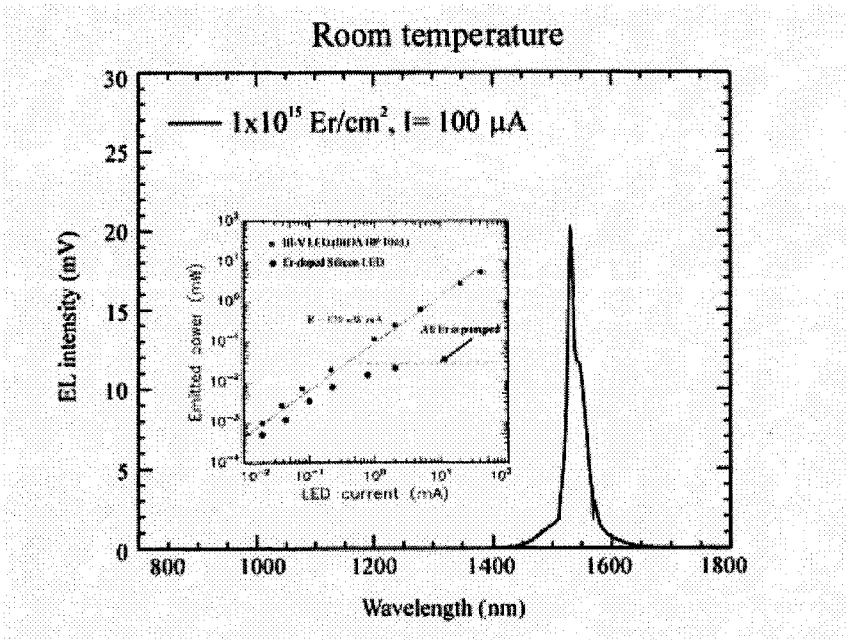


Figure D.3: Room temperature EL spectrum measured at 100 μA on a MOS device with Er-doped gate oxide. In the inset: emitted power versus current for Er-doped MOS and III-V semiconductor LED [87].

for Er emission from an amorphous matrix, see Fig. D.3 [86]. The emitted power was measured using a calibrated power meter positioned immediately above the device in order to collect most of the emitted light. The light measurements demonstrate an external quantum efficiency of ~10%, identical to that of compound III-V devices [86]. Maximum output is limited by the density of Er ions and the radiative lifetime. The electroluminescence can be written as

$$EL = EL_{\max} \frac{\sigma \tau J}{\sigma \tau J + 1} \qquad \text{(D.1)}$$

where EL_{max} is the maximum EL signal at saturation, σ is the effective cross section of

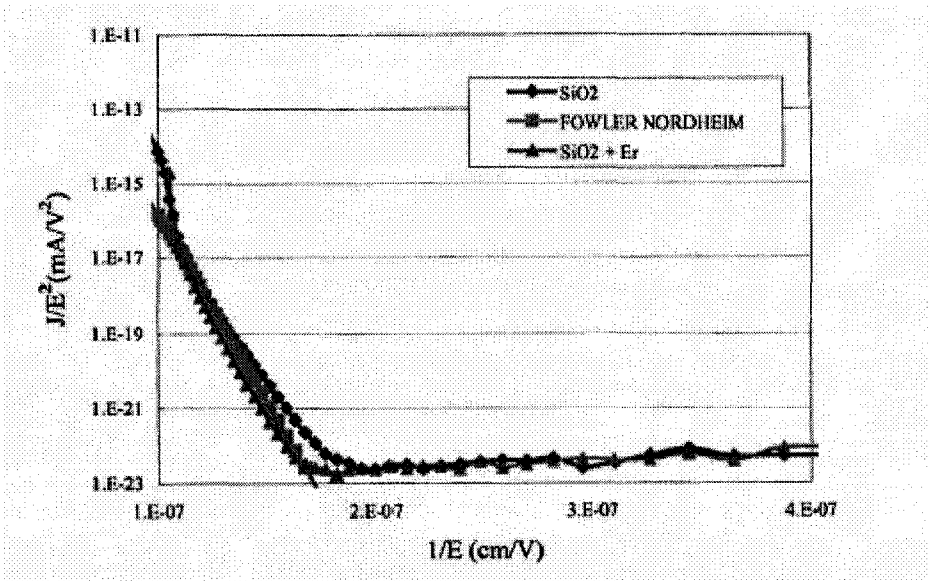


Figure D.4. Fowler-Nordheim plot of the current-voltage measurements performed on undoped (diamonds) and Er-doped (triangles) MOS devices. Expected Fowler-Nordheim behavior is also shown (squares) [87].

excitation, τ is the overall lifetime, and J is the current density [87]. From Fig. D.3, a value of $\sigma\tau = 3 \times 10^{-17}$ cm^2 s is obtained. A measure of the time-decay reveals it is a single exponential with decay time of 3ms. The effective cross-section is 10^{-14} cm^2 [86].

Current-voltage measurements on these devices, see Fig. D.4, reveal the current flow in accumulation conditions can be described as Fowler-Nordheim tunneling. Erbium pumping occurs through impact excitation of hot electrons [90] that are injected into the conduction band of the oxide from the n^+-doped polysilicon layer. These electrons can reach an average energy as high as 5 eV [86]. Analysis of devices doped with Terbium (Tb) and Ytterbium (Yb) demonstrate that Erbium-doped devices emit more power than Tb or Yb-doped devices at the same current levels, see Fig. D.5.

Conclusions

The data presented from results of ST Microelectronics demonstrates the ability to produce an efficient Si-based light-emitting device. Implanting Erbium into the gate oxide of a properly designed and fabricated CMOS capacitor is a proven method used to create these devices. The Sherman Fairchild Laboratory has the facilities and expertise to produce these devices, almost entirely on-site, to ensure quality and creative control.

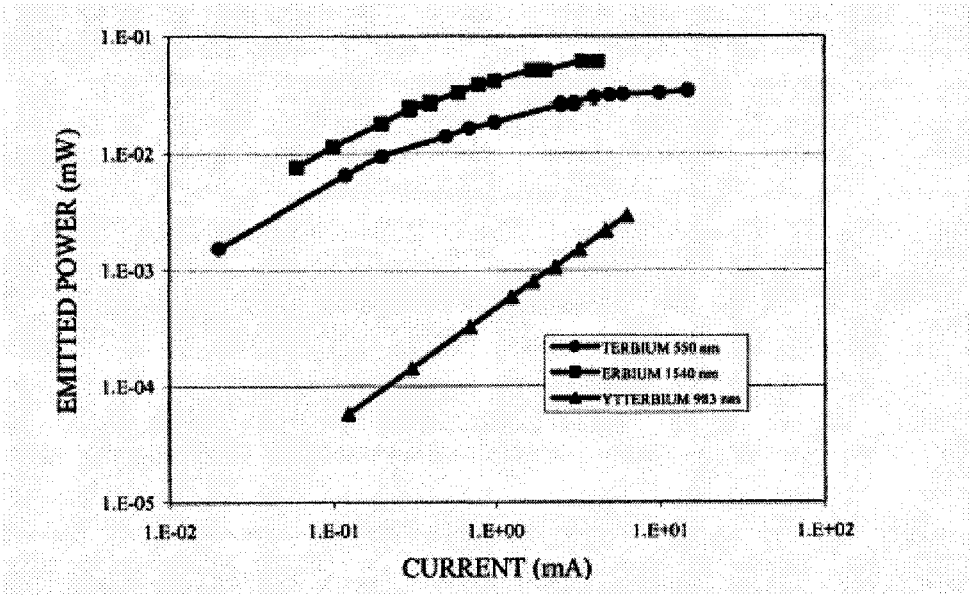


Figure D.5: Emitted power versus current for MOS devices with a rare-earth-doped SiO_2 gate dielectric: Er-doped (squares), Tb-doped (circles), and Yb-doped (triangles) [86].

Appendix E

LabVIEW Program – 4145 Plot Download Program

A LabVIEW™ program has been written to download the data from a 4145 parameter analyzer after a manual sweep has been completed. The majority of the LabVIEW™ programs written at the Sherman Fairchild Lab for use with the 4145 parameter analyzer are specific as to which terminals may have voltages swept and which are used for measuring currents. This program allows the user to set up and run

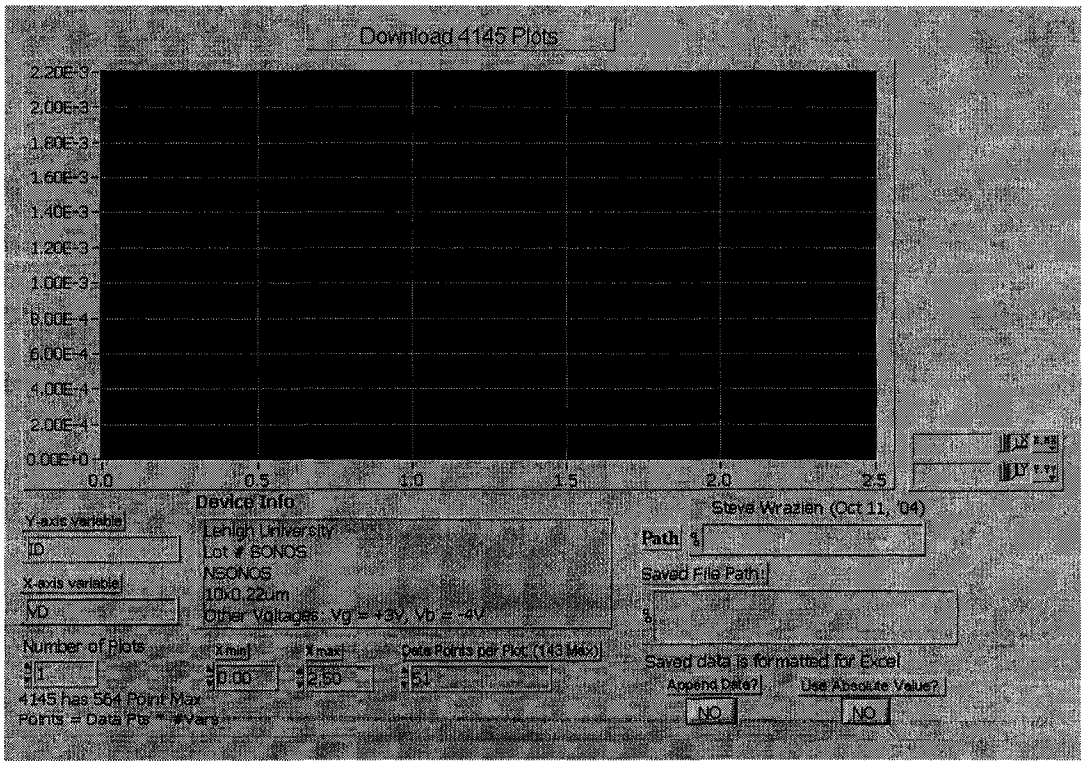


Figure E.1: Front panel of LabVIEW program.

the measurement manually.

The user then connects the 4145 device to the PC running LabVIEW™ and runs the download program. This program downloads the input parameter and saves the data to a spreadsheet with the option to input the axis labels, ranges, and information regarding the measurement. The program downloads a single parameter every time it is run. The program has an append feature which allows the user to run the program several times and save all the data to a single spreadsheet.

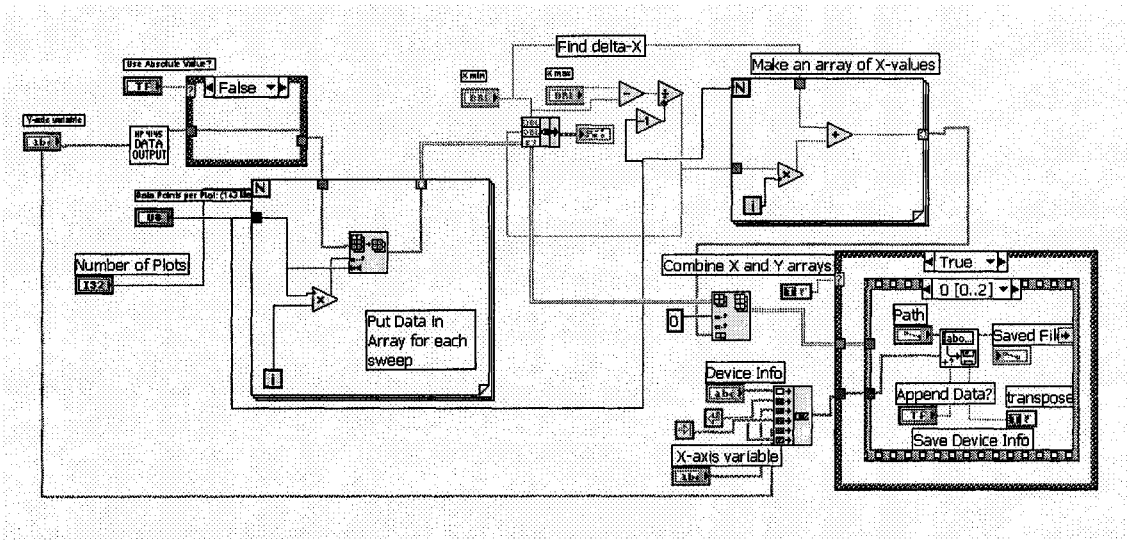


Figure E.2: Back panel of the LabVIEW program.

Vita

Stephen J. Wrazien was born on May 7, 1978 in Hartford, Connecticut to Mr. Carl S. Wrazien and Mrs. Barbara A. Wrazien. He attended the University of Scranton from September 1996 to May 2000 and graduated with a Bachelor's Degree in Electrical Engineering. During his studies at the University of Scranton, he was inducted into the Sigma Pi Sigma and was awarded the Dr. A.J. Cawley Award for Excellence in Electrical Engineering. In the summer of 1999, Mr. Wrazien participated in the research experience for undergraduates program at Lehigh University, working with Dr. Marvin White's research group on SONOS non-volatile memory. In June 2000, he joined Dr. White's research group to begin his graduate career at Lehigh University. In May 2002, he received his Master's Degree from Lehigh University. In the summer of 2003, Mr. Wrazien worked at the Northrop Grumman Advanced Technologies Laboratory in the yield enhancement group supervised by Dennis Adams. During his studies at Lehigh University, Mr. Wrazien received Gotshall and Sherman Fairchild Fellowships and was inducted into the Sigma Xi honor society.

Steve's outside activities include playing on the Physics Softball team (finishing 2nd overall in 1999 and 2000), playing roller hockey in a local league (even recording a hat-trick), volunteering as a DJ at WUSR and WLVR and playing an 'indie rock' show, and recently attempting to golf. He has been a life-long fan of the Boston Red Sox and considers the Sox winning the World Series in 2004 one of the best moments of his life.

Publications

[1] M.H. White, Y. Wang, Y. Zhao, S. Wrazien, *Selected Topics in Electronics and Systems: Advances in Nanoelectronic SONOS Nonvolatile Semiconductor (NVSM) Devices and Technology*, World Scientific Publishing Company, 2005.

[2] D.A.. Adams, J.T. Smith, J.R. Murray, M.H. White, S.J. Wrazien, "Design, Fabrication and Characterization of High Density Radiation Hardened SONOS/CMOS EEPROMs for Space and Military Systems", NVMTS, Orlando, FL, November 2004.

[3] M.H. White, D.A. Adams, J.R. Murray, S.J. Wrazien, Y. Zhao, Y. Wang, B.M. Khan, W.S. Miller, R. Mehrotra, "Characterization of Scaled SONOS EEPROM Memory Devices for Space and Military Systems", NVMTS, Orlando, FL, November 2004.

[4] S.J. Wrazien, Y. Wang, B.M. Khan, M.H. White, "Characterizing Damage to ONO Dielectrics Induced During Programming SONOS/NROM™ Non-Volatile Semiconductor Memory (NVSM) Devices", *Solid-State Electronics*, 48 (10-11), 2004.

[5] S. Wrazien, Y. Wang, B. Khan, M.H. White, "Characterizing Damage to Gate Oxides Induced During Programming SONOS/NROM™ Non-Volatile Semiconductor Memory (NVSM) Devices", ISDRS, Washington D.C., December 10, 2003.

[6] M.H. White, Y. Zhao, Y. Wang, S. Wrazien, J.D. Krayer, D.A. Adams, J.R. Murray, "Scaled SONOS NVSM Devices for Space and Military Applications", NVMTS, San Deigo, CA, November 2003.

[7] S. Wrazien, Y. Zhao, J. Krayer, M.H. White, "Characterization of SONOS Oxynitride Nonvolatile Semiconductor Memory Devices", *Solid-State Electronics*, 47(5), 885-891, 2003.

[8] S. Wrazien, J. Faul and M.H. White, "Characterization of Scaled SONOS NVSM Devices for Space and Military Applications", NVMTS, San Diego, CA, November 2001.