Spring 5-31-2004

Characterization of ultrathin gate dielectrics and multilayer charge injection barriers

Edwin M. Dons
New Jersey Institute of Technology

ABSTRACT

CHARACTERIZATION OF ULTRATHIN GATE DIELECTRICS AND MULTILAYER CHARGE INJECTION BARRIERS

by

Edwin M. Dons

Since the invention of the first integrated circuit, the semiconductor industry has distinguished itself by a phenomenally rapid pace of improvements in device performance. This trend of ever smaller and faster devices is a result of the ability to exponentially reduce feature sizes of integrated circuits, a trend commonly known as "scaling". A reduction of overall feature sizes requires a simultaneous reduction in the thickness of the gate dielectric, SiO_2, of a MOSFET. Gate oxides in the ultrathin regime (<35 Å) feature a large direct tunneling leakage current. The presence of this leakage current requires a reevaluation of standard characterization techniques as well as a reevaluation of the continued usefulness of SiO_2 as the gate dielectric of choice for future applications. On the other hand, a thorough understanding of the dynamics of ultrathin oxides opens up a range of future device applications that were not possible with thicker oxides.

Capacitance-voltage characterization has been the standard technique to study the electrical properties and interface quality of MOS devices. However, the presence of a large leakage current in ultrathin oxides distorts standard C-V measurements, rendering this technique no longer useful. In this work, a leakage compensated charge measurement is developed to overcome this difficulty. This technique produces static C-V curves, even for oxides as thin as 24 Å, thereby permitting C-V characterization well into the direct tunneling regime.

As an extension of this leakage problem, the usefulness of SiO_2 as the gate dielectric of choice for future CMOS devices has been called into question. One solution – but not the only – calls for a new dielectric to replace SiO_2 for future gate applications. This research presents some of the earliest results ever on the electrical properties of MOCVD and ALCVD hafnium oxides as a potential candidate. Electrical characterization revealed that the devices have characteristics such as large leakage currents, dielectric charging under stress, hysteresis and a large flatband voltage shift that is commonly found in materials such as the one that was investigated in this work.

As one example of future device applications that become possible due to the scaling of ultrathin oxides, silicon-based multilayer charge injection barriers have been investigated. These barriers consist of alternating layers of ultrathin SiO_2 and Si. The electrical properties of these structures were studied in detail and revealed that they can be used as an active tunnel dielectric in nonvolatile memory devices.

CHARACTERIZATION OF ULTRATHIN GATE DIELECTRICS AND MULTILAYER CHARGE INJECTION BARRIERS

by
Edwin M. Dons

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology and
Rutgers, The State University of New Jersey-Newark,
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Applied Physics

Federated Physics Department

May 2004

APPROVAL PAGE

CHARACTERIZATION OF ULTRATHIN GATE DIELECTRICS AND MULTILAYER CHARGE INJECTION BARRIERS

Edwin M. Dons

Dr. K. R. Farmer II, Dissertation Advisor — Date
Associate Professor of Physics, New Jersey Institute of Technology
Chief Technology Officer, TSI Incorporated

Dr. K. K. Chin, Committee Member — Date
Professor of Physics, New Jersey Institute of Technology
Director, NJIT-Rutgers (Newark) Joint Applied Physics M.S./Ph.D Program

Dr. G. S. Tompa, Committee Member — Date
President, Structured Materials Industries, Inc.

Dr. Z. Wu, Committee Member — Date
Associate Professor of Physics and Department Chair, Rutgers University

Dr. Z. Xiao, Committee Member — Date
Assistant Research Professor, New Jersey Institute of Technology

BIOGRAPHICAL SKETCH

Author: Edwin M Dons

Degree: Doctor of Philosophy

Date: May 2004

Undergraduate and Graduate Education:

- Doctor of Philosophy in Applied Physics
  New Jersey Institute of Technology, Newark, NJ, 2004

- Bachelor of Science in Applied Physics
  University of Utrecht, Utrecht, The Netherlands, 1994

Major: Applied Physics

Presentations and Publications:

E.M. Dons, C.S. Skowronski and K.R. Farmer,
"Electrical characterization of a double barrier direct tunneling diode structure,"
Proceedings of the Materials Research Society Symposium, San Fransisco, CA, May 1999.

E. M. Dons, C. S. Skowronski and K. R. Farmer,
"Direct tunneling diode structure with a multilayer charge injection barrier,"
Applied Physics Letters **73**, 3712 (1998).

Hui Song, Edwin Dons, Xi Qing Sun and K.R. Farmer,
"Leakage compensated charge method for determining static C-V characteristics of ultrathin MOS capacitors,"
AIP Conference Proceedings, Gaithersburg, MD, October 1998.

A. Weber, R. Nikulski, C.P. Klages, M.E. Gross, W.L. Brown, E. Dons, R.M. Charatan,
"Low Temperature Deposition of TiN Using Tetrakis(dimethylamido)-Titanium in an Electron Cyclotron Resonance Plasma Process"
Journal of the Electrochemical Society, **141**, 849 (1994).

A. G. Schins, E. M. Dons, A. F. M. Arts,
"Aging in two-dimensional Ising spin glasses"
Physical. Review B **48**, 16524 (1993).

Dedicated to my wife, Nelly

ACKNOWLEDGMENT

I would like to thank my advisor, Dr. Kenneth R. Farmer, for his never ending support, patience and confidence. I also want to thank the members of my committee, Dr. Tompa, Dr. Chin, Dr. Xiao and Dr. Wu for their participation.

In addition, I would like to express my appreciation to Dr. Veena Misra and her students for helping me during my cleanroom work at NC State. I would also like to acknowledge the staff of Structured Materials Industries, Inc. for their collaboration on the hafnium processing.

My former lab mates deserve special thanks for their help and advice, especially the funny young man Michael Beggans.

This work was partially supported by the National Science Foundation, ECS-9624798.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

Since the invention of the first integrated circuit, the semiconductor industry has seen a phenomenally rapid pace of improvements in device performance. This trend of ever smaller and faster devices is a result of the ability to exponentially reduce feature sizes of integrated circuits, a trend commonly known as "scaling". An indication of the pace at which devices are scaled down is known as Moore's Law. It was first introduced as an observation that the number of transistors per chip seemed to double every 24 months, but Moore's Law has been remarkably accurate in predicting the scaling trend for future device generations.

There are two very clear-cut reasons for scaling: smaller is cheaper and smaller is faster. There is no doubt that an important reason behind the scaling of transistor devices is an economic one; more transistors per wafer makes each individual transistor cheaper to produce. But a more important reason for scaling is the fact that the basic building block of CMOS technology, the MOSFET, operates faster when the overall size is shrunk.

Fundamental road blocks for continued scaling are appearing, calling into question the sustainability of the trend towards ever faster devices. On the other hand, scaling may also reveal device properties that were unknown. This offers opportunities to develop new devices and applications that were either unimaginable or impossible before.

1.1 Scope of Research

A reduction in overall feature sizes of a MOSFET requires a simultaneous reduction in the thickness of its gate dielectric, SiO_2. As the thickness of the gate oxide is already in the ultrathin regime (<35 Å), and is expected to shrink even further, quantum mechanical tunneling, or direct tunneling, through the gate oxide becomes prohibitively large. As a result, characterization techniques need to be reevaluated. One particularly important technique, capacitance-voltage measurements, is the de facto industry standard to determine the electrical properties and integrity of the gate oxide. However, the presence of the direct tunnel leakage current renders this technique useless in the ultrathin regime. Characterization of ultrathin oxides is still crucial because of the importance of the properties of the gate dielectric on overall device performance. A new technique called leakage compensated charge measurement is introduced to that end.

Besides making characterization of ultrathin oxides problematic, the direct tunnel leakage current has obvious adverse effects on battery life (mobile applications) as well. Continued scaling of the dielectric thickness may well make the presence of the leakage current prohibitively large, endangering a further scaling down of feature sizes and, therefore, device performance. A large body of research exists that has reported on materials with a high dielectric constant as possible candidates to replace the gate dielectric of choice, silicon dioxide. One particularly promising candidate is hafnium oxide; this work will present one of the earliest research efforts into evaluating the properties of this material for such applications.

These first two topics dealt with the adverse effects of scaling and potential solutions to it. Scaling also offers opportunities. One such opportunity is the emergence

of new device applications based on the properties of ultrathin layers, in particular ultrathin layers of silicon and ultrathin layers of silicon dioxide. This research will focus on one particular system based on alternating layers of ultrathin silicon and ultrathin silicon dioxide. One example of new device applications that are possible with this system will be developed, a nonvolatile memory structure.

1.2 Statement of Purpose

The purpose of the research presented here has been to (1) develop a leakage compensated charge measurement for characterization of ultrathin silicon dioxides, (2) investigate the electrical properties of hafnium oxide as a possible candidate for advanced gate dielectric applications, (3) investigate the electrical properties of a silicon-based multilayer charge injection barrier and (4) develop a novel nonvolatile memory transistor that incorporates this barrier as the active tunnel dielectric.

1.3 Outline of the Dissertation

The dissertation is divided into seven chapters and two appendices. Chapter 2 presents a review of CMOS technology, a brief history, scaling and a review of the properties of ultrathin silicon dioxide. Chapter 3 describes the principle of a leakage compensated charge measurement and results of such measurements on ultrathin silicon dioxide devices that exhibit a large leakage current. Chapter 4 presents some of the earliest results ever obtained on the electrical properties of hafnium oxide as a candidate for future gate dielectric applications. Chapter 5 introduces a silicon-based multilayer structure and discusses the electrical properties the system has. Chapter 6 describes the fabrication and characterization results of a new application that was made possible by the properties of silicon-based multilayer barriers, a nonvolatile memory transistor.

Chapter 7 presents the conclusions drawn from this research and finishes with suggested future work. Appendix A contains a table of the process parameters used for the deposition of thin hafnium oxide layers discussed in Chapter 4. Finally, Appendix B contains a table of the process parameters used to fabricate the tunnel stack that is incorporated as the active dielectric in the nonvolatile memory transistor fabrication run of Chapter 6.

CHAPTER 2

REVIEW OF CMOS TECHNOLOGY

2.1 Overview

This chapter covers a review of CMOS technology and ultrathin silicon oxides. First it gives a brief introduction and history of CMOS technology; then it will describe the operation and device characteristics of the basic building block of CMOS, the MOSFET. This leads to Moore's Law, the scaling of devices to smaller sizes and its effect on device performance. The final section of this chapter reviews the properties, characterization techniques and fabrication of ultrathin silicon oxide layers.

2.2 Introduction to CMOS Technology

Complementary MOS is so-named because it uses both p- and n-type (complementary) MOS transistors in its circuits. Figure 2.1 depicts a schematic illustration of the basic building block of integrated circuits, the CMOSFET, the Complementary Metal Oxide Semiconductor Field Effect Transistor. The left part of the illustration shows the n-doped source and drain regions in a p-type tub, as well as the n-gate separated from the channel region by a gate dielectric. Applying a positive voltage to the gate electrode will set up an electric field in the semiconductor and induce the forming of an n-type channel underneath the gate dielectric, allowing electrons to travel from the source to the drain. Hence, the name NMOSFET. The right part of the illustration shows the same configuration with the role of electrons now being played by holes, hence PMOSFET. The heart of the structure, the gate dielectric between the gate and the channel is arguably the most important part of the entire transistor structure and is typically formed by a

silicon dioxide layer. It is the unique properties of this silicon – silicon dioxide interface that is solely responsible for the existence of the silicon based microelectronics industry we know today, as will be explained later in this chapter.

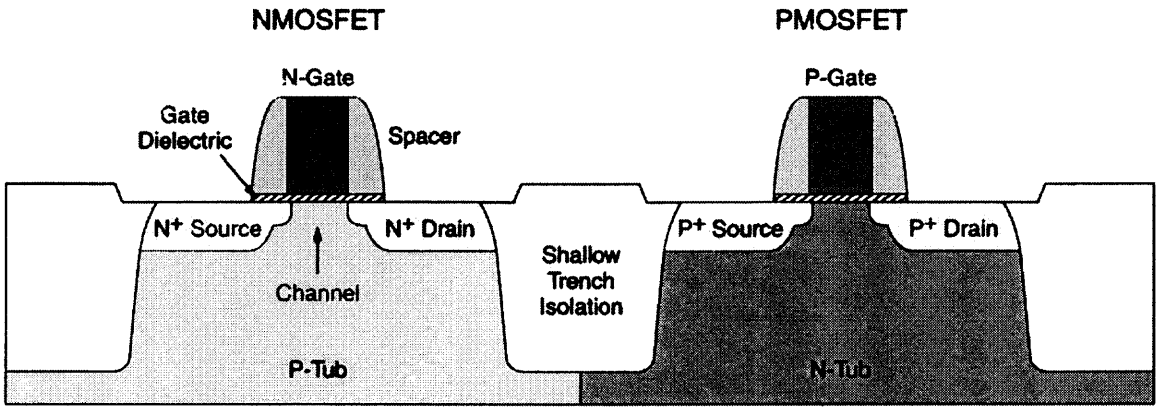

Figure 2.1 Cross-section of a Complementary Metal Oxide Semiconductor Field Effect Transistor (CMOSFET).

2.3 History of CMOS Technology

It is important to note that even though CMOS was introduced in 1963 [1], it was not the technology of choice even as late as the late 1970's; NMOS was. The popularity of NMOS over PMOS was due to the fact that it was cheap to fabricate and faster than PMOS because of the higher electron mobility. However, NMOS gates draw dc power even when no signal is applied, hence an integrated circuit will draw a steady current in the standby mode. Consequently, as the number of transistors on the chip grows, the power being dissipated also increases. Although this was always a limitation of NMOS, it did not represent a drawback for most applications when the number of devices was relatively small. Such was the situation at the level of device integration that existed up

to 1978 when Intel introduced the 8086 processor which was the last processor to be built in NMOS. It had 29,000 devices and dissipated 1.5 W of power at 8 MHz. However, when this same processor was later reintroduced in CMOS technology as the 80C86 the power dissipation dropped to 250 mW.

So why this decrease in power consumption for CMOS technology? In a CMOS transistor only one of the two transistors is driven at any one time. This means that a high impedance path exists from the supply voltage to ground, regardless of whether the transistor is in the on- or off-mode. Hence, very little current flows and almost no dc-power is dissipated. CMOS thus allows the manufacturing of circuits that need very little standby power.

The problem of power dissipation can also be considered from both a chip perspective and a system perspective. From the chip perspective, if microprocessors of the 32-bit generation – which only now are starting to be replaced by 64-bit microprocessors for high-end applications – were built in NMOS, they would dissipate 5 to 6 W of power. This would lead to severe heating and reliability concerns. In addition, expensive packages would be needed to house such chips. However, building these same microprocessors in CMOS reduces power consumption to about 1 W. From the system perspective, let's consider memory chips. Although a 1-Mbit DRAM may consume only 120 mW of power in NMOS, it consumes even less, ~ 50 mW, in CMOS. Since there may be thousands of memory chips in a system, the ramifications of lower power dissipation are significant. Smaller power supplies and smaller cooling fans are two important ramifications.

Even though the most important advantage of CMOS is its significantly reduced power density and dissipation, there are other advantages as well, such as device performance, reliability, circuit design and cost [2].

2.4 Moore's Law and Scaling

For four decades, the semiconductor industry has distinguished itself by the rapid pace of improvement in its products. The principal categories of improvement trends are shown in the table below with an example of each.

Table 2.1 Scaling trends

TREND	EXAMPLE
Integration Level	Components/Chip (Moore's Law)
Cost	Cost per function
Speed	Clock Rate, GHz
Power	Laptop or cell phone battery life
Compactness	Small and light-weight products
Functionality	Nonvolatile memory

Source: International Technology Roadmap for Semiconductors (Semiconductor Industry Association), Santa Clara, CA, 2003 [3].

All of these trends, sometimes referred to as "scaling" have resulted principally from the industry's ability to exponentially decrease the minimum feature sizes used to fabricate integrated circuits. Of course, the most frequently cited trend is in integration

level, which is usually expressed as Moore's Law [4,5], i.e. the number of components per chip doubles every 24 months. The most significant trend for society is the decreasing cost per function, which has led to significant improvements in productivity and quality of life through proliferation of computers, communication devices and consumer electronics. The author would like to point out that, even though he is fully aware of the near sanctified status that the phrase "Moore's Law" has attained in the semiconductor community, the term Moore's *Law* is a misplaced term, since it describes merely an observation made by Mr. Moore in December 1975, *not* a *law* in the mathematical sense of the word that there is a proof for it. As a matter of fact, Moore's Law is very often erroneously described as a doubling of components per chip every 18 months.

2.4.1 MOSFET Operation

As noted previously, the Metal Oxide Semiconductor Field Effect Transistor (MOSFET) forms the basic building block of the microelectronics industry. As a matter of fact, more than 99% of all integrated circuits are MOSFETs used for random-access memory (RAM), flash memory, microprocessors and application specific integrated circuits (ASIC). Figure 2.2 shows a cross-sectional view of an n-channel MOSFET and will be used to illustrate the operation of the device [6].

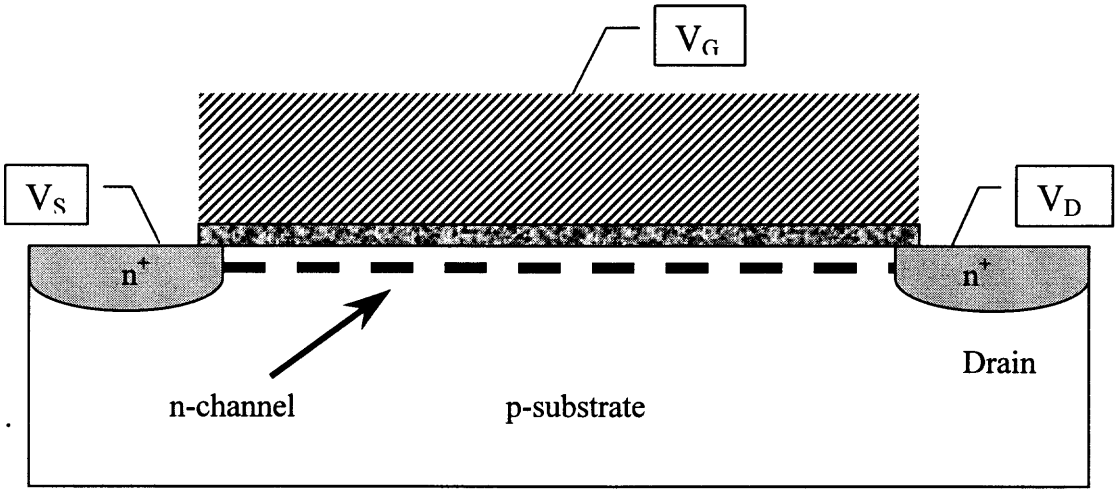

Figure 2.2 Cross-section of an n-channel MOSFET

The device has a gate terminal to which the input signal is applied as well as source and drain terminals across which the output voltage is developed and through which the output current flows. A channel region in the silicon substrate under the gate electrode separates the source and the drain. The substrate is also physically separated from the gate electrode by an insulating layer – typically SiO_2 – so that no current flows between the gate electrode and the semiconductor.

In simplest terms, the operation of a MOSFET involves the application of an input voltage to the gate, which sets up a transverse electric field in the channel region of the device. By varying this transverse electric field, it is possible to modulate the conductance in the channel region. Since an electric field controls current flow, such devices are called *field effect transistors*. If no gate bias is applied, the electrical path between source and drain consists of two back-to-back pn junctions in series, one of which will be in reverse direction. The channel current I_D will only consist of the reverse-bias diode leakage current and hence will be considered negligibly small.

When positive bias is applied to the gate electrode, electrons will be attracted to the channel region and holes (the majority carriers in a p-type substrate) will be repelled. Once enough electrons have been drawn into the channel by the positive gate voltage to exceed the hole concentration, the region behaves like an n-type semiconductor. Under these circumstances, an n-type channel connects the source and the drain regions. Current will flow if a voltage V_{DS} is applied between the source and the drain terminals. The voltage-induced n-type channel does not form unless the voltage applied to the gate exceeds a certain threshold voltage V_T. A device as is described above is referred to as an enhancement mode (or normally OFF) transistor. It is also possible to build MOSFETs in which a conducting channel region exists when $V_G = 0$ V. Such devices are referred to as depletion mode (or normally ON) transistors, since a bias voltage is needed to deplete the channel region of majority carriers.

2.4.2 MOSFET Characteristics

This section will show the equations that describe the current-voltage characteristics of an NMOSFET. In the simplest model, if V_G is smaller than V_T, no channel exists and no current is assumed to flow between the source and the drain. If V_G is greater than V_T, a conducting channel is present and V_{DS} causes a drain current I_D to flow from source to drain. For small values of V_{DS}, the drain current I_D is linearly related to V_{DS}. In this so-called *linear region* of operation, the equation for the drain current is:

$$I_D = \frac{1}{2} \cdot \mu_n \cdot C_{ox} \cdot \frac{W}{L} \left[2(V_G - V_T)V_{DS} - V_{DS}^2 \right] \qquad \text{(Equation 2.1)}$$

where μ_n is the mobility of electrons in the channel, C_{ox} the oxide capacitance, W the width of the channel region and L the length of channel region between source and drain, or *gate length*. This expression is valid when $V_G > V_T$ and $V_{DS} < V_G - V_T$.

As the value of V_{DS} increases, the induced conducting-channel charge decreases near the drain. When V_{DS} equals or exceeds $V_G - V_T$, the channel is said to be pinched off. Increases above this critical voltage produce little change in I_D and Eq. 2.1 no longer applies. The value of I_D in this region is given by the following expression:

$$I_D = \frac{1}{2} \cdot \mu_n \cdot C_{ox} \cdot \frac{W}{L} (V_G - V_T)^2 \qquad \text{(Equation 2.2)}$$

This is the so-called *saturation region* of operation.

A plot of I_D versus V_{DS} (with V_G as parameter) for a NMOSFET as described by the previous two equations is depicted in Figure 2.3. If the value of V_G is smaller than V_T, the transistor is said to be in *cutoff*. In the simplified model given here, I_D is assumed to be zero in cutoff. It is important to note that this model does not take into account two very important factors that can significantly affect device performance, namely short-channel effects and subthreshold currents. Discussion of these effects lies outside of the scope of this work.

Another important parameter is the so-called *saturation transconductance* g_m and is defined as $g_m = dI_D / dV_G$. As such, it can be viewed as a measure of the easiness with which charge carriers can drift from the source to the drain. From the previous equation it can be shown that the transconductance follows this expression:

$$g_m = 2 \cdot \mu_n \cdot C_{ox} \cdot \frac{W}{L} (V_G - V_T) \qquad \text{(Equation 2.3)}$$

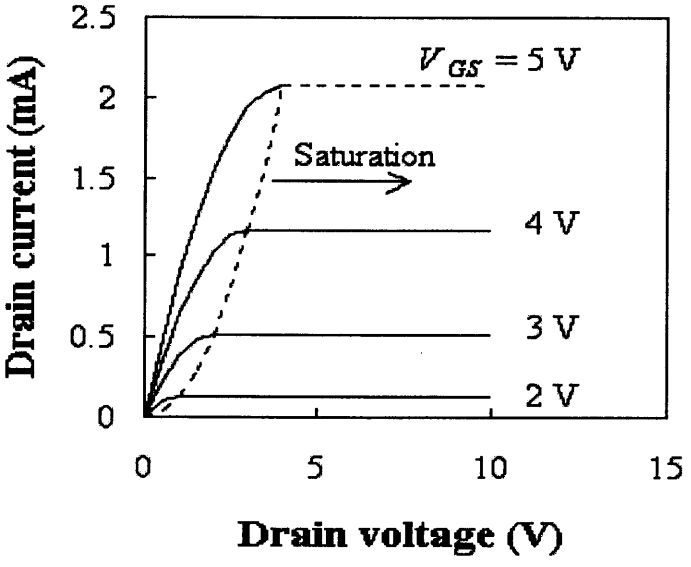

Figure 2.3 Drain current I_D vs. Drain voltage for different gate voltages. The drain current saturates at a value that is described by Equation 2.2.

A critical metric for transistor speed is the intrinsic switching frequency f_i. It can be shown that this is not limited by the time it takes the charge carriers to cross the channel from source to drain, the channel transit time, but rather by the intrinsic delay time τ_i required to charge and discharge the load capacitance of the transistor that exists between device electrodes and between the interconnecting lines of the circuit. The equation that describes the intrinsic delay time τ_i is as follows:

$$\tau_i = \frac{1}{f_i} = \frac{C_L \cdot V_{dd}}{I_D} \qquad \text{(Equation 2.4)}$$

where C_L is the load capacitance of the transistor and V_{dd} the power supply voltage.

The equations described in this section are instructive in explaining how a reduction in transistor size will increase device performance as will become clear in the following section.

2.4.3 Why Scaling?

As mentioned earlier in this chapter, Moore's Law describes a historic observation of a doubling of components per chip every 24 months. Even though the economic reasons behind this endless effort to ever smaller devices can not be dismissed – smaller is cheaper – it has really been the quest for ever faster devices as described by Equation 2.4 that has been the driver behind the steady performance improvement of CMOS devices.

From Equation 2.4 it is clear that there are three ways to reduce the delay time τ_i, and therefore increase the speed of the transistor: a reduction in the load capacitance C_L, a reduction in the power supply voltage V_{dd} and an increase in the drive current I_D. While all three trends occur simultaneously and are important parameters in the scaling of transistors, it is the increase in I_D that is the most instructive in showing that in order to improve the performance of the device, one has to reduce the device in size.

Equation 2.2 predicts that in order to increase I_D, one can simply increase the dimension of the gate width W. However, when minimum-sized devices are preferred for economic reasons, this is not an option. The drive current I_D also is inversely proportional to the channel length L and minimum channel lengths are therefore required. From the dependence of I_D on the electron mobility μ_n it is clear that the electron mobility must be as high as possible. Since the mobility of carriers decreases as the doping concentration of the channel increases, lightly doped channel regions are important. In addition, many efforts are under way to increase the carrier mobility through alternative means such as using strained-silicon substrates [7].

Equation 2.2 also shows that the gate oxide capacitance C_{ox} is proportional to I_D. Since C_{ox} is inversely proportional to the gate oxide thickness t_{ox}, as thin a gate oxide as

possible is needed, commensurate with oxide breakdown and reliability considerations. However, continued scaling of gate oxide thicknesses results in a significant increase in gate leakage current due to a sharply increasing direct tunneling leakage current which has adverse effects on device performance, in particular for low-power applications.

Even though all the factors mentioned here are important aspects of device scaling, it is the issues related to the scaling of the gate oxide that will be discussed in detail in this work.

2.5 Ultrathin Silicon Dioxide

From the previous section it has become clear that the gate dielectric, usually silicon dioxide, is perhaps the most critical part of a MOSFET. It is largely the wonderful characteristics of silicon dioxide that have enabled an aggressive scaling of MOSFETs to ever smaller devices. The table below shows some selected properties of SiO_2.

Table 2.2 Properties of silicon dioxide

Selected properties of SiO_2 gate dielectric layers
Native to silicon (SiO_2 is the only stable phase on Si)
Low interfacial (Si/ SiO_2) defect density (~$10^{10} eV^{-1} cm^{-2}$)
High melting point (1730 °C)
Large energy band gap (9 eV)
High resistivity (10^{15} Ω cm)
High dielectric strength (15 MV/cm)
Dielectric constant = 3.9

SiO_2 is native to silicon, and with it, forms a low defect density interface. It also has high resistivity, excellent dielectric strength, a large band gap, and a high melting point. These properties of SiO_2 are in large part responsible for enabling the microelectronics revolution. Indeed, the first transistor made in 1947 was not made with silicon but with germanium. It was not selected as the semiconducting material of choice, mainly due to the lack of a stable native oxide and a low defect density interface. The ease of fabrication of SiO_2 gate dielectrics and the well passivated Si/ SiO_2 interface that have made this possible. SiO_2 has been and continues to be the gate dielectric of choice for the MOSFET. In spite of its many attributes, however, SiO_2 suffers from a relatively low dielectric constant, $\kappa = 3.9$. Since high gate dielectric capacitance is necessary to produce the required drive currents for submicron devices [8], and further, since capacitance is inversely proportional to gate dielectric thickness, the SiO_2 layers have of necessity been scaled to ever thinner dimensions, as is shown in Figure 2.4.

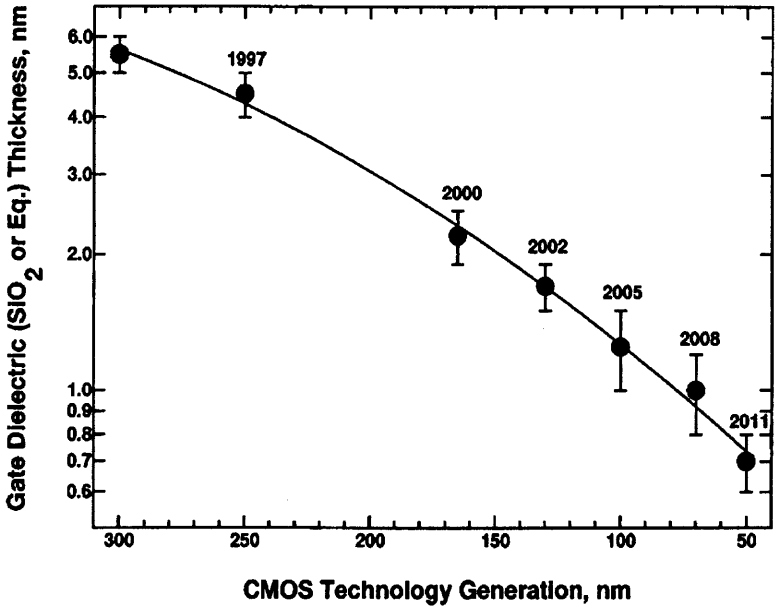

Figure 2.4 Scaling of gate dielectric thickness [9].

This gives rise to a number of problems, including impurity penetration through the SiO_2, enhanced scattering of carriers in the channel, possible reliability degradation, high gate leakage current, the need to grow ultrathin and uniform SiO_2 layers and the need to devise characterization techniques that can be used for these ultrathin layers. Any of these effects may ultimately pose a fundamental limit to the continued scaling of SiO_2.

2.5.1 Fundamental limits?

Due to the large band gap of SiO_2 , ~9 eV, and the low density of traps and defects in the bulk of the material, the carrier current passing through the dielectric layer is normally very low. For ultrathin films this is no longer the case. When the physical thickness between the gate electrode and doped Si substrate becomes thinner than ~30 Å, direct tunneling through the dielectric barrier dominates leakage current [10,11] According to fundamental quantum mechanical laws, the tunneling current increases exponentially with decreasing oxide thickness. Figure 2.5 depicts gate leakage currents for oxide thicknesses ranging from 32 Å to 6 Å:

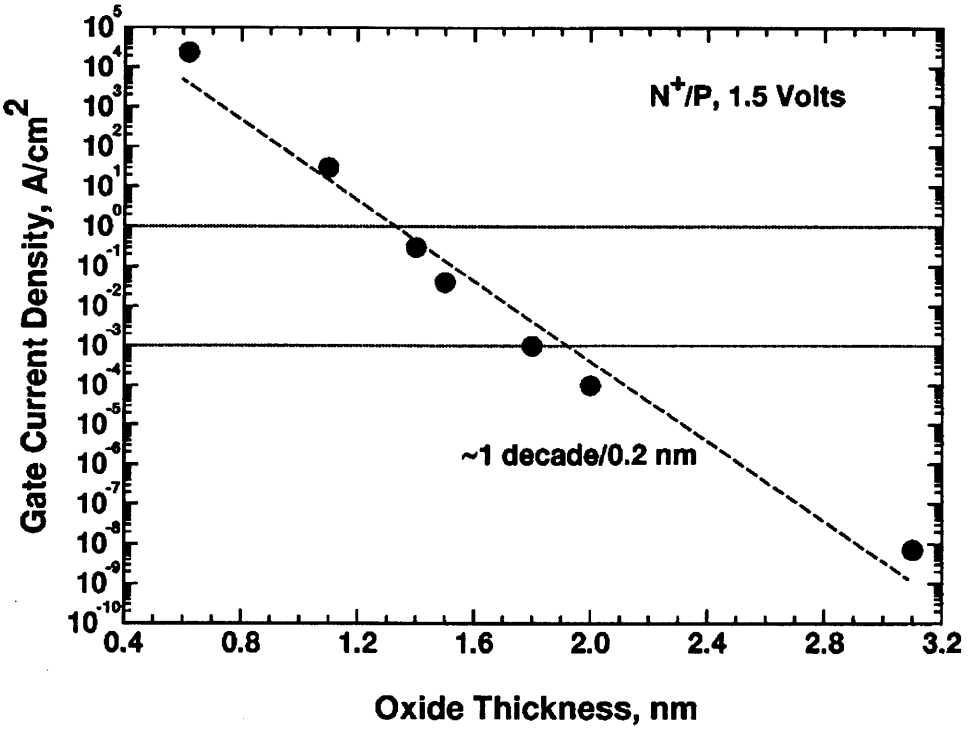

Figure 2.5 Gate leakage current density as a function of oxide thickness ranging from 32 Å to 6 Å [12].

The leakage current is seen to increase by one order of magnitude for each 2 Å thickness decrease. Assuming a maximum allowable gate current density of 1 A/cm^2 for desktop computer applications, and 10^{-3} A/cm^2 for portable applications, minimum acceptable SiO_2 physical thicknesses would be approximately 13 and 19 Å, respectively.

Reliability, which is the lifetime to breakdown, of ultrathin SiO_2 is a major concern for oxide scaling into the sub-20 Å regime. Electrons traveling through the SiO_2 layer may create defects such as electron traps and interface states [13] that in turn, upon accumulation to some critical density, degrade the insulating properties of the oxide. It has been predicted that reliability may attain unacceptable levels at thicknesses as low 10 Å [14], but considerable work is still being done in this field.

In addition, a reduced drive current has been reported in small transistors with ultrathin gate dielectrics less than 13 Å [15]. Thus for SiO_2 layers thinner than 13 Å there is no advantage in performance for incurring the additional burden of an ever increasing gate leakage current. The cause of the decreased drive current is not fully understood. One possibility is an additional scattering component from the upper SiO_2/Si interface. Another cause could be a universal mobility effect, i.e. a lowered mobility due to enhanced scattering because of carrier confinement in the inversion layer of the ultrathin oxide layer.

Thus, the fundamental limits imposed on SiO_2 are excessively high leakage current, reduced drive current and reliability. All three effects suggest that the fundamental scaling limit is somewhere in the range from 10 to 13 Å. In any case, it is important to point out that it has been reported that the fundamental physical limit to SiO_2 is believed to be 8 Å [16]. Oxides thinner than 8 Å can no longer be considered SiO_2, just two interfacial layers sandwiched in between the silicon substrate and the polycrystalline silicon gate. Thus, layers this thin will no longer have the same properties as bulk SiO_2.

2.5.2 Electrical Characterization

There are many analytical techniques available to study both the physical as well as the electrical properties of SiO_2. Physical characterization techniques can be grouped in optical techniques (e.g. ellipsometry), X-ray techniques (e.g. X-ray photoemission spectroscopy (XPS)), ion beam based techniques (e.g. medium energy ion scattering spectroscopy (MEIS) and secondary ion mass spectroscopy (SIMS)), electron microscopy (e.g. transmission electron microscopy (TEM)) and scanning probe microscopic

techniques (e.g. scanning tunneling microscopy (STM)). While all of these physical characterization techniques have been and continue to be very valuable in understanding the properties of ultrathin SiO_2 and the Si/SiO_2 interface, it is the electrical properties that have been a focus of this work. In this section, the application of electrical characterization techniques to ultrathin SiO_2 will be discussed.

The most frequently used electrical technique to assess the properties of both the thin oxide layer and its interface with Si is the C-V measurement. In thicker oxide layers C-V curves can be fitted satisfactorily with classical models[17] The C-V technique can be used to determine flatband and threshold voltage, fixed charge, and interface state density. It is also often used to determine the oxide thickness. In sub-40 Å oxide layers, C-V measurements provide the same information, but the interpretation of the data requires considerable caution. The assumptions needed to construct the ''classical model'' are no longer valid, and quantum mechanical corrections become mandatory, thus increasing the complexity of the analytical treatment. First, several authors have demonstrated that for ultrathin layers, Maxwell–Boltzman statistics no longer describe the charge density in the inversion and accumulation layers satisfactorily, and should be replaced by Fermi–Dirac statistics. In addition, band bending in the inversion layer near the semiconductor–insulator interface becomes very strong, and a potential well is formed by the interface barrier and the electrostatic potential in the semiconductor. This potential well may be narrow enough to give rise to electron confinement at discrete energy levels [18,19]. One of the main effects of the quantum mechanical treatment of the inversion layer is a considerable shift of the inversion charge centroid away from the

semiconductor–insulator interface and can be treated as an additional capacitor in series with the oxide capacitance.

A similar effect is generated by polycrystalline Si depletion on the gate side of the capacitor of a MOS transistor [20]. This effect is related to both the high fields at the insulator surface as well as the incomplete activation of the dopants near the polycrystalline Si/SiO_2 interface. A carrier concentration profile with a finite width, having a centroid several tenths of a nanometer away from this interface, results. This effect can also be modeled as an additional capacitance in series with the oxide capacitance. As a consequence of quantum mechanical effects and polycrystalline Si depletion, the measured capacitance is smaller than the expected "physical" oxide capacitance, and the difference becomes very significant for ultrathin layers. This also implies that oxide thickness extraction from C-V measurements becomes more difficult. For very thin oxides, typically sub-30 Å, the huge leakage current through the oxide, due to direct tunneling of electrons creates an additional complication in the interpretation of C-V curves. A sharp drop in the capacitance is observed as the voltage increases. This effect is illustrated in Figure 2.6 below. As a matter of fact, the next chapter provides an answer to this problem through a new C-V technique that utilizes a tunnel leakage compensation circuit, allowing the use of standard C-V characterization even for ultrathin layers.

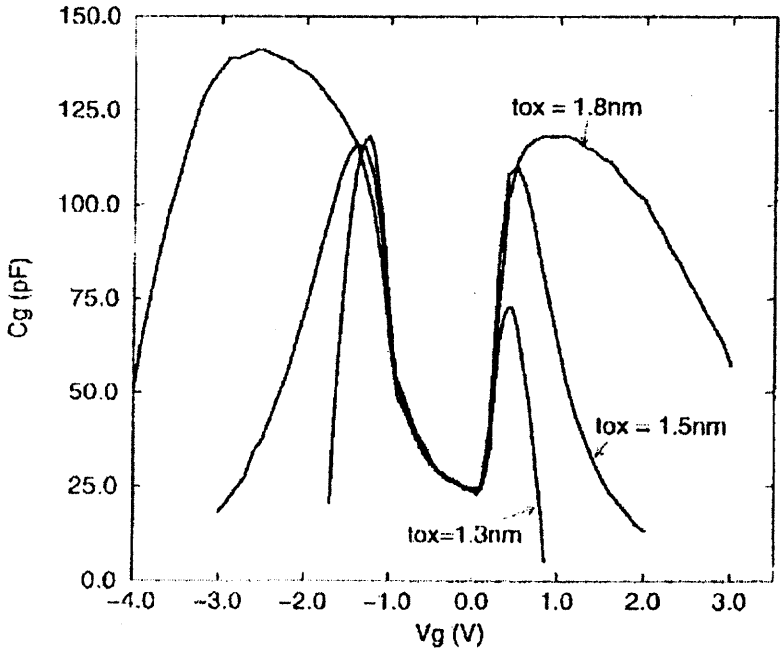

Figure 2.6 C-V characteristics distorted by leakage currents through the ultrathin gate oxide.

Another critical technique to study the bulk and interfacial properties of these layers is the measurement of the tunnel current. When a voltage V is applied across an oxide layer with thickness t_{ox}, the resulting oxide field, $E_{ox} = V_{ox} / t_{ox}$, gives rise to a current flow through the oxide. This current originates from electrons that tunnel quantum mechanically through the Si/SiO_2 potential barrier from the Si conduction band to the SiO_2 conduction band as illustrated in the figure below. When tunneling occurs through a triangular barrier, Figure 2.7(a), the conduction mode is called Fowler-Nordheim (FN) tunneling and the measured current density, J_{FN}, can be described by the following formula:

$$J_{FN} = A \cdot E_{ox}^{2} \cdot e^{\left(-\frac{B}{E_{ox}}\right)} \qquad \text{(Equation 2.5)}$$

where A and B are constants given by:

$$A = \frac{q^3}{8\pi\varphi_b}\frac{m}{m^*} \qquad \text{(Equation 2.6)}$$

$$B = 4\sqrt{2m^*}\frac{\varphi_b^{3/2}}{3\hbar q} \qquad \text{(Equation 2.7)}$$

in which q is the charge of a single electron ($1.6\text{x}10^{-19}$ C), m the mass of a free electron ($9.1\text{x}10^{-31}$ kg), m^* the effective mass of a free electron in the bandgap of SiO_2 (0.42m), h Planck's constant and φ_b the energy barrier at the interface (3.2 eV for Si/SiO_2).

When the oxide voltage drops below 3.2 V, the electron barrier height, electrons can no longer enter the oxide conduction band, but tunnel directly from the gate to the silicon substrate as can be seen in Figure. 2.7(b). In state-of-the art CMOS technology, direct tunneling is the dominant current conduction mechanism at operating voltage. The direct current density cannot be described easily in a closed analytical form, but several approximate formulas and simulations have been proposed [21].

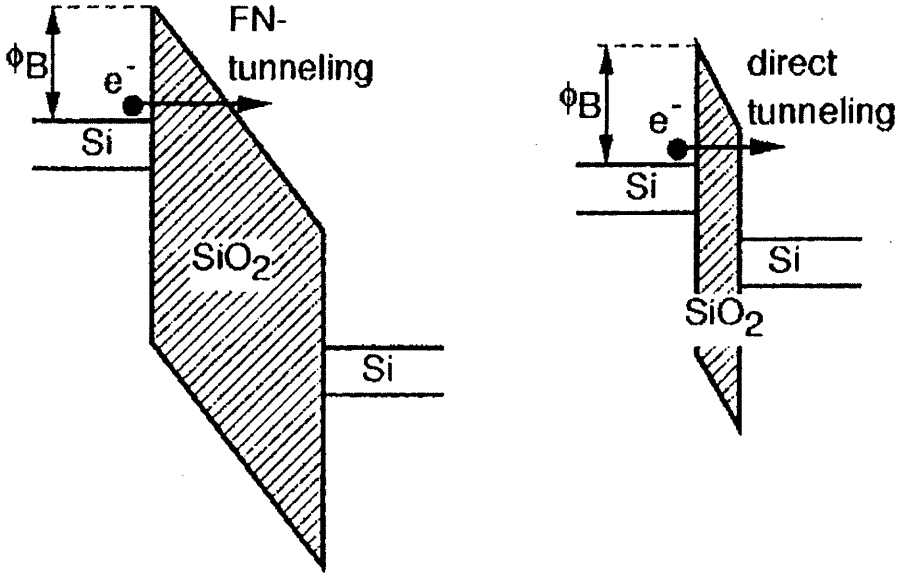

Figure 2.7 Conduction mechanisms through SiO_2: (a) Fowler-Nordheim tunneling and (b) direct tunneling.

In addition there are other electrical characterization techniques such as charge pumping, conductance measurements and inelastic tunneling spectroscopy. Since these techniques have not been used in this research, it is outside the scope of this document to discuss them here.

2.5.3 Oxide Degradation and Breakdown during Electrical Stress

In this section oxide degradation during electrical stress, ultimately leading to breakdown, is briefly discussed. Oxide degradation is defined as the continuous, gradual deterioration of the oxide properties, resulting from structural damage generated in the oxide by electrical stress. Breakdown is triggered when the accumulated damage reaches a critical level.

Degradation can manifest itself in a variety of phenomena. During high field oxide stressing, interface traps are created at the Si/SiO_2 interface [22.]. Their density D_{it} can be obtained from either C-V or charge pumping measurements. It has been claimed that the interface trap density reaches a critical density at the moment of oxide breakdown. In addition, traps are created in the bulk of the oxide that that can be filled with either holes or electrons . This leads to a net positive or negative charging of the oxide. This net charging manifests itself in a net increase of the tunnel current for positive charging or a net decrease for negative charging.

Another phenomenon that occurs during oxide degradation is the generation of stress induced leakage current (SILC) through the gate [23]. It manifests itself as a current that is obtained at ever decreasing applied voltages when the fluence of injected charge carriers increases. The SILC is caused by trap-assisted tunneling from the gate to the substrate. The traps act as "stepping stones" for electrons to tunnel through the oxide.

This SILC itself is in some cases an important reliability problem. Leakage current through the gate translates in a power waste problem in MOSFETs, resulting in reduced battery life. It is also detrimental to the performance of traditional floating gate nonvolatile memory devices, since it causes the charged that is stored on the floating gate to leak off after continuous cycling. This is a problem that will be addressed in more detail in later chapters.

It is clear that trap generation is the key factor in determining oxide degradation and breakdown. Different models have been proposed that describe this process, but the exact physical nature of trap generation is still under considerable debate.

2.5.4 Fabrication of Ultrathin Oxides

The gate dielectric's ultimate electrical performance is determined not only by its composition and fabrication method (growth or deposition), but also by pregrowth surface preparation and postfabrication processing such as plasma etching of the gate stack. The interdependence between the various steps, especially surface preparation, becomes more prominent for ultrathin gate dielectric layers, since the Si/SiO_2 interface is a more significant part of the layer as it gets thinner.

Surface preparation is a more appropriate term than cleaning, since preparation of the Si surface for subsequent oxidation is far more involved than merely removing contamination. In fact, conditioning of the surface to result in its smoothest and cleanest state is just as important a step as the actual dielectric fabrication. Among the important physical attributes of ultrathin SiO_2 and the Si/SiO_2 interface that can be influenced by surface preparation are interfacial roughness, interfacial transition layer width, contamination level of the SiO_2 and the interface and chemical bonding structure at the

interface . Wet cleaning currently dominates pregate oxidation clean applications. The so-called ''RCA clean" is the most widely used clean for removing organic compounds and metals from Si wafers. Subsequent processing in HF removes the chemical oxide that results from the RCA clean. Dry cleaning technology, so called because it involves vapor or gas, and not liquid cleaning of the wafer surfaces, is presently the subject of much research. This is driven primarily by the drive towards *in situ*, cluster tool processing. Some research has already demonstrated oxide reliability increases due to dry pregate oxide cleans.

Fabrication of ultrathin dielectric layers may be accomplished by growth or deposition. Growth refers to thermal oxidation of the silicon. Deposition usually refers to chemical or physical generation of the layer, not involving a reaction with the Si substrate. The utter simplicity of growing thermal SiO_2 by exposing Si to O_2 at elevated temperatures, as well as the perfection of the resulting interface, are in large part responsible for the success of Si as the integrated circuit material of choice. Virtually all SiO_2 gate dielectrics are grown by thermal oxidation, using O_2 or H_2O as the oxidant species. Since oxidation in H_2O enhances oxidation kinetics, it is not generally used for the growth of ultrathin films. Thermal oxides consume Si during growth, thereby continuously creating a new and fresh interface. Thermal growth usually takes place at a higher temperature than chemical or physical deposition, and higher fabrication temperature has been associated with improved dielectric properties.

While the seminal paper by Deal and Grove [24] laid out the mechanism for silicon oxidation, much still remains under debate, especially in the ultrathin regime. The Deal-Grove model treats Si oxidation as the reaction of Si and O at the Si/SiO_2 interface,

accomplished by diffusion of O_2 through the growing oxide. For thick films, the model predicts a parabolic dependence of oxidation time on thickness because the growth is limited by diffusion. In thinner films, the model shows that the reaction rate at the interface governs the growth and results in a linear relationship between the oxidation time and thickness.

There are currently two primary thermal techniques for growing SiO_2, furnace or rapid thermal oxidation, RTO. Oxidation usually takes place in the temperature range of 750–1100 °C. Furnace technology is still the manufacturing standard for ultrathin oxide growth. Furnaces are robust and reliable and offer excellent thickness uniformity. In addition it allows wafers to be processed in "batches". On the other hand, RTO offers better absolute thickness control for oxides <20 Å, a greater processing temperature range, and is ''cluster-friendly,'' i.e. processing chambers can be integrated for control of interfaces. However, wafers are processed individually instead of in batches.

Chemical deposition processes are usually used when a lower thermal budget for the dielectric growth step is desired. Since deposition kinetics are slow at such temperatures (typically 350–600 °C), a plasma source is commonly used to activate the reaction. Chemical deposition methods do not consume the substrate, unlike thermal oxidation, and interfacial properties are usually inferior to those of thermal oxides. High temperature anneals (>750 °C) are usually necessary to bring the electrical performance up to the level of thermal oxides. SiO_2 layers have been deposited by chemical vapor deposition (CVD). However, their application to ultrathin gate dielectrics will be limited due to difficulties in controlling deposited layer thickness uniformity across large wafers.

Atomic layer deposition (ALD), in which films are grown approximately one monolayer at a time, has been used to grow ultrathin SiO_2 layers. This may be an important technique to grow <5 Å SiO_2 layers, useful as buffer layers between Si and high- k gate dielectrics. ALD has the outstanding advantages of superb conformal coverage as well as precise thickness control.

Crucial electrical performance parameters such as mobility and interface state density are directly related to physical structure and chemical bonding at the Si/ SiO_2 interface. This interface does not reach its final configuration after oxidation, but rather after all postoxidation processing has been completed. Since the interface is defined by the last SiO_2 to form and the last thermal treatment it is exposed to, postoxidation processing, which involves among other steps implant activation annealing, polycrystalline Si deposition, and plasma etching and deposition, greatly impacts the properties of the gate dielectric. It is generally agreed that annealing parameters play an important role in device manufacturing. However, in most cases these parameters are chosen empirically and not optimized.

CHAPTER 3

C-V CHARACTERIZATION OF ULTRATHIN SIO_2

3.1 Overview

For over two decades, capacitance-voltage (C-V) measurements of MOS structures have been a key analysis tool for the semiconductor industry in understanding the electrical properties of silicon dioxide as well as the properties of the Si/SiO_2 interface. However, with continued scaling of the gate dielectric thickness into the direct tunneling regime, <3.5 nm, the transport of carriers through the dielectric distorts the C-V characteristics, rendering this technique no longer useful for future device characterization [25-33]. This chapter describes an approach to overcome this difficulty using a leakage-compensated charge method (LCCV) which can produce true static C-V curves even when the leakage current is over five orders of magnitude larger than the displacement current that would be expected in a corresponding quasistatic measurement.

In the first part of this chapter, the MOS system is introduced and its capacitance behavior under a bias voltage. The following section explains the principle of the leakage compensated charge measurement. Experimental results on oxides as thin as 2.4 nm will be presented as well as a discussion of the results.

3.2 The MOS System

Figure 3.1 provides a schematic overview of the various regions associated with the gate stack of a MOSFET.

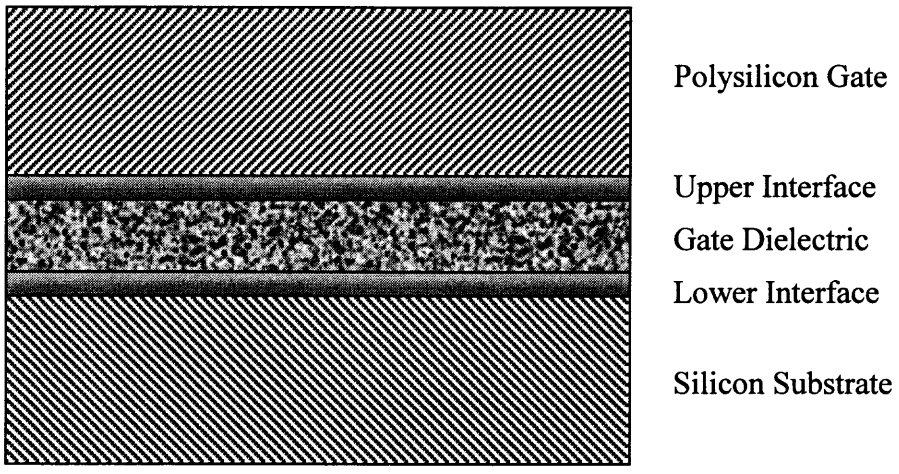

Figure 3.1 Cross-section of the gate stack in a MOSFET structure.

The gate dielectric (typically SiO_2) insulates the gate electrode from the silicon substrate. Gate electrodes in modern CMOS technology are composed of highly doped polycrystalline silicon. The interfacial regions between the gate dielectric and the gate (upper interface) and between the gate dielectric and the substrate (lower interface) are typically around 0.5 nm thick. These interfaces are particularly important with respect to device performance. For instance, they represent a capacitance that can be relevant if the thickness of the interface is substantial to the overall thickness of the dielectric. The lower interface must be of high quality, i.e. low interface trap density in order to minimize scattering in the top layer of the substrate, the channel region. Scattering in the channel region adversely affects the mobility of charge carriers in the channel and therefore the transistor drive current.

It is instructive to consider the energy band diagrams for an ideal MOS structure with zero applied voltage depicted in Figure. 3.2.

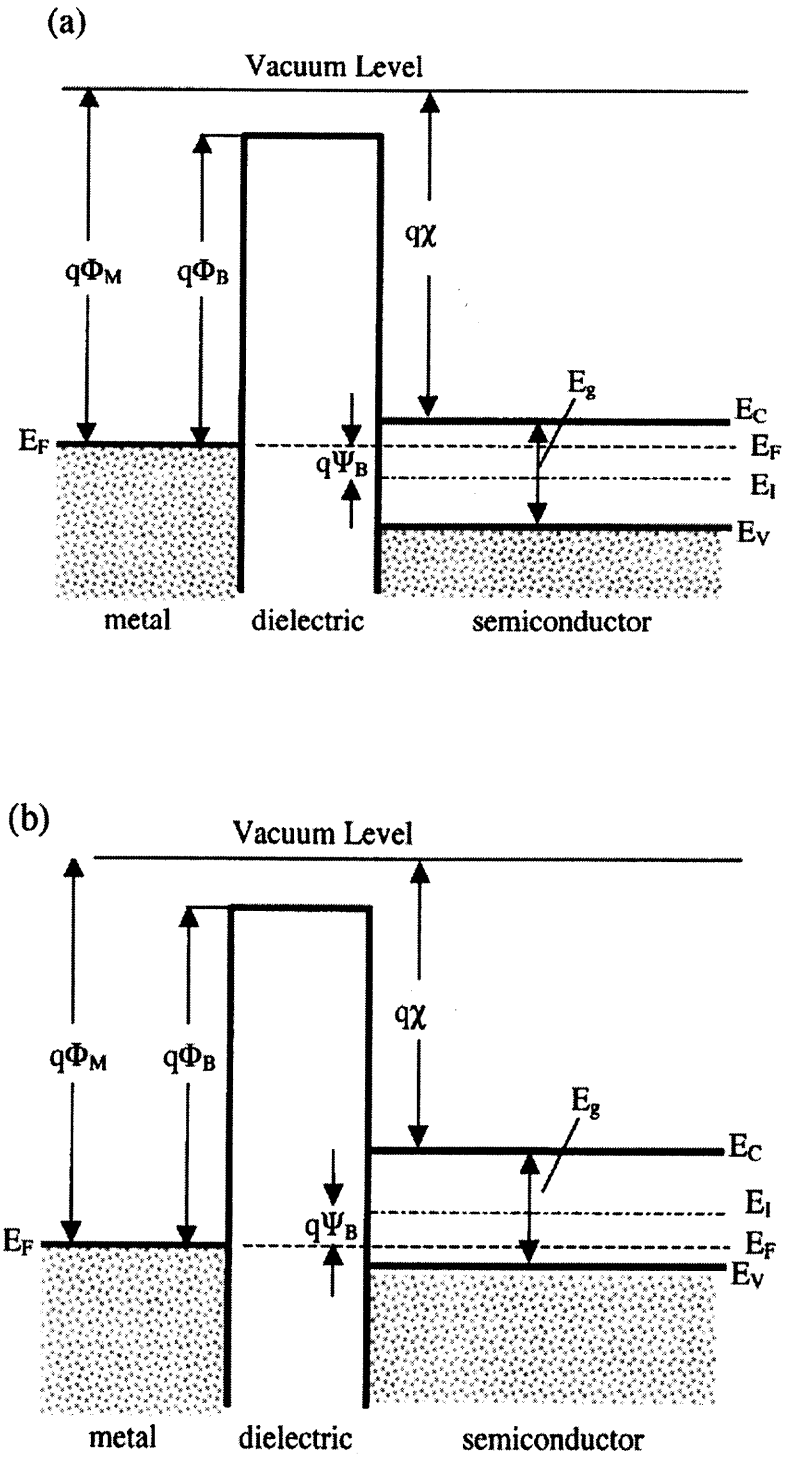

Figure 3.2 Energy band diagrams for an ideal MOS structure with zero applied voltage; (a) on n-type silicon and (b) on p-type silicon.

Figure 3.2(a) shows the energy band diagram for an MOS structure on an n-type substrate, Figure 3.2(b) for a p-type substrate. For these ideal structures, at V = 0 applied voltage on the gate, the work function difference between the gate and semiconductor substrate, Φ_{ms}, is zero. Or,

$$\Phi_{ms} = \Phi_m \cdot \left(\chi + \frac{E_g}{2q} - \Psi_B \right) \qquad \text{(n-type)} \qquad \text{(Equation 3.1)}$$

and

$$\Phi_{ms} = \Phi_m \cdot \left(\chi + \frac{E_g}{2q} + \Psi_B \right) \qquad \text{(p-type)} \qquad \text{(Equation 3.2)}$$

where Φ_m is the metal work function, χ the electron affinity, E_g the bandgap of the semiconductor (1.1 eV for Si) and ψ_B the potential difference between the Fermi level E_F and the intrinsic Fermi level E_i. Under these conditions, the energy bands are flat across the structure. A more typical case is that the Fermi levels of the gate and substrate are misaligned by an energy difference and a voltage V_{FB}, the flatband voltage, must be applied to bring the Fermi levels into alignment. Many dielectrics, including SiO_2, exhibit a fixed charge in the oxide, Q_F, however, resulting in a required $V = V_{FB} \neq 0$ to achieve a flat band condition.

3.3 The MOS System under Bias Voltage

Under a bias voltage the MOS structure acts as a nonlinear capacitor, i.e. the capacitance varies with the applied voltage. There are three basic regimes of operation: accumulation, depletion and inversion. An MOS structure on p-type silicon (Figure 3.2 (b)) will be used to illustrate the different regimes.

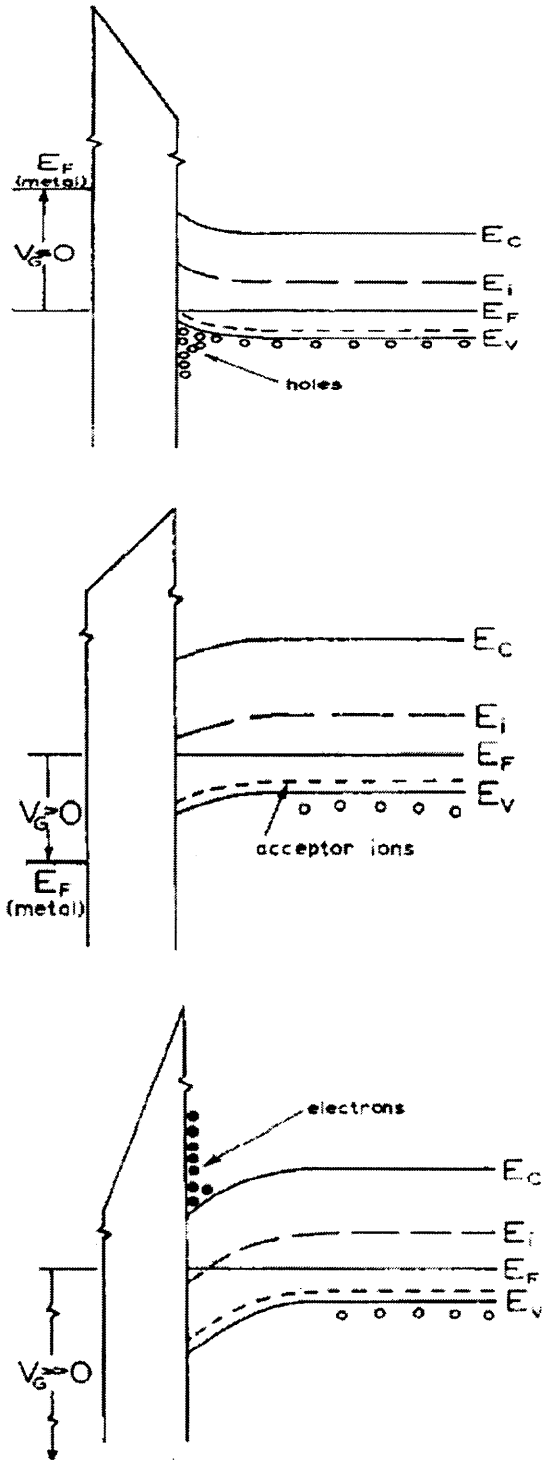

Figure 3.3 A p-type MOS system under bias: (a) accumulation mode (V < 0V), (b).depletion mode and (c) inversion mode (V > 0V).

When a bias voltage $V < 0$ V is applied to the gate, holes will be attracted to the interface between the gate dielectric and the substrate. The conduction and valence bands bend upwards (Figure 3.3(a)) and since the majority carrier density (holes) depends exponentially on the energy difference $E_F - E_v$, holes will accumulate at the substrate side of the gate dielectric. This is the "accumulation" mode. In this case, the gate dielectric acts as a straightforward parallel plate capacitor whose capacitance is fully determined by the gate dielectric.

$$C = \frac{\varepsilon \cdot A}{t_{ox}} \qquad \text{(Equation 3.3)}$$

Upon increasing the bias voltage – making it less negative at first, then slightly positive – the energy bands will start bending downwards (Figure 3.3(b)). This effectively increases the energy difference $E_F - E_v$, causing the substrate region immediately adjacent to the gate dielectric to be depleted of majority carriers. Hence the term "depletion mode". Further increases in applied voltage will increase the width of the depletion region. In this mode the gate dielectric no longer acts as a parallel plate capacitor. Rather, the capacitance decreases with voltage as a result of an increasing width of the depletion region. The capacitance of the MOS structure reaches its minimum at a voltage at which maximum depletion is reached.

Further increases to larger positive gate voltages will bend the bands even further (Figure 3.3(c)), so that the Fermi level will cross over the intrinsic level E_i. At this point, the minority carriers (electrons) in the region immediately next to the gate dielectric will outnumber the majority carriers and the surface will become inverted. Hence the term "inversion mode". The minority carriers appear at the surface by ways of a thermal generation process that is time dependent. It is in this regime that the frequency

dependence of the MOS capacitance primarily occurs since a certain time is needed to generate the minority carriers in the inversion layer. Thermal equilibrium is therefore not immediately obtained.

If the frequency of the AC signal with which the capacitance is measured, is low enough, the minority carriers will be able to respond to the changing gate bias. As a result, the inversion layer will be built up and the measured capacitance is equal to the capacitance of the oxide. On the other hand, if the frequency of the AC signal is high, the minority carriers will not be able to respond and the inversion layer will not build up. Consequently, the measured capacitance will remain at the minimum value it achieved in depletion mode. The voltage dependence of the capacitance of a p-type MOS structure as described here, is shown in Fig. 3.4. Obviously, a similar behavior will be obtained for an n-type MOS structure, with a C-V curve that is a mirrored curve of the one shown below.

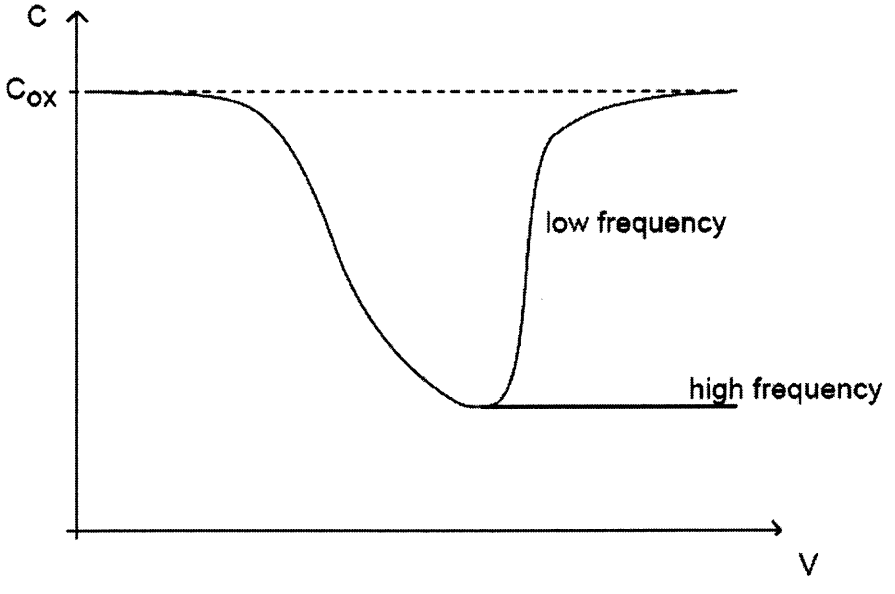

Figure 3.4 Typical frequency dependence of C-V characteristics for a p-type MOS structure.

3.4 Leakage Compensated Capacitance Measurement

The capacitance characteristics for MOS structures as described in the previous section are valid as long as the oxide is relatively thick, > 3.5 nm. However, as CMOS technologies have evolved to include gate dielectrics that extend into the ultrathin regime, < 3.5 nm, a significant direct tunneling current severely distorts the C-V characteristics of the device, as will become clear from the next two figures. As can be seen in Figure 3.5, the direct tunnel current through a 6.25×10^{-4} cm^2, 3.5 nm thick oxide reaches 10 pA at 1.5 V, a bias voltage where capacitance would be measured for the purpose of characterization. In a 2.4 nm oxide of the same area, the leakage is more than five orders of magnitude higher.

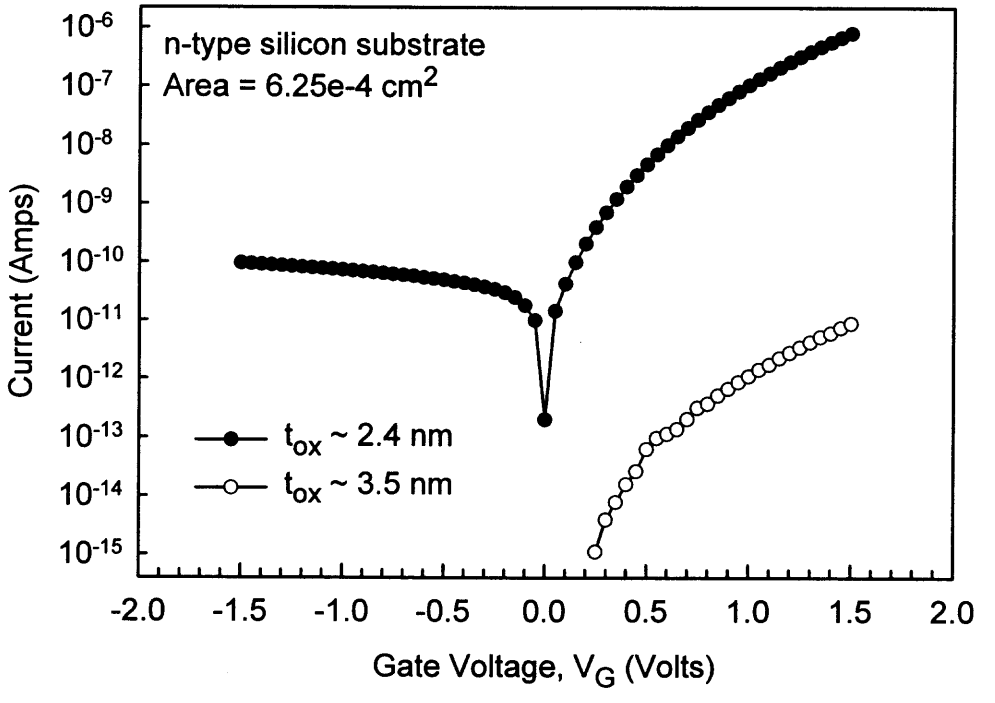

Figure 3.5 Leakage current through a 2.4nm and 3.5 nm ultrathin oxide.

Figure 3.6 demonstrates the effect that this direct tunnel leakage has on the capacitance characteristics. The graph shows the quasistatic C-V curve of a 6.25×10^{-4} cm^2, 3.5 nm thick oxide measured at the relatively fast sweep rate of 50 mV/s, i.e. the same device whose I-V characteristics are represented by the open bullets in the previous figure. It is apparent that even a leakage current of several pA severely distorts the C-V characteristics. It is obvious that the significantly higher leakage current of the 2.4 nm oxide, as shown in the previous figure, will lead to an even greater distortion, making C-V characterization of such oxides all but impossible.

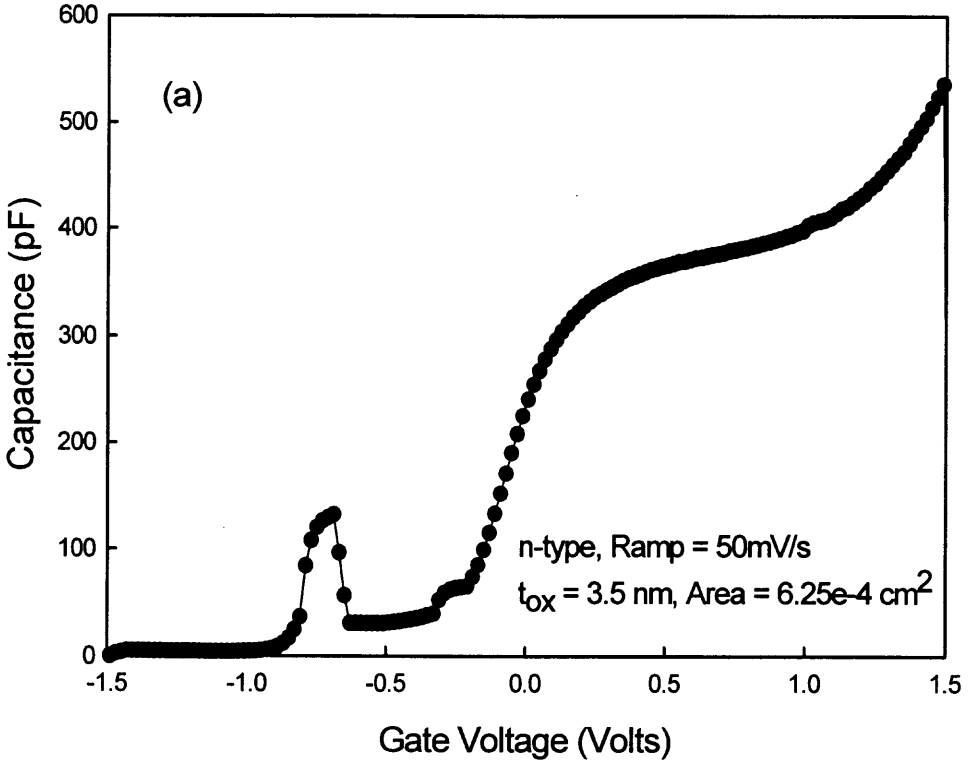

Figure 3.6 Severely distorted Low Frequency C-V curve of a 3.5 nm oxide due to leakage current.

This problem can be overcome by using a new measurement method that makes use of electronic circuitry to compensate for the leakage current. As will become clear from the rest of this chapter, this measurement method produces true static C-V curves, even for oxides as thin as the ones that have been discussed above.

The principle of a leakage compensated charge measurement is illustrated in Figure 3.7. A small amplitude square wave signal $V_g(t)$ is applied across a MOS capacitor at DC bias level V_G. The total device current contains both a displacement current component responding to the square wave and a DC tunnel leakage current component. The device current is integrated, with the DC component being eliminated using a feedback leakage compensation scheme. The output of the circuit V_0 is thus a transient charge response waveform, Q(t) generated by the applied small excitation signal. If the response waveform is allowed to saturate, then the amplitude of the waveform at a given bias is proportional to the static device capacitance, that is $C(V) = Q_{max}(V)/V_g$, where V_g is the amplitude of the excitation square wave and Q_{max} is the maximum or saturation value of the transient charge response curve. Therefore, a C-V curve is obtained by measuring Q_{max} as a function of a varying bias voltage that sweeps the MOS structure from accumulation, through depletion into inversion, or vice versa.

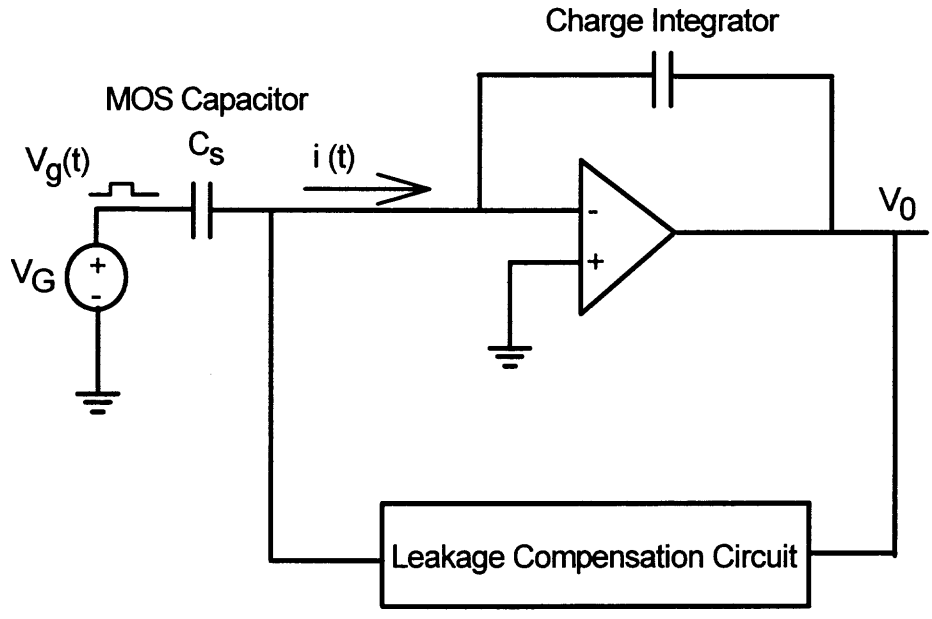

Figure 3.7 Principle of a Leakage Compensated Charge Measurement.

3.5 Experimental Details

To fabricate the devices used in this work, active SiO_2 tunnel oxides of thickness 2.4, 2.8 and 3.5 nm are formed at 700 °C in dry O_2 in windows opened in a field oxide grown on <100> oriented, 0.005-0.05 ohm-cm, n-type silicon wafers. The active oxide is then immediately covered by depositing polycrystalline silicon, which is phosphorous doped approximately to degeneracy. Devices exhibit current-voltage (I-V) characteristics which scale well both with device area and oxide thickness. Experimental data were acquired using a standard personal computer that was equipped with a National Instruments AT-MIO-16E-10 analog-digital/digital-analog converter board. LabVIEW was used to develop fully automated measurement procedures. In particular, LabVIEW-based programs were developed for the generation of the analog excitation signal and the

acquisition of the digital response waveform. The test devices were mounted in a HP16055A test box. High frequency C-V curves were measured using a HP4285 LCR meter.

3.6 Experimental Results

3.6.1 LCCV on a 3.5 nm Oxide

This first section will discuss the results of a leakage compensated charge measurement on an MOS structure with a 3.5 nm oxide. Figure 3.8 illustrates the transient waveforms Q(t) in response to a 50 mV square waveform with a frequency of 1 Hz. As is clear from the figure, the shape and amplitude of the output waveform vary with V_G, for example displaying a slow rise time in inversion where the minority carrier response time is long, and conversely a fast rise time in accumulation where majority carriers can easily follow the applied square wave signal. It must also be pointed out that the saturation of minority carrier response is easily obtained as can be seen from the flatness at the end of each section of the charge waveform. This assures that the measured curve is a static C-V curve, instead of "quasistatic" as in the conventional approach.

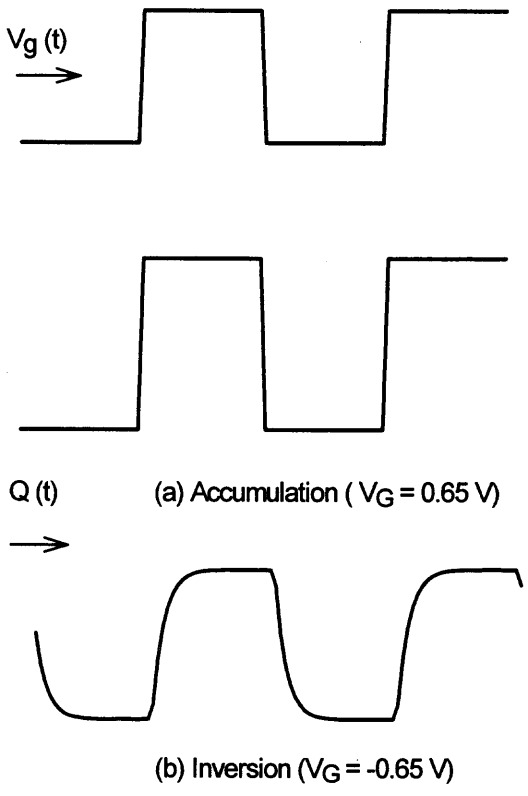

Figure 3.8 Transient waveforms Q(t) in accumulation and inversion for a 50mV, 1 Hz excitation signal on a 3.5nm oxide.

Figure 3.9 is a 3-D graph of these transient waveforms Q(t) as a function of applied bias voltage V_G, with the "front axis" being V_G and the second planar axis being time. The transient waveforms Q(t) are represented by the dashed curves. The "right" part of the graph depicts the device in accumulation as can be seen from the square wave transient responses. The "left" part depicts the device in inversion as can be seen from the slow rise time of the waveform. It is clear that the amplitude of the waveform in the center of the graph is significantly smaller than that in accumulation or inversion. This is expected since this part of the graph corresponds to the depletion regime of the device. As explained in the previous section, the amplitude of the waveforms is proportional to the capacitance (the vertical axis). Against the backwall of the graph these amplitudes

are traced out for each bias voltage (the solid curve). This curve, therefore, traces out the static C-V curve of the device.

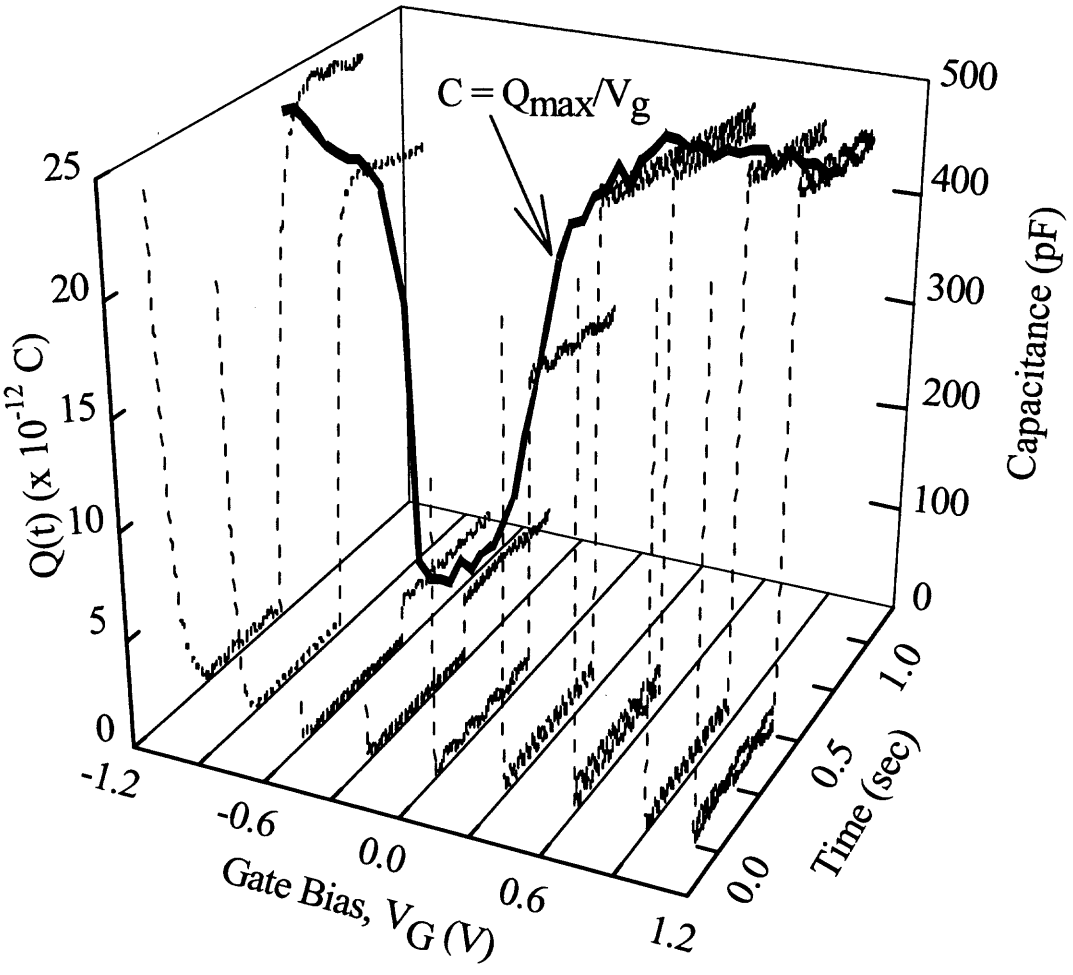

Figure 3.9 3-D graph of transient waveforms Q(t) as a function of bias voltage V_G.

The C-V curve displays the expected static characteristics. In accumulation and inversion, the capacitance saturates at a value that is typically the oxide capacitance, but can also include quantum effects due to the finite width of the inversion layer as well as polycrystalline silicon depletion when the substrate is in accumulation. As expected, the

capacitance goes through a minimum when the silicon is swept through depletion. It is quite remarkable that this C-V curve can be obtained despite the presence of a large leakage current.

The expected transition from a low frequency C-V characteristic to high frequency behavior can be demonstrated by varying the frequency of the excitation square wave signal. If the frequency is too high, then the minority carrier response in inversion does not saturate, and thus the capacitance extracted from the the values of Q_{max} in this bias region will not be the saturation value. This effect is shown in Figure 3.10, where the minority carriers in the inversion layer are unable to reach equilibrium at frequencies less than ~2 Hz, and do not respond at all above ~10 kHz.

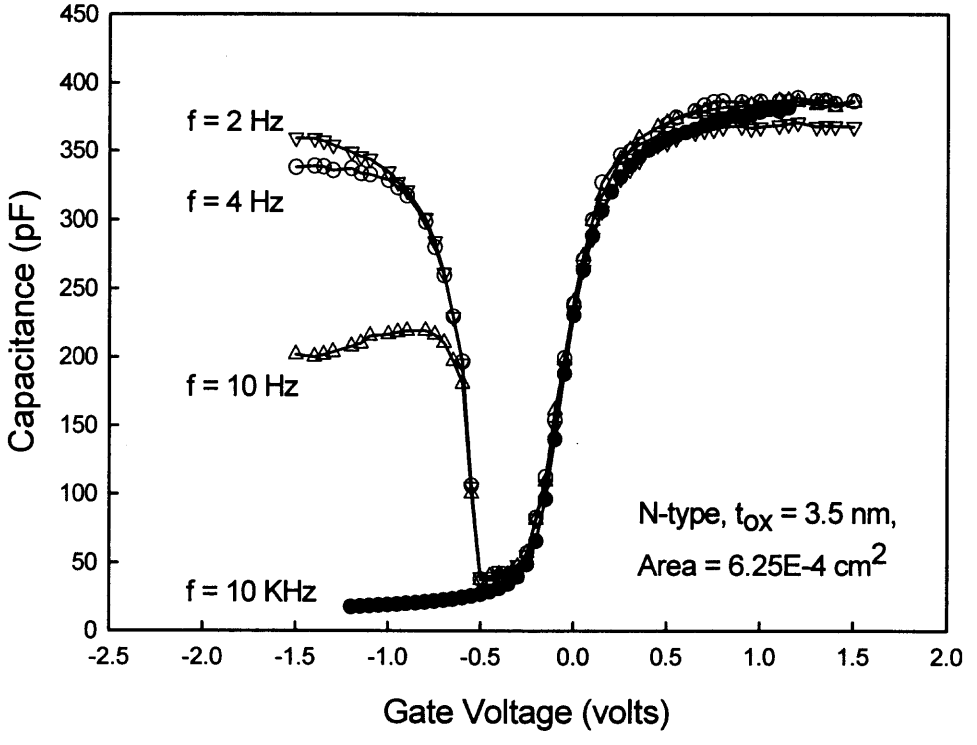

Figure 3.10 Frequency dependence of C-V characteristics of 3.5 nm oxide measured by LCCV method.

3.6.2 LCCV on a 2.4 nm Oxide

This section presents the results of the LCCV method on oxides with a thickness of 2.4 nm. For oxides this thin, the tunnel leakage becomes so large that the circuitry alone is not sufficient to compensate all of the device leakage. In this situation, the transient response curves do not saturate as in Figure 3.8, but rather they display the behavior presented in Figure 3.11. In both cased, response curves are shown for five different excitation signals amplitudes ranging from 30 mV (solid curves) to 70 mV (dotted curves). The charge transient amplitude for a 2.4 nm oxide increases indefinitely, varying linearly with time once "saturation" is achieved.

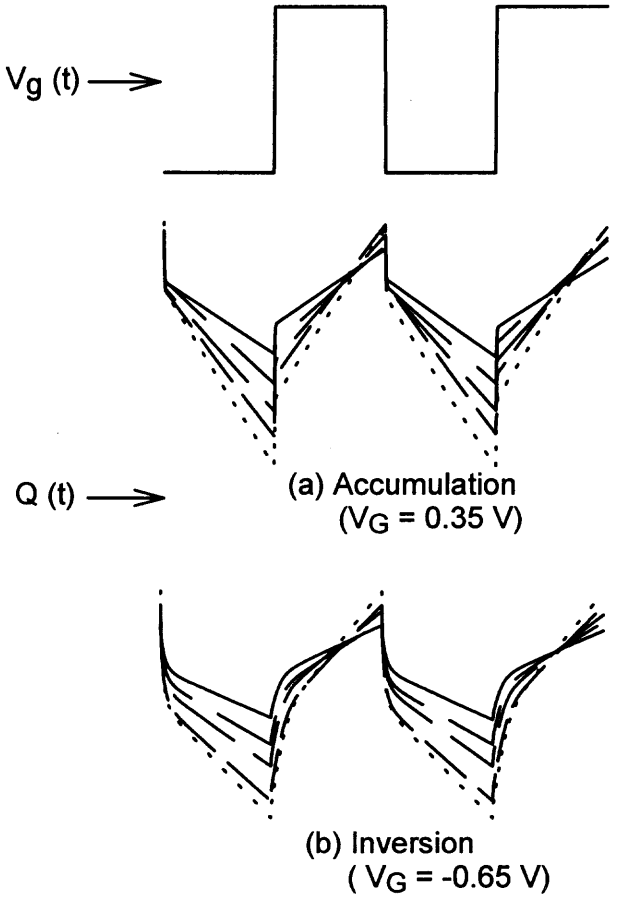

Figure 3.11 Transient waveforms Q(t) in accumulation and inversion for a 2.4nm oxide.

The slope of this variation, dQ/dt is found to increase linearly with the excitation pulse amplitude, as shown in Figure 3.12. We term dQ/dt an "excess" leakage current since it is not accounted for by our circuitry.

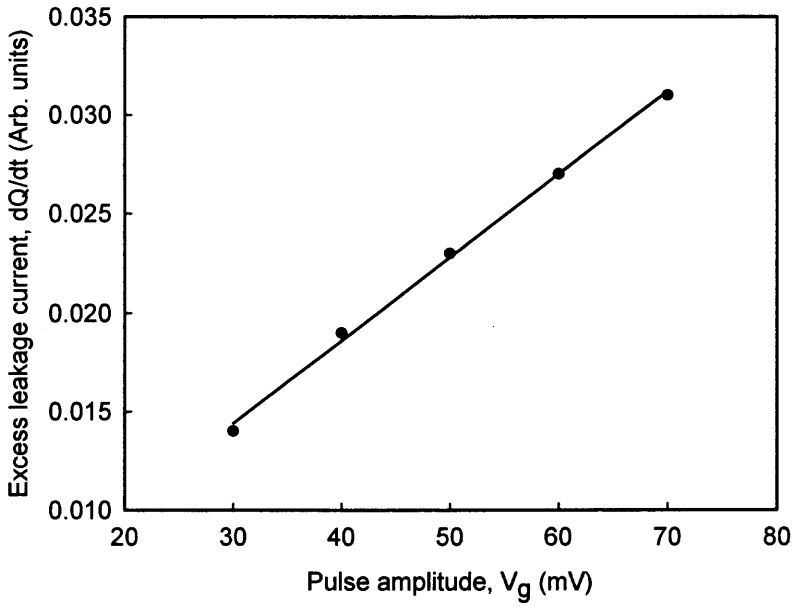

Figure 3.12 Excess leakage current dQ/dt as a function of pulse amplitude V_g.

Because of the straightforward relationship with the excitation pulse amplitude, the excess leakage is easy to account for numerically. Independent of our choice of excitation pulse height, subtracting the linearly increasing component of the response curve leads to recovery of the "normal" behavior of Figure 3.8. It is important to note that the total tunnel leakage through this oxide can be as high as 10^{-7} A, while a quasistatic capacitance measurement would have to detect a displacement current on the order of ~1 pA. Thus our charge compensation circuitry has eliminated most of the leakage current, and left only a small "excess" leakage to be dealt with numerically.

In Figure 3.13 we present a high frequency C-V for the 2.4 nm oxide as well as the static C-V's obtained both without and with numerically compensating the "excess" leakage. The fully compensated static curve (solid triangles) shows that even at the measurement frequency of 2 Hz, which was sufficient to achieve minority carrier equilibrium in 3.5 nm oxides, the minority carriers in the thinner oxide device cannot yet achieve equilibrium. This might be expected because of the higher fields present in the thinner system. More importantly, the fully compensated data show that it is possible to measure static capacitance even in oxides as thin as 2.4 nm. Clearly from the high and low frequency curves presented for this device it will be possible to use the leakage compensated charge method to extract defect and interface state information for devices having such a thin oxide.

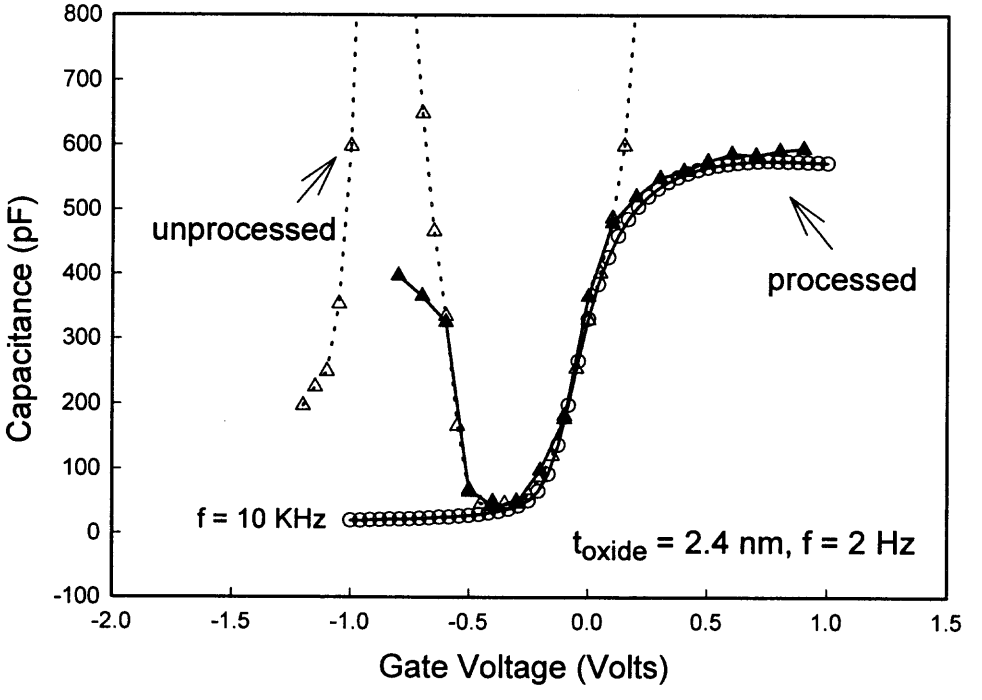

Figure 3.13 C-V characteristics of 2.4 nm oxide before and after numerical processing of the excess leakage current.

3.7 Summary

In this chapter a charge measurement method has been described to determine C-V characteristics of ultrathin silicon oxide dielectric, MOS device structures in which significant leakage current is present. The leakage is accounted for using a current compensation circuitry.

C-V curves have been measured for oxides between 2.4 and 3.5 nm thin. From the results on 3.5 nm oxides it is found that the minority carrier response saturates easily, making this a true, static C-V measurement as opposed to the standard quasistatic technique. In addition, the expected frequency dependence of the capacitance has been observed. This allows the classical high frequency - low frequency capacitance (HLCV) method for obtaining the interface state level.

In the thinnest oxides (2.4 nm), an "excess" leakage current is observed that can not by compensated for by the circuitry alone. However, the excess leakage current is found to increase linearly with the excitation signal amplitude and can therefore be easily compensated for by additional numerical processing. The significance of these findings lie in the fact that this measurement method extends the range of application of capacitance characterization of silicon oxides well into the direct tunnel thickness regime.

CHAPTER 4

CHARACTERIZATION OF HAFNIUM OXIDE GATE DIELECTRICS

4.1 Overview

The rapid shrinking of the MOSFET feature size has forced the gate dielectric thickness to decrease simultaneously in order to increase device performance. The current CMOS gate dielectric, SiO_2, can probably scale to a thickness of the order of 10 Å. Further scaling of the SiO_2 thickness will not be possible. In addition, dielectrics that are this thin exhibit a large direct tunnel leakage current. This may not be a prohibiting factor for improvement of high performance devices (desktops), but certainly is a concern for low power (laptops) applications, where the leakage current is a drain on battery life. One solution may be replacing the gate dielectric of choice, SiO_2, with a material that has a higher dielectric constant. Such an alternative gate dielectric would allow a thicker physical gate dielectric to be used without a reduction in oxide capacitance, and therefore performance.

This chapter will first explore the possible candidates to replace SiO_2 and their materials properties. It will show that one particularly promising candidate is HfO_2. The rest of the chapter is devoted to our own exploration in this field, the fabrication and characterization of thin HfO_2 films that are deposited by Metal Organic Chemical Vapor Deposition (MOCVD) and Alternating Layer Chemical Vapor Deposition (ALCVD). These techniques were chosen because they offer a high level of controllability over the deposition process.

4.2 Materials Properties Considerations

It is mentioned in the introduction above that an alternative gate dielectric with a higher dielectric constant than SiO_2 allows the physical thickness of the gate dielectric to be larger without losing device performance. This can be understood by considering the gate dielectric as a parallel plate capacitor. It must be noted that this is an accurate assumption as long as quantum effects and polysilicon depletion from the substrate and gate are small enough [20]

$$C = \frac{\kappa \cdot \varepsilon_0 \cdot A}{t_{ox}} \qquad \text{(Equation 4.1)}$$

Where κ is the relative dielectric constant (3.9 for SiO_2), ε_0 the permittivity of free space (8.85x 10e-11 fF/um), A the area under the gate dielectric and t_{ox} the thickness of the gate dielectric. It is clear from this equation that a layer of a material with a dielectric constant that is, for instance, 5 times larger than SiO_2, can have a physical thickness that is 5 times larger in order to get the same capacitance. In other words, a 50 Å layer of a material with a dielectric constant equal to 20 will have an equivalent oxide thickness t_{eq} of 10 Å.

Selecting a gate dielectric with a higher permittivity than that of SiO_2 is clearly essential. It is expected that the permittivity of the gate dielectric of choice will have to be larger than 15, although a lower permittivity may be a short term solution in the absence of a material with a high permittivity. The required permittivity must be balanced, however, against the barrier height for the tunneling process. For electrons travelling from the silicon substrate to the gate, the barrier height is given by the conduction band offset: $\Delta E_c = q(\chi - (\Phi_M - \Phi_B))$. For electrons travelling from the gate to

the substrate it is simply Φ_B. Since direct tunnel leakage current exponentially increases with decreasing barrier height, it is essential that the bandgap and conduction band offset are sufficiently high.

A third requirement is thermal stability of the alternative gate dielectric on silicon, i.e. they do not react with the substrate to form an undesirable interfacial layer. Such a layer may improve the interfacial quality of the dielectric with the substrate, but it severely limits the maximum attainable gate stack capacitance and therefore, the minimum attainable t_{eq}. This effect of reduced capacitance can be seen by noting that, when the structure contains several dielectrics in series, the lowest capacitance layer will dominate the overall capacitance. The largest benefit of using SiO_2 as the underlayer of a stack (at the channel interface) is that the unparalleled quality of the SiO_2 – Si interface will help maintain a high channel carrier mobility. However, this would introduce a second interface between the underlayer and the high-k dielectric that can potentially be another source of charge trapping.

A clear goal of any potential high-k gate dielectric is to attain a sufficiently high-quality interface with the silicon substrate, as close as possible to that of SiO_2. It is difficult to imagine any material to have a better interface, since typical CMOS SiO_2 gate dielectrics have a midgap interface state level density of 10^{10} eV^{-1} cm^{-2}. Almost all high-k materials reported to date have an interface state density of 10^{11}, 10^{12} eV^{-1} cm^{-2} and exhibit a flatband voltage shift larger than 300 mV, caused by a fixed dielectric charge in excess of 10^{12} eV^{-1} cm^{-2}. If alternative gate dielectrics are ever to be employed in device structures, it is crucial that the interfacial properties are significantly improved.

Most of the advanced gate dielectrics studied to date are either polycrystalline or single crystal films. It may be desirable to select a material which remains in an amorphous phase throughout the necessary processing treatments. Polycrystalline gate dielectrics may be problematic because grain boundaries serve as high-leakage paths and this may lead to the need for an amorphous interfacial layer to reduce leakage current. Single crystal oxides grown by MBE or ALCVD methods can in principle avoid grain boundaries while providing a good interface, but these materials also require submonolayer deposition control. Given the concerns regarding polycrystalline and single crystal films, it appears that an amorphous film structure is the ideal one for the gate dielectric.

A significant issue for integrating any advanced gate dielectric into standard CMOS is that the dielectric should be compatible with silicon-based gates, rather than *require* a metal gate. Silicon-based gates are desirable because dopant implant conditions can be tuned to create the desired threshold voltage for both NMOS and PMOS, and the process integration schemes are well established in industry. Nearly all of the potential advanced gate dielectrics investigated to this point, however, require metal gates. This is expected because the same instability with silicon will exist at both the channel and the poly-Si gate interfaces. Metal gates are very desirable for eliminating dopant depletion effects.

There are two basic approaches toward achieving successful insertion of metal electrodes: a single midgap metal or two separate metals for NMOS and PMOS. The first approach is to use a metal, such as TiN, that has a work function that places its Fermi level at the midgap of the Si substrate. These are generally referred to as ‘‘midgap

metals.'' The main advantage of employing a midgap metal arises from a symmetrical value of the threshold voltage for both NMOS and PMOS, because by definition the same energy difference exists between the metal Fermi level and the conduction and valence bands of Si. The second main approach toward metal electrodes involves two separate metals, one for PMOS and one for NMOS devices. Two metals could be chosen by their work functions, Φ_M, such that their Fermi levels line up favorably with the conduction and valence bands of Si, respectively.

A crucial factor in determining the final film quality and properties is the method by which the dielectrics are deposited in a fabrication process. The deposition process must be compatible with current or expected CMOS processing. Alternative dielectrics have been deposited through a variety of processes, such as PVD (sputtering), MOCVD, MBE and ALCVD. Especially films deposited using ALCVD have shown very promising electrical and structural properties, largely due to the submonolayer deposition control that is attained by ALCVD.

A final factor that is important to the selection of the appropriate gate dielectric is dielectric reliability. To date, almost all gate dielectrics have breakdown voltages that are too low (< 5 MV/cm) for practical use.

4.3 Status of Alternative Gate Dielectrics

Table 4.1 is a compilation of potential high-k candidates and their materials properties. It is immediately clear that all the alternative gate dielectrics have a bandgap and a conduction band offset smaller than SiO_2. This is important to point out, since both of these properties are a measure of barrier height. A lower barrier height enhances direct tunneling through the dielectric, and therefore, part of the advantage of using a thicker dielectric will be eliminated due to increased tunnel leakage.

Table 4.1 Properties of the most studied alternative gate dielectrics.

Material	**Dielectric Constant**	**Bandgap (eV)**	**Band Offset to Si (eV)**	**Morphology**
SiO_2	3.9	8.9	3.2	Amorphous
Si_3N_4	7	5.1	2	Amorphous
Al_2O_3	9	8.7	2.8	Amorphous
Ta_2O_5	26	4.5	1-1.5	Orthorhombic
La_2O_3	30	4.3	2.3	Hexagonal/Cubic
TiO_2	80	3.5	1.2	Tetragonal
ZrO_2	25	7.8	1.4	Tetragonal/Cubic
HfO_2	25	5.7	1.5	Tetragonal/Cubic

Source: G.D Wilk, R.M Wallace, J.M. Anthony "High-k gate dielectrics: Current status and materials properties considerations," *J. Appl. Phys.*, **89**, 10 (2001) [34].

Many of the materials initially chosen as potential alternative gate dielectric candidates were inspired by memory capacitor applications. [42] The most commonly studied high-κ gate dielectric candidates had been materials systems such as Ta_2O_5,[36-

41], $SrTiO_3$,[42-44] and Al_2O_3,[45-49] which have dielectric constants ranging from 10 to 80, and have been employed mainly due to their maturity in memory capacitor applications.

With the exception of Al_2O_3, however, these materials are not thermodynamically stable in direct contact with silicon. This thermodynamic stability is not a requirement for memory capacitors, since the dielectric is in contact with the electrodes. Oxynitrides and oxide/nitride reaction barriers between these high-κ metal oxide materials and silicon in an attempt to prevent or at least minimize reaction with the underlying silicon. It is important to note, however, that using an interfacial layer of SiO_2 or another low permittivity material, will limit the highest possible gate stack capacitance, or equivalently, the lowest achievable t_{eq}.

Alumina (Al_2O_3) is a very stable and robust material, and has been extensively studied for many applications. Regarding its usefulness as an alternative gate dielectric, Al_2O_3 has many favorable properties, as shown in the table above, including a high band gap (8.7 eV), thermodynamic stability on Si up to high temperatures, and is amorphous under the conditions of interest. The drawback is that alumina only has a dielectric constant of 8 to 10 making it a relatively short-term solution.

Atomic layer CVD (ALCVD) Al_2O_3 has been studied both physically and electrically, in particular to better understand the interface formed between Si and Al_2O_3 deposited by this technique [50]. It was shown that using this technique Al_2O_3 could be deposited *without* forming an interfacial SiO_2 layer. In addition, transistor data, revealed a leakage current density of 0.1 A/cm2 for an equivalent oxide thickness of 13 Å, showing a reduction in leakage current of two orders of magnitude. The channel carrier

mobility was found to be a factor two smaller. Furthermore, a flatband voltage shift on the order of 700 mV was observed corresponding to a negative fixed oxide charge of 10^{12} /cm^2. Most high-k films show a large flatband voltage shift.

A substantial amount of investigation has gone into the metal oxides TiO_2,[51-53], ZrO_2,[54-57] and HfO_2,[58-65] as these systems have shown much promise in overall materials properties as candidates to replace SiO_2. The TiO_2 system has been heavily studied for high- k applications both for memory capacitors and in transistors. It is attractive because it has a high permittivity of 80– 110, depending on the crystal structure and method of deposition. Full transistors using CVD TiO_2 as the gate dielectric displayed relatively large interface state density (10^{12}/cm^2 eV) and leakage currents were also unacceptably high in the transistors.

Encouraging results of both CVD and sputtering of ZrO_2 and HfO_2 have been reported by others. Using ALCVD highly uniform layers as thin as 20 A have been deposited on a very thin layer of SiO_2. The thin oxide layer was intentionally grown in a thermal anneal. This layer serves a dual purpose by providing a high quality interface and at the same time a reactive surface on which to deposit the dielectric. Very low leakage currents of 10^{-4} A/cm^2 were reported and C-V characterization revealed hysteresis effects on the order of 80 mV, indicating the presence of charges in the films. Additionally, low electrical breakdown fields of 4 MV/cm were found.

More recent work has been reported on another class of candidates, so called metal silicates, such as $Hf_xSi_yO_z$ or $Zr_xSi_yO_z$, indicating that such materials exhibit encouraging gate dielectric properties [66-71]. Both materials have the same underlying principle of mixing a high-k crystalline metal oxide such as HfO_2 or ZrO_2 with an

amorphous, stable lower-k materials such as SiO_2 to obtain a morphology with suitable properties for a CMOS gate dielectric. The effect of adding SiO_2 to metal oxides is to produce an amorphous film that is thermodynamically stable on silicon. The overall permittivity of the alloy is inevitably lower than that of the pure metal oxide, but this tradeoff can be adequate for better stability.

It is tempting to wonder though, whether the work on silicates is an attempt at improving the dielectric properties of pure metal oxides simply by making them "look more like SiO_2". It can surely be expected that by adding SiO_2 to pure metal oxides, important properties such as interface quality will improve, since the silicate has a chemical composition closer to SiO_2 than a pure metal oxide.

4.4 Experimental Details

The devices that have been investigated in this research were largely fabricated in the NJIT cleanroom facility. The deposition of the hafnium oxide was done in the reactor of Structured Materials Industries, Inc. Windows were opened in a field oxide grown on <100> oriented, 0.005-0.05 ohm-cm, n-type silicon wafers. Subsequently, the hafnium oxide was deposited in a Rotating Disk Reactor Chemical Vapor Deposition (RDR-CVD) system. The deposition was carried out at a temperature of 350°C, a chamber pressure of 20 Torr and a deposition time of four minutes. The precursor used during deposition was hafnium-tertiary-butoxide ($Hf(C_4H_9O)_4$)) while the oxidizer used was O_2. Two modes of deposition were evaluated for this study: MOCVD, where precursor and oxidizer flowed simultaneously in the reactor, and ALCVD, where the precursor and oxidizer alternately flowed with a 50/50 duty cycle or a 90/10 duty cycle. More detailed deposition parameters can be found in Appendix A. Upon finishing the deposition, patterned front

side aluminum contact pads were sputtered on. The final step in the process was aluminum sputtering of the backside of the wafer.

All electrical measurements were done using an HP4285 LCR meter and an HP4140B pA meter. Devices were mounted for testing in a probe station. The data were acquired using a standard personal computer, equipped with a GPIB data acquisition card. In addition, measurements were fully automated using LabVIEW programming software.

4.5 Results and Discussion

4.5.1 Hafnium Oxide Thickness

Figure 4.1 and 4.2 show images of an MOCVD wafer after deposition (Figure 4.1) and an ALCVD wafer after deposition (Figure 4.2). It is clear that the ALCVD wafer looks a little darker than the MOCVD wafer, indicating a thicker deposited layer. While this was generally found to be true, it is more important to notice the concentric ring in each of the images. This concentric ring is the area of the wafer where significant, measurable hafnium oxide deposition took place. Naturally, this suggests a strong thickness nonuniformity across the wafer. While an optimised process that produces thickness uniformity obviously is desirable, it does not prevent us from characterizing the wafers. It merely limits the number of devices available for characterization since the devices that lie outside the concentric ring lack sufficient deposited hafnium oxide.

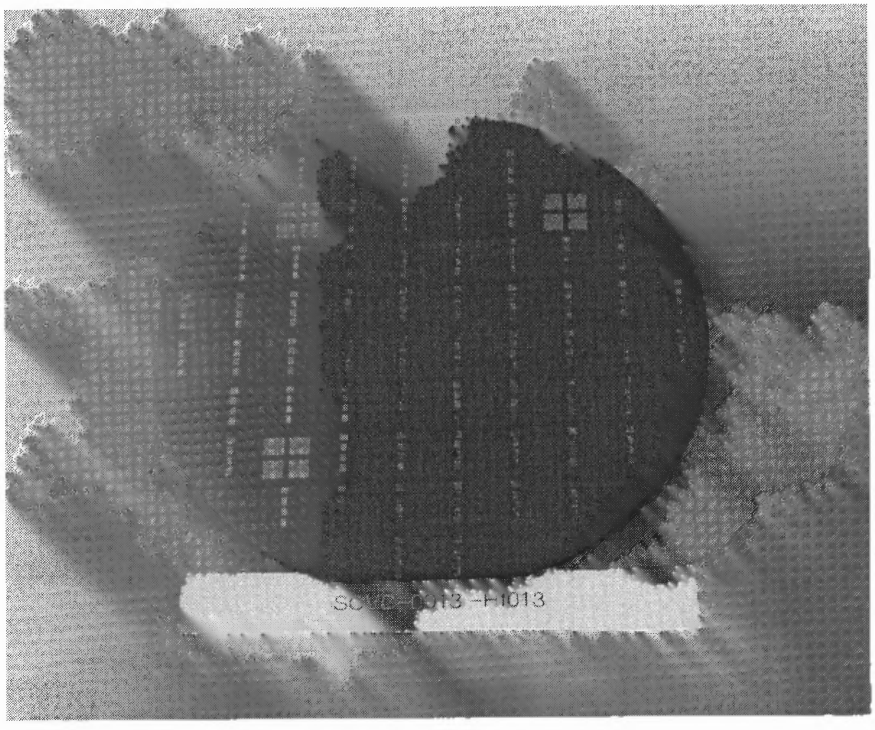

Figure 4.1 Picture of wafer after MOCVD deposition of hafnium oxide.

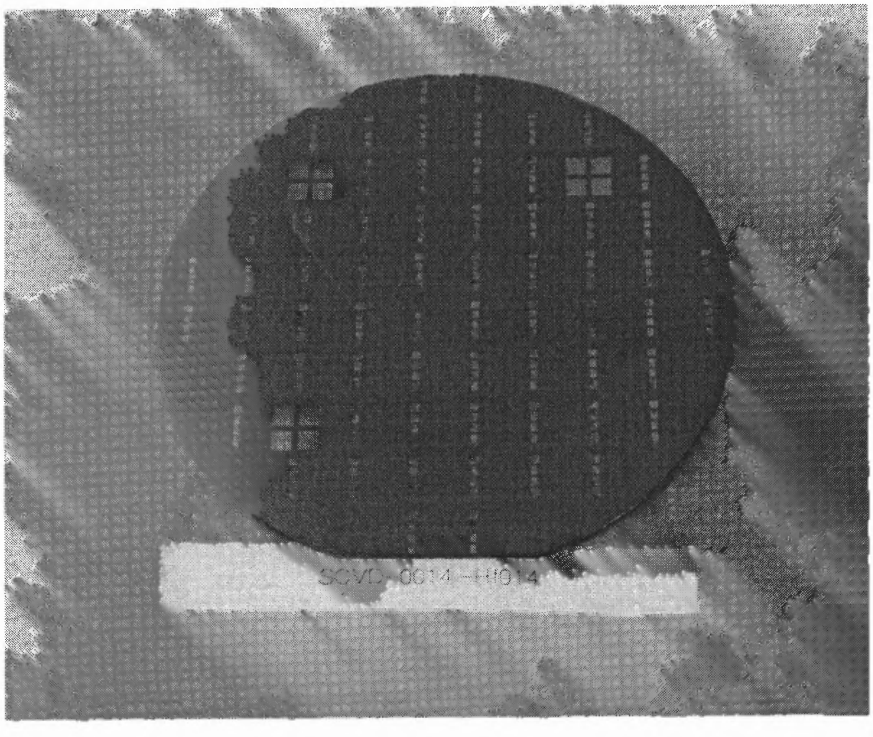

Figure 4.2 Picture of wafer after ALCVD deposition of hafnium oxide.

Ellipsometry was performed on the wafers to further investigate the thickness of the deposited layers. This revealed that the ALCVD wafers indeed had thicknesses that were typically on the order of 20 Å higher than the MOCVD wafers as was already suspected from the slight color difference. However, it is more important to note that the thickness variation *within* the concentric rings was much larger than the thickness variation *between* the wafers. The thickness of the deposited hafnium oxide layer on the MOCVD wafer ranged from 110 Å to 280 Å. On the ALCVD wafer, the thickness within the concentric ring varied between approximately 130 Å and 300 Å.

4.5.2 Electrical Characterization

Capacitance-voltage and current-voltage measurements were performed on devices from both wafers. Figure 4.3 displays two high frequency C-V curves taken on a 10x10 μm^2 device with a hafnium oxide thickness of 140 Å. The first curve was measured from –3 V to +3 V, the second subsequently from +3 V going back to –3 V. The first observation is that the C-V curve is a well behaved curve, i.e. it shows the accumulation regime, a decreasing capacitance in depletion and finally a leveling off at a low capacitance in inversion. It becomes immediately clear that the curves display a hysteresis in their capacitance behavior. Undoubtedly this is due to trapping and detrapping of charge carriers in the interfacial region and possibly the bulk of the hafnium oxide. The flatband voltage shift can be estimated to be on the order of –600 mV, indicating a presence of a fixed positive oxide charge. While one prefers not to observe phenomena such as hysteresis and a relatively large shift in the flatband voltage, it is important to point out that the observations as described above are common in alternative gate dielectrics.

In accumulation the capacitance is seen to level off at a value of 20 pF. From the parallel plate approximation $C = \kappa \varepsilon_0 A / t_{ox}$, one can calculate the dielectric constant:

$\kappa = 13$,

where C = 20 pF, A = 100 μm^2, t_{ox} = 140 Å and ε_0 = 8.85x 10^{-15} fF/μm. A dielectric constant of 13 for hafnium oxide devices is on the low end of what has been reported [72-74]. The is most likely due to the formation of a thin SiO_2 layer at the silicon – hafnium oxide interface. In addition, a low dielectric constant of 13 suggests that the hafnium oxide film is not of a pure HfO_2 nature. Most likely the film has a lower density than one would expect for a true HfO_2 dielectric.

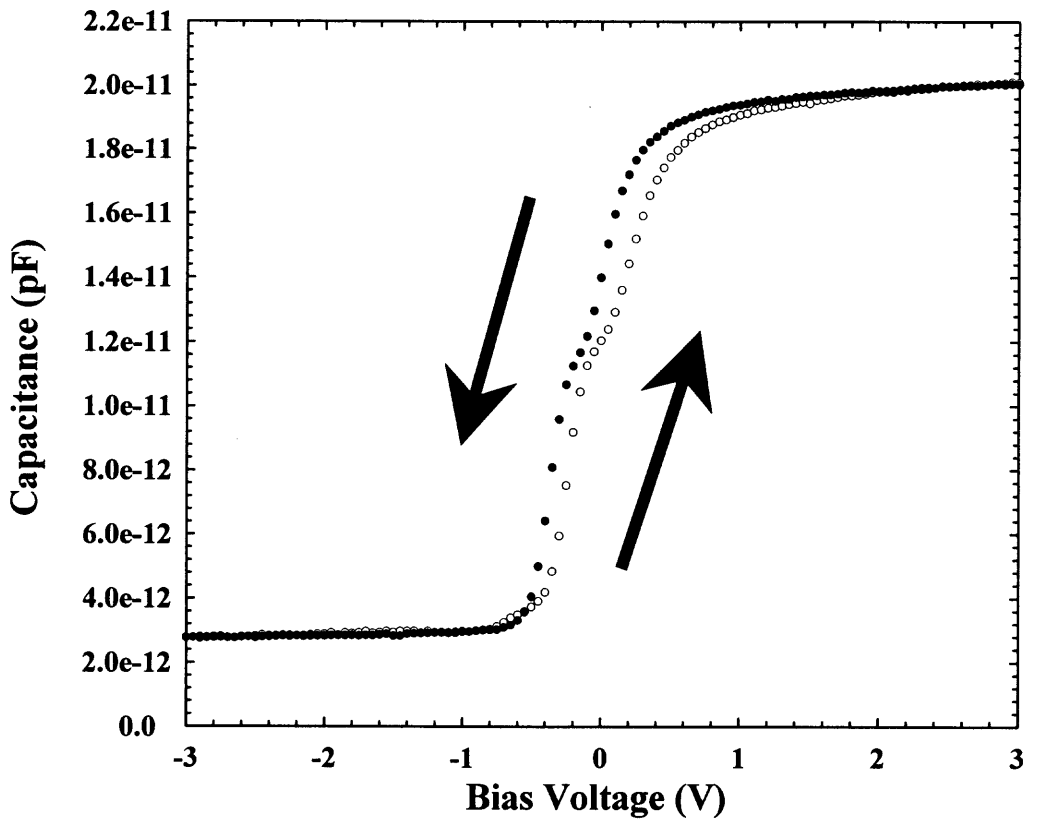

Figure 4.3 C-V characteristics of a 140 Å hafnium oxide film.

Figure 4.4 shows the current-voltage characteristics of a 10x10 μm^2 device on the same chip as the previous device. This device can thus be expected to have a hafnium oxide thickness of approximately 140 Å as well. It is immediately apparent that the current is rather high considering this is a relatively thick oxide. Quantum mechanical tunneling is virtually nonexistent for an oxide that has a thickness of 140 Å. The rapid increase in the current at low voltages is probably due to electrons traversing the hafnium oxide from the gate to the silicon substrate by hopping from bulk trap to bulk trap. As became clear from the C-V characteristics in the previous figure, these devices have a large trap density, and therefore such a hopping scenario is likely. The additional "hump" that is observed at 3.5 V is probably related to defects at the interface. The scattering that starts at 5.5 V is the onset of breakdown of the dielectric and at 6 V the dielectric has broken down. For this particular oxide, this translates into a breakdown field of 3.5 MV/cm, again a value that is not uncommon to what others have reported for hafnium oxides deposited by chemical vapor deposition[72-74].

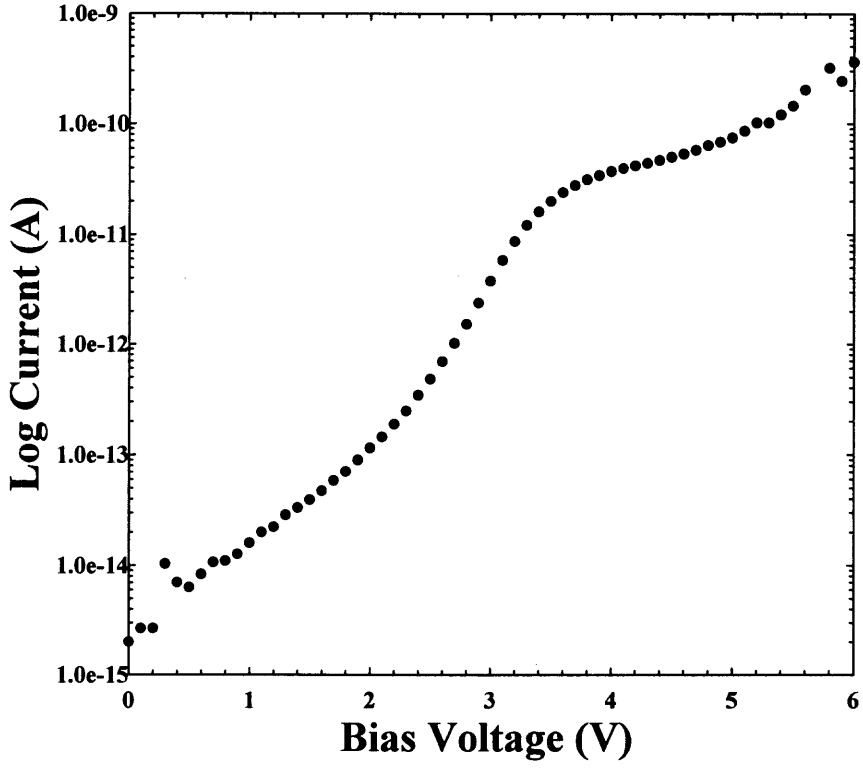

Figure 4.4 I-V characteristics of a 140 Å hafnium oxide film.

Figure 4.5 shows sequential current-voltage characteristics on the same device as the previous figure. The current was measured as a function of bias voltage for three consecutive runs from 0 V to + 4 V and back to 0 V. A measurement like this gives insight into the charging dynamics of the device. Since these devices have large trap densities it is expected that significant charging will take place. The initial current response (solid triangles up) to an increasing bias voltage is the same as the one shown in the previous figure. Note how the "hump" is also clearly visible in this graph. On reducing the bias voltage back to 0 V (open triangles up), the current is less than during the first run up. This is evidence that the device has become charged negatively during

its initial run up. On the second run to + 4 V (solid circles) and back to 0 V (open circles) the same effect is observed, but less pronounced. Note how both currents during the second run are lower than those of the first run. This indicates that during the first run traps in the bulk oxide and possibly the interface have filled with electrons and the device as a whole stayed negatively charged after the first run. During the third and final run the currents during the positive ramp and the negative ramp more or less coincide. This indicates that the charging of the device reaches a saturation. Subsequent runs would only reinforce that view.

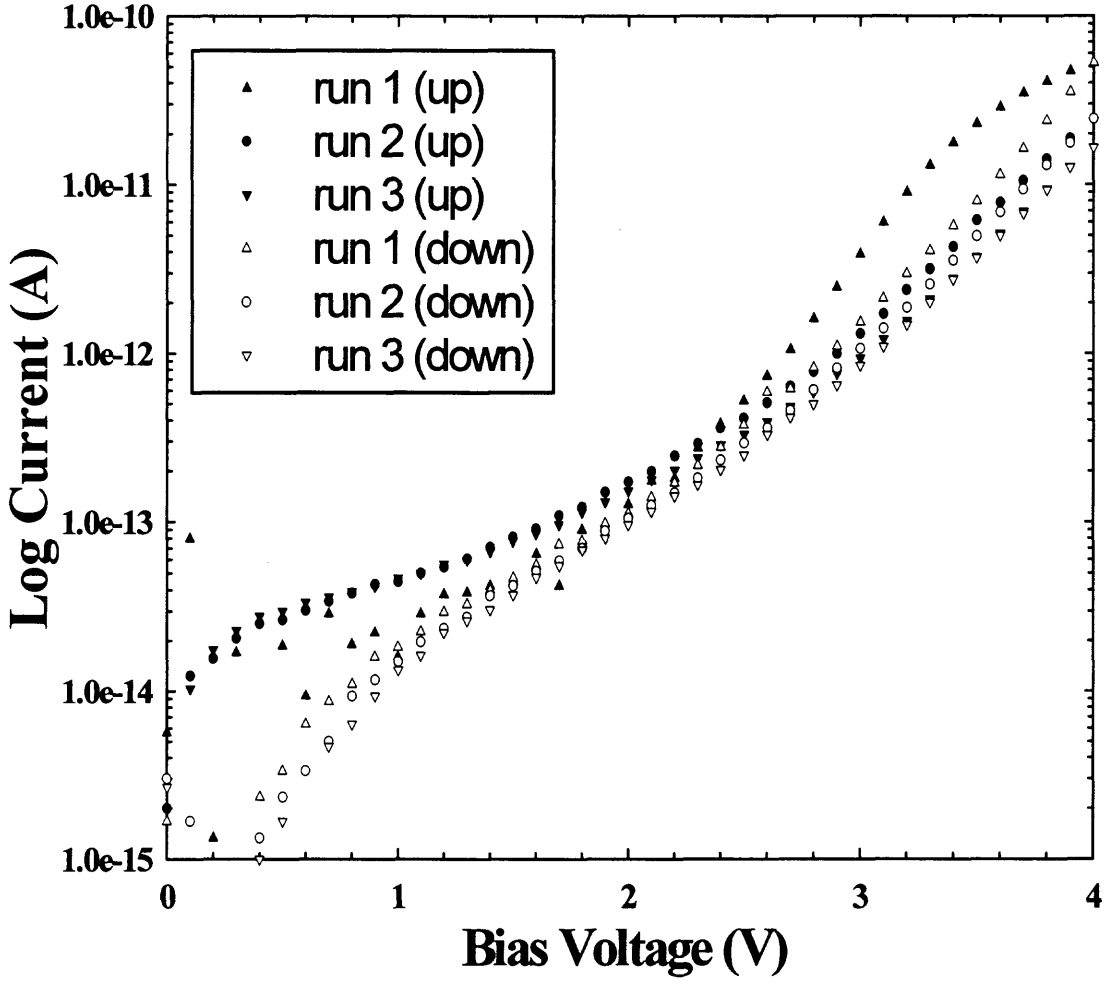

Figure 4.5 Sequential I-V runs showing the charging dynamics of the dielectric.

4.6 Summary

This chapter describes the fabrication and characterization of thin hafnium oxide dielectrics. Hafnium oxide was deposited using a Rotating Disk Reactor CVD system. Films were deposited by an MOCVD and ALCVD process. Both processes were found to yield essentially the same results. The deposition produced a strongly nonuniform film with significant, measurable hafnium oxide concentrated in a concentric ring. The range of thicknesses within the concentric ring was found to be between 110 Å to 280 Å.

Capacitance-voltage characterization showed the device to have electrical characteristics such as hysteresis and a large flatband voltage shift that is commonly found in materials such as the one that was investigated in this work. Additionally, a dielectric constant equal to 13 was found from the oxide capacitance. Current-voltage characteristics reveal a large leakage current caused by a high trap density in the bulk oxide and significant negative charging effects.

The electrical characteristics that are reported here suggest on the one hand the presence of an interfacial SiO_2 layer that has an adverse effect on the dielectric constant. In addition, it must be concluded that the physical character of the hafnium oxide film is most likely of a less than pure HfO_2 nature.

This work entailed a possible solution to overcome the presence of a large direct tunnel leakage current in conventional silicon dioxides by replacing it with a material with a higher dielectric constant. This work, and the knowledge of what others have reported, leave the author believing that alternative gate dielectrics may not be the answer. The answer will most likely be found in alternative silicon-based transistor structures, such as vertical transistors or dual-gate transistors.

CHAPTER 5

THE MULTILAYER CHARGE INJECTION BARRIER

5.1 Overview

The ability to grow stacks of alternating layers of ultrathin silicon and silicon dioxide, or MultiLayer Charge Injection Barriers (MLCIBs), opens many new possibilities to develop silicon-based device concepts that were previously impossible to fabricate. MLCIBs consist of alternating, ultrathin layers of silicon and silicon dioxide, with thicknesses typically ranging from 10 Å to 40 Å. At this thickness, the insulator is in the direct tunneling (DT) charge transport regime. These barriers have been used in several electronic applications such as stacked vertical tunnel transistors[75], resonant tunneling structures with the possibility of room temperature operation[76], silicon-based single electron memory structures[77], and even coupled quantum devices with possible uses in quantum computing[78]. Optical studies have also been performed on silicon/silicon dioxide heterostructures but optical applications have not yet been developed [79,80].

To date, no one has presented a detailed study of the electronic properties and in particular the charge transport dynamics of silicon-based MLCIBs. This chapter will fill that void. Two systems have been studied in detail: a double barrier $SiO_2/Si/SiO_2$ on an n-type silicon substrate and the same structure on a p-type substrate.

5.2 Introduction

While the use of multilayer structures in III-V semiconductors is widespread, largely because of the existence of Molecular Beam Epitaxy (MBE), their incorporation in silicon-based devices is not common, due in part to the issue of how to fabricate ultrathin, continuous silicon layers separated by quality tunnel oxides. A number of device concepts have been developed using MLCIBs, with the earliest applications involving resonant tunneling and quantum wells. Early attempts to observe resonant tunneling in silicon-based material focussed on silicon microcrystallites embedded in a matrix of amorphous silicon dioxide [81]. These observations indicated that resonant tunneling might occur at low temperatures. However, producing distributions and sizes of microcrystallites embedded in silicon dioxide in a controllable way proved to be too challenging. Extensive work in silicon based resonant devices has been done by Tsu [82,83] and independently by Seabaugh [76] to form continuous thin layers of *crystalline* silicon using epitaxy. Their work shows that it is possible to form thin continuous layers of silicon sandwiched between layers of amorphous silicon dioxide (MLCIBs), and indications of resonant tunneling have been observed, but not at room temperature. Interface roughness is thought to diffuse the resonant tunneling.

5.3 Double Barrier on n-Type Substrate

This section will present in detail the electronic properties and charge transport dynamics of a double barrier silicon/silicon dioxide structure on n-type silicon.

5.3.1 Device Fabrication

Single crystal silicon wafers doped with ~$3x10^{16}$ phosphorous atoms per cm^3 (n-type) are used as substrates. On each wafer, windows are opened in a field oxide in which an amorphous layer of SiO_2, nominally ~35 Å thick, is grown by a rapid thermal oxidation (RTO) in O_2 at 1050 °C. Without removing the sample to atmosphere, a thin, ~70 Å, undoped layer of Si is then deposited by Rapid Thermal Chemical Vapor Deposition (RTCVD). This layer is amorphous and continuous after deposition, but becomes nanocrystalline during subsequent processing. Next, a second amorphous SiO_2 layer, again ~35 Å thick, is formed at 1050 °C by RTO of the RTCVD Si layer. The growth of this second SiO_2 layer consumes about 44 % of the undoped Si layer. Finally a 2000 Å layer of in-situ doped, nearly degenerate polycrystalline Si, ~10^{20} phosphorous atoms per cm^3, is deposited by RTCVD. Al/Ti gate contact pads are formed by a lift-off process and aluminum is deposited for the back electrode.

Electrical characterization of the diodes is performed using an HP4140B picoammeter/DC voltage source and an HP4284A precision LCR meter/DC voltage source. All data was acquired using a standard personal computer running LabVIEW.

A high-resolution transmission electron micrograph of the barrier portion of the diode is shown in Figure 5.1.

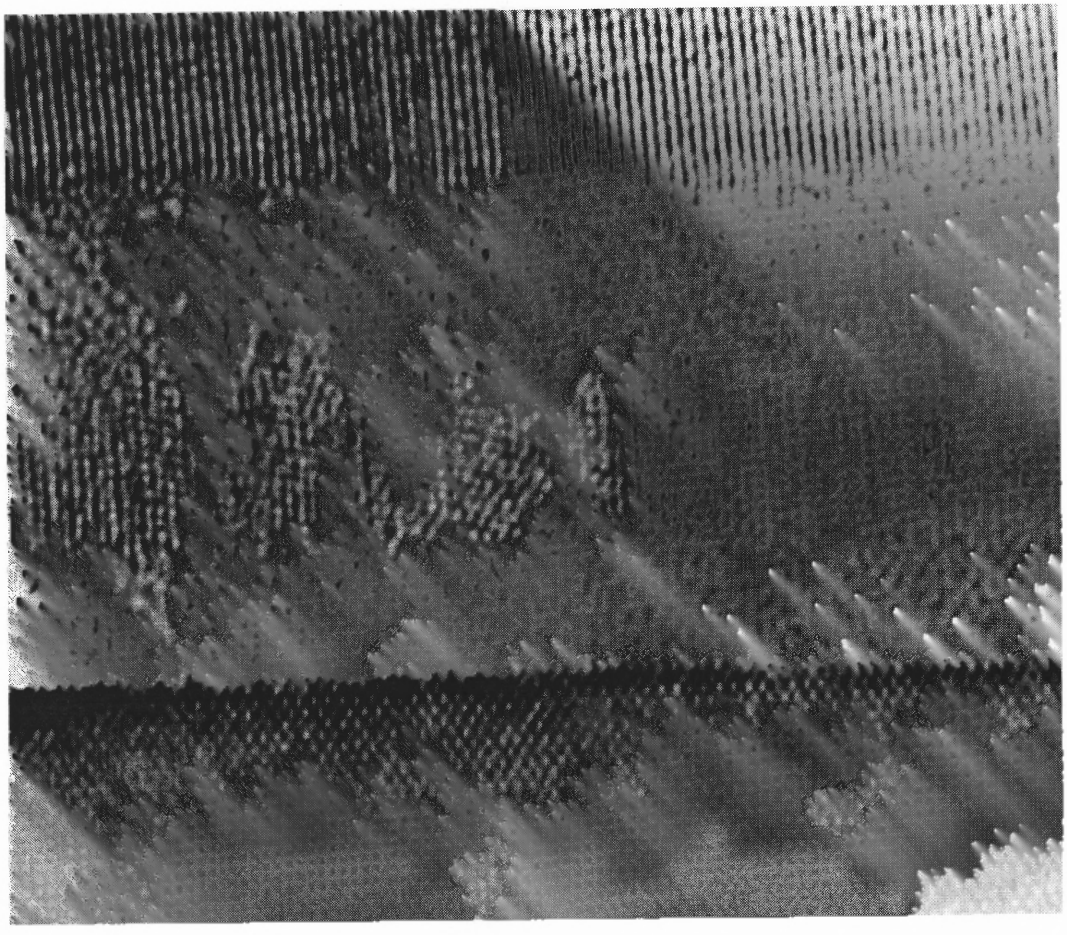

Figure 5.1 High resolution TEM of SiO_2/nc-Si/SiO_2 double barrier structure.

Since the Si lattice constant is ~5.34 Å and the crystalline substrate is <100> oriented, the diagonal rows of atoms in the bottom portion of Figure 1 are <111> planes spaced ~3.84 Å apart. This allows us to estimate the physical thickness of the different layers. The first amorphous SiO_2 layer is approximately 35 Å thick, while both the intermediate nanocrystalline Si layer and the second amorphous SiO_2 layer are approximately 50 Å thick. It is important to point out that the presence of parallel planes of atoms in the center Si layer clearly indicates that while this film is continuous, it is nonetheless made up of nanocrystalline grains which might be expected to exhibit bulk

properties such as the existence of a bandgap and the usual Si-SiO_2 interface potential barrier height. The work presented in this paper confirms this expectation. We note that because of nonuniformities in the thickness of the two oxide layers, particularly the second layer which is grown on the intermediate nanocrystalline Si, and because of the exponential dependence of tunneling on thickness which causes most of the current to tunnel through the thinnest part of the barrier, it is not unreasonable to approximate both barriers as having roughly the same electrical thickness, ~35 Å.

5.3.2 Current-Voltage Characterization

In order to interpret the current characteristics of this structure, it is instructive to first consider the band diagrams.

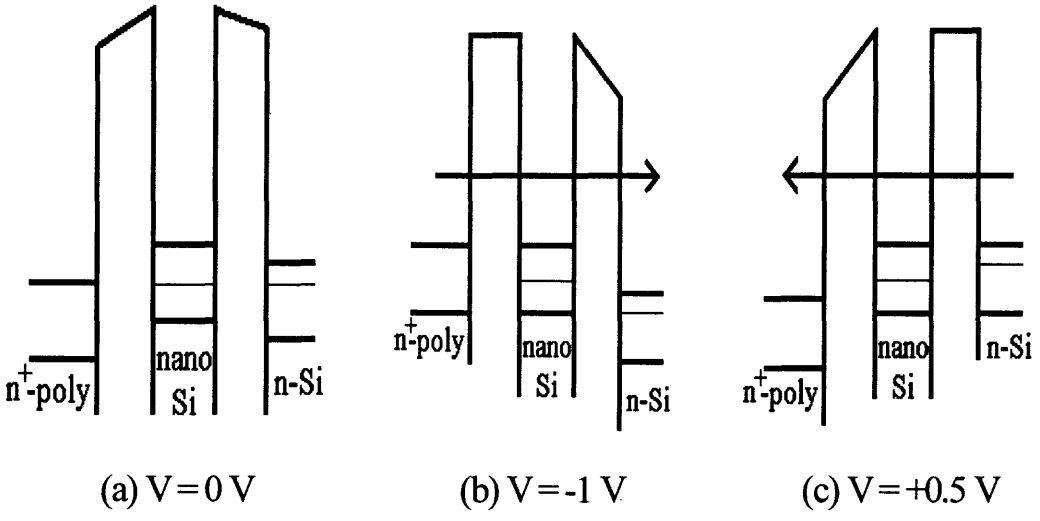

Figure 5.2 Energy band diagrams for a double barrier structure on an n-type substrate.

Figure 5.2 shows a schematic energy band representation of the double barrier structure on an n-type substrate at zero applied bias. The multilayer dielectric is sandwiched between the n+ polycrystalline Si gate contact (left) and the n-type substrate (right). The undoped, intermediate nanocrystalline Si layer separates both ultrathin oxides. This layer is thin enough that it can be thought of as a quantum well, but with a continuum of levels above the conduction band edge due to thermal and well thickness variations. At zero bias, little or no charge is expected to tunnel between the substrate and the gate. Also, because the gate Fermi level is opposite the bandgap of the intermediate Si well, there are no states available in the well through which electrons can tunnel. When a negative voltage is applied to the gate such that the conduction band of the polycrystalline gate contact lines up with the conduction band of the well, a strong turn-on of the current is expected, because electrons will be able to tunnel directly to the substrate via available states in the well. On the other hand, at a positive voltage such that the conduction band of the Si substrate lines up with the conduction band of the well, another strong turn-on of the current is expected, because electrons will be able to tunnel directly to the gate via available states in the well.

A measured I-V curve that confirms this operational description of the structure is shown in Figure 5.3. The significance of this confirmation lies in the fact that it shows that ultrathin silicon dioxide layers have a bandgap and potential barrier height that are similar to that in bulk silicon dioxide. Note that a strong turn-on of the current is observed at a gate bias of ~–0.8 V corresponding to tunneling from the gate to the substrate via the well. Also, a strong turn-on of the current is observed at ~+0.3 V corresponding to tunneling from the substrate to the gate, again via the well. The

scattering in the range from –0.8 V to +0.3 V is noise at the resolution limit of our measurement system, indicating that negligible transport occurs across the double barrier in this bias region, thus creating a "window" of operation in the I-V curve, analogous to that seen for FN tunneling through thicker single oxides. The saturation of the current at negative gate bias levels beyond –1.2 V is due to limited charge generation in the substrate, which can be minimized if a more heavily doped substrate is used. This would have the additional benefit of a shift in the current turn-on voltage for positive biases to a higher value, widening the "write/erase window". The use of multiple intermediate silicon layers would have a similar effect.

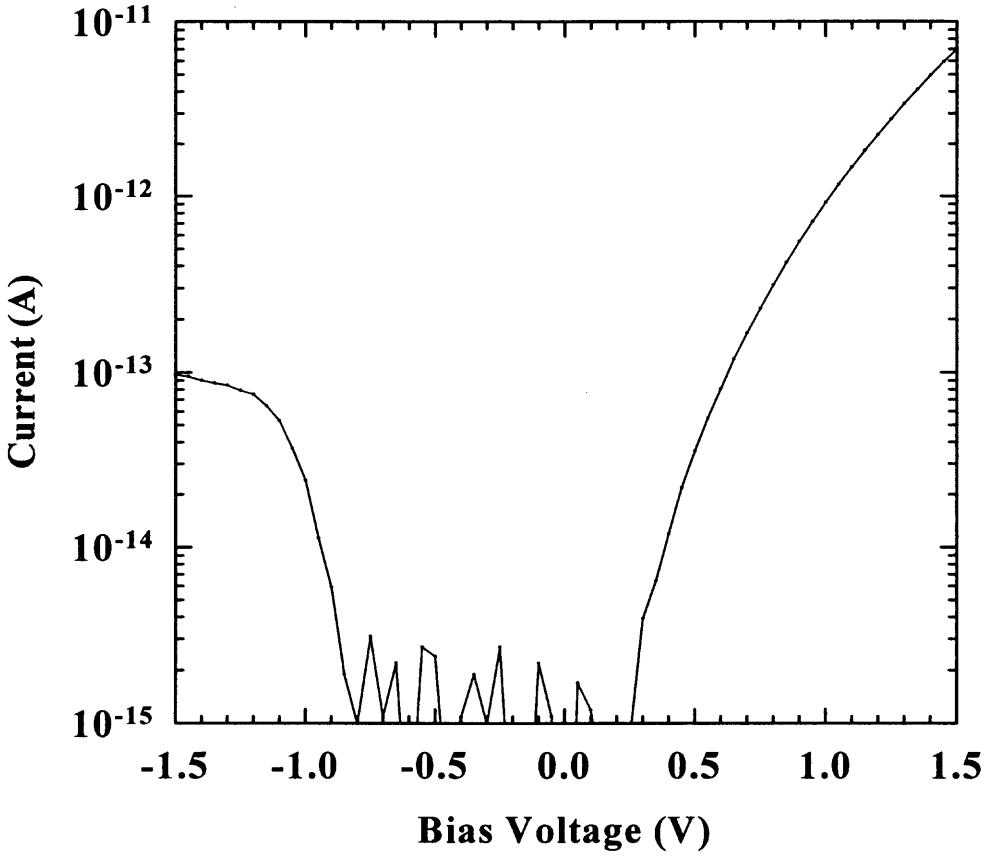

Figure 5.3 Current vs. Voltage for double barrier on n-type substrate.

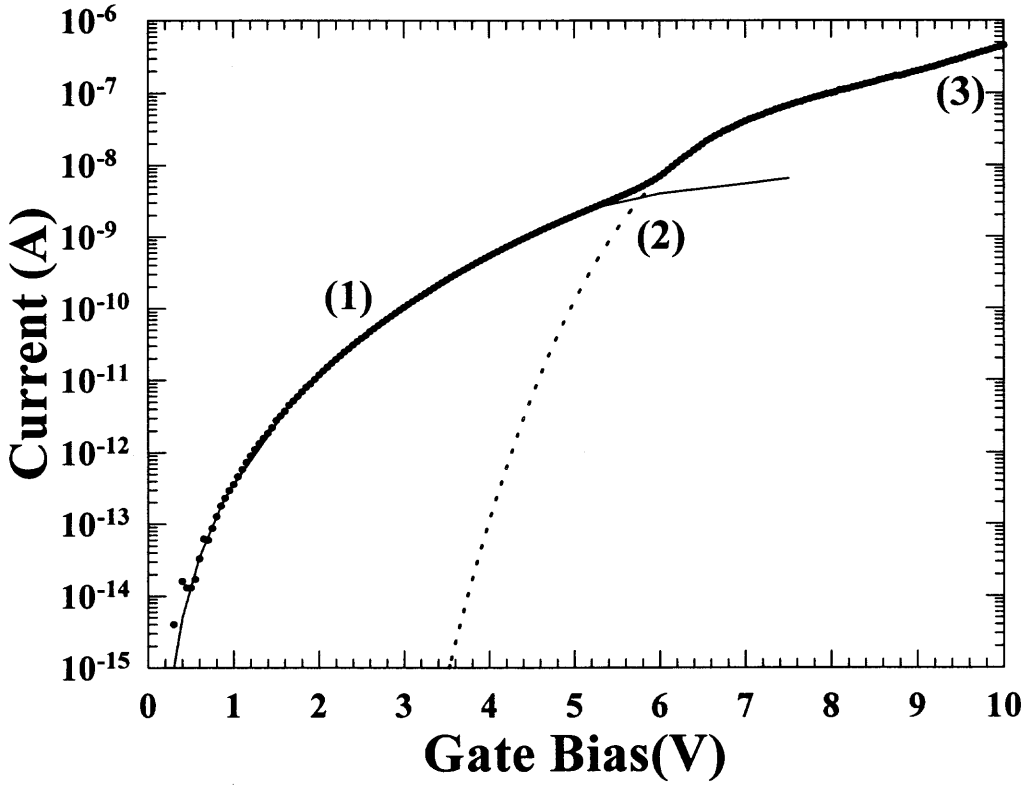

Figure 5.4 Current vs. Voltage curve for a double barrier on an n-type substrate between 0 V and 10 V.

In Figure 5.4 an I-V curve is measured for gate biases between zero and +10.0 V. The initial gradual increase in current is due to an increase in direct tunneling of electrons through the two trapezoidal oxide barriers, from the substrate to the gate via the well. At +6 V the current increases at a higher rate, corresponding to the onset of FN tunneling across one of the oxides. This is illustrated by the dashed lines in Figure 5.4 and by the band diagram of Figure 5.5(a). The apparent current increase at +6 V actually begins at roughly +3.5 V and is consistent with a transition from direct tunneling through both of the barriers at lower voltages to FN tunneling through the barrier nearest the gate. It is important to note that in this interpretation conduction via the well includes tunneling via

the highest accessible states. We also note that the onset of a current increase at +9 V supports this interpretation. This increase is just observable in the figure as a slight upturn in the I-V, and is consistent with a transition at ~6.5 V to FN tunneling through the barrier nearest the substrate, as illustrated in Figure 5.5(b), and perhaps also to injection over the barrier nearest the gate.

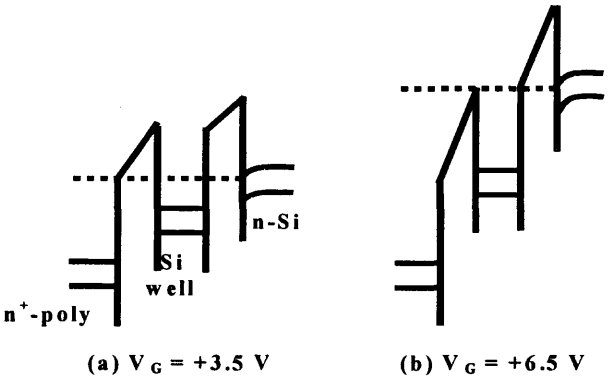

Figure 5.5 Schematic energy band diagrams for V = 3.5 V and V = 6.5 V.

5.3.3 Capacitance-Voltage Characterization

Further characterization of the diode structure has been performed through capacitance-voltage (C-V) measurements. Figure 5.6 shows both high frequency (HF) and low frequency (LF), or quasi-static, C-V curves. At 100 kHz a conventional HFCV curve is obtained featuring the expected accumulation, depletion and deep depletion regions. With the substrate in accumulation, the saturation capacitance of 11.6 pF is in close agreement with a calculated barrier capacitance of 12 pF assuming a total oxide thickness of 70 Å. Also shown are LFCV curves at six different sweep rates ranging from 15 mV/s to 125 mV/s. The total device current measured in any sweep direction consists of both tunnel and displacement current components. Since the polarity of the displacement

current depends on the sweep direction, the quasistatic capacitance is extracted from the total measured current in both sweep directions by subtracting the total measured current for the negative sweep direction, I_{tot-}, from the total measured current for the positive sweep direction, I_{tot+}, and dividing the difference by twice the magnitude of the sweep rate. That is:

$$C = I_{displacement}/|dV/dt| = [(I_{tot+} - I_{tot-})/2]/|dV/dt|.$$

The LFCV curves appear to reach their saturation capacitance at ~0.4 V. This saturation is consistent with that measured at 100 kHz, though it occurs at a much lower bias level than in the HFCV. In the voltage range from +0.5 V to –0.5 V the LFCV curves sweep from accumulation, through depletion into inversion. The peak structure at voltages beyond –0.5 V is largest for the slowest sweep rate and is attributed to a decreasing ability of the slow minority carriers to respond to the linear voltage ramp at increasing sweep rates. By comparing the HF and LF C-V curves near the flatband condition we estimate that the density of interface state levels at the crystalline silicon surface is on the order of 10^{10} eV^{-1} cm^{-2}.

An interesting feature of the LFCV's is the quick rise to saturation at the onset of accumulation, which corresponds to the onset of direct tunneling, followed by a rollover to a steadily decreasing capacitance. This may be attributed to a polycrystalline Si depletion effect, which may be the case, but it is important to note that the rollover is only seen in the LFCV measurements. The HFCV shows the usual gradual rise to saturation, so if polycrystalline Si depletion is occurring, it is not due to doping, but rather due to an effect which turns on with the transition to accumulation, i.e., with the onset of direct tunneling via states in the well. Further insight into this effect can be

obtained by examining our LFCV measurement technique more closely. The LFCV's initial rise above the HFCV, followed by a fall below the HFCV is simply an indication of a small hysteresis in our measured response of the structure to the voltage ramp swept in the two directions. While this is a small effect in these devices, it is worth noting, and the origin of the effect will become clearer with our investigation of the double barrier fabricated on p-type substrates.

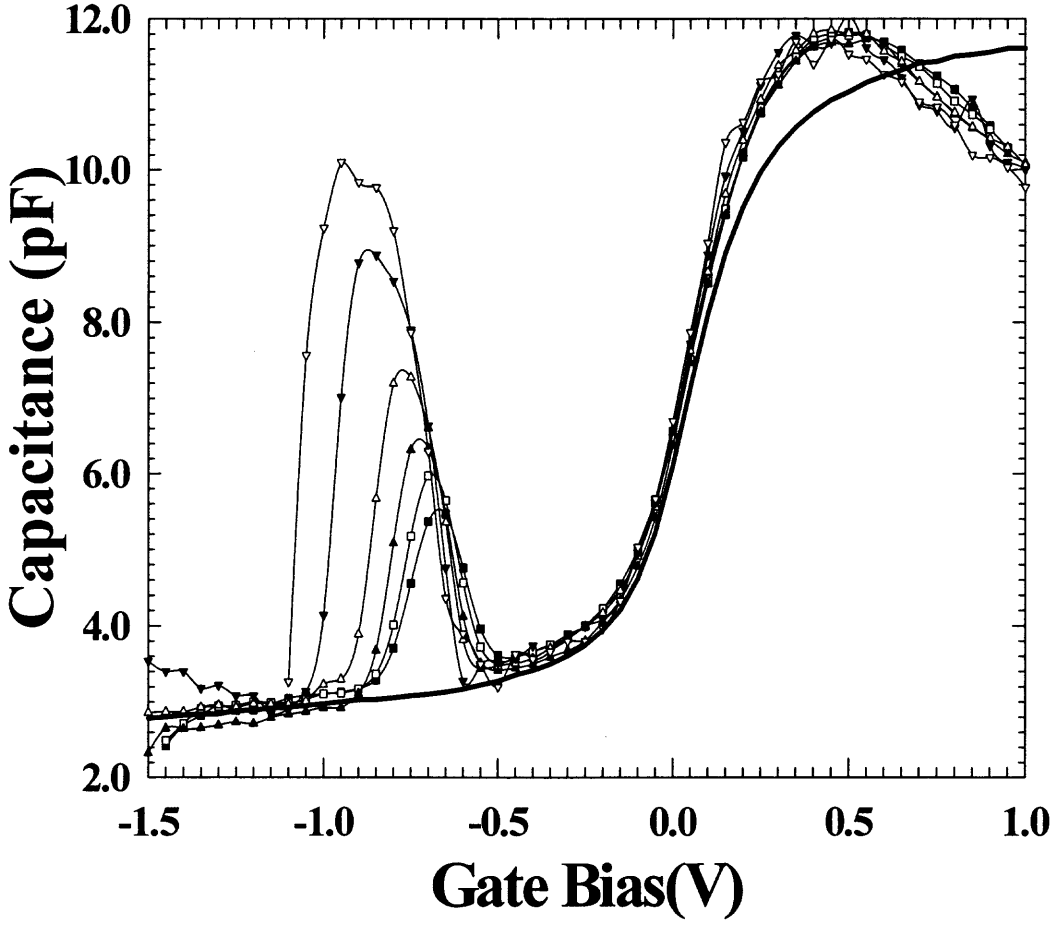

Figure 5.6 C-V characteristics of double barrier on n-type substrate.

5.3.4 Degradation

Shown in Figure 5.7 are I-V curves measured after 10, 10^2, 10^3, 10^4 and 10^5 electrical stress. The bias was switched between -1.5 V and + 1.5 V at a frequency of 0.3 Hz. It is apparent that the I-V characteristics are identical up to 10^5 electrical stress cycles, indicating that the repeated injection of charge across the double barrier has not degraded it at all. This is consistent with the expectation that charge carriers can tunnel through an ultrathin oxide without damaging it as long as the voltage across it is held within the range $-3\ V < V_{gate} < +1\ V$ [84]. This also opens the possibility to a new device structure based on these layered tunnel dielectrics, as will become clear further in this section.

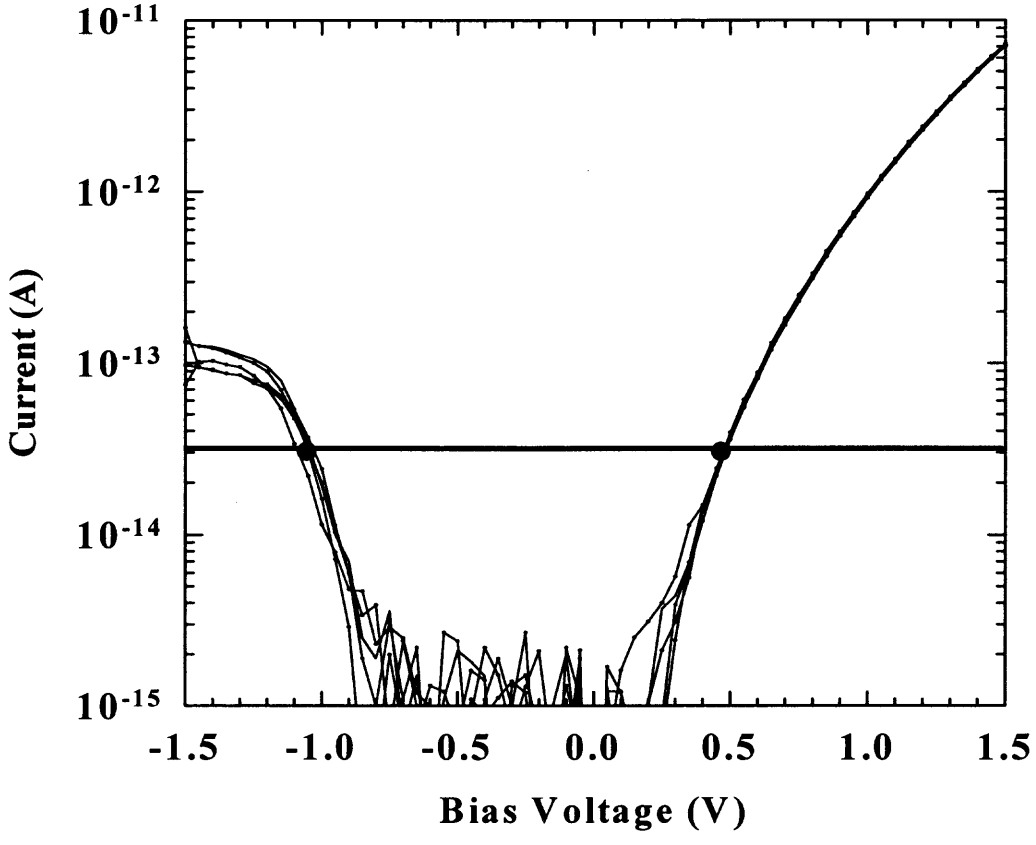

Figure 5.7 Degradation I-V curves.

The spread between the positive and negative voltage at which the current is equal to $3.16 \cdot 10^{-14}$ A (as indicated by the line) is defined as the "window" of operation of the device. It is clear that the current within this window is below the noise level. It is obvious from Figure 5.7 that the window is unaffected by repeated stress. The following figure, Figure 5.8, shows how the window is affected by up to 10^9 stress cycles. Vertically depicted are the voltages at which the current reaches our threshold value of $3.16 \cdot 10^{-14}$ A. The horizontal axis represents the number of stress cycles. It may appear that the window has widened slightly. This is not the case for the following reason. The first five points (10 cycles up to 10^5 cycles) are measured on the same device and are taken from Figure 5.7. However, the last point at 10^9 stress cycles was measured on a different device that had a slightly wider window to begin with. Figure 5.8 shows us that the barrier remains intact even after 10^9 cyclesThis is a remarkable result.

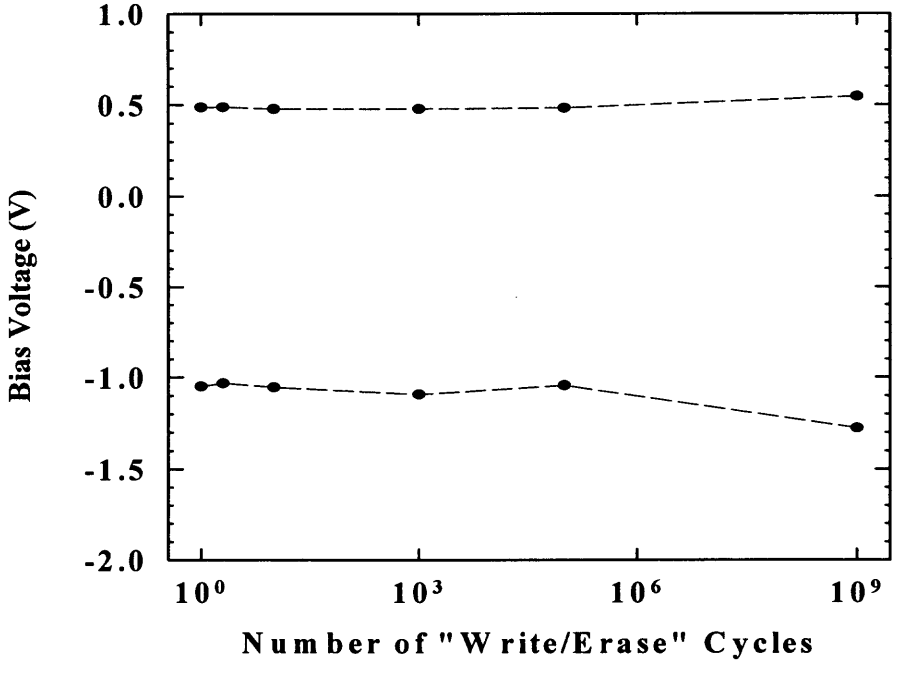

Figure 5.8 Width of the window vs. number of cycles.

5.3.5 Double Barrier as Tunnel Dielectric in Nonvolatile Memory Devices

From the previous sections it has become clear that two key properties of the multilayer charge injection barrier have been established: first, an inherent resistance to degradation due to the use of ultrathin silicon dioxide and secondly, a window of operation in the current characteristics, in which there is negligible conductance. These two key properties suggest that this dielectric may be used as the active tunnel dielectric in traditional floating gate nonvolatile memory devices. These properties of the structure therefore facilitate a new device application for MLCIBs. This is a topic that will be discussed in the next chapter.

5.4 Double Barrier on p-Type Substrate

Using n-type substrates, the most detailed transport information is obtained for injection from the substrate. In order to examine in more detail the case of injection from the gate, p-type substrates have been used. In Figure 5.9 a schematic energy band representation of the diode structure on a p-type substrate is shown for zero gate voltage.

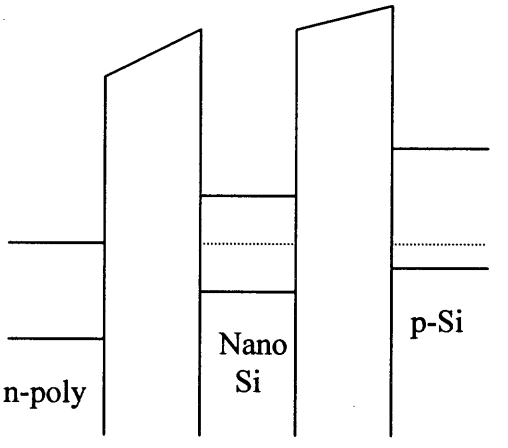

Figure 5.9 Energy band diagram for a double barrier structure on an p-type substrate.

At this bias little or no charge will tunnel between the substrate and gate, since the total barrier thickness is too large and there are no available states in the bandgap of the center Si layer. At a negative gate bias such that the conduction band in the polycrystalline Si gate lines up with the conduction band edge of the well, a strong turn-on of the current is expected corresponding to electrons tunneling directly from the gate to the substrate via states in the well. At positive gate bias, no significant electron tunneling current is expected because it is limited mainly by recombination-generation and diffusion processes in the p-type substrate. This behavior is confirmed by the measured I-V curve shown in Figure 5.10, where the only strong turn-on of tunneling is seen for negative gate biases beyond ~—1.4 V. The plateau near –2.0 V will be discussed in conjunction with Figure 5.12.

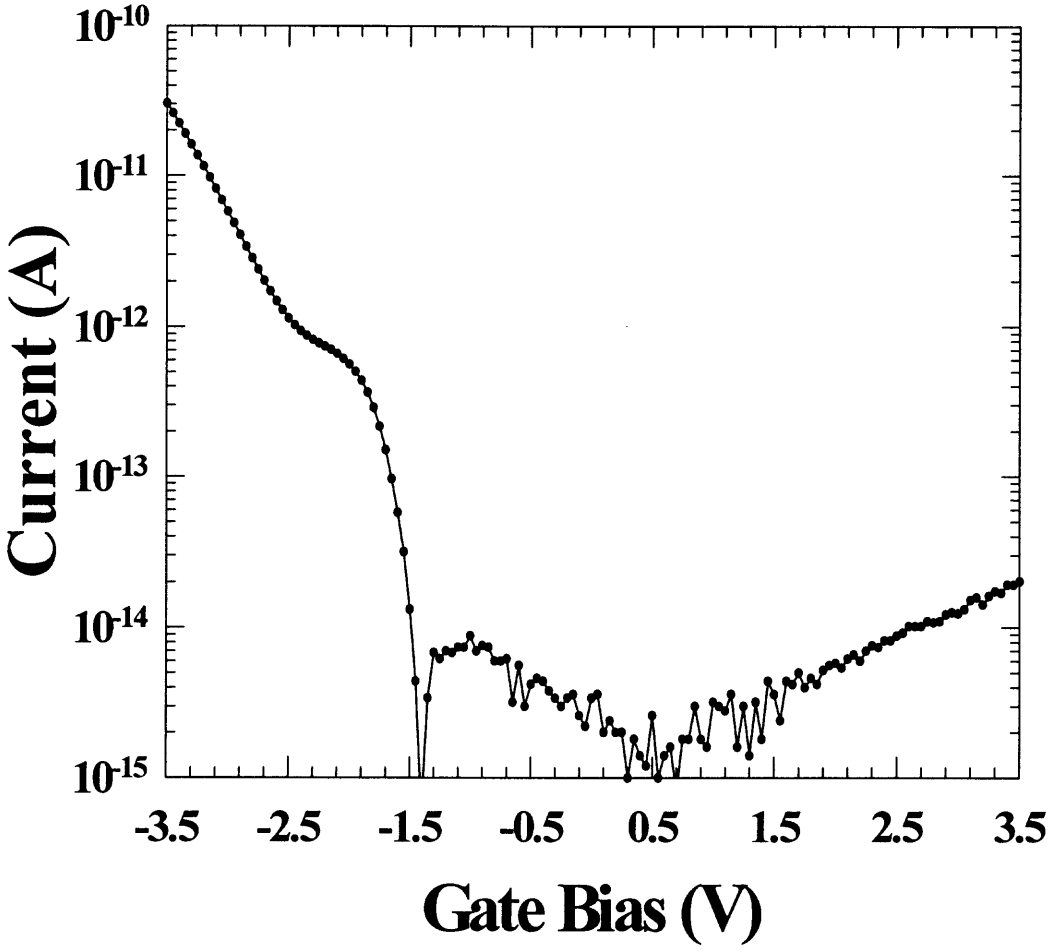

Figure 5.10 Current vs. Voltage for double barrier on p-type substrate.

LFCV results for the p-type substrates are shown in Figure 5.11 together with an HFCV curve. These results are obtained using the same C-V measurement techniques as for the n-type substrates. The HFCV curve is measured at 100 kHz and features the expected accumulation, depletion and deep depletion regions, with the transition from depletion to accumulation occurring at a gate bias of roughly –1.4 V. The capacitance saturates at approximately 11 pF corresponding to a total oxide thickness of ~80 Å, in close agreement with the n-type devices. The LFCV curves are measured at sweep rates

of 40, 60, 80, and 100 mV/s and show (1) deep depletion rather than a saturation oxide inversion capacitance because we cannot stimulate formation of the inversion layer using our LFCV measurement technique, and (2) dispersion in the depletion dip indicating the presence of interface states. By comparing the HFCV curve with the LFCV measured at 40 mV/s near the flatband voltage, we estimate the density of interface state levels to be ~10^{12} $eV^{-1}cm^{-2}$, ~100 times higher than what is seen in the n-type samples.

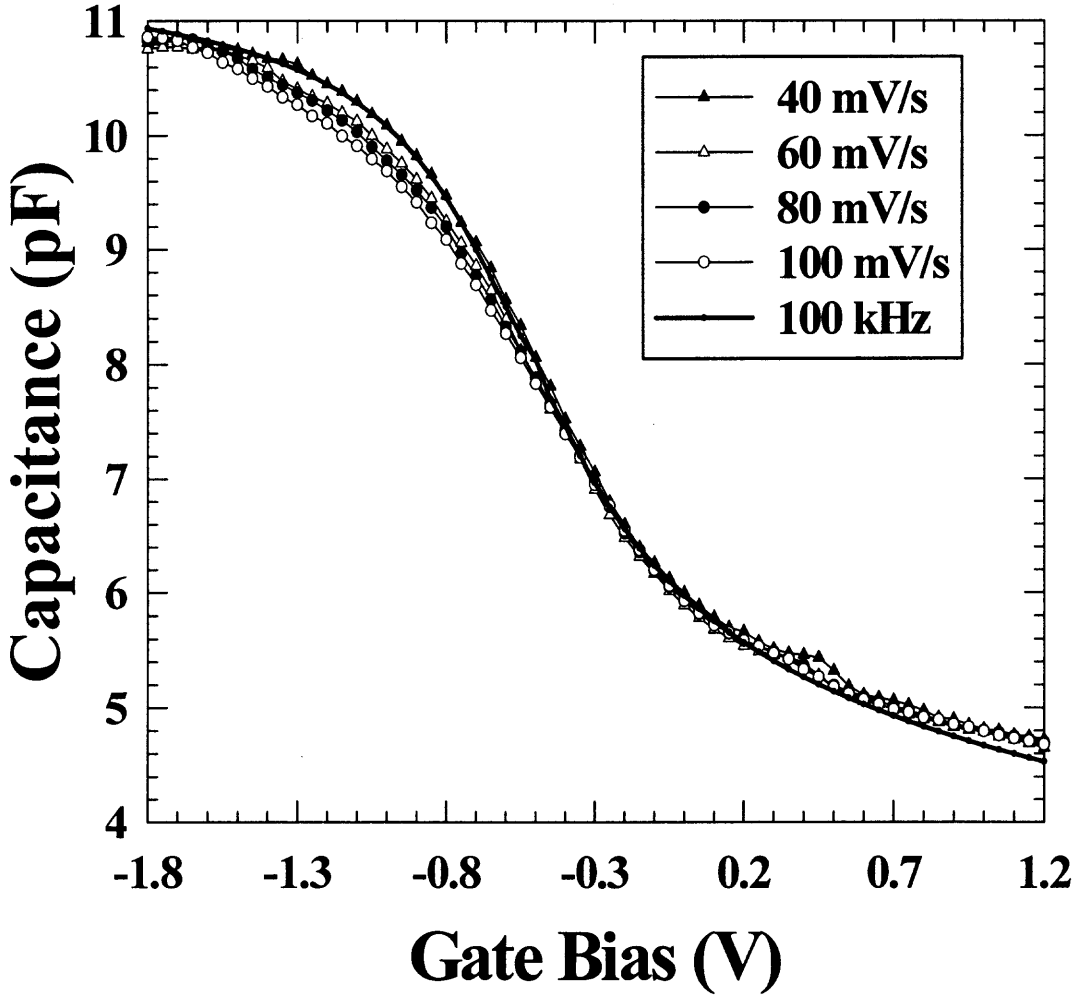

Figure 5.11 C-V characteristics of double barrier on p-type substrate.

In Figure 5.12 step I-V curves are shown, measured for both positive and negative voltage step directions at three different very slow step rates. These curves display an obvious hysteresis effect that is analogous to, though much larger than that seen in the n-type devices at a higher equivalent sweep rate. The magnitude of the total measured current is initially higher as the device is being swept from depletion to accumulation, then in accumulation the opposite is true. In other words, as the direct tunneling is turning on, the barrier is initially more transparent compared to when the tunneling is turning off, then becomes less transparent at higher voltages.

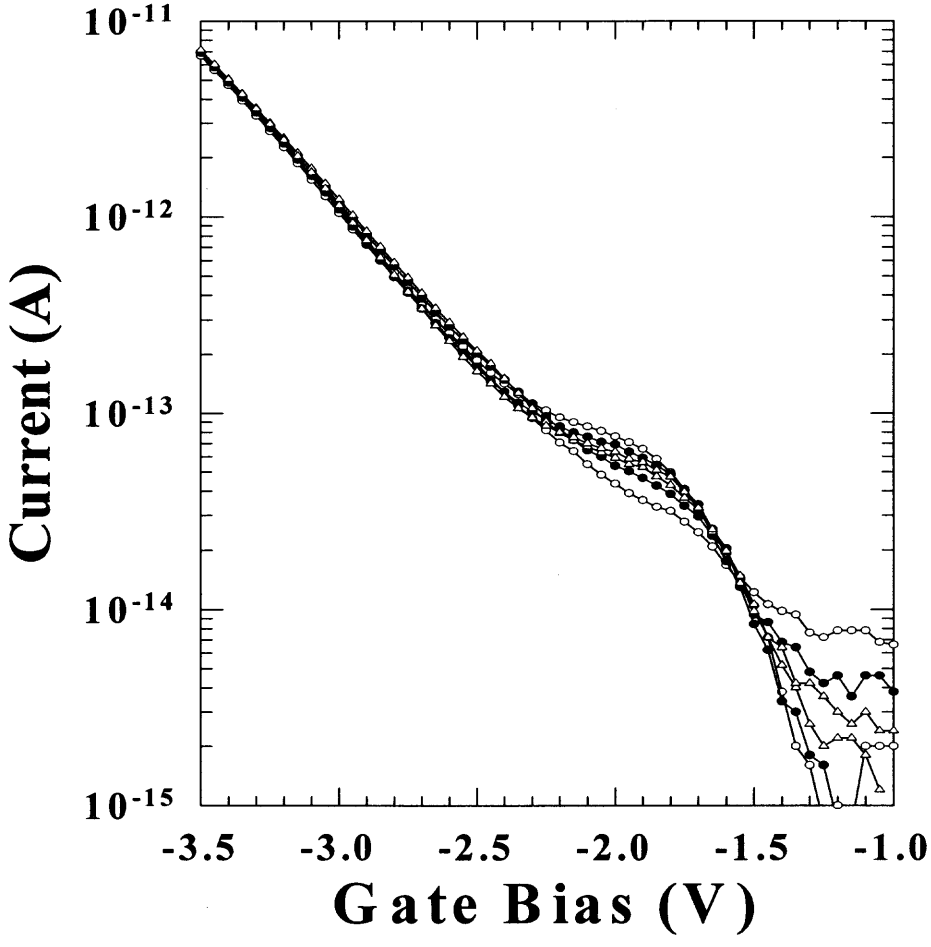

Figure 5.12 Sequential I-V curves for negative voltages.

Accompanying the onset of direct tunneling in the p-type devices is a plateau in the I-V curves between ~–1.7 V and ~–2.4 V. Upon first consideration, this plateau might be presumed to be the familiar depletion plateau seen in single tunnel oxides on p-type Si substrates. The depletion plateau occurs when incremental increases in the bias across the device are dropped, not across the oxide, but across the Si substrate as bands are being swept from inversion to accumulation. However, the C-V data clearly indicate that the Si substrate is already in accumulation at ~–1.4 V.

In this situation, perhaps the most plausible explanation for the plateau in the p-type devices may be that at the onset of direct tunneling, incremental voltage increases are initially dropped across the well. This would be possible if, for example, the charge states of trap levels at the well-oxide interfaces change over this bias region. Such a scenario is illustrated in Figure 5.13 which shows that a high density of charged levels at the well-oxide interfaces can result in a voltage drop across the well. The hysteresis would be determined by the relative rates of donor and acceptor interface state charging in response to changes in the bias, and the size of the effect would be determined by the density of well-oxide interface states. The differences observed in the n-type and p-types samples are consistent with this explanation.

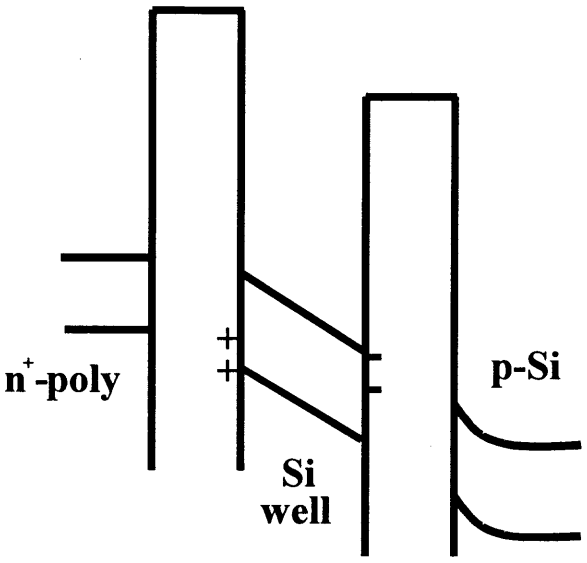

Figure 5.13 Energy band diagram after the interfaces have been charged.

5.5 Summary

In this chapter, silicon-based multilayer charge injection barriers have been fabricated and characterized. Dielectrics on both n-type and p-type substrates were used. Current-voltage characteristics were in correspondence to what was expected from band diagrams for both types of devices. This proves that ultrathin silicon dioxide layers "behave" as bulk oxide, i.e. have a bandgap and an interface potential barrier height. Devices on n-type substrates in particular, were found to be of extraordinary electrical quality with an interface state density of 10^{10} eV^{-1} cm^{-2} and a breakdown field strength of 15 MV/cm. Current-voltage characteristics and subsequent degradation studies revealed this structure to have two key properties: an intrinsic resistance to degradation and a window of operation in which there is negligible conductance. These properties suggest that this structure can act as the active tunnel dielectric in traditional floating gate nonvolatile memory devices.

CHAPTER 6

NONVOLATILE MEMORY TRANSISTOR

6.1 Overview

Earlier in this dissertation it was explained how scaling of devices leads to improvements in device performance. This was based on the device characteristics of MOSFETs. There is another class of devices for which scaling has not been possible, namely, floating gate nonvolatile memory devices. These devices suffer from two performance limitations. One can not continuously scale the dimension of the tunnel dielectric – typically a thin oxide in the range of 5.0 to 7.0 nm – since it acts as a barrier to prevent stored charges from leaking off the floating gate. In addition, such devices suffer inherent degradation problems caused by the charge transport mechanisms (Fowler-Nordheim tunneling or Channel Hot Electron injection) that are used to transport charges to the floating gate.

In the previous chapter it was established that a double barrier $SiO_2/Si/SiO_2$ structure on an n-type silicon substrate has two important characteristics that can be exploited in a nonvolatile memory structure: resistance to degradation and a window of operation that allows for charge retention. This chapter reports on the effort to fabricate and characterize a nonvolatile memory device with silicon-based multilayer charge injection barriers acting as the active dielectric. In the first part, the principle of nonvolatility is detailed, followed by a review of state-of-the-art nonvolatile memory technology. The rest of the chapter covers fabrication details and results.

6.2 Introduction

6.2.1 Principle of Nonvolatile Memory

As seen from the cross section in Figure 6.1, nonvolatile semiconductor memory devices are similar to MOSFETs except that a floating gate has been placed between the control gate and the substrate. The floating gate is completely insulated from the substrate by a tunnel dielectric and from the control gate by an interpoly dielectric. Two classes of floating gate devices exist. The first class is based on a continuous semiconducting layer, typically polycrystalline silicon. In the second class of devices, charge is stored in discrete trapping centers of an appropriate dielectric layer, usually silicon nitride. These devices are generally referred to as charge trapping devices. This chapter focuses on the first class of devices.

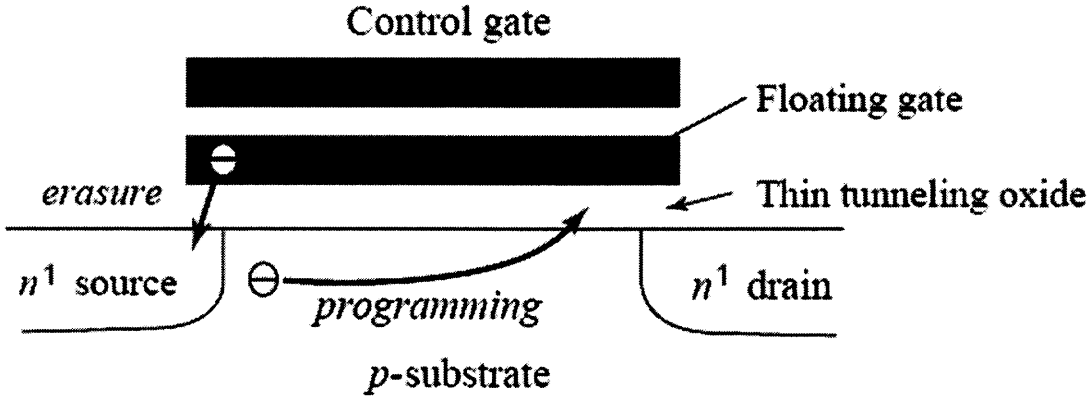

Figure 6.1 Schematic cross-section of a traditional floating gate memory transistor.

The memory action is accomplished by storing charge on the floating gate. Since the floating gate is completely insulated, charges ideally remain on the floating gate when the power supply is removed, making this a nonvolatile memory device. It must be pointed

out that many different device structures exist with different write and erase mechanisms. The one described is generally referred to as FETMOS and is only used for illustrative purposes.

The presence of stored charges on the floating gate alters the threshold voltage of the device. Two distinct values of the threshold voltage can be defined: the erased state, or "0" state and the programmed state, or "1" state. This is illustrated in the figure 6.2.

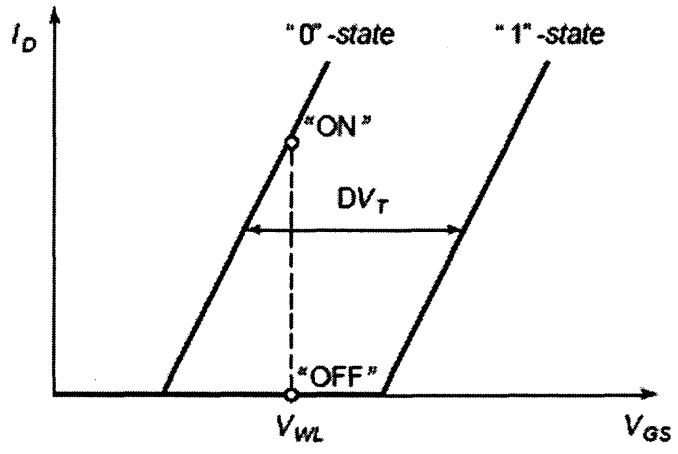

Figure 6.2 Threshold voltage shift in I-V curve of a non-volatile memory transistor as a result of charge storage on the floating gate.

Obviously, apart from being able to program and erase the memory device, one needs to be able to sense, or "read", the state of the device. The state of the device is detected by applying a gate voltage V_{WL} with a value in between the two threshold voltages. By detecting the conduction between the source and the drain as a result of V_{WL}, one can determine the state of the device. If the device is in the "1" state, no conduction will be observed since V_{WL} is smaller than the threshold voltage in that case. Alternatively, conduction will be observed when the device is in the "0" state.

6.2.2 Program Operation

Using FETMOS (Figure 6.1), two programming methods are possible: Fowler-Nordheim tunneling and Channel Hot Electron injection. In the case of Fowler-Nordheim tunneling, a large positive bias is placed on the control gate with the substrate, drain and source grounded. This allows electrons to tunnel through the triangular barrier of the tunnel dielectric onto the floating gate. Electrons are injected uniformly across the entire channel of the device.

In the case of Channel Hot Electron Injection, a very large positive bias is applied to the drain, a positive bias to the control gate while the substrate and source are grounded. Under this large lateral electric field, electrons in the channel travelling from the source to the drain gain kinetic energy and become "hot". A few of these "hot" electrons will have gained enough kinetic energy to surmount the 3.2 eV potential barrier between the substrate and the oxide. These electrons will be attracted to the floating gate.

Due to the inefficiency of this mechanism, a large drain current is required. Therefore, devices exploiting this mechanism consume a lot more power than devices exploiting Fowler-Nordheim tunneling. However, the programming speed is on the order of 10 ns while that of a Fowler-Nordheim device is about 100 ns.

As the device is programmed, the amount of charge on the floating gate increases. This has the effect that the electric field across the tunnel dielectric, and thus the programming current. As this current diminishes, the threshold voltage begins to saturate constituting a self-limited mechanism.

6.2.3 Erase Operation

Fowler-Nordheim tunneling is typically used to remove electrons from the floating gate. This can be accomplished by uniform injection across the entire channel or nonuniform injection into the source or drain regions. In the case of uniform erasure, a large negative bias is placed on the control gate with the drain, source and substrate grounded. Electrons are injected uniformly form the floating gate to the substrate. In the case of nonuniform injection, the large negative voltage is split between the control gate and the junction where the electrons will be erased to. The erase operation is also self-limiting.

6.2.4 Cycling Endurance

The cycling endurance of a nonvolatile memory device is defined as the number of program/erase cycles the device can withstand before it can no longer retain its memory state. Ideally, one would like the device to have infinite cycling endurance. However, the charge transport mechanisms that are used to program and erase the device as discussed above, severely limit the endurance of the device. The large voltages needed to program and erase the device cause charge trapping and interface state generation. Charges can therefore be trapped in these states, causing a shift in both the programming threshold voltage as well as the erase threshold voltage. The spread between these two voltages defines the program/erase window of operation. The net effect of charge trapping and interface state generation is a closing of the program/erase window. This is illustrated by Figure 6.3. Typically, complete window closing occurs after 10^6 cycles, rendering the device useless, since it no longer acts as a memory device. It is this inherent endurance problem of floating gate devices that our work has addressed, as will become clear later in this chapter.

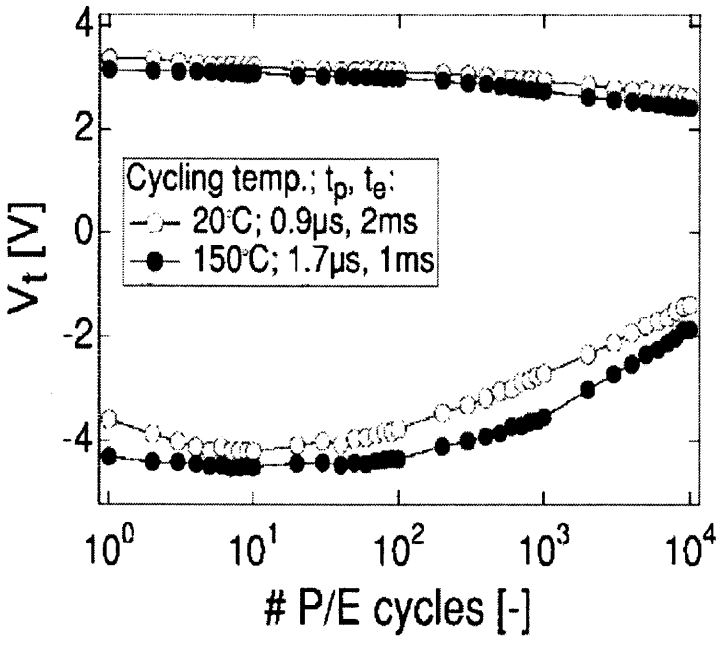

Figure 6.3 Threshold voltage window closing as a result of program/erase cycles.

6.2.5 Data Retention

After being programmed, an ideal memory device should be able to retain charge indefinitely. Generally, a memory device is called nonvolatile when it has a data retention time of 10 years. This allows a leakage of only one electron per day. Typically, electrons leak through the tunnel dielectric or interpoly dielectric. If a large number of charge traps are available in the tunnel dielectric, trap-assisted tunneling becomes a source of leakage. In this case, electrons will first tunnel into a trap and subsequently to the substrate. Initially, tunnel dielectrics will have a negligible number of oxide traps. However, the large voltages needed to program and erase the device, will cause the generation of charge trapping centers in the tunnel dielectric. As a result, the number of

program/erase cycles will increase the leakage current from the floating gate. This phenomenon is called Stress Induced Leakage Current (SILC).

It is clear that the transport mechanisms used in conventional floating gate devices cause a severe degradation of the tunnel dielectric. This has negative effects on both cycling endurance and data retention. The search for potential solutions to these problems is an ongoing research effort. The last sections of this chapter will describe our own approach to solving these problems. But before that, a review is presented of alternative device structures that try to answer the issues mentioned above.

6.3 Alternative Device Structures

Continuing efforts are under way to improve the performance of traditional floating gate transistors that use a single tunnel oxide, without resorting to alternative structures. These efforts typically focus on fabricating higher quality tunnel oxides or fabricating thinner oxides. Even though scaling of the tunnel dielectric is crucial for low voltage, low power and high endurance applications, reduced data retention limits this effort. While there is certainly room for improvement, it is hard to imagine that much progress can be made in improving endurance, due to the inherent degradation problems of the charge transport mechanisms. Much more room for device improvement can be found in alternative structures.

One type of nonvolatile memory transistor that shows a much larger cycling endurance has already been mentioned: charge trapping devices[85-87]. It is misleading to see charge trapping devices such as SONOS as an answer to the performance limitations of traditional floating gate devices, since this device was actually the first nonvolatile memory transistor [88]. The SONOS transistor (Silicon Oxide Nitride Oxide

Silicon) stores charges in localized traps. In order to get a sufficiently high speed of charge transfer, ultrathin oxides are used as tunnel dielectrics. Endurance in charge trapping devices is typically reported to be on the order of 10^9 to 10^{10} write/erase cycles [87]. This is at least a factor 1000 higher for two reasons: localized charge storage will prevent all the charge from leaking away in case of localized oxide breakdowns and secondly, direct tunneling through ultrathin oxides is an inherently less damaging charge transfer mechanism. However, charge trapping devices are inherently slow.

Floating gate devices and charge trapping devices are currently the two state-of-the-art nonvolatile memory devices. For applications that require high endurance, charge trapping devices are used. In case a high speed is needed, floating gate devices are used.

One particularly promising memory device that combines high endurance with high write/erase speed, is the so called ferroelectric RAM (FRAM). In this device the memory state is written as the polarization of a domain. However, these devices require relatively large voltages and are therefore less useful for low power applications [89].

As mentioned before, the use of ultrathin silicon dioxide as the active tunnel dielectric has the advantage of an inherent resistance to degradation. Of course, simply using a single layer of ultrathin dioxide as the active dielectric is not feasible, since the charge on the floating gate will immediately leak off. One way to overcome this problem, and one that has commended a lot of interest, is the use of silicon nanocrystals to store charge[90-95]. While many have reported cycling endurance on the order of 10^9 to 10^{12} cycles, data retention was always problematic. At first, it was believed that a Coulomb blockade effect was responsible for the ability of the nanocrystals to hold charge. Later it was reported that it may be interface traps at the nanocrystal interface to

cause the memory effect. In addition, fabrication of nanocrystals of predictable, controllable sizes is extremely difficult, throwing the future of these devices in doubt.

6.4 Multilayer Charge Injection Barriers

By now it is well established that the use of ultrathin silicon dioxides and therefore direct tunneling as the means of charge transportation, significantly improves cycling endurance. This has become clear from both the work that is presented in the previous chapter, as well as by the work on charge trapping devices and silicon nanocrystals. However, what about data retention? This is the area where the multilayer approach becomes crucial. The presence of a thin, continuous silicon layer that has bulk like properties such as a bandgap and surface potential "opens up" the current-voltage characteristics. This allows for data retention, without having to resort to relatively thick oxides. In addition, this approach has the added advantage that the window width can be tailored by increasing the number of layers in the tunnel dielectric stack. Each additional layer will shift the positive and negative turn-on voltage by 0.5 V, causing a net 1 V widening of the window. Furthermore, tailoring of doping concentrations in the substrate or the gate can be exploited to change the width of the window as well.

Figure 6.4 depicts a cross section of a nonvolatile memory transistor that has multilayer tunnel barriers incorporated as active tunnel dielectric. Notice how in this approach the floating gate is separated from the substrate by a double barrier tunnel dielectric.

The rest of this chapter will detail the results of the very first attempt that was made in trying to fabricate such a device.

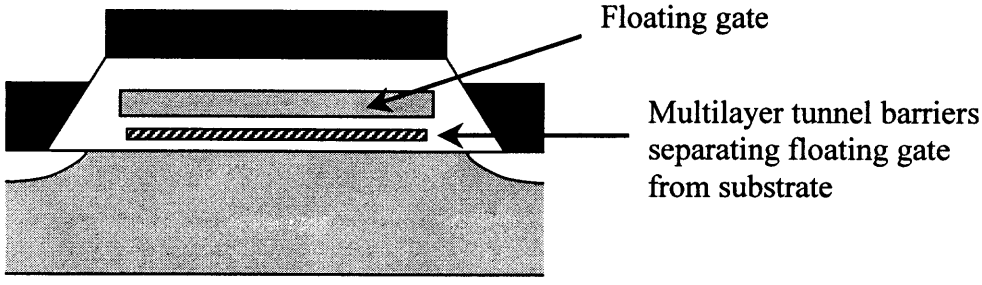

Figure 6.4 Nonvolatile memory transistor with a multilayer tunnel barrier acting as the active.dielectric.

6.5 Device Fabrication

All device fabrication was done in the class 10 cleanroom of the Microelectronics Research Facility at North Carolina State University. A standard floating gate transistor process flow was used in combination with an existing mask set. Only the standard, single step thermal growth of the tunnel dielectric was replaced by a multistep process to grow and deposit stacks of ultrathin silicon and ultrathin silicon dioxide. The mask set contains FETMOS and ETOX floating gate transistors, overlap and non-overlap tunnel dielectric test capacitors and interpoly dielectric test capacitors. The minimum feature size of the mask is 1μm.

Single crystal silicon wafers, <100> oriented, 0.005-0.05 ohm-cm, n-type wafers were used as substrates. A 3200 Å field oxide was grown and wet etched to form the active area.

Since the tunnel stack is such a critical part of the device, processing conditions of the stack will be discussed in more detail here. A table with the actual processing conditions can be found in Appendix B. The tunnel stack was processed in a Rapid Thermal Processing (RTP) reactor. Many calibration runs have been done to determine the correct parameters for the growth of RTO oxides as well as the RTCVD deposition of

thin amorphous silicon layers. Oxide thickness was found to be remarkably uniform across the wafer to within typically 2 Å. The RTO oxides were targeted to grow at a temperature of 950 °C, but was generally found to vary between runs ranging from 930 °C to 980 °C. The pressure in the chamber was 100 Torr. Under these conditions, a 30 second oxidation is expected to grow approximately 25 Å of silicon dioxide.

The thin amorphous silicon layers were targeted to be deposited at 500 °C using disilane. The pressure in the chamber was maintained at 1 Torr. The thickness of the silicon layers was found to be much less uniform across the wafer. From calibration runs it was clear that variations of thickness could easily reach 15 Å or more across a wafer. All amorphous silicon layers were targeted to ultimately be 50 Å thin after the second oxidation. Taking silicon consumption in consideration, silicon layers ranging in thickness from 60 to 65 Å were targeted. A 60 seconds deposition at a temperature of 500 °C and a pressure of 1 Torr will deposit approximately that. It must be noted that temperature deviation of 15 C to 20 °C from the targeted 500 °C were usually observed. This was significant enough to compensate for a lower (or higher) actual deposition temperature by increasing (or decreasing) the deposition time on the spot. It is obvious that the adaptation of deposition time to actual processing circumstances is educated guesswork, amorphous silicon layers ranging in thickness from 40 to 60 Å may be expected.

The floating gate was formed using 1500 Å of amorphous silicon which was then doped with phosphorous (n-type) in a diffusion furnace at 900 °C. Following this was the wet etch step of the floating gate and tunnel stack. HF etch was used to etch through the complete stack. Etch rates were determined by first etching dummy wafers for a set

number of minutes and measuring the remaining silicon (or silicon dioxide). Silicon was found to etch at a rate of 13 Å/s, silicon dioxide at 0.7 Å/s. Etching for the proper amount of time, obviously, is not straightforward since one does not have exact information on the thickness of the different layers. At this stage it is good to point out that future processing attempts may want to consider using dry etching techniques instead, since overetching the tunnel stack with wet processes is certainly a possibility.

Following etching of the stack, a 220 Å dry polyoxide was grown at 1000 °C, followed by LPCVD of the control gate. Doping of the control gate with phosphor was followed by a wet etch of the control gate polysilicon and the interpoly dielectric. Source and drain junctions were formed by dopant diffusion and followed by a 15 minute drive-in at 900 °C. After a passivation capping oxide was deposited, the wafers underwent a forming gas anneal at 500 °C for 30 minutes. Finally, contact pads were formed by an aluminum/titanium liftoff process.

6.6 Results and Discussion

Current-voltage characteristics of tunnel stack capacitors were first tested. Figure 6.5 shows I-V curves on a double barrier capacitor of 25μm^2 device with 25 Å silicon dioxides. The graphs are separated for clarification. The left curve is an I-V curve measured between 0 and –3 V in steps of 0.05 V. The curve on the right shows the I-V characteristics of the same chip measured from 0 to 3 V. It is immediately clear that the current increases steadily from 0 V on for both positive and negative bias voltages. From the results presented in the previous chapter, one would expect to see a delay in the onset of a tunnel current due to the presence of the intermediate silicon layer. In other words,

this device does not exhibit a window of negligible tunnel current in between two bias voltages of strong tunnel current turn-on. Before attempting to interpret this absence of a

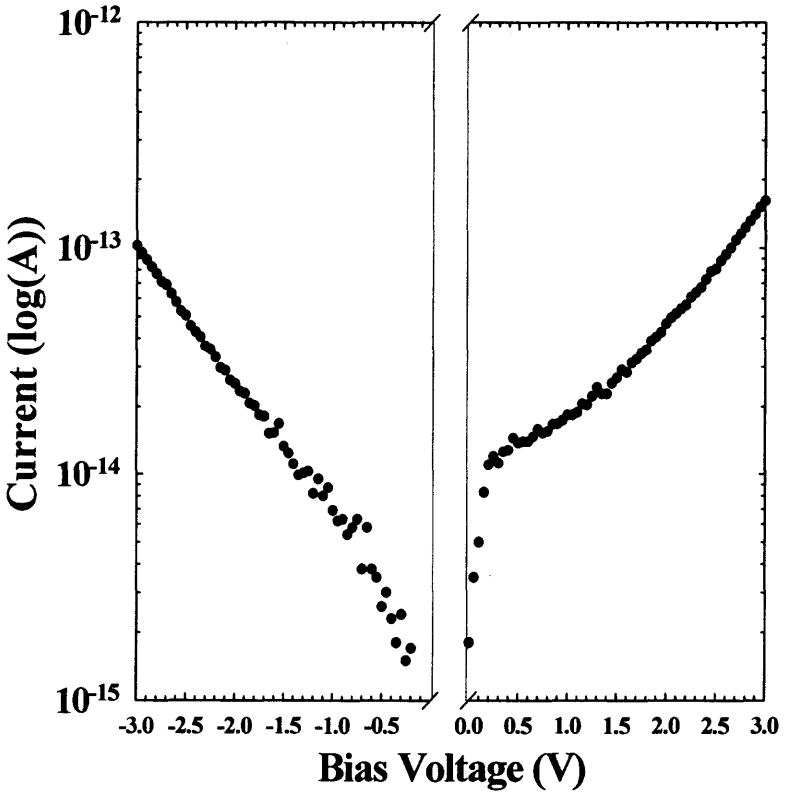

Figure 6.5 Current-voltage characteristics of a multilayer tunnel dielectric on transistor wafer.

window, another significant observation needs to be pointed out. Considering that this is a device with oxides that are approximately 10 A thinner than those presented in the previous chapter, it is clear that very little current flows through the device. At +1.5 V only 20 fA of current is observed, as opposed to 8 pA for the device shown in the previous chapter. Taking into account that the device of figure 3.4 is 4 times smaller, and thus a current of 80 fA would be expected for a device of the same size, the current is still

approximately two orders of magnitude smaller. This is surprising since an oxide that is 10 A thinner is expected to display a current five orders of magnitude larger.

The current behavior for V > +1.5 V is probably of a direct tunneling nature. Careful extrapolation of this component may suggest that this current exhibits a delayed on-set, but the effect is obscured by the rapid increase in current at 0 V. This component is large and most likely due to defects at the interface. The continuing increase in current for negative bias voltages is not well understood.

Figure 6.6 gives us insight into the charging dynamics of the device after electrical stress. The figure shows three I-V curves measured between +3 V and 0 V after waiting 10 seconds, 20 seconds and 40 seconds respectively. The bias voltage at which the device was stressed before starting the measurement, was chosen at +3 V, since this was just below the breakdown voltage. Hence, considerable stress was expected to take place at +3 V. The curve with the longest wait time (triangle) is seen to have the lowest current at 3 V. This indicates that the electrical stress put on the device during the wait time causes it to charge negatively, i.e. a net filling of traps and interfacial defects with electrons occurs. The spread in curves for voltages below 2 V can be explained as a transient effect during the initial ramp up towards 3 V. It is clear that the transient effect saturates when the wait time is long enough.

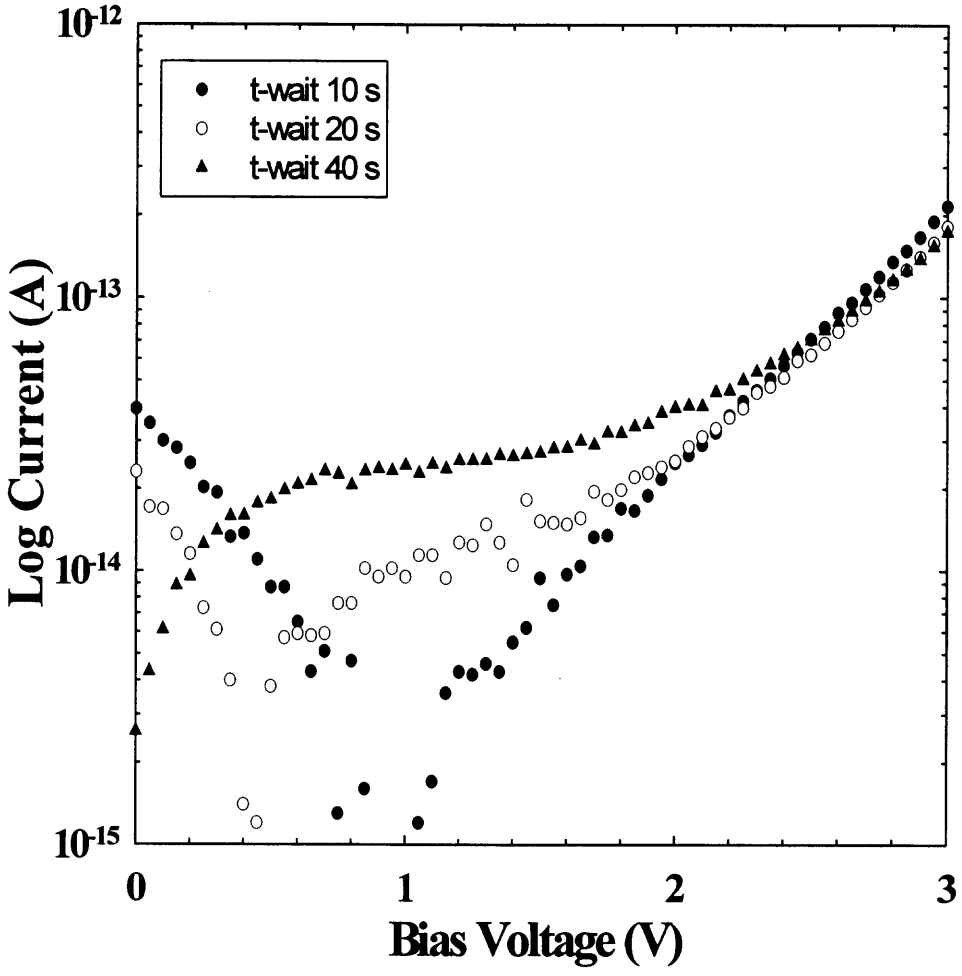

Figure 6.6 Sequential I-V curves after different waiting times at 3 V.

The observations as shown and described above are only reported for one particular chip, but are seen across all wafers. The immediate effect is that in the absence of a window of negligible current around 0 V, these devices will not show any memory effect, the goal of this study. It is obvious that with a significant current present at 0 V, the charge that one might be able to store on the floating gate will immediately leak off.

It is clear from the experimental results that these devices exhibit strong charging effects and a low breakdown field on the order of 4 to 5 MV/cm due to large numbers of

defects and interface states. This however, does not explain the reduced current density with respect to the devices that were described in the previous chapter. One might immediately conclude that a reduced overall current indicates that the oxides are thicker than assumed. A possible explanation is that the curves simply indicate that all the intermediate silicon has been consumed during the second oxidation. This may be true for devices on the outer areas of the wafers due to the nonuniformity of the deposited silicon layer. However, test runs revealed that sufficient silicon was deposited in the center of the wafer to accomodate the silicon consumption during the second oxidation.

Rather, the explanation can most likely be found in a lack of sufficient crystallization of the deposited amorphous silicon layer during the second oxidation. The devices that are discussed in this chapter were targeted to have thinner oxides than the ones described in the previous chapter. As a consequence, the thin amorphous silicon layer may not have crystallized enough – or not at all – during the second, shorter oxidation. Therefore, one might call into question whether the thin intermediate silicon layers in these particular devices really have the properties of bulk silicon, such as a bandgap and surface potential. In the absence of such bulk properties, a window of negligible conductance around 0 V is not expected.

6.7 Summary

In this investigation we have tested double and triple barrier tunnel dielectric stacks of silicon and silicon dioxide. These devices were fabricated as part of a full nonvolatile transistor memory process. The current-voltage characteristics revealed that the devices have large numbers of defects and interface states that prevents the retention of charges on the floating gate. In addition, these devices have a sharply reduced current density. It is believed that one possible explanation for these observations is the lack of sufficient crystallization of the thin silicon layer during processing conditions. It is important to note that these findings by no means prove that it is not possible to fabricate a nonvolatile memory transistor using silicon based multilayer charge injection barriers as the tunnel dielectric. It simply proves that much work remains to be done in understanding the processing conditions under which thin amorphous layers of silicon crystallize.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

In this dissertation, some of the fundamental problems of continued scaling of MOS structures have been identified and possible solutions have been studied in detail. On the other hand, scaling also opens up new venues of device applications that were once unimaginable or simply impossible. An investigation into one of those applications has been carried out and presented here. Below the conclusions and suggestions for future work are presented.

7.1 Leakage Compensated Charge Measurement

7.1.1 Conclusions

A charge measurement has been developed that employs an electronic compensation circuitry, allowing measurement of C-V characteristics of ultrathin MOS device structures, even in the presence of significant leakage current. C-V curves have been measured for oxides between 2.4 and 3.5 nm thin. From the results on 3.5 nm oxides it is found that the minority carrier response saturates easily, making this a true, static C-V measurement as opposed to the standard quasistatic technique. In addition, the expected frequency dependence of the capacitance has been observed. Consequently, the classical high frequency - low frequency capacitance (HLCV) method for obtaining the interface state level density can still be used.

In the thinnest oxides (2.4 nm), an "excess" leakage current is observed that can not by compensated for by the circuitry alone. However, the excess leakage current is

found to increase linearly with the excitation signal amplitude and can therefore be easily compensated for by additional numerical processing.

The significance of these findings lie in the fact that this measurement method extends the range of application of capacitance characterization of silicon oxides well into the direct tunnel thickness regime.

7.1.2 Future Work

With the continued scaling of CMOS devices, this work will remain relevant in the foreseeable future. However, as indicated above, the lower limit of oxides that can be characterized by the measurement method in its current state seems to be somewhere around 24 Å. For these oxides, an additional, numerical compensation technique is already necessary. Further improvement of the circuit is therefore necessary. In its current state the circuit compensates a DC, or first order, current component. Improvement of the circuit may lie in circuit design that compensates for higher order current components as well.

7.2 Alternative Gate Dielectrics

7.2.1 Conclusions

Some of the earliest experiments on hafnium oxide for gate dielectric applications were performed. Thin hafnium oxide dielectrics were fabricated and characterized. The dielectrics were deposited in a Rotating Disk Reactor CVD system. Both an MOCVD and an ALCVD process were employed. Similar results were found for both processes. The deposition produced a strongly nonuniform film with significant, measurable

hafnium oxide concentrated in a concentric ring. The range of thicknesses within the concentric ring was found to be between 110 Å to 280 Å.

Electrical characterization revealed that the devices have characteristics such as large leakage currents, dielectric charging under stress, hysteresis and a large flatband voltage shift that is commonly found in materials such as the one that was investigated in this work. Additionally, a relatively low dielectric constant equal to 13 was found from the oxide capacitance. This can undoubtedly be explained by the formation of a thin SiO_2 layer at the silicon – hafnium oxide interface. In addition, the physical nature of the hafnium oxide is most likely not that of a pure HfO_2 nature.

While one of the alternative gate dielectrics, possibly a silicate, may be a short-term solution for leakage current requirements in low-power applications, it is the author's firm believe that alternative gate dielectrics are *not* the long term solution to the leakage current problem of ultrathin gate dielectrics. The answer will be found in alternative silicon-based transistor structures, such as vertical transistors or dual-gate transistors.

7.2.2 Future Work

As indicated above, there may be a need for alternative gate dielectrics as a short term solution. Therefore, from a business perspective, there may be commercial value in pursuing this line of research. Obviously one would have to improve the process to be able to obtain uniform coatings. But even with the limited work that was done on process optimization, already devices were fabricated that are at least as good as what other have reported. From a university point of view, this may prove to be a line of research with a dead end.

7.3 Silicon-based MLCIBs

7.3.1 Conclusions

Silicon-based double barrier tunnel structures were fabricated and characterized. Especially the structures on n-type substrates proved to be of an extraordinary electrical quality, i.e. low interface state level density of 10^{10} eV^{-1} cm^{-2} and high breakdown field 15 MV/cm. I-V characteristics were in agreement to what was expected from band diagrams for both types of devices. This proves that ultrathin silicon dioxide layers "behave" as bulk oxide, i.e. have a bandgap and an interface potential barrier height. I-V characteristics and subsequent degradation studies revealed this structure to have two key properties: an intrinsic resistance to degradation and a window of operation in which there is negligible conductance.

Subsequently, these structures were incorporated as the active tunnel dielectric in a floating gate nonvolatile memory transistor. A full transistor fabrication process run was done. Wafers with double and triple barrier tunnel dielectric stacks were fabricated and tested. Characterization of test capacitors revealed that the devices have large numbers of defects and interface states that prevents the retention of charges on the floating gate, i.e. no window of operation was observed. In addition, these devices have a sharply reduced current density.

One possible explanation for these observations is that not enough silicon was deposited for the intermediate layer and consequently may all have been consumed by the second oxide growth. While this may be true on certain parts of the wafers, especially the outer parts, due to nonuniformity of the deposited silicon layer, it is certain that, even after the second oxidation, there is still a thin layer of silicon left, especially at the center

of the wafer. Rather, the explanation can most likely be found in a lack of sufficient crystallization of the deposited amorphous silicon layer during the second oxidation. It is important to note that these findings by no means prove that it is not possible to fabricate a nonvolatile memory transistor using silicon based multilayer charge injection barriers as the tunnel dielectric.

7.3.2 Future Work

While one might be tempted to immediately run back into the clean room to attempt a second try at fabricating a transistor, it might be more fruitful to first consider a detailed study of the crystallization parameters of the intermediate ultrathin silicon layer. Not much, if anything, is known about how, when and under what circumstances the ultrathin silicon layer crystallizes. We simply know it does during "subsequent processing". But what is that "subsequent processing"? In addition, another transistor run would probably have to be preceded by a detailed design analysis. This must involve process and device simulation and may involve the designing of a new mask set.

APPENDIX A

PROCESSING CONDITIONS HAFNIUM OXIDE DEPOSITION

The table below lists the processing parameters used during the hafnium oxide deposition that was described in section 4.4. During the first four wafer runs the hafnium precursor and oxygen flowed simultaneously (MOCVD); during the next four wafer runs the hafnium precursor flowed continuously while the oxygen was turned on and off in cycles (ALCVD).

Wafer #	Deposition	P (Torr)	T (°C)	Speed (rpm)	Time
SCVD-0010	MOCVD	20	250	300	4 min
SCVD-0011	MOCVD	20	250	300	4 min
SCVD-0012	MOCVD	20	350	300	4 min
SCVD-0013	MOCVD	20	350	300	4 min

Wafer #	Deposition	P (Torr)	T (°C)	Speed (rpm)	Time	Duty Cycles
SCVD-0014	ALCVD	20	350	300	4 min	30 s. ON / 30 s. OFF
SCVD-0015	ALCVD	20	350	300	4 min	30 s. ON / 30 s. OFF
SCVD-0016	ALCVD	20	350	300	4 min	6 s. ON / 54 s. OFF
SCVD-0017	ALCVD	20	350	300	4 min	6 s. ON / 54 s. OFF

APPENDIX B

PROCESSING CONDITIONS TUNNEL STACK

The table below lists the recorded processing parameters of the tunnel stack for the eight different wafers that were fabricated in the nonvolatile memory transistor run. RTO refers to the Rapid Thermal Oxidation of the SiO_2 layers; RTCVD refers to the deposition of the ultrathin Si layer.

Wafer #	Process	Time (s)	T (°C)
W-11	RTO	45	980
	RTCVD	75	480
	RTO	60	950
W-16	RTO	40	980
	RTCVD	70	510
	RTO	60	940
W-23	RTO	30	950
	RTCVD	60	520
	RTO	30	920
W-6	RTO	30	950
	RTCVD	75	490
	RTO	30	950
W-14	RTO	25	950
	RTCVD	70	480
	RTO	25	950
	RTCVD	70	490
	RTO	25	935
W-8	RTO	25	945
	RTCVD	75	480
	RTO	25	950
	RTCVD	75	490
	RTO	25	950
W-18	RTO	45	940
	RTCVD	70	520
	RTO	45	930
W-12	RTO	30	940
	RTCVD	40	520
	RTO	30	940

REFERENCES

[1] F. M. Wanlass and C. T. Sah, in *Solid-State Circuits Conference. Dig. of Tech. Papers, VI*, 32 (IEEE, New York, 1963).

[2] S. Wolf, *Silicon Processing for the VLSI Era, Volume 2 – Process Integration* (Lattice Press, Sunset Beach, CA 1990).

[3] *International Technology Roadmap for Semiconductors*, Semiconductor Industry Association, Santa Clara (2003).

[4] G. E. Moore, *Electron. Mag.* **38**, 114 (1965).

[5] P. Peercy, *Nature* **406**, 1042 (2000).

[6] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. (Wiley, New York, 1981).

[7] J. L. Hoyt, H. M. Nayfeh, S. Eguchi, I. Aberg, G. Xia, T. Drake, E. A. Fitzgerald and D. A. Antoniadis, in *Tech. Dig. Int. Electron Devices Meet.*, 23 (IEEE, New York, 2002).

[8] G. Timp, R. E. Howard and P. M. Mankiewich, in *Nanotechnology*, edited by G. Timp (Springer-Verlag, New York, 1998).

[9] M. L. Green, E. P. Gusev, R. Degraeve and E. L. Garfunkel, *J. Appl. Phys.* **90**, 2057 (2001).

[10] D. A. Buchanan, *IBM J. Res. Dev.* **43**, 245 (1999).

[11] S. H. Lo, D. A. Buchanan, Y. Taur and W. Wang, *IEEE Electron Device Lett.* ***ED-18***, 209 (1997).

[12] M. L. Green, T. W. Sorsch, G. L. Timp, D. A. Muller, B. E. Weir, P. J. Silverman, S. V. Moccio and Y. O. Kim, *Microelectron. Eng.* **48**, 25 (1999).

[13] D. A. Buchanan and D. J. DiMaria, *J. Appl. Phys.* **67**, 7439 (1990).

[14] M. Alam, J. Bude, B. Weir, P. Silverman, A. Ghetti, D. Monroe, K. P. Cheung and S. Moccio, in *Tech. Dig. Int. Electron Devices Meet.*, 715 (IEEE, New York, 1999).

[15] G. Timp, A. Agarwal, F. H. Baumann, T. Boone, M. Buonanno, R. Cirelli, V. Donelly, M. Foad, D. Grant, M. L. Green and B. Weir, in *Tech. Dig. Int. Electron Devices Meet.*, 930 (IEEE, New York, 2002).

[16] S. Tang, R. M. Wallace, A. Seabaugh and D. King-Smith, *Appl. Surf. Sci.* **135**, 137 (1998).

[17] E. H. Nicollian and J. R. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology* (Wiley, New York, 1982).

[18] M. J. van Dort, P. H. Woerlee and A. J. Walker, *Solid-State Electron.* **37**, 411 (1994).

[19] S. H. Lo, D. A. Buchanan and Y. Taur, *IBM J. Res. Dev.* **43**, 327 (1999).

[20] R. Rios, N. D. Arora and C. L. Huang, *IEEE Electron Device Lett.* **15**, 129 (1994).

[21] L. F. Register, E. Rosenbaum and K. Yang, *Appl. Phys. Lett.* **74**, 457 (1999).

[22] G. Groeseneken, H. E. Maes, N. Beltra`n and R. F. D. Keersmaecker, *IEEE Trans. Electron Devices 31*, 42 (1984).

[23] R. Degraeve, B. Kaczer and G. Groeseneken, *Microelectron. Reliab.* **39**, 1445 (1999).

[24] B. E. Deal and A. S. Grove, *J. Appl. Phys.* **36**, 3770 (1965).

[25] D. P. Norton, *Solid-State Electr.*, **47**, 801 (2003).

[26] K. J. Yang and H. Chenming, *IEEE Trans. Electron Devices* **46**, 1500 (1999).

[27] Y. Fu and M. Willander, *Appl. Phys. A: Materials Science and Processing* **76**, 27 (2003).

[28] R. Clerc, B. De Salvo, G. Ghibaudo, G. Reimbold and G. Pananakakis, *Solid-State Electr.* **46**, 407 (2002).

[29] R. Clerc, A. S. Spinelli, G. Ghibaudo, C. Leroux and G. Pananakakis, *Microelectr. Reliab.* **41,** 1027 (2001).

[30] C. H. Choi, J. S. Goo, T. Y. Oh, Z. Yu, R. W. Dutton, A. C. Bayoumi, P. Vande Voorde, D. Vook and C. H. Diaz, *IEEE Electron Device Lett.* **20**, 292 (1999).

[31] C. H. Choi, Y. Wu, J. S. Goo, Z. Yu and R. W Dutton, *IEEE Trans. Electron Devices* **47**, 1843 (2000).

[32] S. H. Lo, D. A. Buchanan and Y. Taur, *IBM J. Res. Dev.* **43**, 327 (1999).

[33] K. Ahmed, E. Ibok, G. C. Yeap, C. F. Geoffrey, Q. Xiang, B. Ogle, J. Wortman and J. Hauser, *IEEE Trans. Electron Devices* **46**, 1650 (1999).

[34] G. D Wilk, R. M Wallace and J. M. Anthony, *J. Appl. Phys.* **89**, 10 (2001).

[35] A. I. Kingon, J. P. Maria and S. K. Streiffer, *Nature* **406,** 1032 (2000).

[36] H. Gruger, C. Kunath, E. Kurth, S. Sorge, W. Pufe and T. Pechstein, *Thin Solid Films* **447,** 509 (2004).

[37] C. K. Maiti, S. Chatterjee, G. K. Dalapati and S. K. Samanta, *Electr. Letters* **39,** 497 (2003).

[38] A. N. Chandorkar and P. Krishnamoorthi, *Proc. of SPIE - The International Society for Optical Engineering* **4746 II,** 1307 (2002).

[39] J. H. Cho, T. H. Cha, K. Y. Lim, D. G. Park and J. W. Park, *J. Electrochem. Soc.* **149,** G403 (2002).

[40] C. L. Benjamin, J. C. Yu and J. Y. Lee, *IEEE Electron Device Lett.* **22,** 221 (2001).

[41] B. C. Lai and J. Y. Lee, *IEEE Electron Device Lett.* **22,** 142 (2001).

[42] C. Liu and T. Y. Tseng, *J. of the European Ceramic Soc.* **24,** 1449 (2004).

[43] J. Legrand, S. Lhostis, Y Chang, L. Militaru, C. Lamard, A. Sibai, L. Auvray, M. Audier, J. P. Senateur, C. Dubourdieu and F. Ducroquet, *Microelectr. Eng.* **72,** 310 (2004).

[44] K. Ueno, I. H. Inoue, H. Akoh, M. Kawasaki, Y. Tokura and H. Takagi, *Appl. Phys. Lett.* **83,** 1775 (2003).

[45] J. A. Felix, J. R. Schwank, D. M. Fleetwood, M. R. Shaneyfelt and E. P. Gusev, *Microelectron. Reliab.* **44,** 563 (2004).

[46] S. Skordas, F. Papadatos, Z. Patel, N. Zubin, G. Nuesca, E. Eisenbraun, E. Gusev and A. E. Kaloyeros, *Materials Research Society Symposium – Proc.* **716,** 183 (2002).

[47] M. Copel, E. Cartier, E. Gusev, S. Guha, N. Bojarczuk and M. Poppeller, *Appl. Phys. Lett.* **78,** 2670 (2001).

[48] J. Lee, J. H. Kim and S. Seongil, *J. Appl. Phys.* **95,** 3733 (2004).

[49] K. Gloos, P. J. Koppinen and J. P. Pekola, *J. Phys. Cond. Matter* **15,** 1733 (2003).

[50] E. P. Gusev, M. Copel, E. Cartier, I. J. R. Baumvol, C. Krug and M. A. Gribelyuk, *Appl. Phys. Lett.* **76,** 176 (2000).

[51] C. K. Maiti, S. K. Samanta, G. K. Dalapati, S. K. Nandi and S. Chatterjee, *Microelectron. Eng.* **72,** 253 (2004).

[52] M. Kadoshima, M. Hiratani, Y. Shimamoto, K. Torii, M. Kazuyoshi, K. Hiroshi, N. Shinichiro and T. Nabatame, *Thin Solid Films* **424,** 224 (2003).

[53] T. Kauerauf, R. Degraeve, E. Eduard, C. Soens and G. Groeseneken, *IEEE Electron Device Lett.* **23**, 215 (2002).

[54] J. H. Han, G. Gao, Y. Widjaja, E. Garfunkel and C. B. Musgrave, *Surf. Sci.* **550,** 199 (2004).

[55] I. Kim, S. K. Han and C. M. Osburn, *J. Electrochem. Soc.* **151**, F22 (2004).

[56] D. W. Kim, F. E. Prins, T. Kim, S. Hwang, D. L. Kwong, S. K. Banerjee and C. H. Lee, *IEEE Trans. Electron Devices* **50**, 510 (2003).

[57] D. Wolfe, K. Flock, R. Therrien, R. Johnson, B. Rayner, L. Gunther, N. Brown, B. Claflin and G. Lucovsky, *Materials Research Society Symposium – Proc.* **567**, 343 (1999).

[58] H. Gruger, C. Kunath, E. Kurth, S. Sorge, W. Pufe and T. Pechstein, *Thin Solid Films* **447**, 509 (2004).

[59] S. Ramanathan, P. C. McIntyre, S. Guha and E. Gusev, *Appl. Phys. Lett.* **84**, 389 (2004).

[60] S. Zafar, B.H. Lee and J. Stathis, *IEEE Electron Device Lett.* **25**, 153 (2004).

[61] A. Kerber, E. Cartier, L. Pantisano, R. Degraeve, G. Groeseneken, H. E. Maes and U Schwalke, *Microelectron. Eng.* **72**, 267 (2004).

[62] E. P. Gusev, C. Cabral Jr., M. Copel, C. D'Emic and M. Gribelyuk, *Microelectron. Eng.* **69**, 145 (2003).

[63] K. J. Choi, W. C. Shin and S. G. Yoon, *J. Materials Research* **18**, 60 (2003).

[64] M. M. Frank, Y. J. Chabal, M. L. Green, A. Delabie, B. Brijs, G. D. Wilk, M. Y. Ho, E. B. O. Da Rosa, I. J. R. Baumvol and F. C. Stedlle, *Appl. Phys. Lett.* **83**, 740 (2003).

[65] S. Sayan, S. Aravamudhan, B. W. Busch, W. H. Schulte, F. Cosandey, G. D. Wilk, T. Gustafsson and E. Garfunkel, *J. Vac. Sci. and Tech., Part A: Vacuum, Surfaces and Films* **20**, 507 (2002).

[66] P. S. Lysaght, B. Foran, G. Bersuker, L. Larson, R. W. Murto and H. R. Huff, *Materials Research Society Symposium – Proc.* **747**, 133 (2003).

[67] P. S. Lysaght, B. Foran, S. Stemmer, G. Bersuker, J. Bennett, R. Tichy, L. Larson and H. R. Huff, *Microelectron. Eng.* **69**, 182 (2003).

[68] T. Gougousi, M. J. Kelly and G. N. Parsons, *Materials Research Society Symposium – Proc.* **765**, 79 (2003).

[69] S. Gopalan, K. Onishi, R. Nieh, C. S. Kang, R. Choi, H. J. Cho, S. Krishnan and J. C. Lee, *Appl. Phys. Lett.* **80**, 4416 (2002).

[70] Z. M. Rittersma, E. Naburgh, T. Dao, A. H. C. Hendriks, W. F. A. Besling, E. Tois, E. Vainonen-Ahlgren, M. Tuominen and S. Haukka, *Electrochem. Solid-State Lett.* **6**, F21 (2003).

[71] E. Dharmarajan, W. J. Qi, R. Nieh, L. Kang, K. Onishi and J. C. Lee, *Materials Research Society Symposium – Proc.* **648,** P6161 (2001).

[72] S. Van Elshocht, M. Caymax, S. De Gendt, T. Conard, J. Petry, M. Claes, T. Witters, C. Zhao, B. Brijs, O. Richard, H. Bender, W. Vandervorst, R. Carter, J. Kluth, L. Date, D. Pique and M. M. Heyns, *Materials Research Society Symposium – Proc.* **745**, 197 (2002)

[73] B. C. Hendrix, A. S. Borovik, Z. Wang, C. Xu, J. F. Roeder, T. H. Baum, M. J. Bevan, M. R. Visokay, J. J. Chambers, A. L. P. Rotondaro, H. Bu and L. Colombo, *Materials Research Society Symposium – Proc.* **716**, 273 (2002).

[74] J. L. Roberts, P. A. Williams, A. C. Jones, P. Marshall, P. R. Chalker, J. F. Bickley, H. O. Davies and L. M. Smith, *Materials Research Society Symposium – Proc.* **745**, 173 (2002).

[75] K. Nakazato, K. Itoh, H. Mizuta and H. Ahmed, *Electr. Lett.* **35,** 848 (1999).

[76] Y. Wei, R. M. Wallace and A. C. Seabaugh, *J. Appl. Phys.* **81,** 6415 (1997).

[77] A. Dutta, S. P. Lee, S. Hatani and S. Oda, *Appl. Phys. Lett.* **75,** 1422 (1999).

[78] T. Tanamoto, *Phys. Rev. A* **61,** 022305 (2000).

[79] P. M. Fauchet, J. Ruan and J. Chen, *Materials Research Society Symposium – Proc.* **737**, 407 (2003).

[80] S. Zollner, A. Konkar, R. Liu, H. Yapa, P. F. Dryer, V. A. Neeley, Q. Xie, G. F. Grom, Q. Zhu, R. Krishnan, P. M. Fauchet and L. V. Tsybeskov, *Materials Research Society Symposium – Proc.* **638**, F511 (2001).

[81] S. Y. Chou and A. E. Gordon, *Appl. Phys. Lett.* **60,** 1827 (1992).

[82] Q. Y. Ye, R. Tsu and E. H. Nicollian, *Phys. Rev. B* **44,** 1806 (1991).

[83] E. H. Nicollian and R. Tsu, *J. Appl. Phys.* **74,** 4020 (1993).

[84] K. R. Farmer, M. O. Andersson and O. Engstrom, *Appl. Phys. Lett.* **60,** 730 (1992).

[85] S. J. Wrazien, Y. Zhao, J. D. Krayer and M. H. White, *Solid-State Electr.* **47**, 885 (2003).

[86] M. She, H. Takeuchi and T. J. King, *IEEE Electron Device Lett.* **24**, 309 (2003).

[87] Y. Yang, A. Purwar and M. H. White, *Solid-State Electr.* **43**, 2025 (1999).

[88] H. A. R. Wegener, A. J. Lincoln, H. C. Pao, M. R. O'Connell and R. E. Oleksiak, in *Tech. Dig. Int. Electron Devices Meet.* (IEEE, Washington D.C., 1967).

[89] T. P. Ma and J. P. Han, *IEEE Electron Device Lett.* **23**, 386(2002).

[90] S. Tiwari, F. Rana, K. Chan, H. Hanafi, W. Chan and D. A. Buchanan, in *Tech. Dig. Int. Electron Devices Meet.*, 521 (IEEE, New York, 1995).

[91] H. Hanafi, S. Tiwari and I. Khan, *Trans. Elec. Dev.* **43**, 1553 (1996).

[92] R. Krishnan, T. D. Krauss and P. M. Fauchet, *Materials Research Society Symposium – Proc.* **737**, 253 (2003).

[93] M. L. Ostraat and J. W. De Blauwe, *Materials Research Society Symposium – Proc.* **686**, 135 (2002).

[94] M. L. Ostraat, J. W. De Blauwe, M. L. Green, L. D. Bell, M. L. Brongersma, J. Casperson, R. C. Flagan and H. A. Atwater, *Appl. Phys. Lett.* **79**, 433 (2001).

[95] K. Han, I. Kim and H. Shin, *IEEE Electron Device Lett.* **21**, 313 (2000).