The Dissertation Committee for Yang-Yu Fan
certifies that this is the approved version of the following dissertation:

Voltage and Temperature Dependent Gate Capacitance and Current Model for High-K Gate Dielectric Stack

Committee:

Sanjay K. Banerjee, Supervisor

Leonard F. Register, Co-Supervisor

Jack C. Lee

Dim-Lee Kwong

Graham F. Carey

Voltage and Temperature Dependent Gate Capacitance and Current Model for High-K Gate Dielectric Stack

by

Yang-Yu Fan, B.S., M.S.E.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2002

To my parents

Acknowledgments

I would like to express my great gratitude to the many people who helped make this dissertation possible. First, I would like to thank my supervisors Professor Sanjay K. Banerjee and Leonard F. Register for their guidance and the freedom and independence they encourage in the device modeling group at University of Texas at Austin. Under such environment, I was able to explore this research work to the extent that I never imagined. Those efforts would not have resulted in any fruitful outcome without their insightful and timely guidance. I am thankful to Professor Jack C. Lee, Dim-Lee Kwong, and Gerald Lucovsky, and their students for the assistance in experiment and the discussion. I am also indebted to George A. Brown at International SEMATECH. My gratitude goes to my doctoral committee and to the professors, students, and staff at Microelectronics Research Center. Especially, I would like to thank Jean Toll, Renee Nieh, Wen-Jie Qi, Siva Mudanai, Xin Wang, Mukund Swaminathan, Dong-Won Kim, and Katsunori Onishi. I am thankful to Douglas Roberts at Motorola, Inc., Judy An and Qi Xiang at AMD, Inc. for their guidance during the summer projects. I am grateful to my friends for the encouragement, advice, and the unforgettable times we spent together. I would like to thank Yi-Chang, Amy, Lai, Maga, Hung-Ming, I-Chi, and Daisy for their time and energy to endure everything I dumped to them. Finally, I dedicate this dissertation and everything I ever accomplished to my parents.

Voltage and Temperature Dependent Gate Capacitance and Current Model for High-K Gate Dielectric Stack

Publication No. _____

Yang-Yu Fan, Ph.D.
The University of Texas at Austin, 2002

Supervisors: Sanjay K. Banerjee
Leonard F. Register

High-dielectric-constant (High-K) materials are being pursued as alternative gate dielectrics to SiO_2 for next-generation Metal-Oxide-Silicon Field-Effect Transistors (MOSFETs). Great efforts are being made to optimize these materials for transistor applications. However, a major hurdle arises from the uncertainties about material properties, especially those of the thin films, and the physical mechanisms that affect the gate capacitance and gate current. In this work, based on the energy-dispersion relation in different regions of the Gate-Dielectric-Silicon system, a tunneling model is developed to understand the gate current as a function of voltage and temperature. A Franz 2-band dispersion relation is assumed in the bandgap to characterize the gate dielectric for tunneling. Quantum confinement effects in the silicon channel, direct and Fowler-Nordheim tunneling, and thermionic emission gate currents are considered simultaneously. The gate capacitance is self-consistently calculated from

the Schrodinger and Poisson equations subject to Fermi-Dirac statistics, using the same band structure in the silicon as used for tunneling injection. A self-consistent gate capacitance-current vs. voltage/temperature model is thus established. The model is implemented for both the silicon conduction and valence bands, and both gate- and substrate-injected currents. SiO_2 devices and multi-layer high-K gate dielectric stack structures such as ZrO_2 and HO_2 are studied by an integrated simulation and experimental investigation. A gate capacitance-current analysis method is explored for understanding the gate current mechanisms and extracting device structure parameters. A trend study for tradeoffs between low equivalent oxide thickness (EOT) and low gate current is performed qualitatively based on reasonable experimental data. Write/Erase/Retention time-dependent characteristics of non-volatile floating gate devices are modeled. Tend study on scaling the control and tunnel oxide is performed for SiO_2 and ZrO_2.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The purpose of this research is to study the leakage current through high-dielectric-constant (high-K) stacked gate dielectric structures. The intent is to establish a general physical model for Metal-Oxide-Silicon (MOS) transistor device design issues. Through simulation and experimental analysis, the device behavior will be studied.

1.1 MOSFET Scaling

Continued MOSFET scaling is the primary driving force that boosts the semiconductor industry [38, 47, 54]. As paced by Moore's Law, consistent improvements in the density, power, and performance have been achieved for more than 30 years. Such a successful scaling mandates that the following rule be satisfied: while the device performance is improved as a result of a shorter channel length, the characteristics of the long-channel MOSFETs such as low off-state leakage currents and good sub-threshold characteristics must be retained [1, 5, 40].

Regardless of the challenges in fabrication, adverse device behavior which is not seen in long channel devices, generally regarded as short channel

effects (SCE), require careful consideration of the device design. Concerns of the device reliability should also be considered. Two of these are hot carrier effects and gate leakage currents. The hot carrier effects degrade the device performance and shorten the device operating lifetime; gate leakage currents can cause the gate dielectric breakdown and increase the power consumption rate. With the demands of more-functionality and faster microprocessors and the application of mobile appliances, power consumption becomes a major concern for both device engineering and circuit design. Thus, the ability to overcome current physical technology limits as well as tradeoffs in circuit design will determine if MOS transistors can continue to be scaled down in the future.

To suppress short channel effects and to obtain maximum performance MOSFETs concurrently, one needs to take into account the tradeoffs among critical parameters such as critical channel length L_c, gate dielectric thickness t_{ox}, channel doping N_a, and junction depth X_j [1, 5, 40]. One scaling formula suggested as a design rule is [40],

$$L_c = 2.2\mu m^{-2} \left(\frac{\delta V_t}{\delta V_{ds}} \right)^{-0.37} (t_{ox} + 0.012\mu m) (W_{sd} + 0.15\mu m) (X_j + 2.9\mu m) \tag{1.1}$$

where $\delta V_t / \delta V_{ds}$, representing the drain-induced barrier lowering (DIBL) effect, is obtained from the parallel shift of the $\log I_d$ vs. V_{gs} curves at a given drain current level; W_{sd} is the sum of the depletion widths of the source and drain which is a function of channel doping, N_a; the gate dielectric is silicon dioxide.

The success of scaling down MOSFETs has been achieved by aggressive engineering of the source/drain and well regions and decreasing gate oxide thickness as well as the development of new lithography tools, masks, photoresist materials and critical dimension etch processes. However, despite these advances in crucial process technologies and the resultant smaller feature sizes, it has become clear that the scaled device performance has been compromised [1]. The reason is that the traditional materials, silicon (Si), silicon dioxide (SiO_2), and polysilicon (Poly-Si) have been pushed to their fundamental limits. Further device scaling thus requires new materials and device structures.

1.2 SiO_2 Scaling and its Challenges

Traditionally, SiO_2 has been used as the gate dielectric due to its process advantages and good interfacial properties with Si [2, 16, 29, 30]. With increasing channel doping concentration to suppress short channel effects, the SiO_2 thickness has been reduced, which results in a larger gate capacitance, to maintain gate control over the channel. Large gate capacitance is needed to invert the silicon surface to a sufficient sheet charge density to obtain the desired drive current for the given supply voltage. This requirement is further demanded for deep-sub-micron or even-smaller MOSFETs.

In the deep-submicron or smaller regime, quantum confinement effects of carriers in the channel and polysilicon depletion effects in the poly gate become more important. These effects cause the gate charge and inversion layer charge centroids be located further away from the SiO_2/poly-Si and SiO_2/Si

interfaces, respectively. As a result, fewer carriers are induced in the channel to contribute to the drive current. In Intel's 0.25 or 0.18μm-generation technology, the increase between the two centroids in addition to the SiO_2 physical thickness is $\sim$ 7Å in inversion at the operating voltage. A SiO_2 physical thickness of 16Å thus is utilized to achieve an effective oxide thickness of 23Å in order to obtain sufficiently high drive currents required in such generations [48].

Higher gate capacitance may become crucial to obtain high drive currents for even smaller devices. It has been reported that the drive (drain) current is ultimately limited by the thermal injection current from the source into the channel [34]. The velocity overshoot in the channel does not extend such a limit. As a result, increasing the gate capacitance, i.e. increasing the carriers in the channel, becomes one important key to further scaling the MOS transistor to obtain better performance.

If the increased gate capacitance is solely obtained by decreasing the oxide thickness, in the year 2004 [1] the required oxide thickness is projected to be $\sim$ 10Å, which is about three atomic layers. MOS devices fabricated with a SiO_2 physical thickness of 15Å have been reported [39]. However, concerns of their manufacturability for ULSI should be considered. Those include thickness variation, penetration of impurities from the gate through the gate dielectric, and the reliability and lifetimes of devices made with such thin films [6].

Another concern is the gate leakage current, which is the most detri-

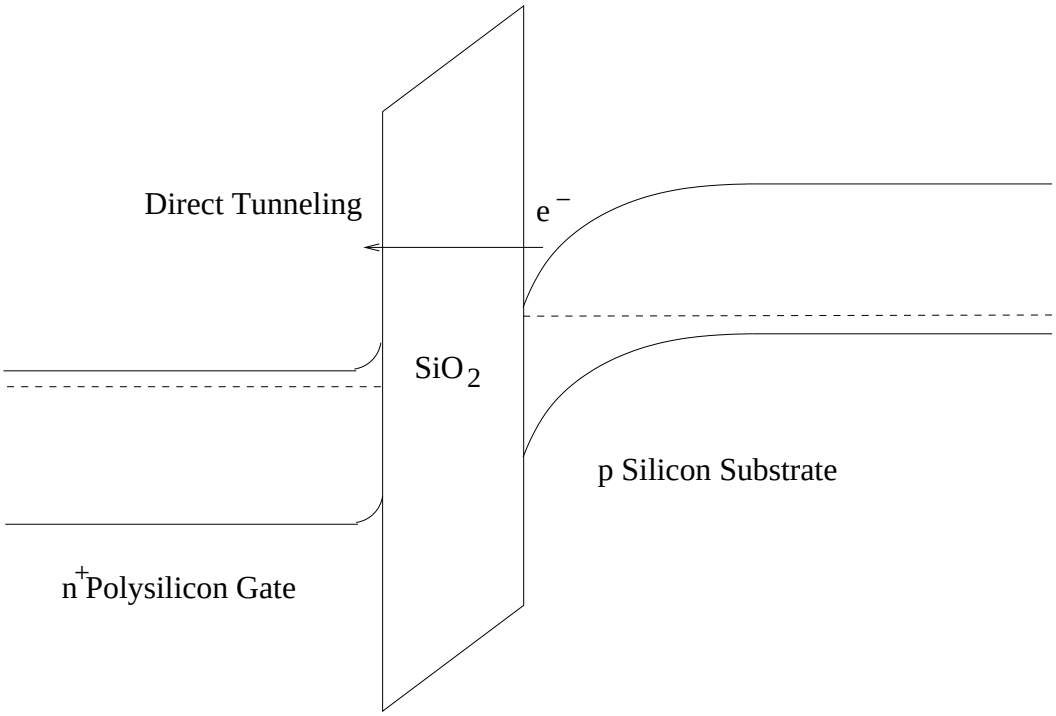


Figure 1.1: Direct tunneling currents in a NMOSFET.

mental effect of further SiO_2 scaling and which poses the fundamental limit upon scaling SiO_2 thickness [21]. Figure 1.1 shows the band diagram of a MOS structure and the direct tunneling mechanism. Since the barrier height between conduction bands of Si and SiO_2 is finite, $\sim$3.15eV, direct tunneling becomes non-negligible for SiO_2 thickness less than 30Å. In the region where direct tunneling dominates, the gate leakage current is exponentially dependent upon the SiO_2 thickness [36]. In fact, at a gate bias $\sim$ 1V, the gate leakage current changes from $\sim 10^{-5}$A/cm^2 at SiO_2 thickness of 30Å to $\sim 10^{2}$A/cm^2 at 15Å [21, 39].

1.3 Alternative Gate Dielectric: High-K Dielectric Stack

The capacitance is proportional to the dielectric constant and inversely proportional to the thickness of a thin film. Thus, by using high-K materials as the gate dielectric, the required large gate capacitance can still be obtained while the direct tunneling current can be reduced as a result of the greater physical thickness. For a given equivalent oxide thickness (EOT), the required physical thickness of the high-K film is

$$t_{\mathrm{hk}} = \mathrm{EOT} \cdot \frac{\varepsilon_{\mathrm{hk}}}{\varepsilon_{\mathrm{ox}}}, \tag{1.2}$$

where $\varepsilon_{\mathrm{hk}}$ and $\varepsilon_{\mathrm{ox}}$ are the dielectric constants of the high-K material and SiO_2, respectively.

Many high-K materials have been suggested as alternative gate dielectrics. Metal oxides such as ZrO_2, HfO_2, Ta_2O_5, TiO_2, Y_2O_3, and Al_2O_3 [3, 7, 17, 27, 31, 41], and ferroelectric materials such as BST (barium strontium titanate) [25] have been suggested as potential candidates. However, the challenge of integrating these materials with the current CMOS process still remains. These include thermal stability, compatibility with the gate, and interfacial properties with silicon. One important process issue is that a silicate layer is generally found between the high-K layer and the silicon substrate. This interfacial layer usually has a lower dielectric constant, and hence degrades the capacitance of the gate stack. Furthermore, the band gaps of these materials along with their band offsets with silicon are important material properties for reducing the gate leakage current. Unfortunately, smaller band

	Band Gap(eV)	ε_∞	Conduction Band Offset(eV)
SiO_2	9	2.25	3.5
Si_3N_4	5.3	4.1	2.4
Ta_2O_5	4.4	4.84	0.3
$BaTiO_3$	3.3	6.1	-0.1
$BaZrO_3$	5.3	4	0.8
TiO_2	3.05	7.8	0.05
ZrO_2	5.8	4.8	1.4
HfO_2	6	4	1.5
Al_2O_3	8.8	3.4	2.8
Y_2O_3	6	4.4	2.3
La_2O_3	6	4	2.3
$ZrSiO_4$	6	3.8	1.5
$SrBi_2Ta_2O_9$	4.1	5.3	0

Table 1.1: Experimental values of conduction band offsets of several high K materials with Si, obtained by Robertson et al [43]. ε_∞ is obtained as the square root of the refractive index.

gaps are usually observed in materials with higher dielectric constants [7]. This generally indirectly implies that lower potential barriers are expected for high-K materials, and thicker high-K films are required to reduce the gate current. Table 1.1 lists the band gaps and band offsets of several high K materials, which were obtained by a charge-neutrality-level method [43].

The requirement of a greater physical thickness has another impact on the MOSFET electrical characteristics. In parallel-plate capacitor theory, as the ratio of the distance between the plates to the length or width of the plates becomes large, fringing fields can no longer be neglected. With this high fringing field, the short channel effects will be aggravated due to the thick high-K film [22]. Thus tradeoffs between the gate leakage current and

short channel effects are required in designing next-generation MOSFETs if a higher-K gate dielectric is used [22, 24].

1.4 Scope and Organization

Low EOT (high gate capacitance) and gate currents are simultaneously pursued in device design. However, compromises are required because electrical property trends of high-K materials do not favor such a pursuit . Through simulation, optimum device design can be achieved with low cost and short technology development cycle times.

To achieve such a goal, simultaneous matching of gate capacitance and current simulation with experiment is a key. In this work, an energy-dispersion-based self-consistent gate capacitance and current model will be first established and applied to study gate capacitance and current behavior of ZrO_2 devices (Chapter 2). Dielectric stack engineering is explored by considering parameters to vary the energy dispersion; thus, tradeoffs between the EOT and gate current is qualitatively studied (Chapter 3). Through integrated simulation and experimental analysis, the gate capacitance-current behavior will be studied. Device structure parameters will be extracted and compared with those values obtained by other methods (Chapter 4). Using the established gate capacitance and current model, a trend study on scaling the control and tunnel oxide in floating-gate nonvolatile devices will be performed (Chapter 5). Time-dependent characteristics for write, erase, and retention voltages are also modeled. Finally, conclusion swill be drawn in Chapter 6.

Chapter 2

Fundamental Issues on High-K Gate Capacitance and Current Modeling

A voltage- and temperature-dependent gate capacitance and current model is presented and applied to study ZrO_2 NMOSCAPs to illustrate the fundamental issues regarding the high-K gate stack structures.

2.1 Introduction

Gate capacitance has been used for extracting the EOT [32, 42, 55], but the capacitance-voltage behavior should be considered concurrently with current-voltage calculation. The charge redistribution in the Gate-Dielectric-Silicon structure must be determined self-consistently by solving the Poisson-Schrodinger equations. The gate capacitance provides an experimental probe of the charge distribution at different gate voltages, and this is key to modeling the gate current. Temperature-dependence study of gate currents and gate capacitance of MOS capacitors with SiO_2 shows this correlation: the temperature-dependent region of the gate capacitance, which is near the flat-band, is the same as that of the gate current, (Figure 4.5). Such temperature dependence is attributed to temperature dependence of the charge distribu-

tion or "supply function" instead of the temperature-dependence of the carrier transport in the SiO_2. In this work, the *simultaneous* agreement of gate capacitance and gate current simulation with experiment has been obtained.

High-K materials generally have smaller band gaps and smaller band offsets with silicon than SiO_2 [7, 9]. In addition to direct tunneling, Fowler-Nordheim (FN) tunneling can be important because of the smaller band offsets. The thermionic emission also needs to be considered for materials with extremely small band offsets and for transistors in which hot carriers are of concern. Gate currents from the silicon conduction band and valence band may be non-negligible at the same time. Furthermore, an interfacial layer is generally found between the high-K layer and the silicon substrate. These all impose difficulties in terms of understanding the experimental gate currents.

With increasing normal fields as a result of high doping concentrations and reduced EOTs, quantum confinement effects can no longer be neglected [13, 32, 42, 55]. Based on the energy-dispersion (E versus k) in different regions of the Gate-Dielectric-Silicon system, a gate current model is established to consider quantum confinement effects in the silicon channel, direct, FN tunneling, and thermionic emission gate currents. Such a gate current model is applied to the inversion and accumulation regions and both the conduction and valence band components [12].

2.2 Models

The independent particle approach [19] is adopted to model the tunneling currents through the gate dielectric. If the energy dispersion relation in each region is available, the tunneling probability is obtained by a WKB method. In the originally proposed approach, the wave number vector parallel to region interfaces, $k_{||}$, is required to be conserved. However, well-known inconsistencies between the assumption and experiment are found between gate currents in SiO_2-MOS devices fabricated on Si(110) and Si(111) [51], and this issue has not been resolved. In this work, simultaneous agreement of gate current simulation with experiment in inversion and accumulation is achieved for SiO_2 devices fabricated on Si(100) with $k_{||}$ conserved.

A physical picture of this model can be described in terms of electrons tunneling from the silicon substrate to the gate, (Figure 2.1, 2.2). The incident component of the electron wave function, subject to the band structure, (E, k_x, k_y, k_z), tunnels through the barrier and contributes to the transmitted currents. The energy-dispersion relation in the dielectric bandgap and available states in the gate determine the transmission probability. The continuity of electron flux allows the gate-injected electron current to be calculated by the transmission of electrons into the silicon at silicon-dielectric interface.

2.2.1 Energy Dispersion

The carrier states in the silicon channel are roughly divided into bound states and unbound states. Bound states are quantized states inside the quan-

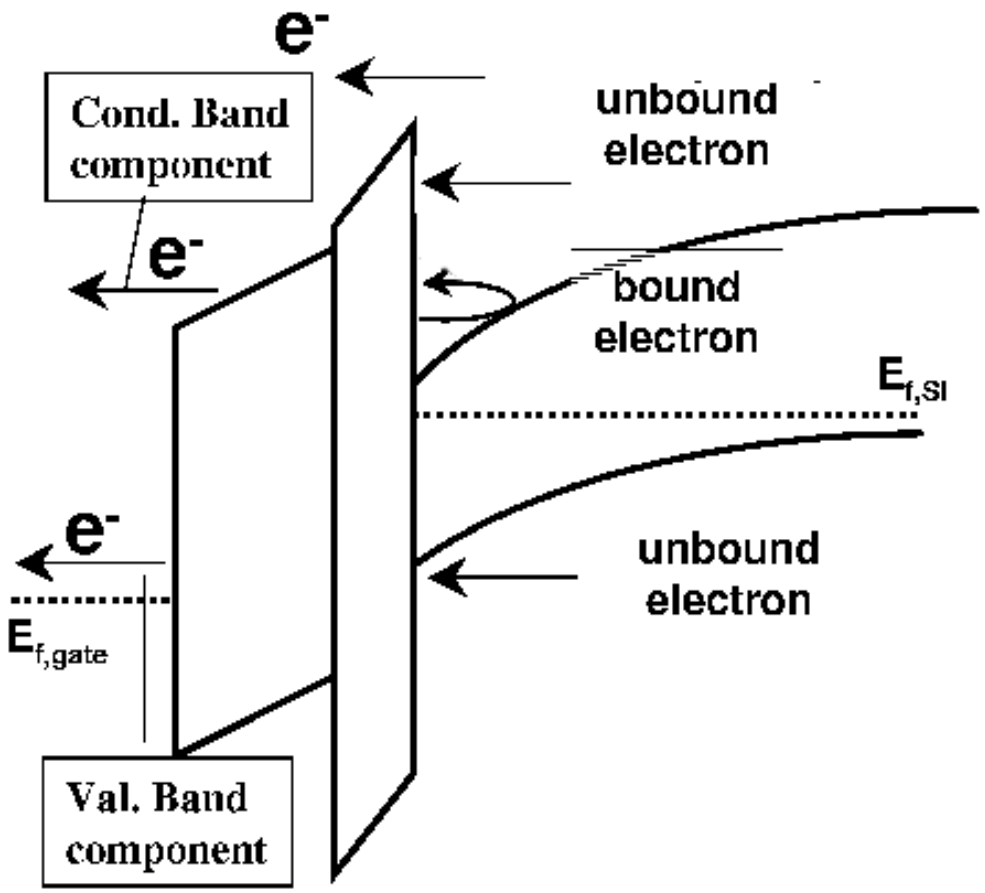


Figure 2.1: Gate current model band diagram of a Gate-Dielectric-Silicon system being biased in the inversion region. Direct, Fowler-Nordheim tunneling, and thermionic emission gate currents are simultaneously considered.

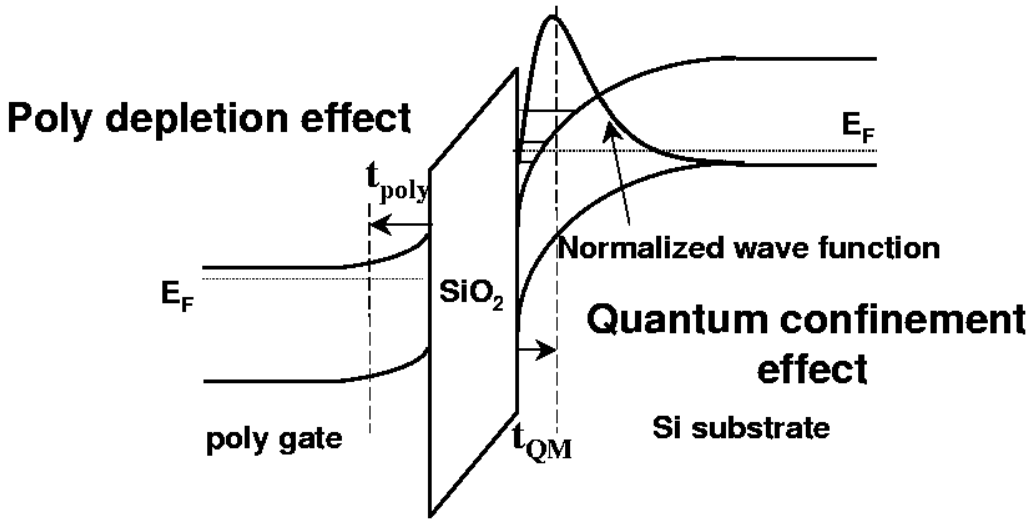


Figure 2.2: Gate capacitance model band diagram of a Gate-Dielectric-Silicon system being biased in the inversion region.

tum well formed by the silicon band edge. Subband structures and envelope functions are obtained by solving the Schrodinger equation in k-space using a full-band formalism [23, 49]. Those states outside the quantum well are approximated as free and unbound, and the energy-dispersion at the silicon-dielectric interface is approximated by the bulk-silicon band structure. The same charge distribution and subband structures in the silicon channel are used both for gate capacitance and gate current calculation.

The apparent energy dispersion "seen" by the tunneling electron in the dielectric band gap is approximated by Franz's two-band model [14],

$$\frac{1}{k^2} = \frac{\hbar^2}{2m_c(E - E_c)} + \frac{\hbar^2}{2m_v(E_v - E)}, \tag{2.1}$$

where E is the total energy of the electron, m_c and m_v are *band-edge* effective

masses for the conduction and valence bands, respectively, of the dielectric, and E_c and E_v are the energies of the respective band edges. Such dispersion relation has previously been shown to be effective approximation for conventional oxides [35, 37]. m_c and m_v are unknown parameters to be extracted by comparing gate current simulation with experiment.

2.2.2 Gate Current Calculation

The tunneling probability, $T(E, k_{||})$, through the dielectric is calculated by the WKB method [26] when $k_z^2 < 0$,

$$-\ln T(E, k_{||}) = \int_{\text{dielectric}} |k_z|\, dz, \tag{2.2}$$

with

$$k_z^2 = k^2 - k_{||}^2, \tag{2.3}$$

where k^2 is calculated from Franz dispersion, and $k_{||}$-conservation is applied.

The total energy E is used, thus direct, FN tunneling, and thermionic emission are considered simultaneously:

- Direct tunneling: $k_z^2 < 0$ across the whole dielectric
- FN tunneling: k_z^2 passes through 0 within the dielectric.
- Thermionic emission: $k_z^2 > 0$ across the whole dielectric

These mechanisms are all considered elastic as the total energy and $k_{||}$ are conserved during transport through the dielectric. The image force is lower in

higher-K materials and reduced by the quantum mechanical repulsion of both bound and unbound carriers from the interface. Thus, its effect on barrier lowering is not considered.

Finally, the total gate current density, J_g, is calculated as

$$J_g = J_{\text{Si-conduction}} + J_{\text{Si-valence}}, \tag{2.4}$$

where $J_{\text{Si-conduction}}$ is from the conduction-band longitudinal valleys. With $k_{||}$-conservation, the effective barrier for transverse valleys is higher and tunneling from them is negligible.

Each component J_i, where i designates the band, is calculated separately for unbound states and bound states. For unbound states in the conduction band

$$J_{i,\text{unbound}} = \frac{2q}{(2\pi)^3} \int_{E_{z,\text{min,c}}}^{\infty} dE_z \int_{k_{||}=k_{||,min}}^{\infty} dk_{||}^2 T(E, k_{||})(f_{FD,\text{gate}} - f_{FD,\text{Si}}), \tag{2.5}$$

and for unbound states in the valence band

$$J_{i,\text{unbound}} = \frac{2q}{(2\pi)^3} \int_{-\infty}^{E_{z,\text{max,v}}} dE_z \int_{k_{||}=(0,0)}^{\infty} dk_{||}^2 T(E, k_{||})(f_{FD,\text{gate}} - f_{FD,\text{Si}}), \tag{2.6}$$

where $f_{FD,\text{gate}}$ and $f_{FD,\text{Si}}$ are the Fermi-Dirac distribution functions for the gate and silicon substrate, respectively; $E_{z,\text{min,c}}(E_{z,\text{max,v}})$ is the lowest (highest) unbound state in the conduction (valence) band; q is the charge of electron. While for bound states,

$$J_{i,\text{bound}} = \frac{2q}{(2\pi)^2} \sum_{\mu} \tau_{\mu}^{-1} \cdot \int_{k_{||}=k_{||,min}}^{\infty} dk_{||}^2 T(E, k_{||})(f_{FD,\text{gate}} - f_{FD,\text{Si}}), \tag{2.7}$$

where τ_μ^{-1} is approximated by the surface impact frequency for a 2D subband μ [46],

$$\tau_\mu^{-1} = \left(\frac{\partial I}{\partial E}\right)^{-1}_{\text{subband}\,\mu} \tag{2.8}$$

with the action integral between the classical turning points

$$I = \oint k_z^{(\mu)} dz, \tag{2.9}$$

and

$$k_z^{(\mu)}(z) = \frac{1}{\hbar}\sqrt{2m_z(E_\mu - E_{\text{band min.}}(z))}, \tag{2.10}$$

where the contour integral is around the quantum well formed in the channel; E_μ is the minimum energy of μ^{th} subband, $E_{\text{band min.}}$ the band edge in which the subbands formed, and m_z the effective mass along the tunneling direction z.

$m_z = 0.2m_e$ and $m_z = 0.9m_e$ are used for the conduction-band longitudinal and transverse valleys, respectively, where m_e is the free electron mass. Because the effective mass approximation is not very accurate for the valence band, m_z then is treated as a subband-independent adjustable parameter. From comparison of gate current simulation with experiment for a SiO_2-PMOSFET in inversion, it is found that $m_z = 0.10 \sim 0.25m_e$ gives good agreement, (Figure 2.3). The SiO_2 thickness is extracted by comparison of C_g-V_g simulation with experiment in accumulation, (Figure 2.3). Similar agreement is also achieved for a SiO_2-NMOSFET, which was fabricated on Si(100), (Figure 2.4). The gate current simulation agrees well with experiment both in inversion and accumulation. The band gap and conduction band

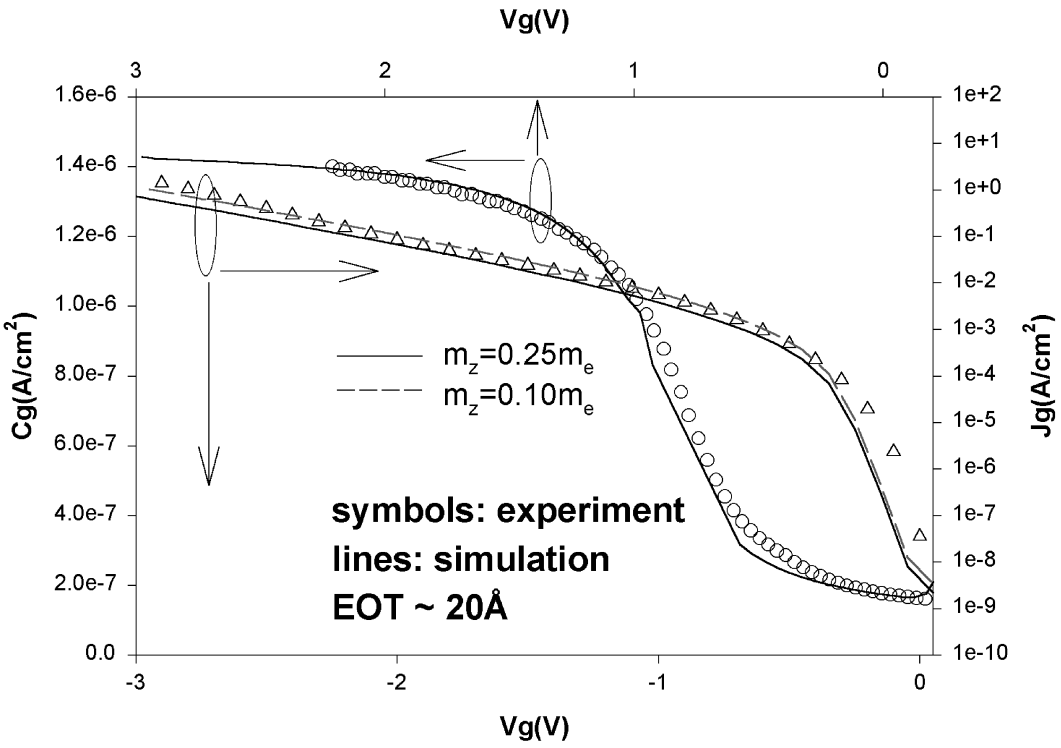


Figure 2.3: Comparison of gate current and gate capacitance simulation with the experimental data for a SiO_2-PMOSFET with EOT of $\sim$20Å. Area = 10^{-4}cm^2.

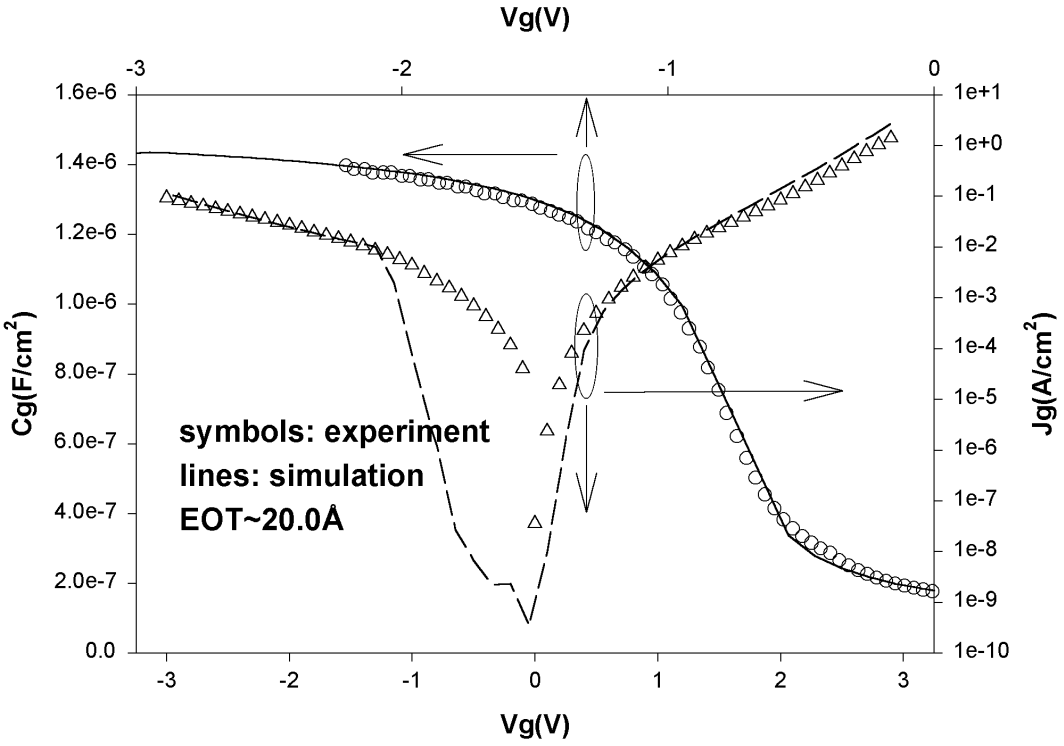


Figure 2.4: Comparison of gate current and gate capacitance simulation with the experimental data for a SiO_2-NMOSFET with EOT of ~20Å. Area = 10^{-4}cm^2.

offset with silicon of SiO_2 are 9.0eV and 3.15eV, respectively. The band-edge effective masses are adjustable parameters but remain the same for the silicon conduction band and valence band components, respectively, whether for substrate-injected or gate-injected currents. More about SiO_2 devices will be discussed in Chapter 4.

2.2.3 Gate Current vs. Voltage/Temperature Behavior

Near the flat-band region, $V_g \sim -0.8$V in Figure 2.4, the quantum well is wide and shallow. Quantization effects due to the quantum well formed by the band edge become less important. Thus, a clear division between the bound and unbound states can cause significant errors in gate current calculation, which result in abnormal I_g-V_g behavior in simulation. The error may originate from either the impact frequency or the supply of carriers. Below flat-band, only energetic carries can tunnel elastically, greatly reducing the calculated current. In practice, other mechanisms may dominate. A possible mechanism might be related to the interface states, which is not considered in the model. Near flat-band, the carrier density is much lower than in accumulation or inversion. The trapped charges can contribute appreciably to the total gate currents; empty traps may assist the carrier tunneling into/out of the silicon substrate. Contribution from the overlap regions with the source and drain may also be important in that region.

Gate current characteristics at different temperatures are shown in Figure 2.5. NMOSFETs with single-layer dielectric of 3.9 with different thicknesses and barrier heights are simulated. The flat band voltages of these devices are ~ -0.8V. When the potential barrier of dielectric is high and/or thin, direct tunneling prevails. However, when the potential barrier is low, transition from direct to FN tunneling is seen. With the potential barrier of 1.0eV high and 40Å wide, such transition occurs around $V_g = -2$V in accumulation and $V_g = 0.8$V in inversion.

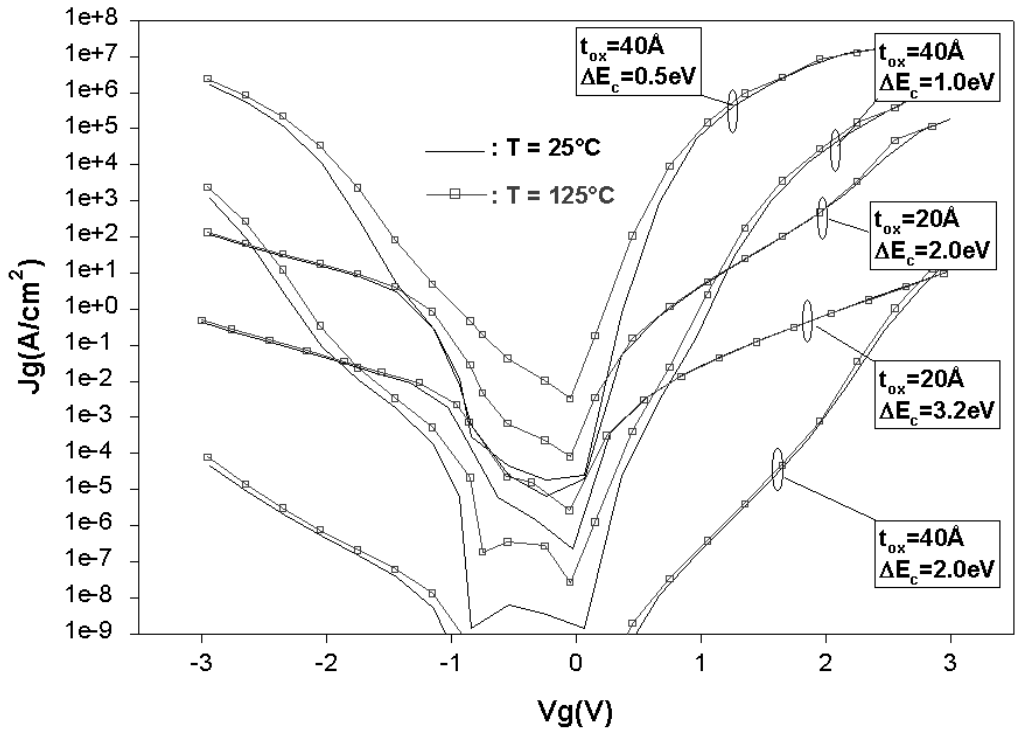


Figure 2.5: Simulated I_g-V_g at 25°C and 125°C for n-channel MOSFETs with different dielectric thicknesses and conduction band offsets with silicon. Single layer of dielectric of dielectric constant of 3.9 is assumed in the devices.

More energetic carriers see a relatively small barrier. Therefore, the hotter the carrier distribution, the greater the tunneling. This field-assisted thermionic effect is particularly significant when the barrier is low and/or thick, (Figure 2.5). Polarity effects are also seen in the temperature dependence of gate current, which is stronger in accumulation than in inversion. Such effects are not due to the transport in the dielectric but the charge supply and asymmetry of the gate stack.

2.3 ZrO_2 and HfO_2 NMOSCAP C_g,I_g-V_g Analysis

Different materials such as ZrO_2 and HfO_2 have shown promise for transistor application [20, 28, 41]. Owing to the complex fabrication process of transistors, most study has been conducted on MOSCAPs. The uncertainty of the material properties, especially for the thin films, imposes another difficulty in terms of understanding the experimental results. In this work, through C_g-V_g and I_g-V_g simulations, the material properties of ZrO_2-NMOSCAPs are extracted in the accumulation region.

2.3.1 C_g vs. V_g

Two samples of different thicknesses of ZrO_2 were fabricated, labeled as A and B. According to TEM analysis, an interfacial Zr-silicate layer exists between the ZrO_2-layer and the silicon substrate as reported in [52]. The physical thickness of each layer is found as: for sample A, $t_{ZrO_2} = 31 \sim 39$Å, and $t_{int} = 6 \sim 10$Å; for sample B, $t_{ZrO_2} = 36 \sim 40$Å, and $t_{int} = 7 \sim 9$Å, where t_{ZrO_2} and t_{int} are physical thicknesses of the ZrO_2 and interfacial silicate layers, respectively. Because of the thickness variation, the experimental data being analyzed for either sample were measured from the same device.

To analyze the devices, C_g-V_g simulation is first compared with the experimental data to extract the related device structural parameters, (Figure 2.6). At this step, EOT is the only physical parameter that can be extracted and is required for the dielectric stack. The interface state effects are neglected. Effects of fixed charges are compensated by adjusting the metal work function

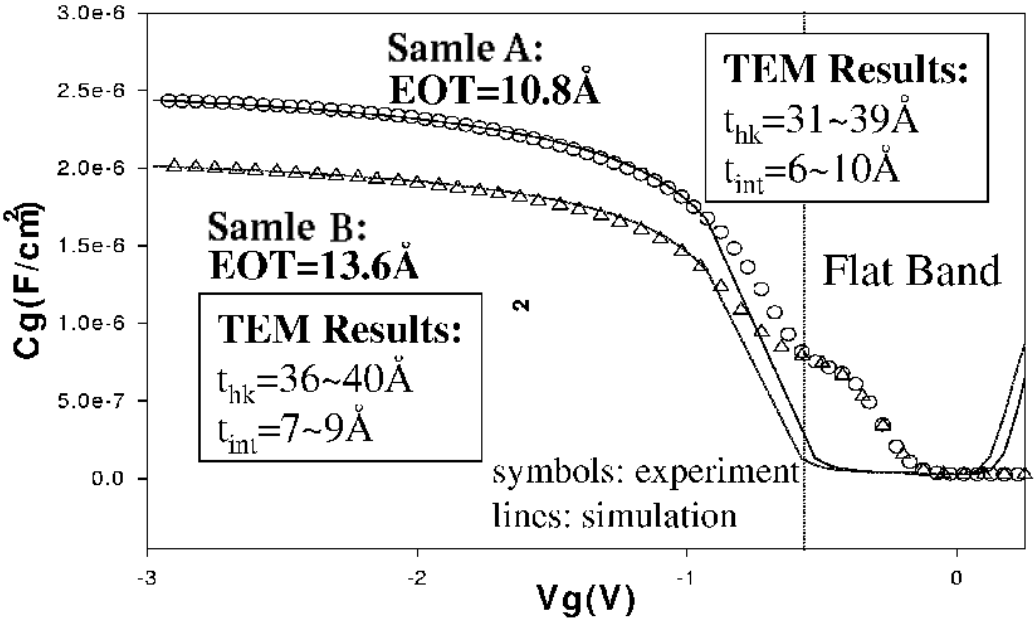


Figure 2.6: Comparison of C_g-V_g simulation with the experimental data in accumulation for ZrO_2-NMOSCAPs with different dielectric thicknesses, samples A and B.

to fit the experimental data. For the fabrication process described above, it was found that the substrate doping concentration is $5 \times 10^{15}\text{cm}^{-3}$, the metal work function is 4.50eV for sample A and 4.45eV for sample B, the EOT of sample A is 10.8Å, and 13.6Å for sample B. The difference of metal work functions indicates slightly different fixed charge densities at the dielectric-silicon interfaces for A and B. The flat-band voltages are ~ -0.55V and and ~ -0.60V for samples A and B, respectively.

If abrupt interfaces are assumed, the EOT of the dielectric stack can

be calculated by

$$\text{EOT} = t_{\text{ZrO}_2} \times \frac{\epsilon_{\text{SiO}_2}}{\epsilon_{\text{ZrO}_2}} + t_{\text{int}} \times \frac{\epsilon_{\text{SiO}_2}}{\epsilon_{\text{int}}}, \tag{2.11}$$

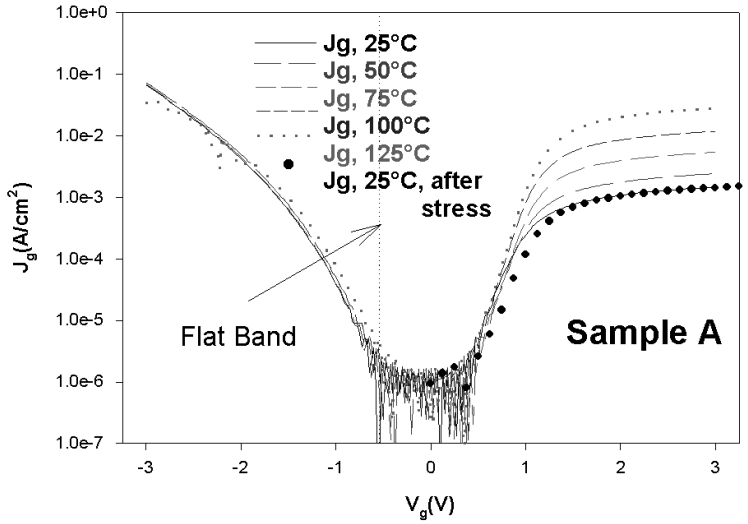
where ϵ_{SiO_2}, ϵ_{ZrO_2}, and ϵ_{int} are dielectric constants of SiO_2, ZrO_2, and Zr-silicate, respectively; $\epsilon_{\text{SiO}_2} = 3.9$.

From ranges of dielectric-layer thicknesses, EOTs from C_g-V_g simulation, $\epsilon_{\text{ZrO}_2} = 20$, and to satisfy Equation (2.11) for both samples A and B, $\epsilon_{\text{int}} \sim 5.5$ is found. The physical thickness of each layer is determined as: $t_{\text{ZrO}_2} = 33.6\text{Å}$ and $t_{\text{int}} = 6.0\text{Å}$ for sample A, and $t_{\text{ZrO}_2} = 38.5\text{Å}$ and $t_{\text{int}} = 8.6\text{Å}$ for sample B. These values are within ranges obtained from TEM but are not the medium values. To satisfy Equation (2.11) with the medium values will result in negative ε_{int} and small $\varepsilon_{\text{ZrO2}}$ (~ 4.18). With limited experimental information available, the above approximation is the most reasonable. These parameters, along with the ones obtained from C_g-V_g simulation, are used later for gate current simulation.

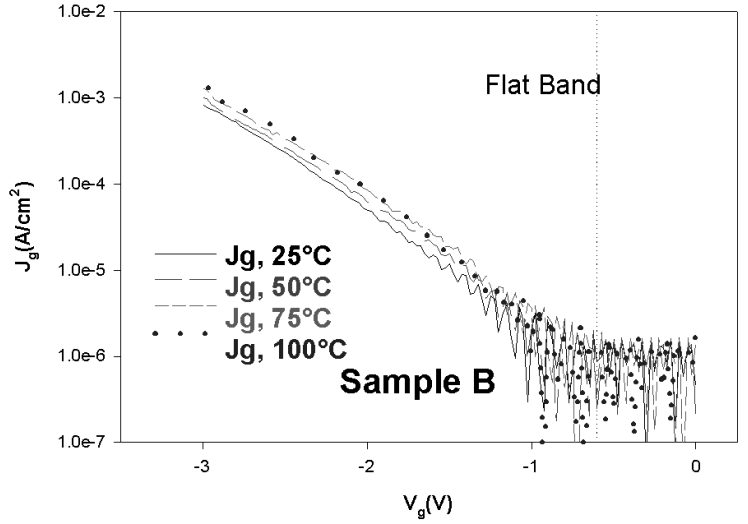
2.3.2 I_g vs. V_g/Temperature

Temperature-dependence of gate current is shown in Figure 2.7. Sample A, with $t_{\text{ZrO}_2} + t_{\text{int}} = 39.6\text{Å}$, shows temperature-independence in accumulation but temperature-dependence in inversion. Sample B, with $t_{\text{ZrO}_2} + t_{\text{int}} = 47.1\text{Å}$, shows weak temperature-dependence in accumulation.

In principle, the temperature-dependence could be contributed from either the transport mechanism, or by the charge distribution in the system, or both. In an NMOSCAP, electrons in the conduction band are minority



(a)



(b)

Figure 2.7: Experimental gate current vs. voltage at different temperatures for (a) Sample A and (b) Sample B.

carriers in the silicon channel. Without a source/drain as external supply, its concentration cannot be maintained at the thermodynamic-equilibrium value. As the temperature is increased, substantially more minority carriers are available, and, as a result, the gate currents show strong temperature-dependence in inversion. This strong temperature dependence contrasts markedly to the much weak temperature dependence for the field-assisted thermionic current for accumulation shown in Figure 2.5, thus is attributed to the temperature-dependence of the charge distribution. The strong temperature-dependence in inversion also indicates that the gate current is primarily from the conduction band.

Yamaguchi et al reported that band gaps of ZrO_2 and Zr-silicate layers are 5.7eV and 4.5eV, respectively, from XPS analysis for sputter-deposited films [52]. Using these values in Franz dispersion, the gate current simulation agrees well with experiment for both samples, (Figure 2.8). The conduction band offsets of ZrO_2 and Zr-silicate are found as 1.45eV and 1.0eV, respectively. The band-edge effective masses are $m_c = m_v = 0.35m_e$ for both ZrO_2 and Zr-silicate.

Yamaguchi assumed Frenkel-Poole conduction through the dielectric stack and reported the conduction band offsets with silicon of the ZrO_2 and Zr-silicate layers as 1.5eV and 1.0eV, respectively. For Frenkel-Poole conduction [15], the conducting electrons are thermally emitted out of the traps in the dielectric. The trapping potential barrier will be reduced if an electric field is applied. As a result, the current will be increased by a factor proportional

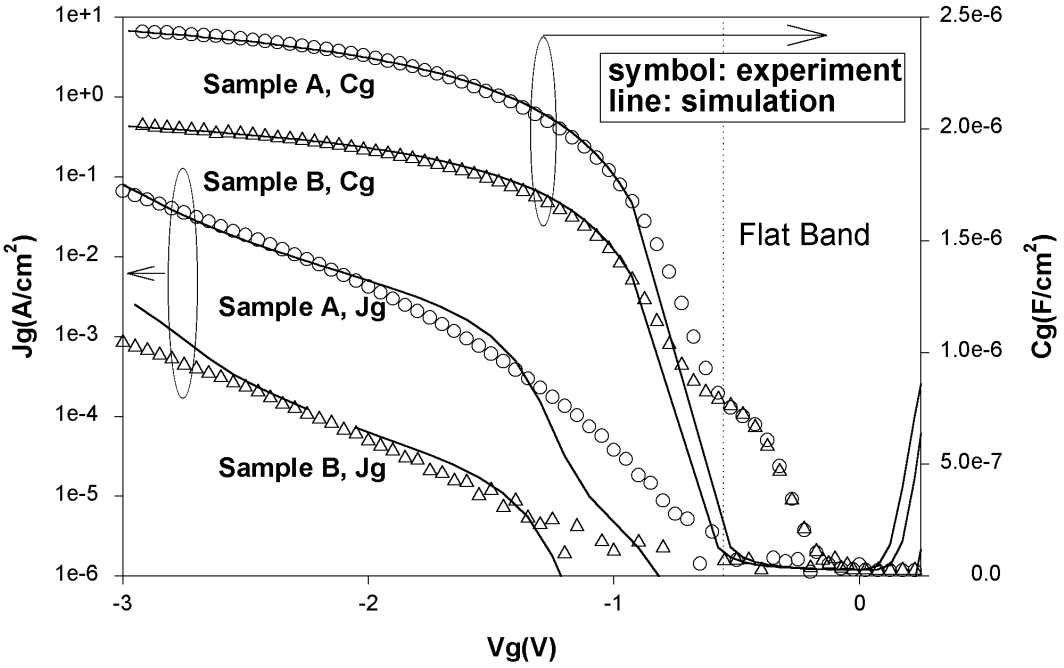


Figure 2.8: Comparison of gate current and gate capacitance simulations with the experimental data for samples A and B.

to $\exp(\frac{\Delta E_i}{k_B T})$, where $k_B T$ is the thermal energy, and ΔE_i is the decrease of the trapping potential barrier due to the electric field. The barrier is further decreased if the electric field is increased. Thus, stronger field/voltage dependence of current is to be expected at low temperatures. Such characteristic was not observed in the gate currents of ZrO_2-NMOSCAPs in our experiments. We believe, therefore, the temperature dependence observed in sample B is not primarily attributed by the the Frenkel-Poole mechanism.

Although simulation and experiment agree quite well, it is not conclusively clear that direct tunneling or FN tunneling, which are the temperature-independent conductions being modeled in this work, is the primary transport mechanism. This is due to the uncertainties about the device structural information, especially those of the dielectric stack. Other temperature-insensitive or weakly dependent transport mechanisms such as trap-assisted tunneling remain possibilities.

Comparison between simulation and experiment of gate currents at different temperatures is shown in Figure 2.9. In simulation, the gate current of sample B has stronger temperature dependence than sample A, consistent with the model for thicker barriers, (Figure 2.5). However, the temperature dependence predicted by the model is weaker than experiment. Thus, some thermally assisting process(es) should be involved, but its influence is weak. Such thermal processes can either be related to the transport through the dielectric or the supply of carriers. If it is related to the transport, it should be weakly dependent on the electric field in the dielectric.

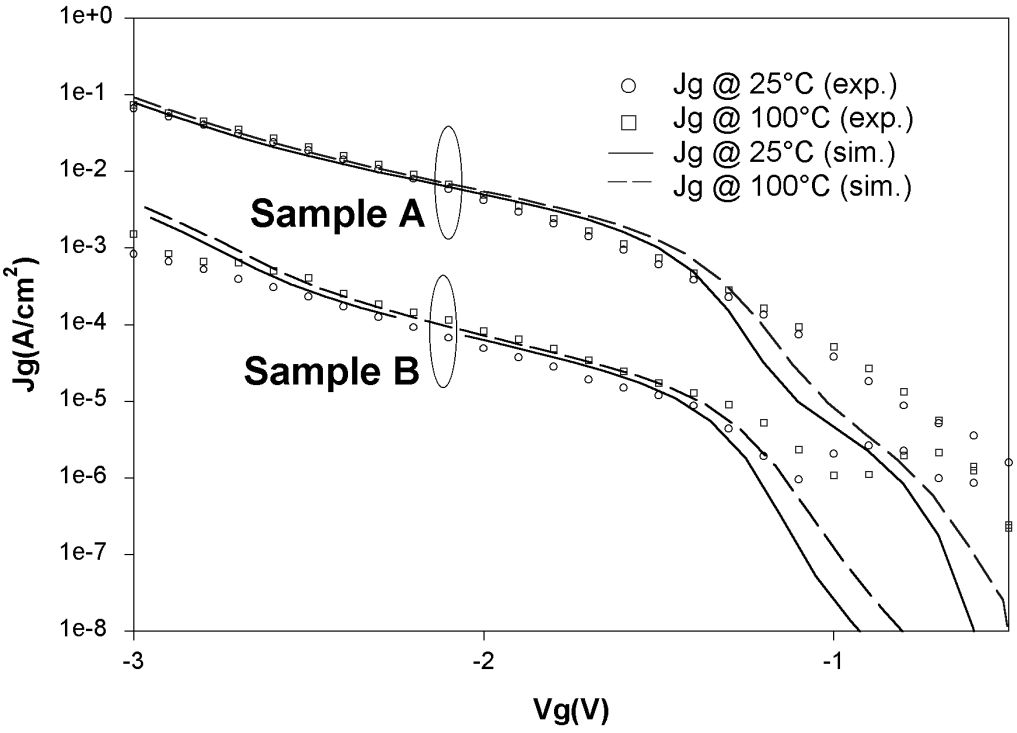


Figure 2.9: Comparison of simulation and experiment for gate currents at different temperatures for samples A and B.

At low gate voltages, oscillations of gate currents were observed near the flat-band and depletion region. Kinks in C_g-V_g were also observed. These might be caused by the interface states between regions (Gate/Dielectric or Dielectric/Silicon). The high-density of interface states might significantly affect the gate capacitance and gate current by trapping or de-trapping the electrons in the absence of inversion or accumulation layers.

The I_g-V_g characteristics from sample A were measured at 25°C before and after thermal and electrical stresses. Gate voltages from -3V to 3V were applied at temperatures of 25°C to 125°C. The measured gate current is similar to that of the fresh device, (Figure 2.7(a)). This implies good quality of the thin films and few charges are trapped or fresh traps are created within the dielectric stack. This is yet another piece of circumstantial evidence of the unimportance of Frenkel-Poole transport in these samples, at least in accumulation.

The comparison of gate current simulation with experiment for a TaN / HfO_2 / p-Si capacitor is shown in Figure 2.10. The simulation agrees well with experiment in accumulation. The EOT is ~10.5Å obtained from C_g-V_g simulation. $E_g = 5.7$eV, $\Delta E_c = 1.6$eV, and $m_c = m_v = 0.32m_e$ are used in the Franz dispersion for the HfO_2 layer, and $E_g = 4.5$eV, $\Delta E_c = 1.2$eV, and $m_c = m_v = 0.5m_e$ for the interfacial layer.

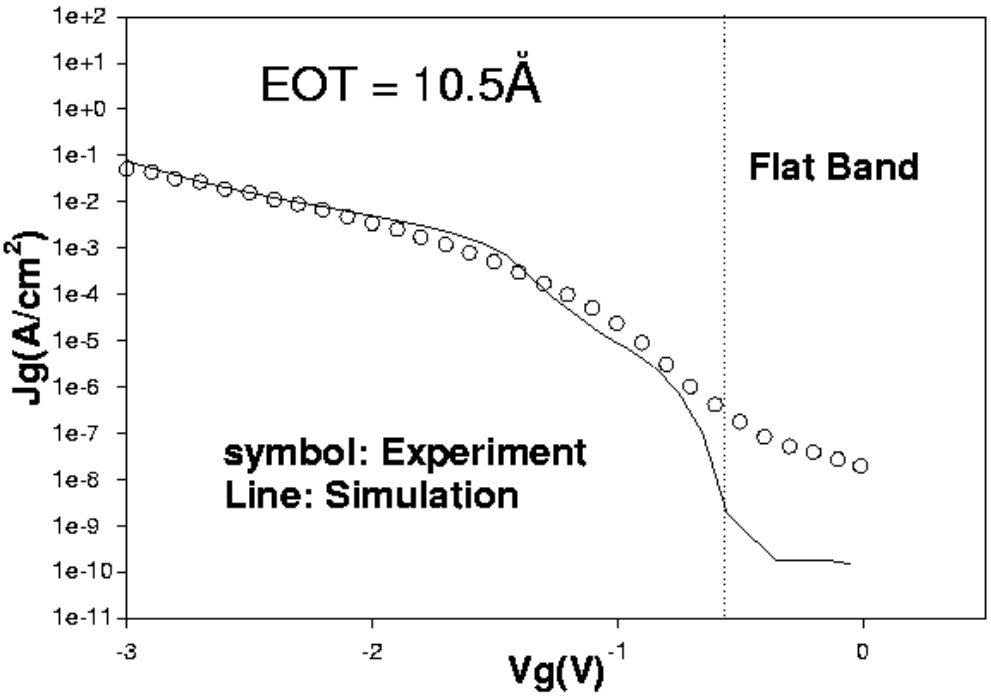


Figure 2.10: Comparison of gate current simulation with experimental data for a TaN/HfO_2/p-Si capacitor.

2.4 Summary and Conclusions

A gate current model has been developed taking into account the quantum confinement effects in the silicon channel, direct and Fowler-Nordheim tunneling, and thermionic emission transport through the gate dielectric. Both gate- and substrate-injected currents are modeled for the silicon conduction- and valence-band components. Subject to the energy dispersion relation in each region of the Gate-Dielectric-Silicon system and available carriers and empty states in either side of the dielectric, the gate current is determined. The energy dispersion in the dielectric band gap is approximated by Franz dispersion. The subband structures and carrier distribution in energy and position in the silicon channel are obtained by solving the Schrodinger and Poisson equations self-consistently, both for gate capacitance and gate current calculations. A self-consistent C_g,I_g-V_g model thus is established. This model was validated using SiO_2 dielectric devices, and good agreement with experiment is achieved.

Device structure information of ZrO_2 NMOSCAPs is extracted in accumulation using C_g-V_g and I_g-V_g simulation. The simulation agrees well with the experiment for different thicknesses of dielectric stack, using consistent parameters. The extracted band gaps and band offsets with silicon are comparable with those that have been reported. The gate current simulation of a HfO_2 NMOSCAP also agrees well with the experiment.

Characteristics of gate current transport mechanism were studied for ZrO_2 NMOSCAPs. The temperature-dependence study shows that the gate

current is primarily contributed from the silicon conduction band, and tunneling is the most likely primary transport mechanism. However, other tunneling processes such as trap-assisted tunneling may be possible. More temperature-dependent processes are also not completely excluded, but their effects are weak. Interface states between different regions might significantly affect the gate capacitance and gate current at low voltages. Kinks in gate capacitance-voltage and oscillations of gate currents were observed. The gate current does not change much after the device is stressed electrically and thermally. This indicates good quality of the film and few charges or traps created in the dielectric stack as a result of the stress.

Chapter 3

Impact of Interfacial Layer and Transition Region on Gate Current: Its Tradeoff with Gate Capacitance

Stacked gate dielectrics are modeled with respect to the impact on leakage current of interfacial layers and transition regions. Using Franz model, the gate dielectric stack is characterized for considering gate capacitance and current performance. Low-EOT and low-gate-current regimes are explored theoretically using reasonable estimates guided by experimental data. Transition layer values of each parameter are qualitatively explored for oxynitride, Si_3N_4/SiO_2, and high-K stacks.

3.1 Introduction

High-K materials such as HfO_2 and ZrO_2 ($\varepsilon \sim 20$) are being considered as alternative gate dielectrics to SiO_2 [28, 41] to obtain the same EOT with a greater physical thickness. High gate capacitance or low EOT thus can be achieved simultaneously with low gate currents. High gate capacitance allows good gate control over the channel and increases the drive current; low gate currents reduce standby power dissipation for low-power application [1, 21].

However, an interfacial layer generally found between the high-K layer and silicon substrate can have an opposing effect on this pursuit of low EOTs and low gate currents. This layer can be an unavoidable silicate layer created during processing [52] or an intentionally deposited/grown layer such as SiO_2 to improve interface properties with silicon [7]. Its dielectric constant is generally lower than that of the high-K layer; thus the EOT of the dielectric stack will be increased. On the other hand, since the band gap is generally inversely proportional to the dielectric constant [9], this layer can lower the gate current if its band offset with silicon is also proportionally larger.

Although an abrupt interface between the high-K layer and interfacial layer is usually assumed, a gradual transition may occur, (Figure 3.1). Yet its role in the gate dielectric stack has not been extensively explored. In this work, effects of the transition regions between the high-K layer and silicon substrate on the tradeoff between EOTs and gate currents are investigated theoretically, based on reasonable assumptions guided by experimental data.

A major hurdle to this study arises from uncertainties of material properties, especially those of the thin films, and the physical mechanisms that affect the gate capacitance and gate current. A self-consistent gate-capacitance and gate-current model which simultaneously considers quantum confinement effects in the silicon channel, direct tunneling, Fowler-Nordheim tunneling and thermionic emission gate currents has been reported and shows good simulation results compared with experiment [12]. Although uncertainties about the physical mechanisms involved still exist, a semi-quantitative investigation can

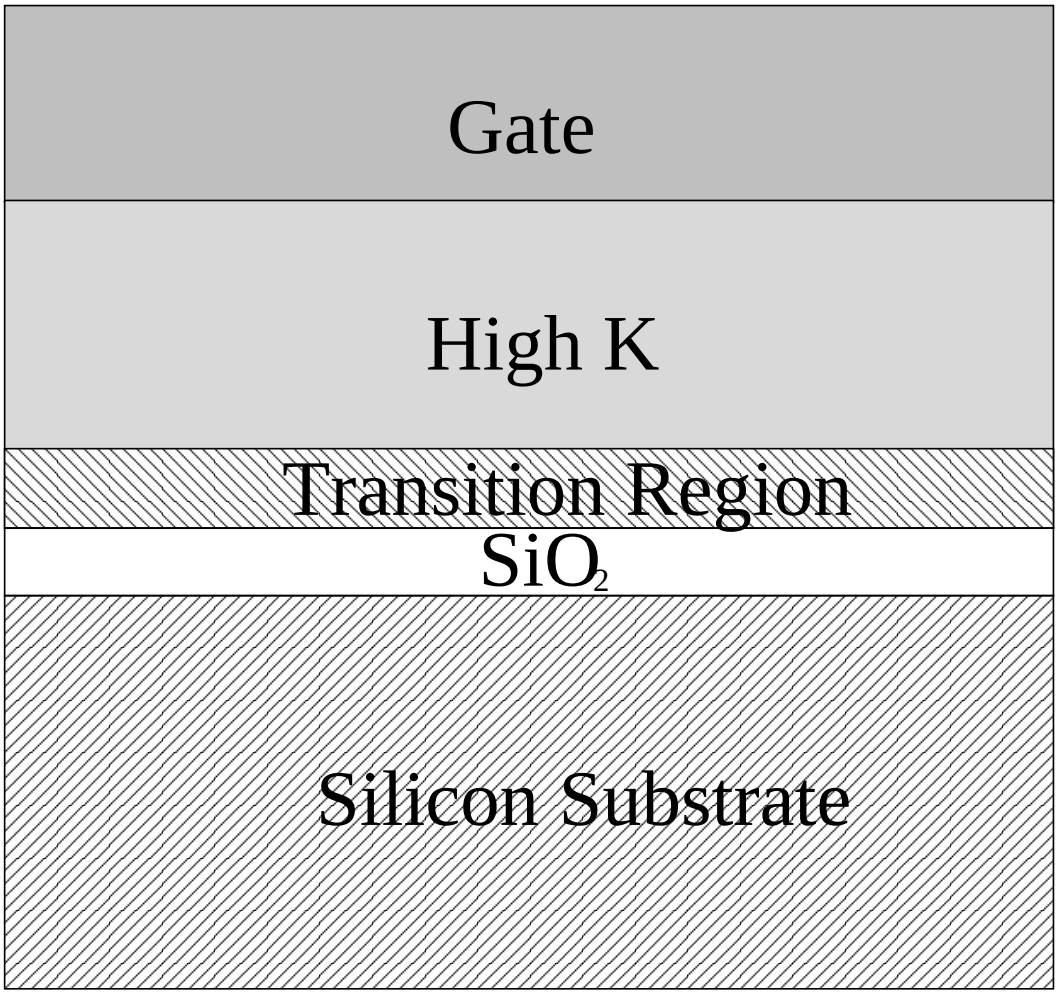


Figure 3.1: Stacked gate dielectric structure that consists of Gate/High-K/Transition Region/SiO_2/Si.

be derived from such work with additional physically-based assumptions.

It has been reported that there is a linear relation between the nitrogen/oxygen concentration ratio in the oxynitride, its dielectric constant, and the potential barrier for electrons in the silicon conduction band [18]. If such linear relation can also be assumed for other physical parameters, effects of the nitrogen concentration profile in the Si_3N_4/SiO_2 stack on the EOT and gate current can be theoretically studied. Furthermore, unless specified otherwise, we assume (1) gate dielectric stack consisting of High-K(or Si_3N_4)/Transition-Layer/SiO_2, as shown in Figure 3.1, (2) linear variation of physical parameters in the transition layer, and (3) no free charges trapped in the dielectric stack in the quasi-static state for the purpose of simplicity and qualitative trend study.

In section 3.2, the gate-capacitance and gate-current models are described. The emphasis is on the characterization of the dielectric stack with the interfacial layer and transition region considered. The tradeoff between the EOT and gate current for oxynitride, Si_3N_4/SiO_2, and high-K gate dielectric stacks is analyzed for NMOSFETs in section 3.3. Finally, conclusions are drawn in section 3.4.

3.2 Theory

The charge distribution in the Gate-Dielectric-Silicon system and E-k dispersion in the silicon channel are determined by solving Schrodinger and Poisson equations self-consistently subject to Fermi-Dirac statistics [23, 49].

The gate capacitance as well as EOT is obtained by assuming quasi-static-thermodynamic equilibrium in the system. Using the same charge distribution and E-k dispersion relation, the gate current can be obtained by a WKB-based method [12].

Linear transition of physical parameters is assumed between the high-K layer and interfacial layer in this work. Figure 3.2 shows assumed variations of those parameters that are considered to affect the gate capacitance and gate current for a Si_3N_4/SiO_2 stack. Conduction and valence band-edge effective masses of the dielectric are set equal to each other in this work for simplicity ($m_{\text{diel}} = m_c = m_v$). $\varepsilon = 3.9$, $m_{\text{diel}} = 0.53m_e$, $E_g = 9.0$eV, and $\Delta E_c = 3.15$eV are used for SiO_2, where m_e is the free electron mass; $\varepsilon = 7.5$, $m_{\text{diel}} = 0.2m_e$, $E_g = 5.0$eV, and $\Delta E_c = 2.2$eV for Si_3N_4. For the high-K study, $\varepsilon_{\text{hk}} = 20$ is assumed and other values are assumed to be the same as for Si_3N_4, unless specified otherwise. $m_{\text{diel}} = 0.2m_e$ is chosen to model less insulating high-K materials as compared with SiO_2 and only serves as a qualitative study. In fact, it will be seen from the oxynitride study that if $m_{\text{diel}} < 0.236m_e$, Si_3N_4 would be unappealing as a gate dielectric. The gate current of a Si_3N_4 device becomes higher than that of a SiO_2 device with the same EOT, as can also be concluded from Figure 3.3. Further discussion of Figure 3.3 is presented in section 3.3.

From Gauss' Law, assuming no free charges within the dielectric stack,

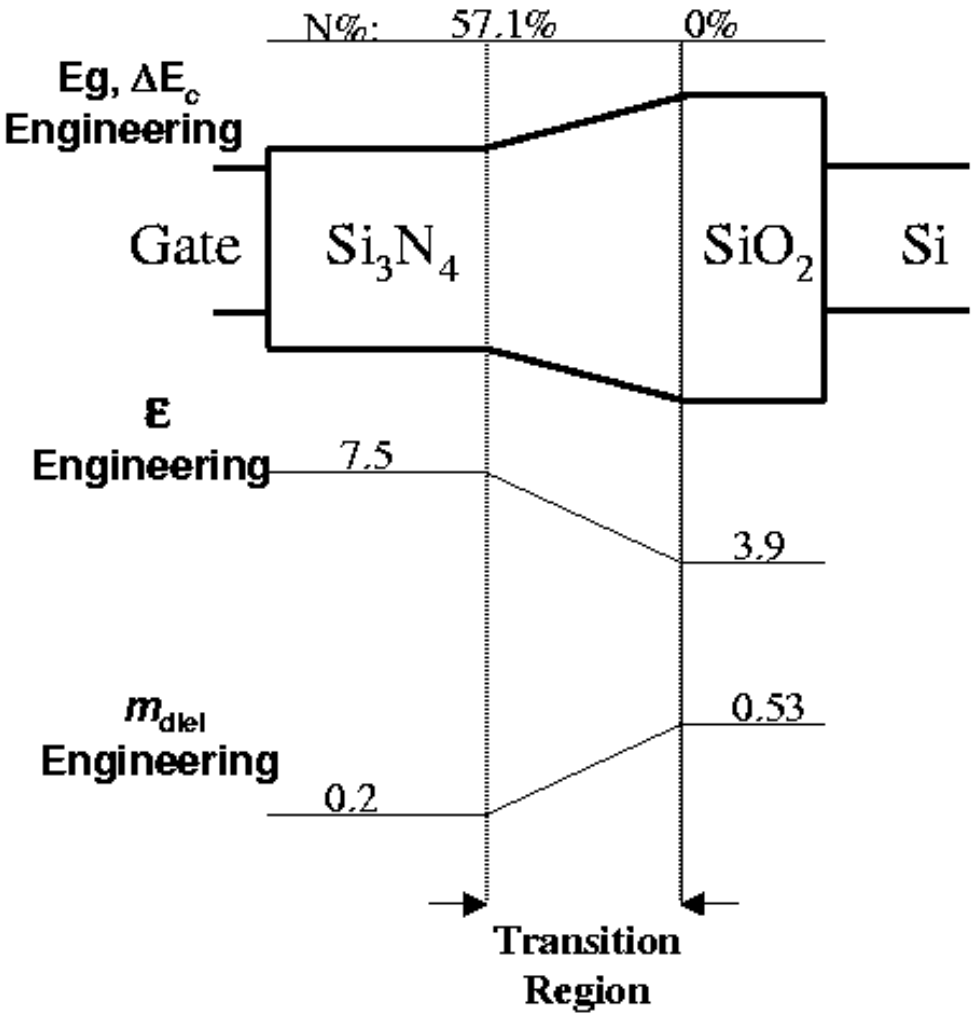


Figure 3.2: Profiles of physical parameters in a Si_3N_4/SiO_2 stack. ε, E_g, ΔE_c, and $m_{\text{diel}}(= m_c = m_v)$ are considered to affect gate capacitance and gate current. Values assumed for Si_3N_4 are only for qualitative study.

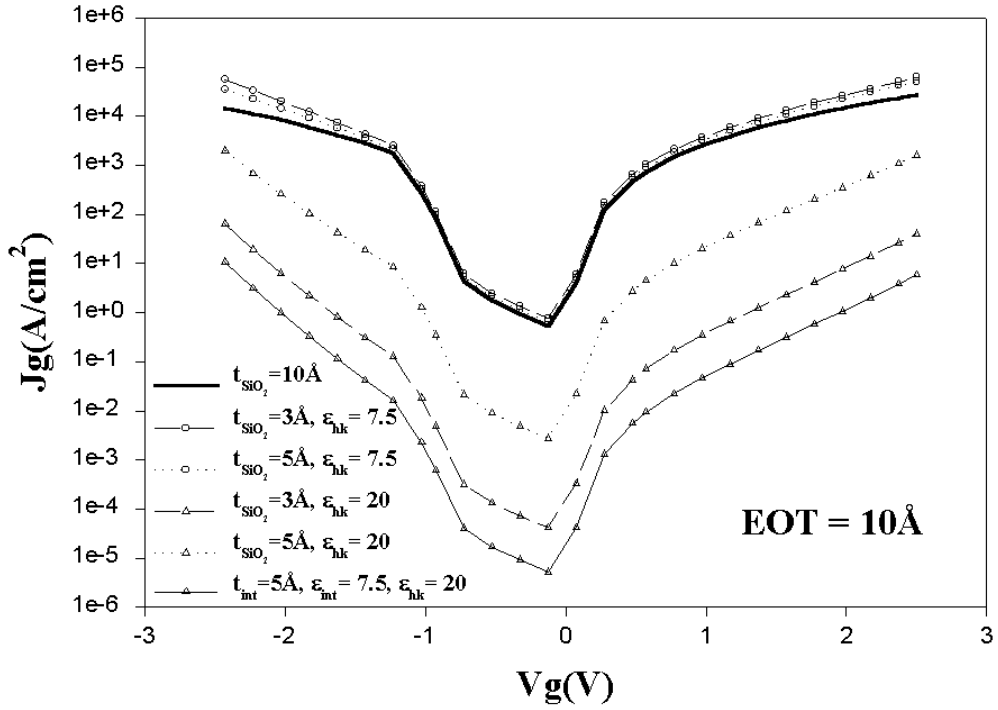


Figure 3.3: Effects of interfacial layer in stacked gate dielectrics without a transition region on gate currents. EOT=10Å is maintained.

$\nabla \cdot \mathbf{D} = 0$, the EOT of the dielectric stack is determined as

$$\mathrm{EOT} = t_{\mathrm{SiO_2}} + t_{\mathrm{hk}} \cdot \frac{\varepsilon_{\mathrm{SiO_2}}}{\varepsilon_{\mathrm{hk}}} + \varepsilon_{\mathrm{SiO_2}} \cdot \int\limits_{\mathrm{trans}} \frac{1}{\varepsilon_{\mathrm{trans}}(z)} \mathrm{d}z, \tag{3.1}$$

where $t_{\mathrm{SiO_2}}$ and t_{hk} are physical thicknesses of SiO_2 and high-K layers, respectively; $\varepsilon_{\mathrm{SiO_2}}$, $\varepsilon_{\mathrm{hk}}$, and $\varepsilon_{\mathrm{trans}}(z)$ are the dielectric constants of SiO_2, high-K, and the transition layer, respectively; the integration is performed between transition region interfaces with SiO_2 and high-K; z is the normal direction to the silicon substrate along which the external electric field is applied. With the assumption of linear variation of the dielectric constant in the transition layer, the EOT of the dielectric stack becomes

$$\mathrm{EOT}_{\mathrm{stack}} = t_{\mathrm{SiO_2}} + t_{\mathrm{hk}} \cdot \frac{\varepsilon_{\mathrm{SiO_2}}}{\varepsilon_{\mathrm{hk}}} + t_{\mathrm{trans}} \cdot \frac{\varepsilon_{\mathrm{SiO_2}}}{\varepsilon_{\mathrm{av}}}, \tag{3.2}$$

where the final term represents contribution of the transition layer to the EOT with

$$\frac{1}{\varepsilon_{\mathrm{av}}} \equiv \frac{1}{\varepsilon_{\mathrm{hk}} - \varepsilon_{\mathrm{SiO_2}}} \cdot \ln \frac{\varepsilon_{\mathrm{hk}}}{\varepsilon_{\mathrm{SiO_2}}}, \tag{3.3}$$

where t_{trans} and $\varepsilon_{\mathrm{av}}$ are the physical thickness and average dielectric constant of the transition region, respectively.

A key to the gate current modeling is the apparent energy dispersion (E-k) in the dielectric band gap seen by the tunneling electrons. The Franz 2-band dispersion model is used for approximating such energy dispersion

$$\frac{1}{k^2} = \frac{\hbar^2}{2m_c(E - E_c)} + \frac{\hbar^2}{2m_v(E_v - E)}, \tag{3.4}$$

where E is the total energy of the electron, m_c and m_v are band-edge effective masses for the conduction and valence bands, respectively, of the dielectric, and E_c and E_v are the energies of the respective band edges [14]. The tunneling probability, $T(E, k_{||})$, through the dielectric is calculated by WKB when $k_z^2 < 0$ [26],

$$-\ln T(E, k_{||}) = \int_{\text{dielectric}} |k_z|\, dz, \tag{3.5}$$

with

$$k_z^2 = k^2 - k_{||}^2, \tag{3.6}$$

where k^2 is calculated from the Franz dispersion with $k_{||}$ being conserved through the dielectric stack and across region interfaces in the Gate-Dielectric-Silicon system. $k_{||}$ is the wave vector component parallel to the region interfaces. The total energy, E, and parallel wave vector, $k_{||}$, are constant in Equation (3.5) when only elastic tunneling is considered. The band edge effective masses, m_c, m_v, are intrinsic material properties. E_c and E_v profiles "seen" by the tunneling electrons are adjusted by both the external electric field and dielectric constant profiles in addition to the intrinsic band structure. As a result, the energy dispersion relation in the dielectric band gap determines the tunneling probability in a non-trivial way, especially when an external electric field is applied.

Within the context of the work presented here, the dielectric stack is characterized by the physical thickness of each layer (t_{hk}, t_{trans}, t_{SiO_2}), the

profiles of its dielectric constant ($\varepsilon(z)$), band gap ($E_g(z)$), band offset ($\Delta E_c(z)$) with silicon, band-edge effective masses ($m_{\text{diel}}(z)(= m_c(z) = m_v(z))$), for the purpose of assessing capacitance/current vs. voltage performance. In the next section, tradeoffs between EOTs and gate currents for oxynitride, Si_3N_4/SiO_2 and high-K stacked gate dielectrics are studied.

3.3 Results and Discussion

Ma *et al. [18]* reported that the dielectric constant and potential barrier of oxynitride vary linearly with the oxygen/nitrogen concentration ratio. In their work, WKB and effective mass approximations (parabolic band structure of dielectric) were applied to study such dependence of tunneling gate currents. A nitrogen concentration independent effective mass was used. However, in this work, the potential effects of variation of the oxynitride effective mass, whatever that may be, are also explored.

Using a scheme similar to Ma's, the dielectric constants, conduction band offsets, and band gaps of the oxynitride are calculated, in the spirit of Vegard's law [50], as

$$\varepsilon_{\text{oxyn}}(\gamma) = 3.9 + 0.063\gamma, \tag{3.7}$$

$$E_{g,\text{oxyn}}(\gamma) = 9.0 - 0.07\gamma, \tag{3.8}$$

$$\Delta E_{c,\text{oxyn}}(\gamma) = 3.15 - 0.017\gamma, \tag{3.9}$$

where $\gamma\%$ is the atomic nitrogen concentration in the oxynitride; $\gamma = 57.14$ and $\gamma = 0$ correspond to Si_3N_4 and SiO_2, respectively. According to Equation

(3.4), the parameter yet to be determined for the tunneling probability is the band-edge effective mass, $m_{\mathrm{oxyn}}(\gamma)$. A direct approach is to find $m_{\mathrm{oxyn}}(\gamma)$ to fit the simulated gate current with experiments. However, this effort is impeded by experimental difficulties, especially the accurate determination of the composition of an ultra-thin film. Therefore, the actual values of $m_{\mathrm{oxyn}}(\gamma)$ remain unknown and the best we can do is to characterize the effects of possible $m_{\mathrm{oxyn}}(\gamma)$ variations.

For the above purpose, a reference band-edge effective mass $m_{\mathrm{oxyn,cc}}(\gamma)$ is chosen to provide similar gate currents ($\sim 10^{-1}\mathrm{A/cm}^2$ @$\sim$ 1.0V) in simulation for NMOSFETs with the same EOT (20Å) but different nitrogen concentrations, (Figure 3.4). Thus, the variation in $m_{\mathrm{oxyn,cc}}(\gamma)$, in turn, is that required to balance changes in the gate current due to changes in the other parameters—physical thickness, dielectric constant, band gap, and band offset—while holding the EOT constant. Variations in the actual mass from this reference mass, $m_{\mathrm{oxyn}}(\gamma) - m_{\mathrm{oxyn,cc}}(\gamma)$, will be indicative of an exponential reduction, if positive, or increase, if negative, of gate currents due to introduction of nitrogen for a given EOT. The device structure considered in simulation consists of a single-layer oxynitride gate dielectric. Assuming the same models for the gate and silicon substrate, with the same EOT, the tunneling probability is only affected by the thin film thickness and composition. Also shown for comparison in Figure 3.4 are $m_{\mathrm{oxyn,lin}}(\gamma)$ and $m_{\mathrm{oxyn,inv}}(\gamma)$ that represent a linear variation of the mass and the inverse of the mass, respectively, between the shown endpoints of $m_{\mathrm{oxyn,cc}}(\gamma)$. Finally, $m_{\mathrm{oxyn,const}} = 0.53m_e$ is also

included as a nitrogen-concentration-independent band-edge mass. It is seen that the constant-current $m_{\text{oxyn,cc}}(\gamma)$ exhibits neither a linear nor an inverse relation with γ.

In the following study, multi-layer gate dielectric systems are considered. $m_{\text{diel}} = 0.2m_e$ is assumed for both Si_3N_4 and high-K materials to manifest the less insulating alternative gate dielectrics for qualitative study. Such a value is lower than $m_{\text{oxyn,cc}}(57.14) = 0.236$, which corresponds to Si_3N_4. As can be seen, it will result in higher gate current for a Si_3N_4(K=7.5)/SiO_2 device but lower gate current for a high-K(K=20)/SiO_2 device (Figure 3.3). Gate dielectric stacks consisting of Si_3N_4/SiO_2 are first studied.

The transition region between the Si_3N_4 layer and SiO_2 layer consists of a oxynitride layer of variable nitrogen concentration. A continuous transition is assumed, *i.e.* $\gamma = 0$ at the oxynitride/SiO_2 interface and $\gamma = 57.14$ at the oxynitride/Si_3N_4 interface. An accurate description of the effects of such a region requires knowledge of $m_{\text{oxyn}}(\gamma)$. Linear variation of γ in the transition layer will lead to linear variation of ε and ΔE_c [18], but not necessarily E_g and m_{diel}. For simplicity and qualitatively theoretical study, linear variation of each parameter from the SiO_2-layer to the Si_3N_4-layer is assumed. Gate currents of NMOSFETs with different transition layer thicknesses but the same EOT (10Å) and t_{SiO_2}(3Å) are studied in simulation. Effects for each parameter in terms of gate current as a function of the transition layer thickness are analyzed individually, using the profile of each parameter shown in Figure 3.2 with the other parameters held constant through the dielectric stack:

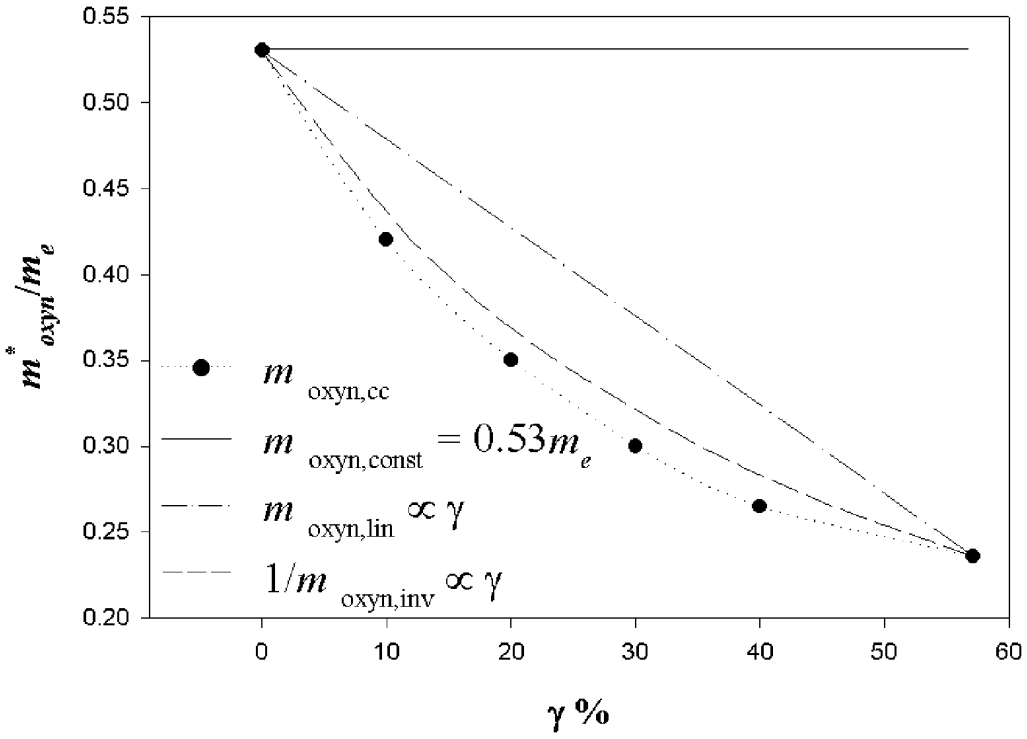


Figure 3.4: $m_{\text{oxyn,cc}}$ vs. γ. $m_{\text{oxyn,cc}}$ are chosen to provide similar gate currents (10^{-1}A/cm^2) at $V_g \sim 1$V in simulation of different nitrogen atomic concentrations ($\gamma\%$) in oxynitride. $m_{\text{oxyn,lin}}$ and $m_{\text{oxyn,inv}}$ represent linear variation of the mass and inverse of the mass, respectively, in relation with γ between the endpoints of $m_{\text{oxyn,cc}}$. $m_{\text{oxyn,const}}$ represents nitrogen-independent band-edge mass. EOT=20Å is maintained. Single-layer oxynitride gate dielectric is assumed.

- ε-transition: $\varepsilon(z)$ as in Figure 3.2 but constant E_g, ΔE_c, and m_{diel}
- E_g,ΔE_c-transition: $E_g(z)$ and $\Delta E_c(z)$ as in Figure 3.2 but constant ε and m_{diel}
- m_{diel}-transition: $m_{\text{diel}}(z)$ as in Figure 3.2 but constant ε, E_g and ΔE_c

The constant values used in each case were those of SiO_2, but they can be either those of SiO_2 or Si_3N_4 and the final results are not very different. Those values are held the same through the SiO_2, transition, and Si_3N_4 layers. The parameters being studied in each case vary linearly in the transition layer with values of SiO_2 and Si_3N_4 at the interfaces with SiO_2 and Si_3N_4 layers, respectively. Instead of using γ, t_{trans} is used to benchmark the gradient of the nitrogen concentration profile in the Si_3N_4/SiO_2 stack. The reference for comparison is the gate current of a Si_3N_4/SiO_2 stack without a transition layer. All simulations were reference at $V_g = 1.2$V. The results are shown in Figure 3.5. Also shown in the figure is the collective effect of the transitions of all of the parameters on the gate current. It is found that ε-transition increases the gate current, but E_g,ΔE_c- and m_{diel}-transitions decrease the gate current. If transition of each parameter is considered, given the values of parameters assumed, the gate current will decrease when the thickness of the transition region is increased; the reduction is $\sim 0.42\times$ at $t_{\text{trans}} = 8.0$Å. Such reduction will be improved if the band-edge mass of Si_3N_4 is larger (it is assumed $0.2m_e$ in the above study). The explanation is presented later in this section.

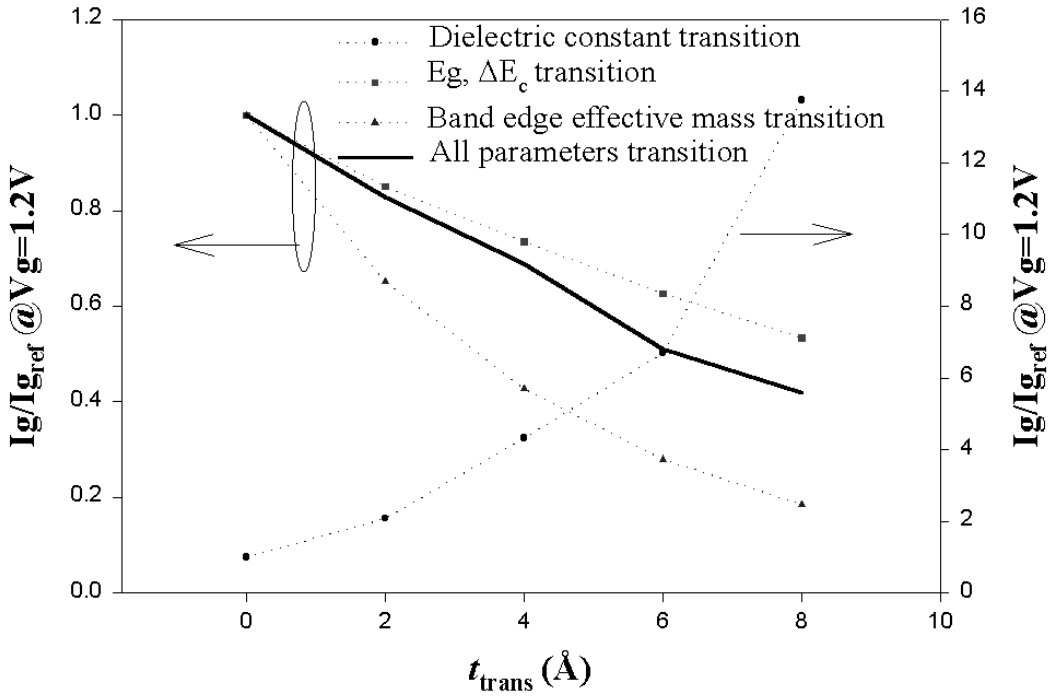


Figure 3.5: Individual and collective effects of linear variation of each parameter in the transition layer on gate currents for Si_3N_4/SiO_2 stacks, benchmarked by the transition layer thickness. The reference is gate current at $V_g = 1.2V$ of a Si_3N_4/SiO_2 stack without a transition layer. EOT=10Å is maintained.

If the Si_3N_4 layer is replaced by a higher-K material and the same analysis is applied, it is found that the effects of E_g,ΔE_c- and m_{diel}-transition, both qualitatively and semi-quantitatively, are quite similar to those of the Si_3N_4/SiO_2 stack. However, effects of ε-transition become more pronounced when the dielectric constant of high-K is increased, especially when greater E_g, ΔE_c and m_{diel} are assumed, *i.e.* a more insulating high-K layer. If parameters of SiO_2 are adopted for the fixed parameters in the ε-transition study, the gate current is greatly increased when the transition layer thickness is greater than 2Å, (Figure 3.6). However, when lower values like those assumed for Si_3N_4 are considered, the increase of gate current for the high-K(ε_{hk}=20)/SiO_2 stack becomes $\sim 2.7\times$, $7.5\times$, $20.4\times$, and $55.1\times$ for transition layer thicknesses of 2Å, 4Å, 6Å, and 8Å, respectively, in the ε-transition study. Such increasing effects on gate currents can be counter-balanced by the band structure (E_g, ΔE_c, m_{diel}) transition. The gate current is increased by $6.5\times$ at $t_{\text{trans}} = 8.0$Å if the linear variation of each parameter in the transition layer is considered, compared with the $0.42\times$ reduction for the Si_3N_4/SiO_2 stack. Figure 3.7 shows the results of the same analysis applied to m_{diel} equal to $0.3m_0$, $0.4m_0$, $0,53m_0$, respectively, for different t_{trans}. The transition layer becomes more important when m_{diel} increases. The above phenomena can be explained as follows. When a transition layer exists, given a targeted EOT and the same interfacial SiO_2 thickness, the increase of the transition layer results in reduction of the physical thickness of the dielectric stack: at $t_{\text{trans}} = 8.0$Å with EOT=10Å, the reduction is 17% and 21% for $\varepsilon_{\text{hk}} = 7.5$

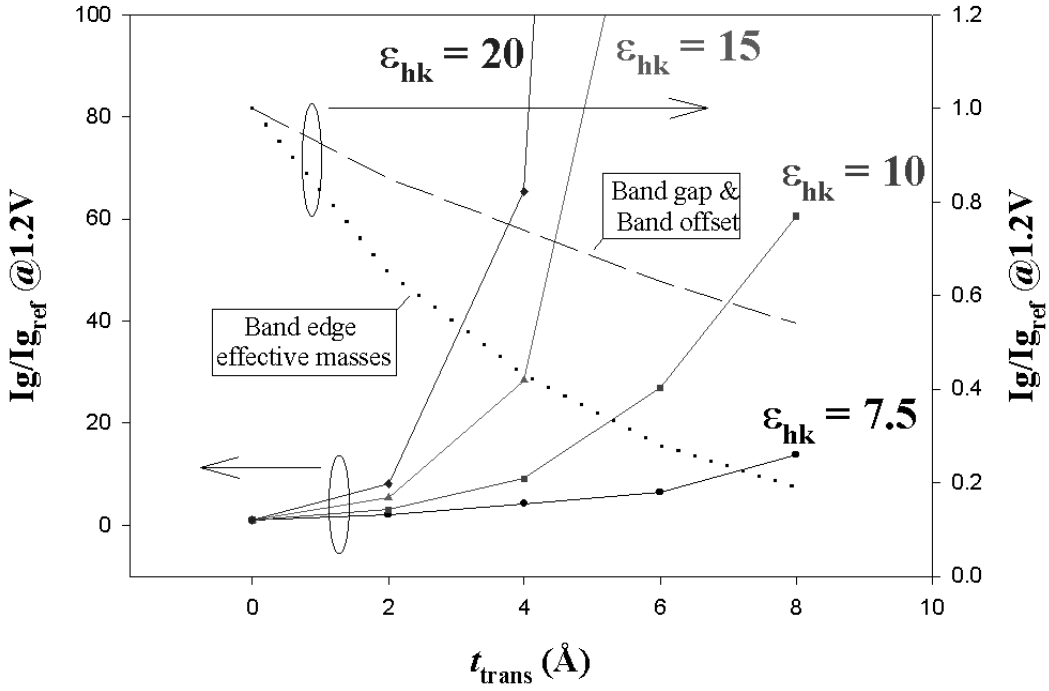


Figure 3.6: Individual and collective effects of linear variation of each parameter in the transition layer on gate currents for High-K/SiO_2 stacks, benchmarked by the transition layer thickness. The reference is gate current at $V_g = 1.2V$ of a Si_3N_4/SiO_2 stack without a transition layer. EOT=10Å is maintained.

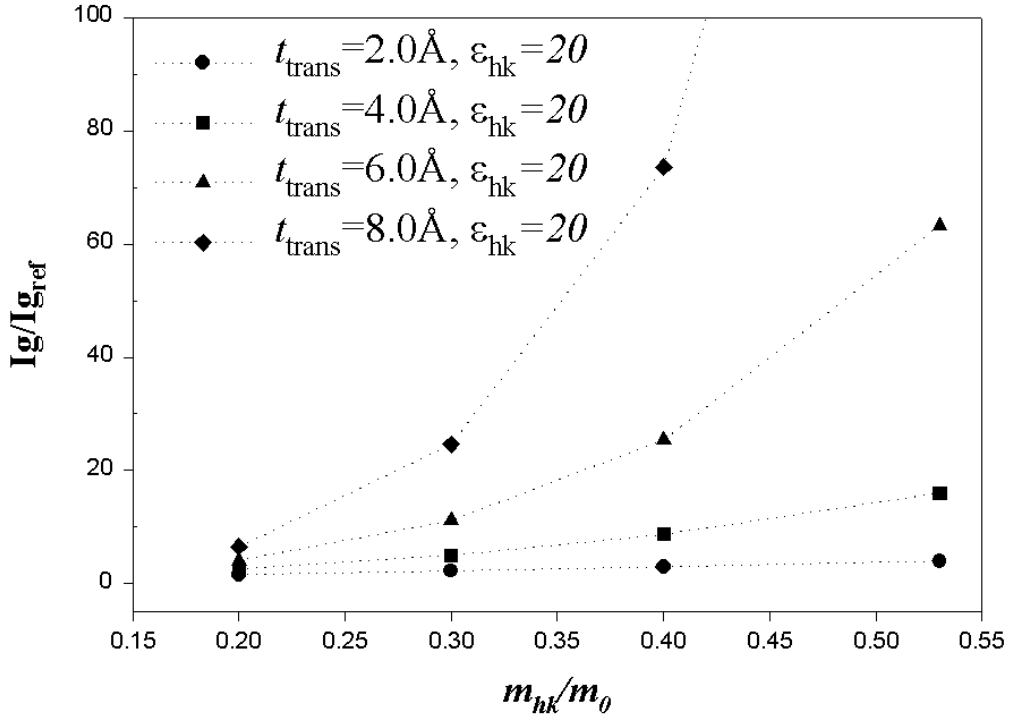


Figure 3.7: Band-edge effective masses of high-K layer on effects of the transition region. The more insulating (higher m_{diel}) high-K layer results in more importance of the transition layer. EOT=10Å is maintained.

and $\varepsilon_{\text{hk}} = 20$, respectively. Since most thickness reduction occurs in the high-K layer, when the high-K layer is more insulating, the gate current will be greatly increased in comparison to the gate current of a stack without a transition layer. Though m_{diel} is chosen as the variable determining resistivity of the high-K layer, similar analysis for E_g,ΔE_c variations can be performed to obtain similar conclusions.

Finally, in Figure 3.3, high-K and Si_3N_4/SiO_2 stacks of EOT=10Å with-

out a transition layer are considered. Gate currents of $Si_3N_4/SiO_2(3Å)$ are found to be higher than for pure SiO_2 devices because of the assumption of small band-edge masses ($0.2m_e$). The gate current is further increased when the Si_3N_4 thickness is increased. When Si_3N_4 is replaced by high-K of $\varepsilon_{hk} = 20$, the trend reverses. If the dielectric constant of the interfacial layer is increased from 3.9 to 7.5, the gate current will be significantly decreased.

3.4 Summary and Conclusions

Stacked gate dielectrics are characterized using dielectric constants, band gaps, band offsets with silicon, and band edge effective masses with a Franz 2-band energy dispersion relation, along with the physical thickness of each layer to assess capacitance/current vs. voltage behavior. Tradeoffs between EOTs and gate currents were studied, considering effects of each parameter as well as the interfacial layer and transition region. Oxynitride, Si_3N_4/SiO_2, and high-K stacked dielectrics were qualitatively studied. More insulating and higher dielectric constant materials are desired for each layer of the stack; however, the role of transition region in the dielectric stack will then become more important.

Given the complicated effects of the transition region on the capacitance/current vs. voltage behavior, increasing the dielectric constant of the interfacial layer seems the most feasible approach in the short-term. However, the possibly accompanying negative effects such as mobility degradation may be of concern.

For a targeted EOT, ε-transition increases gate currents, but E_g,ΔE_c and m_{diel} transition decreases gate currents. The effect of ε-transition becomes dominant when the dielectric constant of the high-K layer is increased. For medium-K stacks, the gate current might be decreased due to the E_g, ΔE_C and m_{diel} transitions dominating the ε-transition. For high-K materials (e.g. $\varepsilon_{\text{hk}} \sim 20$), the gate current is generally increased with a greater transition layer thickness.

Increasing the nitrogen concentration in the oxynitride gate dielectric needs to be carefully assessed. Saturating nitrogen in oxynitride is less appealing and even worse if the band-edge masses are decreased simultaneously. Other negative impacts due to nitrogen introduction such as mobility degradation should also be considered.

Chapter 4

Conduction Mechanisms and Parameter Extraction from C-V and I-V Simulations and Experiments

From C_g, I_g-V_g simulations on experiments, gate leakage current versus voltage behavior can be studied. Furthermore, assuming Franz dispersion to characterize the gate dielectric (stack) for capacitance and current behavior, device structure parameters can be extracted.

4.1 Introduction

High-dielectric-constant (High-K) materials are being pursued as alternative gate dielectrics for next-generation MOSFETs. Great efforts are being made for optimizing these materials for transistor applications [20, 28, 41]. However, a major hurdle arises from the uncertainties about material properties, especially those of the thin films, and the physical mechanisms that affect the gate capacitance and gate current. To overcome such a hurdle, a systematic approach that combines gate capacitance and current analyses is applied to understand their voltage- and temperature-dependent behavior and to extract the device structure parameters. An important element of this work

is the simultaneous matching of the gate capacitance and current simulation with experiment; this matching is achieved via a self-consistent gate capacitance and current model [12]. Detailed discussion of the model was presented in Chapter 2 and 3.

This analysis method will be demonstrated to provide consistent results for different devices fabricated by different processes. The analysis was performed in a systematic way for each study case without violating assumptions made by the underlying model.

4.2 Model and Observations

Neglecting the wave function penetration from the silicon into the gate dielectric, the charge distribution in position and the energy dispersion in the channel can be determined by self-consistently calculating Poisson and Schrodinger equations and by applying the boundary conditions of the substrate. The silicon substrate is pre-defined by its material properties such as the doping concentration. Thus, EOT is the only intrinsic parameter needed from the gate dielectric (stack) for such calculation. Other parameters to model the intrinsic properties of the gate dielectric are the band gap (E_g), conduction band offset (ΔE_c), band-edge effective masses (m_c and m_v), which are required in Franz 2-band dispersion [14], the dielectric constant (ϵ), and physical thickness of each thin-film dielectric layer. Using the potential profile in the dielectric (stack) obtained after the charge distribution is calculated, the tunneling probability can then be obtained.

The quasi-Fermi levels of the gate and the silicon substrate ($E_{f,gate}$ and $E_{f,Si}$) at each gate voltage are self-consistently calculated. The quasi-Fermi levels thus determined are also regarded as chemical potentials of the gate and silicon substrate, which are separated by the gate dielectric but connected by the external circuit. Because the exchange of carriers through the gate current is negligible as compared with that through the external circuit, the quasi-Fermi levels of the gate and substrate are only determined by the external bias:

$$E_{f,Si} - E_{f,gate} = V_{gs}, \tag{4.1}$$

where V_{gs} (or V_g if the substrate is grounded) is the applied gate voltage. For gate current calculation, $(E_{f,Si} - E_{f,gate})$ not only determines the direction of the electron current, but also is another factor to determine the final calculated gate current, (Equation 2.5 and 2.6). As a result, at the zero gate voltage, the gate current should vanish unless other mechanisms cause a chemical potential force between the gate and silicon substrate.

Gate capacitance-current vs. voltage from which the unknown parameters of the system to be extracted, its dependence on the operating temperature in different bias regions and the physical thickness of each gate dielectric layer impose the constraints (boundary conditions) on the parameter space. Varying the voltage and temperature allows the charges to redistribute in position and energy. As a result, the primary transport mechanisms and supply of carriers for the gate current may not be the same at different voltages or temperatures. Thus, in order to study the gate current, source of the carriers should be first

identified. The transport mechanism then can be determined according to its voltage-, temperature-, and dielectric-thickness-dependence behavior.

Parameters that can be extracted by this approach are those that can affect the gate capacitance and current. The space of each parameter to be determined will be refined if other parameters are known with good accuracy. In this work, the parameters of the primary interest are those needed for Franz dispersion.

4.3 Results and Discussion

Detailed discussion of ZrO_2 and HfO_2 devices were presented in Chapter 2. Only important results will be rephrased in this chapter. SiO_2 devices are the primary focus to better illustrate the proposed gate capacitance-current analysis method. Three sets of SiO_2 devices fabricated by different research groups will be discussed. Through out this work, well-known values of the dielectric constant, band gap, conduction band offset with silicon of SiO_2 are assumed in simulation: $\epsilon = 3.9$, $E_g = 9.0$eV, and $\Delta E_c = 3.15$eV. Thus, the band-edge effective masses are the primary parameters to be extracted for SiO_2.

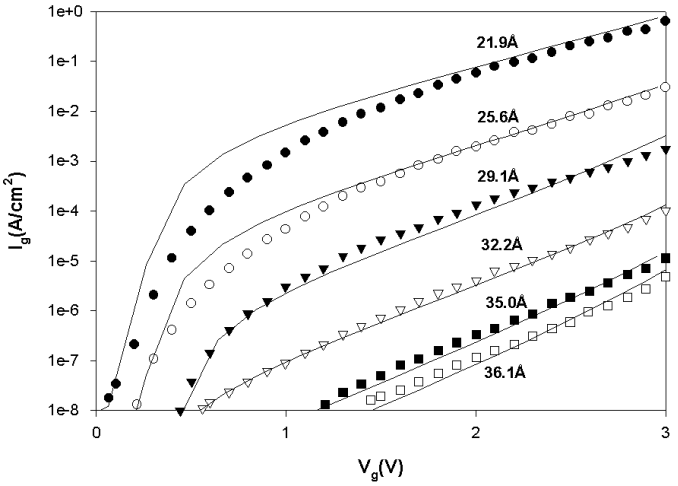
4.3.1 SiO_2 NMOSFETs vs. SiO_2 PMOSCAPs: Substrate-Injected Electron Currents from p-Si Substrate and n-Si Substrate

Figure 4.1 shows the comparison of gate current simulation with experiment for n+-polysilicon-gate/SiO_2/p-Si n-FETs (t_{ox} = 21.9-36.1Å) and

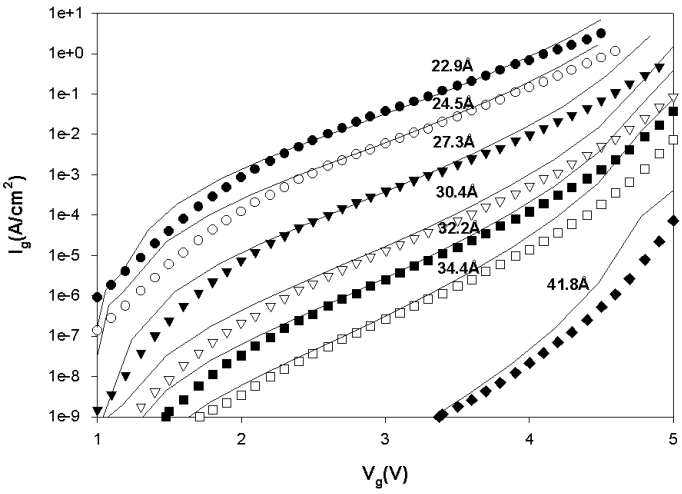
p+-polysilicon-gate/SiO_2/n-Si p-MOS capacitors (t_{ox} = 22.9-41.8Å). The experimental data are from [33]. The SiO_2 thicknesses used in simulation are the same as those provided. The band-edge effective masses of SiO_2 are $m_c = m_v = 0.53m_e$, where m_e is the free electron mass. Figure 4.2 shows the comparison of the gate capacitance simulation with experiment. The SiO_2 thickness of 35.1Å is used in simulation and is the reported value.

Biased at positive gate voltages, the NMOSFETs are in inversion, and the PMOSCAPs are in accumulation. For NMOSFETs, electrons in the silicon conduction band are minority carriers in the channel. The electron concentration is maintained at the equilibrium value by the external supply from the source and drain. Only electron currents from the conduction band of the silicon substrate to the gate are considered for the voltage range being studied. As for PMOSCAPs, there are no source and drain connected to the channel. As a result, the minority holes concentration in the channel is not maintained. Thus, majority electrons in the conduction band dominate the gate current. Since the Fermi-Dirac statistics is applied, only the electrons with energies near or below the quasi-Fermi level are important for gate current in inversion and accumulation.

It should also be noted that the Franz dispersion parameters used to characterize SiO_2 were found to be the same for all devices. Given that only electrons near or below the quasi-Fermi level of the silicon substrate but above the silicon conduction band edge are important, the Franz 2-band model seems a good approximation at least in such energy region. Although it is p-type-



(a)



(b)

Figure 4.1: Gate current simulation compared with experimental data for (a) n+poly/ SiO_2 / p-Si MOSFETs and (b) p+poly / SiO_2 / n-Si MOSCAPs. The experimental data are from [33]. The thicknesses shown in the figures are used in simulation as reported in [33].

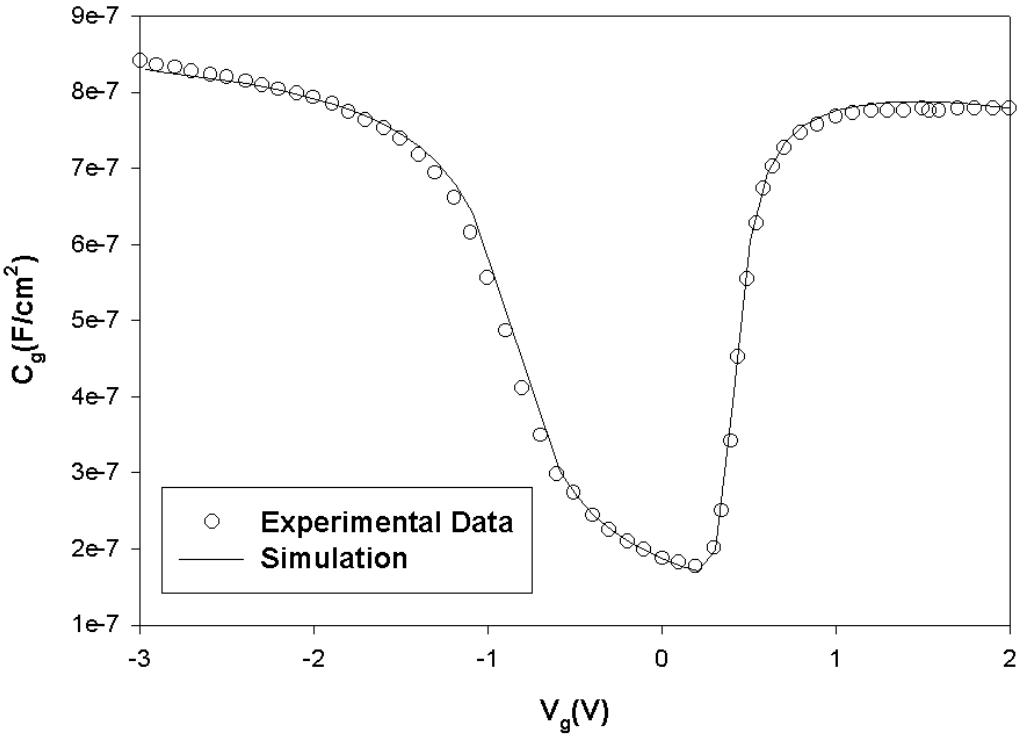


Figure 4.2: Gate capacitance simulation compared with experimental data for a n^+ gate SiO_2 NMOSFETs fabricated by IBM (1.8V and 0.20-μm technology) [33]. The SiO_2 thickness was reported 35.1Å, which is used in simulation in this work.

substrates for NMOSFETs and n-type-substrates for PMOSCAPs, the same energy dispersion relation applies.

4.3.2 SiO_2 NMOSFET vs. SiO_2 PMOSFET: Conduction-Band Component and Valence-Band Component of Gate Currents

Figures 2.3, 2.4 and 4.3 show the comparison of gate capacitance and current simulation with experiment for a SiO_2 NMOSFET and a SiO_2 PMOSFET. The devices were reported in [53]. In gate capacitance simulation, the physical thickness used is 19Å for the NMOSFET and 19.2Å for the PMOSFET. Additional fixed charges at the SiO_2/Si interface are assumed in the gate capacitance simulation in order to fit the flat-band voltage. The band-edge effective masses used in Franz dispersion for the conduction-band gate current component for both devices are $m_c = m_v = 0.72m_e$. As for the valence-band component, $m_c = 0.34m_e$ and $m_v \gg m_c$ are used. When $m_v \gg m_c$, the simulated gate currents saturate to a certain value.

In simulation, as long as m_c/m_v remains the same, the slight change of m_c and m_v, e.g. at least in the range of $0.53m_e \sim 0.72m_e$ for the above n- and p-MOSFETs considered, will only change the magnitude of gate current significantly, (Figure 4.4). But change in voltage dependence behavior, *i.e.* dI_g/dV_g, is not apparent in inversion and accumulation. Slightly changing the SiO_2 thickness used in simulation has similar effects as changing m_c and m_v with m_c/m_v constant, (Figure 4.4) . As a result, if other parameters are assumed the same, the accuracy of the physical thickness will affect that of the

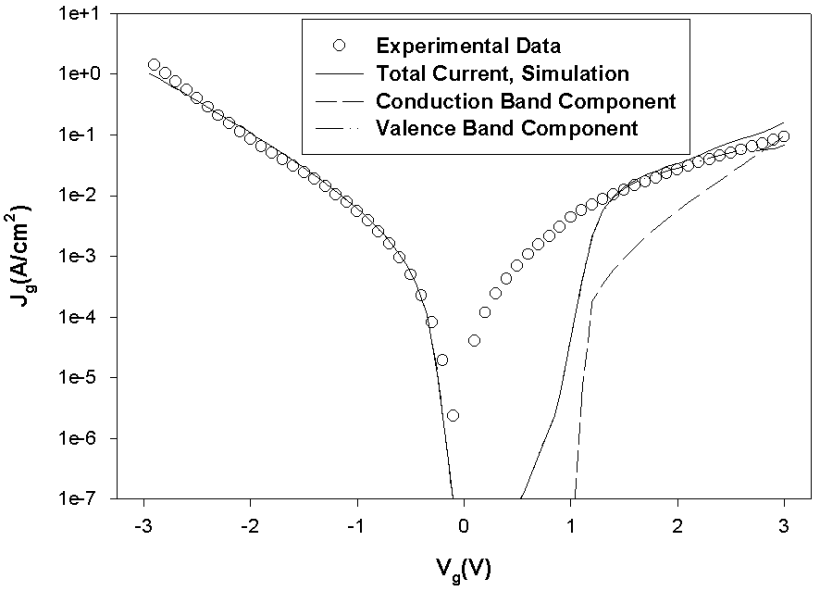


Figure 4.3: Gate current simulation compared with experiment in both the inversion and accumulation for a SiO_2 poly-PMOSFET. $t_{ox} \sim 20$Å.

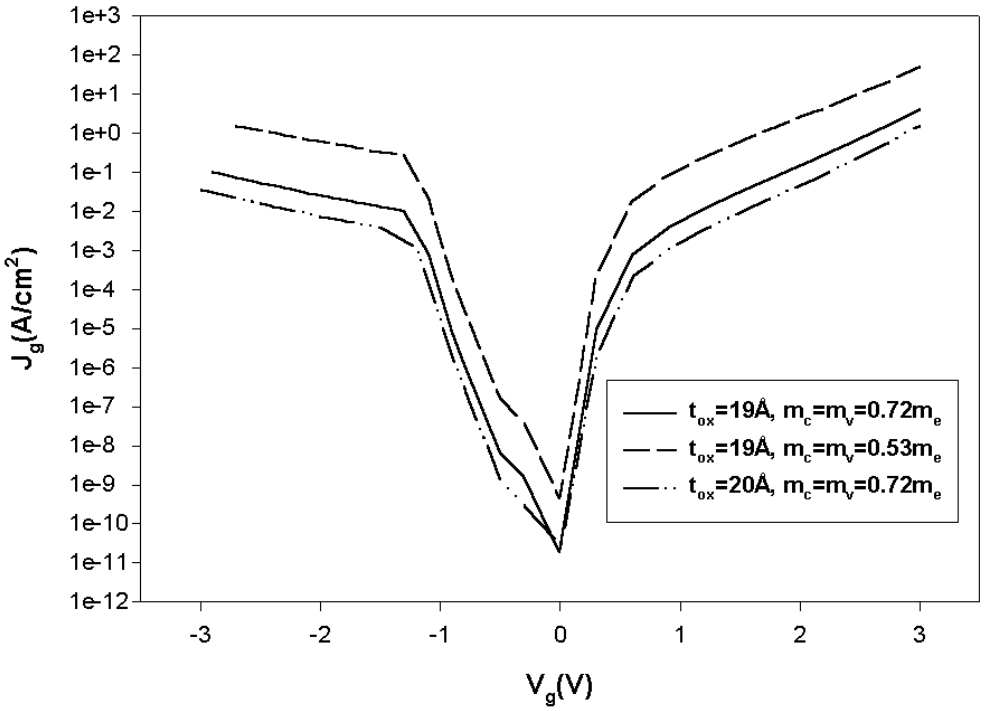


Figure 4.4: Comparison of gate current simulation for small changes in t_{ox}, m_c and m_v with $m_c/m_v = 1$. The simulated device is a SiO_2 n^+ poly NMOSFET.

band-edge effective masses, and vice versa. If the SiO_2 thickness is underestimated, the band-edge effective masses will be overestimated. It was reported in [53] that the SiO_2 thickness is $\sim$ 20Å, but $\sim$ 19Å was extracted in this work from comparing gate capacitance simulation with experiment in accumulation. Thus, the extracted $m_c = m_v = 0.72m_e$ for the conduction-band component and $m_v = 0.34m_e$ for the valence-band component may be the over-estimation of the real values of SiO_2.

One set of Franz dispersion parameters is used for the conduction-band

component of gate current for both devices, whether it is substrate-injected or gate-injected. A different set is used for the valence-band component.

Because the total energy of electrons in the conduction-band component is at least 1.1eV greater than that of the valence-band component. The previous results, (Section 4.3.1), only prove that Franz dispersion works well for a limited energy range in the SiO_2. When the energy range of interest is great, e.g. 1.1eV, different sets of parameters may be needed for different energy ranges. However, the energy dispersion used remains the same for electrons whether they tunnel from the substrate to the gate or from the gate to the substrate.

At low voltages in the accumulation of the PMOSFET in Figure 4.3, the valence-band component of gate current is greater than the conduction-band component. When the gate voltage is greater than $\sim 2.8V$, the conduction-band component becomes dominate. The valence-band component was not seen in the PMOSCAP (Figure 4.1(b)) at low voltages because there are no source/drain to deplete the increased minority holes in the channel.

4.3.3 SiO_2 NMOSFETs : C_g, I_g-V_g vs. Thickness and Temperature

Two SiO_2 n+poly NMOSFETs with different thicknesses were fabricated for studying the temperature-, voltage-, and thickness-dependent gate capacitance and current behavior. Comparison of C_g, I_g-V_g simulation with experiment for a SiO_2 poly NMOSFET at different temperatures is shown in Figure 4.5. The EOT of the SiO_2 layer is 20Å from C_g-V_g simulation. Good

agreement is achieved in inversion and accumulation for both the gate capacitance and current simulation at different temperatures. As seen from both the experiment and simulation, the temperature dependence region of the gate capacitance vs. voltage is the same as that of the gate current vs. voltage. Thus, it can be concluded that the temperature dependence of gate current is attributed to that of the charge distribution in the system rather than the transport mechanism through the SiO_2 layer.

At the flat-band, $V_g \sim -0.85$V, transition of gate current mechanism is seen in both the experiment and simulation. A quantum well is formed in the conduction band when the gate voltage is greater than the flat-band voltage. When the gate voltage is lower than the flat-band voltage, the quantum well is formed in the valence band. Without a quantum well formed for the carriers to reside in, quantum confinement effects of the carriers are less important and are neglected in this work. Such a transition in the model can also contribute to the apparent transition in the modeled gate current behavior.

Figure 4.6 shows the thickness dependence of C_g, I_g-V_g simulation compared with experiment. Both devices were fabricated by the same process. In simulation, only the SiO_2 thicknesses are different: one is 20Å, and the other is 17.5Å, both obtained from C_g-V_g. The experimental data were measured at 25°C as used in simulation. Using $m_c = m_v = 0.53m_e$, the gate current simulation in inversion agrees with experiment in a similar manner as for the NMOSFET with $t_{ox} = 21.8$Å in Figure 4.1. The thickness-dependence is also modeled well. However, $m_c = 0.4m_e$ and $m_v \gg m_c$ can provide a better fit at

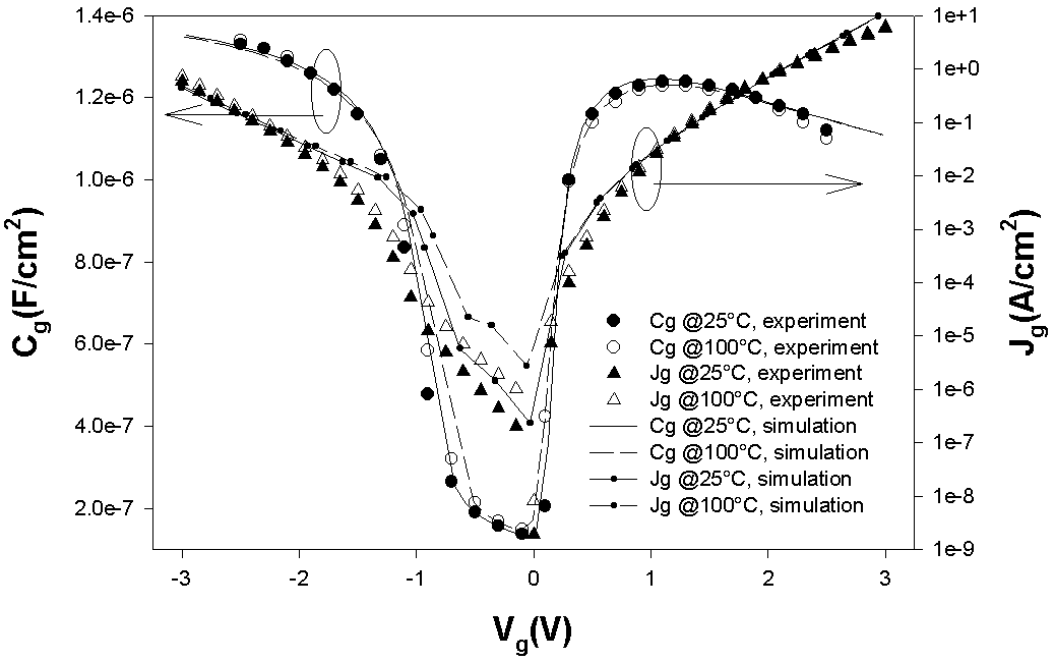


Figure 4.5: Comparison of C_g, I_g-V_g at different temperatures for a SiO_2 poly NMOSFET (sample I). EOT was found to be 20.0Åfrom C_g-V_g simulation.

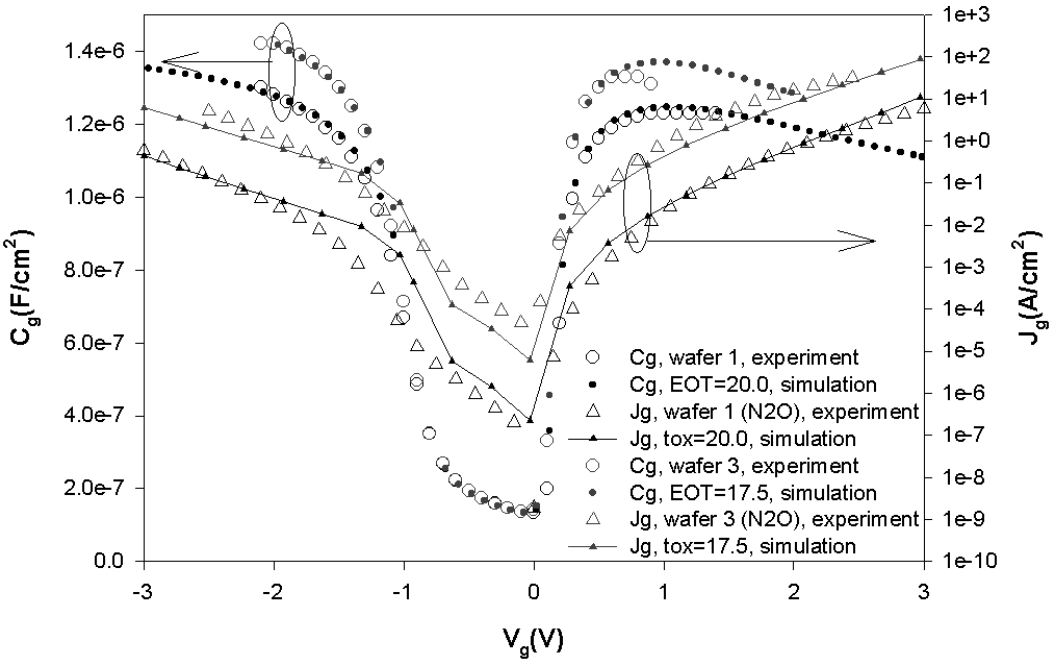


Figure 4.6: Comparison of C_g, I_g-V_g simulation with experiment for SiO_2 poly NMOSFETs with different thicknesses. EOTs of sample I and II are 20Å and 17.5Å, respectively, which were obtained from C_g-V_g simulation. The temperatures in experiment and simulation are both 25°C.

low gate voltages in inversion, and the accumulation gate current can also be simultaneously fitted, (Figure 4.6).

4.3.4 $k_{||}$-Conservation in SiO_2 MOS Devices

It was pointed out that $k_{||}$-conservation has been controversial because of the well-known inconsistencies between the assumption and experiment were found between gate currents in SiO_2-MOS devices fabricated on Si(110) and Si(111) [51]. According to Equation (2.4), because $k_{||}$ is conserved, much

lower gate current (from Si conduction band) is expected for Si(111) because of the effectively higher barrier seen by the tunneling electron from the off-axis valleys, (Equation 2.4). However, not so significantly lower gate current is observed in experiment, e.g. Reference [51].

In the previous discussion, different sets of parameters have to be used in Franz dispersion respectively for the conduction-band and valence-band component of gate currents. This requirement of different sets of parameters is because that Franz dispersion does not well approximate the energy dispersion of SiO_2 over the energy range from the silicon conduction band edge to the silicon valence band edge. Because the tunneling electron conserves its total energy before, during, and after it crosses the SiO_2 layer, the energy dispersion of SiO_2 it sees is determined by its total energy. If the energy range of electrons of interest is small, (Section 4.3.1), then the Franz dispersion relation is a good approximation using the same parameters. If the energy range of interest is large, different sets of Franz dispersion parameters are needed for different ranges of energies, (Section 4.3.2).

By the same token, because $k_{||}$ is conserved, the energy dispersion seen by the tunneling electron is determined by its $k_{||}$. When the range of $k_{||}$ of interest is small, e.g. that in the small neighborhood near one conduction-band valley minimum, Franz dispersion relation is good approximation. However, when the range of $k_{||}$ of interest is as great as that between the minima of the longitudinal valley and transverse valley in Si(100), different sets of Franz dispersion parameters may be needed for electrons from different valleys even

their total energies are the same. Thus, for electrons with the same energy but in different off-axis valley, tunneling probabilities may be different because of different sets of Franz dispersion parameters are needed.

Figure 4.7 shows the comparison of gate current simulation between $k_{||}$-conserved and $k_{||}$-relaxed models for a n+poly SiO_2 NMOSFET fabricated on Si(100) with t_{ox}=19Å. Also shown is a modified $k_{||}$-conserved calculation, which will be discussed later. In the $k_{||}$-conserved calculation, the same Franz dispersion is applied for the longitudinal valley as used for the transverse valley. The parameters are those used in Section 4.3.2. Because of $k_{||}$-conservation, gate currents from the transverse valley are neglected. In the $k_{||}$-relaxed calculation, the same Franz dispersion is used but $k_{||}$ is set to be always zero when calculating the tunneling probability. Setting $k_{||} = 0$ allows the total energy of electron, whether it is in the longitudinal valley or transverse valley, to contribute to the tunneling. Thus, both electrons from the longitudinal valleys and transverse valleys contribute to the gate current. However, Figure 4.7 shows that very small difference between these two calculated gate currents is seen in inversion, but great difference in accumulation.

In inversion, because of quantum confinement effects, the first subband energy in the longitudinal valley is much lower than the subbands in the transverse valley. This is because of the greater effective mass along the (100)-direction in the longitudinal valley, $0.9m_e$ for simulation in Figure 4.7, than that in the transverse valley, $0.2m_e$. As a result, most carriers in the channel reside in the longitudinal valley. Thus, considering only the gate current from

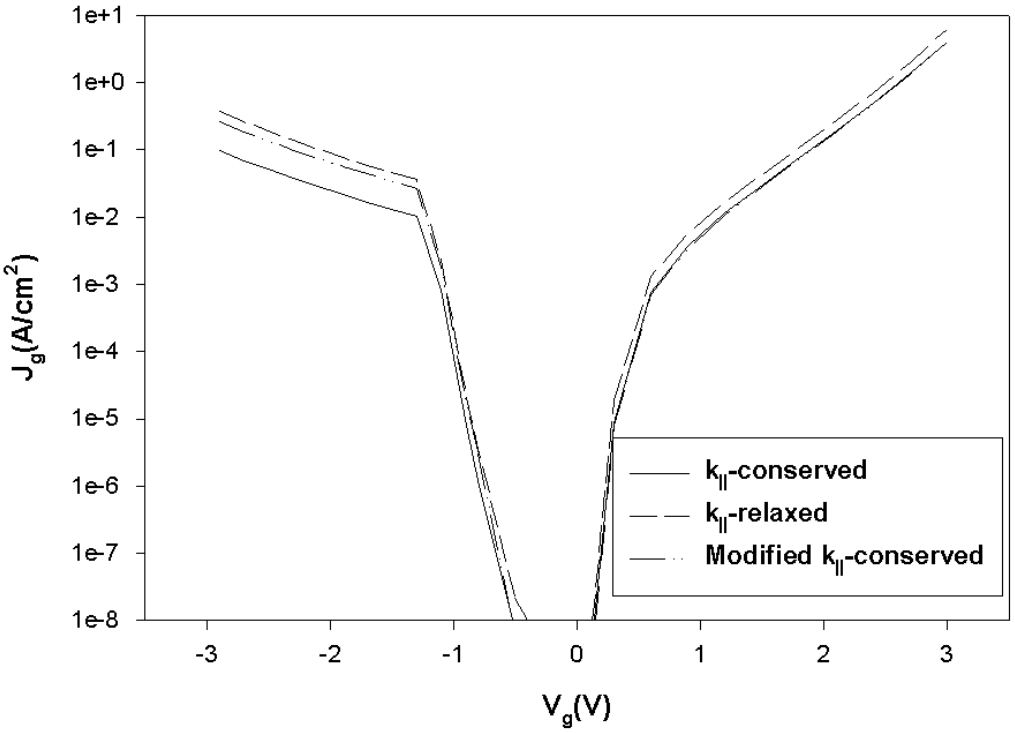


Figure 4.7: Gate currents are calculated with three assumptions: $k_{||}$-conserved, $k_{||}$-relaxed, and modified $k_{||}$-conserved. The simulated device is a n^+ poly SiO_2 NMOSFET fabricated on Si(100) with SiO_2 thickness of 19Å.

the longitudinal valleys is good approximation in inversion. In accumulation, however, the quantum confinement effects on the conduction band carriers are less important and are neglected in calculation. More electrons reside in the four transverse valleys than the two longitudinal valleys. If $k_{||}$-conservation is relaxed, gate currents from the transverse valleys dominate.

In Figure 4.6, when $k_{||}$ is conserved, gate current simulation agrees well with experiment simultaneously in inversion and accumulation using the same Franz dispersion. However, this does not conclude that the same Franz dispersion can be applied for both the conduction-band longitudinal valleys and the transverse valleys, but that neglecting the gate current from the transverse valleys can be a good approximation for SiO_2 MOS devices fabricated on Si(100).

If a different set of Franz dispersion is needed for the transverse valleys, Equation (2.3) is modified as

$$k_z^2 = k^2 - (k_{||} - k_{||,min})^2, \tag{4.2}$$

where $k_{||,min}$ is the $k_{||}$ of the valley minimum. The simulation result is shown in Figure 4.7 for comparison.

4.3.5 ZrO_2 Metal Gate NMOSCAPs

Two ZrO_2 metal-gate NMOSCAPs with different thicknesses were fabricated. From C_g-V_g simulation, the EOTs can be extracted. Because of fabrication variance, from TEM, only the range of the physical thickness of each

dielectric layer can be determined. From this information, and assuming the dielectric constant of ZrO_2 as 20, the dielectric constant of the interfacial layer and the physical thickness of each dielectric layer can be determined by imposing that all the parameters need to satisfy Equation (2.11) for both devices.

From the voltage- and temperature-dependent behavior of the gate current, it can be concluded that tunneling is the primary transport mechanism for the electrons through the high-K stack. Direct, Fowler-Nordheim tunneling, and thermionic emission were simultaneously considered in the gate current model. Consistent parameters were used in simulation to provide good fit with experiment for both devices at different temperatures.

From the gate current simulations, the Franz dispersion parameters for ZrO_2 were found as E_g=5.7eV, ΔE_c=1.45eV, and $m_c = m_v = 0.35m_e$. As for the interfacial silicate, E_g=4.5eV, ΔE_c=1.0eV, and $m_c = m_v = 0.35m_e$.

A HfO_2 metal-gate NMOSCAP was also studied. The parameters found to provide good fit with experiment by simulation are E_g=5.7eV, ΔE_c=1.6eV, and $m_c = m_v = 0.32m_e$ for the HfO_2 layer; E_g=4.5eV, ΔE_c=1.2eV, and $m_c = m_v = 0.5m_e$ for the Hf-silicate layer.

For comparison, other methods have been reported to obtain the band gaps and band offsets for different high-K materials. In [43], the band gaps and conduction band offsets of the ZrO_2 and HfO_2 films were calculated from Schottky barrier height based on metal- induced gap states and charge neutrality levels. It was found that $E_g(ZrO_2)$=5.8eV, $\Delta E_c(ZrO_2)$=1.4eV, $E_g(HfO_2)$=6eV,

and $\Delta E_c(HfO_2)$=1.5eV. Yamaguchi et al performed XPS analysis to obtain the band gap and band offset of the ZrO_2 and Zr-silicate layer [52].

4.4 Conclusions and Summary

A gate capacitance-current analysis method was proposed to better understand the gate capacitance and current vs. voltage/temperature behavior. Applied to SiO_2 devices, it was shown that the extracted Franz dispersion parameters for gate current are consistent for each fabrication process. Gate capacitance and current thus can be considered simultaneously for better device design. ZrO_2 devices with different thicknesses were also studied. Consistent parameters were used in simulation to fit the experimental data in accumulation. A HfO_2 metal-gate NMOSCAP was also studied and parameters were extracted. These parameters were compared with those obtained by other methods.

From comparing simulation with experiment for different SiO_2 devices, the $k_{||}$-conservation rule was explored. Within the context of the energy-dispersion model, the widely-known controversy about $k_{||}$ conservation was not seen for devices fabricated on Si(100). For SiO_2 NMOSFETs fabricated on Si(100), the calculated gate current could fit the experimental data in inversion and accumulation when neglecting the current from the transverse valleys.

Chapter 5

Scaling ZrO_2 Control/Tunnel Oxide for Floating-Gate Nonvolatile Memory Devices

Time-dependent characteristics of a floating-gate nonvolatile memory device are modeled for write, erase, and retention voltages. Effects of the floating quasi-Fermi level of the floating gate are considered by self-consistently calculating the charge distribution of the whole system. A trend study of scaling ZrO_2 and SiO_2 as the control/tunnel oxide is performed based on reasonable experimental data.

5.1 Introduction

A thick control oxide layer is required in floating-gate nonvolatile memory devices to assure low leakage currents between the control gate and floating gate for good retention [8, 44, 45]. To enhance the write and erase speeds and to reduce the programming voltage, the tunnel oxide thickness has been reduced [1]. However, a scaling limit of $\sim$ 8nm is seen for SiO_2 tunnel oxide, which also imposes limitation on scaling down the programming voltage [1, 10]. In order to further scale down the programming voltage, ZrO_2 was reported as the control oxide and shows good write and erase performance [4]. With

the high dielectric constant (high K) of ~ 20, a low equivalent oxide thickness (EOT) of control oxide layer with large physical thickness can be achieved. The large physical thickness ensures good retention; owing to the low EOT, a large portion of the applied control gate voltage drops over the thin tunnel oxide layer. Thus, high currents across the tunnel oxide can be obtained at low control gate voltages during write and erase, and good retention can be maintained when the device is idle.

In this chapter, the time-dependent characteristics of a n-channel floating gate nonvolatile memory device are modeled operating at write, erase, and retention voltages. Scaling trends of both SiO_2 and ZrO_2 as control or tunnel oxide are qualitatively studied based on reasonable experimental data. The leakage current across the tunnel (control) oxide is calculated by an energy-dispersion-based model [12]. Direct, Fowler-Nordheim (FN) tunneling, and thermionic emission currents are considered simultaneously. Table 5.1 shows the material parameters of SiO_2 and ZrO_2 used in leakage current calculation. These values were obtained from fitting simulation with experiment for MOS devices [11, 12].

5.2 Modeling

Figure 5.1 shows the device structure of a Metal/Control Oxide/Floating Gate (n-polysilicon)/Tunnel Oxide/p-Si device being modeled in this work. Unless otherwise stated, the following parameters are used in simulation: metal work function of 4.1eV, doping concentrations of the floating polysilicon

Table 5.1: Material parameters of SiO_2 and ZrO_2 for calculating leakage currents in floating gate devices.

	SiO_2	**ZrO_2**
Dielectric Constant	3.9	20
Conduction Band Offset (eV)	3.15	1.45
Band-edge Effective Mass (m_e)	0.53	0.35
Band Gap (eV)	9	5.7

gate and the silicon substrate of $2 \times 10^{19}\text{cm}^{-3}$ and $7 \times 10^{17}\text{cm}^{-3}$, respectively. The device is assumed to operate in quasi-thermodynamic equilibrium; thus, non-equilibrium hot channel electrons are not considered in this work. The Schrodinger Equation is solved in the silicon channel, and it is solved self-consistently with Poisson Equation, which is solved by considering the charge distribution of the whole system [49]. The effects of the charge distribution in the floating gate are not considered in this model, although the total amount of charge accumulated is considered, for the voltage drop across the floating gate is negligible and the charge distribution in the floating gate has minor effects on the time-dependent write, read, and retention characteristics as compared with others being investigated in this work.

Fixing the quasi-Fermi level of Si substrate, $E_{f,Si}$, as reference, the quasi-Fermi level of the control gate is determined by the applied control gate voltage, V_{cg},

$$E_{f,fg}(V_{cg}) = E_{f,Si} - V_{cg}, \tag{5.1}$$

where the Si substrate is grounded. The quasi-Fermi level (chemical potential) of the polysilicon floating-gate, $E_{f,fg}$, is "floating" because it is isolated by the control and tunnel oxides from the external applied voltage. Its value is determined according to the intrinsic doping concentration and the potential profile of the whole system:

$$E_{f,fg}(V_{cg}) = E_{f,fg}(V_{tnlox} = 0) - V_{tnlox}, \tag{5.2}$$

where V_{tnlox} is the voltage drop across the tunnel oxide layer; $E_{f,fg}(V_{tnlox} = 0)$

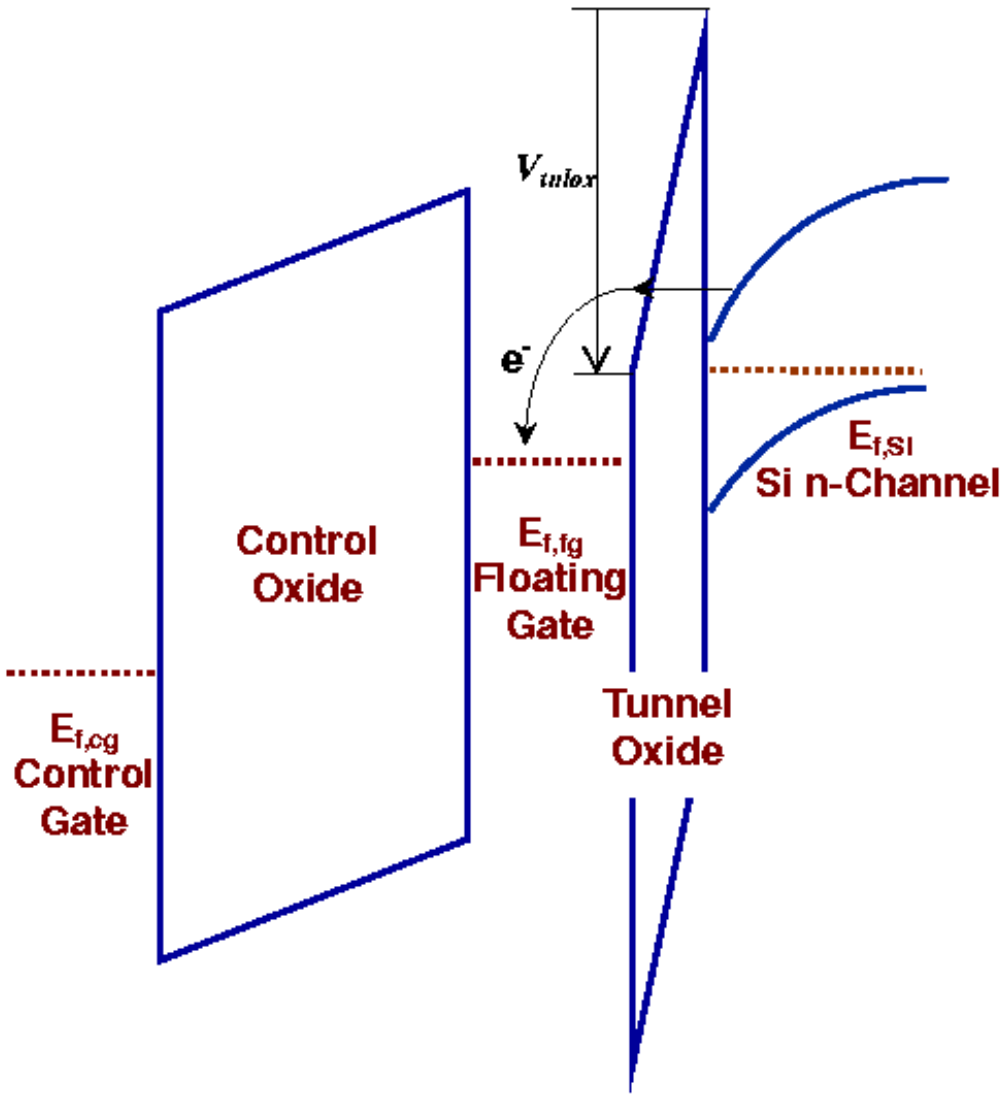


Figure 5.1: Device structure of the modeled floating gate nonvolatile memory device.

is the value at the flat band and is determined intrinsically by the doping concentration. Then the current is calculated according to the energy dispersion of the oxide, and the available carriers and empty states in either side of the oxide. The electron current flows from the higher quasi-Fermi-level side to the lower one.

The current between the floating gate and the control gate is neglected when modeling the time-dependent characteristics. Coulomb blockade effects are not considered. However, because the system is solved self-consistently, the voltage drop between the floating gate and silicon substrate will change when net charges are added into or removed from the floating gate even with the external voltage fixed. As a result, the potential barrier and the floating quasi-Fermi level will change, and the current will also change accordingly. Coulomb blockade effects can be neglected based on the assumption that the floating gate considered in this work is a macroscopic system, and the amount of particles added into or removed from such a system is comparatively negligible and will not alone change its chemical potential.

In the write and erase regimes, FN tunneling is the primary transport mechanism, but direct tunneling is dominant in the retention regime. Figure 5.2 shows the time-dependent characteristics of the modeled device operating at the write voltage, $V_{cg} = 18$V. In such a device, the tunnel oxide is SiO_2; the EOTs of the control and tunnel oxide layers are 13.5nm and 7.5nm, respectively. The write time for flat-band voltage shift of 4V requires $\sim$0.02sec.

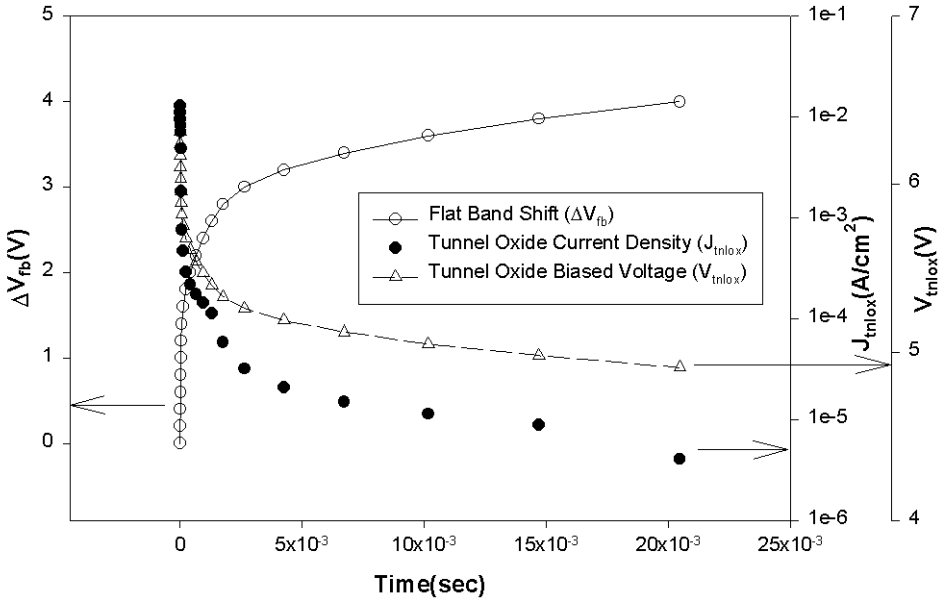
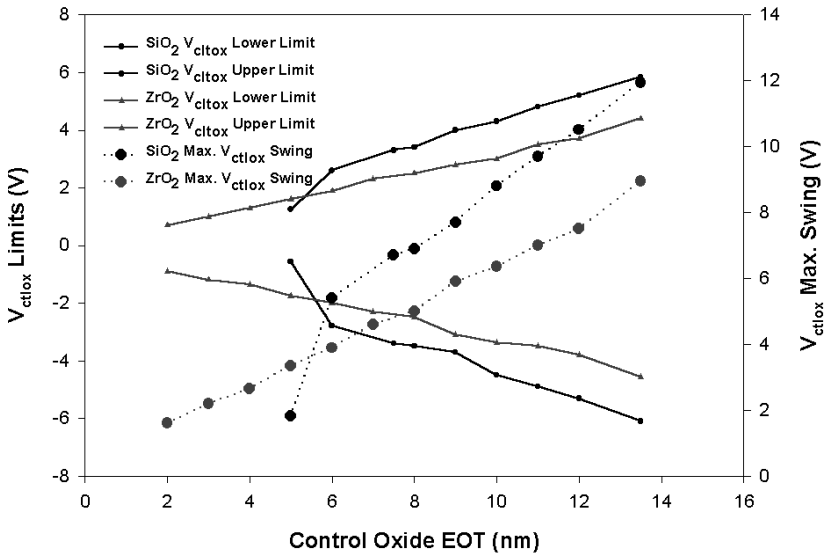


Figure 5.2: Time-dependent characteristics of a floating gate nonvolatile memory device biased at a constant $V_{cg} = 18V$.

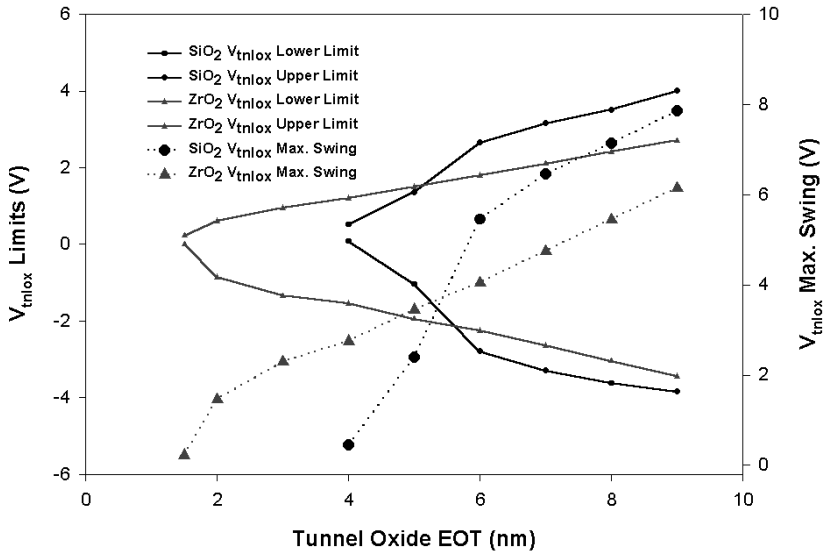
5.3 Discussion and Conclusions

A trend study is performed for ZrO_2 and SiO_2 as the control and tunnel oxides. Fowler-Nordheim tunneling is the primary transport mechanism for assessing write and erase performance, although direct tunneling and thermionic emission are also considered simultaneously. Metal/Control Oxide/n-Si and n-Poly/Tunnel Oxide/p-Si devices are simulated for the trend study for scaling the control oxide and tunnel oxide, respectively.

Retention performance is first assessed for obtaining the scaling limits of SiO_2 and ZrO_2 as the tunnel oxide and control oxide, (Figure 3). In this work, leakage current lower than 10^{-14}A/cm^2 is assumed as the requirement of the control and tunnel oxide. When the value of the voltage across the control (tunnel) oxide is increased, the leakage current is increased. Thus, the voltage swing range, which is between the upper (+) and lower (-) voltage limits to reach current density of 10^{-14}A/cm^2, is used as the parameter to assess the retention performance. Other methods can be used, but this approach allows us to obtain the voltage range that can be applied across the control (tunnel) oxide and is a better approach for considering scaling down the control gate voltage. With a constant leakage current of 10^{-14}A/cm^2, it takes 10^8 seconds ($\sim$10 years) to remove all the charges that are required in the floating gate to cause flat-band voltage shift by the order of 1V with a control oxide EOT of $\sim$10nm. Using the parameters in Table 5.1, it is found that the scaling limits of SiO_2 and ZrO_2 are reached at EOT of $\sim$4.0nm and $\sim$1.5nm as the control and tunnel oxide, (Figure 5.3).



(a)

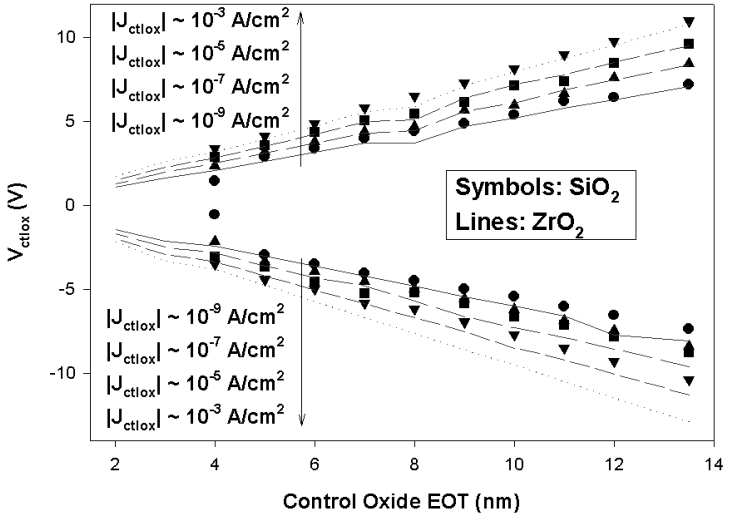


(b)

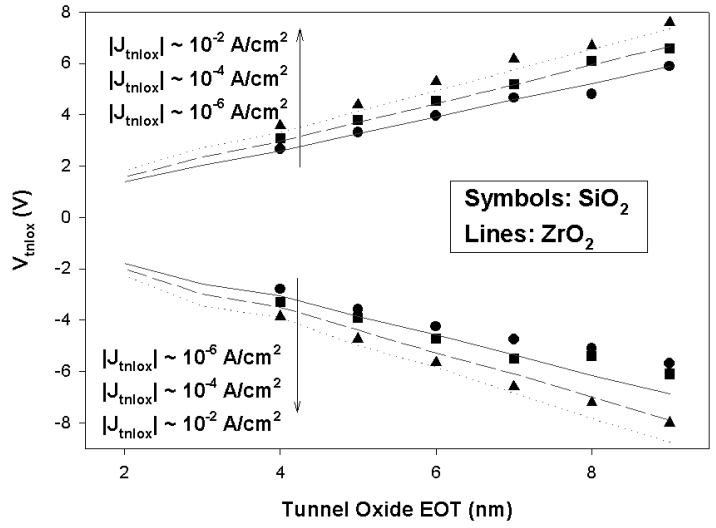
Figure 5.3: Retention scaling trend for SiO_2 and ZrO_2 as (a) control oxide and (b) tunnel oxide.

The same approach is applied to study the scaling trend for write and erase, (Figure 5.4). It is mandated that the leakage current through the control oxide should be much less than that through the tunnel oxide at the write and erase voltages to ensure net charges are added or removed. When EOT>4.0nm, SiO_2 is both better control and tunnel oxides for both write and erase because of its greater voltage ranges across the control oxide and smaller voltage ranges across the tunnel oxide (greater sensitivity to read and write voltages) for given required currents. However, when EOT<4.0nm, ZrO_2 is a better choice as the control oxide. It is not only because of its better retention characteristics, but also because of the greater ranges of voltages across the control oxide. With the parameters used, there is not a great difference between ZrO_2 and SiO_2 as tunnel oxide in the FN tunneling regime, in which the write or erase is performed. Though ZrO_2 has greater dielectric constant, $\sim$20, its barrier height is lower, $\sim$1.45eV. However, for the same reason, FN tunneling is achieved at a lower voltage. Once the FN regime is reached, tunneling through ZrO_2 becomes much greater than that through SiO_2.

With the EOT of control oxide fixed, scaling down the tunnel oxide thickness in general reduces the erase time with the control gate voltage fixed. For a device that consists of control oxide of EOT=13.5nm and SiO_2 as the tunnel oxide being biased at $V_{cg} = -15$V, the erase time increases instead of decreasing when the tunnel oxide is scaled down to $\sim$8.0nm, (Figure 5.5). This reversal of trend is a result of competing effects between the strong thickness and field dependencies of FN tunneling. FN tunneling may be greater for a



(a)



(b)

Figure 5.4: Write/Erase scaling trend for SiO_2 and ZrO_2 as (a) control oxide and (b) tunnel oxide.

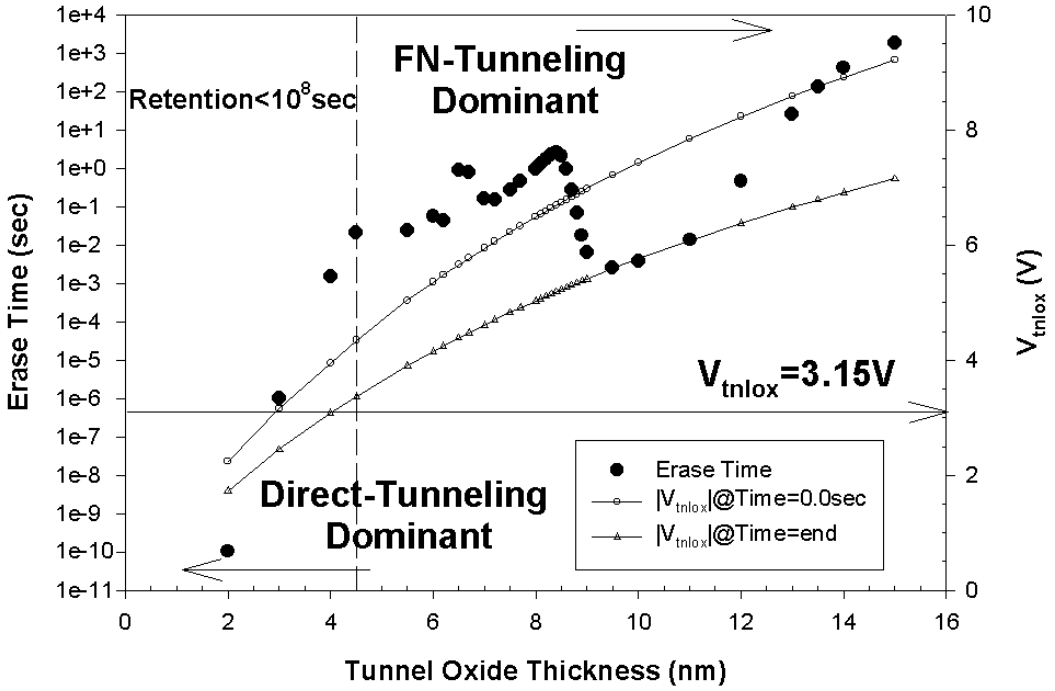


Figure 5.5: Erase time vs. Tunnel oxide thickness.

thicker tunnel oxide with a greater electric field than a thinner oxide with a lower electric field. The EOT ratio between the control oxide and the tunnel oxide determines their relative biased voltages. When the EOT of the tunnel oxide is decreased, less voltage drop is across the tunnel oxide for a fixed control gate voltage. As a result, the erase time may increase when the tunnel oxide thickness is being scaled down in the regime where FN tunneling dominates.

In conclusion, in order to scale down the control gate voltage and at the same time keep or reduce the erase time, the EOT of control oxide has to be scaled down along with scaling down the tunnel oxide. However, their

scaling limits are also imposed by the retention requirements. These results are obtained qualitatively based on reasonable experimental data. A more reliable quantitative result requires experimental verification along with a qualitative trend study of simulation.

Chapter 6

Conclusions and Recommendations

6.1 Summary and Conclusions

An energy-dispersion-based gate current model was established to simultaneously take into account the quantum confinement effects in the silicon channel, direct and Fowler-Nordheim tunneling, and thermionic emission transport through the gate dielectric. Both gate- and substrate-injected currents are modeled for the silicon conduction- and valence-band components. The energy dispersion in the dielectric band gap is approximated by the Franz 2-band dispersion model. The subband structures and carrier distribution in energy and position in the silicon channel are obtained by solving the Schrodinger and Poisson equations self-consistently, both for gate capacitance and gate current calculations. A self-consistent C_g,I_g-V_g model thus has been established.

This model was first validated by applying to SiO_2 devices. Simultaneous matching of gate capacitance and current simulation with experiment was achieved in inversion and accumulation using consistent device structure parameters. Based on such results, a gate capacitance-current analysis method was proposed to better understand the gate capacitance and current vs. volt-

age/temperature behavior. It was shown that the extracted Franz dispersion parameters for gate current are consistent for different SiO_2 devices fabricated, respectively, by different research groups. Gate capacitance and current thus can be considered simultaneously for extracting parameters and better device design.

ZrO_2 NMOSCAPs with different thicknesses were studied through integrated simulation and experimental analysis. Tunneling was found to be the primary transport mechanism of gate current in those devices. Consistent parameters were used in simulation to fit the experimental data in accumulation. A HfO_2 metal-gate NMOSCAP was also studied and parameters were extracted. These parameters were compared with those obtained by other methods.

Stacked gate dielectrics were characterized using dielectric constants, band gaps, band offsets with silicon, and band edge effective masses with a Franz 2-band energy dispersion relation, along with the physical thickness of each layer to assess capacitance/current vs. voltage behavior. Dielectric stack engineering thus can be possible for optimizing the tradeoff between EOTs and gate currents. Oxynitride, Si_3N_4/SiO_2, and high-K stacked dielectrics were qualitatively studied. More insulating and higher dielectric constant materials are desired for each layer of the stack; however, the role of transition region in the dielectric stack will then become more important.

Given the complicated effects of the transition region on the capacitance/current vs. voltage behavior, increasing the dielectric constant of the

interfacial layer seems the most feasible approach in the short-term. However, the possibly accompanying negative effects such as mobility degradation may be of concern.

From comparing simulation with experiment for different SiO_2 devices, the $k_{||}$-conservation rule was explored. Within the context of the energy-dispersion model, the widely-known controversy about $k_{||}$ conservation was not seen for devices fabricated on Si(100). For SiO_2 NMOSFETs fabricated on Si(100), the calculated gate current could fit the experimental data in inversion and accumulation when neglecting the current from the transverse valleys.

Time-dependent characteristics of a floating-gate nonvolatile memory device were modeled at write, erase, and retention voltages. Effects of the floating quasi-Fermi level of the floating gate were considered by self-consistently calculating the charge distribution of the whole system. in order to scale down the control gate voltage and at the same time keep or reduce the erase time, the EOT of control oxide has to be scaled down along with scaling down the tunnel oxide. However, their scaling limits are also imposed by the retention requirements. A trend study was performed to scale SiO_2 and ZrO_2 as the control oxide and tunnel oxide. When the EOT remains high, SiO_2 has better performance than ZrO_2 as both the control and tunnel oxide. When the EOT is scaled below $\sim$ 4.0nm, ZrO_2 should be applied.

6.2 Recommendations for Future Work

Important physics will not be revealed when the phenomena are not carefully observed. Further integration of the simulation and experimental analysis is required to define the scope of the proposed gate capacitance-current model and for improvements. Some important physics were also neglected to simplify the problem and require future development:

- Quantum confinement effects near the flat-band region
- Defect-related gate current mechanisms such as trap-assisted tunneling
- Wave function penetration effects on energy dispersion relations in different regions, especially for the near-interface regions
- Impact frequency for quantized holes in the silicon channel for gate current calculation.
- Interface states effects on the gate capacitance and current

For each new model, a new methodology may be required for validation and understanding its impact on the device behavior.

The gate current component from the transverse valleys of the silicon conduction band has more impact in accumulation than in inversion for devices fabricated Si(100). Exploring its behavior in accumulation from experiment and modeling will improve our understanding of the band structures of the gate dielectric.

More detailed physics pertaining to the floating gate in the nonvolatile memory devices should be further explored. Their relation with the floating quasi-Fermi level is important, especially for the device behavior during retention.

Bibliography

[1] The International Technology Roadmap for Semiconductors, 2001.

[2] M. M. Atalla, E. Tannenbaum, and E. J. Scheibner. Stablization on silicon surfaces by thermally grown oxides. *The Bell System Technical Journal*, page 749, 1959.

[3] J. L. Autran, R. Devine, C. Chaneliere, and B. Balland. Fabrication and charaterization of si-MOSFETs with PECVD amorphous Ta2O5 gate insulator. *IEEE Electron Device Letters*, pages 447–449, 1997.

[4] P. Blomme, B. Govoreanu, M. Rosmeulen, J. V. Houdt, and K. De Meyer. Multilayer tunneling barriers for nonvolatile memory applications. In *Proceedings for Device Research Conference*, pages 153–154, 2002.

[5] J. R. Brews, W. Fichtner, E. H. Nicollian, and S. M. Sze. Generalized guide for MOSFET miniaturization. *IEEE Electron Device Letters*, EDL-1(1):2–4, 1980.

[6] D. A. Buchanan and S.-H. Lo. Reliability and integration of ultra-thin gate dielectrics for advanced CMOS. *Microelectronics Engineering*, 36:13–20, 1997.

[7] S. A. Campbell, D. C. Gilmer, X. C. Wang, M. T. Hsieh, H. S. Campbell, W. Gladfelter, and J. Yan. MOSFET transistors fabricated with high

permittivity TiO2 dielectrics. *IEEE Transactions on Electron Devices*, 44(1):104–109, 1997.

[8] P. Candelier, B. De Salvo, F. Martin, B. Guillaumot, and G. Reimbold. Thinning oxide-nitride-oxide interpoly dielectric (11-13 nm) for 0.25 um flash cell memories. In *Proc. ESSDERC*, page 264, 1997.

[9] J. A. Duffy. *Bonding Energy Levels and Bands in Inorganic Solids*. New York: Wiley, 1990.

[10] T. Endoh, K. Shimizu, H. Iizuka, and F. Masuoka. A new write/erase method to improve the read disturb characterisctics based on the decay phenomena of stress leakage current for flash memories. *IEEE Transactions on Electron Devices*, 45:98–104, 1998.

[11] Y.-Y. Fan, J. C. Lee, G. Lucovsky, J. An, Q. Xiang, L. F. Register, and S. K. Banerjee. Conduction mechanisms and parameter extraction from C-V and I-V simulations and experiments for high-k gate dielectric stacks. In *Gate Stack Workshop Symposium*, 2002.

[12] Y.-Y. Fan, R. E. Nieh, J. C. Lee, G. Lucovsky, G. A. Brown, L. F. Register, and S. K. Banerjee. Voltage- and temperature-dependent gate capacitance and current model application to ZrO2 MOSCAPs. *IEEE Transactions on Electron Devices*, 49(11):000–000, 2002.

[13] M. V. Fischetti and S. E. Laux. Monte carlo study of electron transport in silicon inversion layers. *Physical Review B*, 48(4):2244–2274, 1993.

[14] W. Franz. *Handbych der Physik*, volume 17, page 155. 1956.

[15] J. Frenkel. On pre-breakdown phenomena in insulators and electronic semi-conductors. *Physical Review*, 43:647–648, 1938.

[16] C. J. Frosch and L. Derick. Surface protection and selective masking during diffusion in silicon. In *Proceedings of the Electrochemical Society*, page 547, 1957.

[17] S. M. George, O. Sneh, and J. D. Way. Atomic layer controlled deposition of SiO2 and Al2O3 using ABAB ... binary reaction sequence chemistry. In *Applied Surface Science*, page 460, 1994.

[18] X. Guo and T. P. Ma. Tunneling leakage current in oxynitride: Dependence on Oxygen/Nitrogen content. *IEEE Electron Device Letters*, 19(6):207–209, 1998.

[19] W. Harrison. Tunneling from an independent-particle point of view. *Physical Review*, 123(1):85–89, 1961.

[20] C. Hobbs, H. Tseng, K. Reid, B. Taylor, L. Dip, L. Hebert, R. Garcia, R. Hedge, J. Grant, D. Gilmer, A. Franks, V. Dhandapani, M. Azrak, L. Prabhu, R. Rai, S. Bagchi, J. Conner, S. Backer, F. Dumbuya, B. Nguyen, and P. Tobin. 80 nm poly-si gate CMOS with HfO2 gate dielectric. In *IEDM*, page 30.1, 2001.

[21] C. Hu. Gate oxide scaling limits and projection. In *IEDM*, pages 319–322, 1996.

[22] A. Inani, R. V. Rao, B. Cheng, and J. Woo. Gate stack architecture analysis and channle engineering in deep sub-micron MOSFETs. *Japanese Journal of Applied Physics*, 38:2266–2271, 1999.

[23] S. Jallepalli, J. Bude, W.-K. Shih, M. R. Pinto, C. M Maziar, and A. F. Tasch. Electron and hole quantization and their impact on deep submicron silion p- and n-MOSFET characteristics. *IEEE Transactions on Electron Devices*, 44(2):297–303, 1997.

[24] D. L. Kencke, W. Chen, H. Wang, S. Mudanai, Q. Ouyang, A. Tasch, and S. K. Banerjee. Source-side barrier effects with very high-k dielectrics in 50nm si MOSFETs. In *DRC*, pages 22–23, 1999.

[25] D. E. Kotecki. High-k dielectric materials for DRAM capacitors. *Semiconductor International*, page 109, 1996.

[26] L. D. Landau and E. M Lifshitz. *Quantum Mechanics: Non-Relativistic Theory*. Pergamon Press Ltd., 3 edition, 1977.

[27] B. H. Lee, Laegu Kang, W. Qi, R. Nieh, K. Onishi, and J. C. Lee. Ultrathin hafnium oxide with low leakage and excellent reliability for alternative gate dielectric application. In *IEDM*, pages 133–134, 1999.

[28] C. H. Lee, Y. H. Kim, H. F. Luan, S. J. Lee, T. S. Jeon, W. P. Bai, and D.-L. Kwong. MOS devices with high quality ultra thin CVD ZrO2 gate dielectrics and self-aligned TaN and TaN/poly-si gate electrodes. In *VLSI*, pages 137–138, 2001.

[29] J. R. Ligenza and W. G. Spitzer. The mechanisms for silicon oxidation in steam and oxygen. *Journal of Physics andl Chemistry of Solids*, page 131, 1960.

[30] J. R. Ligenza and W. G. Spitzer. Effects of crystal orientation on oxidation in steam and oxygen. *Journal of Physical Chemistry*, 65:2011, 1961.

[31] C. H. Ling. Interfacial polarization in al-Y2O3-SiO2-si capacitor. *Electron Letters*, page 1676, 1993.

[32] S.-H. Lo, A. Buchanan, Y. Taur, and W. Wang. Quantum-mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide nMOSFETs. *IEEE Electron Device Letters*, 18(5):209–211, 1997.

[33] S.-H. Lo, D. A. Buchanan, and Y. Taur. Modeling and characterization of quantization. polysilicon depletion, and direct tunneling effects in MOSFET's with ultra thin oxides. *IBM Journal of Research and Development*, 43:327–337, 1999.

[34] M. Lundstrom. Elementary scattering theory of the Si MOSFET. *IEEE Electron Device Letters*, 18(7):361–363, 1997.

[35] G. Luwicki and C. A. Mead. Experimental determination of E-K relationship in electron tunneling. *Physical Review Letters*, 16(21):939–941, 1966.

[36] J. Maserjian. Tunneling in thin MOS structures. *Journal of Vacuum Science and Technology*, 11(6), 1974.

[37] J. Maserjian and G. P. Peterson. Tunneling through thin MOS structures: Dependence on energy (e-kappa. *Applied Physics Letters*, 25(1):50–52, 1974.

[38] M. Mehrotra, J. C. Hu, A. Jain, W. Shiau, S. Hattangady, V. Reddy, S. Aur, and M. Rodder. A 1.2v, sub-0.08um gate length CMOS technology. In *IEDM*, pages 419–422, 1999.

[39] H. S. Momose, M. Ono, T. Yoshitomi, T. Ohguro, S. Nakamura, M. Saito, and H. Iwai. Tunneling gate oxide approach to ultra-high current drive in small-geometry MOSFETS. In *IEDM*, pages 593–596, 1994.

[40] K. K. Ng, S.A. Eshraghi, and T. D. Stanik. An improved generalized guide for MOSFET scaling. *IEEE Transactions on Electron Devices*, 40:1895–1897, 1993.

[41] W. Qi, R. Nieh, B. H. Lee, L. Kang, Y. Jeon, K. Onishi, T. Ngai, S.K. Banerjee, and J. C. Lee. MOSCAP and MOSFET characteristics using ZrO2 gate dielectric deposited directly on Si. In *IEDM*, pages 145–148, 1999.

[42] F. Rana, S. Tiwari, and D. A. Buchanan. Self-consistent modeling of accumulation layers and tunneling currents through very thin oxides. *Applied Physics Letters*, 69:1104–1106, August 1996.

[43] J. Roberston. Band offsets of wide-band-gap oxides and implications for future electronic devices. *Journal of Vacuum Science and Technology B*, 18(3):1785–1791, 2000.

[44] E. F. Runnion, S. M. Gladstone, R. S. Scott, and D. J. Dumin. Limitation on oxide thicknesses in ash EEPROM applications. In *Proc. IEEE/IRPS*, page 93, 1996.

[45] B. D. Salvo, G. Ghibaudo, G. Pananakakis, G. Reimbold, F. Mondond, B. Guillaumot, and P. Candelier. Experimental and theoretical investigation of nonvolatile memory data-retention. *IEEE Transactions on Electron Devices*, 46(7):1518–1524, 1999.

[46] W.-K Shih, E. X. Wang, S. Jallepalli, F. Leon, C. M. Maziar, and A. F. Tasch. Modeling gate leakage current in nMOS structures due to tunneling through an ultra-thin oxide. *Solid-State Electronics*, 42(6):997–1006, 1998.

[47] S. Song, W. K. Kim, J. M. Ha, G. G. Lee, J.-H. Ku, H. S. Kim, C. S. Kim, C. J. Choi, T. H. Choe, J. Y. Yoo, W. S. Song, J. W. Park, S. H. Jeong, D. H. Baek, K. Fujihara, H. K. Kang, S. I. Lee, and M. Y. Lee. High performance transistors with state-of-the-art CMOS technologies. In *IEDM*, pages 427–430, 1999.

[48] S. Thompson, P. Packan, and M. Bohr. MOS scaling: Transistor challenges for the 21st century. *Intel Technology Journal*, Q398, 1998.

[49] The University of Texas at Austin. *UTQUANT Manual*, 1997.

[50] L. Vegard. The constitution of mixed crystals and the space occupied by atoms. *Z. Physik*, 5:17–26, 1921.

[51] Z. A. Weinberg. On modeling oxide tunneling current. *Journal of Applied Physics*, 53(7):5052–5056, 1982.

[52] T. Yamaguchi, H. Satake, N. Fukushima, and A. Toriumi. Band diagram and carrier conduction mechanism in ZrO2/Zr-silicate/Si MIS structure fabricated by pulsed-laser-ablation deposition. In *IEDM*, pages 19–22, 2000.

[53] H. Yang and G. Lucovsky. Integration of ultrathin (1.6 2.0nm) RPECVD oxynitride gate dielectrics into dual poly-Si gates into submicron CMOSFETs. In *IEDM*, pages 245–248, 1999.

[54] I. Y. Yang, K. Chen, P. Smeys, J. Sleight, L. Lin, M. Ieong, E. Nowak, S. Fung, E. Maciejewski, P. Varekamp, W. Chu, H. Park, P. Agnello, S. Crowder, F. Assaderaghi, and L. Su. Sub-60nm physical gate length SOI CMOS. In *IEDM*, pages 431–434, 1999.

[55] N. Yang, W. K. Henson, J. R. Hauser, and J. J. Wortman. Modeling study of ultrathin gate oxides using direct tunneling current and capacitance-voltage measurements in MOS devices. *IEEE Transactions on Electron Devices*, 46(7):1464–1471, 1999.

Vita

Yang-Yu Fan was born in Taipei, Taiwan on 29 March 1972, the son of Bin-Kwong Fan and Jui-Chin Chen. He received the Bachelor of Science degree in Physics from National Taiwan University. After serving two years in the army, he entered University of Texas at Austin in Spring, 1997 to pursue a Ph.D. degree in Physics. He started his career in electrical engineering as a research assistant working on surface-micromachining a micro-opto-electrical-mechanical device. Then he transferred to the Department of Electrical and Computer Engineering in 1998. After completing his master's project, he joined the device modeling group under supervision of Professor Sanjay K. Banerjee and Leonard F. Register. Since 1999, he started pursuing Ph.D. by performing research on high-K gate Modeling. He was honored by the army in 1996 for distinguished services. He was also recognized as an outstanding intern by Motorola for his summer project in 2000. He is a member of IEEE.

Permanent address: No. 3 Ln. 103 Tien-Yu St., Taipei, Taiwan,
R.O.C.

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.