2003

Electrical characterization of polysilicon-oxide-nitride-oxide-silicon (SONOS) nonvolatile semiconductor memory devices

Yijie Zhao
Lehigh University

Zhao, Yijie

Electrical Characterization of Polysilicon-Oxide-Nitride-Oxide-Silicon (SONOS)...

May 2003

Electrical Characterization of Polysilicon-Oxide-Nitride-Oxide-Silicon (SONOS) Nonvolatile Semiconductor Memory Devices

by

Yijie Zhao

A Thesis

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Master of Science

In

Electrical Engineering

Lehigh University

April 2003

This thesis is accepted and approved in partial fulfillment of the requirements for the Master of Science in Electrical Engineering.

April 24, 2003
(date)

Dr. Marvin H. White
Thesis Advisor

Dr. Donald M. Bolle
Chairman of Department

Acknowledgements

I wish to express my deep respect and gratitude to my thesis advisor, Dr. Marvin White, for his inspirational guidance, never ending support and encouragement in both professional and personal aspects during my study at Lehigh University. I would like to thank Stephen Wrazien for familiarizing me with the fabrication and measurements involved in the project, and many helpful discussions from time to time. I would like to thank Northrop Grumman for their continued support of the SONOS project and providing devices for characterization. I would also like to thank the National Science Foundation and the Sherman Fairchild Foundation for their continued funding of our research.

I have had the good fortune to work with and learn from many talented and dedicated graduate students and research scientists in the Microelectronics Lab: Dr. Ling Xie, Yu (Anne) Zeng, Santosh Pandey, Yu (Richard) Wang and Joel Krayer. I have benefited immensely from the previous work on SONOS projects by former students, especially Dr. Jiankang Bu and Dr. Yang Yang. I would like to acknowledge the technical assistance from Mr. Ray Filozof and Dr. Floyd Miller in the device processing. I would also like to thank Ms. Linda Dreisbach for her administrative assistance.

I dedicate this thesis to my husband, Zhenlei and my parents for their tremendous support, encouragement, patience and love in my life. Last but not least, I would like to thank all of my friends who share my worries and happiness all the time, and my colleagues at Lucent Technologies who used to work with me shoulder to shoulder and taught me a lot of valuable experience.

Contents

Acknowledgements iii

List of Tables vii

List of Figures viii

Abstract 1

1.0 Introduction 3

1.1 Nonvolatile Semiconductor Memory Devices 3

1.2 Floating Gate Devices 6

1.3 SONOS Devices 9

2.0 Theory of the SONOS Device 12

2.1 MOSFET Drain Current and Transconductance Model 12

2.2 Threshold Voltage Shift. 19

2.3 Retention Model for SONOS Devices 22

3.0 Measurement Techniques 25

3.1 LabVIEW™ 25

3.2 Charge Pumping Measurement 26

3.3 Linear Voltage Ramp Measurement 30

3.4 Dynamic Measurements 33

3.4.1 Erase/Write Measurement 35

3.4.2 Retention Measurement 36

3.4.3 Endurance Measurement 37

4.0 Electrical Characterization of Scaled SONOS NVSM Devices 39

4.1 Device parameters extraction 39

4.2 Simulation Result 44

4.3 SONOS device with protective gate diodes 46

4.4 SONON Retention at Elevated Temperatures 49

5.0 Conclusions 53

5.1 Conclusion from this Thesis 53

5.2 Future Considerations 55

6.0 References 57

7.0 Appendices 60

A. SONOS device pictures. 60

B. Measurement Procedures (Step by Step) 61

B.1 Current Versus Voltage Measurement 61

B.2 Charge Pumping Measurement 63

B.3 Retention Measurement 65

B.4 Erase/Write Cycling Measurement 67

C. SONOS Device Fabrication (Step-by-Step) 68

C.1 Starting Material 68

C.2 N-Well Formation 68

C.3 LOCOS Isolation 68

C.4 Channel Doping 69

C.5 Gate Formation and Polysilicon Deposition 70

C.6 Source and Drain Implants 70

C.7 Contact Window 71

C.8 Pre-metal High Temperature deuterium Anneal 71

C.9 Metalization 72

C.10 Post-Metal Anneal (PMA) 72

8.0 Vitae **73**

9.0 Publications **73**

List of Tables

4.1 Extracted parameters for SONOS single transistors. 44

B.1 Drain Current versus Gate Voltage Measurement Pin Setup. 61

List of Figures

1.1 Floating gate vs. SONOS device structure. 4

1.2 Schematic of a basic ETOX flash memory device showing the program and erase operation [8]. 7

1.3 Energy-band diagrams. (a) Electron flow during programming by CHE injection. (b) Electron flow during erasing by FN tunneling to the source. 7

1.4 SONOS memory device structure showing paths of electron transport during memory operation [8]. 10

2.1 (a) N-channel MOSFET cross section. . . . . (b) Band diagram. 13

2.2 Write/Erase threshold voltage shift of SONOS device. 22

2.3 (a) Three charge states (D^+, D^0, D^-) and two transitional energies (E_{TA} and E_{TD}) associated with an amphoteric trap for electron and hole processes. (b) Four charge decay mechanisms are illustrated for the discussion of charge retention. 23

3.1 Charge pumping setup [1]. 26

3.2 (a) The recombined charge per cycle Q_{cp} as a function of the base voltage for a SONOS transistor (W=50μm L=0.8μm) with 1.8nm tunnel oxide, 10nm nitride storage layer and 4.5nm blocking oxide. 28

(b) The recombined charge per cycle Q_{cp} as a function of frequency for a thin tunnel oxide SONOS device (W=10μm L=0.8μm). The increase in Q_{cp} at low frequency is attributed to an additional charge pumping current

associated with nitride traps, which are only 1.8 nm away from the Si-SiO_2 interface. 28

3.3 Linear Voltage Ramp Setup [1]. 31

3.4 Front panel of LVR measurement program [22]. 31

3.5 Result of an LVR measurement of a SONOS device. A static memory of 3 V is observed for a programming voltage of 5V. 32

3.6 Measurement system functional block diagram. 33

3.7 A schematic diagram of the DUT when the Erase/Write/Read circuit is at different operational modes [23]. 34

3.8 Write/Erase curve for an SONOS transistor (W = 100 μm, L = 5 μm) using an ultra-thin tunnel oxide with programming voltages of 7V (blue), 6V (teal), and 5V (red) [1]. 35

3.9 Threshold voltage of an ultra-thin tunnel oxide SONOS device versus read delay time for temperatures ranging from 22 C to 175 C [25]. 37

3.10 (a) Interface trap density versus stressing cycles on a SONOS transistor. (b) Threshold Voltage Versus Time under different stressing cycles. The SONOS transistor (W=50 μm, L=0.8 μm) has 1.8 nm tunnel oxide, 10 nm nitride and 4.5 nm blocking oxide. 38

4.1 Drain current and transconductance versus gate voltage with stepped reverse bias on the bulk (V_{DS} = 50 mV). . . . . (a) Single SONOS transistor with dimension of W=50 μm, L=0.8 μm. (b) Single SONOS transistor with dimension of W=50 μm, L=50 μm. 41

4.2 Procedures for the determination of SONOS device modeling parameters. 42

4.3 Comparison between experiment and simulation result for

(a) Long channel NSONOS device (W=50 μm, L=50 μm).

(b) Short channel NSONOS device (W=50 μm, L=0.8 μm). 45

4.4 SONOS structure with protective gate diodes. 47

4.5 (a) Single SONOS Memory Transistor Retention --10 Year Extrapolation for TP1 (without gated diodes) and TP18 (with gated diodes) at Room Temperature. . . . . .

(b) Single SONOS Memory Transistor Charge Pumping Comparison -- TP1 (without gate diodes) has an increased Interface Trap Density, D_{it}, as compared with TP18 (with gate diodes) due to Plasma Damage. 48

4.6 Threshold Voltage decay vs. Temperatures for a scaled SONOS device. The SONOS transistor (W=50 μm, L=0.8 μm) has 1.8 nm tunnel oxide, 10 nm nitride layer and 4.5 nm blocking oxide. 49

4.7 Extracted value of the nitride trap density, $g(E_{TA})$, vs. trap energy, E_{TA}, at different temperatures for a SONOS transistor with Oxynitride. 51

4.8 A Comparison of Trap Densities in 'silicon-rich' and 'oxynitride' SONOS NVSM devices [24]. 52

A.1 SONOS device pictures. . . . . .

(a) Single SONOS transistor of W=50 um, L=0.8 um. . . . . .

(b) SONOS Array consisting of 1000 single devices with each device having W=5 um, L=0.8 um. 60

Abstract

Polysilicon-Oxide-Nitride-Oxide-Silicon (SONOS) nano-devices are used widely as nonvolatile semiconductor memories (NVSMs) in a wide range of applications ranging from NASA space missions, military and commercial satellites to mobile electronics, such as cellular phones and digital cameras. This so-called SONOS technology offers improved performance with a small cell size to meet the challenge of high density, low power, low voltage and radiation hard operations. In addition, it is highly compatible with today's state-of-the-art CMOS technology.

In this thesis we utilize a series of measurement techniques to characterize the electrical behavior and reliability of the SONOS devices. I-V measurements are used to derive key device parameters such as transistor gain, flatband voltage and substrate doping. Charge pumping and linear voltage ramp techniques are used to determine the distribution of traps in the Si-SiO_2 interface and in the nitride storage layer, respectively. To study the memory properties, dynamic measurements are performed to obtain erase/write, data retention, and endurance characteristics.

We have developed a preliminary SONOS mobility model, based on Coulomb scattering and Surface Roughness scattering to characterize electron transport in the surface inversion layer of the SONOS device. With this model we are able to extract modeling parameters to describe the I-V characteristics of SONOS nonvolatile semiconductor memory devices.

We also examined the scaled SONOS devices with protective gate diodes provided by Northrup Grumman. These devices have oxygen-rich films called oxynitride (SiO_xN_y). In

these devices, the data retention and endurance characteristics have improved compared with devices without gated diode protection. A 10-year memory window of 0.5 V is extrapolated from characteristics generated with a +7 V, 2.5 msec write pulse and -7 V, 7.5 msec erase pulse at room temperature for an uncycled SONOS device.

Finally, we have studied the degradation of retention characteristic at high temperatures. Based on a model, which assumes thermal excitation of trapped electrons is the dominant discharging mechanism at high temperatures, we have extracted the trap density profile in oxynitride films with our retention measurements. We show the trap density for an 'oxynitride' storage layer is roughly three times less than the trap density for a silicon-rich nitride layer.

Chapter 1

Introduction

1.1 Nonvolatile Semiconductor Memory Devices

With the advent of the System-On-Chip (SOC) concept, Nonvolatile Semiconductor Memories (NVSMs), such as EEPROMs (Electrical Erasable and Programmable Read Only Memories), have become increasingly attractive for their fast in-system program and erase capability. Their small size and low-power consumption make them ideal for a myriad of portable applications such as cellular phones, digital cameras, personal data assistants and compact smart cards.

The basic characteristics of EEPROM devices include programming speed, data retention and endurance. Programming speed represents the needed amount of time to change the memory device from the written state to the erased state, or vise versa. Data retention is used to characterize the ability of the device to store and recover information after a number of program/erase cycles at a specified temperature. Endurance refers to the ability of the device to withstand repeated program/erase cycles and still meet the electrical specifications.

There are two major types of EEPROM's; floating gate devices and floating trap SONOS devices. Both of the devices are basically Metal-Oxide-Semiconductor-Field-Effect-Transistor (MOSFET) with modified gate stack structures. The floating gate devices store charge in a polysilicon gate electrode as free carriers with a continuous

spatial distribution in the conduction band, while the SONOS (Polysilicon-Oxide-Nitride-Oxide-Silicon) devices store charge in spatially isolated deep level traps (Fig. 1.1).

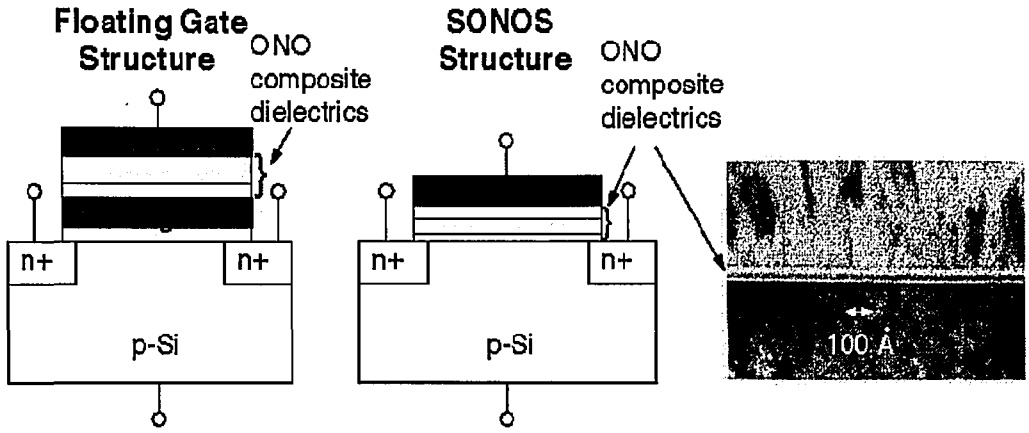


Figure 1.1 Floating gate vs. SONOS device structure.

N-channel devices are used commonly in both technologies because the higher mobility of electrons allows a faster readout. The main differences between the floating gate and the SONOS technologies are as follows [1]:

1) In memory operations, SONOS devices employ both electrons and holes. Floating gate devices use only electrons.
2) SONOS has a simpler structure because it only has one gate with one composite dielectric. Floating gate consists of two floating gates and two dielectrics (oxide and composite).
3) SONOS gate voltage is controlled directly. Floating gate voltage is coupled capacitively from the control gate.
4) SONOS structure is linearly scalable with programming voltage. Floating gate structure is not easily scaled.

spatial distribution in the conduction band, while the SONOS (Polysilicon-Oxide-Nitride-Oxide-Silicon) devices store charge in spatially isolated deep level traps (Fig. 1.1).

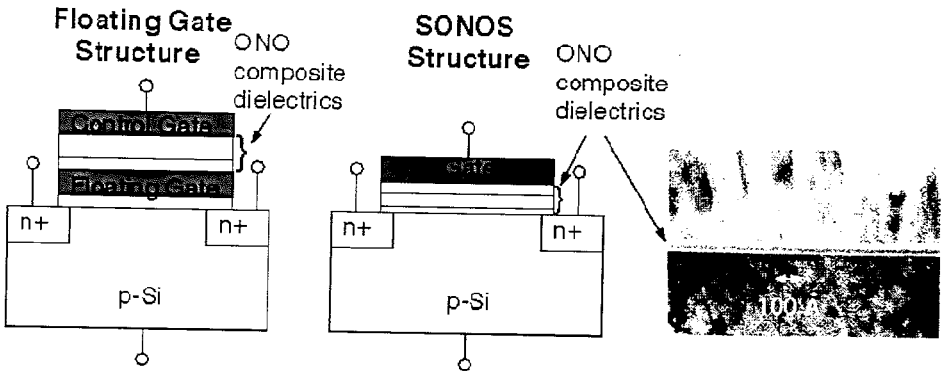


Figure 1.1 Floating gate vs. SONOS device structure.

N-channel devices are used commonly in both technologies because the higher mobility of electrons allows a faster readout. The main differences between the floating gate and the SONOS technologies are as follows [1]:

1) In memory operations, SONOS devices employ both electrons and holes. Floating gate devices use only electrons.
2) SONOS has a simpler structure because it only has one gate with one composite dielectric. Floating gate consists of two floating gates and two dielectrics (oxide and composite).
3) SONOS gate voltage is controlled directly. Floating gate voltage is coupled capacitively from the control gate.
4) SONOS structure is linearly scalable with programming voltage. Floating gate structure is not easily scaled.

5) The SONOS device structure is immune to single defect memory loss because charge is stored in traps distributed throughout the dielectric. Floating gate structure is susceptible to single defect memory loss where a single defect in the tunnel oxide allows free charge to escape into the bulk.

6) Floating gate has thicker tunnel oxide than SONOS devices, generally around the range of 7-10 nm. This leads to better retention for floating gate devices at the expense of higher programming voltages.

7) SONOS devices operate with reduced electric fields in the tunnel oxide (Modified Fowler-Nordheim and Direct tunneling) as opposed to floating gate devices which use exclusively Fowler-Nordheim tunneling.

8) Recent developments with the use of hot electron injection (HEI) have enabled SONOS devices [31] to (a) store two bits of information spatially localized over the source and drain junctions of the device, and (b) permit the shrinking of device cell size to $2.5F^2$, where F is the so-called 'feature' size – a consequence of sharing source lines on adjacent cells.

Future NVSMs require an endurance of 10^6 program/erase cycles, 10-year charge retention at 85°C, low programming voltages of 5-7V and compatibility with standard CMOS technology – a technology with only single-level polysilicon. To achieve low power, low voltage operation while still retaining programming speed and device reliability presents a major challenge for our research in this area. The NVSM program and erase voltages currently used are far too large, especially in floating gate devices, compared with aggressively scaled CMOS logic devices. The net result is the need for high-voltage generator charge-pumping circuits, which degrade array efficiency and

easily break down in radiation environments. In addition, high-voltage devices require additional fabrication steps and reduce yield in manufacturing high-density NVSMs.

To meet the challenges for the next generation NVSM, several new technologies have emerged. FRAM, or Ferroelectric RAM, is a technology that uses high currents to orient the magnetic direction of a tiny strip of magnetic material to record data bits [2]. Unfortunately, fabricating FRAMs in silicon-based facilities will exposes the various processes to contamination. The nanocrystal floating gate memory is based on direct quantum tunneling by electrons from the Si channel to a nanocrystal. For all its advantages, technical issues remain to be solved, such as (a) the control of disturbances and thermal fluctuations in low-voltage operation, (b) understanding the mechanisms for trapping in nanocrystals and their influence on memory performances, and (c) the influence of fixed surface charge on the characteristics of the nanocrystals [3]. The NROM (Nitride Read Only Memory), mentioned previously, seems to be the most promising device structure for high-density, nonvolatile semiconductor memory, which is compatible with standard CMOS – a feature important for cost effectiveness and reliability. However, for NROM to be a reproducible technology suitable for mass production, we need to have a clearer picture of the charge transport and trapping mechanisms in these devices.

1.2 Floating Gate Devices

Floating gate memory, a standard commercial flash device, is similar to the Intel ETOX (EPROM Tunnel Oxide) structure (Fig. 1.2).

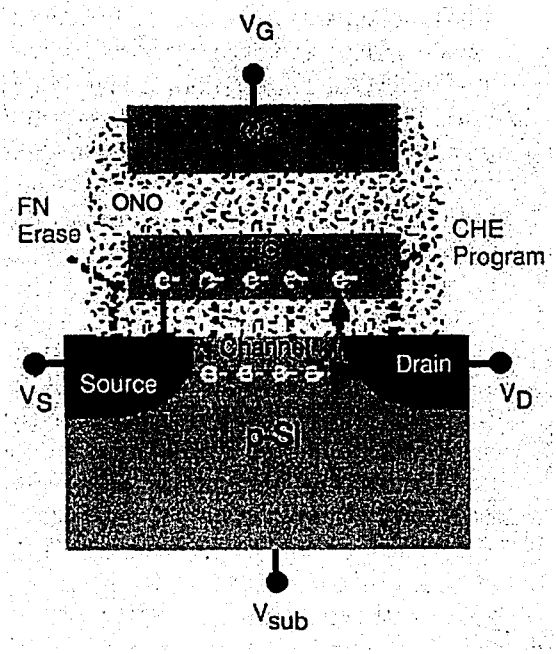
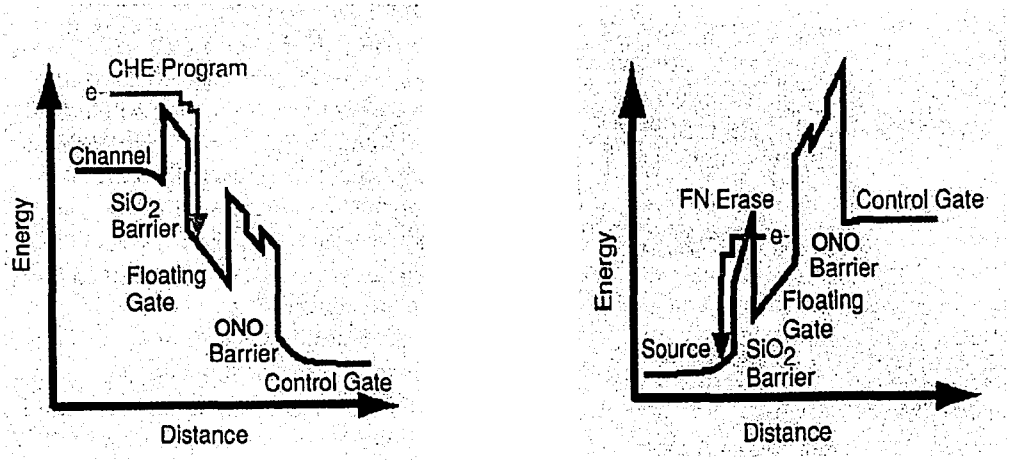


Figure 1.2 Schematic of a basic ETOX flash memory device showing the program and erase operation [8].



(a) CHE programming (b) FN erase

Figure 1.3 Energy-band diagrams. (a) Electron flow during programming by CHE injection. (b) Electron flow during erasing by FN tunneling to the source.

The memory storage element is the isolated floating gate disconnected from terminal voltages. It rests between the control gate and the channel [4]. Fig. 1.3(a) shows a device

cross-section representing the electron energy as a function of the vertical distance from the control gate to the channel near the drain region of the MOSFET. The first SiO_2 energy barrier between the floating gate and the channel prevents electrons from leaking into the channel. The second ONO barrier between the floating gate and the control gate prevents the electrons from escaping to the control gate.

Under typical programming conditions, positive drain and control gate biases are applied to produce a large transverse electric field. This field sets the conditions for a current to flow from the source to the drain. A fraction of the electrons, accelerated by the high field in the channel near the drain region, gain enough kinetic energy to overcome the 3.2-eV SiO_2 barrier and reach the floating gate. This programming method is known as *Channel-Hot-Electron* (CHE) injection: the hot electrons refer to the highly energetic electrons induced by the high electric field. Upon surmounting the barrier, the electrons rapidly lose their kinetic energy (i.e. "*thermalize*") through interactions with the atomic lattice vibrations and other electrons. They "trickle" down and are collected in the floating gate.

Under "memory" erase conditions, a large positive voltage is applied to the source terminal and a small negative voltage may be applied to the control gate. Fig. 1.3(b) shows the electron energy as a function of the vertical distance from the control gate to the source of the MOSFET. The applied electric field between the control gate and the source increases the slope of the energy barrier, effectively thinning the barrier top. At the same time, the high field imparts kinetic energy to some of the floating gate electrons. The electrons move energetically up towards the SiO_2 top where they experience a reduced triangular barrier to *tunnel* quantum-mechanically to the source. This mechanism

is known as Fowler-Nordheim tunneling. Once the tunneling electrons penetrate to the source, they rapidly "*thermalize*" through interactions with the source electrons and the crystal lattice.

Utilizing CHE programming and FN erase mechanisms, floating gate memory devices have major limitations with respect to cell size scaling and programming voltage. The main problems are related to high field stressing and the leaky scaled-down oxide barrier, since a single defect can cause the discharge of the whole memory due to the conductive properties of the floating gate [5]. Currently, there are efforts to find high-permittivity (high-K) dielectric materials to replace SiO_2 as the gate dielectric, with the purpose of minimizing the leakage. However, a thicker high-K dielectric would only improve memory data retention at the expense of program speed [4].

1.3 SONOS Devices

The SONOS technology holds the promise to circumvent the limitations mentioned above. The memory elements of SONOS device are the charge traps distributed throughout the volume of the Si_3N_4 layer that stores both electrons and holes injected from the channel. A typical trap density is in the order of 10^{18}-10^{19} cm^{-3} [6]. These nitride-based memory devices were extensively studied in the early 70s after the first metal-gate nitride device MNOS (Metal-Nitride-Oxide-Silicon) was reported in 1967 by Wegener et al [7]. SONOS research has been motivated largely by the relative small change of threshold voltage to ionizing radiation. As such, they are valued for many military and space applications, where radiation hardness is of primary concern.

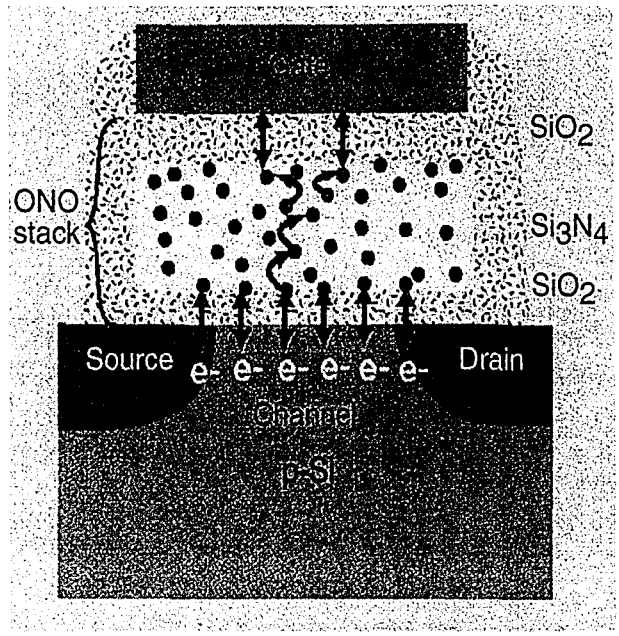


Figure 1.4 SONOS memory device structure showing paths of electron transport during memory operation [8].

In the SONOS device, an oxide layer is introduced between the gate and the nitride region. Thus, it forms the ONO (SiO_2-Si_3N_4-SiO_2) gate dielectric stack (Fig. 1.4) instead of capping the nitride layer with just a metal or semiconductor gate. The purpose of the top blocking oxide is to prevent charge escaping from the nitride layer to the control gate and to reduce the charge injection from the control gate into the nitride layer, which limit the memory window of both MNOS and SNOS devices. The gate dielectrics of SONOS devices can endure the effects of aggressive scaling, which have revived interest in scaled-down nitride memory devices for mainstream NVM applications.

During programming, the gate is biased positively so electrons from the channel can tunnel across the SiO_2 into the nitride layer. Most electrons get trapped in the nitride layer, and some may even tunnel through the blocking oxide, into the gate. The trapped charges produce the same threshold voltage shifts as in conventional floating gate devices. The

charge transport mechanisms involve Modified Fowler-Nordheim tunneling, direct tunneling and Frenkel-Poole emissions. (A Frenkel-Poole emission is the thermal excitation of trapped electrons to the conduction band with the assistance of the electric field. [8]) The erase process occurs when the control gate is biased negatively for holes to tunnel into the nitride film from the silicon bulk.

In our studies, SONOS nonvolatile memory devices are fabricated with a gate dielectric consisting of an 18 Å 'tunneling' oxide, 100 Å oxynitride, and a 45 Å 'blocking' oxide underneath a phosphorus-doped polysilicon gate [9]. We program the devices by applying +7 V to the gate terminal for 2.5 milliseconds and erase the devices by applying –7 V for 7.5 milliseconds. Source, drain and bulk terminals are all grounded during programming, erase and retention operations.

Chapter 2

Theory of the SONOS Device

The SONOS device operates similar to a conventional MOSFET except for one major difference – a conventional MOSFET has a single dielectric - an oxide between the gate and the substrate, while the SONOS device has three dielectrics - a tunnel oxide, a nitride and a blocking oxide. The extra dielectric layers in the SONOS device allow for the trapping of charge in the nitride, which produces a shift in the threshold voltage to serve as a memory element. The conventional MOSFET device has only one constant threshold voltage. Thus, to understand the electrical behavior of SONOS device, we first examine the I-V characteristics of a MOSFET.

2.1 MOSFET Drain Current and Transconductance Model

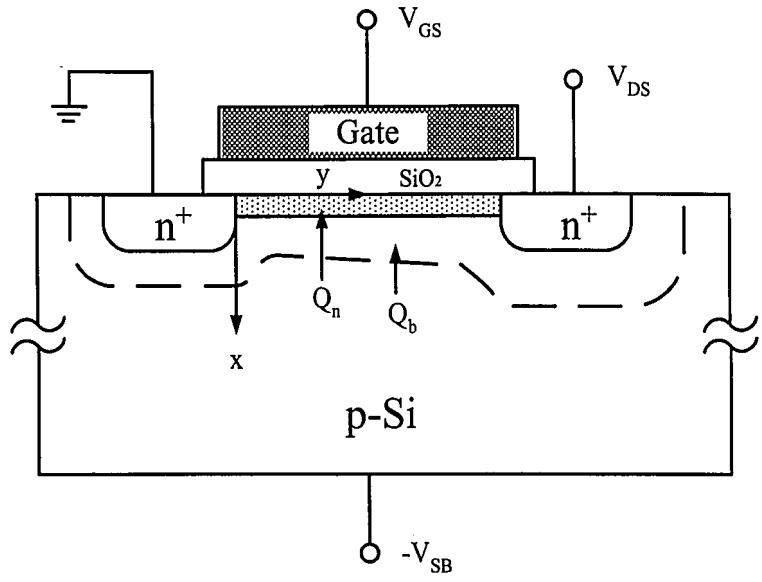
Let us consider a surface channel MOSFET with a uniform doping (or a doping profile which can be approximated by an effective uniform doping). We assume the following [10]:

1) One-dimensional current flow.

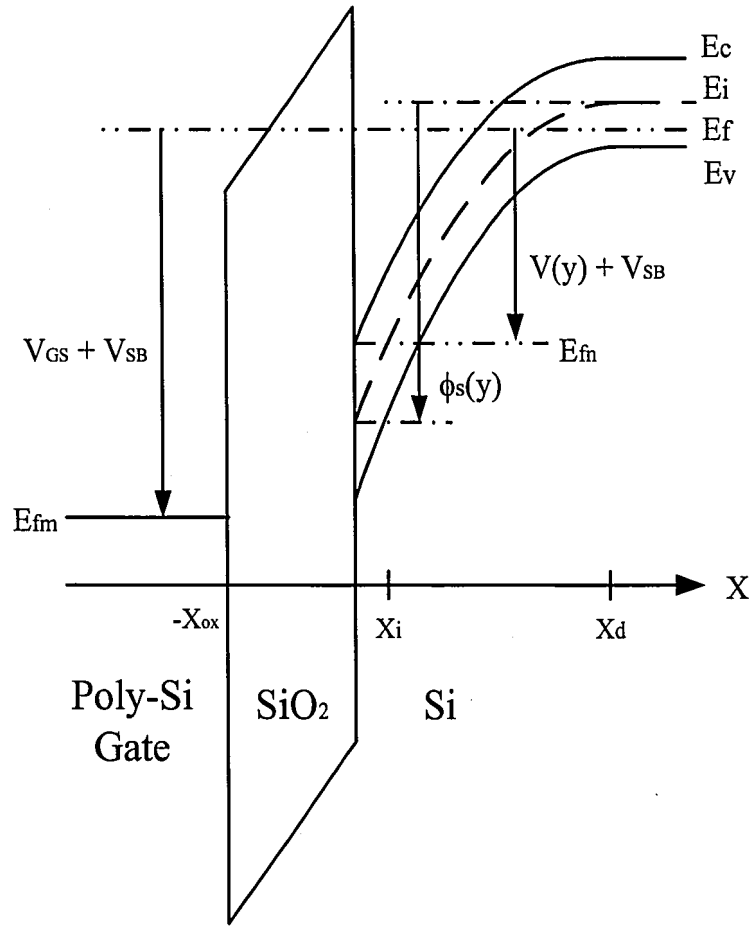
2) Gradual channel approximation, i.e. $\nabla^2\phi \cong \partial^2\phi/\partial x^2$, where ϕ is the potential with espect to the intrinsic Fermi level *Ei*.

3) A 3-D density of states in the conduction band.

4) Maxwell-Boltzmann statistics.



(a)



(b)

Figure 2.1 (a) N-channel MOSFET cross section. (b) Band diagram.

If we employ the notations and the coordinates of Fig. 2.1, then the drain current of an n-channel MOSFET can be written in the Pao-Sah formulation [11]:

$$I_D = q(\frac{W}{L})\int_0^{V_{DS}} dV \int_0^{x_i} \mu_n(x,y)n(x,y)dx$$

$$= q(\frac{W}{L})\int_0^{V_{DS}} dV \int_{\phi_s}^{\phi_F} \frac{\mu_n(x,y)n(x,y)}{d\phi/dx}d\phi \qquad (2.1)$$

where x_i is the inversion layer thickness, n is the electron volume density, μ_n is the electron surface channel conductivity mobility, $V(y)$ is the channel potential, ϕ_s is the surface potential, ϕ_F is the Fermi potential and W and L are the electrical channel width and length, respectively.

The Pao-Sah drain current formulation includes both drift and diffusion currents to model the drain current in the region between subthreshold and strong inversion. This formulation can be modified to incorporate the variation of mobility with normal electric field, fixed oxide charge, and interface states. From Matthiessen's rule, we have

$$\mu_n = \frac{\mu_0}{1 + \frac{\mu_0}{\mu_{SR}} + \frac{\mu_0}{\mu_c}} \qquad (2.2)$$

where μ_{SR} is the surface roughness mobility, μ_C is the coulomb scattering mobility, μ_0 is the low field mobility, and μ_0 includes phonon and impurity scattering in the bulk silicon.

In a quantum mechanical approach, which employs a triangular-well approximation of the potential at the silicon surface [12], the surface roughness mobility (μ_{SR}) at room temperature is inversely proportional to E_{eff}, the effective normal electric field at the interface [13]

$$E_{eff} = (Q_b + \frac{1}{2}Q_n)/k_s\varepsilon_0 \tag{2.3}$$

where $Q_b = Q_b(\phi_s, V, V_{SB})$, $Q_n = Q_n(\phi_s, V, V_{SB})$ are the bulk charge and inversion charge per unit area, respectively and $Q_s = Q_b + Q_n$ is the semiconductor charge density. We introduce E_{crit} as the critical field to describe the influence of surface roughness on the electron mobility. The critical field is related to the electron transport effective mass, the mean asperity height, Δ, and correlation length, L_c, of an assumed Gaussian distributed surface undulation of the oxide/silicon interface [13] and may be written as

$$E_{crit} = \frac{9\hbar L_c}{4\mu_0\Delta^2}\left(\frac{\pi}{m_l m_t}\right)^{1/2} \tag{2.4}$$

where m_l and m_t are the longitudinal and transverse effective masses, respectively for transport in the (100) crystallographic plane of a 2D silicon lattice. We can write

$$\frac{\mu_0}{\mu_{SR}} = \frac{E_{eff}}{E_{crit}} = (\frac{\theta_s}{C_{ox}})\left|2Q_b(\phi_s, V, V_{SB}) + Q_n(\phi_s, V, V_{SB})\right|$$

$$= (\frac{\theta_s}{C_{ox}})\left|2Q_s(\phi_s, V, V_{SB}) - Q_n(\phi_s, V, V_{SB})\right| \tag{2.5}$$

where $\theta_s = \dfrac{k_0}{2k_s x_0 E_{crit}}$ is the so-called *surface roughness parameter.*

The Coulomb scattering mobility, μ_C, accounts for the Coulomb scattering from oxide charges, interface charges and lattice ions. If we consider only the charges in the plane of Si/SiO_2 interface, which is prominent in most cases, then with a quantum mechanical calculation we can write [14]

$$\frac{\mu_0}{\mu_C} = \alpha\left[N_{I0} + \int_0^{\phi_s(y)} D_{it}(\phi)d\phi\right] \tag{2.6}$$

where $\alpha = \frac{\mu_0 m_t e^3}{32\hbar k T \bar{\varepsilon}^2}$ and N_{I0} is the total number of charges per unit area (including fixed charges Q_f and interface charge Q_{it}) in the plane of the Si-SiO_2 interface, at zero band-bending ($\phi_s = 0$), and is independent of ϕ_s. D_{it} is the density of interface traps per unit area per eV, m_t is the transverse electron mass in the silicon and $\bar{\varepsilon}$ is the average of the dielectric constants of Si and SiO_2. Substituting Eqns. (2.5), (2.6) into Eqn. (2.1) yields

$$I_D = q\left(\frac{W}{L}\right)\int_0^{V_{DS}} dV \int_{\phi_s}^{\phi_F} \frac{\mu_0 n(x,y)}{\left\{1+\left(\frac{\theta_s}{C_{ox}}\right)\left|2Q_s(\phi_s,V,V_{SB})-Q_n(\phi_s,V,V_{SB})\right|+\alpha\left[N_{I0}+\int_0^{\phi_s(y)} D_{it}(\phi)d\phi\right]\right\}\left(\frac{d\phi}{dx}\right)} d\phi \tag{2.7}$$

For a given set of terminal voltages, the '*band-bending*', ϕ_s, at the Si-SiO_2 interface determines uniquely Q_s, Q_b, and Q_n (and hence the drain current), which can be calculated by an iterative solution of

$$V_{GB} = V_{GS} + V_{SB} = V_{FB} + \phi_s - \frac{Q_s(\phi_s,V,V_{SB})}{C_{ox}} - \frac{Q_{it}(\phi_s,V,V_{SB})}{C_{ox}} \tag{2.8}$$

where V_{FB} is the flatband voltage. Until now, we have not specified formerly the expressions for the semiconductor charge, Q_s, and the inversion charge, Q_n. Therefore, Eqns. (2.7) and (2.8) are applicable generally for all gate voltages and operational temperatures within the limits of Eqns. (2.1), (2.5) and (2.6), taking into account the effect of mobility variation with electric field.

If we use the four assumptions stated at the beginning of this section, then we can solve the one-dimensional Poisson's equation and calculate Q_s, Q_n and $(\frac{d\phi}{dx})$

$$Q_s = -k_s\varepsilon_0(\frac{d\phi}{dx})\Big|_{\phi=\phi_s} = -\lambda C_{ox}\sqrt{\frac{kT}{q}}F(\phi_s,V,V_{SB}) \tag{2.9}$$

$$Q_n = -\frac{1}{2}\lambda C_{ox}\sqrt{\frac{q}{kT}}G(\phi_s,V,V_{SB}) \tag{2.10}$$

$$(\frac{d\phi}{dx}) = \frac{\sqrt{2}}{L_D}(\frac{kT}{q})F(\phi,V,V_{SB}) \tag{2.11}$$

$$n(x,y) = \frac{n_i^2}{N_A}e^{q[\phi-(V+V_{SB})]/kT} = N_A e^{q[\phi-(2\phi_F+V+V_{SB})]/kT} \tag{2.12}$$

where

$$F(\phi,V,V_{SB}) = \left\{\frac{q\phi}{kT} - 1 + e^{-q\phi/kT} + e^{-q(2\phi_F+V+V_{SB})/kT}(e^{q\phi/kT} - \frac{q\phi}{kT}e^{q(V+V_{SB})/kT} - 1)\right\}^{1/2} \tag{2.13}$$

$$G(\phi_s,V,V_{SB}) = \int_{\phi_F}^{\phi_s}\frac{\exp\{q[\phi-(2\phi_F+V+V_{SB})]/kT\}}{F(\phi,V,V_{SB})}d\phi \tag{2.14}$$

$\lambda = \dfrac{\sqrt{2k_s\varepsilon_0 qN_A}}{C_{ox}}$ is the 'body effect' factor (2.15)

$L_D = \sqrt{\dfrac{kTk_s\varepsilon_0}{q^2N_A}}$ is the 'extrinsic' Debye length (2.16)

$C_{ox} = k_0\varepsilon_0/x_0$ is the oxide capacitance per unit area (2.17)

k_s and k_0 are the relative dielectric constants of silicon and silicon dioxide, respectively, x_0 is the oxide thickness, and N_A is the bulk doping concentration.

If we assume low drain bias ($V_{DS} < 2k_BT/q$) and neglect D_{it} (assume $D_{it} << N_{I0}$), then substituting Q_s, Q_n, $n(x,y)$ and $(\frac{d\phi}{dx})$ into Eqn. (2.7), we can write the drain current for an n-channel MOSFET as

$$I_D = -q\left(\frac{W}{L}\right)\mu(\phi_s, V_{SB})V_{DS}\int_{\phi_F}^{\phi_s}\frac{N_A \exp\{[q(\phi-(2\phi_F+V_{SB})]/kT\}}{\frac{\sqrt{2}}{L_D}\left(\frac{kT}{q}\right)F(\phi,0,V_{SB})}d\phi \tag{2.18}$$

where

$$\mu(\phi_s, V_{SB}) = \frac{\mu_0}{1+\frac{\theta_s}{C_{ox}}[-2Q_s(\phi_s,0,V_{SB})+Q_n(\phi_s,0,V_{SB})]+\alpha N_{I0}} \tag{2.19}$$

The drain current is then differentiated with respect to gate-to-source voltage (V_{GS}) to obtain the transconductance:

$$g_m(V_{GS}) \equiv (\frac{\partial I_D}{\partial V_{GS}})\Big|_{V_{DS},V_{SB}} = (\frac{\partial I_D}{\partial \phi_s})\Big|_{V_{DS},V_{SB}}(\frac{\partial \phi_s}{\partial V_{GS}})\Big|_{V_{DS},V_{SB}} \tag{2.20}$$

From Eqn (2.8), we have

$$(\frac{\partial \phi_s}{\partial V_{GS}})\Big|_{V_{DS},V_{SB}} = \frac{C_{ox}}{C_{ox}+C_s+C_{it}} \tag{2.21}$$

where

$$C_s \equiv -\frac{\partial Q_s}{\partial \phi_s}; \quad C_i \equiv -\frac{\partial Q_n}{\partial \phi_s}; \qquad C_{it} \equiv -\frac{\partial Q_{it}}{\partial \phi_s}; \tag{2.22}$$

C_s, C_i, and C_{it} are the semiconductor space charge, inversion charge and interface charge capacitances per unit area, respectively. Using Eqns. (2.20), (2.21) and differentiating Eqn. (2.18) with respect to ϕ_s, we obtain

$$g_m = q(\frac{W}{L})V_{DS}\frac{C_{ox}}{C_{ox}+C_s+C_{it}}\left\{-\mu(\phi_s,V_{SB})\frac{N_A\exp\{q[\phi_s-(2\phi_F+V_{SB})]/kT\}}{\frac{\sqrt{2}}{L_D}\left(\frac{kT}{q}\right)F(\phi_s,0,V_{SB})}\right.$$

$$\left.+\frac{\mu^2(\phi_s,V_{SB})\frac{\theta_s}{C_{ox}}(2C_s-C_i)}{\mu_0}\int_{\phi_F}^{\phi_s}\frac{N_A\exp\{q[\phi-(2\phi_F+V_{SB})]/kT\}}{\frac{\sqrt{2}}{L_D}(\frac{kT}{q})F(\phi,0,V_{SB})}d\phi\right\} \qquad (2.23)$$

2.2 Threshold Voltage Shift

The threshold voltage of a MOSFET is defined at the point of strong inversion as

$$V_{GS}(\phi_s=2\phi_F+V_{SB})=V_{th}=V_{FB}+2\phi_F+\frac{\sqrt{2\varepsilon_{si}qN_A(2\phi_F+V_{SB})}}{C_{ox}}-\frac{Q_{it}}{C_{ox}} \qquad (2.24)$$

with the Fermi potential defined as

$$\phi_F=\frac{kT}{q}\ln(\frac{N_A}{n_i}) \qquad (2.25)$$

where Q_{it} is the interface trap charge defined at the point of strong inversion, and $C_{ox} = \varepsilon_{ox}/x_o$ is the oxide capacitance per unit area. V_{FB}, a generalized flatband voltage, may be written as

$$V_{FB}=\phi_{GS}-\frac{1}{\varepsilon_{OX}}\int_0^{x_o}x\rho_{ox}(x)dx \qquad (2.26)$$

where ϕ_{GS} is the gate to semiconductor work function difference, ε_{OX} is the dielectric constant of the silicon dioxide, ρ_{ox} is the variable oxide charge, and x_o is the oxide thickness. In general, the interface trap charge at strong inversion, Q_{it}, is neglected in conventional MOSFET theory, but may be significant in SONOS devices. In Eqns. (2.24)

and (2.26) we consider the oxide charge to be fixed and not mobile under bias. Thus, the flatband voltage, $V_{FB,}$ is considered to be constant in a conventional MOSFET. Therefore, the threshold voltage will not change as we vary the gate and substrate biases.

In contrast with a conventional MOSFET, as we have discussed in Chapter I, the SONOS device can have a variable threshold voltage since the net charge in the nitride film can be made either positive or negative by suitably programming or erasing the device, respectively. The flatband voltage shift, ΔV_{FB}, includes information on the charge distribution within the nitride and may be written as [15]

$$\Delta V_{FB} = -\int_0^{x_N} \left\{ \frac{x_{OB}}{\varepsilon_{OX}} + \frac{x_N - x}{\varepsilon_N} \right\} \Delta\rho_n(x)dx = -\left\{ \frac{x_{OB}}{\varepsilon_{OX}} + \frac{x_N - \bar{x}}{\varepsilon_N} \right\} \Delta Q_n \tag{2.27}$$

where,

$$\Delta Q_n = \int_0^{x_N} \Delta\rho_n(x)dx \tag{2.28}$$

is the change density per unit area in the silicon nitride due to a change in the volume charge density $\Delta\rho_n$ as a result of program and erase operations.

$$\bar{x} = \frac{\int_0^{x_N} x\Delta\rho_n(x)dx}{\int_0^{x_N} \Delta\rho_n(x)dx} \tag{2.29}$$

is the *incremental charge centroid*, x_N is the nitride thickness, x_{OB} is the blocking oxide thickness, and ε_N is the dielectric constant of nitride. If we assume an initial charge state for the silicon nitride with a uniform trap density completely occupied, either by electrons or holes, then the charge centroid becomes $\bar{x} = x_N/2$ if we completely convert all of the traps to the opposite charge state. As a result, the threshold voltage of SONOS device can be expressed as

$$V_{th} = \phi_{GS} + 2\phi_F + \frac{\sqrt{2\varepsilon_{si} q N_A (2\phi_F + V_{SB})}}{C_{eff}} - \left[\frac{x_{OB}}{\varepsilon_{OX}} + \frac{x_N}{2\varepsilon_N}\right] Q_n - \frac{Q_f + Q_{it}}{C_{eff}} \qquad (2.30)$$

where we have assumed the oxide charge in the tunnel oxide can be represented by a fixed charge, Q_f, at the Si-SiO_2 interface, and the oxide capacitance per unit area is replaced by an effective capacitance, $C_{eff} = \varepsilon_{OX}/x_{eff}$, where $x_{eff} = x_{OT} + x_{OB} + (\varepsilon_{OX} / \varepsilon_N) x_N$ is the effective ONO thickness.

In the programming operation for the SONOS device, when negative charge carriers (electrons) are injected into the nitride, the threshold voltage shifts in the positive direction. In the erase operation, when positive charge carriers (holes) are injected into nitride, the threshold voltage shifts in the negative direction. Fig. 2.2 illustrates the drain current (I_D) and transconductance (g_m), as a function of gate voltage (V_{GS}) for a small drain bias (V_{DS} = 50 mV). The three curves depict the so-called fresh or virgin state of the SONOS device, the written and the erased states. The threshold voltage, to first order, can be determined by drawing a tangent to the point on the I_D curve where the g_m goes through a maximum. This tangent is extrapolated to zero current where the voltage V_{GS} = V_{th}. If we examine Fig. 2.2, then the memory window is determined by the difference between the measured threshold voltage in the written state and the measured threshold voltage in the erased state. The shift in threshold voltage may be written as

$$\Delta V_{th} = \left(\frac{x_{OB}}{\varepsilon_{OX}} + \frac{x_N}{2\varepsilon_N}\right) \Delta Q_N \qquad (2.31)$$

and the shift will reach maximum for a uniform trap density N_T when all the traps are alternately filled with electrons and holes by $\Delta Q_N = 2qN_T x_N$. Eqn (2.31) assumes the threshold voltage shift is caused only by trapped charge in the silicon nitride film.

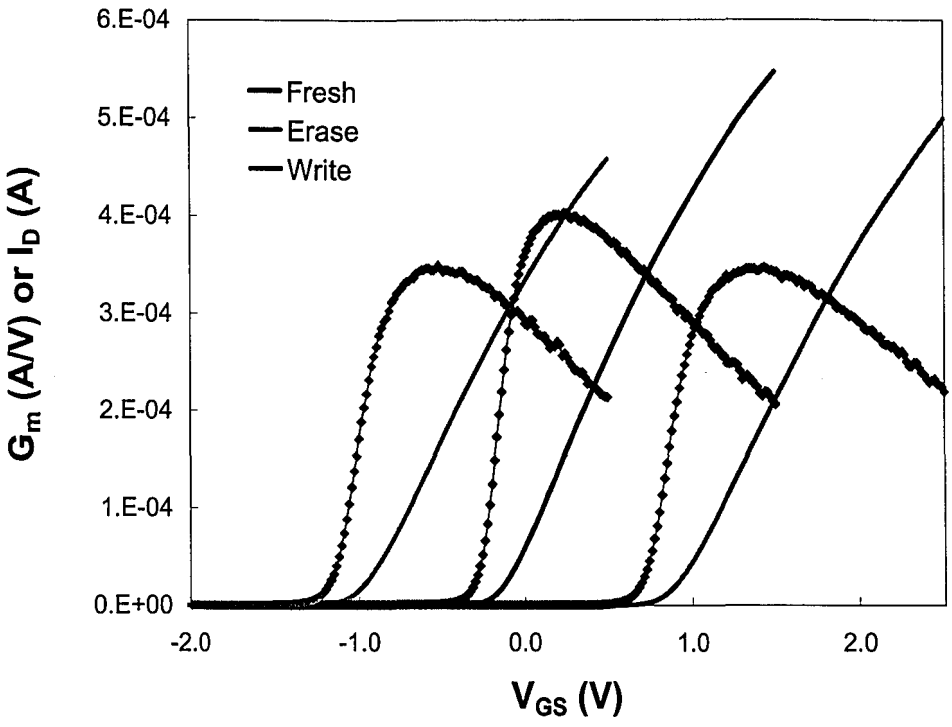


Figure 2.2 Write/Erase threshold voltage shift of SONOS device

2.3 Retention Model for SONOS devices

After a SONOS device is written or erased and resides or rests in the retention mode (all terminals are grounded), the charge stored in the nitride will leak out or 'back-tunnel' with a corresponding decay in the threshold voltage. We will use an amphoteric model to describe the trapping and detrapping phenomenon. In the amphoteric trap model traps of both charge states are attributed to a trivalent silicon center [16-17]. This center consists of a silicon atom with a dangling bond, while the remaining three bonds are bonded with nitrogen atoms. The charge state is neutral (D^0) when a single electron is attached to the dangling bond, negative (D^-) when two electrons are attached to the dangling bond, and positive (D^+) in the absence of electron bonding (i.e. a hole is attached to the bond). The

interaction of an amphoteric trap with electrons and holes is illustrated in Fig. 2.3(a). E_{TA} and E_{TD} are the two transitional energy levels.

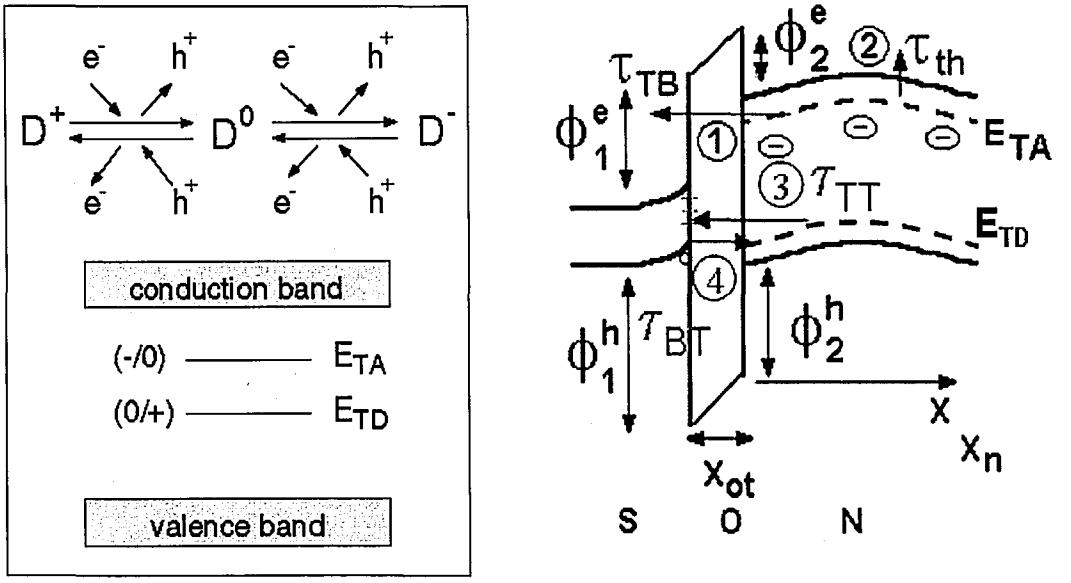


Figure 2.3 (a) Three charge states (D^+, D^0, D^-) and two transitional energies (E_{TA} and E_{TD}) associated with an amphoteric trap for electron and hole processes. (b) Four charge decay mechanisms are illustrated for the discussion of charge retention.

If we assume the nitride film is initially filled with injected electron, all the traps are in the D^- state, and the blocking oxide is thick enough to prevent any charge passage, then the threshold voltage decay can be associated with the following processes [1]:

1) Trap-to-band tunneling associated with the doubly occupied D^- state to the silicon conduction band with time constant τ_{TB}.

2) Thermal excitation of the trapped electrons associated with the doubly occupied D^- state to the nitride conduction band, followed by Frenkel-Poole drift conduction through the silicon nitride to the tunnel oxide-nitride interface. Subsequently, these carriers tunnel to the silicon substrate under the influence of the internal oxide electric field (see τ_{th} in

Fig. 2.3(b)). Thermal excitation of electrons from the singly occupied D^0 state is neglected since the corresponding energy level is much deeper in the silicon nitride band gap.

3) Trap-to-trap tunneling of electrons from the singly occupied D^0 state to a $Si\text{-}SiO_2$ interface state with time constant τ_{TT}.

4) Band-to-trap tunneling of holes from the silicon valence into the nitride traps under the influence of the internal oxide electric field with time constant τ_{BT}.

Chapter 3

Measurement Techniques

Almost all the measurement techniques performed on conventional MOS transistors can be applied to SONOS nonvolatile memory devices. For example, we can study the transfer characteristics of a SONOS device with an I-V measurement and study interface traps at the Si-SiO_2 interface with a charge pumping technique. In addition, to study the memory properties of a SONOS device we need to have some special characterization methods. These include the Linear Voltage Ramp (LVR) and charge separation technique to study the static memory window, and a set of dynamic measurements associated with Erase/Write, data retention and endurance. Using these techniques, we can determine how well the SONOS memory device stores charge under different conditions.

3.1 LabVIEW™

All the measurements presented in this thesis are performed with the aid of a computer program called LabVIEW™. LabVIEW™ is a software platform developed by National Instruments. In the LabVIEW™ environment, a graphical programming language is used to write programs to control test equipments, such as oscilloscope, electrometer, and function generator. These programs are usually referred to as "virtual instruments (VIs)" because their appearance and operation imitate actual instruments. With a VI, users can control and monitor simultaneously several test equipments to perform a measurement. Communication between the computer, the instruments and the

device is accomplished through a GPIB cable. LabVIEW™ also allows data to be sent back to the computer, where it can be displayed on the screen and written to a file.

3.2 Charge Pumping Measurement

The charge pumping measurement is a widely accepted technique used to determine the interface trap parameters in MOS devices [1]. We use this technique on SONOS devices to measure the interface trap density before and after the device has been Erase/Write cycled. Figure 3.1 demonstrates a bi-level charge pumping setup where a SONOS device is connected in a gated diode configuration by shorting the source and drain together. A pulse train is applied to the gate of the device from a functional generator, which drives the surface repetitively from accumulation to inversion. The interface traps serve as recombination centers in the process and a net recombination current can be measured at the drain/source terminal.

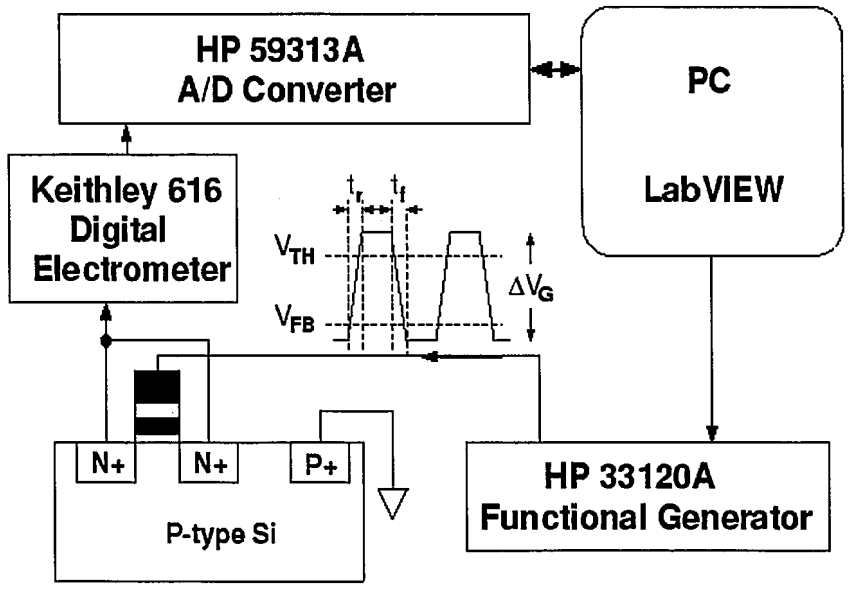


Figure 3.1 Charge pumping setup [1].

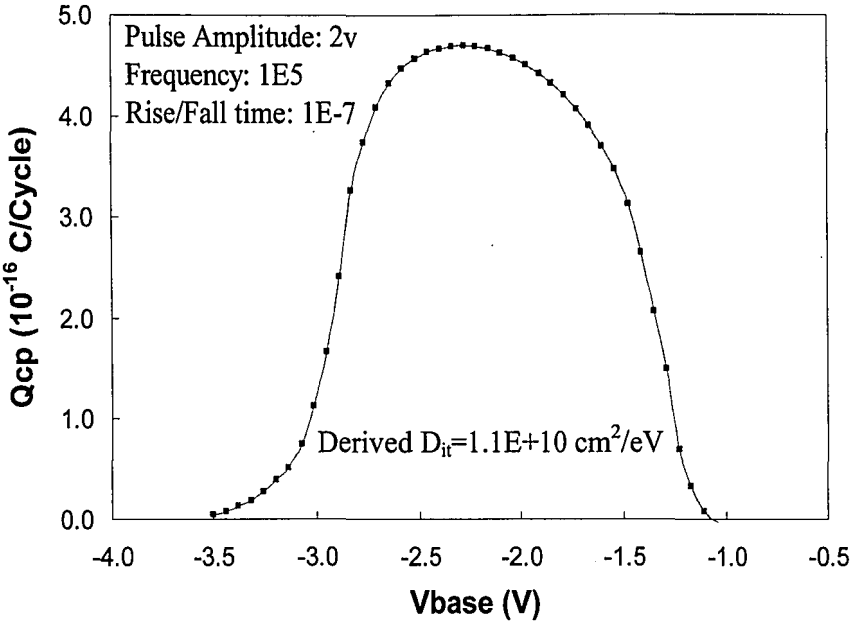
The charge pumping current, I_{cp}, can be derived as [18]:

$$I_{cp} = 2q\overline{D_{it}} f A_G kT \ln\left[v_{th} n_i \left| \frac{V_{FB} - V_{th}}{\Delta V_G} \right| \sqrt{\sigma_n \sigma_p t_r t_f} \right] \tag{3.1}$$

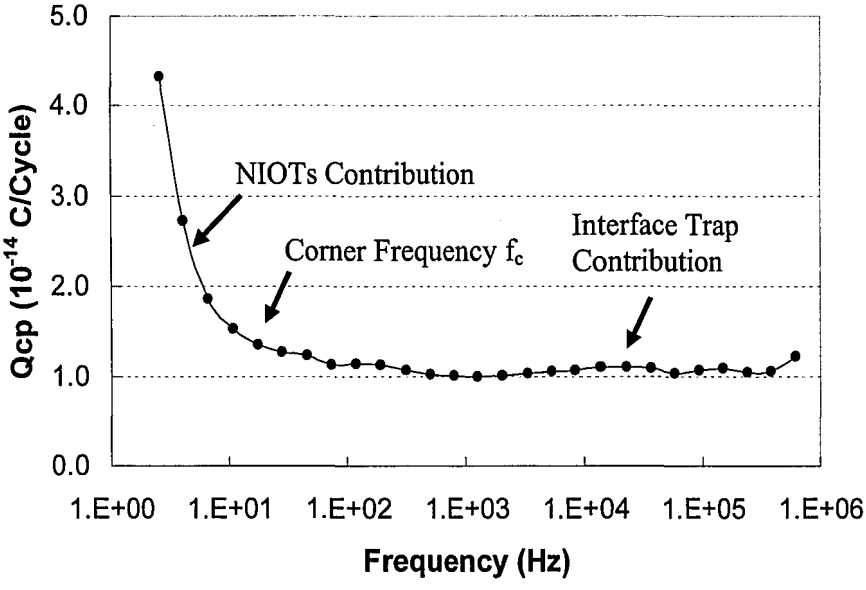
where A_G is the area of the gate, v_{th} is the thermal velocity of the carriers, σ_n and σ_p are the capture cross sections for electrons and holes, V_{FB} and V_{th} are the flatband and threshold voltages of the device, ΔV_G, t_r, and t_f are the amplitude, rise time, and fall time of the pulse applied to the gate, respectively.

As seen in Eqn. (3.1), the charge pumping current is directly proportional to the average density of interface states $\overline{D_{it}}$ and the frequency of the pumping pulse, f. In order to remove the frequency dependence, the charge recombined per cycle, $Q_{cp}=I_{cp}/f$ is examined, which is generally constant in the frequency range from 100 KHz to 1 MHz. At these frequencies, only the traps within 10 Å of the Si-SiO_2 interface can respond and contribute to the charge pumping current. However, for SONOS nonvolatile memory devices, which possess a high density of near-interface oxide traps (NIOTs), if the device is held in inversion for a period of time longer than the tunneling time constant, then the trapped carriers at the interface will have enough time to communicate with the NIOTs. This results in an increase of charge recombined per cycle as the frequency of the applied pulse is lowered [19].

Variable frequency charge pumping technique is used to monitor the charge pumping current over a wide range of gate pulse frequencies. A typical plot of charge pumped per cycle, Q_{cp}, versus pumping frequency is shown in Fig. 3.2 (b).



(a)



(b)

Figure 3.2 (a) The recombined charge per cycle Q_{cp} as a function of the base voltage for a SONOS transistor (W=50μm L=0.8μm) with 1.8nm tunnel oxide, 10nm nitride storage layer and 4.5nm blocking oxide. (b) The recombined charge per cycle Q_{cp} as a function of frequency for a thin tunnel oxide SONOS device (W=10μm L=0.8μm). The increase in Q_{cp} at low frequency is attributed to an additional charge pumping current associated with nitride traps, which are only 1.8 nm away from the Si-SiO_2 interface.

The data obtained from the slope and corner frequency in the variable frequency charge pumping measurement provides information regarding the density and location of the NIOTs and, thus, the memory traps in the silicon nitride film. If we assume the carriers in the silicon bands communicate with the NIOTs through an intermediate interface trap (i.e. a two-step process), then we can write an expression for the contribution of mono-energetically, distributed NIOT's to the charge pumped per cycle, Q_{cp} (NIOT), in the manner [20]

$$Q_{cp}(NIOT) = qA_g \int_{x_{\min}}^{x_m(f)} dx N_T(x)[1 - e^{-\frac{1}{2\tau_{TT}(E_{TO},x)}}] \tag{3.2}$$

where

$$\tau_{TT}(E_{TO},x) = \frac{(m^*_{ox,e})^2 x(1+\frac{1}{2\kappa_1 x})}{2\pi^2 \kappa_2 \hbar^3 D_{it}} e^{2\kappa_1 x} \approx \tau_0 e^{2\kappa_1 x} \tag{3.3}$$

is the trap-to-trap tunneling time constant, which describes the carrier tunneling phenomenon between the interface traps and the NIOT's. The tunneling time constant, τ_0, is inversely proportional to $\overline{D_{it}}$ and is weakly dependent on x. $N_T(x)$, is the spatial trap density of NIOT's and $x_{\min}$ is the minimum distance at which a NIOT is distinguishable from an interface trap, which is determined by an experimental 'corner' frequency, $f = f_c$, in the measurement process. $x_m(f) = \frac{1}{\kappa_1}\ln[\frac{1}{2\ln 2 f\tau_o}]$ is the maximum distance into a dielectric that a gate pulse of frequency f can probe and $\hbar$ is the reduced Planck's constant. $\kappa_1 = \sqrt{2m^*_{ox,e}(\phi_B + E_c - E_T)}/\hbar$ and $\kappa_2 = \sqrt{2m^*_{si,e}(E_c - E_T)}/\hbar$ 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 [20]. Using the slope of the experimental curve (Fig. 3.2), the density of NIOT's at the maximum tunneling distance x_m can be determined with the expression

$$N_T(x_m) \approx \frac{2\kappa_1}{2.3qA_G}\left(-\frac{\partial Q_{cp}}{\partial \log f}\right) \tag{3.4}$$

From the previous discussion, we can vary the gate pulse frequency f, and with the aid of Eqn. (3.4) we can determine the spatial distribution of NIOT's. The assumption of a two-step process for carrier communication to NIOTs has an alternative model, which assumes the carriers tunnel from the silicon bands to a virtual energy state in the insulator with subsequent relaxation to a NIOT ground state [32, 33]. Rather than employ a trap-to-trap tunneling time constant, which is exponentially related to the trap distance, this model assumes a trap capture cross section with an exponential dependence on the distance into the insulator.

3.3 Linear Voltage Ramp Measurement

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. In a typical LVR setup (Fig. 3.3), the source and drain of the device are tied to the bulk. A functional generator controlled by the LabVIEW™ sends a ramping voltage to the bulk of the device. The voltage is applied to the bulk rather than the gate electrode to reduce the system noise caused by capacitive loading. The ramp rate, α, is small to maintain quasi-thermal equilibrium in the device. The gate current is measured with an electrometer. Fig. 3.4 shows the front panel of the control program.

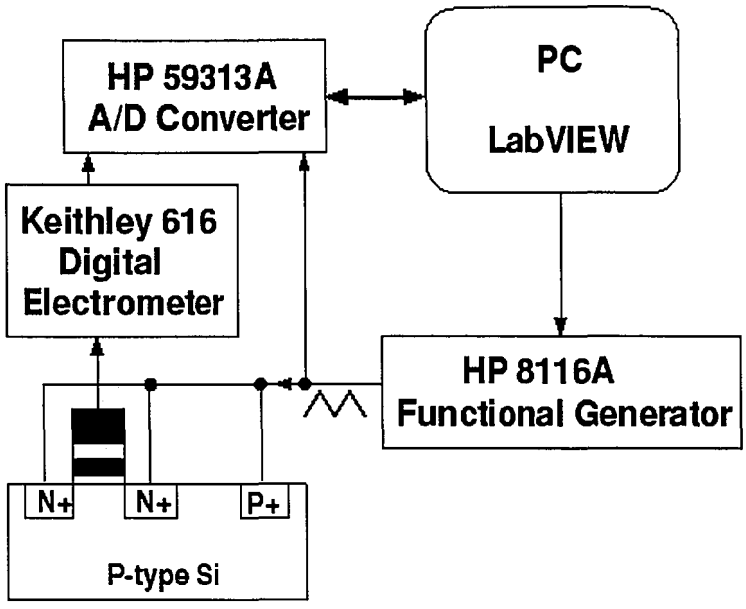


Figure 3.3 Linear Voltage Ramp Setup [1]

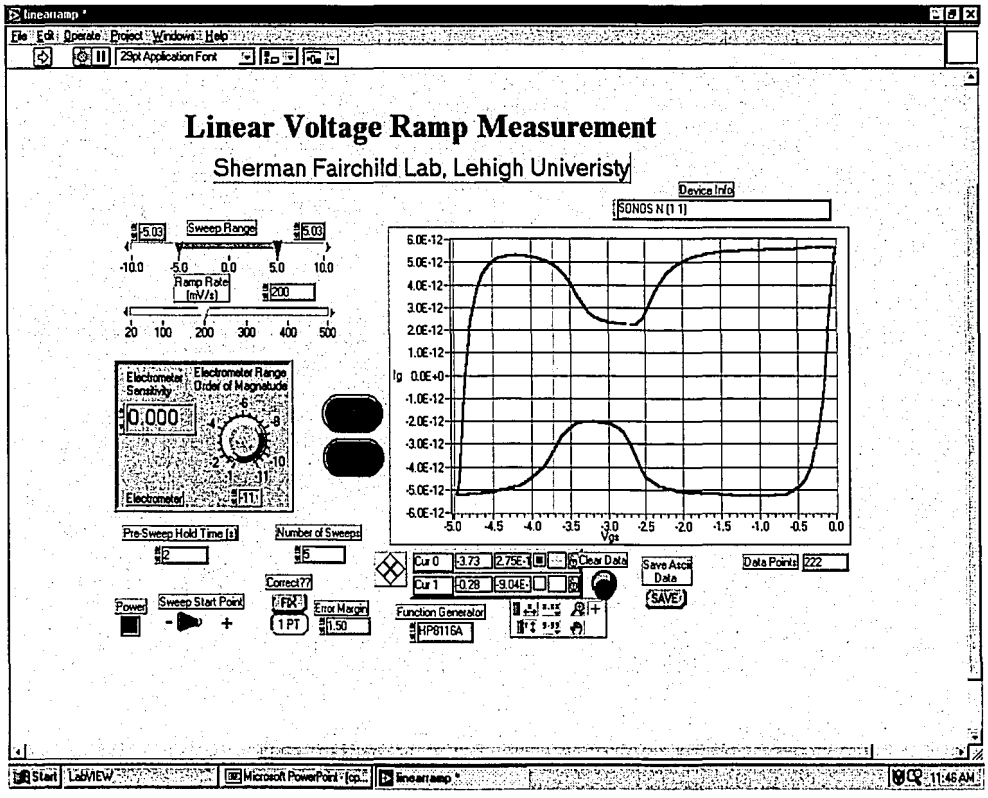


Figure 3.4 Front panel of LVR measurement program [22].

The ramping voltage can be written as

$$V_{GB} = V_O \pm \alpha t \tag{3.5}$$

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

$$I_g = \frac{\partial Q_g}{\partial t} A_G = \frac{\partial Q_g}{\partial V_{gb}} \frac{dV_{gb}}{\partial t} A_G = \alpha C_{eff} A_G = \alpha \frac{\varepsilon_{ox}}{x_{eff}} A_G \tag{3.6}$$

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 dielectric constant, and A_G is the gate area of the device. In addition, the LVR measurement can be used to examine the static memory window of a SONOS device.

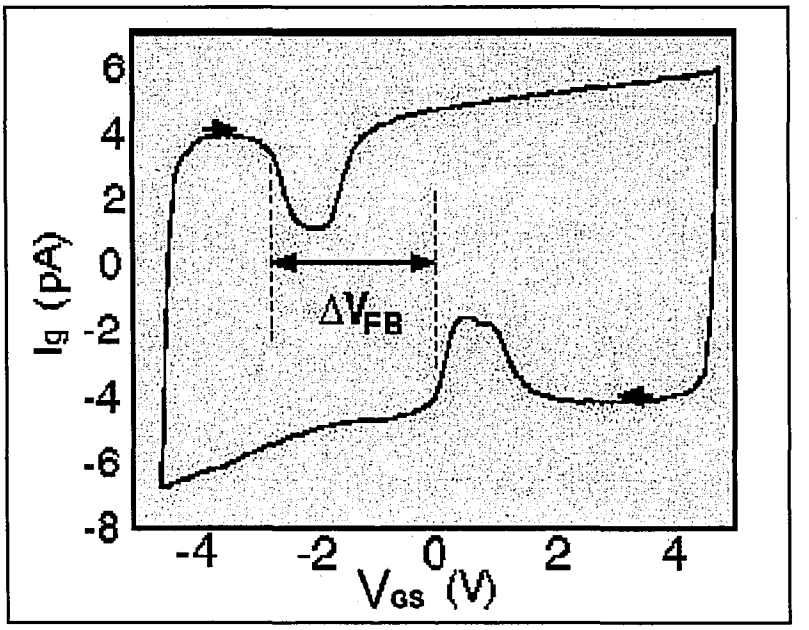


Figure 3.5 Result of an LVR measurement of a SONOS device. A static memory window of 3 V is observed for a programming voltage of 5 V.

As shown in Fig. 3.5, suppose the voltage ramp starts at a positive value where the semiconductor surface is in accumulation, so the nitride traps are filled with holes. As the bulk voltage is swept from the positive value to a negative value, the surface of the

semiconductor goes from accumulation to inversion. During this process, electrons are injected and trapped in the nitride, thereby, shifting the flatband voltage in the positive direction. In the reverse sweep, holes injection and trapping shift the flatband voltage negatively.

3.4 Dynamic Measurements

From a user's point of view, erase/write (how fast), retention (how long) and endurance (how robust) are the critical tests for NVSM devices. In our laboratory, all three measurements can be conducted with an Erase/Write/Read setup originally designed and implemented by Roy [21], but modified by Banerjee [22] and more recently by Bu [23]. Modifications have been made to enhance the automation level and to improve the transient response of the Erase/Write/Read circuit. Fig. 3.6 represents the functional block diagram of the experimental setup.

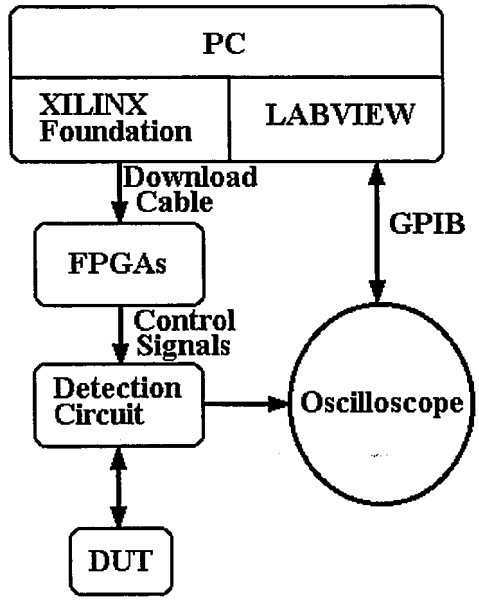


Figure 3.6 Measurement system functional block diagram.

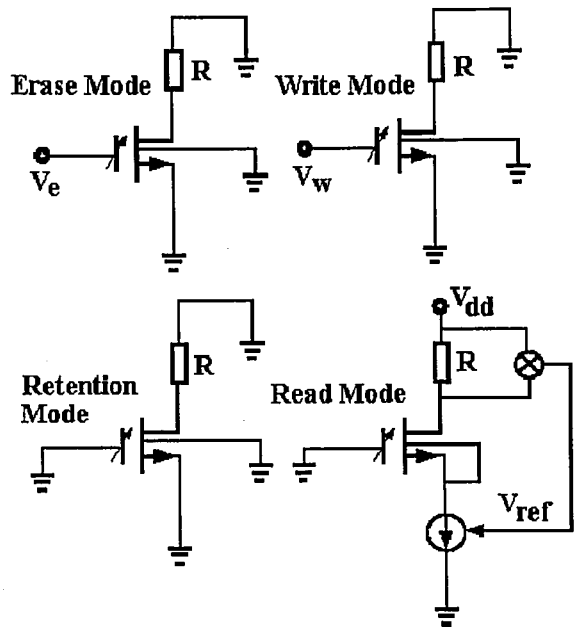


Figure 3.7 A schematic diagram of the DUT when the Erase/Write/Read circuit is at different operational modes [23].

Fig. 3.7 illustrates the different states of the device under test (DUT), when the Erase/Write/Read circuit in at different operational modes. To read out the threshold voltage of a device, a current of 10 μA is forced through the channel while the gate is grounded. The source-gate voltage is measured with a Tektronix TDS 460 digital oscilloscope, where the negative of the voltage equals the threshold voltage of the device.

$$V_{GS} = V_{th} + \sqrt{\frac{2I_{DS}}{\beta_n}} \tag{3.7}$$

$$\beta_n = \mu_{eff}(\frac{W}{L})C_{eff} \tag{3.8}$$

where μ_{eff} is the effective, voltage-dependent, carrier mobility, W/L is the width to length ratio of the device under test, and C_{eff} is the effective capacitance defined previously.

3.4.1 Erase/Write Measurement

The Erase/Write test is to measure of the programming speed of a device. The larger the amplitude of the applied voltage pulse, the quicker the device will be erased or written. The typical voltage pulse duration is in the range of 1 ms to 10 ms.

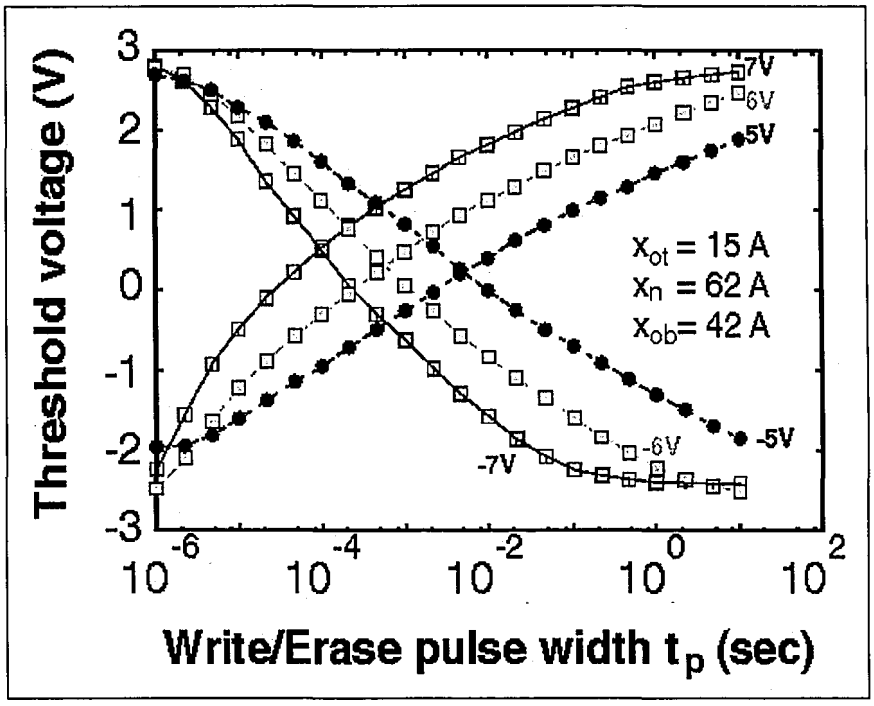


Figure 3.8 Write/Erase curves for a SONOS transistor (W = 100μm, L = 5 μm) with an ultra-thin tunnel oxide with program voltages of 7V (blue), 6V (teal), and 5V (red) [1].

Fig. 3.8 shows the result of an Erase/Write measurement on an n-channel SONOS transistor using erase and write voltages of 5, 6, 7 volts. The transistor has a virgin threshold of 0.2 V. For each point on the write curve, the device is first erased by a long Erase pulse (10s) to saturate the nitride traps with holes, and then a Write pulse with different duration, t_p, is applied to the gate of the device. In the meantime, the source, drain and bulk are tied together to ground. After a short read-delay time t_{rd}, the threshold voltage is extracted by a Read operation. The Erase measurement is performed in the

same manner, except for a reversal of the voltage polarity. The 'cross-over' point, where the Write curve meets the Erase curve, is an indication of the programming speed of the device for a specified gate voltage. As the programming voltage is increased from 5 to 7 volts, the 'cross-over' point decreases from 10 ms to 0.1 ms, a rate of 1 decade/V [1].

3.4.2 Retention Measurement

Retention measurements are used to characterize the ability of a SONOS device to store and recover data over time. As the tunnel oxide of SONOS device is scaled down, the device will have a faster programming speed and lower program and erase voltages, but data retention and endurance will be degraded.

To perform the retention measurement, the device is first written or erased. After a pre-determined read delay time, we apply a Read signal to the Erase/Write/Read circuit to read out the threshold voltage. Next, the device is reset to the same original state and the same Read operation is performed after another read-delay time. The read-delay time is between 10^{-7} seconds and 10^{5} seconds.

The retention measurement can be performed at elevated temperatures (Fig. 3.9) with a hot chuck in the laboratory. Experiments indicate the decay rate of the write state (excess electron state) is sensitive to temperature, while the decay rate of the erase state (excess hole state) remains stable. Using the data, we can analyze the nitride trap density for the write state [24-25]. We can also use the last 3 or 4 points and a straight line fit to obtain an extrapolated memory window at 10 years (3E+8 sec). The commercial standard for data retention is a 0.5 V window at 10 years, at 85 C.

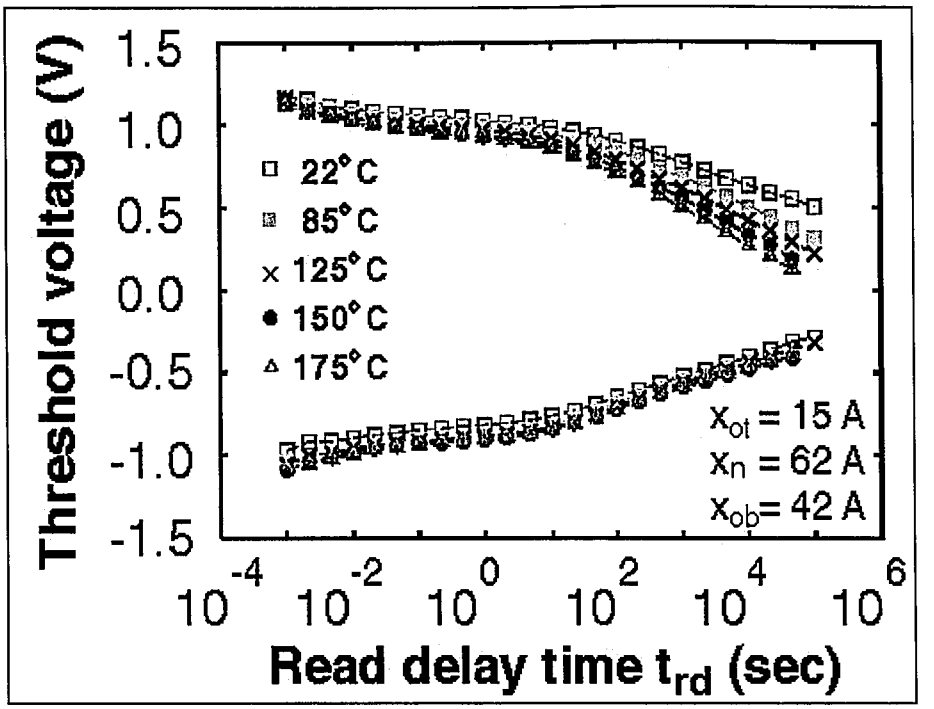
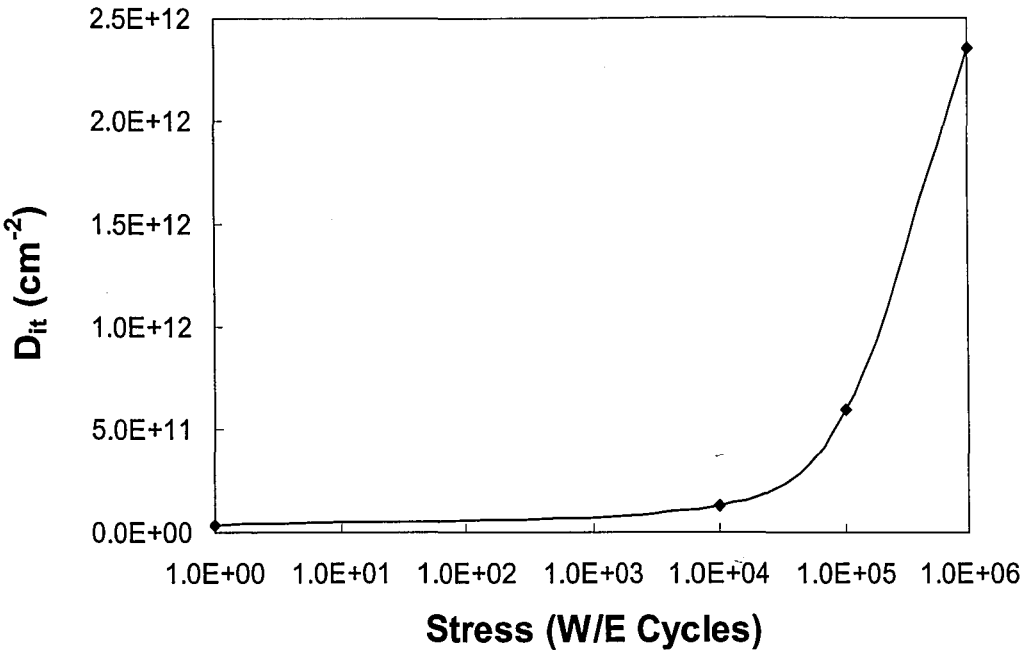


Figure 3.9 Threshold voltage of an ultra-thin tunnel oxide SONOS device versus read delay time for temperatures ranging from 22 C to 175 C [25].

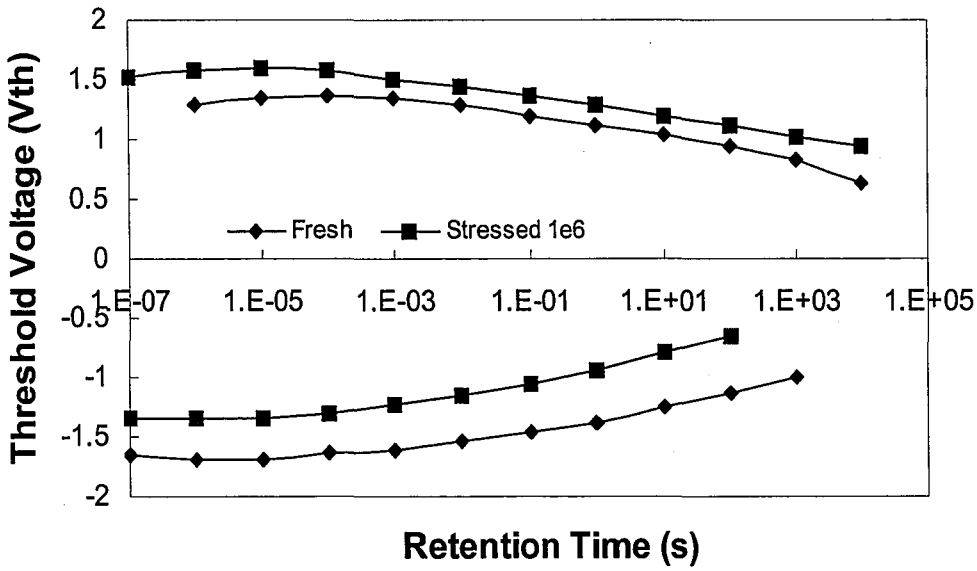
3.4.3 Endurance Measurement

Endurance is a measure of the device's ability to sustain the damage caused by repetitive Erase/Write cycling. Each time an Erase/Write operation is performed on the device, charges pass through the tunnel oxide with the possibility of deteriorating the Si-SiO_2 interface and the nitride film.

To investigate the effects of Erase/Write cycling, we measure the threshold voltage decay curve and interface trap density D_{it} for a virgin device. Next, the device is subjected to 10^4, 10^5 or 10^6 Erase/Write cycles using a write voltage of +7 V for 2.5 ms and an erase voltage of -7 V for 7.5 ms. Charge retention and charge pumping measurements are performed after each cycling. Fig. 3.10 (a) and (b) show the effect of stressing on the density of interface traps and threshold voltage, respectively.



(a)



(b)

Figure 3.10 (a) Interface trap density versus stressing cycles on a SONOS transistor. (b) Threshold Voltage Versus Time under different stressing cycles. The SONOS transistor (W=50 μm, L=0.8 μm) has 1.8 nm tunnel oxide, 10 nm nitride and 4.5 nm blocking oxide.

Chapter 4
Electrical Characterization of Scaled SONOS NVSM Devices

Electrical Characterization is critical to the success of SONOS device modeling and optimization studies. In this chapter, we present our results on the characterization of scaled SONOS devices. We begin with I-V measurement to extract device modeling parameters, such as substrate doping and flatband voltage. Using the derived parameters, we can simulate the drain current and transconductance behavior using the theoretical model described in Chapter 2. It is the foundation to understand the physics of SONOS devices. Next, basic memory characteristics of n-channel SONOS transistors with gate diodes are presented. In particular, one of the most prominent reliability issues, the degradation of data retention at elevated temperature, is discussed.

4.1 Device parameters extraction

Many key parameters that describe the behavior of SONOS transistors can be derived from the Drain Current vs. Gate Voltage curves. According to an accurate linear region model for single-device MOSFET parameter extraction in strong inversion [26], the drain current for an n-channel SONOS may be written

$$I_D = \frac{\beta_0\left[(V_{GS}-V_{th})-\frac{V_{DS}}{2}(1+\delta_n)\right]V_{DS}}{1+(\theta_s+2\beta_0 R_S)\left[V_{GS}-V_{th}+2\lambda_n\sqrt{2\phi_F+V_{SB}}-\frac{V_{DS}}{2}(1-\delta_n)\right]-\beta_0 R_S(4\lambda_n\sqrt{2\phi_F+V_{SB}}+3V_{DS}\delta_n)} \quad (4.1)$$

where

$$V_{th} = V_{FB} + 2\phi_F + \lambda_n \sqrt{2\phi_F + V_{SB}} \tag{4.2}$$

$$\delta_n = \frac{\lambda_n}{2\sqrt{2\phi_F + V_{SB}}} \tag{4.3}$$

$$\lambda_n = \frac{\sqrt{2k_s \varepsilon_0 q N_A}}{C_{eff}} \tag{4.4}$$

$$\phi_F = \frac{kT}{q} \ln\left(\frac{N_A}{n_i}\right) \tag{4.5}$$

This model includes second-order drain voltage effects [27] and a series resistance at the source and drain ($R_S = R_D$) [28]. The $Q_{it}(\phi_s)$ term is included in V_{FB} and variations with surface potential are neglected. All other constants have their usual meanings.

A family of I_D-V_{GS} curves with V_{SB} as a parameter taken at low drain bias ($V_{DS} < 2kT/q$) is shown in Fig. 4.1. The gate voltage is controlled to prevent electron tunneling. From this plot a linear extrapolation at the point of maximum slope to $I_D = 0$ gives the threshold voltage under different V_{SB}. Rearranging (4.2), we have

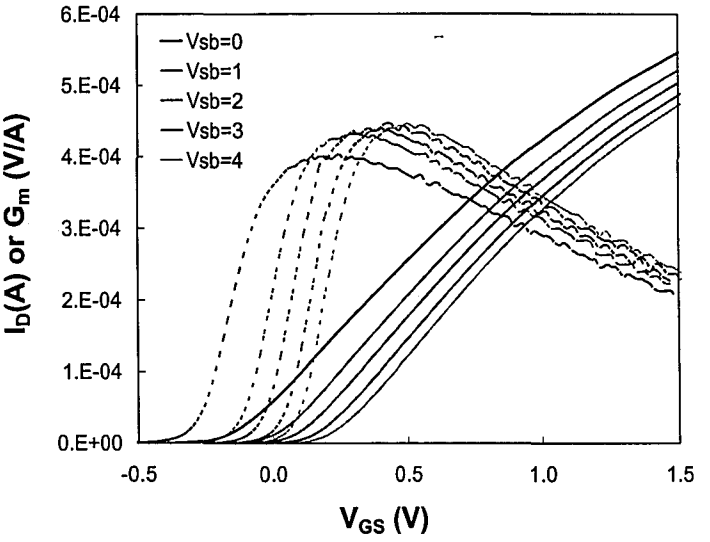
$$V_{th} = V_{t0} + \lambda_n (\sqrt{V_{SB} + 2\phi_F} - \sqrt{2\phi_F}) \tag{4.6}$$

where

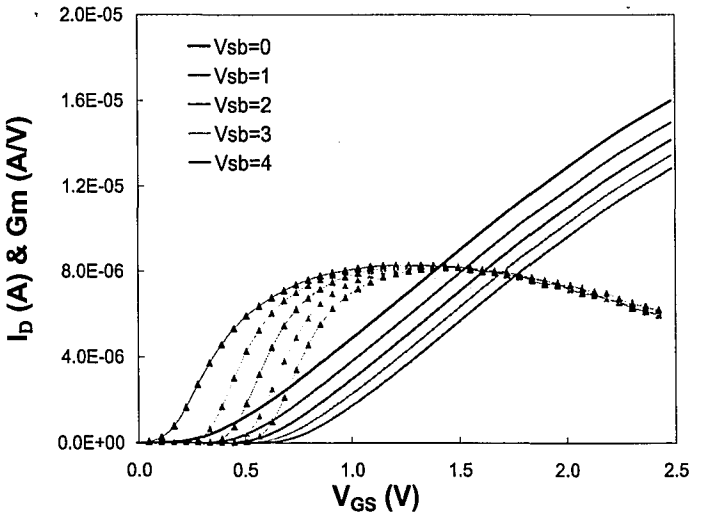
$$V_{t0} = V_{FB} + 2\phi_F + \lambda_n \sqrt{2\phi_F} \tag{4.7}$$

If we assume an initial value of N_A with a known value of effective capacitance C_{eff}, then we can plot V_{th} vs. $(\sqrt{V_{SB} + 2\phi_F} - \sqrt{2\phi_F})$ as shown in Fig. 4.2 (a). The body-effect parameter (λ_n) is determined by the slope and the threshold voltage (V_{to}) is determined by the y-axis intercept. The corrected value of doping concentration (N_A^*) can be

calculated using the derived value of λ_n according to Eqn. (4.4). This whole process is repeated until the assumed value for (N_A) is equivalent to the actual value (N_A^*). The final results, obtained by iteration, for λ_n and V_{to} yield the flatband voltage V_{FB} and bulk doping concentration N_A.



(a)



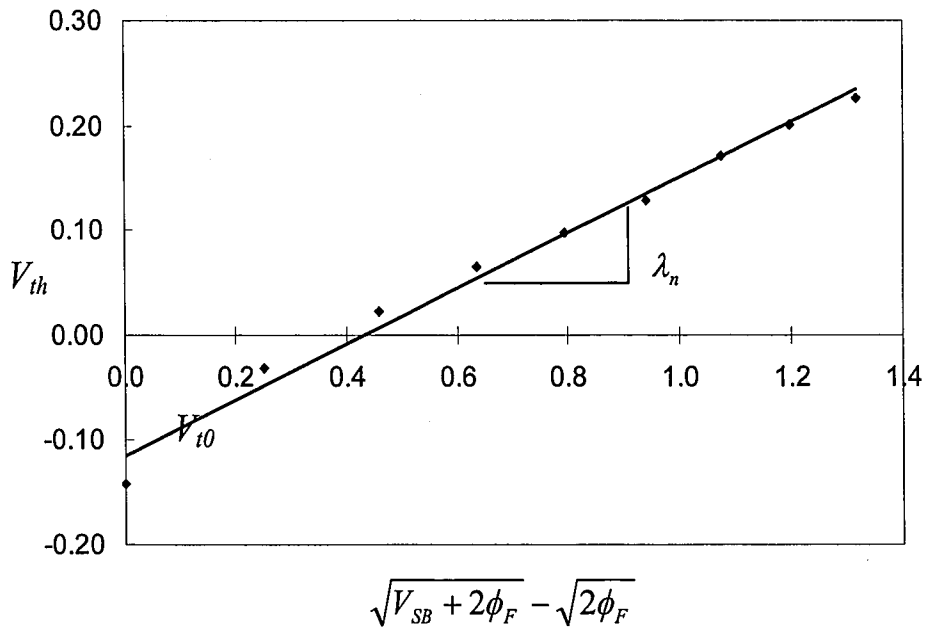
(b)

Figure 4.1 Drain current and transconductance vs. gate voltage with stepped reverse bias on the bulk (V_{DS} = 50 mV). (a) Single SONOS transistor with dimension of W=50 μm, L=0.8 μm. (b) Single SONOS transistor with dimension of W=50 μm, L=50 μm. Both devices have effective capacitance, C_{eff}=2.8E-7 fd/cm^2.

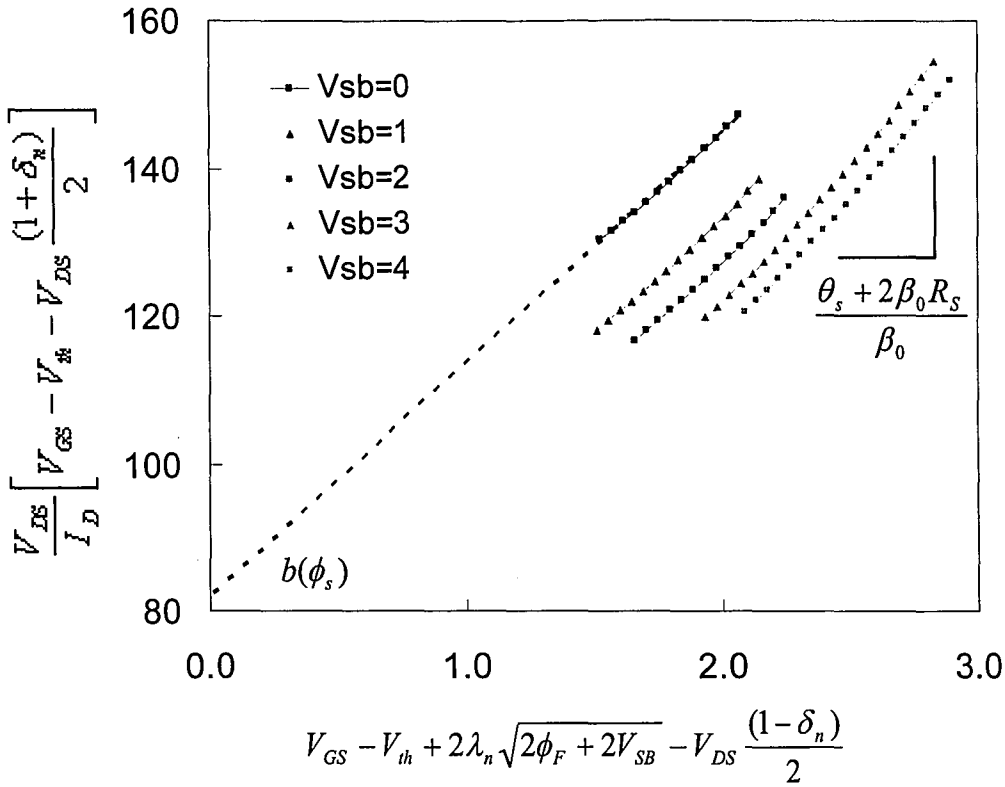
The parameters R_S, θ_s, and β_0 can be determined through two linear extractions as follows: First, plot the graph denoted in Fig. 4.2 (b) to obtain a slope of $(\theta_s + 2\beta_0 R_S)/\beta_0$ and a series of intercepts $b(\phi_s)$ defined by

$$b(\phi_s) = \frac{1}{\beta_0} - R_S(4\lambda_n\sqrt{2\phi_F + V_{SB}} + 3V_{DS}\delta_n) \tag{4.8}$$

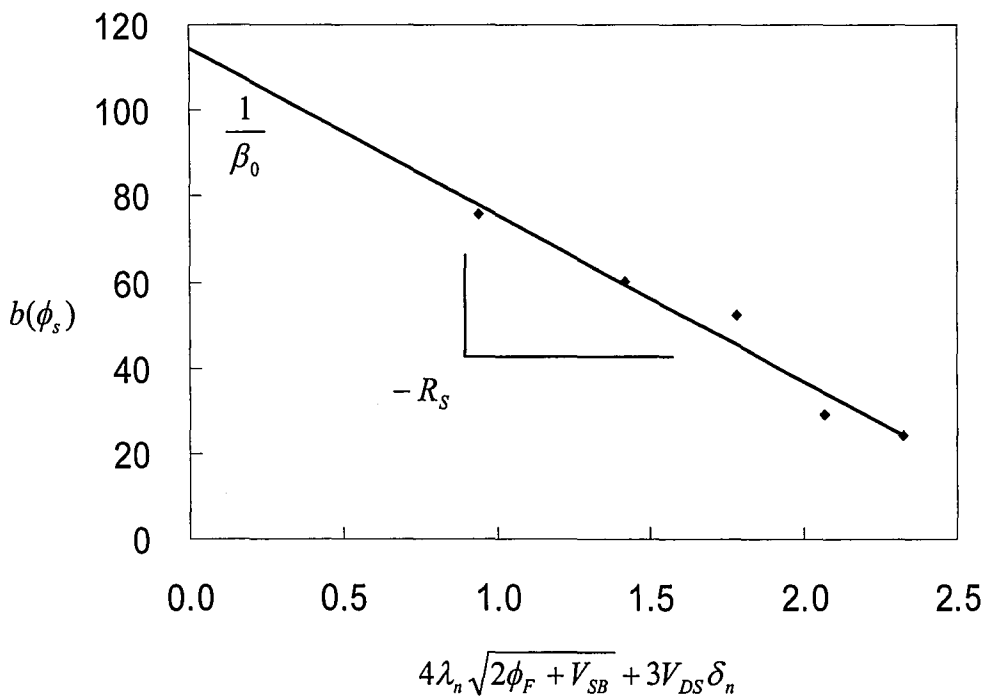
Then we can plot $b(\phi_s)$ versus $4\lambda_n\sqrt{2\phi_F + V_{SB}} + 3V_{DS}\delta_n$ as in Fig. 4.2 (c) and find $1/\beta_0$ as an intercept and R_S as the slope. These values can be inserted into the slope of Fig. 4.2 (b) to determine the surface roughness factor (θ_s).



(a)



(b)



(c)

Figure 4.2 Procedures for the determination of SONOS device modeling parameters.

We use this method to extract device parameters for both short channel and long channel SONOS devices. In short channel devices, we have observed a gradual increase in the peak transconductance with increase of substrate bias. This effect can be explained by channel length modulation and charge sharing effects, which requires a 2-D description of the current flow. Below is the listing of the parameters we extracted:

	N_A (cm^{-3})	Cox (fd/cm^2)	Φ_f (V)	λ_n ($\sqrt{V}$)	V_{FB} (V)	β_0 (A/V^2)	Rs (Ω)
SONOS Transistor 1 (W=50μm, L=50 μm)	2.E+16	2.8E-07	0.38	0.3	-0.61	2.1E-4	270
SONOS Transistor 2 (W=50μm, L=0.8μm)	2.E+16	2.8E-07	0.37	0.27	-1.09	8.7E-3	38

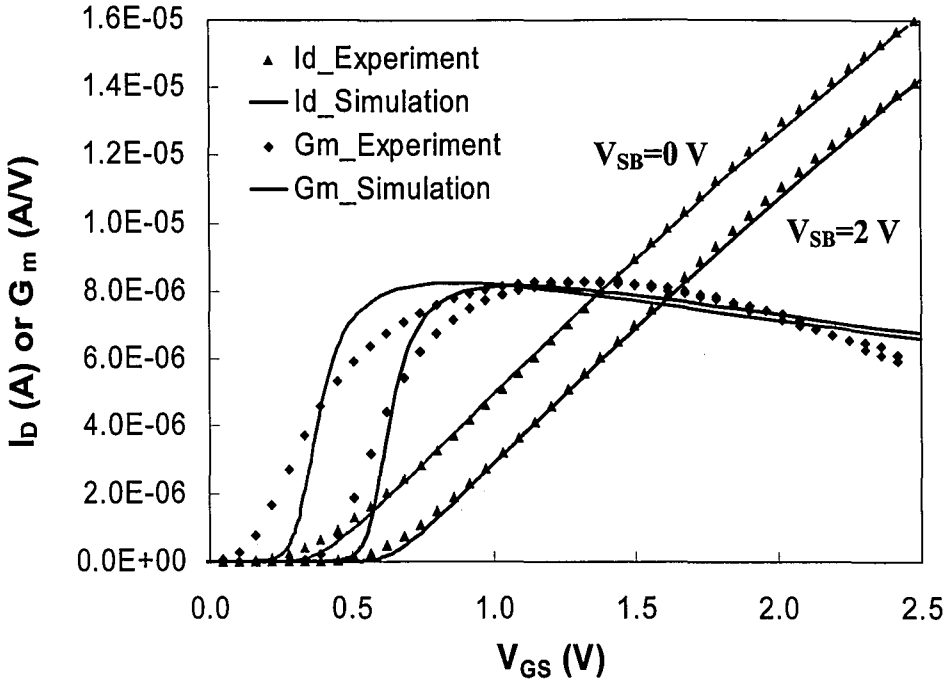
Table 4.1 Extracted parameters for SONOS single transistors.

From our observations, the series resistance plays a more important role than the surface roughness parameter in the deterioration of the drain current with gate and substrate bias.

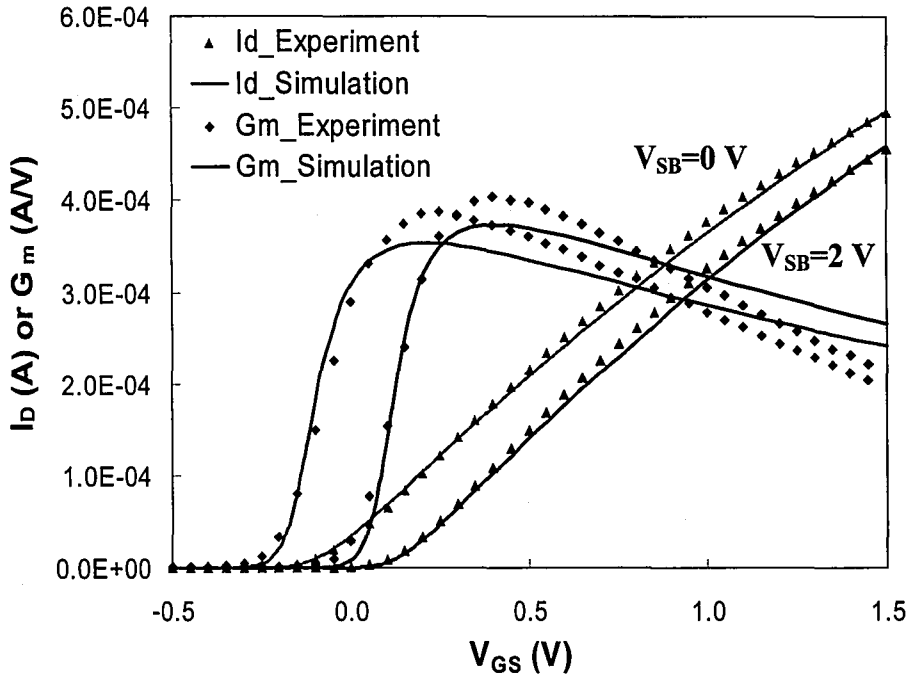
4.2 Simulation Result

The parameters are employed to generate theoretical I_D and g_m curves. The mean density of interface traps (D_{it}) is determined by charge pumping measurement and has a typical value of 1.1×10^{10} $cm^{-2}eV^{-1}$. The effect of series resistance (R_S) can be modeled by an effective θ_s , where $\theta_{s(effective)} = \theta_s + 2\beta_0 R_S$ contributes to the degradation of the transconductance [28].

Fig. 4.3 shows the comparison between experiment and simulation results on the drain current and transconductance behavior for both short channel and long channel NSONOS devices.



(a)



(b)

Figure 4.3 Comparison between experiment and simulation result for (a) long channel NSONOS device (W=50 μm, L=50 μm), (b) short channel NSONOS device (W=50 μm, L=0.8 μm). Both devices have 1.8nm tunnel oxide, 10nm nitride storage layer and 4.5nm blocking oxide.

As shown in Fig. 4.3, the simulated drain current curves lie close to the experimental ones from the subthreshold region to the inversion region. As the transconductance represents the derivative of the drain current, modeling of transconductance is more difficult since the transconductance is very sensitive to the slightest shift in the horizontal axis, which can take place when charge injection occurs in the silicon nitride. For some reason, in the long channel SONOS devices, the Gm measurement results are not ideal and different from the simulation results. We didn't see the same behavior in our previous devices and we will figure it out later.

4.3 SONOS device with protective gate diodes

The step-by-step process of SONOS device fabrication at the Sherman Fairchild Microelectronic Lab is summarized in Appendix C. The polysilicon etch is one of the most important operations in semiconductor device fabrication, especially for state-of-the-art small geometry devices.

Plasma process is conventionally used to etch polysilicon, which results in less undercutting as compared with a chemical etch process. However, it has been widely recognized that plasma processes cause current to flow through thin oxides, resulting in charge trapping in the oxide and trap generation at the Si-SiO_2 interface. Although many of these traps can be passivated by a forming gas anneal, latent damage exists [29]. In addition, some of the traps can be reactivated by the subsequent electrical stressing. Plasma-induced damage enhances the vulnerability of gate oxide to hot-carrier-induced degradation and time-dependent dielectric breakdown (TDDB) [30].

In the SONOS device area, it has been suggested that plasma etching causes a reduction in device retention because the Si-SiO_2 interface traps act as the stepping stones in charge decay process [17]. In order to improve the SONOS device performance, researchers at Northrop Grumman and Sandia National Laboratories design a SONOS structure with two diodes connected back to back and to the gate of the device (Fig. 4.4). The protective gate diodes can prevent charge accumulation and improve device performance. From our measurements, it is clear that both the retention characteristic and interface trap density have improved compared with devices without protective gate diodes. Fig. 4.5 shows the result. These SONOS devices are fabricated with a gate dielectric consisting of an 18 Å tunneling oxide, a 100 Å "oxynitride" layer and a 45 Å blocking oxide underneath a phosphorus-doped polysilicon gate. We program the devices by applying +7 V to the gate terminal for 2.5 msec, and erase them by applying -7 V pulse for 7.5 msec.

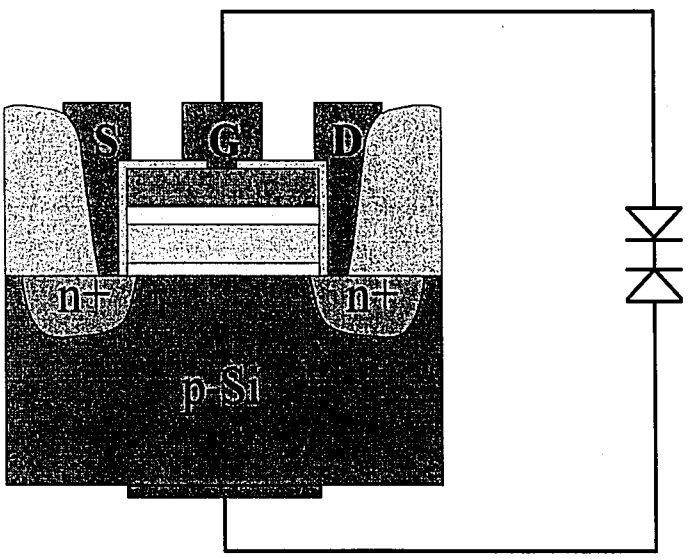


Figure 4.4 SONOS structure with protective gate diodes

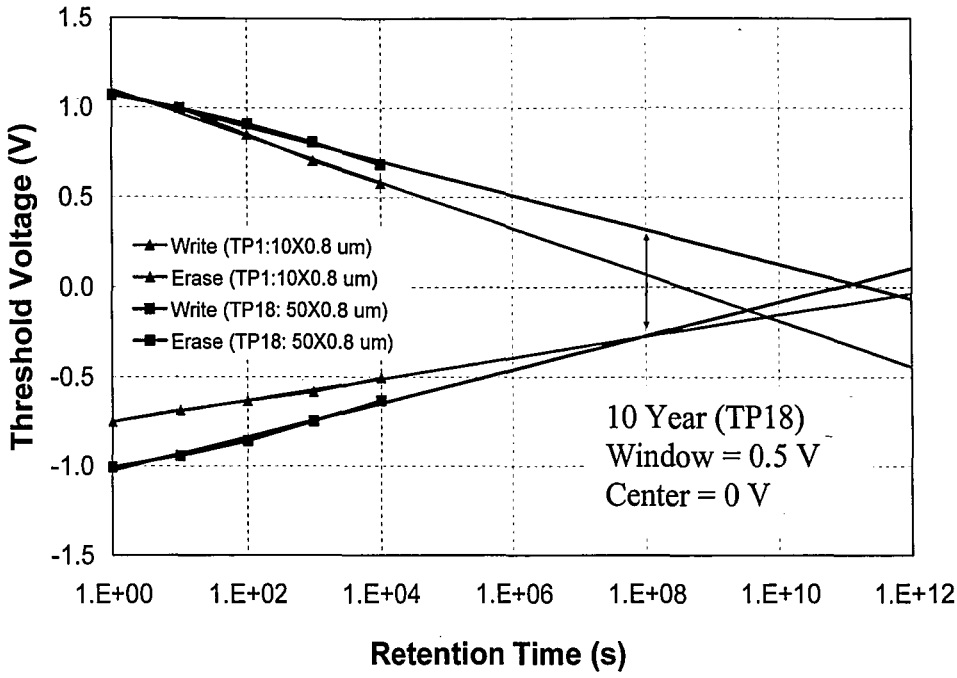


Figure 4.5 (a) Single SONOS Memory Transistor Retention --10 Year Extrapolation for TP1 (without gated diodes) and TP18 (with gated diodes) at Room Temperature.

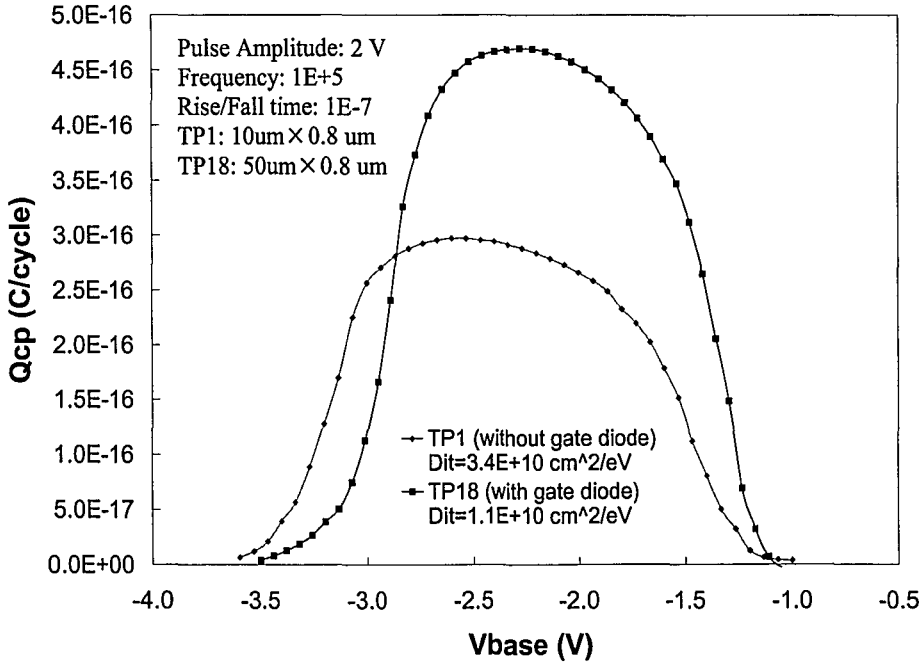


Figure 4.5 (b) Single SONOS Memory Transistor Charge Pumping Comparison -- TP1 (without gate diodes) has an increased Interface Trap Density, D_{it}, as compared with TP18 (with gate diodes) due to Plasma Damage.

4.4 SONON Retention at Elevated Temperatures

To study the charge retention at elevated temperatures, we subjected a single SONOS device to temperatures ranging from 22 degrees Celsius to 225 degrees Celsius while performing the retention measurement for both the write and erase state. From the plotted result (Fig. 4.5), we can see that the elevated temperatures have a large effect on the write state of the device and barely a noticeable effect on the erase state. The reason for this phenomenon is that the electrons are trapped in shallow energy wells and, as the temperature goes up, the trapped electrons in the nitride can escape the traps and back tunnel through an ultra-thin oxide to the silicon. The holes, on the other hand, are trapped in deep energy wells and cannot escape by thermal excitation.

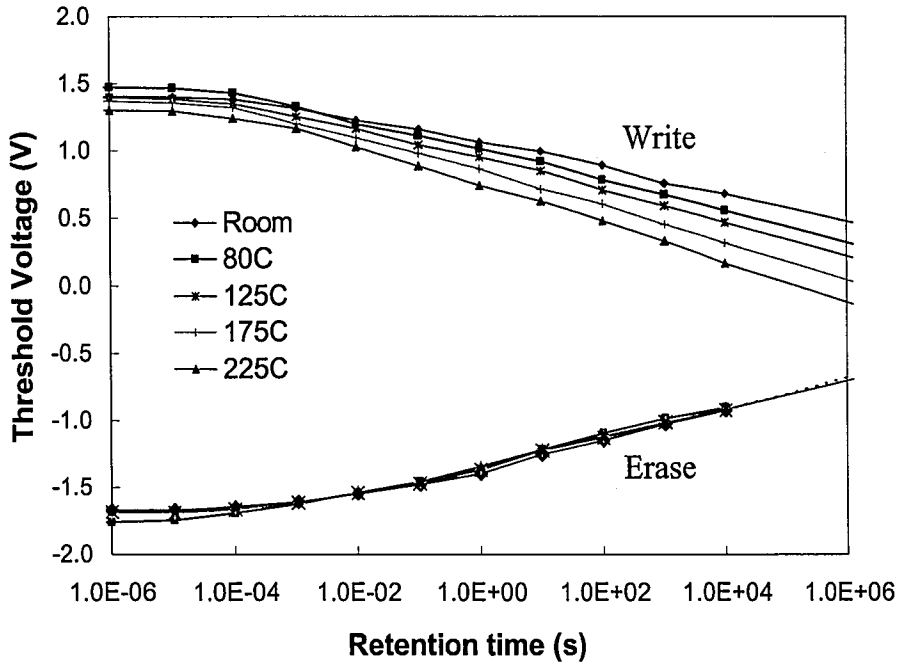


Figure 4.6 Threshold Voltage decay vs. Temperatures for a scaled SONOS device. The SONOS transistor (W=50 μm, L=0.8 μm) has 1.8 nm tunnel oxide, 10 nm nitride layer and 4.5 nm blocking oxide.

The characterization of charge retention in a SONOS device at elevated temperatures has been investigated by Yang and White [25] using an amphoteric trap model, which attributes both electron and hole traps to a silicon-dangling bond. Using this trap model, Yang and White derived an expression for the charge trapped in the nitride,

$$\rho_n(x, E_{TA}, t) = -qg(x, E_{TA})f^- \tag{4.9}$$

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 [25],

$$\frac{\partial \Delta V_{th}}{\partial \log(t)} = -2.3k_B T X_N \left(\frac{X_N}{2\varepsilon_N} + \frac{X_{OB}}{\varepsilon_{OX}} \right) g(E_{TA}) \tag{4.10}$$

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. Equation (4.10) assumes a uniform distribution of nitride traps and the activation energy responsible for the decay rate is [25]:

$$E_{TA} = k_B T \ln(AT^2 t) \tag{4.11}$$

where A is a constant given by

$$A = 2\sigma_n \sqrt{\frac{3k_B}{m^*}} \left[\frac{2\pi m^* k_B}{h^2} \right]^{3/2} \tag{4.12}$$

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 were used to interpret measurements of retention at elevated temperatures for retention times out to 10^4 seconds for the written

state of a scaled SONOS device (Fig. 4.6). For example, according to Eqn. (4.11), at a given temperature 175 C, the read delay time ranges from 1 millisecond to 10^4 seconds, corresponding to trap energy level from 0.84 eV to 2.08 eV. The trap energy depth is measured from the conduction band edge. Eqn. (4.10) shows the trap density can be probed at the same temperature by measuring the slope of the threshold voltage decay characteristics as a function of retention time. The nitride trap density vs. the trap energy at different temperatures is plotted in Fig. 4.7.

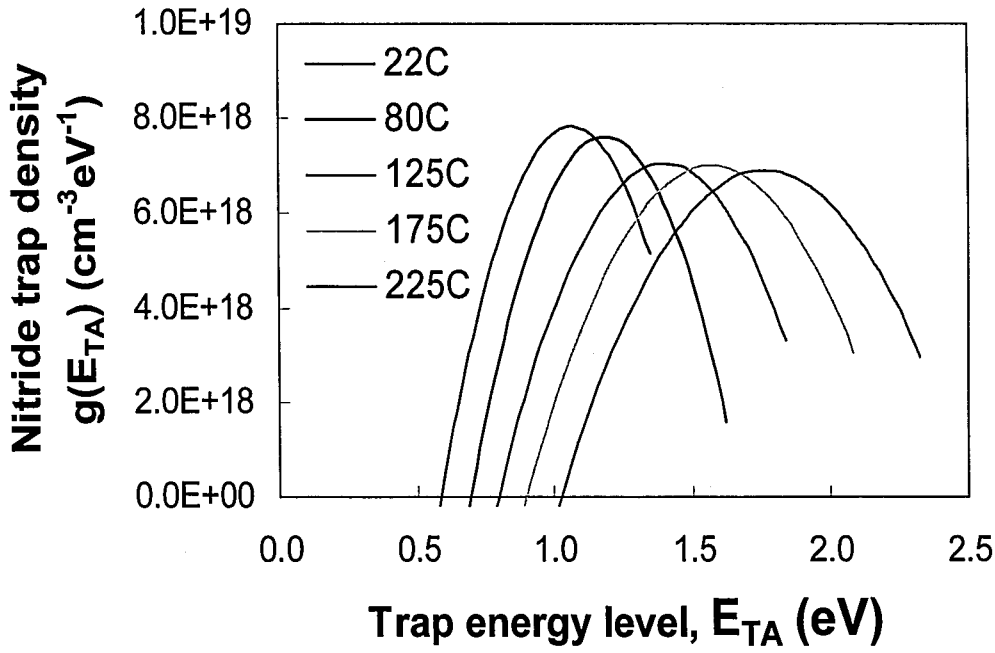


Figure 4.7 Extracted value of the nitride trap density, $g(E_{TA})$, vs. trap energy, E_{TA}, at different temperatures for a SONOS transistor with Oxynitride.

The SONOS devices we measured are consisting of oxygen-rich films called oxynitride (SiO_xN_y). In previous studies of elevated temperature affects, Yang and White [25] have extrapolated the electron trap density for a SONOS device with a 'silicon-rich' nitride layer at 175 C. The work of Wrazien and Zhao *et al.* [24] shows the trap density

for a silicon-rich nitride layer is roughly three times greater than the density for an 'oxynitride' layer as shown in a comparison chart of Fig. 4.8. This is expected as the presence of oxygen will 'tie-up' silicon dangling bonds, which cause the memory traps in the nitride. The 'oxynitride' and 'silicon-rich' nitride trap densities peak at the same activation energy, 1.1 eV.

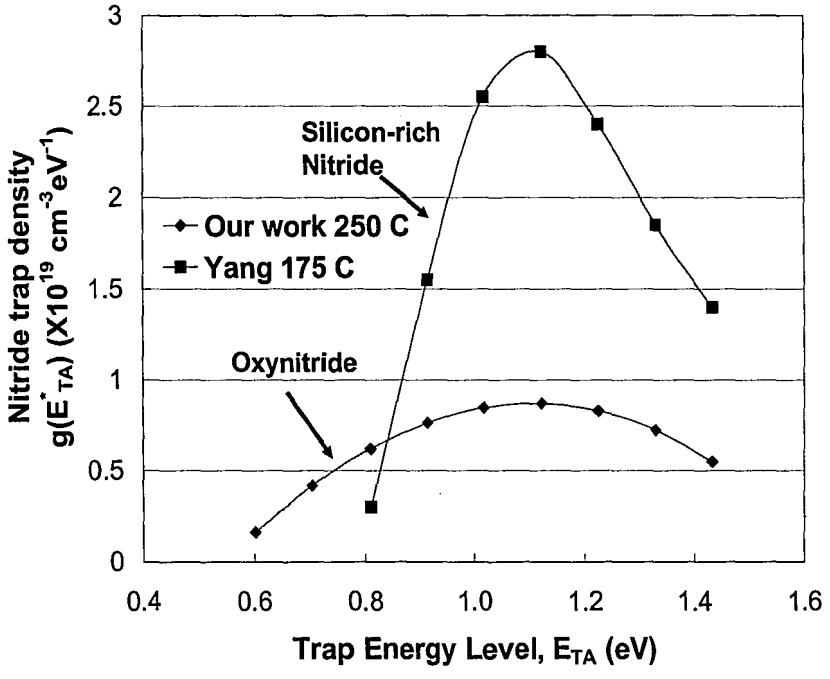


Figure 4.8 A Comparison of Trap Densities in 'silicon-rich' and 'oxynitride' SONOS NVSM devices [24].

Chapter 5

Conclusions

This thesis addresses the electrical characterization of scaled SONOS nonvolatile memory devices. In this chapter, I will summarize our results on device modeling and experimental observations. Also include a brief description of future investigation.

5.1 Conclusions from this Thesis

- A parameter extraction procedure based on physical concepts is applied to determine SONOS device constants and model transconductance degradation for a uniform doping profile. Taking into account the effect of R_S, the intrinsic value of β_0 and θ_S can be extracted from a single dc I_D-V_{GS} measurement. In our devices, series resistance (R_S) rather than surface roughness scattering (θ_S) has a pronounced effect on the degradation of drain current and transconductance. Important device parameters extracted for both short channel and long channel SONOS devices are listed in Table 4.1.

- With the extracted parameters, we can simulate the drain current and transconductance behavior of SONOS devices in program and erase states with the modification of a model used for MOSFET devices (Fig. 4.3). The simulated drain current curves lie close to the experimental ones from the subthreshold region to the inversion region. As the transconductance represents the derivative

of the drain current, modeling of transconductance is more difficult since the transconductance is very sensitive to the slightest shift in the horizontal axis, which can take place when charge injection occurs in the silicon nitride. For some reason, in the long channel SONOS devices, the Gm measurement results are not ideal and different from the simulation results. We didn't see the same behavior in our previous devices and we will figure it out later.

- Plasma damage has been observed in SONOS devices fabricated at Northrop Grumman. During the plasma etching process, electrons build up in the gate of the device, which results in charge trapping in the gate dielectric and a trap generation at the Si-SiO_2 interface. SONOS structure with two diodes connected back to back and to the gate of the device (the protective gate diode configuration shown in Fig. 4.4) can be used to prevent charge accumulation and improve device performance. According to our measurements, both the retention characteristic and interface trap density are improved compared with SONOS devices without protective gate diodes (Fig. 4.5).

- At elevated temperatures, the write (program) state of the scaled N-channel SONOS memory device shows thermal emission of electrons (Fig. 4.6). As the temperature increases, electrons are excited and free from traps in the nitride and tunnel back to Si, causing an accelerated loss of charge at high temperatures. Assuming thermal excitation of trapped electrons is the dominant discharging mechanism at high temperatures, the nitride trap density vs. the trap energy at

different temperatures can be extracted (Fig. 4.7). The result shows that the trap density for an 'oxynitride' storage layer is roughly three times less than the density for a silicon-rich nitride layer (Fig. 4.8). The excess 'hole' state does not show appreciable loss of charge with temperature, which seems to indicate the 'hole' trap levels are located deeper in the silicon nitride energy gap than their electron counterparts so thermal emission is not observed.

- With different stressing cycles, an increase in Si-SiO_2 interface trap density (D_{it}) is observed. The threshold voltage decay rate of excess hole state is degraded with cycling, while the decay rate of the excess electron state is not affected (Fig. 3.10). This indicates the excess hole state is more sensitive to the trap-to-trap tunneling, which is associated with Si-SiO_2 interface traps. We have also observed both the program and erase curves shift in the positive direction because of the increase of interface charges.

5.2 Future Considerations

- Oxynitride and silicon-rich nitride (SRN) can be further studies in terms of retention. They can be deposited in sequence on top of tunnel oxide with LPCVD to provide a charge storage layer. This dielectric combination can increase charge back-tunneling distance and have improved charge retention.

- Materials with high dielectric constants (high-K materials) can be used as a charge storage layer. They are electrically thinner than nitride, which is advantageous for low-voltage operation.

- SONOS devices, called NROM devices, use hot-carrier injection to program and store charge locally over source and drain junctions. This NROM technology can double the memory density of the conventional SONOS device [31] and needs to be further investigated. A thicker tunnel oxide is employed in NROM devices to increase memory retention comparable to the conventional floating gate technology.

References

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

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

[3] S. Tiwari, F. Rana, K. Chan, H. Hanafi, W. Chan, and Doug Buchanan, "Volatile and non-Volatile memories in silicon with nano-crystal storage," IEDM Tech. Dig., pp. 521-524 Dec. 1995.

[4] S. Lai, "Tunnel Oxide and ETOX Flash Scaling Limitation," IEEE International Non-Volatile Memory Conference, 1998.

[5] H. Wann and C. Hu, "High-endurance ultra-thin tunnel oxide in MONOS device structure for dynamic memory application," IEEE Electron Device Lett., vol. 16, p. 491, 1995.

[6] Y. Yang, A. Purwar, M. H. White, "Reliability Considerations in Scaled SONOS Nonvolatile Memory Devices," Solid-State Electronics, Vol. 43, pg. 2025-2032, 1999.

[7] H. A. R. Wegener, A. J. Lincoln, H. C. Pao, M. R. O'Connel, and R. E. Oleksiak, "The variable threshold transistor, a new electrically alterable, non-destructive read-only storage device," IEEE IEDM Tech. Dig., p. 70, 1967.

[8] A. Thean and J. Leburton, "Flash memory: towards single-electronics," IEEE potentials, Nov. 2002

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

[10] H. Wong, M. H. White, T. J. Krutsick and R. V. Booth, "Modeling of transconductance degradation and extraction of threshold voltage in thin oxide MOSFET's", Solid-State Electronics, Vol. 30, pg. 953-968, 1987.

[11] H. C. Pao and C. T. Sah, "Effects of diffusion current on characteristics of metal-oxide (insulator)-semiconductor transistors", Solid-State Electronics, Vol. 9, pg. 927-937, 1966.

[12] F. Stern, "Self-consistent results for n-type Si inversion layers", Physical Review B, Vol. 5, pg. 4891-4899, 1972.

[13] C. T. Sah, T. H. Ning and L. L. Tschopp, "The scattering of electrons by surface oxide charges and by lattice vibrations at the silicon-silicon dioxide interface", Surface Science, Vol. 32, pg. 561-575, 1972.

[14] S. C. Sun and J. D. Plummer, "Electron Mobility in Inversion and Accumulation layers on Thermally Oxidized Silicon Surface", IEEE Trans. Electron Dev., Vol. ED-27, pg. 1497-1508, 1980.

[15] C. C. Chao and M. H. White, "Characterization of Charge injection and trapping in Scaled SONOS/MONOS Memory Devices", Solid-State Electronics, Vol. 30, pg. 307-319, 1987.

[16] 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, pg. 1083-1089, 1991.

[17] Y. Hu and M. H. White, "Charge retention in scaled SONOS nonvolatile semiconductor memory devices – Modeling and characterization", Solid-State Electronics, Vol. 36, pg. 1401-1416, 1993.

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

[19] R. E. Paulsen, M. H. White, "Theory and Application of Charge Pumping for the Characterization of Si-SiO_2 Interface and Near-Interface Oxide Traps," IEEE Trans. Electron Devices, vol. 41, no. 7, 1994.

[20] N. E. Cohen, R. E. Paulsen, M. H. White, "Observation and Characterization of Near-Interface Oxide Traps with C-V Techniques," IEEE Trans. Electron Devices, vol. 42, no. 11, 2004-2009, 1995.

[21] A. Roy, "Retention endurance and interface traps in MONOS memory transistors," Master's thesis, Lehigh University, 1985.

[22] A. K. Banerjee, "Automated measurements of memory devices," Master's thesis, Lehigh University, 1993.

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

[24] S. J. Wrazien, Y. Zhao, J. D. Krayer, M. H. White, "Characterization of SONOS oxynitride nonvolatile semiconductor memory devices," Solid-State Electronics, Vol.47, pg. 885-891, 2003

[25] Y. Yang and M. H. White, "Charge Retention of Scaled SONOS Nonvolatile Memory Devices at Elevated Temperatures", Solid-State Electronics, Vol. 44, pg. 949-958, 2000.

[26] T. J. Krutsick, M. H. White, H. S. Wong and R. V. H. Booth, "An Improved Method of MOSFET Modeling and Parameter Extraction", IEEE Trans. Electron Devices, vol. 34, no. 8, 1676-1680, 1987.

[27] F. M. Klaassen, "A MOS model for computer-aided design", Philips Res. Rep., vol.31, pp. 71-83, 1976

[28] J. Greenfield, R. Dutton, E. Demoulin, P. Chatterjee, and A. Tasch, "Process statistics of submicron MOSFET's", in IEDM Tech. Dig., 1979.

[29] H. Shin, N. Jha, X. Qian, G. W. Hills and C. Hu, "Plasma etching charge-up damage to thin oxides", Solid State Technology, August, 29-36, 1993

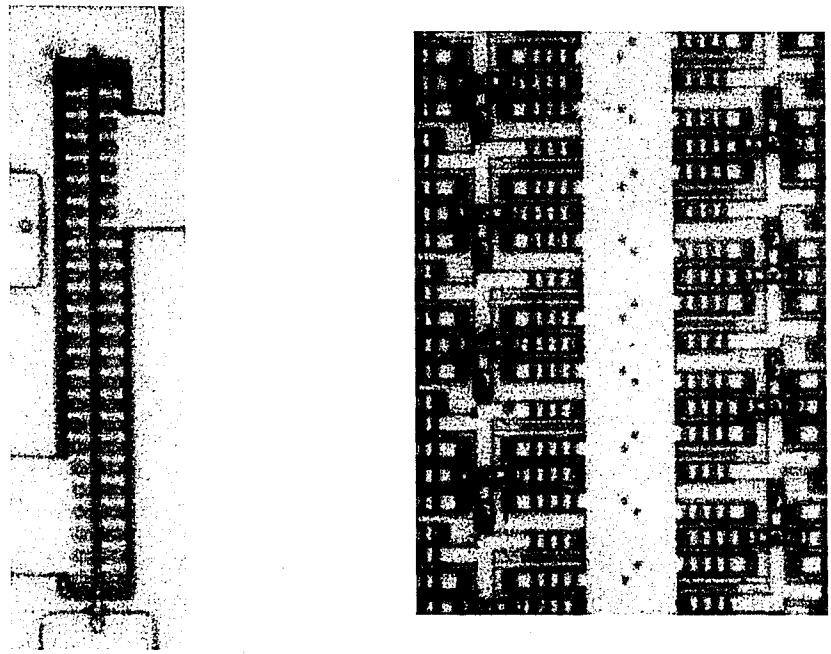
[30] S. Fang and J. McVittie, "Thin-oxide damage from gate charging during plasma processing", IEEE Elec. Dev. Lett., 13(5), pp. 288-301, 1992.

[31] I. Bloom, P. Pavan and B. Eitan, "NROM™ – a new technology for non-volatile memory products", Solid-State Electronics, Vol. 46, pg. 1757-1763, 2002.

[32] Y. Maneglia and D. Bauza, "Extraction of slow oxide trap concentration profiles in metal-oxide-semiconductor transistors using the charge pumping method", J. Appl. Phys., 79(8), pg. 4187-4192, 1996.

[33] D. Bauza and Y. Maneglia, "In-depth exploration of Si-SiO_2 interface traps in MOS transistors using the charge pumping technique", IEEE Trans. Electron Devices, vol. 44, no. 12, 2262-2266, 1997.

Appendix A – SONOS device pictures



(a) (b)

Figure A.1 (a) Single SONOS transistor of W=50 μm, L=0.8 μm. (b) SONOS Array consisting of 1000 single devices with each device having W=5 μm, L=0.8 μm.

Appendix B - Measurement Procedures (step by step)

B. 1 Drain Current versus Gate Voltage Measurement

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

1) Setup

A) Verify that the computer is connected to the 4145 scope and the Keithley 7001 Switch System using the GPIB cables. Place the wafer onto the probing station.

B) Locate a single SONOS device.

C) Make contact with the device at pads indicated in the table.

Table B.1 – Drain Current versus Gate Voltage Measurement Pin Setup

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

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

E) 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.

F) 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.

G) 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 "\\Calvin\pcdir\labview\programs\SONOS".

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 μm and L = 0.8 μm.

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

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 ""\\Calvin\pcdir\labview\programs\SONOS".

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 Erase/Write Cycling Measurement

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 - SONOS Device Fabrication – (Step-by-Step)

C. 1 Starting Material

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

C. 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

C. 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)

C. 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)

C. 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

C. 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)

C. 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

C. 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)

C. 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

C. 10 Post-Metal Anneal (PMA)

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

Vitae

Yijie Zhao was born on July 24th, 1977 in Shanghai, People's Republic of China to Mr. Zhaolin Zhao and Mrs. Weiwen Lin.

She attended Shanghai Jiaotong University, P. R. China from September 1995 to June 1999 and graduated with a Bachelor's Degree in Electrical Engineering. From July 1999 to July 2001, she worked as a hardware engineer at Lucent Technologies of Shanghai. In August 2001, she enrolled as a full time student at Lehigh University to pursue a M.S. Degree in Electrical Engineering followed by a PhD degree. During her undergraduate and graduate studies, she has been a recipient of various fellowships including a University Fellowship from September 1995 to June 1999, a Graduate School Fellowship from August 2001 to May 2002 and a Sherman Fairchild Fellowship from August 2002 to the present time.

She is an IEEE student member of the Electron Device Society and an associate member of the Scientific Research Society of Sigma Xi.

Publications

S. J. Wrazien, Y. Zhao, J. D. Krayer, M. H. White, "Characterization of SONOS oxynitride nonvolatile semiconductor memory devices," Solid-State Electronics, Vol.47, pg. 885-891, 2003

END OF TITLE