Modeling of direct tunneling gate current and gate capacitance in deep submicron MOSFETs with high-K dielectric.

A thesis submitted to
the department of Electrical and Electronic Engineering
of
Bangladesh University of Engineering and Technology
in partial fulfillment of the requirement
for the degree of
MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING

by
Mohammad Mojammel Al Hakim

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY
2002

i

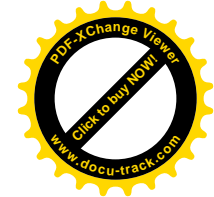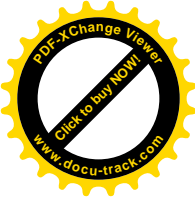The thesis titled "**Modeling of direct tunneling gate current and gate capacitance in deep submicron MOSFETs with high-K dielectric**" submitted by Mohammad Mojammel Al Hakim, Roll No.: 040006216P, Session: April 2000 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING on March 03, 2002.

BOARD OF EXAMINERS

1. _____

   Dr. Anisul Haque   **Chairman**
   Associate Professor   (Supervisor)
   Department of Electrical and Electronic Engineering
   BUET, Dhaka-1000, Bangladesh.

2. _____

   Dr. M. Rezwan Khan   **Member**
   Professor
   Department of Electrical and Electronic Engineering
   BUET, Dhaka-1000, Bangladesh.

3. _____

   Dr. M. Jahangir Alam   **Member**
   Assistant Professor
   Department of Electrical and Electronic Engineering
   BUET, Dhaka-1000, Bangladesh.

4. _____

   Dr. M. M. Shahidul Hassan   **Member**
   Professor and Head   (Ex-officio)
   Department of Electrical and Electronic Engineering
   BUET, Dhaka-1000, Bangladesh.

5. _____

   Dr. A. K. M. Akhter Hossain   **Member**
   Assistant Professor   (External)
   Department of Physics
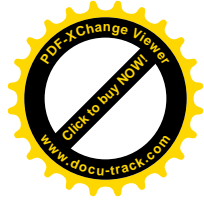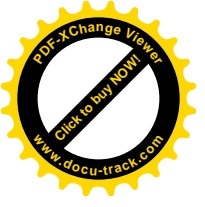   BUET, Dhaka-1000, Bangladesh.

Declaration

I hereby declare that this thesis or any part of it has not been submitted elsewhere for

the award of any degree or diploma.

Signature of the candidate
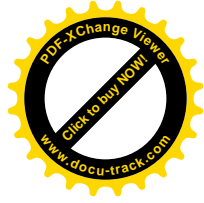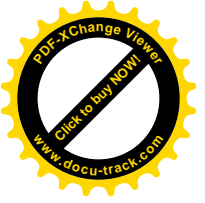
_____

(Mohammad Mojammel Al Hakim)
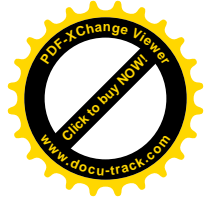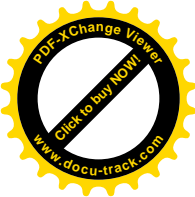
Dedication

To My Parents

Contents

List of Figures

vii

List of tables

Acknowledgement

I wish to convey my heartiest gratitude and profound respect to my supervisor Dr. Anisul Haque, Associate Professor, Department of Electrical and Electronic Engineering (EEE), Bangladesh University of Engineering and Technology (BUET), Bangladesh, for his continuous guidance, suggestions and wholehearted supervision throughout the progress of this work, without which this thesis never be materialized. I am grateful to him for acquainting me with the world of advanced research.

I am grateful to Dr. M. M. Shahidul Hassan, Professor and Head, Dept. of EEE, Bangladesh University of Engineering and Technology (BUET), who provided with the computing facilities of the department and cooperation to complete the work. I want to thank M. Zahed Kauser, former Lecturer, Dept. of EEE, BUET for his continuous suggestions, inspiration, encouragement and many fruitful discussions.

I want to thank my friends Khairul, Yunus and many others, who were directly or indirectly related to this work, for their support and encouragement. I also thank all the personnel at the departmental library, BUET reference library and xerox section for providing me with the valuable journals and thesis papers to complete this work.
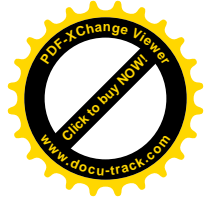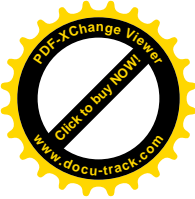
Abstract

Scaling down of MOS device dimensions is accompanied by a decrease in gate-oxide thickness and an increase in substrate doping density. When gate oxide thickness becomes less than 2 nm, a subtantial current follows through gate-oxide due to direct tunneling. In order to reduce this current, International Technology Roadmap for Semiconductors (ITRS) has suggested replacement of SiO_2 gate insulator layer by high-K dielectrics. For a given equivalent oxide thickness (EOT), high-K dielectrics offer greater physical thickness. The direct tunneling (DT) current and the gate capacitance for an inverted n-MOS device with different dielectrics used as gate insulator is studied. Coupled Schrodinger's and Poisson's equations are solved self-consistently. Open boundary conditions, taking account the wavefunction tail inside the gate dielectric within the self-consistent loop are used to solve Schrodinger's equation. DT current increases exponentially with the decrease of conduction band offset for electrons travelling from silicon substrate to dielectric. As general trend of dielectrics is to decrease of conduction band offset with the increase of dielectric constant, use of high-K material as gate insulator results in prominent influence of direct tunneling of carriers on potential profile. Therefore in DT current calculation effect of wavefunction penetration on potential profile is incorporated within self-consistent loop. Results of this simulation is compared with published experimental results and also with the results of the simulation where penetration effect on potential profile is neglected. Results show that neglect of wavefunction penetration effect on potential profile causes underestimation of DT current. A comprehensive analysis of the effect of wavefunction penetration on the gate capacitance of the MOSFETs with high-K dielectrics is also done. Gate capacitance from conventional modeling is found to be independent of dielectric materials for a given EOT. The study reveals that accounting for wavefunction penetration into the gate dielectric causes gate capacitance to vary from material to material for a given EOT. Consequently wavefunction penetration effects must be considered to determine properties of future generation devices where high-K dielectrics will be employed as gate insulator.

Chapter 1

Introduction

Due to the rapid advancement of the semiconductor device fabrication technology, metal-oxide-semiconductor (MOS) devices recently approached the deep submicron regime [1] and aggressive scaling of gate dielectric thickness has continued. According to the International Technology Roadmap for Semiconductors (ITRS) 2001, in sub-100 nm generation devices, an equivalent gate oxide thickness of less than 2 nm will be required [2]. Different dimensions (gate length, oxide thickness etc) are scaled down and the substrate doping concentration is increased. This results in large surface electric fields and steep potential wells even near threshold leading to significant quantization of carriers in the direction perpendicular to the channel. Due to quantization of energies of the inversion layer carriers in quantum well, the distribution of carriers can no longer be represented by semi-classical models [3,4], rather quantum mechanical (QM) models must be used. QM approach is essential for the calculation of several parameters; such as gate capacitance, carrier concentration, effective oxide thickness, drive current, on-state series resistance, polysilicon work function, gate leakage current etc. Since carriers in the inversion layer, localized near Si/SiO_2 interface, are quantized into two dimensional (2-D) subband structure, the finite thickness of the wave function of 2-D carriers results in the inversion-layer capacitance, C_{inv}, which acts as a series capacitance with the oxide capacitance. As a consequence, C_{inv} degrades the gate capacitance and the gate oxide thickness determined from the gate capacitance becomes thicker than the physical thickness.

As potential barrier and thickness of oxide layer is finite, inversion electrons may tunnel through the oxide barrier and reach gate electrode. In deep sub-micron MOSFETs, the tunneling probability is high and a dc gate leakage current due to the quantum mechanical tunneling flows in these devices. As the thickness of the oxide layer decreases, the tunneling current increases in a nearly exponential manner. The increased current not only adversely affects the MOS device performance but also greatly increases the standby power consumption of a highly integrated chip. Therefore, there has been much interest in finding a high-permittivity gate insulator with equivalent SiO_2 thickness and sufficient barrier height as a replacement for SiO_2. Physical thickness of such an insulating layer will be much higher than that of the corresponding SiO_2 layer, thereby reducing direct tunneling current considerably. But reducing oxide thickness down to less than 2 nm or using high-K dielectrics make the increase of the gate-oxide capacitance indispensable. In MOSFETs with such high capacitance, inversion layer thickness or inversion layer capacitance can have a significant influence on the electrical characteristics. However, since the barrier height of a dielectric material tends to decrease

with increasing dielectric constant [5], strong tradeoffs exist between various alternative dielectrics.

1.1 Literature Review

The rapid progress of complementary metal-oxide-semiconductor (CMOS) integrated circuit technology since the late 1980's has enabled the silicon based microelectronics industry to simultaneously meet several technological requirements to fuel market expansion. These requirements include performance (speed), low static (off-state) power, and a wide range of power supply and output voltages. This has been accomplished by developing the ability to perform a calculated reduction of the dimensions of the fundamental active device in the circuit: the field effect transistor (FET)-a practice termed 'scaling'. The key element enabling the scaling of the silicon based metal-oxide-semiconductor field effect transistor (MOSFET) is the materials (and resultant electrical) properties associated with the dielectric employed to isolate the transistor gate from the Si channel in CMOS devices for decades: silicon dioxide. Since its use as gate insulator in 1957, many research works have been devoted to the conduction through silicon dioxide [6]. In the last three decades two approaches are taken in explaining the physical origin of the experimental gate current. One explains with band to band tunneling and the other is with trap assisted tunneling. In 1969 Lezlinger and Snow [7] showed that the current through thermally grown silicon dioxide is 'electrode limited'. Like other similar works done at that time, they overlooked quantum mechanical effects in the silicon inversion layer and their model under estimated the measured current. Stern and Howard [8] first considered QM treatment and solved Schrödinger's equation mostly concerning with the lowest levels of subbands split up from the conduction band. They assumed the potential profile near the interface as triangular and calculated wave functions from the variational principle. Later Stern [9] solved the self-consistent solution of Schrödinger's and Poisson's equations. Self consistent results for energy levels, populations and charge distributions were given for n-type inversion layer on p-type silicon. QM effects were taken into account in the effective mass approximation and the envelope wave function was assumed to vanish at the oxide-semiconductor interface. This assumption is equivalent to assuming an infinite potential barrier at the interface. It is considered to be justified since the barrier height at the interface is high enough (~ 3.1 *eV)* and hence, the contribution of the wave function tail inside the oxide is extremely small. Moglestue [10] extended the self-consistent calculations to hole inversion charges in the same way as Stern did. The charge distribution for holes and electrons was calculated for the {100} oriented Si-SiO$_2$ interface, both for weak and strong inversion, by solving Schrodinger's and Poisson's equations for particles obeying Fermi-Dirac statistics. It was shown that for weak inversion, the self consistent result did not deviate significantly from those obtained assuming a triangular potential well, but for strong inversion region carriers tend to move closer to the oxide.

Ohkura [11] studied the dependence of self-consistent solution on the substrate doping concentration. In his results he found shift of subthreshold curves with the increase in doping level, especially for impurity concentration greater than 10^{17} /cm^3. He also proposed a method for transferring the QM effects into the semi-classical calculation through the modification of gate oxide thickness and flat band voltage. Quantization of

2

electrons in inversion layer modifies the electron density. Therefore gate voltage dependency of surface potential at the interface must be calculated considering the QM effects. Actually it is found that, due to this dependence, both the non-degenerate and degenerate approximations have deviated increasingly from the actual quantum case with increasing gate voltage. This deviation also increases with decreasing oxide thickness.

Dort *et al.* proposed a simple model for explaining measurements in the high doping level regime where the conventional model fails to reproduce the experimental results [12]. The model proposed by them uses the same drift-diffusion approximation used by semi-classical model [3]. This model modifies the intrinsic carrier concentration for the silicon bandgap inversion conditions. They showed that their model agreed with the results given by self-consistent QM calculation.

Rana *et al.* [13] modeled gate current including the accumulation regime self-consistently and showed good agreements with measured tunneling current. In their work they also explained why classical calculations were successful in predicting the tunneling currents. Self consistent solutions give a greater substrate potential drop and classical cal-culations give larger electric field in the oxide. The tunneling rate is dependent exponentially on both oxide electric field and the energy of incident electron. With increased gate voltage the difference in oxide electric fields predicted by the classical and self consistent models increases but at the same time the energies of electrons increase in self consistent models making the predicted tunneling currents similar in both models. They used finite element method for the purpose. To include wave function penetration into gate-oxide, they assumed that the wave function goes zero at gate-electrode-gate-metal interface instead of Si/SiO$_2$ interface.

Tunneling current from MOS inversion layer was studied by Lo *et al.* [14]. They have showed that for the purpose of modeling tunneling characteristics of electrons exhibiting quasi two-dimensional character, the transmission probability applicable to an incident Fermi gas of free electrons is no more an acceptable concept and the well known WKB (Wentzel-Krammer-Brioullion) approximation or the numerical integration of Airy function is not valid. In their work an accurate determination of the physical oxide thickness was achieved by fitting experimentally measured capacitance-versus-voltage curves to quantum-mechanically simulated capacitance-versus-voltage results. The lifetimes of quasi-bound states and the direct tunneling current was calculated using a transverse resonant method, applicable for electromagnetic waves in a non-uniform waveguide for solving the Schrödinger's equation.

Shih and others [15] have shown the viability of WKB approximation at low biases. Tunneling rate calculations invoking the WKB approximation and the numerical solution to the effective-mass Schrodinger equation have been independently carried out in a post-processing fashion following a self-consistent Schrödinger-Poisson solver. The direct-tunneling currents calculated by these two distinct approaches have been compared with oxide thickness ranging from 1.5 to 4 nm. Their findings suggest that the WKB approximation, in spite of its simplicity, remains a viable approach for predicting the gate leakage current in the direct tunneling regime. However, in the Fowler-Nordheim (FN) tunneling regime, the WKB approximation fails to capture the interference effects near the SiO$_2$-metal interface, and the results obtained from the fully numerical approach are more consistent with experimental findings.

A modified WKB approximation has been proposed by Register et *al.* [16]. It modifies the usual tunneling probability predicted by WKB, by accounting the reflections

from potential discontinuity. In this model, the barrier height to tunneling is taken to be a function only of the total electron energy and the Si bandgap dispersion relation is modeled as a two band Franz-type.

Using the QM calculations and a modified WKB method, Yang *et al.*[17] have demonstrated an efficient model for direct tunneling current for ultra thin gate oxides. They have used it in conjunction with the C-V measurements as a complementary characterization technique for estimating physical oxide thickness. The combined C-V and tunneling current technique has been shown to provide an efficient and reasonably accurate method for characterizing ultra-thin oxides in the direct tunneling regime. The doping effects on the tunneling currents have been studied in this work, where quantum mechanical calculations of the substrate have been used to obtain accurate information about the substrate inversion layers. It was shown that higher levels of substrate doping help to reduce the direct tunneling currents at the same applied gate voltage. However, for the self-consistent loop, they have neglected effects of wave function tail in oxide on electrostatic potential by assuming the barrier at the Si/SiO_2 interface to be infinite. As the tunneling current across ultra thin silicon dioxide becomes too large for devices to function, they have suggested that alternative dielectric materials (such as high-K materials) will be required for the replacement of silicon dioxide

Recently Ghetti and others [18] have investigated different components of gate current using self consistent potential profile. They have calculated transmission probability through an exact solution of Schrödinger equation. Effective mass approximation was used in terms of Airy's function following the transfer matrix method.

Eric M. Vogel and others [19] modeled the tunneling currents for insulators with an effective oxide thickness of 2.0 nm using a numerical calculation of the WKB tunneling current. Their model was shown to agree with previously determined analytical WKB formulations of tunneling current for SiO_2. Their tunneling model was applied to alternative dielectrics having different barrier heights and dielectric constants. Their results indicated that alternative dielectrics with higher dielectric constants resulted in lower currents at low biases. They also mentioned that if an alternative high dielectric constant material is to replace SiO_2, then it will be necessary to find one with a barrier height greater than the applied voltage and one that can he fabricated with a few atomic layers (or less) of SiO_2 at the interface. This may prove to be a difficult challenge for future IC manufacturing.

In search of suitable high-K material, Yeo and others [20] calculated direct tunneling gate currents through Si_3N_4 in both n-and pMOSFETs using an analytical model. Agreement between model and experimental data was shown. Important differences between Si_3N_4 and SiO_2 parameters were highlighted. It was projected that the equivalent oxide thickness of Si_3N_4 gate dielectric fabricated by jet-vapor deposition (JVD) technique can be scaled down to 0.65 nm and 1.13 nm for high-performance and low-power applications, respectively before it is limited by excessive tunneling gate leakage current. In both the previous work nothing was told about the effect of penetration wheather on SiO_2 or high-K dielectrics.

Later J. Zhang and others [21] gave an effective analytical model to evaluate the leakage currents for different dielectric stacked gate deep sub micron MOSFETs. They showed that for a given equivalent oxide thickness of a stacked gate, the gate leakage current decreases with an increase of high-K dielectric thickness or a decrease of interlayer thickness. Turning points at high gate biases of the *I-V* curves are observed for

Si_3N_4/SiO_2, Ta_2O_5/SiO_2, $Ta_2O_5/SiO_{2-y}N_y$, Ta_2O_5/Si_3N_4 and TiO_2/SiO_2 stacked gates except for Al_2O_3/SiO_2 structure. Design optimization for the stacked gate architecture to obtain the minimum gate leakage current was evaluated.

Most recently Mudanai and others [22] calculated the direct tunneling current from an inverted p-substrate through different gate dielectrics by numerically solving Schrodinger's equation and allowing for wave-function penetration into the gate-dielectric stack. Their numerical solution adopted a first-order perturbation approach to calculate the lifetime of the quasi-bound states. This approach was been verified to be valid even for extremely thin dielectrics (0.5 nm). They showed that for the same effective oxide thickness (EOT) the direct tunneling current decreases with increasing dielectric constant, as expected. But they did not consider effect of wave function penetration on the potential profile. For the self-consistent loop, the Schrodinger's equation was solved assuming that the wave function goes to zero at the dielectric-silicon substrate boundary on one side and at some point deep in the substrate on the other side. Once they found the self-consistent potential, the Schrodinger's equation was again solved assuming that the wave function now penetrates the gate dielectric and gate electrode. Thus, the effect of the wave function penetration into the oxide on the electrostatic potential has been ignored in this work and tunneling current was calculated in post processing fashion.

The inversion layer capacitance (C_{inv}) inherent to MOS structure, plays a significant role in the gate capacitance of ultra-thin MOSFETs. The total gate capacitance (C_g) is the series combination of the oxide capacitance (C_{ox}) and C_{inv}. Since device scaling rule demands thinner gate oxide and the resultant larger C_{ox}, the degradation of C_g due to C_{inv} becomes larger with the shrinkage of devices. As a result transconductance of MOSFET's which is determined by C_g, is reduced. There are two physical origins of C_{inv}, one origin is the finite band bending required to increase inversion charge density (N_S), due to finite effective density-of-states in the band and the other due to the finite inversion layer thickness.

In the paper of Takagi et al. [23] the inversion-layer capacitance in n-channel MOSFET's was studied experimentally and theoretically with emphasis on the surface carrier concentration dependence of C_{inv} which is important in the quantitative description of the inversion layer capacitance. Based on the experimental N_S and temperature dependencies, the physical origin C_{inv} was discussed. It was shown that, at lower N_S, the C_{inv} is determined by the finite effective density of states, while, at higher N_S, C_{inv} is determined quantum mechanically by the finite inversion-layer thickness. Also, the results of the surface orientation dependence of C_{inv} were presented as the first direct evidence for the fact that surface quantization plays a significant role in C_{inv} even at room temperature. They also studied the effect of substrate doping and verified that C_{inv} increases with substrate doping. But their work did not take account of inversion electron penetration into the gate dielectric.

Later Takagi et al. [24] compared the amount of C_{inv} and its influence on the gate capacitance between electron and hole inversion layers. It was experimentally verified that, under same physical thickness of gate oxides, the electrical gate oxide thickness, determined from the gate capacitance, is larger for inversion layer holes than that for inversion-layer electrons, because of smaller values of C_{inv} for inversion-layer holes.

In another paper of Takagi et al. [25] the influence of inversion-layer capacitance on supply voltage (V_{dd}) of n- and p-MOSFET's was quantitatively examined. The

physical origin of the effect of C_{inv} on V_{dd} consists in the band bending of a Si substrate in the inversion condition due to C_{inv}, which is not scaled with a reduction in gate oxide thickness. The amount and the impact of the band bending is accurately evaluated on a basis of one-dimensional (1-D) self-consistent calculations including two-dimensional (2-D) subband structure of inversion electrons and holes.

Jallepalli et al.[26] have presented the effects of hole quantization on p-MOS device characteristics and improvement on earlier results for electron quantization. A compact analytical model has been provided to quantitatively describe the threshold voltage shifts due to quantum effects as a function of the doping concentration and oxide thickness. The temperature dependence of the threshold shifts was also illustrated. The increase in the "effective" oxide thickness due to quantization has been presented as a function of the oxide field for a range of doping concentrations. To describe the quantization effects in a simple fashion, constant quantization and density of states effective-mass have also been presented. This simplified description has then been used to study the significance of quantization for buried channel p-MOS structures.

Ritcher, Hefner and Vogel have very recently compared the results of an extensive ensemble of the most advanced available QM capacitance-voltage simulators and analysis packages for a range of MOS device parameters [27]. They found that, for different simulators, in the accumulation capacitance region, extracted equivalent oxide thickness (EOT) shows on the order of 0.2 nm variations for total SiO_2 thickness in the range of 1.0 to 3.0 nm. Their study emphasizes on the fact that, when reporting experimentally derived electrical thickness results, it is essential to describe fully how these values were obtained. The same experimental curve can lead to different extracted EOT depending upon which quantum mechanical software is used for the analysis.

In Ref. [28], an empirical model for the DC charge centroid in terms of the gate voltage, threshold voltage and oxide thickness was developed. This model developed a one dimensional simulator using a self-consistent method by solving Schrödinger's and Poisson's equations iteratively along with Fermi-Dirac Statistics. Employing this model, a universal expression for this DC charge centroid from the accumulation to depletion region was later derived by Liu et.al. [29]. They considered the finite charge layer thickness due to quantization effects in all operating regions and thereby proposed an accurate model for C-V characterization.

Fiegna and Abramo have analyzed the QM effects on gate capacitance, threshold voltages and effective mobility of electrons in nonuniformly doped MOS structures [30]. Their results show that, with the introduction of a low doped region at the device surface, it is possible to manipulate the threshold voltage according to the circuit application. Also for a given charge sheet density, the introduction of low doped epitaxial region produces a reduction of the electron effective field, resulting in the improvement of electron effective mobility. But it leaves the total gate capacitance unaffected.

In the study of QM effects in MOS inversion layer, conventionally the wavefunction penetration inside oxide region is neglected. Since present MOS devices approach the deep submicron dimensions, this boundary condition is no longer justified. In ref [22] though wavefunction penetration was allowed but they did not consider wavefunction penetration within self-consistent loop. In ref [31] a new boundary condition encountering wavefunction penetration was introduced. In that work both conventional and the proposed boundary condition were employed in calculation of N_S, DC charge centroid shift from the oxide semiconductor interface (Z_{avg}) and effective gate

oxide capacitance (C_{EFF}). A comparative study of the effects of these boundary conditions made in this work suggested that the choice of boundary condition for deep submicron devices is significant and the choice becomes increasingly more important as the devices are scaled down.

Most recently, Mudanai et al. [32] have performed a comprehensive analysis of the effects of wavefunction penetration on the capacitance of NMOS capacitors for the first time, using a self-consistent Schrödinger-Poisson solver. Their study revealed that accounting for wave function penetration into the gate dielectric causes carrier profile to be shifted closer to the gate dielectric reducing the electrical oxide thickness. They observed that this shift increases with increasing gate voltage and results in an increased gate capacitance. In this work the increase in capacitance observed in the inversion region was relatively insignificant when a poly gate electrode with a doping of less than $1 \, x \, 10^{20}$ cm^{-3} was used due to the poly-depletion effect. A physical picture of the effect of physical thickness on the tunneling current was also presented.

1.2 Objective of the work

We already know in deep sub-micron MOSFET, quantum effects of inversion carriers become prominent which has two important adverse effects 1) a high gate current flow due to the direct tunneling (DT) of inversion carriers and 2) inversion-layer capacitance. It has been suggested that in order to reduce direct tunneling current, the gate dielectric in devices with sub-2 nm should be a high-K dielectric instead of SiO_2. The thickness of high-K dielectric should be such that the physical thickness must increase, while equivalent oxide thickness must continue to scale down. A number of studies have been reported in the literature on the modeling of gate capacitance, which is the series combination of the oxide capacitance and inversion layer capacitance, of MOS structures [23-27] and on the modeling of direct tunneling gate current [13-21]. The electrostatic potential in silicon near silicon-gate-oxide interface is typically determined from the self-consistent solution of Schrödinger's and Poisson's equations. Closed boundary conditions commonly used for the solution of Schrödinger's equation are that the wave function goes to zero at silicon-gate-oxide interface and at some point deep inside the bulk [9]. However, due to the finite potential barrier height, some penetration of the wave function into the gate-oxide occurs [22,31,32]. This penetration is actually responsible for the DT gate current. Because of the computational involvement associated with the common solution techniques of Schrödinger's equation with open boundary conditions, in many studies, the potential profile is determined self-consistently neglecting the DT of the carriers, and the tunneling current[13-21] and gate capacitance[23-27] is calculated in a postprocessor outside the self-consistent loop. Effect of wave function penetration into the gate dielectric was considered during calculation of the gate current and gate capacitance in ref [22] and ref [32] respectively. But effect of DT of inversion electron on potential profile was neglected in these works as mentioned in review section. Also a comparison between closed and open boundary models for gate current calculation has been performed in Ref. [33], but in this study too, while using open boundary conditions, the effects of tunneling within the self-consistent loop are neglected. It is already known that the effects of carrier tunneling on potential profile is non-trivial [34, 35]. This raises questions on the accuracy of calculation of the tunneling current in a post-processor.

Providing high-K dielectric material as gate insulator is taken as alternate solution of the problem of high off state leakage current. The fabrication rule dictated by ITRS roadmap is that the width of high-K material should be such that would provide same gate capacitance of their SiO_2 counterpart. As general trend of high-K dielectric is to decrease of conduction band offset with the increase of dielectric constant strong trade off exists various alternate dielectrics. We will investigate the validity of the assumption of neglecting carrier tunneling effects on self-consistent electrostatic potential in calculating direct tunneling gate current and gate capacitance in deep sub-micron MOSFETs with both SiO_2 and high-K dielectrics as the gate insulator. Comparison between simulated and experimental results will be shown. A numerically efficient technique, proposed recently [36], will be applied in the calculation of the DT current considering tunneling within the self-consistent loop. The relative error in gate current and gate capacitance due to neglect of carrier tunneling will be calculated to find out penetration pattern with gate voltages and with oxide thickness. The results will shed light on the accurate modeling of direct tunneling current and gate capacitance of MOSFETs with high-K dielectric materials used as gate insulator. We will also study the substrate doping effect on the modeling of gate capacitance considering wave function penetration in the gate dielectric.

1.3   Organization of the Thesis

Chapter 2 reviews quantum effect in MOS devices and discusses transmission line analogy and Green's function formalism. Chapter 3 deals with the available high-k materials, their properties, interface quality etc. Chapter 4 narrates calculation of gate leakage current from density of states and results of direct tunneling current for various gate dielectrics. Chapter 5 narrates results of gate capacitance and inversion capacitance using high-k gate dielectric materials. Concluding remarks of this work will be presented in chapter 6.

Chapter 2

Quantum Mechanical Analysis of MOS Devices

Quantum mechanical effects in MOSFETs arise when the quantization of the inversion layer carriers occur due to the presence of a high electric field (F_S) at the surface of the semiconductor. In this chapter the theory of the quantum mechanical analysis of MOS structures are discussed. The Structure of an n channel enhancement type MOS is shown in the Fig. 2.1. It is doped p type and the gate is in contact with positively biased metal electrode. The positive potential of the metal causes the depletion of holes at the surface. If the bias is sufficiently strong, the bands will bent enough to make the conduction band come below the Fermi level, in which case an inversion layer of electrons forms near the interface. Even when the Fermi level is just below the conduction band edge, there will be a thermally induced inversion layer of electrons. The following analysis also applicable to an n-type semiconductor with a negatively biased insulator, in which case holes collect near the interface. In that case it is assumed that holes behave like particles.

2.1   Self-consistent Solution

The self consistent solution of coupled Schrödinger's and Poisson's equations as proposed by Stern [9] is presented in this section which is based on three major approximations. It is assumed 1) that the effective mass approximation is valid, so that the periodic potential can be neglected. The effective masses and the dielectric constant of the perfect crystal can be used. 2) that the envelope wavefunction vanishes at the surface. Neither approximation is likely to be valid at high electric field. The third major
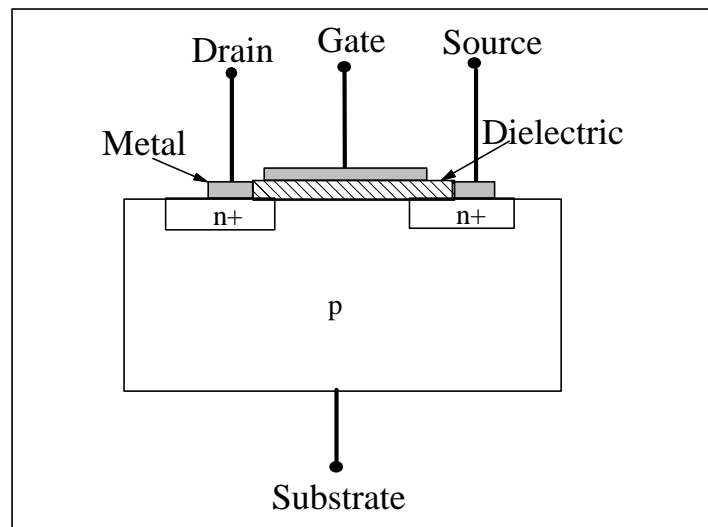

Fig 2.1: A typical nMOS structure.

9

Fig. 2.2: A typical conduction band profile of an NMOS structure.

approximation 3) is that surface states can be neglected and the effect of charges in the oxide or insulator adjacent to semiconductor can be replaced by an equivalent electric field. A typical conduction band profile for an nMOS structure is shown in Fig. 2.2

The band bending of a semiconductor can be characterized by a potential $f(z)$. In the effective mass approximation, the electronic wavefunction for the ith subband is the product of the Bloch function at the bottom of the conduction band and envelope function. Envelope function $V_i(z)$ is the solution of

$$\frac{d^2 V_i}{dz^2} + \frac{2m_3}{\hbar^2}[E_i + ef(z)]V_i(z) = 0 \tag{2.1}$$

Here m_3 is the effective mass in the direction perpendicular to the interface and E_i is the energy of the ith bound state in the same direction. Boundary conditions commonly used for the solution of Eq. (2.1) are $V_i(\infty) = 0$ and at the semiconductor-oxide interface, $z_i(0) = 0$. Each solution of Eq. (2.1) gives the bottom of a continuum of levels called a subband. There can be as many as three values of m_3 depending on the surface orientation because the conduction band of Si has six ellipsoid valleys along the (100) family of direction. In the effective mass approximation, the valleys are degenerate in pairs. Solution of Eq. (2.1) gives the eigenenergy E_i and the envelope function $z_i(z)$.

The potential $f(z)$ is found from the solution of Poisson's equation, which is as follows.

$$\frac{d^2 f(z)}{dz^2} = -[r_{depl}(z) - e\sum_i N_i z_i^2(z)]/e_{si}e_o \tag{2.2}$$

here, e_{si} is the dielectric constant of the semiconductor, N_i is the carrier concentration in the ith subband .N_i is given by the following equation,

$$N_i = \frac{n_{vi}m_{di}kT}{p\hbar^2} In\left[1 + \exp\left(\frac{E_F - E_i}{kT}\right)\right] \tag{2.3}$$

10
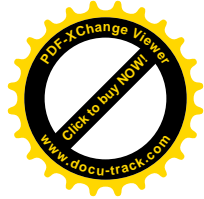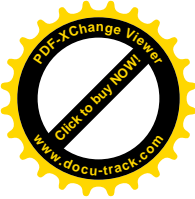
where, n_{vi} is the valley degeneracy of ith valley, m_{di} is the density of states effective mass per valley and E_F is the Fermi energy.

$r_{depl}(z)$ is the charge density in the depletion layer, which is taken to be.

$$r_{depl}(z) = -e(N_A - N_D), \qquad 0 < z < z_d$$
$$r_{depl}(z) = 0, \qquad\qquad z > z_d \tag{2.4}$$

here, z_d is the depletion layer thickness given by [9] as following,

$$z_d = \sqrt{\frac{2e_{si}e_o f_d}{e(N_A - N_D)}} \tag{2.5}$$

f_d is the band bending due to depletion charge only. f_d can be calculated from [4] as following,

$$f_d = f_s - \frac{kT}{e} - \frac{eN_{inv}Z_{av}}{e_{si}e_O} \tag{2.6}$$

Where $N_{inv} = \sum_i N_i$ is the total number of charge per unit area in the inversion layer and z_{av} is the average penetration of inversion charge density into Si. The two boundary conditions for solution of Eq. (2.2) are df/dz=0 for large z and at the surface, its value is F_s. From [9], F_5 is given by,

$$F_S = \frac{e(N_{inv} + N_{depl})}{e_{si}e_o} \tag{2.7}$$

$N_{depl} = z_d(N_A - N_D)$ is the number of charge per unit area in the depletion layer. In a self-consistent formulation, Eqs. (2.1)-(2.6) are solved iteratively for a given F_S until results converge.

2.2 Transmission Line Analogy

After Stern's publication [9] it is evident that QM effects can no longer be neglected to make a balance between the theoretical model and the experimental results with the increasing trend of scaling down the MOS feature size. To accurately model the QM effects, it is a must to solve Schrödinger's equation. But this generally requires lengthy matrix manipulation. Hence, some assumptions are made to reduce the complicacy of the problem. Various researchers worked on to derive a simple solver to reduce the computational hazards in solving the QM approach. Of those, some are mentioned in chapter 1. In this chapter, a quite simple approach is presented. This approach first uses the well-known transmission line concept used in microwave engineering. Realizing the power of this approach to solve QM problems, later Green's function is introduced to calculate eigenenergies and wave functions. When these are known, other quantities of interest can be calculated easily.

11

Transmission line analogy to solve Schrödinger's equation is described briefly as follows [42]. The well known equations for voltage (V) and current (I) used in transmission line theory are, with time variation assumed as e^{-jwt} instead of usual e^{jwt}

$$I(z) = I^+(e^{g_r z} - \Gamma_t e^{-g_r z})$$
$$V(z) = I^+ Z_o (e^{g_r z} + \Gamma_t e^{-g_r z})$$

(2.8)

where, g_r is the propagation constant and G_t is the wave amplitude reflection coefficient given by

$$\Gamma_t = \frac{Z_{lt} - Z_{ot}}{Z_{lt} + Z_{ot}}$$

(2.9)

where Z_{it} and Z_{ot} are the load and the characteristics impedance of the transmission line, respectively.

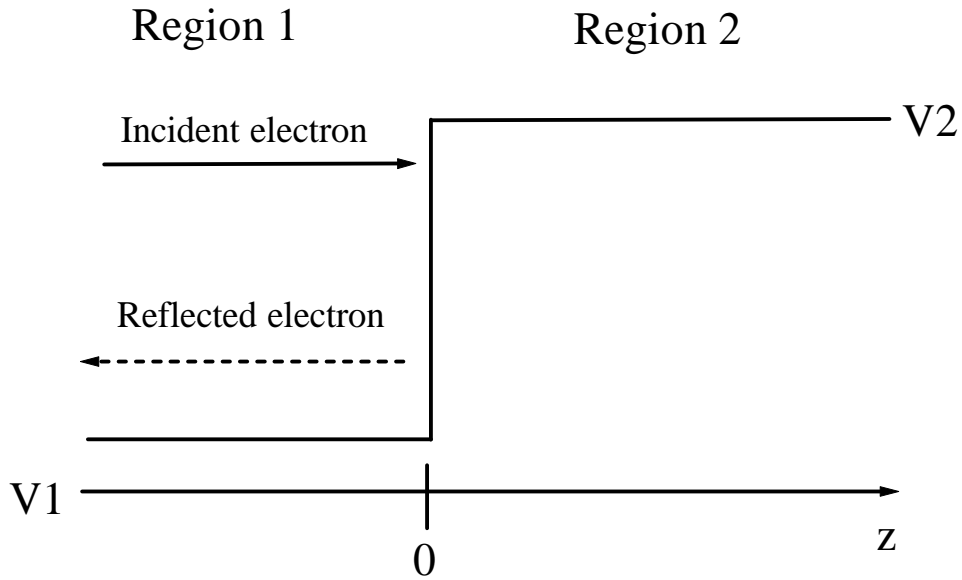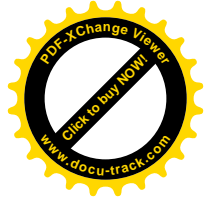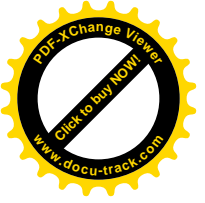Fig. 2.3: A potential barrier with an incident electron (solid lines) and the reflected wave (dashed line).

If an electron with an energy E is incident on the potential barrier, then the corresponding wave function can be expressed as

$$y(z) = A^+(e^{gz} - r e^{gz})$$

(2.10)

where time-variation is implicitly assumed as $e^{-iEt/\hbar}$ and

$$g = a + ib = i\sqrt{\frac{2m^*}{\hbar^2}(E-V)}$$

(2.11)

12

is the propagation constant, m^* is the effective mass, V is the potential and $\mathbf{r}$ is the wave amplitude reflection coefficient. For the two regions shown in figure 2.3 Eq. 2.10 is rewritten as

$$\mathbf{y}_1(z) = A_1^+(e^{\mathbf{g}_1 z} - \mathbf{r}e^{-\mathbf{g}_2 z}) \qquad z < 0$$

$$\mathbf{y}_2(z) = A_2^+ e^{\mathbf{g}_2 z} \qquad\qquad z > 0$$

(2.12)

Here $\mathbf{g}_j = \mathbf{a}_j + i\mathbf{b}_j = i\sqrt{\dfrac{2m^*}{\hbar^2}(E - V_j)}$, m_j and V_j $(j = 1,2)$, are the effective mass and the potential, respectively, for the jth region. Since, for $z > 0$ the region is of infinite extent, it is evident that there is no reflection. The boundary conditions for this problem, with regard to the continuity of wave function, are

$$\mathbf{y}_1(0) = \mathbf{y}_2(0)$$
$$\frac{\mathbf{y}_1'}{m_1^*} = \frac{\mathbf{y}_2'}{m_2^*}$$

(2.13)

Using these boundary conditions an expression for $\mathbf{r}$ is found as

$$\mathbf{r} = \frac{\dfrac{\mathbf{g}_2}{m_2^*} - \dfrac{\mathbf{g}_1}{m_1^*}}{\dfrac{\mathbf{g}_1}{m_1^*} + \dfrac{\mathbf{g}_2}{m_2^*}}$$

(2.14)

If a function, $\mathbf{y}\ddot{a}(z)$ is defined as

$$\mathbf{y}^{TM}(z) = \frac{2\hbar}{im^*}\frac{d\mathbf{y}}{dx}$$

(2.15)

then,

$$\mathbf{y}^{TM}(z) = A^+ Z_o(e^{\mathbf{g}z} + \mathbf{r}e^{-\mathbf{g}z})$$

(2.16)

where,

$$Z_o = \frac{2\mathbf{g}\hbar}{im^*}$$

(2.17)

Using the value of Z_o in Eq. 2.14, it is found that

$$\mathbf{r} = \frac{Z_{o,2} - Z_{o,1}}{Z_{o,2} + Z_{o,1}}$$

(2.18)

where $Z_{0,1}$ and $Z_{0,1}$ are defined for region 1 and region 2, respectively.

From the above discussion, an analogy is found between voltage and current expressed by Eqs. 2.8 with $\mathbf{y}\ddot{a}$ and $\mathbf{y}$ expressed by Eqs. 2.16 and 2.10, respectively. Since, the transmission line impedance is the ratio of V and I, so the ratio of $\mathbf{y}\ddot{a}$ and $\mathbf{y}$

13

is called as wave impedance by using this analogy. This fact can be verified by using another analogy found between the equation of wave amplitude reflection coefficient for electron wave (**r**) and the equation of wave amplitude reflection coefficient for transmission line (G_t), as seen in Eqs 2.18 and 2.9, with Z_{lt} replaced by $Z_{0,2}$ and Z_{ot} replaced by $Z_{o,1}$ Moreover, **y̎** and **y** are continuous across the boundary between the two regions, which directly corresponds to the continuity conditions for voltage and current at the junction between two transmission lines. All these lead to a conclusion that a quantum mechanical wave impedance concept can be introduced analogous to transmission line impedance as given by at any plane z

$$Z(z) = \frac{\mathbf{y}^{TM}(z)}{\mathbf{y}(z)} \tag{2.19}$$

Thus, transmission line concept can be applied for QM calculations. For example, the input impedance, Z_i at $z = -l$ may be expressed in terms of load impedance, Z_l at $z = 0$ as in transmission line

$$Z_i = Z_o \frac{Z_l \cosh(\mathbf{g}l) - Z_o \sinh(\mathbf{g}l)}{Z_o \cosh(\mathbf{g}l) - Z_l \sinh(\mathbf{g}l)} \tag{2.20}$$

An important feature of using transmission line concept in QM calculations is its ability to calculate the eigenenergy of any arbitrary potential well. Becasue, from Eq. 2.19, at any eigenenergy, the wave impedances looking to the right (positive direction) and to the left (negative direction), at any plane z must be equal, i.e.

$$Z_{iR} = Z_{iL} \tag{2.21}$$

2.3   Green's Function Formalism

Green's function formalism is used along with the transmission line analogy to include the effects of energy broadening of the density-of-states due to either finite particle lifetime in a well (i.e. when particles leak out from a quantum well) or the presence of inelastic scattering processes. Although this formalism is based on a complicated Green's function approach, it allows one to calculate easily the normalized wave functions in arbitrary 1-D quantum well structures [36].

 In this formalism the QM wave impedance is redefined in terms of the logarithmic derivative of the retarded Green's function, G^R as

$$Z(z, z'; E) = \frac{2\hbar}{im_z^*} \left[\frac{\partial G^R(z, z'; E)}{\partial z} / G^R(z, z'; E) \right] \tag{2.22}$$

where G^R satisfies the equation,

$$\left[E + \frac{\hbar^2}{2m_z^*}\frac{\partial^2}{\partial z^2} - V(z) + i\boldsymbol{e}\right]G^R(z,z';E) = \boldsymbol{d}(z-z') \tag{2.23}$$

where $\boldsymbol{e}$ is an infinitesimally small positive energy. The Green's function in this context is discussed in detail in Ref. [37]. Owing to the property of G^R, $Z(z, z'; E)$ has a discontinuity at $z = z'$, and one needs two boundary conditions to determine $Z(z, z'; E)$. To obtain these boundary conditions, the potential profile is assumed flat sufficiently far from $z = z'$ in both directions. If V_R is the constant potential at $z = \infty$ and if V_L is the constant potential at $z = -\infty$ the Green's function in these regions can be expressed as [36]:

$$G^R(z \to \infty, z'; E) \approx e^{\boldsymbol{g}R(z-z')}$$
$$G^R(z \to -\infty, z'; E) \approx e^{-\boldsymbol{g}L(z-z')} \tag{2.24}$$

where

$$\boldsymbol{g}_{R(L)} = i\sqrt{\frac{2m_z^*}{\hbar^2}(E - V_{R(L)})} \tag{2.25}$$

$\boldsymbol{g}_{R(L)}$ is imaginary if $E > V_{R(L)}$, else this is real. From the above relationships, the boundary conditions are found as

$$Z(z \to \infty, z'; E) = Z_o(\infty)$$
$$Z(z \to -\infty, z'; E) = -Z_o(-\infty) \tag{2.26}$$

where

$$Z_o(\pm\infty) = \frac{2\hbar}{im_z^*}\boldsymbol{g}_{R(L)} \tag{2.27}$$

From the properties of the Green's functions it can be shown that [37]

$$Z(z, z'; E) = Z_{iR}(z; E) \text{ for all } z' < z$$
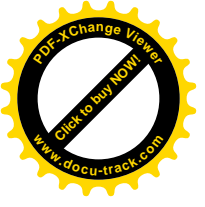$$Z(z, z'; E) = Z_{iL}(z; E) \text{ for all } z' > z \tag{2.28}$$

It is noteworthy that Z_{iR} (Z_{iL}) does not depend on z' as long as $z > z'$ ($z < z'$). Using transmission line analogy, Z_{iR} (Z_{iL}) can be calculated. The eigenenergies of an arbitrary quantum well can be determined using following condition:

$$Z_{iR}(z; E) = Z_{iL}(z; E) \tag{2.29}$$

as from tranmission line analogy, at any eigenenergy and for all values of z inside the quantum well the above equation must be satisfied. Once an eigenenergy is found, the corresponding normalized wavefunction can be calculated using the following relationship:

$$\left|\boldsymbol{j}_n(z)\right|^2 = \frac{4\boldsymbol{e}}{\hbar}\Im\left[\frac{i}{Z_{iR}(z;E_n) - Z_{iL}(z;E_n)}\right] \tag{2.30}$$

By definition, the one dimensional (1D) DOS is given by the imaginary part of the diagonal term of 1D retarded Green's function G^R,

15

$$N(z;E) = -\frac{1}{\boldsymbol{p}} \Im m \left[G^R(z,z;E) \right] \tag{2.31}$$

Relating diagonal part of G^R to quantum mechanical impedence, equation (2.31) reduces to,

$$N(z;E) = \frac{4}{\boldsymbol{p}\hbar} \Im m \left[\frac{j}{Z_{iR}(z;E_n) - Z_{iL}(z;E_n)} \right] \tag{2.32}$$

Equations 2.29 and 2.30 are used to find the eigenenergies and wavefunctions in MOS inversion layers. Note that in calculating the wavefunctions the conventional boundary conditions that the wave function is zero at the oxide-semiconductor interface and inside the bulk are not used. Rather, the boundary conditions used (Equation 2.26) assume that the potential profile is flat sufficiently far from the interface in both directions. These boundary conditions, known as the asymptotic boundary conditions, are definitely physically more acceptable boundary conditions and allow one to include the exponentially decreasing tail of the wave function inside the oxide in the calculation.
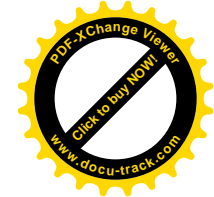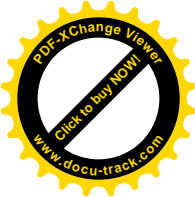
Chapter 3

Properties of high-K dielectrics

The key element enabling the scaling of the Si-based MOSFET is the material (and resultant electrical) properties associated with the dielectric employed to isolate the transistor gate from the Si channel in CMOS devices. For decades, silicon dioxide has been the material of choice. The use of amorphous, thermally grown SiO_2, as a gate dielectric offers several key advantages in CMOS processing including a stable (thermodynamically and electrically), high-quality Si-SiO_2 interface as well as superior electrical isolation properties. These outstanding electrical properties clearly present a significant challenge for any alternative gate dielectric candidate. Many dielectrics appear favorable and show the most promise toward successful integration into the expected processing conditions for future CMOS technologies, especially due to their tendency to form at interfaces with Si (e.g. silicates). A list of these available dielectrics and material properties is shown in Table 3.1 [38].

3.1 Properties of available dielectrics

In SiO_2 the density of states transition from the substrate into the SiO_2 layer indicates that the full band gap of SiO_2 is obtained after only about two mono layers of SiO_2. Within two mono layers of the Si channel interface, oxygen atoms do not have the full arrangement of oxygen neighbors and therefore cannot form the full band gap that exists within the "bulk" of the SiO_2 film. These results set an absolute physical thickness limit of SiO_2 to 0.7 nm. Below this thickness, the Si rich interfacial regions from the channel and polycrystalline Si gate interfaces used in MOSFETs overlap, causing an effective "short" through the dielectric, rendering it useless as an insulator. And it is already mentioned that in sub-2 nm dielectric thickness leakage current is very high. Therefore below 2 nm regime high-K dielectrics are favorable.

In addition to leakage current increasing with scaled oxide thickness, the issue of boron penetration through the oxide is a significant concern. The large boron concentration gradient between the heavily doped poly-Si gate electrode, the undoped oxide and lightly doped Si channel causes boron to diffuse rapidly through a sub-2 nm oxide upon thermal annealing, which results in a higher concentration of boron in the channel region. A change in channel doping then causes a shift in threshold voltage, which clearly alters the intended device properties in an unacceptable way.

The concerns regarding high leakage currents and boron penetration have led to materials structures such as oxy-nitrides and oxide/nitride stacks for near-term gate dielectric alternatives. These structures provide a slightly higher K value than SiO_2 (pure Si_3N_4 has K = 7) for reduced leakage (since the film is physically slightly thicker), reduced boron penetration and better reliability characteristics. The addition of N_2 to SiO_2
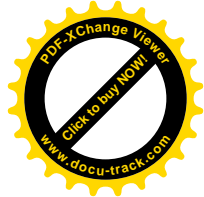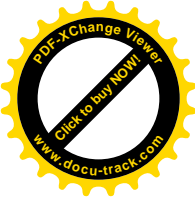
TABLE 3.1. Comparison of relevant properties of high-K dielectrics.

Material	Dielectric constant (k)	Band gap (E_G)	ΔE_C to Si
SiO_2	3.9	8.9	3.2
Si_3N_4	7	5.1	2
Al_2O_3	9	8.7	2.8
Y_2O_3	15	5.6	2.3
La_2O_3	30	4.3	2.3
Ta_2O_5	26	4.5	1-1.5
TiO_2	80	3.5	1.2
HfO_2	25	5.7	1.5
ZrO_2	25	7.8	1.4

greatly reduces boron diffusion through the dielectric, and has been shown to result from the particular Si-O-N network bonding formed in silicon nitride and oxy-nitride. Furthermore, small amounts of N_2 (0.1 at. %) at or near the Si channel interface have been shown to improve device performance.

Scaling with oxy-nitrides/nitrides appears are limited to 1.3 nm. Below this, the effects of gate leakage, reliability or electron channel mobility degradation will most likely prevent further improvements in device performance. According to the most recent industry roadmaps, SiO_xN_y, and Si_xN_y /SiO_2 dielectrics represent current three year near-term solutions for scaling the CMOS transistor.

The most commonly studied high-K gate dielectric candidates have been materials systems such as Ta_2O_5, which have dielectric constants ranging from 10 to 80, DE_C=1-1.5 eV and have been employed mainly due to their maturity in memory capacitor applications. Interfacial reaction has been observed for the case of Ta_2O_5 on Si, therefore its high permittivity could not be fully utilized.

Alumina (Al_2O_3) with DE_C=2.3-2.8 eV is a very stable and robust material, and has been extensively studied for many applications. Al_2O_3 has many favorable properties, as shown in Table 3.1, including a high band gap, thermodynamic stability on Si up to high temperatures, and is amorphous under the conditions of interest. The drawback is that

A1$_2$O$_3$ only has K = 8-10, and would therefore make it a relatively short-term solution for industry's needs (1-2 generations).
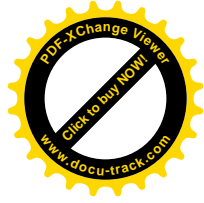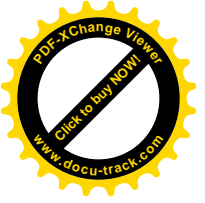
Y$_2$O$_3$ and La$_2$O$_3$ with DE_C=2.3 eV, are also studied for the purposes of high-K gate dielectrics. The dielectric constant of the Y$_2$O$_3$ grown on SiO$_2$ was found to be 17 to 20 but for Y$_2$O$_3$ grown directly on Si, it was found that K= 12. This 1ower measured permittivity value may have resulted from growth of interfacial SiO$_2$, during the thermal oxidation step. La$_2$O$_3$ has a high dielectric constant 30. Pure La is well known to be very volatile and reactive in air, and La$_2$O$_3$ will absorb water vapor from air. These characteristics signify that any *ex situ* exposure of these films to air will certainly result in an uncontrolled reaction

A substantial amount of investigation has gone into TiO$_2$ (DE_C = 1.2eV), ZrO$_2$ (DE_C = 1.4eV) and HfO$_2$ (DE_C = 1.5eV), as these systems have shown much promise in overall materials properties as candidates to replace SiO$_2$. The TiO$_2$, system has been heavily studied for high-K applications both for memory capacitors and in transistors. It is attractive because it has a high permittivity of K=80-110, depending on the crystal structure and method of deposition. This anomalously high permittivity, which arises through a strong contribution from soft phonons involving Ti ions, is not exhibited by the other metal oxides.

HfO$_2$ films in A1/HfO$_2$/Si structures show K= 22-25 and ZrO$_2$ films in Al/Zr0$_2$ /Si structures show K= 17-18. Pure ZrO$_2$ anf HfO$_2$ show mobile ion transport and crystallization of the metal oxides either during the deposition process or upon moderate post-annealing conditions. The interface quality of these systems remains a critical issue, however, since the materials in particular are extremely susceptible to O$_2$ diffusion and reaction at the channel interface. It was observed that processing in inert ambient reduces mobile ion transport. This indicates the possible application of these materials as high-K candidate.

To select a gate dielectric with a higher permittivity than that of SiO$_2$, the required permittivity must be balanced, however, against the barrier height for the tunneling process. For electrons traveling from the Si substrate to the gate, this is the conduction band offset, for electrons traveling from the gate to the Si substrate, this is work function of gate material. This is because leakage current increases exponentially with decreasing barrier height (and thickness) for electron direct tunneling transport. In order to obtain low leakage currents, it is desirable to find a gate dielectric that has a large DE_C value to Si . If the experimental DE_C values for various oxides are much less than 1.0 eV, it will likely preclude using these oxides in gate dielectric applications, since electron transport (either by thermal emission or tunneling) would lead to unacceptably high leakage currents.

In contrast to the general trend of increasing permittivity with increasing atomic number for a given cation in a metal oxide, the band gap E_G of the metal oxides tends to decrease with increasing atomic number, particularly within a particular group in the periodic table. In the cases of Ta$_2$O$_5$ and TiO$_2$, both materials have small E_G values and correspondingly small DE_C values. These small DE_C values directly correlate with high leakage currents for both materials, making pure Ta$_2$O$_5$ and TiO$_2$ unlikely choices for gate dielectrics. Table 3.1 also shows, however, that La$_2$O$_3$, HfO$_2$, and ZrO$_2$ offer relatively high values for both K and E_G.

Chapter 4

Modeling of Direct Tunneling gate current

Due to finite potential barrier height and width of gate dielectric material, inversion electrons can tunnel into the gate electrode from the inversion layer. As a consequence of this tunneling, a leakage current flows in the gate which increases standby power consumption. To restrict this current we have to replace SiO_2 gate dielectric by suitable high-K material in deep sub-micron MOSFET's as predicted by ITRS roadmap [2]. In such leaky quantum well, the lifetime of inversion carriers is finite and the 1D DOS broadens in energy around each eigenvalue. When open boundary conditions are used, the Hamiltonian for a finite system becomes non-Hermitian and the eigenenergies become complex. The real parts give the energies of the quasi-bound states and the imaginary parts are related to lifetime. Since the calculation of complex eigenenergies is numerically complicated, typically Schrodinger's and Poisson's equations are solved self consistently neglecting tunneling of carriers into the gate dielectric and gate current is calculated in a post processing fashion assuming effect of wave function penetration on potential profile to be negligible [13-17, 22]. Therefore, we adopt an alternate technique using the theory developed in chapter 2, and the direct tunneling current is calculated considering effects of wavefunction penetration self-consistent potential. The consequences of neglecting this effect are also investigated.

4.1 Theory of current calculation

In self-consistent solution, Eqs. 2.1 to 2.6 are solved iteratively for a fixed inversion electron density N_S with the help of Green's function. Wave functions are calculated from the logarithmic derivative of the retarded Green's function (Eq. 2.30).
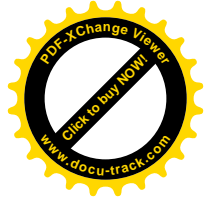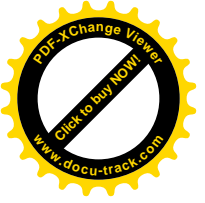
It may be mentioned that once self-consistent potential profile for any fixed inversion condition is obtained, gate voltage required to obtain the inversion condition can be calculated from the following equation:

$$V_G = \boldsymbol{f}_S + T_{OX} F_{OX} + \boldsymbol{f}_{ms} + \boldsymbol{f}_{poly} \tag{4.1}$$

Where $\boldsymbol{f}_S$ is the total band bending, $\boldsymbol{f}_{ms}$ the difference between the metal and semiconductor work functions, $\boldsymbol{j}_{poly}$ is the voltage drop in the poly-silicon depletion region. Here T_{OX} and F_{OX} are oxide width and electric field in the oxide respectively.

The Gate current due to quantum mechanical tunneling is calculated from the carrier concentration and the lifetimes of the carriers of all eigenstates using following relationship,

$$J = \sum_i \frac{qN_i}{\boldsymbol{t}_i} \tag{4.2}$$

Here, N_i is the concentration of the electrons in the ith eigenstate (Eq. 2.3), t_i is the lifetime of the electrons in the ith eigenstate and J is the total gate leakage current. The peaks of the energy broadened 1D DOS occur at the eigenenergies of the quasi-bound states and the lifetimes of the inversion layer electrons are related to the energy broadening of DOS according to,

$$\hat{0}_i = \frac{\hbar}{2\tilde{A}_i} \tag{4.3}$$

Here, G_i is the Full-Width at Half-Maximum (FWHM) of the energy broadening around the ith eigenenergy. The position dependent DOS as a function of energy around each eigenenergy is calculated using Eq. (2.31). The FWHM of the DOS can be evaluated anywhere inside the well since it has been shown that the energy broadening of the DOS is the same at all positions [39].

Inclusion of carrier tunneling within self consistent loop can easily be done by incorporating open boundary condition (Eq. 2.26) while calculating normalized wave functions during solving Schrodinger's equation. The problem of dealing with non Hermitian matrices during calculation of eigenenergies are avoided here as eigenenergies are found by searching peaks of DOS (Eq. 2.31). Once self-consistent potential profile is obtained considering DT effect on potential profile, gate leakage current can easily be determined from Eqn. (4.2) and Eqn. (4.3) by searching FWHM broadening of DOS. Values of conduction band offset used for various dielectrics are shown in the Table 3.1 of chapter 3. Figure 4.1(a) shows typical self-consistent potential profile found for inversion condition of $N_S = 10^{13}$ /cm^2 for 1.5 nm thickness of SiO$_2$ gate material with open boundary condition.

Closed boundary conditions are simulated by using theoretically infinite conduction band offset for electrons tunneling from silicon to gate dielectric. In practice, we have used a barrier height of 83.1 eV to represent closed boundary condition. Figure 4.1 (b) shows the potential profile found for inversion condition of $N_S=10^{13}$/cm^2 for 1.5 nm thickness of SiO$_2$ gate material using closed boundary conditions.

4.2 Result and Discussion

All simulations are performed for {100} Si substarte. For high-K dielectrics, Ta$_2$O$_5$, Si$_3$N$_4$ and Al$_2$O$_3$ are chosen. Values of conduction band offset and dielectric constant are taken from Ref. [38] as mentioned in Table 3.1. Table 4.1 shows values of effectve masses for {100} surface orientation which are taken from Ref. [9]. Effective mass of electrons in gate dielectric region are taken to be 0.50m$_o$ for all the dielectric materials [21]. All calculations are done for room temperature (T=300K). Substrate doping density $N_A=10^{18}$/cm^3 is used and n$^+$-ploysilicon with doping density equal to 10^{20}/cm^3 is used as the gate electrode.
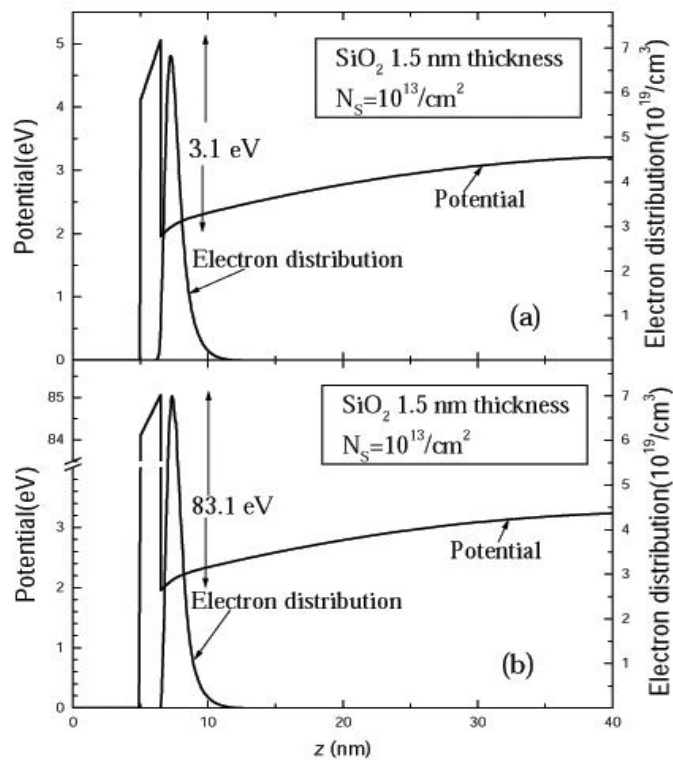
Fig. 4.1: Potential profile obtained self-consistently with SiO_2 gate dielectric of 1.5 nm thickness at inversion charge density of $N_S = 10^{13}/cm^2$ a) open boundary conditions b) closed boundary conditions.

Table 4.1: Parameters of {100} orientation.

Surface	{100}	
Valleys	Lower	Higher
Degeneracy	2	4
Normal mass	0.916	0.190
Density-of-states mass per valley	0.190	0.417

All effective masses are in unit of free electron mass.

22

Figure 4.2 shows modeled direct tunneling currents from the inversion layer for oxide thickness between 1.4 nm to 2 nm as a function of applied gate voltage both considering and without considering carrier tunneling effects on electrostatic potential. Results are also compared with experimental data obtained from Ref. [17]. From Figure 4.2 one can see that excellent agreement between measured and modeled currents are found when we consider carrier tunneling within the self-consistent loop. It is evident that neglect of wavefunction penetration causes under estimation of direct tunneling current. For these devices, good agreement between the modeled and experimental direct tunneling currents is achieved for gate voltages greater than 0.35V. Below this value of gate voltage, experimental current is much higher than the simulated current (results not shown). The discrepancy below 0.35 V is shown to be due to inelastic trap scattering of electrons in gate oxide region [41].
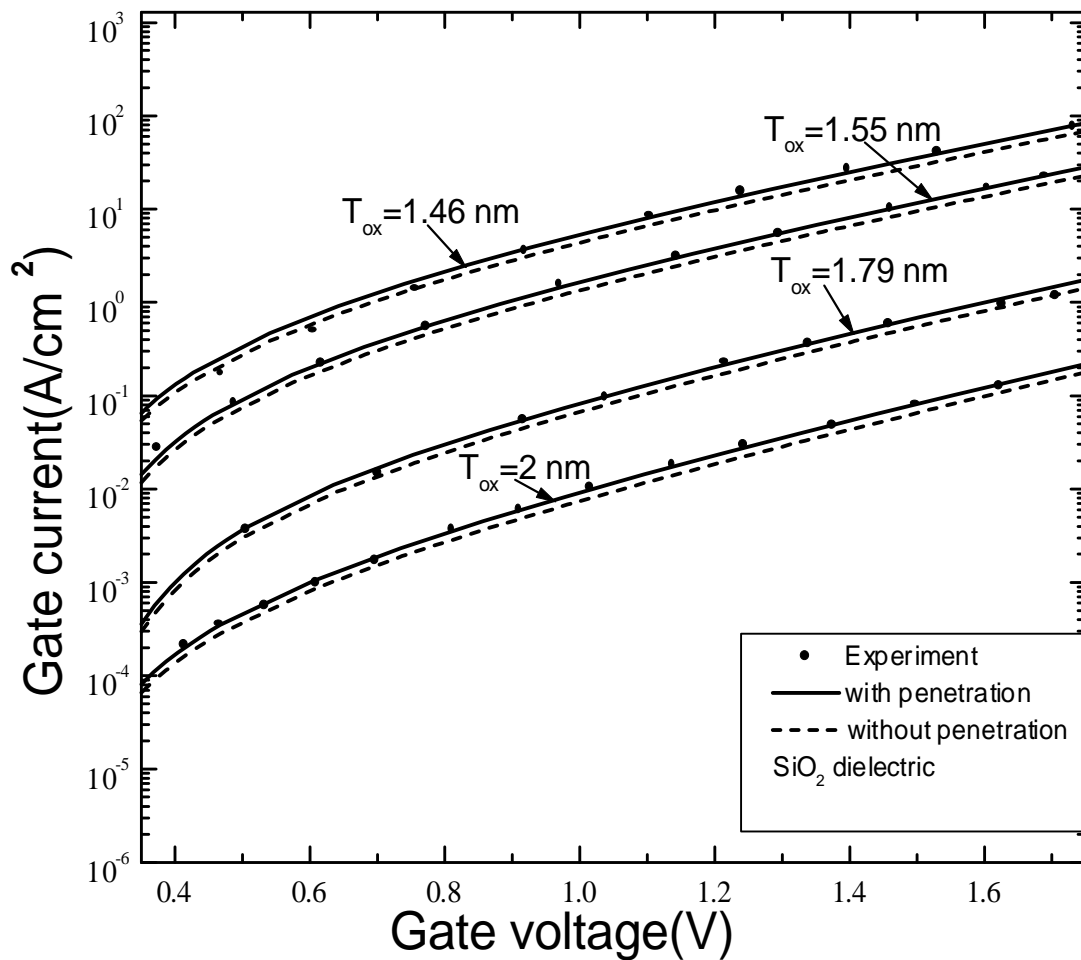

Fig. 4.2: Simulated and experimental direct tunneling currents in nMOSFET's both considering and without considering carrier tunneling effects within the self-consistent loop. Experimental data are taken from Ref. [17]. The poly-silicon doping level is $N_{poly}=10^{20}/cm^3$.

Fig. 4.3 shows the relative error in direct tunneling gate current as a function of gate voltage. We see that the relative error increases with the increase of gate voltage for a given oxide thickness and decreases with the increase of the gate oxide thickness at a fixed gate voltage. This agrees with the fact that DT of inversion electrons increases with the increase of gate voltage for a fixed oxide thickness and for a fixed gate voltage decreases with the increase of gate oxide thickness. We see that this error is well above 40% at high gate voltage in thin gate-oxide devices. The error is found to be significant over the whole voltage regime. In Ref. [22] it is claimed without adequate justification that neglect of penetration effect on potential profile does not have any significant influence on modeling of DT currents. This allows calculation of the gate current in a post processor. But our results show that the neglect of the penetration effects causes under-estimation of the measured current. As the error is non-trivial over the entire gate voltage range, we conclude that for accurate modeling of DT gate current, tunneling or penetration effect on potential profile should be considered within the self-consistent loop. The good agreement between our simulation and experiment also shows the usefulness of our model.
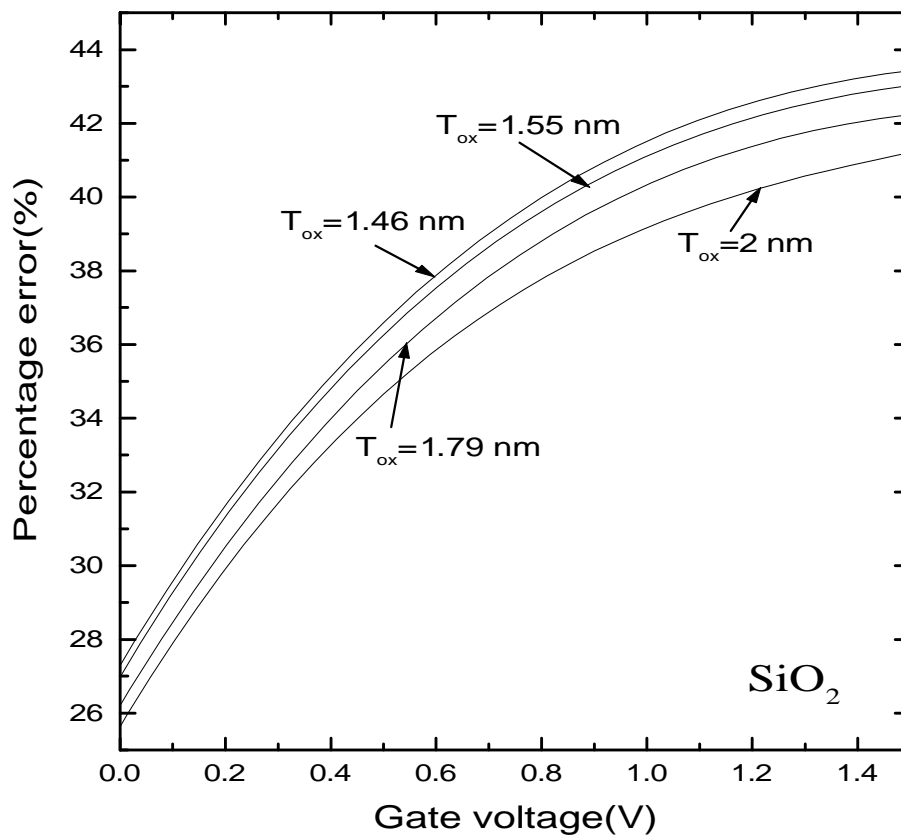

Fig. 4.3: Relative error in direct tunneling gate current shown in Fig. 4.2 due to neglect of tunneling effects within self-consistent loop.

We have already mentioned that as the gate lengths of CMOS devices continue to scale down to the sub-100 nm regime, the device requirements dictate that the gate oxide thickness be scaled down to below 2 nm. However, as the thickness decreases below 2 nm, the direct tunneling current increases rapidly. In order to decrease the leakage current caused by tunneling, the physical thickness of the dielectric layer must increase, while the equivalent oxide thickness must continue to scale down. This is only possible with the introduction of suitable high-dielectric constant (high-K) materials. Hence, ITRS has called for the introduction of high-K materials for MOSFET's below 100-nm channel length due to direct tunneling problems in SiO_2. For this reason we have performed simulation of DT gate current at various EOT with high-K dielectric materials. Materials chosen for simulation are Ta_2O_5, Si_3N_4 and Al_2O_3. This simulation is done taking Al as gate electrode material and dielectric thickness is taken to provide same gate capacitance of SiO_2 counter part (EOT) as dictated by ITRS [2].

It may be mentioned that in a dielectric layer with a given EOT, the DT current decreases with increase in physical oxide thickness of the layer (increase in the value of K). That means FWHM broadening of various eigenstate decreases. When the FWHM becomes too small, due to the precision limit of the compiler, its accurate determination becomes difficult. Therefore, we suggest an alternate indirect technique for evaluating FWHM. To find out the trend of FWHM decrease with increasing EOT, FWHM at various EOT for the same inversion conditions are considered for different materials. Fig. 4.4 shows FWHM as a function of oxide thickness for different inversion conditions for SiO_2 gate dielectric. It is seen that FWHM decreases exponentially with increasing EOT. FWHM vs EOT for two different high-K dielectrics (Al_2O_3 and Ta_2O_5) are also plotted in Figs. 4.5 and 4.6.These figures also confirm the exponential trend observed in SiO_2 case in Fig. 4.4. So FWHM for a given dielectric may be expressed by following equation,

$$\Gamma = \Gamma_o e^{-T_{OX} / L} \tag{4.4}$$

Where Γ_O and L are constants that depend on the dielectric material. We evaluate FWHM at two EOT (small values) from which Γ_O and L are determined. Eq. (4.4) is then used to find FWHM, Γ, at higher EOTs where FWHM may be out of the precision limit of any particular compiler under consideration.

Fig. 4.7 shows the gate current vs gate voltage for different dielectrics with 1 nm EOT. Current considering both penetration and no penetration are shown. We see that for SiO_2 with dielectric constant 3.9, current varies from about 10^{-2} A/cm^2 to around 10^4 A/cm^2 when gate voltage is increased from 0V to 2V. In Si_3N_4 with dielectric constant 7, this variation is from 10^{-3} A/cm^2 to 10^3 A/cm^2, and in case of Al_2O_3 with dielectric constant 7.9, this variation is from 10^{-8} A/cm^2 to 10^{-1} A/cm^2 whereas for Ta_2O_5 this variation is from 10^{-18} A/cm^2 to 10^{-8} A/cm^2. It is evident from this figure that with the increase in dielectric constant significant reduction of gate current is possible. In Fig. 4.7, we observe that in strong inversion, current for different dielectrics have different slopes. It is found that the lower the conduction band offset is, the higher the slope of the current is. We have also performed our simulation for DT gate current through dielectric materials of fig. 4.7 for 1.5 nm and 2 nm EOT. The currents show similar behavior to devices with 1 nm EOT, but the magnitude decreases with increasing thickness of dielectric layer.
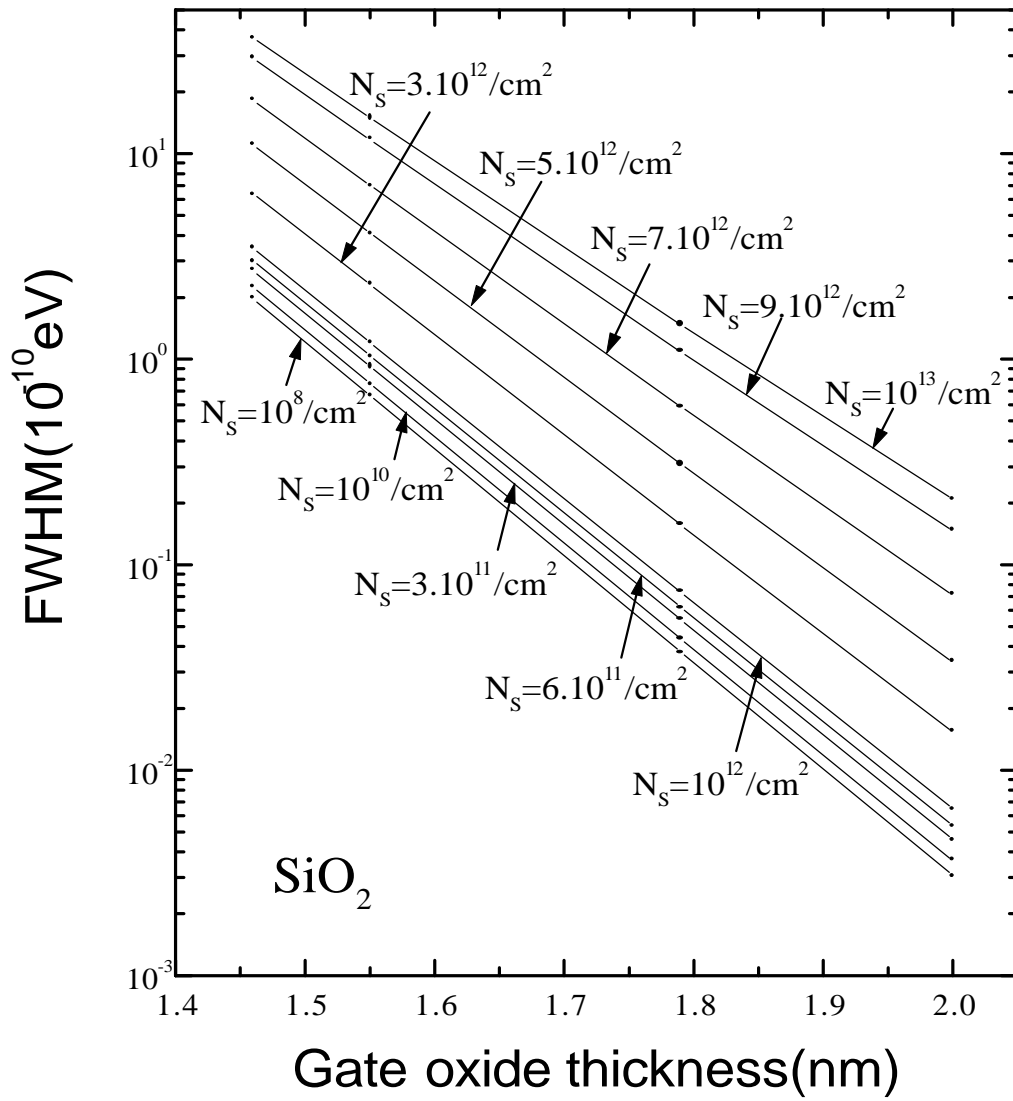
Fig. 4.4: FWHM variation with oxide thickness for different inversion condition for SiO$_2$ dielectric.
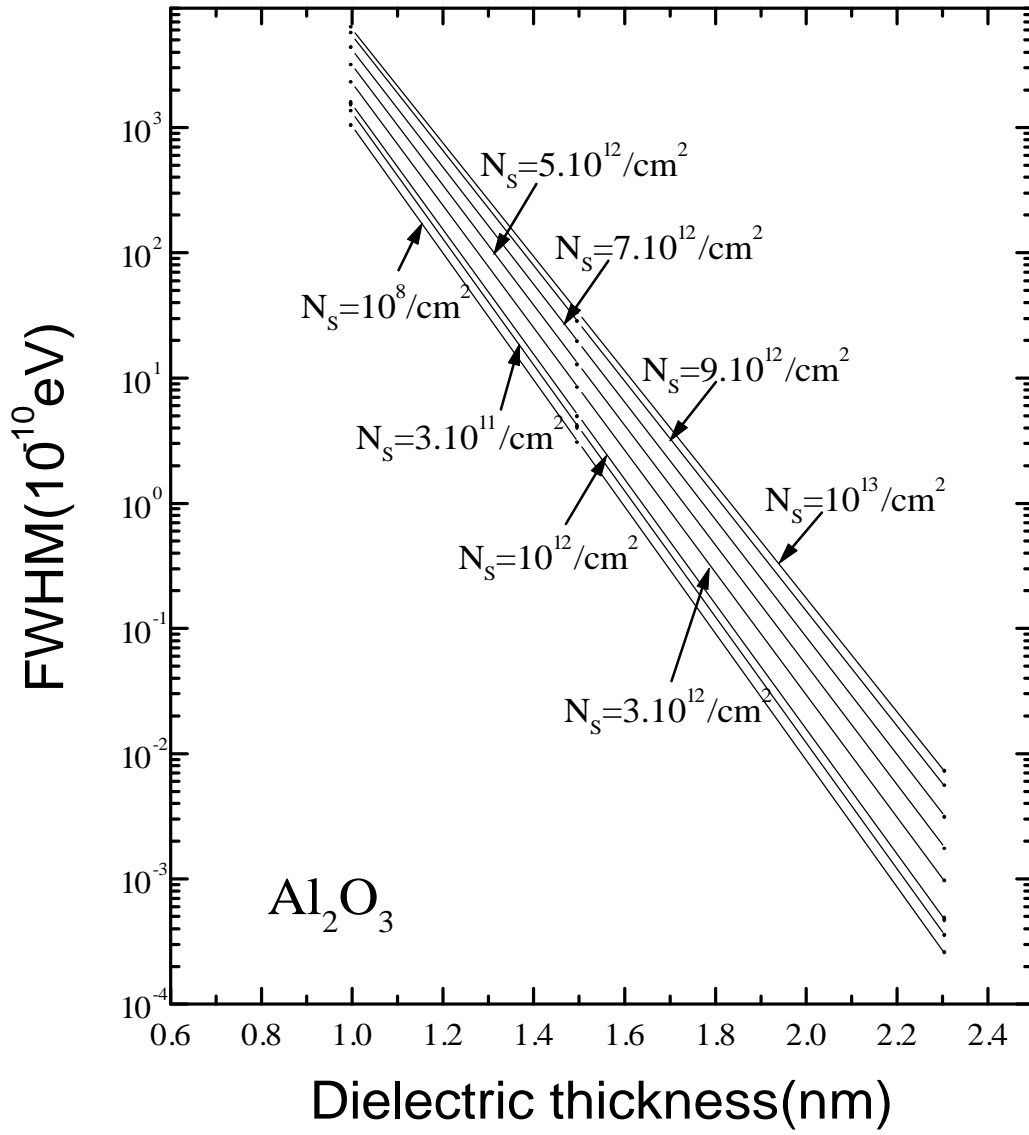
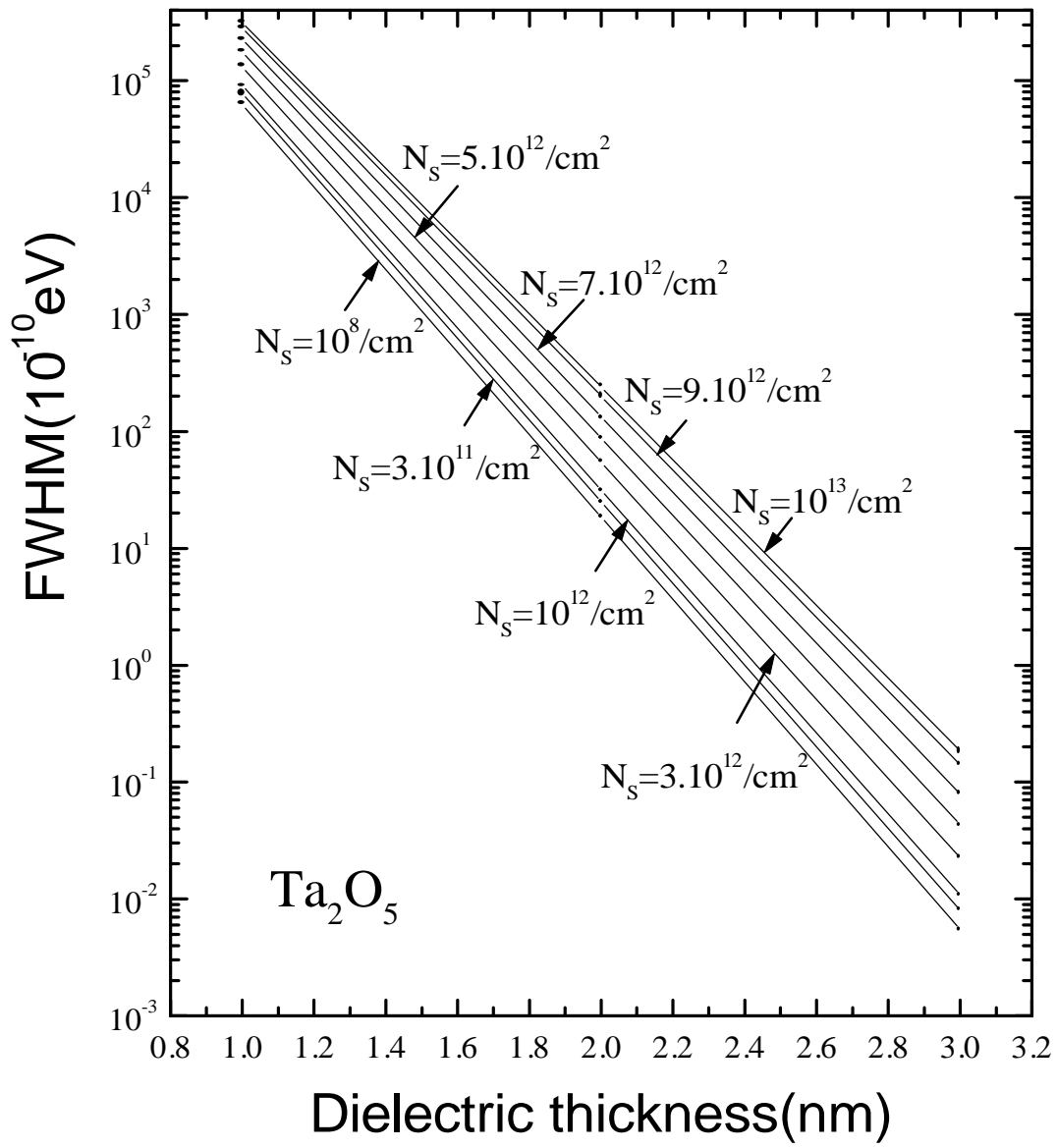Fig. 4.5: FWHM variation with oxide thickness for different inversion condition for Al_2O_3 dielectric.

Fig. 4.6: FWHM variation with oxide thickness for different inversion condition for Ta₂O₅ dielectric.
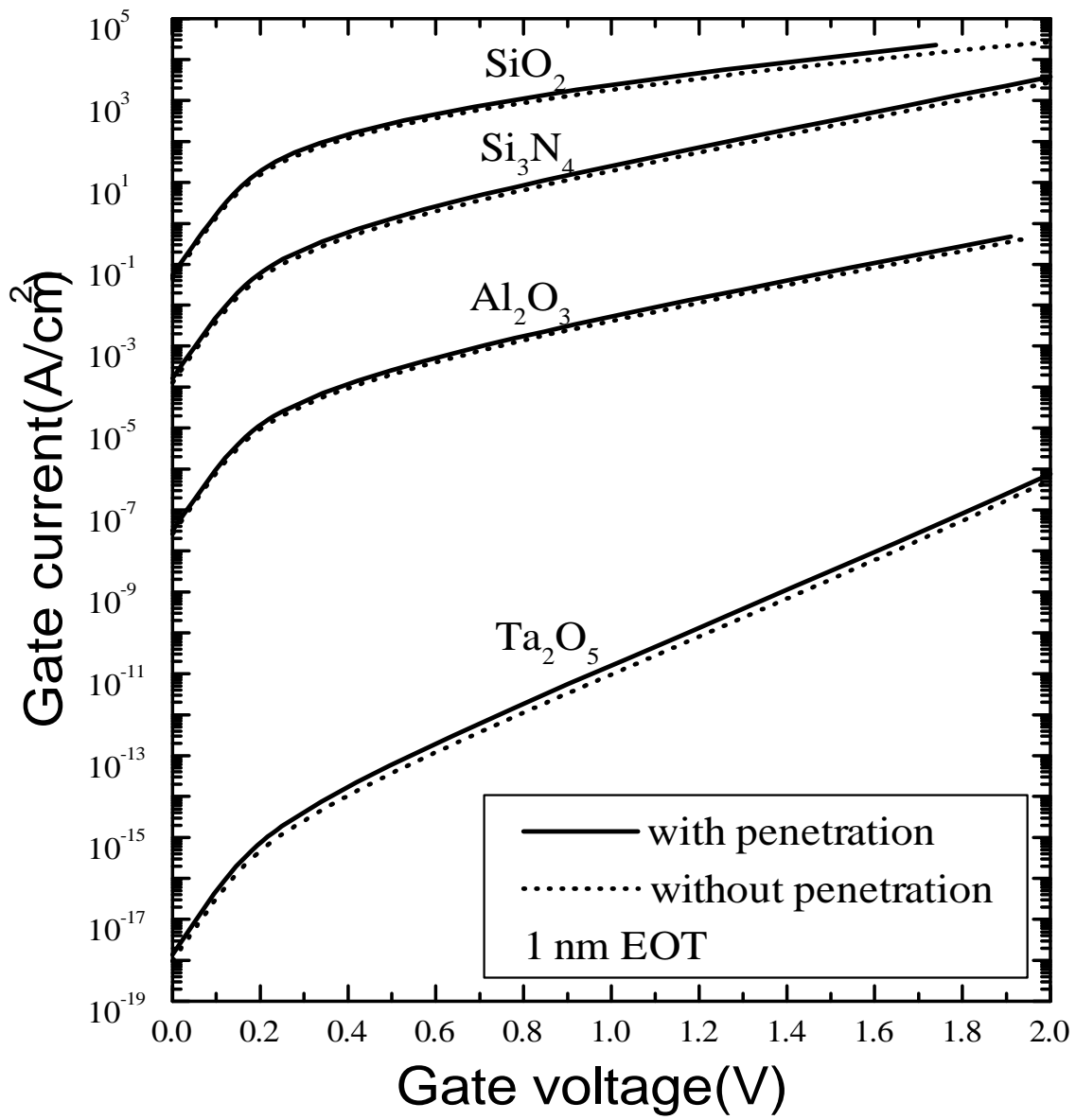
Fig. 4.7: Simulated DT gate leakage current vs gate voltage for different dielectrics for EOT of 1 nm.

To understand importance of wave function penetration effect on potential profile during calculation of direct tunneling gate leakage current, especially in the case of MOSFET's made with high-K dielectric material, the relative error of gate current due to neglect of wave function penetration is considered. Fig 4.8 shows the error for different dielectric materials for three different values of EOT. The errors are significant for all the devices for the entire gate voltage range and can be well above 50% in the worst case considered and indicate the importance of considering wave function penetration within the self-consistent loop during gate current modeling. We see with the decrease of conduction band offset, the error increases for the same EOT in high-K materials. The cause of this increase in error is the increased tunneling of the inversion electrons due to lower barrier height seen by the electrons from Si inversion layer. We can also see that with the decrease of EOT the error in gate current increases for the same dielectric material. Thus, with scaling down of MOSFET dimensions, as higher-K dielectrics are used gate insulators, and as the EOT of the dielectric layers are reduced, the error in DT gate current due to neglect of tunneling effects on potential profile increases. This shows that the necessity of considering penetration effects within the self-consistent loop becomes more critical with device scaling.
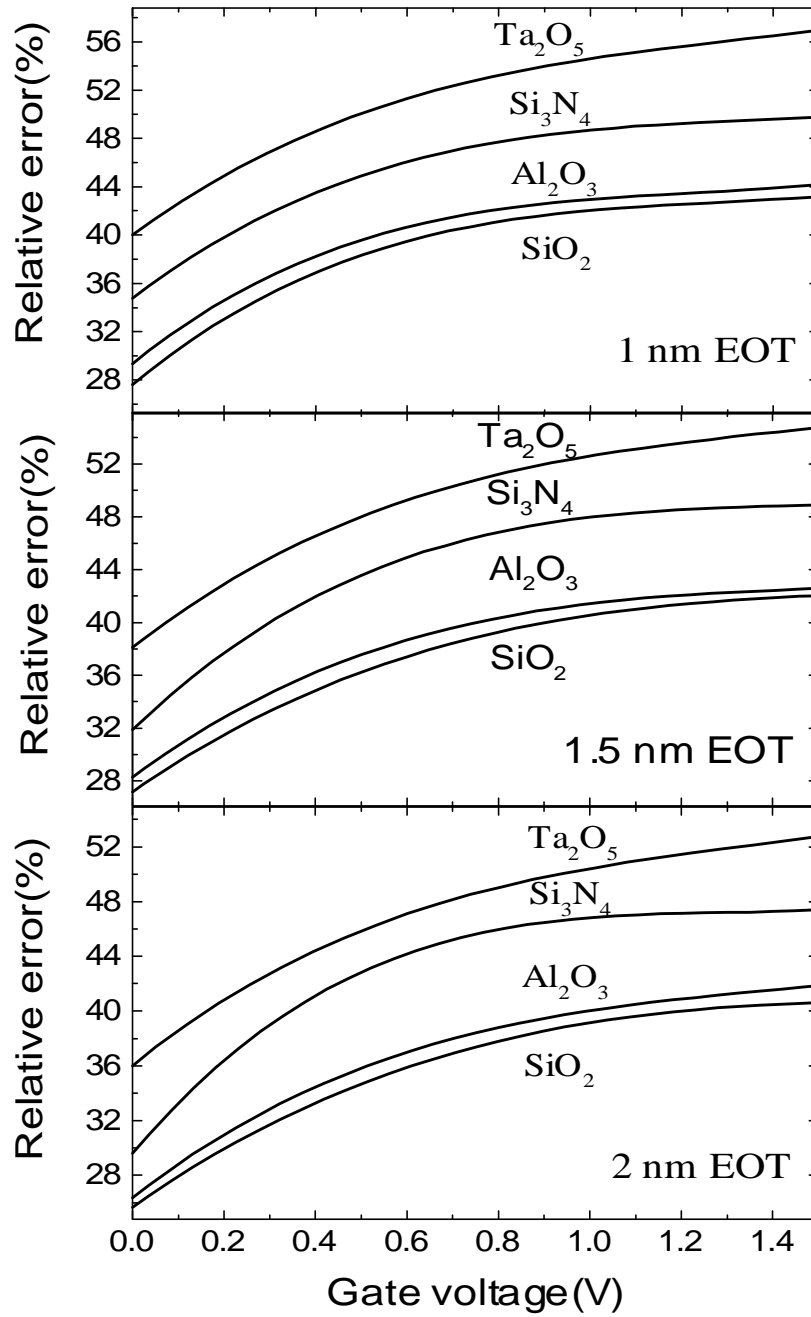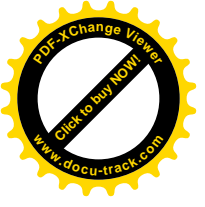
Fig. 4.8: Relative error in DT gate current due to neglect of tunneling effects within self-consistent loop for different dielectrics.

Chapter 5

Modeling of Gate capacitance

It is already mentioned that due to aggressive scaling of MOSFET's, decrease of oxide thickness and increase of channel doping has become indispensable. This results in large surface electric field and steep potential well even near the threshold, leading to significant quantization of carriers in the direction perpendicular to the channel. Threshold voltage shift becomes more important as power supply voltages are reduced and thin oxide increases the relative importance of the capacitive contribution due to channel quantization. In this section, effect on capacitance in the context of thickness of gate dielectric, gate dielectric material, substrate doping and importance of considering wave function penetration and effect of substrate doping are discussed.

5.1 Theory of gate capacitance calculation

Gate capacitance of MOSFET under inversion condition can be represented by the series capacitance of the oxide capacitance per unit area, C_{OX}, and the inversion layer capacitance per unit area, C_{inv}. Inversion layer capacitance has a significant influence on the performance of scaled MOSFET's with thin gate oxides. This is because the total gate capacitance, which determines the transconductance of MOSFET's is reduced by inversion layer capacitance. C_{inv} is defined by,

$$C_{inv} = \frac{q \partial N_S}{\partial \boldsymbol{f}_S} \tag{5.1}$$

Where q is the elemental charge, N_S is the surface carrier concentration, and $\boldsymbol{f}_S$ is the surface potential. The total gate capacitance per unit area, C_g, is defined by,

$$C_g = C_{OX} \frac{1}{1 + \dfrac{C_{OX}}{C_{inv}}} \tag{5.2}$$

Therefore if C_{OX} is comparable to C_{inv}, C_g becomes significantly lower than C_{OX}. In deep submicron MOSFET, gate oxides are very thin that results in larger C_{OX} and degradation of C_g due to C_{inv} becomes larger with the shrinkage of devices. So larger value of C_{inv} is favorable to obtain higher C_g. Another alternate way to achieve increased gate capacitance is to use high-K dielectric material which as predicted by ITRS roadmap[2]. C_{inv} has two physical origins. One origin is the finite band bending associated with an increase in N_S, due to the finite density of states in the bands and the other is finite inversion layer thickness. C_{inv} due to inversion layer thickness is determined by the average distance of the carriers under the 2-D quantization. First one is dominant in week inversion region and second one in strong inversion region [25]. The average

location of inversion carriers from Si/SiO_2 interface physically depends upon gate dielectric material. It will be shown later that in numerical simulation the average location of carriers from Si/SiO_2 interface significantly changes if wavefunction penetration effect is incorporated, especially in the case of high-K dielectrics. As ITRS roadmap has already suggested the replacement of SiO_2 gate dielectric by high-K materials, accurate modeling of gate capacitance incorporating wavefunction penetration effect is necessary.

Detail procedure of our numerical calculation has been discussed in chapter 2. Once self consistent result is obtained, C_{inv} can easily be determined from equation (5.1). Determination of C_{inv} by equation (5.1) automatically incorporates both physical origins of C_{inv} without any specific definition of the inversion layer thickness [25]. For gate capacitance calculation, instead of Eq. (5.2), we use the fundamental definition:

$$C_g = \frac{q\partial(N_S + N_{dep})}{\partial V_G} \tag{5.3}$$

5.2 Results

We know quantization of carriers shifts the peak of carrier concentration away from the interface and thus results in an increase of effective oxide thickness. This increase is manifested as lower transconductance and a lower quasi-static capacitance in strong inversion. In semi-classical calculation, MOS quasi-static capacitance is predicted to be lower than the ideal oxide capacitance. This is due to a finite mean displacement predicted by even the semi-classical charge distribution caused by Fermi-Dirac distribution of carriers in energy and a finite density of states at a given energy. QM simulation shows degraded quasi-static capacitance compared to semi-classical simulation. This degradation due to QM effect becomes increasingly significant as the gate oxide thickness is scaled. We also know C_{inv} due to average location of inversion carriers from Si/SiO_2 interface is dominant in strong inversion region. Therefore the impact of wavefunction penetration effect into the gate dielectric on accurate modeling of capacitance can be understood by studying the average location of inversion carriers from silicon/dielectric interface. This study is done with SiO_2 gate dielectric and also with high-K dielectrics, such as, Si_3N_4, Ta_2O_5 and TiO_2. Conduction band offset and dielectric constant of these materials are taken from ref [38] and are shown in Table 3.1.

Fig. 5.1 shows average DC centroid (Z_{avg}) of inversion carriers from silicon/dielectric interface at various voltages for SiO_2, Si_3N_4, Ta_2O_5 and TiO_2 dielectrics when EOT is 1 nm with and without considering wavefunction penetration into the gate dielectric. To achieve inversion condition $1.5\times10^{13}/cm^2$, average DC centroid of inversion charge found is 1.16 nm for all the materials when no penetration is considered. We see that the Z_{avg} is material independent in this case. The same observation is also made with different EOT when no penetration is considered. Fig. 5.1 also shows DC centroid of inversion carriers when wave function penetration is allowed for an EOT of 1
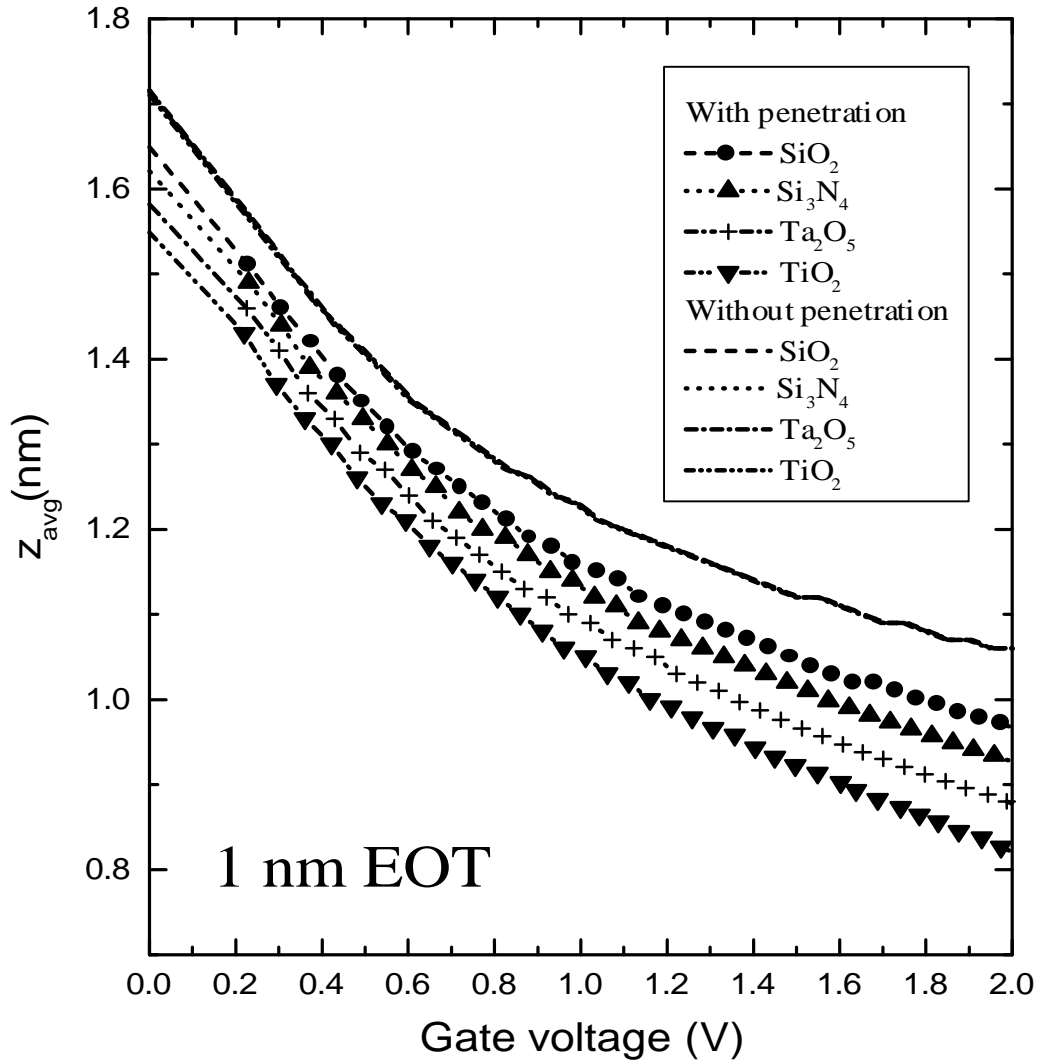
Fig. 5.1: Z_{avg} variation with gate voltage for different dielectrics of 1 nm EOT both considering and without considering penetration.

nm. In this case, to achieve inversion condition of $1.5\times10^{13}/cm^2$, DC centroid of inversion charge found are 1.099 nm, 1.07 nm, 1.032 nm and 0.9911 nm for SiO_2, Si_3N_4, Ta_2O_5 and TiO_2, respectively. While considering penetration, gate voltage required to achieve same inversion condition is not the same for different dielectrics. The gate voltage in this case depends upon the amount of wave function penetration into the gate dielectric, which in turn depends upon the conduction band offset. Therefore, it is found that allowing for wave function penetration results in Z_{avg} to be shifted closer to Si/dielectric interface and this effect also depends upon gate dielectric material. We see that a lower conduction band offset shifts the DC centroid of inversion charge closer to the Si/dielectric interface. To confirm this we also calculate Z_{avg} for EOT 1.5 nm and 2 nm for the chosen dielectrics and similar results are obtained. Z_{avg} for $N_S=1.5\times10^{13}/cm^2$ are shown in Table 5.1. This table confirms that with the decrease of the conduction band offset, penetration becomes more prominent, as Z_{avg} shifts closer to the interface. This will have strong effect on inversion layer capacitance and gate capacitance obtained from quasi-static calculation of dQ/dv, is expected to be larger than what one would obtain assuming zero wave function penetration.

To predict possible behavior of capacitance with different EOT for a given dielectric, Z_{avg} variation with gate voltage is plotted in Fig. 5.2 for EOT of 1nm, 1.5 nm and 2 nm for SiO_2. From this figure it is seen that Zavg is smaller when oxide thickness is lower for the same applied voltage. This is because when the EOT is greater, a greater voltage drop occurs across the dielectric layer and a smaller voltage is available to create inversion. Therefore, same gate voltage in a device with thicker EOT corresponds to a lower inversion carrier density. As Z_{avg} is higher at lower inversion condition [40], the results of Fig. 5.2 can be explained. Similar effect is also found in the case of high-K dielectrics.

TABLE 5.1: Z_{avg} (nm) of different dielectric for inversion condition of $1.5\times10^{13}/cm^2$

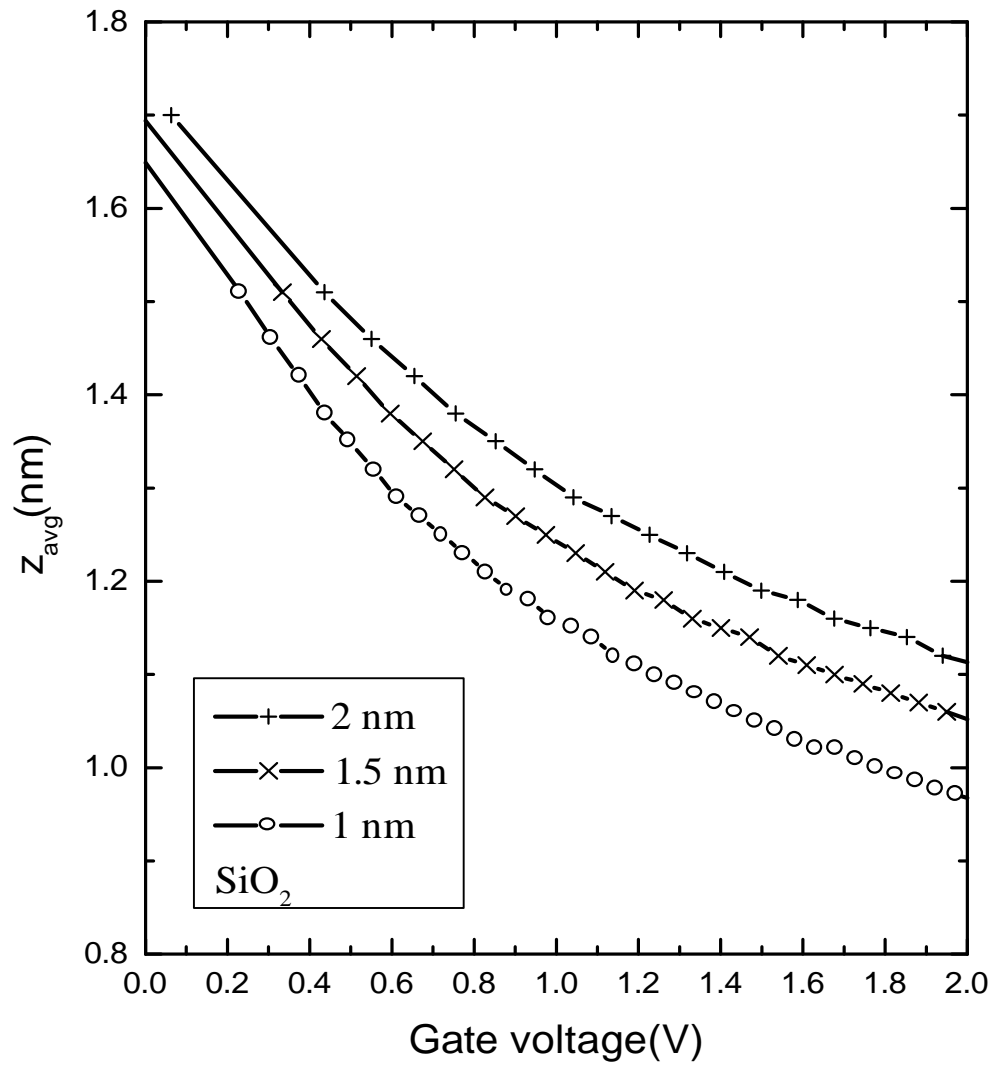| Material	Without penetration	With penetration		
	1 nm EOT	1 nm EOT	1.5 nm EOT	2 nm EOT
SiO_2	1.16	1.099	1.099	1.099
Si_3N_4	1.16	1.07	1.069	1.07
Ta_2O_5	1.16	1.032	1.02	1.02
TiO_2	1.16	0.99	0.98	0.97

35

Fig. 5.2: Z_{avg} variation with gate voltage for different dielectric thickness for SiO_2 when penetration is considered.

Since C_{inv} is dependent on N_S and doping density of Si substrate, N_{SUB}, it is necessary to characterize C_{inv} as a function of N_S and N_{SUB}. Fig. 5.3 shows modeled N_S dependence of C_{inv} for different gate dielectrics with three different gate oxide thickness, from 1 – 2 nm, at room temperature both considering and without considering penetration. Substrate impurity concentration is the same (10^{18}/cm^3) for all the cases. C_{inv} at a given value of N_S should be independent of the oxide thickness if we neglect penetration effect. The identical C_{inv}-N_S curves observed in Fig. 5.3 for the three oxide thickness when penetration effect is neglected reflects this observation. C_{inv} at any particular N_S is larger when penetration is incorporated. We see in each of the dielectric thickness, when penetration effect is incorporated, C_{inv} is larger and this effect is prominent only in the strong inversion region. This is explained using the fact that in strong inversion region C_{inv} is dominated by $Z_{avg.}$ Another important aspect found in Fig. 5.3 is that C_{inv} with penetration is almost the same for a given N_S at different oxide thickness for a given dielectric. Behavior of C_{inv} with N_S in high-K dielectrics are found to be qualitatively identical as that for the SiO$_2$. However like Z_{avg}, penetration effects on C_{inv} is greater for dielectric materials with lower DE_C. It is also found that C_{inv} increases with an increase in N_S, meaning that the influence of C_{inv} upon gate capacitance becomes more severe for lower N_S. With increasing N_S, the N_S dependence on C_{inv} becomes weaker, from almost linear dependence to some order of fractional power dependence demonstrates the change in the physical origin of C_{inv}, from the effect of the finite density of states at lower N_S to the quantum mechanical inversion layer thickness at higher N_S.

To investigate the impact of wavefunction penetration on devices with different gate dielectric materials, relative error in C_{inv} due to neglect of penetration is plotted in Fig. 5.4 for an EOT of 1 nm. It is found that the error is maximum in TiO$_2$ and minimum in the case of SiO$_2$. This is again related to DE_C. The error increases with increasing gate bias and at high gate voltage is significant for all the devices. We have also verified that this error for 1.5 nm and 2 nm EOT for different dielectrics are similar. However the magnitudes are smaller in the devices with larger EOT. This implies that for accurate modeling of C_{inv} in deep sub micron MOSFET's, wavefunction penetration effect must be incorporated, especially for future generation devices where high-K dielectric will be the potential replacement of SiO$_2$.

Gate capacitance of MOSFET is calculated from equation (5.3) which takes account of the both depletion or oxide capacitance and inversion layer capacitance. Fig. 5.5 shows gate capacitance as a function of gate voltage for all four gate dielectrics at 1 nm EOT both considering and without considering penetration. The pattern obtained basically is due to C_{inv} variation. At low voltage, depletion capacitance is larger but due to lower C_{inv}, gate capacitance in this domain is mainly reduced by C_{inv}. At high voltages or in strong inversion region C_{inv} is larger and variation of C_{inv} with gate voltage is small as found in the previous Fig. 5.3. Therefore C_{inv} forces C_g variation with gate voltage to be small, approaching to be almost constant at very high voltages. Effects of wavefunction penetration is also reflected here. When C_g is calculated using conventional closed boundary conditions, for a given EOT, it is independent of gate dielectric material. Fig. 5.5 shows that due to different DE_C for different dielectrics, wavefunction penetration effect are not the same for all the materials. Consequently it is found that the gate capacitance varies from material to material even when the EOT is the same. This is an important observation which emphasizes the need for incorporating
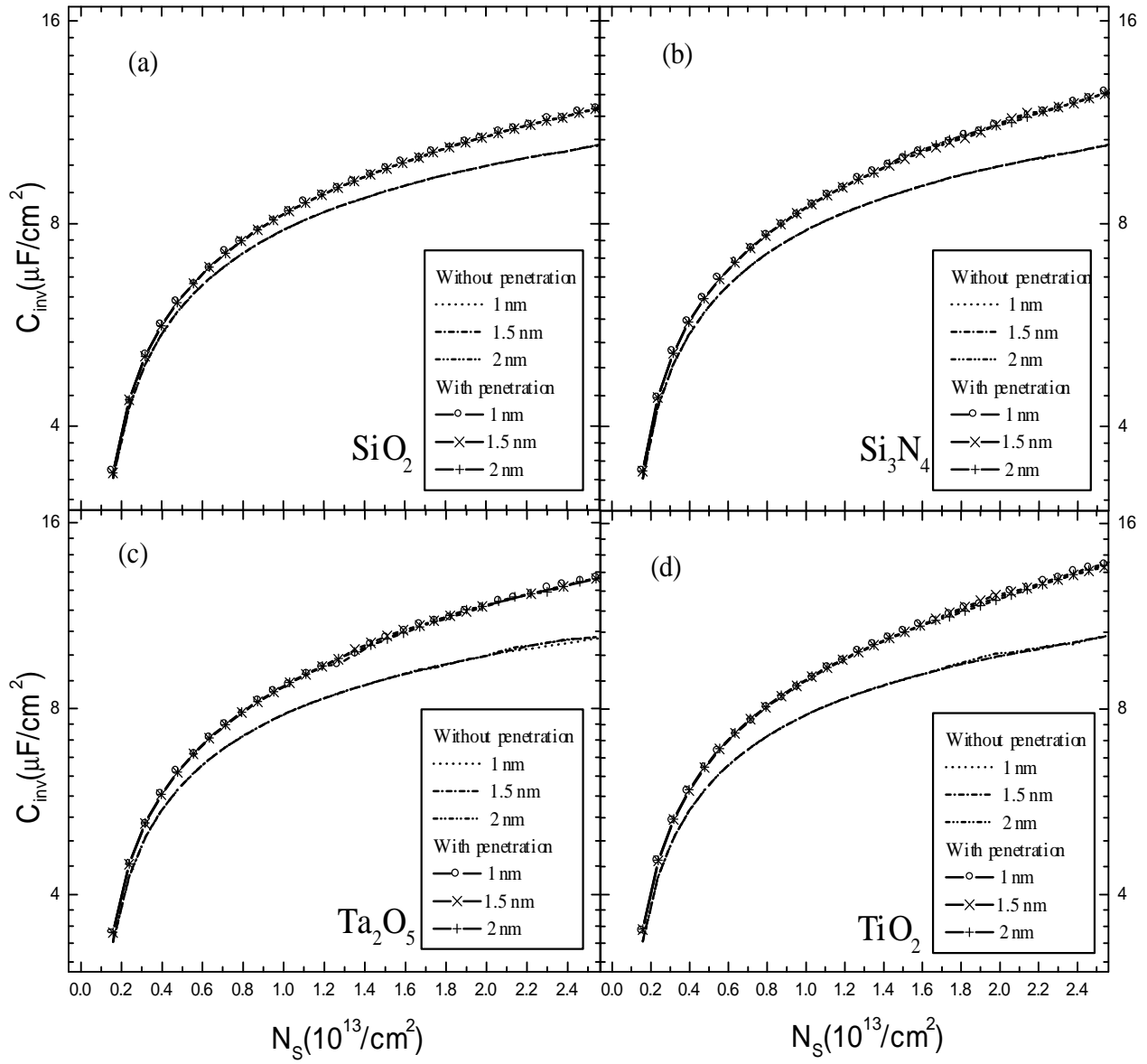
Fig. 5.3: C_{inv} vs N_S for different gate dielectrics of 1 nm EOT both considering and without considering penetration.
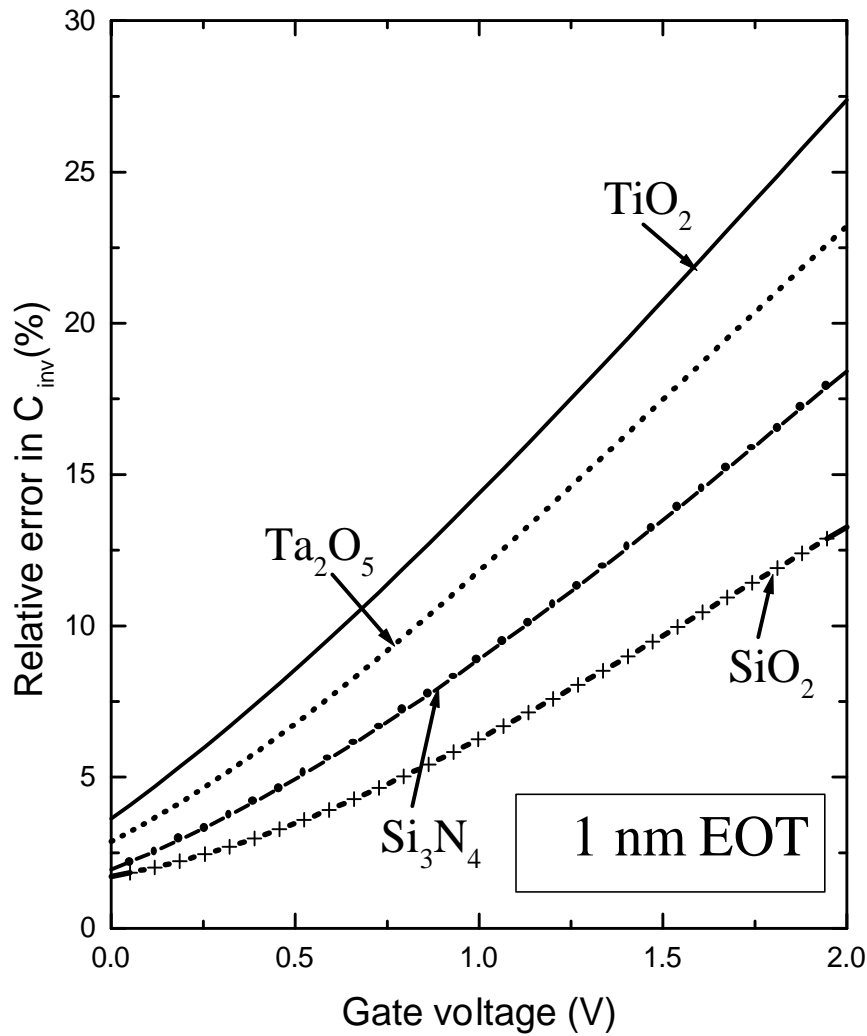
Fig. 5.4: Relative error in calculation of C_{inv} due to neglect of penetration at various voltages for different dielectrics.

penetration effects in *C-V* calculation, particularly for MOSFETs with high-K dielectric material. Fig. 5.6 shows C_g vs V_g for different dielectric materials for an EOT of 1.5 nm. The trend is identical with that for 1 nm EOT except for the fact that penetration effects are more pronounced at lower EOT.

To have a quantitative understanding of penetration effects on C_g, the relative errors of gate capacitance for different dielectric materials for an EOT of 1 nm are shown in Fig. 5.7. As expected, the error is maximum in TiO_2, and minimum for SiO_2. The values of error is quite significant in strong inversion region and can be greater than 6% for a particular material. Even for SiO_2, it is higher than 3%. As the maximum permissible error in modeling C_g is 5% according to ITRS [2], inclusion of penetration effect in gate capacitance model is essential in MOSFETs with ultra-thin gate dielectrics.
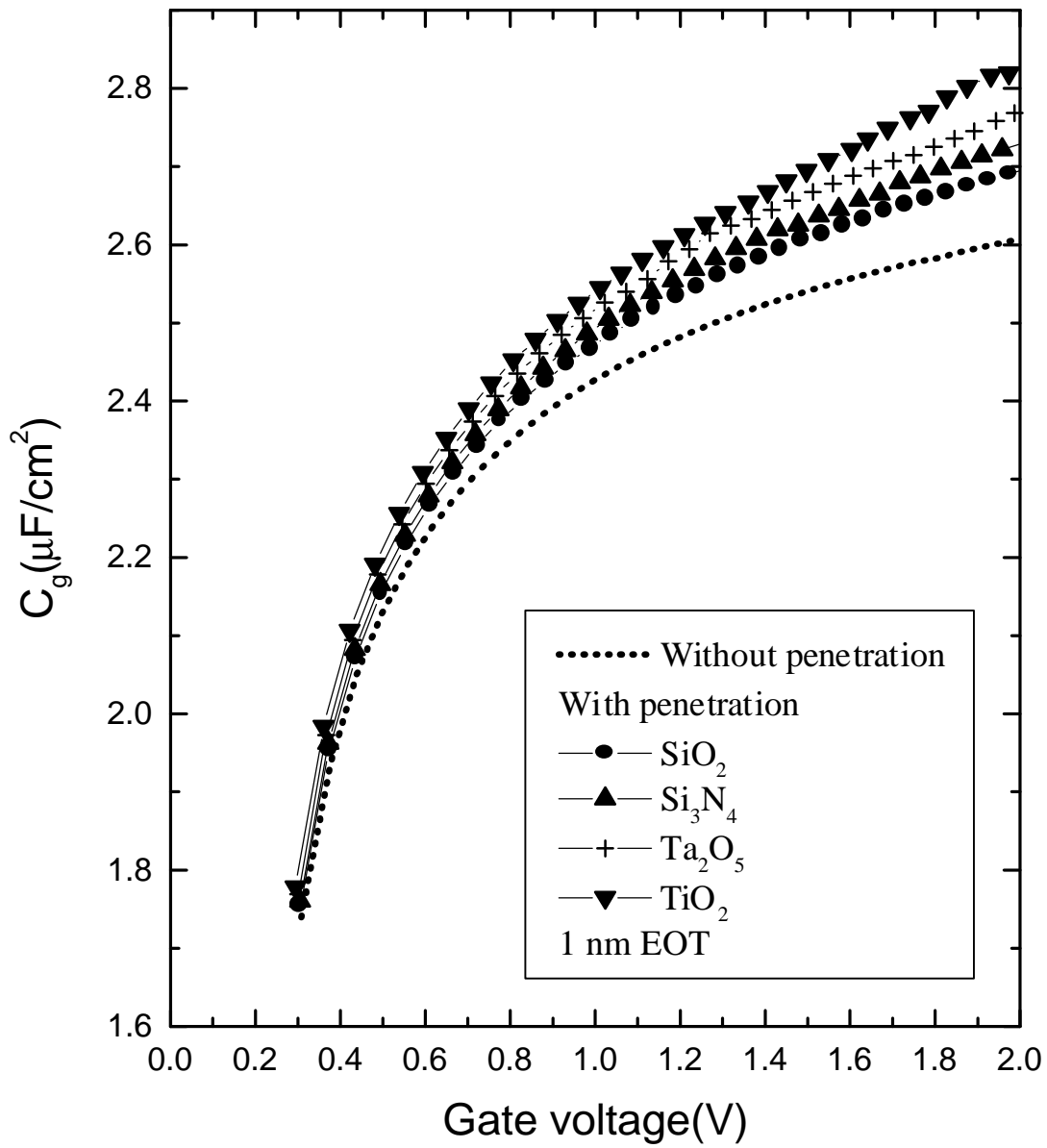
Fig. 5.5: C_g as a function of gate voltage for different dielectrics at 1 nm EOT both considering and without considering penetration.
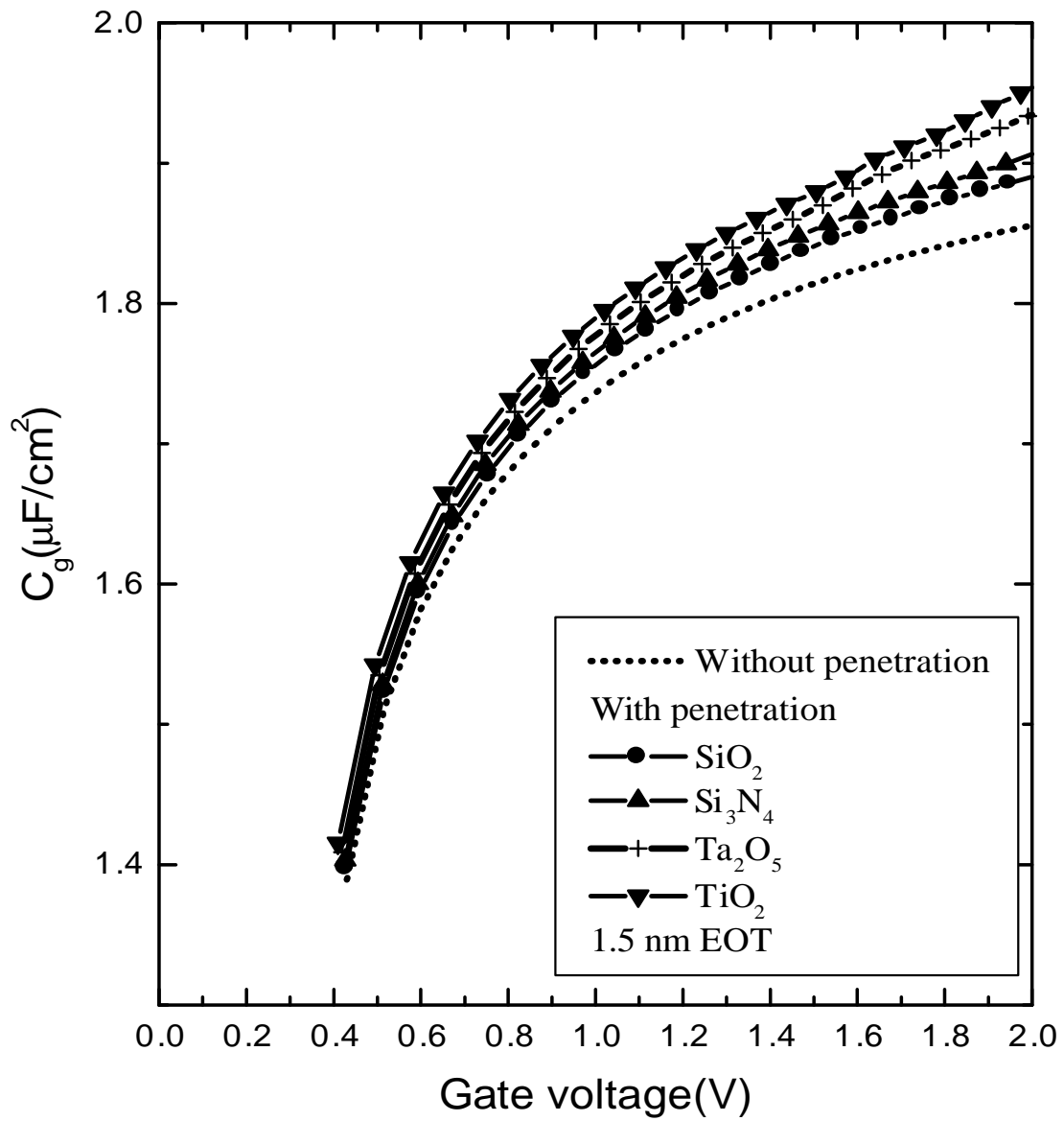
Fig. 5.6: C_g as a function of gate voltage for different dielectrics at 1.5 nm EOT both considering and without considering penetration.
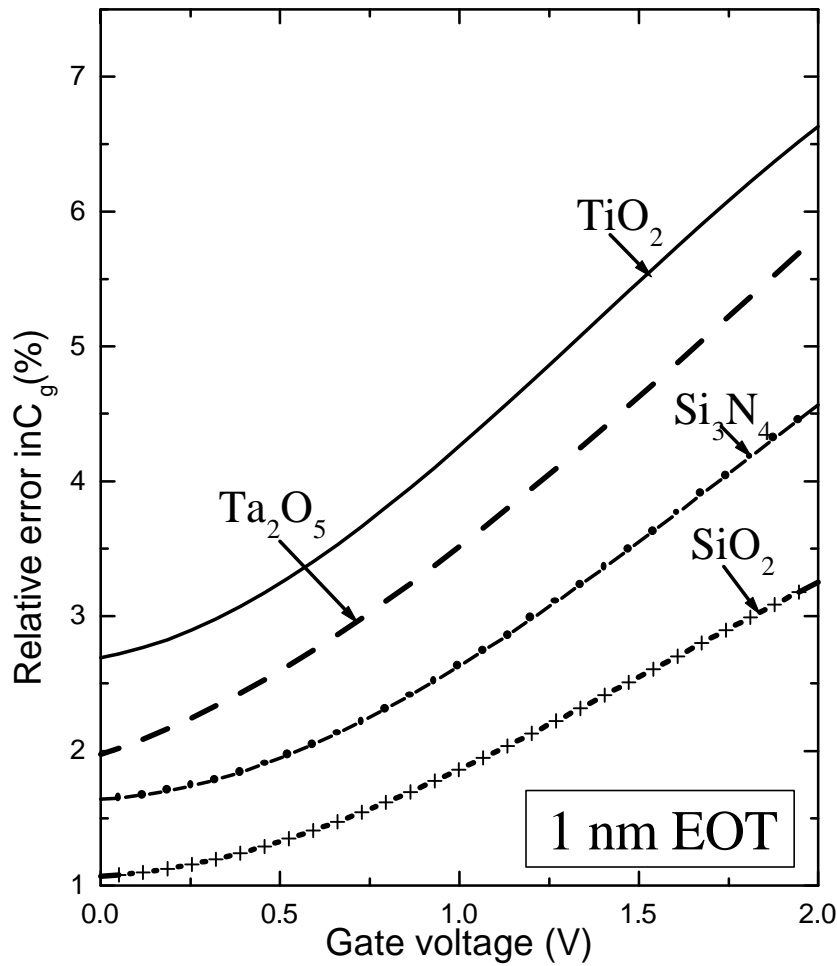
Fig. 5.7: Relative error in gate capacitance due to neglect of penetration for different dielectrics of 1 nm EOT.

Since scaling down of device dimension is accompanied by an increase in substrate doping density, it is important to know the dependence of $C_{inv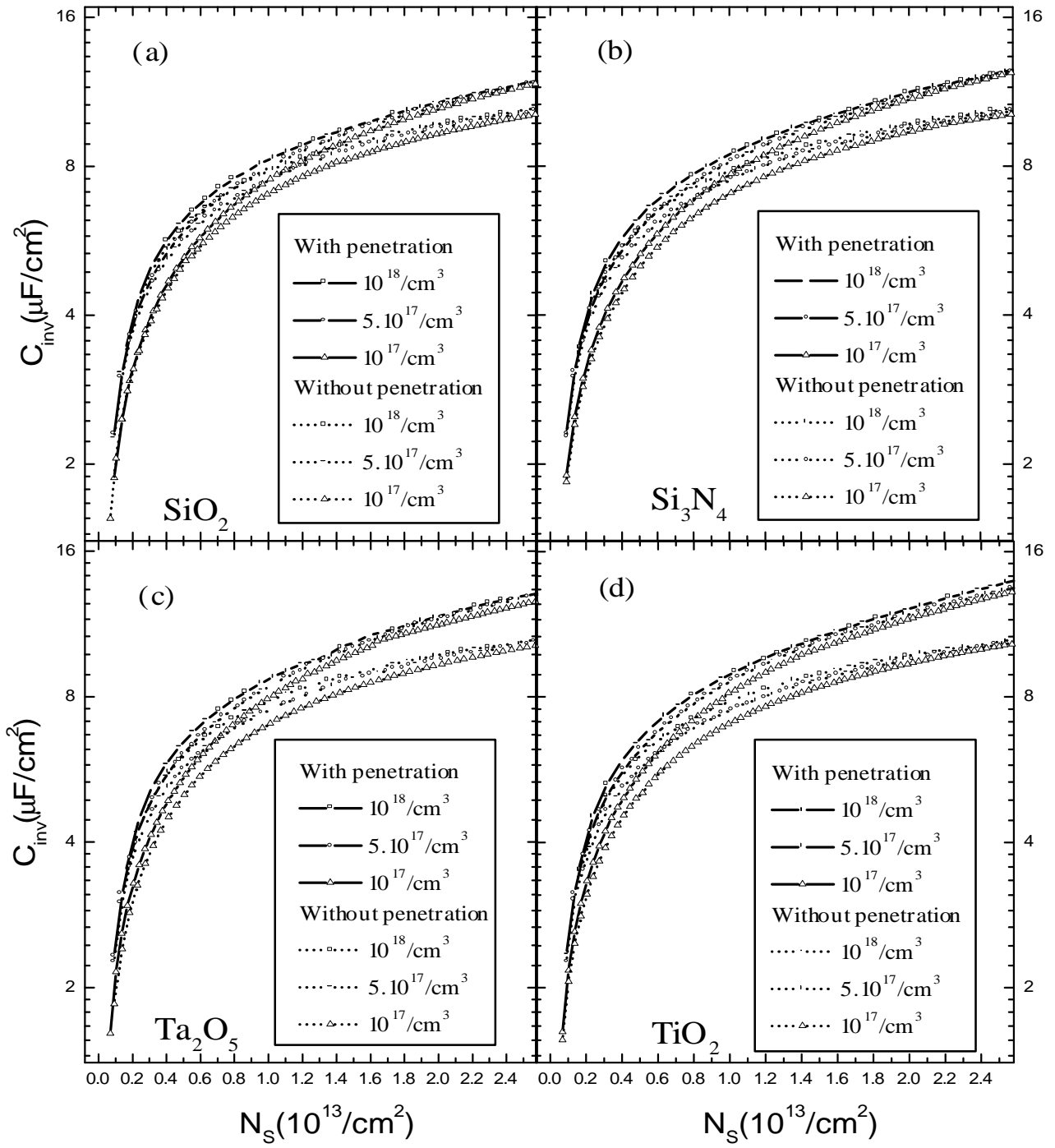}$ on substrate doping concentration. Fig. 5.8 shows the calculated C_{inv}-N_S curves with different substrate doping concentrations both considering and without considering penetration. All four gate dielectric materials are considered with 1 nm EOT. It is found that the increase in the substrate doping concentration leads to an increase in C_{inv} for a fixed inversion charge density. This is consistent the results of Ref. [25]. The increase is observed both in sub threshold and strong inversion region for a same inversion charge density.

Fig. 5.9 shows the relative error in calculating C_{inv} due to neglect of penetration for different dielectrics at different N_{SUB} for 1 nm EOT. We see from this figure the relative error is maximum in case of TiO$_2$ with $N_{SUB} = 10^{17}$/cm^3 and minimum in the case of SiO$_2$ with $N_{SUB} = 10^{18}$/cm^3. For each material, error increases with decreasing N_{SUB}.

42

Fig. 5.8: C_{inv}-N_S curves with different substrate doping concentrations both considering and without considering penetration.
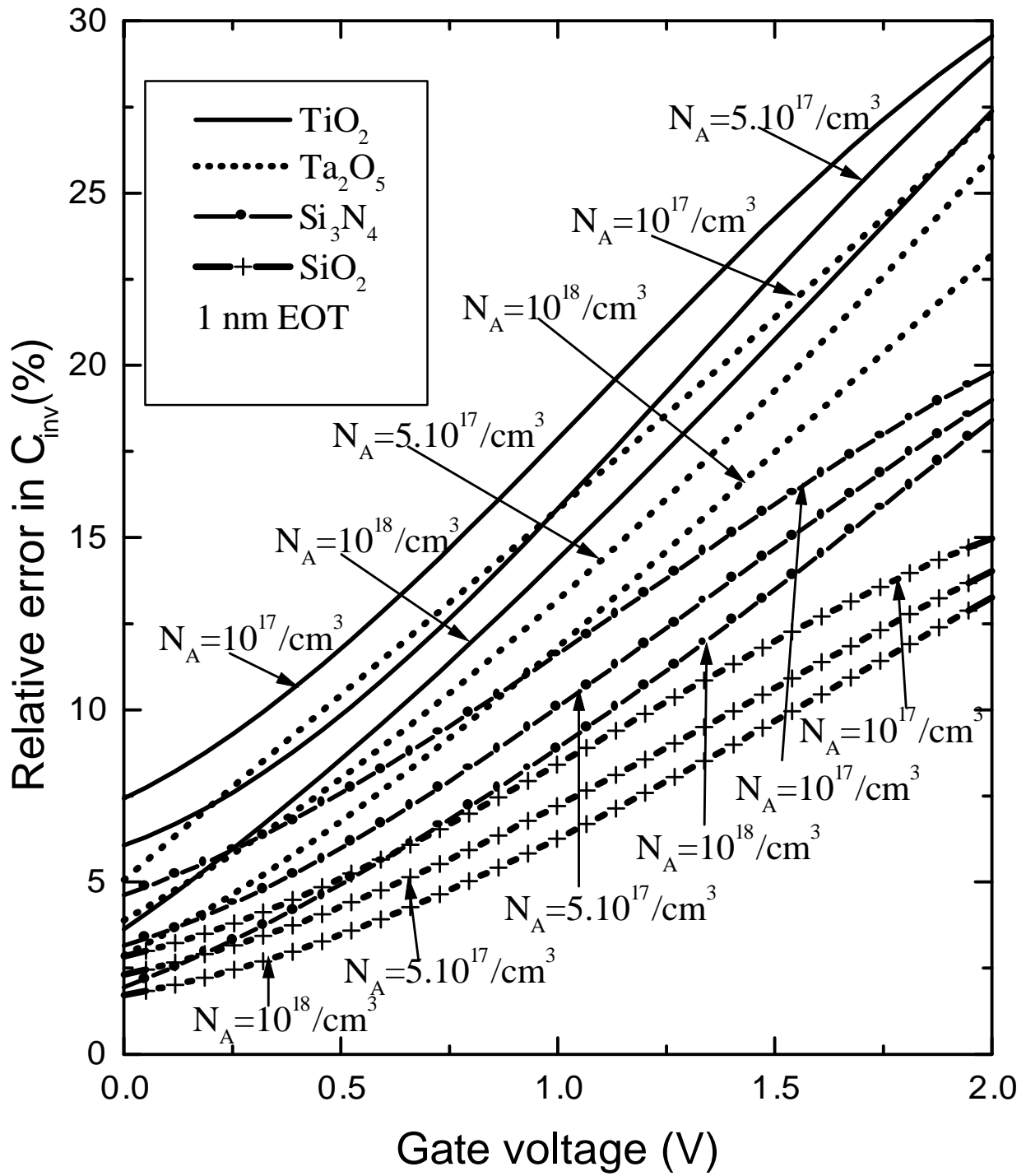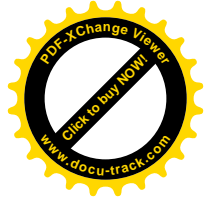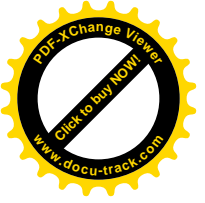
Fig. 5.9: Relative error in C_{inv} due to neglect of penetration for different dielectrics at different N_{SUB} for 1 nm EOT.

Finally, we investigate effect of N_{SUB} on gate capacitance. Fig. 5.10 shows C_g variation with voltage at different N_{SUB} for different gate dielectric materials of 1 nm EOT. These results show that effect of N_{SUB} is also reflected in C_g. At sub threshold region C_g increases with decrease of N_{SUB} and at strong inversion region C_g increases with the increase of substrate doping for a fixed voltage. The cause of this reverse behavior in weak inversion region can be explained from the phenomenon that in weak inversion for a same applied voltage inversion charge density created is greater in the case of low substrate doping or in another way behavior of depletion region is dominant here. After achievement of strong inversion C_{inv} becomes dominant and C_g follows general trend of increase with the increase of N_{SUB} as found in Fig. 5.8. In strong inversion region the increase of C_g due to increase in N_{SUB} at a given gate voltage is observed when penetration effect is neglected. Little variation in C_g with N_{SUB} is observed beyond threshold when penetration effect is incorporated. It can be explained from the fact that Z_{avg} shifts closer to the Si/dielectric interface for an increase in substrate doping [25]. But in our result this Z_{avg} shift due to change in N_{SUB} is not so much prominent when penetration effect is considered. This will be clear from Fig. 5.11 where Z_{avg} shift due to N_{SUB} change from 5×10^{17}/cm^3 to 10^{18}/cm^3 is plotted against voltage for both considering and without considering penetration. We see from this figure that Z_{avg} shift while considering penetration is lower than that of the non penetration case. Therefore increase in C_g with the increase of N_{SUB} for a same applied voltage is almost un-noticable when penetration effect is considered and significant change is observed when penetration effect is neglected. This implies that conventional modeling where penetration effect is neglected will predict overestimated increase in C_g due to increase in N_{SUB} for a given voltage. Difference between C_g at a fixed voltage between penetration and non penetration conditions is larger in lower N_{SUB} and this difference increases with decreasing conduction band offset of high-K materials. We also note that in the sub-threshold region, effects of variation of N_{SUB} on C_g is large, but it decreases as gate voltage increases.

Relative errors in C_g for different dielectrics are plotted in Fig. 5.12 for various substrate doping concentrations for an EOT of 1 nm. We see in this figure that the error is higher at low substrate doping concentration for a given dielectric material and increases with decreasing conduction band offset. We have also checked this at 1.5 nm EOT and 2 nm EOT for the mentioned materials and found similar result of N_{SUB} dependence, but quantitative value of this error decreases with increasing EOT. In the worst case of 1 nm EOT, we see that the error in gate capacitance is below 5% in case of Si_3N_4 and SiO_2 at 2V gate voltage, which is below the maximum error permitted by ITRS for any particular simulator. But in case of Ta_2O_5 and TiO_2, this error is above 5% limit. Therefore, it can be concluded that for accurate modeling of gate capacitance of high-K dielectrics, penetration effect must be considered.
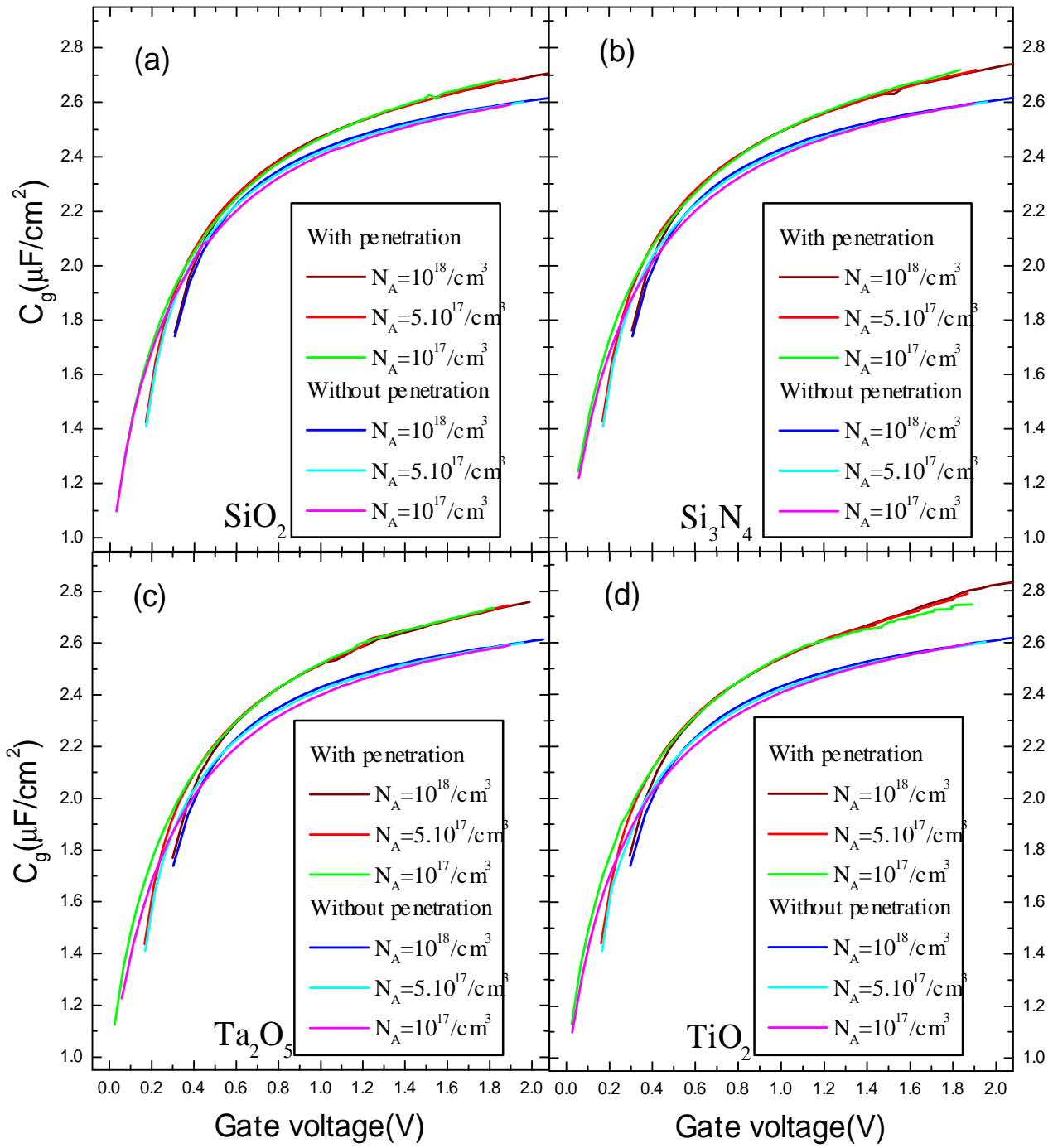
45

Fig. 5.10: C_g variation with voltage at different N_{SUB} for different gate dielectric materials of 1 nm EOT.
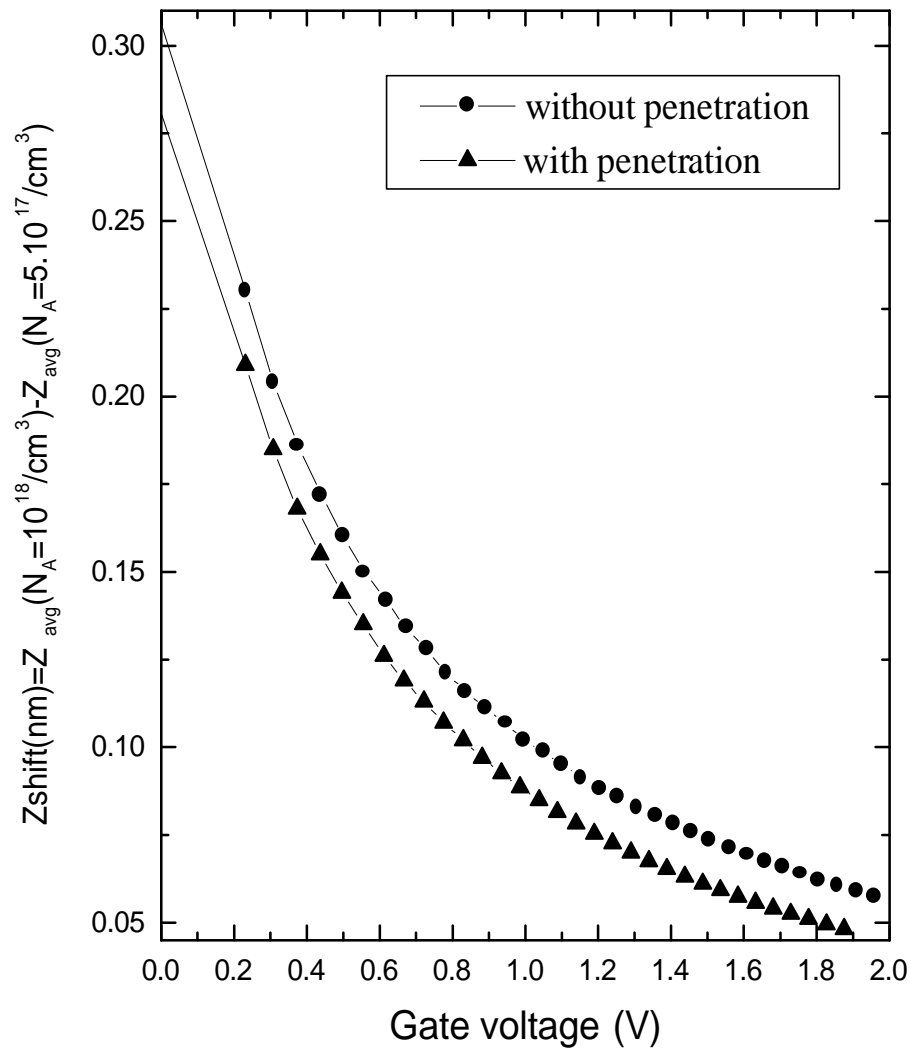
Fig. 5.11: Z_{avg} shift due to change of N_{SUB} at various voltages both considering penetration and without considering penetration for SiO_2 of 1 nm EOT.
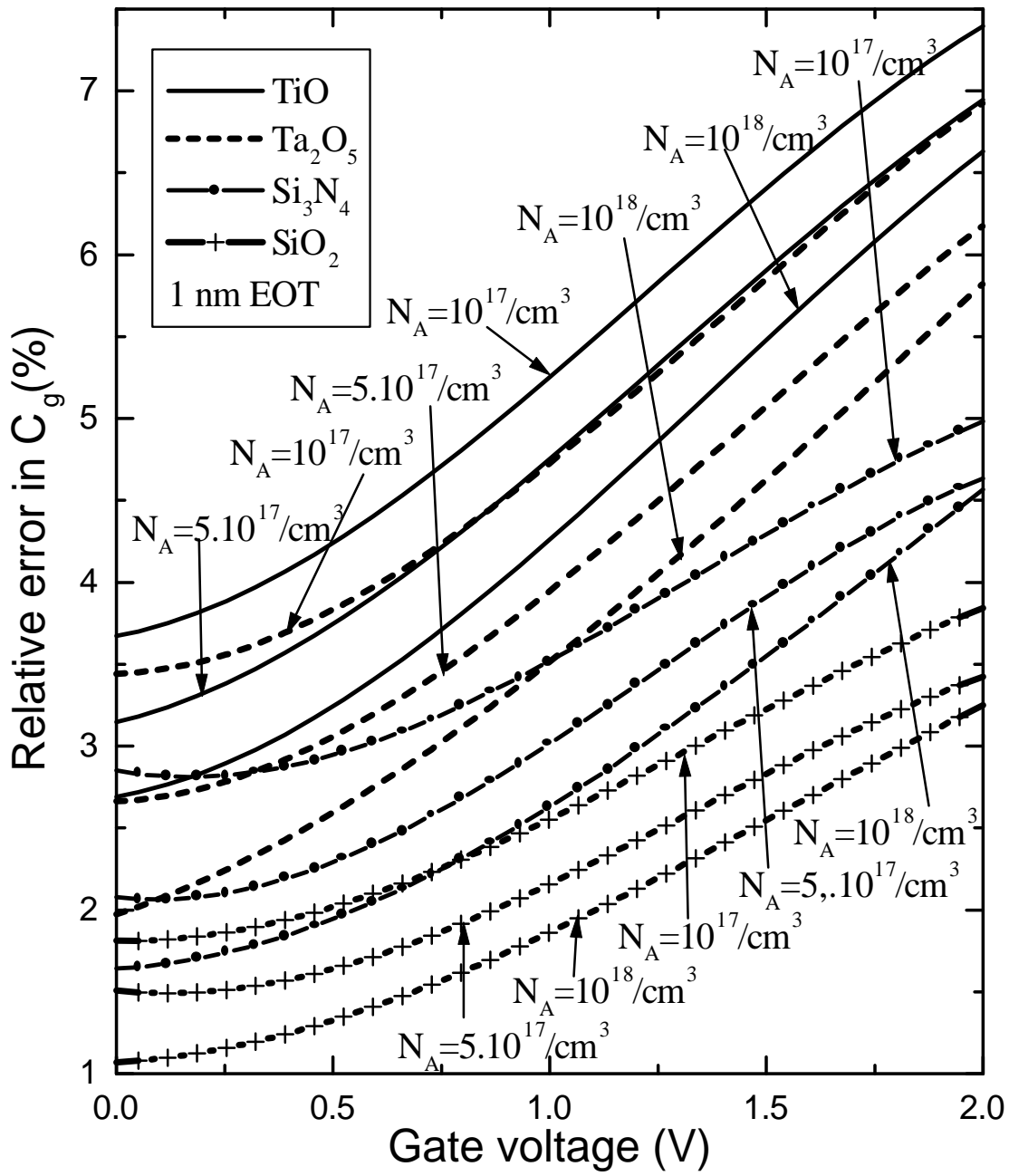
Fig.5.12: Relative error in C_g due to neglect of penetration at different N_{SUB} for various dielectrics of 1 nm EOT.
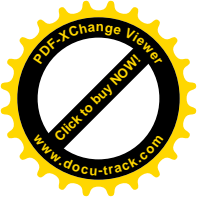
Chapter6

Conclusion

6.1 Summary

Accurate modeling of direct tunneling current and gate capacitance of deep sub-micron MOSFETs with high-K dielectrics used as gate insulator, is presented in this work. The method used, is numerically simple and does not require either lengthy matrix manipulation or calculation of complex eigenenergies. The eigenenergies and lifetimes of quasi-bound states are calculated from the peaks and the broadening of the 1D DOS around each eigenenergy, respectively. It is found that for accurate modeling of DT gate current and gate capacitance penetration effect should be taken into account. DT current is under-estimated when penetration effects on potential profile are neglected. Inclusion of this effect has led to excellent agreement with published experimental results. The relative error in case of DT current is significant over entire range of gate voltage. This error increases with decrease of the potential barrier height at silicon-gate-dielectric interface and decreases with increasing oxide thickness.

   It is found that the value of gate capacitance is independent of gate dielectric material for a given EOT from conventional modeling. This is not the actual case for MOSFETs with high-K dielectrics. Rather, gate capacitance is found to vary from material to material even when the EOT is the same, which emphasizes the need for incorporating penetration effects in simulation, particularly in MOSFETs with high-K dielectric material. Effects of substrate doping density on gate capacitance is found to be negligible if penetration effects are incorporated. It may be claimed that for accurate modeling of device properties of future generation MOSFETs where high-K dielectric materials will play a significant role, wavefunction penetration effects must be incorporated.

6.2   Suggestion for further work

To employ an advanced gate dielectric, good interface quality is highly desirable. The interfaces with either the gate or the silicon region are particularly important in regard to device performance. These regions serve as a transition between the atoms associated with the materials in the gate electrode, gate dielectric and silicon channel. Even after very careful fabrication it is quite natural to exist a SiO_2 layer in between dielectric and silicon channel. Therefore, to study actual device phenomena, this interfacial layer should be taken into account during modeling. In the present work, pure dielectric layer is studied. But it leaves the scope of studying dielectric/SiO_2 gate stack structures.

   During fabrication, owing to disruption in crystal structure, states are introduced deep in the energy gap. This results in trapped ionic states in the silicon /dielectric transition region. Therefore scattering phenomena is quite natural for tunneling electrons colliding with these states, phonons or impurities. In this modeling scattering phenomenon is not
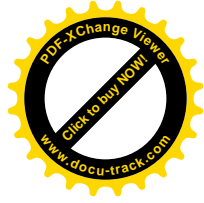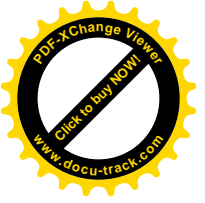
considered which leaves space for more work to be done.

In this simulation the DT current and gate capacitance is calculated for no signal condition and after inversion has started. But calculation may be done for accumulation region also.
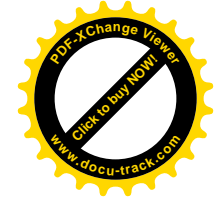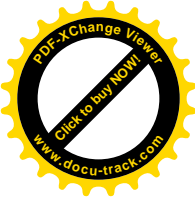
References

[1] H. S. Momose, M. Ono, T. Yoshitomi, T. Ohguro, S. I. Nakamura, M. Saito and H. Iwai, "1.5 nm direct tunneling gate oxide in Si-MOSFET's," IEEE Trans. Electron Devices, vol. 43, pp. 1233-1239, 1996.

[2] "International Technology Roadmap for Semiconductors," *http://public.itrs.net/Files/2001 ITRS/Home.htm.*

[3] S. M. Sze, *Physics of Semiconductor Devices,* Second edition, Wiley Eastern Limited, pp. 362-366, 1987.

[4] Y. Tsividis, *Operation and Modeling of the MOS Transistor,* McGraw-Hill, Ch. 2,1999.

[5] S. A. Campbell, D. C. Gilmer, X. C. Wang, M. T. Hsieh, H. S. Kim, W. L. Glacfelter and J. Yan, "MOSFET transistors fabricated with high permittivity TiO_2 dielectrics," IEEE Trans. Electron Devices, vol. 44, pp. 104-109, 1997.

[6] D. A. Buchanan, "Scaling the gate dielectric: Material, integration and reliability," IBM J. Res. Develop., vol. 43, pp. 245-264, 1999.

[7] M. Lezlinger and E. H. Snow, "Fowler-Nordheim tunneling into thermally grown SiO_2." Journal of Applied Physics, vol. 40, pp. 278-283, 1969.

[8] F. Stern and W.E. Howard, "Properties of semiconductor surface inversion layers in the electric quantum limit", Phys. Rev., vol. 163, pp. 816-835, 1967.

[9] F. Stern, "Self-consistent results for n-type Si inversion layers", Phys. Rev. B, vol. 5, pp. 4891-4899, 1972.

[10] C. Moglestue, "Self-consistent calculation of electron and hole inversion charges at silicon-silicon dioxide interfaces", Journal of Applied Physics, vol. 59, pp. 3175-3183, 1986.

[11] Y. Ohkura, "Quantum effects in Si n-MOS inversion layer at high substrate concentration", Solid-State Electron., vol. 33, pp. 1581-1585, 1990.

[12] M. J. Van Dort. P. H. Woerlee, A. J. Walker, "A simple model for quantization effects in heavily- doped silicon MOSFETs at inversion conditions," Solid-State Electron., vol. 37, pp. 411-414, 1994.

[13] F. Rana, S. Tiwari. and D. A. Buchanan, "Self-consistent modeling of accumulation layers and tunneling currents through very thin oxides," Appl. Phys. Lett.,, vol. 69, pp. 1104-1106, 1996.

[14] S. H. Lo, D. A. Buchanan. Y. Taur, W. Wang, "Quantum-mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide nMOSFET's," IEEE Electron Device Letters, vol. 18, pp. 209-211, 1997.

[15] W.-K. Shih, E. X. Wang, S. Jallepalli, F. Leon, C. M. Maziar, and A. F. Tasch Jr., "Modeling gate leakage current in nMOS structures due to tunnneling through an ultra-thin oxide," Solid-State Electron.., vol. 42, pp. 997-1006, 1998.

[16] Leonard F. Register, Elyse Roesenbaum, Kevin Yang, "Analytical model for direct tunneling current in polysilicon-gate-metal-oxide-semiconductor devices," Appl. Phys Lett, vol. 74, pp. 457-459, 1999.

[17] Nian Yang, W. Kirklel Henson, John R. Hauser, J . J. Wortman, "Modeling study of ultrathin gate oxide using direct tunneling current and capacitance-voltage measurements in MOS devices," IEEE Trans. Electron Devices, vol. 46, pp. 292-294, 1999.

[18] A. Ghetti, "Characterization of tunneling current in ultra-thin gate oxide," Solid-State Electron., vol. 44, pp. 1523-1531, 2000.

[19] Eric M. Vogel, Khaled Z. Ahmed, Brian Hornung, W. Kriklen Henson, Peter K. Mclarty, Gerry Lucovsky, John R. Hauser and Jimmie J. Wortman, "Modeled tunnel currents for high dielectric constant dielectrics," IEEE Trans. Electron Devices, vol. 45, pp. 1350-1355, 1998.

[20] Yee Chia Yeo, Qiang Lu, Wen Chin Lee, Tsu-Jae King, Chenming Hu, Xiewen Wang, Xin Guo and T. P. Ma, "Direct tunneling gate leakage current in transistors with ultrathin silicon nitride gate dielectric," IEEE Electron Device Letters, vol. 21, pp. 540-542, 2000.

[21] J. Zhang, J. S. Yuan, Y. Ma and A. S. Oates, "Design optimization of stacked layer dielectrics for minimum gate leakage currents," Solid-State Electron., vol. 44, pp. 2165-2170, 2000.

[22] Sivakumar Mudanai, Yang-Yu Fan, Qiqing Ouyang, Al. F. Tasch and Sanjay Kumar Banerjee, "Modeling of direct tunneling current through gate dielectric stacks," IEEE Trans. Electron Devices, vol. 47, pp. 1851-1857, 2000.

[23] S. Takagi and A. Toriumi, "Quantitative understanding of inversion-layer capacitance in Si-MOSFETs," IEEE Trans. Electron Devices, vol. 42, pp. 2125-2130, 1995.
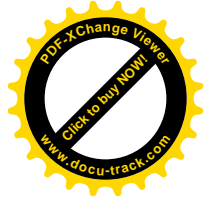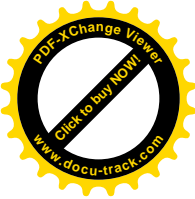
[24] S. Takagi, M. Takayanagi and A. Toriumi. "Characterization of inversion-layer capacitance of holes in Si MOSFET's," IEEE Trans. Electron Devices, vol. 46, pp. 1446-1450, 1999.

[25] S. Takagi, M. Takayanagi and A. Toriumi. "Impact of electron and hole inversion layer capacitance on low voltage operation of scaled n- and p-MOSFET's," IEEE Trans. Electron Devices, vol. 47, pp. 999-1005,2000.

[26] 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 silicon p- and n-MOSFET characteristics," IEEE Trans. Electron Devices, vol. 44, pp. 297-302, 1997.

[27] C. A. Ritcher, A. R. Hefner, E. M. Vogel, "A comparison of quantum mechanical capacitance-voltage simulators," IEEE Electron Device Letters, vol. 22, pp. 35-37, 2001.

[28] Y-C King, H. Fujioka, S. Kamohara, K. Chen and C. Hu, "DC electrical oxide thickness model for quantization of the inversion layer in MOSFET's", Semicond. Sci. Technol., vol. 13, pp. 963-966, 1998.

[29] W. Liu, X. Jin, Y. King and C. Hu, "An efficient and accurate compact model for thin-oxide-MOSFET intrinsic capacitance considering the finite charge layer thickness," IEEE Trans. Electron Devices, vol. 46, pp. 1070-1072, 1999.

[30] C. Fiegna and A. Abramo, "Analysis of quantum effects in nonuniformly doped N'IOS structures," IEEE Trans. Electron Devices, vol. 45, pp. 877-880, 1998.

[31] A. Haque, A. Rahman and I. B. Chowdhury, "On the use of appropriate boundary conditions to calculate the normalized wave functions in the inversion layers of MOSFETs with ultra-thin gate oxides," Solid-State Electron.,vol.44, pp.1833-1836, 2000.

[32] S. Mudanai, L. F. Register, A. F. Tasch and S. K. Banarjee, "Understanding the effects of wave function penetration on the inversion layer capacitance of NMOSFETs," IEEE Electron Device Letters, vol. 22, pp.145-147, 2001.

[33] A. D. Serra, A. Abramo, P. Palestri, L. selmi and F. Widdershoven, "Closed-and open-boundary models for gate-current calculation in n-MOSFETs," IEEE Trans. Electron Devices, Vol. 48, pp. 1811-1815, 2001.

[34] J. Wang, Y. Ma, L. Tian and Z. Li, "Modified Airy function method for modeling of direct tunneling current in metal-oxide-semiconductor structures," Appl. Phys. Lett., Vol. 79, pp. 1831-1833, 2001.

[35] M. Z. Kauser, M. S. Hasan and A. Haque, "Effects of wave function penetration into the gate oxide on self-consistent modeling of scaled MOSEFTs," IEEE Trans. Electron Devices, to be published, March 2002.
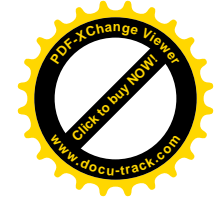
[36]   A. Haque and A. N. Khondker, "An efficient technique to calculate the normalized wave functions in arbitrary one-dimensional quantum well structures," Journal of Applied Physics, Vol. 84, pp. 5802-5804, 1998.

[37] A. Haque and A. N. Khondker, "On the conductance and conductivity of disordered quantum wires", Journal of Applied Physics, vol. 80, pp. 3876-3880, 1996.

[38]   G. D. Wilk, R. M. Wallace and J. M. Anthony, "High-K dielectrics: Current status and materials properties considerations," Journal of Applied Physics, Vol. 89, pp. 5243-5275, 2001.

[39] A. Rahman and A. Haque, "A study into the broadening of quantized inversion layer states in deep submicron MOSFETs," Solid-State Electron., vol. 45, pp. 755-760, 2001.

[40] M. Zahed Kauser, "Effects of wavefunction penetration into the gate oxide on self-consistent modeling of deep submicron MOSFETs," M.Sc. Thesis, Bangladesh University of Engineering and Techmology, Department of Electrical and Electronic Engineering, 2001.

[41] Khairul Alam, "Self-consistent calculation of direct tunneling gate leakage current in deep submicron n- and p-MOSFETs in presence of inelastic scattering," M.Sc. Thesis, Bangladesh University of Engineering and Techmology, Department of Electrical and Electronic Engineering, 2002.

[42] A. N. Khondker, M. Rezwan Khan and A. F. M. Anwar, "Transmission line analogy of resonance tunneling phenomena: the generalized impedance concept," Journal of Applied Physics, vol. 63, pp. 5191-5193, 1988.