MODELLING OF DIRECT TUNNELLING GATE LEAKAGE CURRENT THROUGH HIGH-k/SiO_2 GATE STACKS

Thesis submitted in partial fulfillment of the requirements
for the award of the degree of

Master of Technology

In

VLSI Design

Submitted by

Shrutika.Satyanarayana

Roll. No. 601261026

Under the supervision of

Mr. Arun Kumar Chatterjee

Assistant Professor, ECED

Department of Electronics & Communication Engineering

Thapar University, Patiala

INDIA

July, 2014

CERTIFICATE

I hereby declare that the work which is being presented in the thesis entitled, **"Modelling of direct tunnelling gate leakage current through high-k/SiO_2 gate stacks"** in partial fulfillment of the requirement for the award of degree of Master of Technology (VLSI Design) at the electronics and Communication Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of Mr. Arun Kumar Chatterjee, Assistant Professor, ECED.

The matter presented in this thesis has not been submitted in any other University/Institute for the award of any other degree.

Date: 14/7/2014

Shrutika.Satyanarayana

Roll.No : 601261026

It is certified that the above statement made by the student is correct to the best of my knowledge and belief.

Arun Kumar Chatterjee 14/7/14

Mr. Arun Kumar Chatterjee

Assistant Professor

ECED,Thapar University

Counter signed by:

Dr. Sanjay Sharma **Dr. S. K. Mahopatra**

ACKNOWLEDGEMENT

I take this opportunity to express my profound sense of gratitude and respect to all those who helped me through the duration of this thesis. I acknowledge with gratitude and humility my indebtedness to **Mr. Arun Kumar Chatterjee, Assistant Professor,** Electronics and Communication Engineering Department, Thapar University, Patiala, under whose guidance I had the privilege to complete this thesis. I wish to express my deep gratitude towards him for providing individual guidance and support throughout the thesis work.

I convey my sincere thanks to **Head of the Department, Dr. Sanjay Sharma** as well as **PG Coordinator, Dr.Kulbir Singh, Associate Professor, ECED**, entire faculty and staff of Electronics and Communication Engineering Department for their encouragement and cooperation.

Finally and above everyone else, my heartfelt thanks and gratitude goes to my parents and sister for their constant support and encouragement. I am also thankful to god who bestowed upon ability and strength in me to complete this work.

Shrutika.Satyanarayana

ABSTRACT

The scaling down of device dimensions of MOSFETs is accompanied by a decrease in the gate-oxide thickness and an increase in the substrate doping density. When the gate-oxide thickness becomes less than 2 nm, a substantial amount of current flows through the oxide due to Direct tunnelling.

High-k dielectrics are being extensively studied to replace silicon dioxide as they allow for higher physical thicknesses of oxides (EOT). In order to maintain a good interface between the gate-oxide and the channel and to prevent mobility degradation, there must be a thin interfacial layer of silicon dioxide between the bulk and high-k dielectric which calls for the introduction of gate stacks. In this work, an analytical model has been developed for gate tunnelling through two layer gate stacks. The results obtained from this model show good agreement with the simulation results obtained using Silvacco (Atlas) tool. The effect of variation of interfacial layer thickness and capping layer thickness on the tunnelling current density has been studied. The study reveals that both these thicknesses must be kept as thin as possible for the gate leakage current density to be minimum.

Direct tunnelling through single layer dielectrics has been well studied and analytically modelled for single dielectric layer. The results obtained from this model show good agreement with the experimental results. The impact of variation of gate-oxide thickness and k-value of dielectric on the tunnelling current density has been studied. The tunnelling current density is inversely proportional to the gate-oxide thickness as well as the k-value of the dielectric.

The tunnelling current decreases by an order of magnitude of approximately 10^4 by using gate stacks instead of using only silicon dioxide as dielectric. The polysilicon gate must be replaced by metal gate so as to prevent the formation of interface states between the top gate and high-k layer. An analytical model has been presented for the case of metal gate also.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

MOSFET	Metal oxide semiconductor field effect transistor
CMOS	Complementary Metal oxide semiconductor field effect transistor
NMOS	N-channel Metal oxide semiconductor field effect transistor
PMOS	P-channel Metal oxide semiconductor field effect transistor
SCE	Short channel effect
DIBL	Drain induced barrier lowering
GIDL	Gate induced drain leakage
FN	Fowler Nordheim
SOI	Silicon on insulator
PDSOI	Partially depleted silicon on insulator
FDSOI	Fully depleted silicon on insulator
DGSOI	Double gate silicon on insulator
DGFET	Double gate field effect transistor
BOX	Burried oxide
ITRS	International technology roadmap for semiconductors
IC	Integrated circuit
WKB	Wentzel-Kramers-Brillouin
ECB	Electron tunnelling from conduction band
EVB	Electron tunnelling from valence band
HVB	Hole tunnelling from valence band
JVD	Jet vapour deposition
EOT	Equivalent oxide thickness

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 TRANSISTOR SCALING

Over the past decades, there has been continuous scaling down of the MOS transistor. The typical channel lengths of MOSFETs were once several micrometers, but the channel lengths of MOSFETs incorporated in modern integrated circuits are in tens of nanometres. The term scaling refers to the reduction in device dimensions. All vertical and horizontal transistor dimensions are scaled by the scaling factor s>1. The minimum feature size and the fabrication technology determine the extent of scaling that is achievable. The pace for MOSFET development is set by the "ITRS" which is a roadmap maintained by the semiconductor industry [1]. The fabrication process of semiconductor devices, the need to use extremely low voltages, and poorer electrical performance pose a hindrance to the down scaling of transistor dimensions, thus, necessitating circuit redesign and innovation.

REASONS FOR SCALING

- The key reason for scaling down transistors is to accommodate large number of devices in a given chip area. As a result, the chip exhibits the same functionality in a smaller area, or chips exhibit more functionality in the same area [2].
- Since, the cost of fabrication for a semiconductor wafer is almost fixed, the cost per IC is associated with the no of chips that can be produced per wafer. Hence, more chips can be produced per wafer by using smaller ICs. This reduces the price per chip.
- Smaller transistors switch faster. Width, length and oxide thickness are the main device dimensions that are scaled down .The scaling down of these dimensions by an equal factor results in the scaling down of the gate capacitance as well. This in turn scales down the RC delay of transistor and speed increases [3].

1.2 MOORE'S LAW

Moore's law describes a long-term trend in which the number of transistors that can be incorporated on an integrated circuit has doubled approximately every eighteen months.

This trend was described by Intel co-founder Gordon E. Moore, in his 1965 paper. The paper brought into notice the fact that the number of components placed on integrated circuits had doubled every year since the invention of the integrated circuit in 1958 until 1965 and also predicted that this trend would continue "for at least ten years". His prediction has proved to be exceptionally accurate and is now used in the semiconductor industry to set targets for research and development and to guide long term planning. Moore's law, has undoubtedly been the guiding principle for the continuous reduction of CMOS dimensions since the time Gordon Moore first predicted it in 1965. There has been a scaling down of CMOS devices to the sub-100nm regime over a period of few decades. The "Moore's Law" is a concise description of the continuous, periodic increase in the level of miniaturization [4].

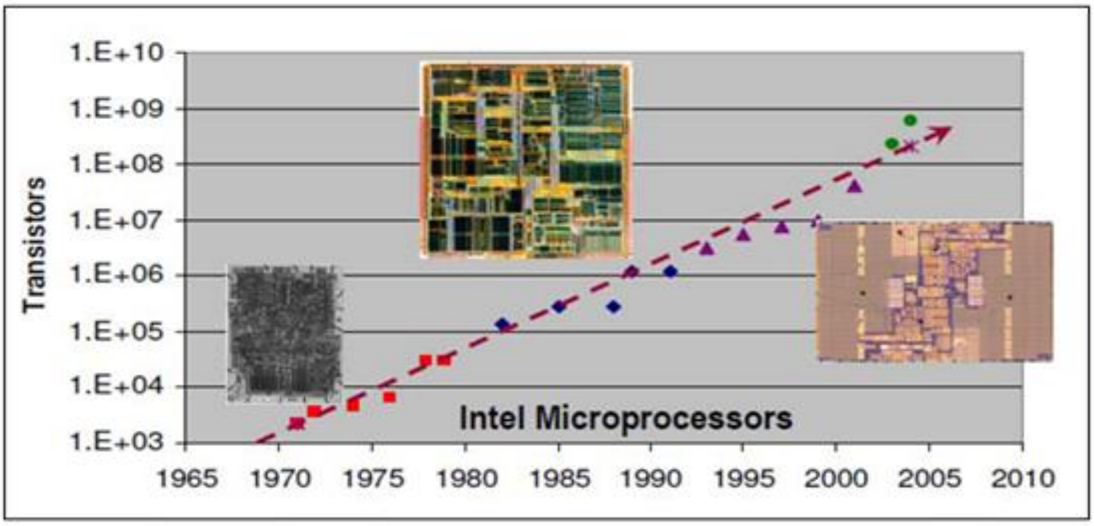

Figure 1.1: Moore's law

1.3 SHORT CHANNEL EFFECTS

A MOSFET device is considered to be short when the channel length is of the same order of magnitude as the depletion layer widths of source and drain junctions. In order to increase the operation speed as well as the number of components per chip, the channel length is reduced. This gives rise to Short channel effects [5]. The various short channel effects are as follows:

1.3.1 SUBTHRESHOLD LEAKAGE

In order to reduce internal electric fields and power consumption, low power supply levels are required by short channel transistors. Maintenance of performance is guaranteed by commensurate scaling of the transistor threshold voltage. When the gate voltage is below the threshold voltage, sub-threshold or weak inversion current flows between drain and source in a MOS transistor.

There is an exponential relation between gate voltage and the drain current. On a semilog plot of I_d versus V_g, this relation is a straight line. The equation for sub-threshold current is

$$I_{ds} = \mu_o c_{ox} \left(\frac{w}{l}\right) * (m-1) * v_t^2 * \left[e^{\left(v_g - v_{th}/m v_t\right)}\right] * \left[1 - e^{(-v_{ds}/v_t)}\right] \qquad (1)$$

where,

$m=1+C_{dm}/C_{ox}=1+3t_{ox}/ w_{dm}$

V_{th}: threshold voltage

V_t : KT/q (thermal voltage)

C_{ox}: gate oxide capacitance

μ_0: zero bias mobility

m: subthreshold swing

w_{dm}: maximum depletion layer width

1.3.2 DRAIN INDUCED BARRIER LOWERING

This occurs when the source and drain depletion regions interact with each other and lower the source potential barrier near the channel surface. Carriers are then injected into the channel surface by the source. DIBL becomes more severe at shorter channel lengths and high drain voltages. DIBL can be reduced by higher surface and channel doping and shallow source/drain junction depths. The higher doping reduces the depletion widths of source and drain and prevents them from interacting with each other.

1.3.3 PUNCH THROUGH

Extremely low substrate doping results in even larger depletion depths and an increase in the penetration of drain electric field through the channel to the source, thus, making the DIBL effect stronger. As a result, barrier to the carriers decreases and more and more carriers enter the channel leading to a drastic increase in the current. The current is so high that it becomes impossible to turn off the device. Punchthrough is of two types [6]:

- SURFACE PUNCHTHROGH: In this the current flows through the surface.
- BULK PUNCHTHROUGH: In this the current follows the subsurface path.

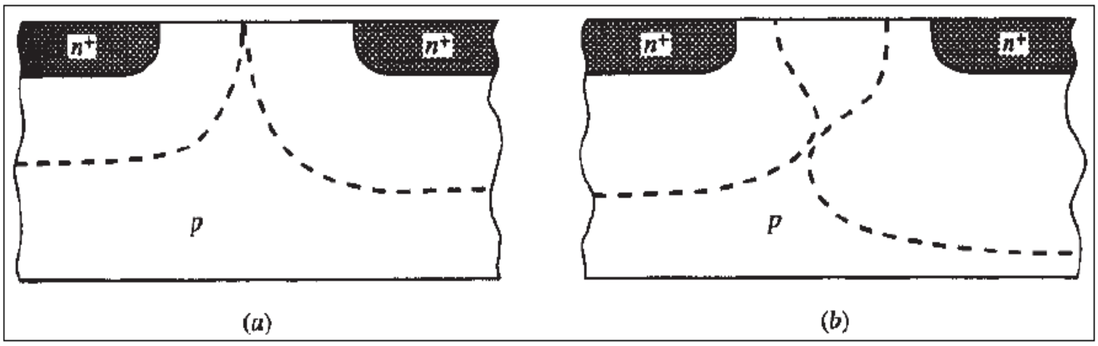

Figure 1.2: Surface and Bulk Punchthrough

1.3.4 HOT CARRIER INJECTON INTO THE GATE OXIDE

In a short channel transistor, the electric field is very high near the Si-SiO_2 interface, due to which electrons or holes can gain sufficient energy and enter into the oxide layer by crossing the interface potential barrier. This effect is called hot-carrier injection. The probability of the injection from Si to SiO_2 is more for electrons than holes, as the effective mass of electrons is lower than that of holes, and the barrier height for holes (4.5eV) is more than that for electrons (3.1eV).

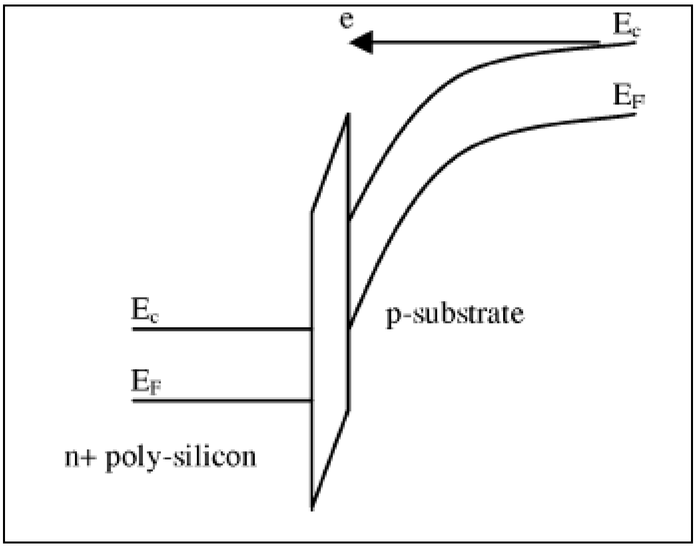

Figure 1.3: Injection of hot carriers from substrate to oxide

1.3.5 GATE INDUCED DRAIN LEAKAGE

High field effect in the drain junction of a MOS transistor gives rise to GIDL. When a negative bias is applied to the gate, there is accumulation of holes at the surface due to which the surface behaves like a p region more heavily doped than the substrate. As a result of this, the depletion layer at the surface is much narrower than elsewhere. This narrowing of the depletion layer at the surface causes an increase in the local electric field, which in turn enhances the high field effects near that region. An even larger negative bias causes the n+ drain region under the gate to be depleted and even inverted. This leads to a peak field increase, resulting in a drastic increase of high field effects such as Band-to-Band tunnelling and avalanche multiplication. These effects lead to the emission of minority carriers in the drain region lying underneath the gate. The minority carriers are swept laterally to the substrate as the substrate is at a lower potential, thus, completing a path for the GIDL [5].

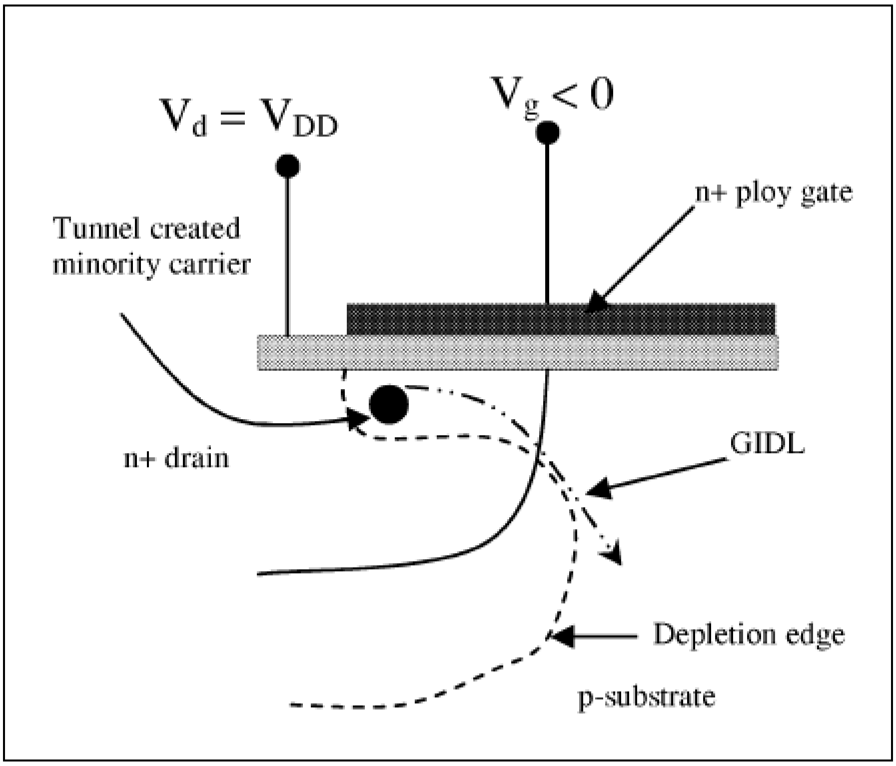

Figure 1.4: Gate Induced Drain Leakage

1.3.6 TUNNELING INTO THE GATE OXIDE

There is an increase in the field across the oxide as the gate oxide thickness is reduced. This high electric field along with low oxide thickness leads to tunnelling of electrons from substrate to gate and also from gate to substrate through the gate oxide, giving rise to gate oxide tunnelling current. The tunnelling mechanism between gate and substrate can be mainly divided into two parts, namely: (1) *Fowler–Nordheim tunneling*; and

(2) *Direct tunnelling*. Electrons tunnel through a triangular potential barrier in the case of F-N tunnelling. On the other hand, electrons tunnel through a trapezoidal potential barrier in the case of Direct tunnelling.

1.4 ADVANCED MOSFET STRUCTURES

SOI technology gives a good alternative to silicon CMOS processes which are reaching their limit in terms of fabrication and device miniaturization. CMOS processes can be scaled to ultimate limits by employing SOI technology. Silicon-on-insulator technology offers many advantages over bulk silicon CMOS processing such as higher speed, lower power dissipation, lower parasitic capacitance, improved short channel effect immunity, high sub-threshold voltage swing etc.

1.4.1 SILICON ON INSULATOR MOSFETS

SOI MOSFETS can be categorized into Partially depleted and Fully depleted MOSFETs based upon the state of the depletion layer [7].

1.4.1.1 PARTIALLY DEPLETED SOI MOSFET

In partially depleted state, silicon film thickness is large, so there is no interaction between the depletion regions from the front and back ends. A piece of neutral silicon lies beneath the front depletion region. If this body is left electrically floating, two parasitic effects will come into picture, the first one is Kink effect or Floating Body effect, and second on is the presence of parasitic open base NPN bipolar transistor between source and drain. A kink is seen in the output characteristics of an SOI MOSFET due to the kink effect. Due to high drain to source voltage the electrons at the drain end create the electron-hole pairs, due to impact ionization, and are collected in the floating body.

1.4.1.2 FULLY DEPLETED SOI MOSFET

In fully depleted state, the silicon film thickness is small. Due to this the depletion regions from front and back ends meet. There is electrostatic coupling between front and back channels during device operation. Fully depleted devices are practically free of kink effect [8].

1.4.2 DOUBLE GATE METAL-OXIDE-SEMICONDUCTOR FIELD EFFECT TRANSISTOR

A double-gate-silicon-on-insulator DGSOI Structure consists, of a silicon slab sandwiched between two oxide layers. Each oxide is contacted by a metal or a polysilicon film. If an appropriate bias is applied an inversion region is formed at the Si-SiO_2 interface at the front as well as back gate. Thus we would have two MOSFETs sharing the substrate, source, and drain [8]. The most notable feature of the DGFET is that it controls short-channel effects by device geometry instead of doping as in the case of bulk FET [10]. This feature provides DGFET the advantages such as 1) reduced 2D short channel effects that allow for a shorter channel length compared to bulk FET; 2) a sharper sub-threshold slope (60mV/dec compared to >80mV/dec for bulk FET) and 3) better carrier transport due to decrease in channel doping . The most common mode of operation of the DGFET is to switch the two gates simultaneously.

1.4.3 GROUND PLANE SOI MOSFET

A ground plane can be formed in the silicon substrate underneath the buried oxide (BOX) in order to prevent the electric field lines from the drain from penetrating into the channel region. The heavily doped electric field stop can be placed in the substrate either under the channel region or underneath the boundary between channel and source/drain. This field stop improves short channel effect immunity and sub-threshold slope [11].

1.4.4 FINFET

The FinFET is a non-planar, double-gate transistor which is built on an SOI substrate. The main distinguishing characteristic of the FinFET is that the conducting channel is wrapped by a thin silicon "fin", which forms the body of the device. The effective channel length of the device is determined by the thickness of the fin (measured in the direction from source to drain) [7].

1.4.5 TRI-GATE TRANSISTOR

In these transistors there is a single gate stacked on top of two vertical gates. This allows for three times the surface area for electrons to travel. Tri-gate transistors reduce leakage and consume far less power than current transistors. The additional gate enables the transistor to switch very quickly between the ON and OFF state [7].

1.5 TRANSISTOR MODELLING

The reliable operation of circuits employing transistors can be ensured by carefully modelling the physical phenomena observed in their operation using transistor models. There exists a wide range of models that vary in purpose and complexity. Transistor models can be divided into two main categories: Models for device design and Models for circuit design [2].

MODELS FOR DEVICE DESIGN

The design of devices entails an in depth understanding of how device manufacturing processes such as impurity diffusion, ion implantation, oxidation, annealing, and etching effect device behaviour. The various manufacturing steps are simulated by the process models which provide a description of device geometry to the device simulator. Geometry does not only refer to the geometrical features but also to the details inside the structure such as the doping profiles after the completion of device processing. Once the device simulator has enough information about what the device looks like, it models the physical processes that take place in the device so as to determine its electrical behaviour in a number of circumstances such as: DC current-voltage behaviour, transient behaviour for both large signal as well as small signal, dependence on device layout.

MODELS FOR CIRCUIT DESIGN (COMPACT MODELS)

Almost all modern electronic design work employs transistor models. Spice which is an analog circuit simulator uses models to predict the behaviour of a design. Most design work is related to integrated circuit designs. The models for circuit design can be divided into two categories: Large-signal nonlinear models and small signal linear models.

LARGE SIGNAL NON-LINEAR MODELS

Non-linear or large signal transistor models are mainly of three types:

- PHYSICAL MODELS

These models are based upon device physics and on the modelling of physical phenomena that occur within a transistor. Parameters used by these models are based upon physical properties such as substrate doping concentrations, oxide thicknesses, carrier mobility etc. These models were used extensively in the past but they have now proved to be inadequate for quantitative design due to the complexity of modern devices.

- EMPERICAL MODELS

These models do not directly model the device physics but are entirely based upon curve fitting, using parameter values that best fit the measured data. The parameters in an empirical model do not have any fundamental basis and will depend only on the fitting procedure used to find them. The fitting procedure is the heart of these models and is the key to their success.

- TABULAR MODELS

A tabular model is a form of look-up table that contains large number of values for common device parameters such as device parasitic and drain current. The indexing of these values is done in reference to their bias voltage combinations. The biggest advantage of this model is decreased simulation time. A limitation of these models is that they work well only for designs which use devices within the table and are not reliable for devices outside the table.

- SMALL-SIGNAL LINEAR MODELS

Small-signal or linear models are used to evaluate gain, noise, stability and bandwidth. A small-signal model can be generated by taking derivatives of the current-voltage curves about a bias point or Q-point. A big advantage of small signal models over large signal nonlinear models is that they can be solved directly, while large signal nonlinear models are solved iteratively, with possible convergence issues. Another advantage is that linear models are easier.

1.6 NEED FOR MODELLING

As we are heading towards higher CMOS technologies, gate oxide thicknesses are becoming extremely thin leading to the dominance of direct tunnelling gate leakage current. The thickness of SiO_2 layer presently used as gate dielectric is becoming so thin (less than 2nm) that the gate leakage current due to direct tunnelling of electrons through SiO_2 will be very high exceeding 1A/cm^2 at 1V. Thus, high-k dielectrics are being studied extensively so as to replace SiO_2 as they allow for higher physical thicknesses. In order to maintain a good interface between the gate oxide and channel and to prevent mobility degradation, it is desirable to have a thin interfacial layer of SiO_2 between the bulk and high-k dielectric. This calls for the introduction of high-k gate stacks and the modelling of direct tunnelling gate leakage current in order to design these stacks.

In this thesis work, an analytical model has been developed for the determination of direct tunnelling gate leakage current through high-k/SiO_2 gate stacks and the effect of various parameters on the gate leakage current density has been analyzed.

1.7 THESIS ORGANIZATION

This thesis comprises of five chapters out of which the first chapter is introduction. A brief description of work done by various researchers throughout the world has been given in second chapter. The third chapter includes the basic concepts related to tunnelling and the various models that have been proposed for the determination of direct tunnelling gate leakage current through ultra-thin gate oxides. The fourth chapter includes the proposed model that has been developed for the determination of direct tunnelling gate leakage current through high-k/SiO_2 gate stacks. Finally, the fifth chapter gives the conclusion and future scope.

CHAPTER 2

LITERATURE SURVEY

- **Kaushik Roy et al. , 2003 [5]**

  The continuous scaling of channel length, gate oxide thickness, and threshold voltage in the deep-submicrometer regime leads to high leakage current which is becoming a serious concern because of the high power dissipation it leads to. Consequently, it is necessary to identify different leakage components and to carefully model them so as to reduce the leakage power. This paper reviews the different transistor intrinsic leakage mechanisms, including sub-threshold leakage, drain-induced barrier lowering, punchthrough, gate-induced drain leakage, hot carrier injection into the oxide, and gate oxide tunnelling. The paper also discusses channel engineering techniques including halo doping and retrograde doping as a means to ameliorate the short-channel effects for continuous scaling down of CMOS devices. Finally, the paper explains various circuit-level techniques to reduce leakage power.

- **Vishwas Jaju , 2004 [8] , Anurag Chaudhry et al. , 2004 [12]**

  These papers explain the issues related to silicon-on-insulator technology. SOI technology gives a good alternative to bulk CMOS processes which are reaching their limits in terms of device miniaturization and fabrication. Different types of SOI MOSFETs have been presented and the related physical concepts have been evaluated. The advantages of SOI MOSFETs have been discussed with the stress on Double Gate MOSFETs. The technological challenges in realizing this new device structure have also been presented. Double Gate MOSFETs provide excellent short channel effect immunity and exhibit a near ideal sub-threshold slope which make them the ultimate scalable device structure. Different structures such as Thin Body FD SOI with raised source and drain, Halo Doped SOI MOSFET, Ground Plane FDSOI MOSFET, Multiple Gate SOI MOSFET have also been discussed.

- **Shunpei Abe et al. , 2009 [13]**

  The root cause of short channel effects in SOI MOSFETs is Drain-Induced barrier lowering at SOI/buried oxide interface. This paper highlights the role of the thickness and permittivity of the buried oxide layer. It analyzes the effect of varying the BOX thickness and permittivity by performing numerical device simulations of SOI MOSFETs and also by visualizing the distribution of current flow lines, dielectric flux lines and contour potential lines in MOSFETs. It is demonstrated that smaller values of BOX thickness and permittivity are effective in suppressing DIBL in deca-nano SOI MOSFETs.

- **FZ Rahou, A.Guen-Bouazza and M.Rahou, 2013 [14]**

  In this paper the electrical characteristics of fully-depleted and partially depleted SOI MOSFET devices have been investigated and compared. The comparisons are focussed on four important electrical characteristics namely threshold voltage, subthreshold voltage, leakage current and kink effect. The device structures are constructed and the characteristics are observed using Silvaco-Atlas. The Results obtained show that the electrical characteristics of fully-depleted SOI devices are better than that of partially depleted SOI devices. In this paper it is concluded that FDSOI MOSFET has lower leakage current than PD SOI n-MOSFET. The main drawback in PDSOI n-MOSFET is kink effect which is eliminated in FDSOI n-MOSFET.

- **Tetsu Tanaka et al., 1991 [15]**

  This paper presents the fabrication and analysis of planar double-gate thin–film SOI MOSFETs. Wafer bonding has been used to fabricate planar p^+ poly Silicon Double-gate thin-film SOI nMOSFETs. The paper confirms that the conduction in double-gate SOI MOSFET stems from a fully flat potential as well as charge injection. Finally, the paper proposes an analytical model that displays electrical characteristics that agree well the characteristics of fabricated devices.

- **Vinay Kumar Yadav, Ashwani K.Rana, 2012 [9]**

  This paper studies the impact of channel engineering on double gate MOSFET by using different channel doping. It is observed in the results that it is possible to adjust the threshold voltage by changing the channel doping. The impact of channel engineering is also observed on performance parameters of the DG-MOSFET such as on current, off current, drain induced barrier lowering, sub-threshold slope and carrier mobility. Thus, an optimized value of the channel doping is desired to be projected for future reference in context of leakage power. The paper concludes that as we scale down the devices, threshold voltage of the device decreases, so to adjust the threshold voltage and other short channel effects within the permissible limits channel engineering can be done. But the channel doping cannot be increased beyond a certain limit because some parameters like subthreshold swing and drain induced barrier lowering, both are opposite in nature. If channel doping is increased, one parameter gets improved while the other gets worse. So, there is a need to optimize their value at particular doping level.

- **K.F Schuegraf et al., 1992 [16]**

  For thinner oxide films, FN tunnelling is not adequate to accurately model the oxide leakage currents. In this paper, modifications have been made to the expression for FN tunnelling to obtain a closed form expression for direct tunnelling. This expression has been derived using WKB and other approximations. The conduction characteristics for insulator layers thinner than 6nm can be modelled accurately using this expression. The paper also takes into consideration the effect of polysilicon depletion. It is concluded that the practical limit for SiO_2 scaling is 4nm.

- **Wen-Chin Lee et al., 2001 [17]**

  In this paper, a semi-empirical model has been proposed in order to quantify the gate tunnelling currents through ultrathin oxides. The simple analytical model is not absolutely accurate and leads to inaccuracies as it includes a number of approximations. A correction function is introduced as a multiplier to this simple model to achieve applicability to a variety of experimental conditions including electron/hole injection from inversion and accumulation layers and different bias polarities. This new model can predict all the current components: electron tunnelling from valence band (EVB), electron tunnelling from conduction band (ECB) and hole tunnelling from valence band (HVB) taking into account polysilicon gate depletion and quantization effects. In addition, this model can also be employed for the determination of physical oxide thickness from I-V data. This model is highly sensitive in ultra-thin oxide regime, where due to the presence of large tunnelling currents C-V extraction happens to be impossible.

- **Yee Chia Yeo et al., 2000 [18]**

  This paper presents a semi-empirical model for the determination of direct-tunnelling leakage current in N-type and P-type MOSFETs with ultrathin Si_3N_4 as the gate dielectric. The Si_3N_4 has been formed by jet-vapour deposition (JVD) technique. Important material parameters such as electron/hole tunnelling effective masses and barrier potentials have been extracted from the measured data. The model is also used to predict the scaling limits of JVD Si_3N_4 gate dielectric based on supply voltages for a number of technology nodes. The paper projects that the thickness of JVD Si_3N_4 can be scaled down to 1.13nm and 0.65nm for low-power and high-performance applications respectively before the tunnelling current exceeds the maximum tolerable limit. Important differences between the parameters of Si_3N_4 and SiO_2 have been highlighted .The most important difference is that the scaling limit for JVD Si_3N_4 will be reached first for PMOSFET, contrary to what has been observed for SiO_2.

- **Yee-Chia Yeo et al., 2003 [19]**

  In this paper, the scaling limits of alternative gate dielectrics have been explored based on their direct-tunnelling characteristics. Material parameters such as tunnelling effective mass have been extracted from the tunnelling characteristics of promising high-k dielectrics. A figure-of-merit has been introduced in order to compare the relative advantages of various high-k dielectrics. The paper also proposes a method for predicting the scaling limits of different gate dielectric candidates. Finally, the paper provides guidelines for the selection of gate dielectrics in order to satisfy the off-state leakage current requirements that have been projected for future low power and high-performance technologies. The paper suggests that Si_3N_4 or SiO_xN_y will be the suitable candidates for high-performance and low-power applications through 2016.

- **J.Robertson , 2004 [20]**

  The scaling of CMOS transistors has led to the silicon dioxide dielectric layer becoming very thin (1.4nm) which in turn has led to extremely large leakage current. It is necessary to replace the silicon dioxide dielectric with a physically thicker layer of 'high-k' gate oxide. This paper basically covers the choice of oxides, their electronic structure, their structural and metallurgical behaviour, their bonding, atomic diffusion, their deposition, interface structure, band offsets, flat band voltage shifts, mobility degradation, and electronic defects. The six main conditions that the new oxides must satisfy in order to be accepted as gate dielectrics are: a high enough k value, relatively high band offset, good interface quality with Si, kinetic stability, thermal stability, and low bulk defect density. There is a need to optimise the oxides further in order to achieve high performance devices. For this the flat band voltage must be improved and the defect densities must be lowered. The root cause of flat band voltage shift is the interface defects. The main defects in the oxides are interstitials and oxygen vacancies. The oxygen vacancies pose a huge problem as they lead to defect levels near the Silicon conduction band. The paper concludes that HfO_2 and Hf silicate are the most suitable candidates for replacing silicon dioxide.

- **Mohan V. Dunga et al., 2003 [21]**

  This paper presents a physical model for gate tunnelling current through multi-layer gate dielectrics. The model extends over a wide range of voltages covering direct tunnelling as well as Fowler Nordheim tunnelling regimes. This model shows that the introduction of High-k dielectric into the gate may lead to a poor performance at certain operating conditions. The paper studies the trade-off between barrier height and dielectric constant. It is a necessary requirement for future high-k dielectrics to have a relatively high barrier height in order to retain their efficacy. It is shown that tunnelling from the gate material is smaller as compared to tunnelling from the substrate. With increase in the dielectric constant, this difference will also increase and restrictions on leakage will arise due to tunnelling from the substrate side.

- **Thomas Kauerauf et al., 2005 [23]**

  The main aim of this paper is to identify the combination of stack architecture and high-k dielectric material with the lowest leakage current. In the first part of the work, the leakage current through high-k/ SiO_2 stacks with various equivalent oxide thicknesses, materials and gate electrodes is calculated assuming tunnelling only. The paper discusses the difference between gate injection and substrate injection and shows the impact of interfacial layer thickness, k-value, barrier height, and work function of the gate material on the tunnelling current. In the second part, the universal relation between barrier height and k-value is considered, and the material which best suits all the requirements is introduced. The leakage current is calculated for different values of EOTs and it is demonstrated that if all the dielectrics follow this universal relation, for a given thickness there exists one material, which gives a minimum value of gate leakage current. Finally, the paper concludes that even for devices with EOTs less than 1nm, the k-value must not exceed 25 and HfO_2 or ZrO_2 are the most suitable candidates if only the leakage current issues are considered.

- **Ghader Darbandy et al., 2011 [22]**

  This paper presents an analytical model for modelling of direct tunnelling leakage currents through gate stacks with SiO_2 as an interfacial layer. This model is based on the modified WKB tunnelling probability approximation. The simple analytical model is not absolutely accurate and leads to inaccuracies as it includes a number of approximations. A correction function is introduced as a multiplier to this simple model to achieve applicability to a variety of experimental conditions. This new model can predict all the current components: electron tunnelling from valence band (EVB), electron tunnelling from conduction band (ECB) and hole tunnelling from valence band (HVB) taking into account polysilicon gate depletion and quantization effects. The paper provides guidelines for the determination of a suitable high-k candidate in the case of high-k/ SiO_2 gate stack. Using the proposed model, the most promising high-k materials are predicted for different conditions, taking into consideration the effects of electron effective mass, equivalent oxide thickness, value of permittivity of high-k dielectric material, barrier height and thickness of interfacial layer. The paper predicts that materials such as Hfo_2, La_2O_3 and Pr_2O_3 would successfully fulfil the considered requirements.

CHAPTER 3

BASIC CONCEPTS OF TUNNELLING

3.1 GATE OXIDE TUNNELING PHENOMENON

As MOSFETs are being scaled down to the nanoscale regime, it becomes necessary to consider quantum mechanical effects in the design and modelling of MOSFETs. In present day CMOS technology, the thickness of gate oxide of a MOSFET is less than 1.5 nm, and the channel doping is as high as $10^{18}\ cm^{-3}$. For MOSFETs with ultrathin oxides and heavily doped channels, the electric field in the oxide can reach very high values of MV/cm. This ultrathin oxide layer decreases the width of the barrier that separates the gate from the channel, thus making the injection of electrons/holes easier. This direct tunnelling gate current could be the major source of leakage in the device, leading to faulty operation of circuits and an increase in standby power consumption in the MOSFET. In MOSFETs with ultra thin oxides, the gate voltage applied plays a big role in the type of leakage current. Larger the value of the applied gate voltage, larger the electric field generated in the substrate will be. As a result, more energy will be given to the carriers present in the substrate, thus increasing the probability of tunnelling. Depending on the magnitude of gate voltage there can be two types of tunnelling. One is the Fowler Nordheim (FN) tunnelling, and the second is Direct tunnelling.

3.1.1 FN TUNNELING IN A MOSFET

At lower gate voltages, FN tunnelling takes place in which electrons tunnel from the conduction band of Silicon substrate to the conduction band of polysilicon gate through the conduction band of SiO_2.In this case, the oxide conduction band is triangular in nature as shown in Figure 3.1. FN tunnelling is valid for $V_{ox} > \phi_b$ where V_{ox} is the voltage drop across the oxide and ϕ_b is the Si/SiO_2 barrier height [5].

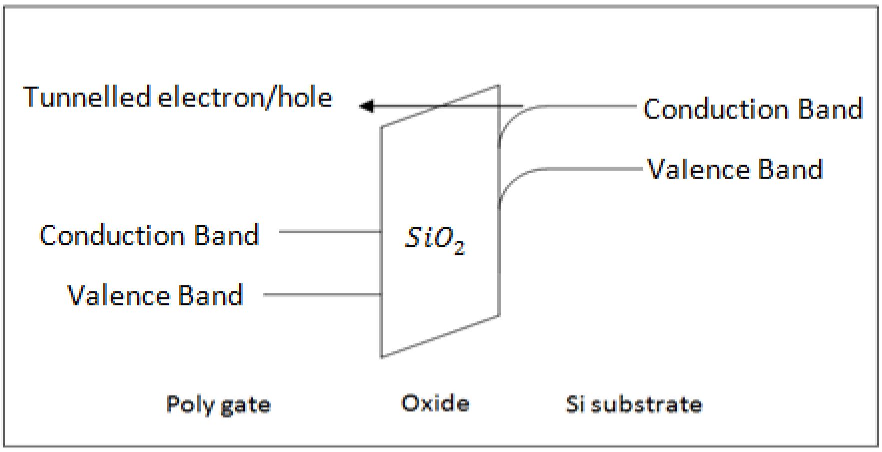

Figure 3.1: FN Tunnelling in a MOSFET

3.1.2 DIRECT TUNNELLING IN A MOSFET

In the case of ultrathin oxide MOSFET, the electric field across the ultrathin gate oxide is so high that direct tunnelling of electrons takes place which is impossible to explain using classical physics. This tunnelling is more pronounced in thin oxides, e.g. 4 nm or less. The electrons tunnel directly to the gate through the forbidden energy gap of the oxide. In this case, the oxide conduction band is trapezoidal in nature as shown in Figure 3.2. Direct tunnelling is valid for $V_{ox} < \phi_b$. It can be modelled only by using quantum mechanical theory. In the gate oxide, the net current is the summation of the FN tunnelling and the direct tunnelling. Usually, the FN current is very small and can be neglected. So the net current contribution is from direct tunnelling only [5].

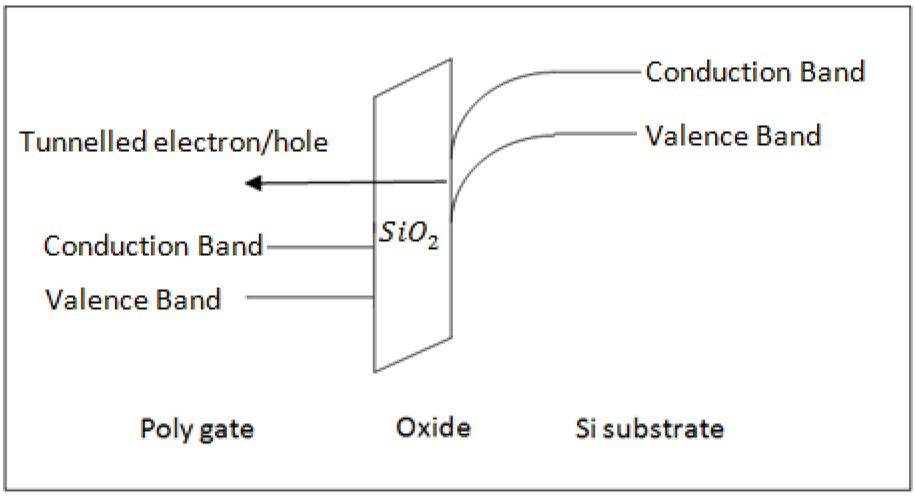

Figure 3.2: Direct Tunnelling in a MOSFET

- MECHANISMS OF DIRECT TUNNELLING

In MOS devices, there are three major mechanisms for direct tunnelling namely: electron tunnelling from conduction band (ECB), electron tunnelling from valence band (EVB), and hole tunnelling from valance band (HVB). In NMOS, the gate to channel tunnelling current is controlled by ECB in inversion, whereas gate-to-body tunnelling is controlled by EVB in depletion-inversion and ECB in accumulation. In PMOS, the gate to Channel tunnelling current is controlled by HVB in inversion, whereas gate-to-body leakage is controlled by EVB in depletion-inversion and ECB in accumulation [25].

- COMPONENTS OF DIRECT TUNNELLING

The five major components of direct tunnelling are gate to inverted channel current, part of which goes to the source and the rest goes to the drain; parasitic leakage current through gate–to-S/D extension overlap region and the gate to the substrate leakage current.

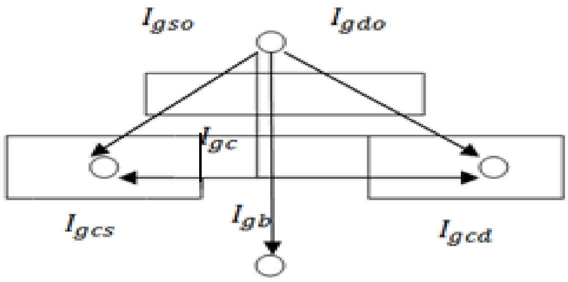

Figure 3.3: Components of Direct Tunnelling [5]

3.2 IMPACT OF GATE OXIDE TUNNELLING

According to the tunnelling theory, the magnitude of tunnelling probability is determined by the width of the potential barrier. In MOSFETs, the gate dielectric acts as the potential barrier separating the carriers in the channel from the gate. When the gate dielectric layer is thick, the carriers are unable to penetrate the potential barrier and extend to the gate. However, when the gate dielectric layer is thin, direct tunnelling will be increased exponentially, leading to significant power dissipation and deterioration in device performance. The gate oxide leakage can subsequently lead to circuit failures as the designs assume that there is no significant gate current. The reduction in the channel current due to the gate oxide leakage results in low drain currents and hence the circuit operation at low power is hampered drastically. The direct tunnelling in the gate oxide is more harmful in MOSFETs in the deep submicron and nanometer regime since the density of inversion charge is reduced due to the quantization of energy bands at the Si/SiO_2 interface.

3.3 EXISTING MODELS FOR DETERMINATION OF DIRECT TUNNELING CURRENT THROUGH ULTRATHIN GATE OXIDES

3.3.1 SIMPLE MODEL

The expression for FN tunnelling current density is given as [16]:

$$J = AE_{ox}^2 e^{\frac{-B}{E_{ox}}} \tag{1}$$

Where, $B = \frac{8\pi\sqrt{2m_{ox}}}{3hq}\phi_b^{\frac{3}{2}}$

The expression for direct tunnelling current density can be obtained by multiplying the

FN expression by the factor: $\frac{1}{\left[1-\left(\frac{\phi_b - V_{ox}}{\phi_b}\right)^{\frac{1}{2}}\right]^2} e^{\frac{B\left(\frac{\phi_b - V_{ox}}{\phi_b}\right)^{\frac{3}{2}}}{E_{ox}}}$

Therefore, the direct tunnelling current density can be expressed as:

$$J_{n=} A\left(\frac{\phi_B}{V_{ox}}\right)\left(\frac{2\phi_B}{V_{ox}} - 1\right)E_{ox}{}^2 \exp\left(-\frac{B}{E_{ox}}\left[1-\left(1-\frac{V_{ox}}{\phi_B}\right)^{\frac{3}{2}}\right]\right) \tag{2}$$

Where

$A = (q^3/8\pi h\phi_B)$;

$B = \left(\frac{8\pi\sqrt{2m_{ox}}\,\phi_B^{\frac{3}{2}}}{3hq}\right)$;

$E_{ox} = V_{ox}/t_{ox}$;

t_{ox}: Oxide thickness;

m_{ox}: Effective mass in the oxide;

The above expression incorporates a number of approximations that lead to inaccuracies. The finite density of electrons or energy states in the semiconductor needs to be taken into account. The effective mass has been assumed to be constant for all energies (all locations at any gate bias and oxide thickness) which also leads to inaccuracy. Also, there must be a consideration of the quantization effects that arise in the semiconductor due to small oxide thicknesses in order to obtain the oxide potentials as a function of the gate voltage. Because of these reasons, Schuegraf's model is unable to fit the tunnelling current for the entire range of gate voltage. There is a need of a correction function to overcome the above mentioned second-order effects.

3.3.2 IMPROVED MODEL

This model can be expressed as [17]:

$$J_n = \frac{q^3}{8\pi h \phi_B \varepsilon_{ox}} C(V_g, V_{ox}, t_{ox}, \phi_B) \exp\left(- \frac{8\pi\sqrt{2m_{ox}}\phi_B^{\frac{3}{2}}}{3qh|E_{ox}|}\left[1 - \left(1 - \frac{V_{ox}}{\phi_B}\right)^{\frac{3}{2}}\right]\right) \qquad (3)$$

$C(V_g, V_{ox}, t_{ox}, \phi_B)$ has been obtained by replacing $\varepsilon_{ox}\left(\frac{\phi_B}{V_{ox}}\right)(\frac{2\phi_B}{V_{ox}} - 1)E_{ox}^2$ term in equation (2).

It has been developed by empirical fitting and is given as:

$$C\left(V_g, V_{ox}, t_{ox}, \phi_B\right) = \exp\left[\frac{20}{\phi_B}\left(\frac{(|V_{ox}| - \phi_B)}{\phi_{BO}} + 1\right)^{\alpha}\left(1 - \frac{|V_{ox}|}{\phi_B}\right)\right] * \left(\frac{V_g}{t_{ox}}\right) * N \qquad (4)$$

where

α: Fitting parameter which depends on the tunnelling process

ϕ_{BO}: Si/SiO_2 barrier height (e.g. 3.1eV for electron and 4.5eV for hole).

ϕ_B: Actual tunnelling barrier height (e.g. 3.1eV for ECB, 4.2eV for EVB, and 4.5eV for HVB with silicon electrode).

Most of the secondary effects which were neglected in Scheugraf's derivation are covered in the exponential term of the correction function. The V_g term in the correction function forces the tunnelling current to be zero when the gate voltage i.e. V_g is zero. Best fit is provided by $\alpha's$ of 0.6,1 and 0.4 for ECB, EVB and HVB, respectively.

I. N for ECB and HVB

N represents the carrier density in the inversion and accumulation layer of the injecting electrode. Both for ECB as well as HVB tunnelling, N in inversion and accumulation regimes is given as:

$$N = \frac{\varepsilon_{ox}}{t_{ox}}\left\{n_{inv} V_t . ln\left[1 + exp\left((\frac{V_{ge} - V_{th}}{n_{inv} V_t}\right)\right] + n_{acc} V_t . ln\left[1 + exp\left[-\left(\frac{V_g - V_{FB}}{n_{acc} V_t}\right)\right]\right]\right\} \qquad (5)$$

where

n_{inv} and n_{acc} : represent swing parameters

V_{th}: threshold voltage

V_{FB}: flatband voltage

V_t: thermal voltage

V_{ge}: effective gate voltage ($V_g - V_{poly}$)

The default values of n_{inv} and n_{acc} :

n_{inv}: S/V_t

S: subthreshold swing and is given as follows [3]

S=2.3*m*V_t

m = 1+$\frac{3t_{ox}}{W_{dm}}$

n_{acc}: 1

II. N for EVB

For EVB tunnelling process N can be given as:

$$N = \frac{\varepsilon_{ox}}{t_{ox}} \left\{ n_{EVB} V_t . ln \left[1 + exp \left(\frac{|V_{ox}| - \phi_g}{n_{EVB} V_t} \right) \right] \right\} \quad (6)$$

where

ϕ_g: E_g/q by default and n_{EVB} is a fitting parameter with a default value of 3.

III. Value of V_{ge}

It is possible to derive V_{ge} by solving the poisson equation in the polysilicon gate under depletion approximation. It is given as [26]:

$$V_{ge} = V_{FB} + \phi_{so} + \frac{q\varepsilon_{si} N_{poly}\, t_{ox}^2}{\varepsilon_{ox}^2} \left(\sqrt{1 + \frac{2\varepsilon_{ox}^2 (V_g - V_{FB} - \phi_{so})}{q\varepsilon_{si} N_{poly}\, t_{ox}^2}} - 1 \right) \quad (7)$$

where

V_{FB}: Flatband voltage

ε_{si} and ε_{ox} : Dielectric constants of Si and SiO_2 respectively

N_{poly} : Polysilicon gate doping concentration

ϕ_{so}: $2\phi_f$ (ϕ_f is Fermi potential)

IV. Value of V_{ox}

$$V_{ox} = V_{ge} - V_{FB} - \phi_s \quad (8)$$

$$\phi_s = \left[\frac{\gamma}{2} \left(-1 + \sqrt{1 + \frac{4(V_g - V_{ge} - V_{FB})}{\gamma^2}} \right) \right]^2 \quad (9)$$

$$\gamma = \frac{\sqrt{2\epsilon_{si} q N_{sub}}}{C_{ox}} \quad (10)$$

where

ϕ_s : Surface band bending of substrate

N_{sub} : Channel doping concentration

γ: Body-effect parameter

3.4 HIGH-k OXIDES

The scaling of CMOS transistors has led to the silicon dioxide dielectric layer becoming very thin (1.4nm) which in turn has led to extremely large leakage current. It is necessary to replace the silicon dioxide dielectric with a physically thicker layer of 'high-k' gate oxide [20].

There are six main requirements which the new oxide must fulfil:

1. It must have a high value of k.
2. The oxide must be thermodynamically stable with the Silicon channel beneath it.
3. It must be kinetically stable.
4. Its band offsets with silicon must be greater than 1eV in order to reduce carrier injection into its bands.
5. It must have a few electrically active defects.
6. It must form a good electrical interface with silicon.

3.4.1 k –VALUE

The first requirement that the oxide must fulfil is that its k value must be greater than 10; preferably 25-30. There has to be a trade off between the value of k and the band offset condition. Table 3.1 shows that there is an inverse relationship between the k value and band gap of candidate oxides.

Table 3.1: Dielectric constant, Experimental band gap and conduction band offset on Silicon of the candidate gate dielectrics.

Oxide	k	Gap (eV)	CB offset (eV)
Si		1.1	
SiO_2	3.9	9	3.2
Si_3N_4	7	5.3	2.4
Al_2O_3	9	8.8	2.8
Ta_2O_3	22	4.4	0.35
TiO_2	80	3.5	0
$SrTiO_3$	2000	3.2	0
ZrO_2	25	5.8	1.5
HfO_2	25	5.8	1.4
La_2O_3	3	6	2.3
Y_2O_3	15	6	2.3

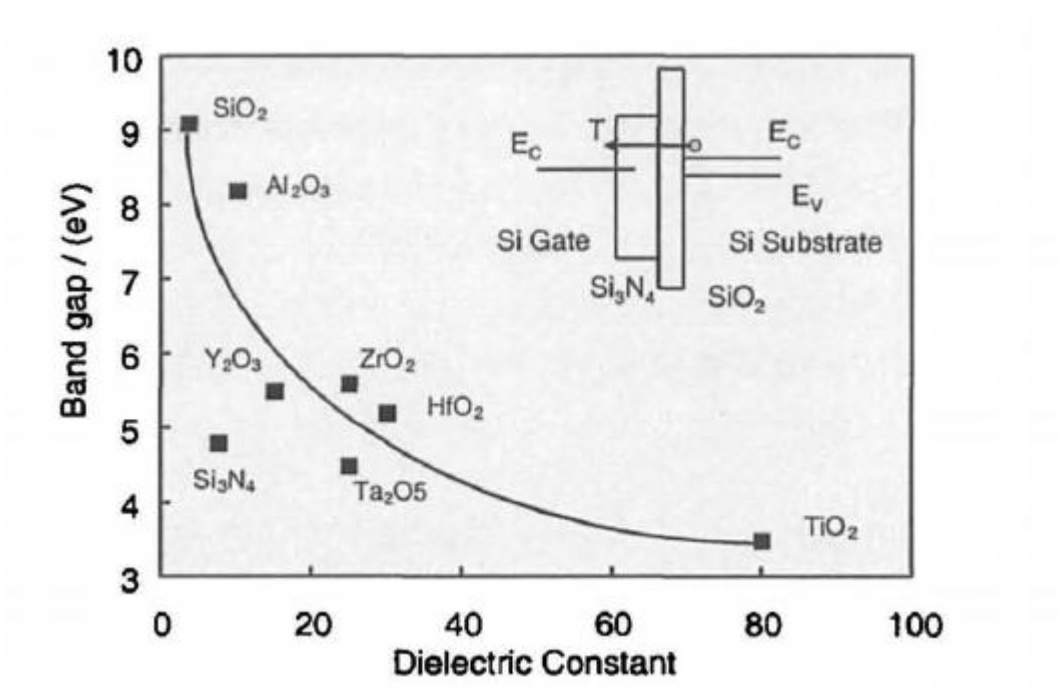

Figure 3.4: Bandgap versus dielectric constant for insulators [20]

3.4.2 BAND OFFSET

The potential barrier at each band of the high-k oxide must be greater than 1 eV in order to prevent the injection of carriers into the bands. The band gap of SiO_2 is 9 eV, so it has high barriers for both electrons and holes. However, if the oxide has a narrow band gap like $SrTiO_3$ which is 3.3eV, then its bands must be aligned with those of Silicon for both the barriers to be greater than 1eV. The conduction band offset is usually smaller than the valence band offset. The various oxides which satisfy this requirement are Al_2O_3,HfO_2, ZrO_2 , La_2O_3, Y_2O_3 etc.

3.4.3 NEED OF GATE-STACK

In order to maintain a good interface between the gate oxide and the channel and to prevent mobility degradation, there must be a thin interfacial layer of SiO_2 between the bulk and the high-k dielectric. Hence, there is a need for the introduction of gate stacks [21].

3.4.4 THE INTERFACIAL LAYER

The series capacitance formula for an overall EOT for a layer 1 of silicon dioxide and layer 2 of high-k oxide is given by [20]:

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2}$$

Therefore, $EOT = t_{SiO_2} + EOT_{hi_k}$

The SiO_2 layer is introduced because it improves the quality of the interface between Silicon and the oxide. The interface between Silicon and SiO_2 is of high quality and is well understood. It can be made with a low defect concentration and it is possible to passivate the defects by gas annealing. The presence of a SiO_2 layer also separates the Si channel from the high-k oxide, which can help to stop the reduction in mobility of carriers that the high-k oxides can lead to. Excessive amounts of trapped charges are present in high-k oxides. Mobility degradation is caused due to scattering by these trapped charges.

CHAPTER 4

MODELLING OF DIRECT TUNNELLING GATE LEAKAGE CURENT THROUGH TWO-LAYER GATE STACKS

4.1 DOUBLE GATE MOSFET STRUCTURE

The double gate MOSFET under analysis is shown in Figure 4.1.It consists of two thin interfacial layers of SiO_2 and two thick high-k layers [22].

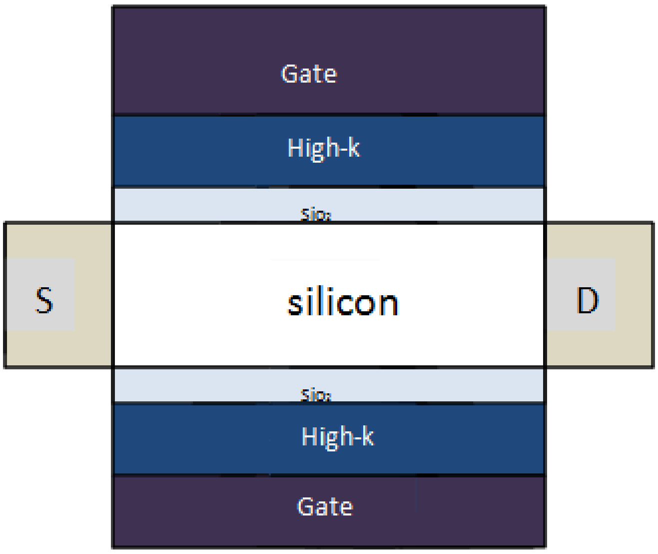

Figure 4.1: Double gate MOSFET structure under analysis

4.2 POTENTIAL PROFILE

The Potential profile which will be analyzed in the analytical model is shown in Figure 4.2. ϕ_{B1} is the Si/SiO_2 barrier height for the carriers and ϕ_{B2} is the Si/high-k barrier height for the carriers. Direct tunnelling takes place through the entire stack only if the oxide voltage is lower than ϕ_{B1} as well as ϕ_{B2} [22].

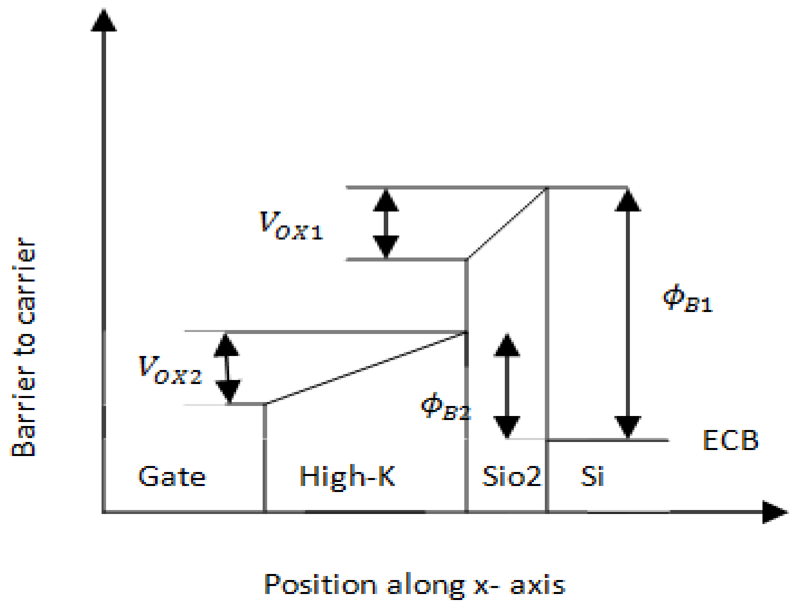

Figure 4.2: Potential profile under analysis

4.3 TUNNELLING PROBABILITY

The tunnelling current models are based on Wentzel-Kramers-Brillouin (WKB) approximation which is used to calculate the tunnelling probability of a carrier through an arbitrary barrier shape [21][24].

Tunneling probability (T) is given as:

$$T = Trans * \exp(-\alpha . t_{oxe}) \qquad (1)$$

where

t_{oxe} : effective oxide thickness

Trans: transmission factor

α : loss factor and is given as:

$$\alpha = \frac{4}{3\hbar}\frac{t_{oxe}}{V_{ox}}\left(\frac{K_1}{K_{ox}}\sqrt{2m_1 q}\left[\phi_{B1}^{3/2} - (\phi_{B1} - fV_{ox})^{3/2}\right] + \frac{K_2}{K_{ox}}\sqrt{2m_2 q}\left[(\phi_{B2} - fV_{ox})^{\frac{3}{2}} - (\phi_{B2} - V_{ox})^{3/2}\right]\right) \qquad (2)$$

So, the tunnelling probability for the two layer gate stack, assuming direct tunnelling through the entire stack is given by:

$$T = \exp\left(-\frac{8\pi tox}{3hvox\,k_2}k_1 d_2 + \quad k_2 d_1\sqrt{2m_1 q}\left[\phi_{B1}^{\frac{3}{2}} - (\phi_{B1} - fvox)^{\frac{3}{2}}\right]\right)$$

$$* \exp\left(-\frac{8\pi tox}{3hvox\,k_1}k_1 d_2 + k_2 d_1\sqrt{2m_2 q}\left[(\phi_{B2} - fV_{ox})^{3/2} - (\phi_{B2} - V_{ox})^{3/2}\right]\right) \quad (3)$$

where

m_1: Effective mass of carriers in the interfacial layer

m_2: Effective mass of carriers in the high-k layer

d_1 : Thickness of interfacial layer

d_2: Thickness of high-k layer

$\phi_{B1:}$ Si/SiO_2 barrier height

$\phi_{B2:}$ Si/high-k barrier height

<u>4.4 SIMPLE MODEL</u>

A simplified direct tunnelling model for SiO_2 as an interfacial layer and different high-k materials for a two layer stack can be formulated as follows [22]:

$$J = \frac{q^3}{8\pi h\phi_{B1}}\left(\frac{\phi_{B1}}{V_{ox1}}\right)\left(\frac{2\phi_{B1}}{V_{ox1}} - 1\right)\frac{V_{ox1}}{d_1}$$

$$* \exp\left(-\frac{8\pi tox}{3hvox\,k_2}k_1 d_2 + \quad k_2 d_1\sqrt{2m_1 q}\left[\phi_{B1}^{\frac{3}{2}} - (\phi_{B1} - fV_{ox})^{\frac{3}{2}}\right]\right)$$

$$* \exp\left(-\frac{8\pi tox}{3hvox\,k_1}k_1 d_2 + k_2 d_1\sqrt{2m_2 q}\left[(\phi_{B2} - fV_{ox})^{3/2} - (\phi_{B2} - V_{ox})^{3/2}\right]\right) \quad (4)$$

where V_{ox1} is the voltage across the interfacial layer.

The applied gate voltage will drop over the interfacial layer as well as the high-k layer. The distribution of voltage depends on the physical thickness and the k-values.

The voltage across the i[th] dielectric is given by [4]:

$$V_{oxi} = \frac{\frac{d_i}{\varepsilon_i}}{\sum_{j=1}^{N}\frac{d_j}{\varepsilon_j}}\,V_{ox} \quad (5)$$

The above expression incorporates a number of approximations that lead to inaccuracies. The finite density of electrons or energy states in the semiconductor needs to be taken into account. The effective mass has been assumed to be constant for all energies (all locations at any gate bias and oxide thickness) which also leads to inaccuracy. Also, there must be a consideration of the quantization effects that arise in the semiconductor due to small oxide thicknesses in order to obtain the oxide potentials as a function of the gate voltage. Because of these reasons, Schuegraf's model is unable to fit the tunnelling current for the entire range of gate voltage. There is a need of a correction function to overcome the above mentioned second-order effects. This calls for the introduction of a new improved analytical model.

4.5 IMPROVED MODEL

The improved model can be formulated as:

$$J = \frac{q^3}{8\pi h\phi_{B1}}\frac{C}{\varepsilon_1} * \exp\left(-\frac{8\pi tox}{3hvox\, k_2}k_1 d_2 + k_2 d_1\sqrt{2m_1 q}\left[\phi_{B1}^{\frac{3}{2}} - (\phi_{B1} - fV_{ox})^{\frac{3}{2}}\right]\right)$$

$$* \exp\left(-\frac{8\pi tox}{3hvox\, k_1}k_1 d_2 + k_2 d_1\sqrt{2m_2 q}\left[(\phi_{B2} - fV_{ox})^{3/2} - (\phi_{B2} - V_{ox})^{3/2}\right]\right) \quad (6)$$

where

ε_1: SiO_2 Permittivity

C is correction function and has been obtained by replacing $\varepsilon_1\left(\frac{\phi_B}{V_{ox}}\right)\left(\frac{2\phi_B}{V_{ox}} - 1\right)\frac{V_{ox\,1}}{d_1}$ term in equation (4).

It has been developed by empirical fitting and is given as:

$$C = exp\left[\frac{20}{\phi_{B1}}\left(\frac{|V_{ox\,1}| - \phi_{B1}}{\phi_o} + 1\right)^{\alpha}\left(1 - \frac{|V_{ox\,1}|}{\phi_{B1}}\right)\right] * \frac{V_g}{d_1} * N \quad (7)$$

where

α: Fitting parameter which depends on the tunnelling process

ϕ_o: Si/SiO_2 barrier height (e.g. 3.1eV for electron and 4.5eV for Hole).

N: Density of carriers in inversion and accumulation layer of the injecting electrode.

- Value of V_{ge}

It is possible to derive V_{ge} by solving the poisson equation in the polysilicon gate under depletion approximation. It is given as:

$$vge = V_{FB} + \phi_S + \frac{q\varepsilon_{Si}N_{poly}\left[\sqrt{1 + \frac{2C_{hi}^2\beta^2(V_{gs} - V_{FB} - \phi_s)}{q\varepsilon_{Si}N_{poly}}} - 1\right]}{\beta^2 C_{hi}^2} \quad (8)$$

where

$$\beta = \frac{\frac{d_{hi}}{\varepsilon_{hi}}}{\frac{d_{hi}}{\varepsilon_{hi}} + \frac{d_{sio\,2}}{\varepsilon_{sio\,2}}} \quad (9)$$

$$C_{hi} = {}^{\varepsilon_{hi}}/_{d_{hi}} \quad (10)$$

• Value of V_{ox}

V_{ox} is given as

$$V_{ox} = V_{ge} - V_{FB} - \phi_s \tag{11}$$

$$\phi_s = \left[\frac{\gamma}{2}\left(-1 + \sqrt{1 + \frac{4(V_g - V_{ge} - V_{FB})}{\gamma^2}}\right)\right]^2 \tag{12}$$

$$\gamma = \frac{\sqrt{2\epsilon_{si} q N_{sub}}}{C_{ox}} \tag{13}$$

where

ϕ_s : Surface band bending of substrate

N_{sub} : Channel doping concentration

γ: Body-effect parameter

4.6 REPLACEMENT OF POLYSILICON GATE BY METAL GATE

As we replace poly-silicon gate with metal gate there is an increase in tunnelling current density. The reason for this is the increase in voltage across the oxide due to absence of poly-silicon depletion in the case of metal gate. Despite of this increase in tunnelling current density, there still exists a need to use metal gates. This is due to the presence of interface states between the top metal gate and the high-k film (especially binary oxides like HfO_2,).In the case of HfO_2, excess Si bonded to Hf can potentially create a deep level as shown in Figure 4.3.

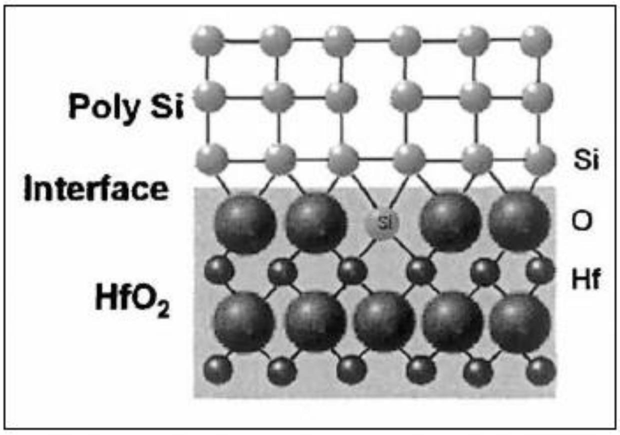

Figure 4.3: Interface states in HfO_2/Si system

A high quality interface requires less than 1% of the bonds to be incorrectly bonded. Because of this very difficult requirement at the top interface, focus has been shifted to metal top gates with high-k.

Using metal gate the equation of V_{ox} will be modified as follows:

$$V_{ox} = V_g - V_{FB} - \phi_s \qquad (14)$$

$$\phi_s = \left[\frac{\gamma}{2}\left(-1 + \sqrt{1 + \frac{4(-V_{FB})}{\gamma^2}}\right)\right]^2 \qquad (15)$$

$$V_{FB} = \phi_M - \phi_S \qquad (16)$$

In the case of metal gate, V_{ox} and ϕ_s are obtained by replacing V_{ge} in equation 11-12 by V_g since there is no polysilicon depletion.

CHAPTER 5

RESULTS AND DISCUSSION

In this thesis work,

- The Improved model for the determination of direct tunnelling gate leakage current through ultra-thin gate-oxides has been tested for a 10µm N^{+}Poly-Si–gated NMOS using SiO_2 of thickness = 2.48nm.
- The Improved model for the determination of direct tunnelling gate leakage current through high-k/SiO_2 gate stacks has been tested for a 10µm N^{+}Poly-Si–gated NMOS using a Al_2O_3/SiO_2 gate stack of EOT =2.5nm.
- The Improved model for the determination of direct tunnelling gate leakage current through high-k/ SiO_2 gate stacks has been tested for a 10µm Aluminium- gated NMOS using a Al_2O_3/SiO_2 gate stack of EOT =2.5nm.

RESULTS FOR MODELLING OF DIRECT TUNNELLING GATE LEAKAGE CURRENT THROUGH ULTRATHIN OXIDES

The improved model presented in chapter 3 has been tested for a 10µm N^{+}Poly-Si–gated NMOS having the parameters shown in Table 5.1.

Table 5.1: Parameters values for 10µm N^{+}Poly-Si–gated NMOS

S.No	Parameter	Value
1.	Channel doping concentration	4.7e17 cm^{-3}
2.	Gate doping concentration	9e19 cm^{-3}
3.	Oxide thickness	2.48 nm

Table 5.2: Model Parameters for ECB tunnelling process

S.No	Parameter	Value
1.	m_{ox}	$0.4m_o$
2.	ϕ_b	3.1 eV
3.	ϕ_{bo}	3.1 eV
4.	A	0.6
5.	n_{inv}	S/V_t
6.	n_{acc}	1

- **J_g versus V_g characteristics**

  The parameters used in the model for the calculation of J_g versus V_g characteristics are shown in Table 5.2.

  Figure 5.1 shows the calculated as well as measured J_g versus V_g characteristics in the inversion region. It can be seen that there is an excellent agreement between the calculated results as well as the experimental results [17].

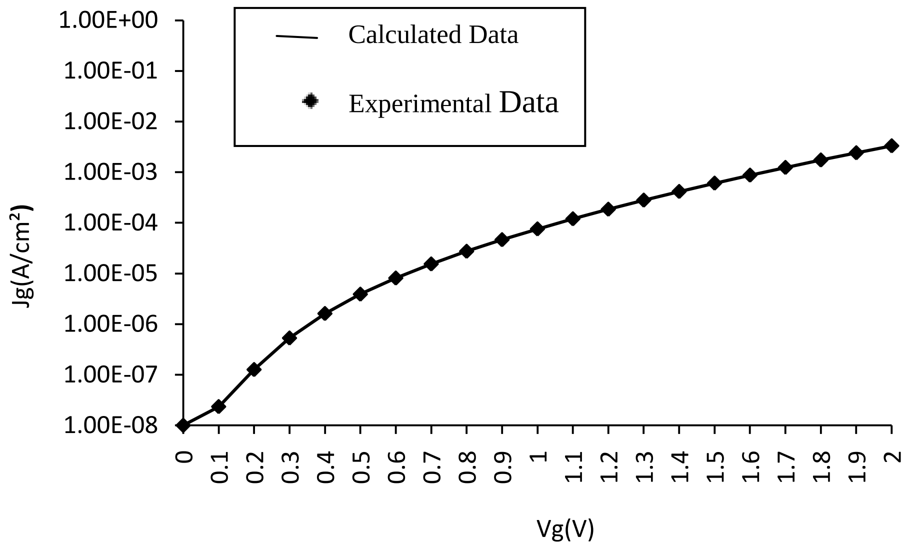

Figure 5.1: Measured (symbols) and calculated (lines) J_g versus V_g characteristics

- **Impact of variation of oxide thickness on Tunnelling current density**

Figure 5.2 shows the J_g versus V_g characteristics for various values of oxide thicknesses. Figure 5.3 shows the variation of tunnelling current density with oxide thickness at V_g=2V. It can be seen that as the value of oxide thickness increases, the tunnelling current density decreases.

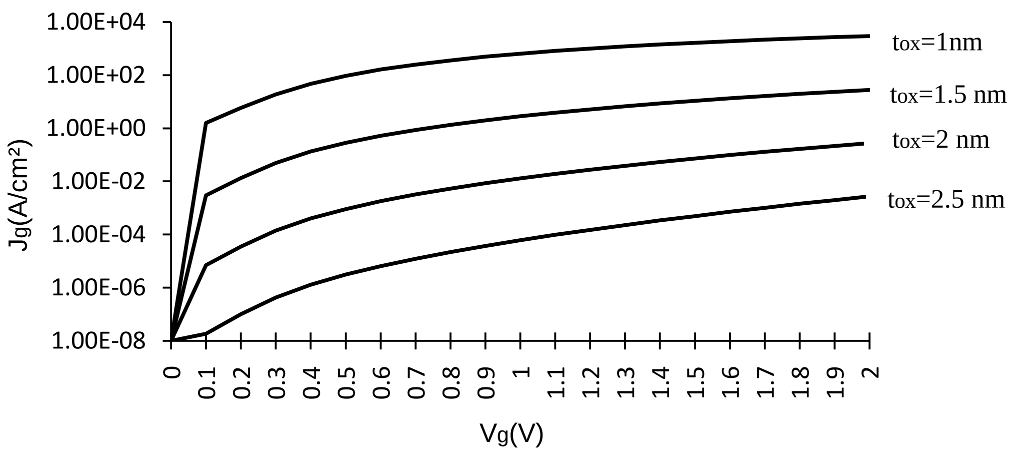

Figure 5.2: J_g versus V_g characteristics for various oxide thicknesses

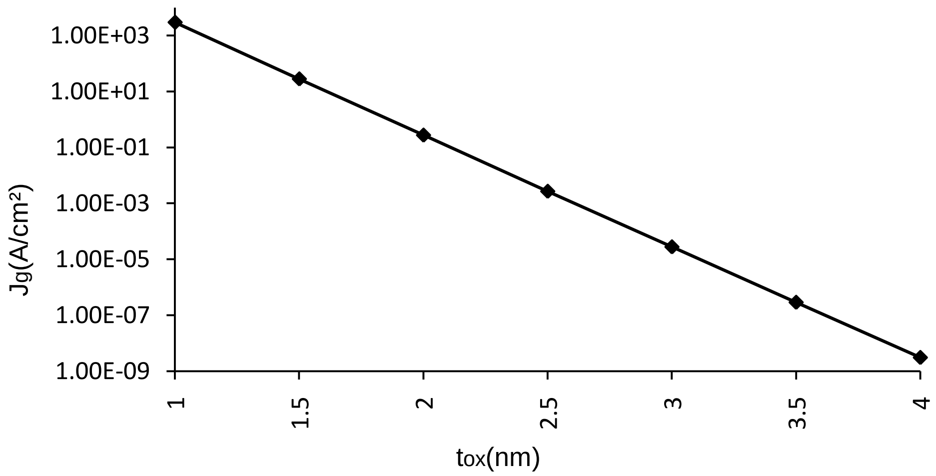

Figure 5.3: J_g versus t_{ox} characteristics at V_g=2 V

The decrease in tunnelling current density with increase in the gate-oxide thickness is attributed to the increase in the thickness of the barrier that separates the gate from the channel.

- **Impact of variation of k-value of gate oxide on Tunnelling current density**

High-k dielectrics allow for higher physical thicknesses as compared to SiO_2 and thereby reduce the gate oxide tunnelling. Figure 5.4 shows the J_g versus V_g characteristics for different k-values. Figure 5.5 shows the variation of tunnelling current density with k-value at V_g=2V. It can be seen that as the value of k increases, the tunnelling current density decreases.

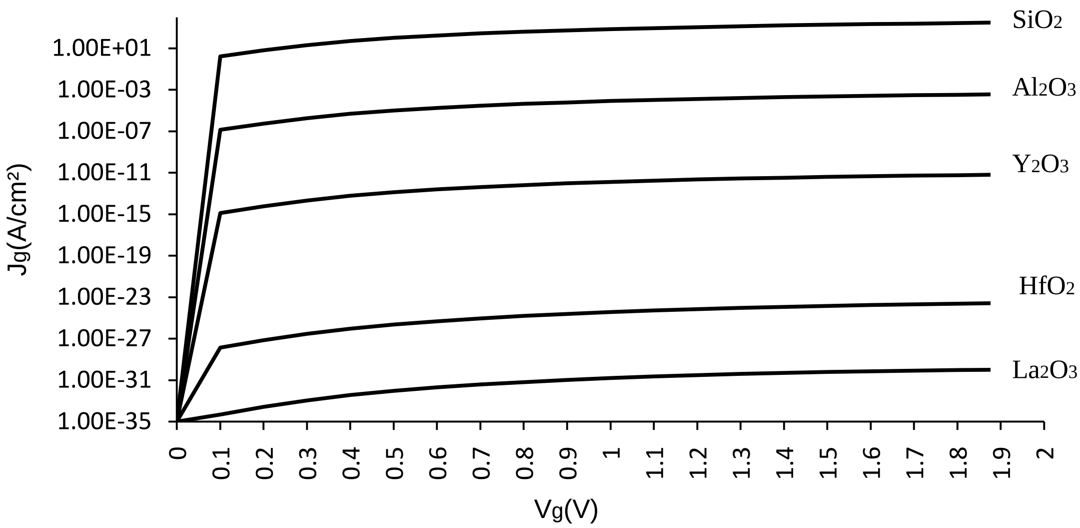

Figure 5.4**:** J_g versus V_g characteristics for various k values

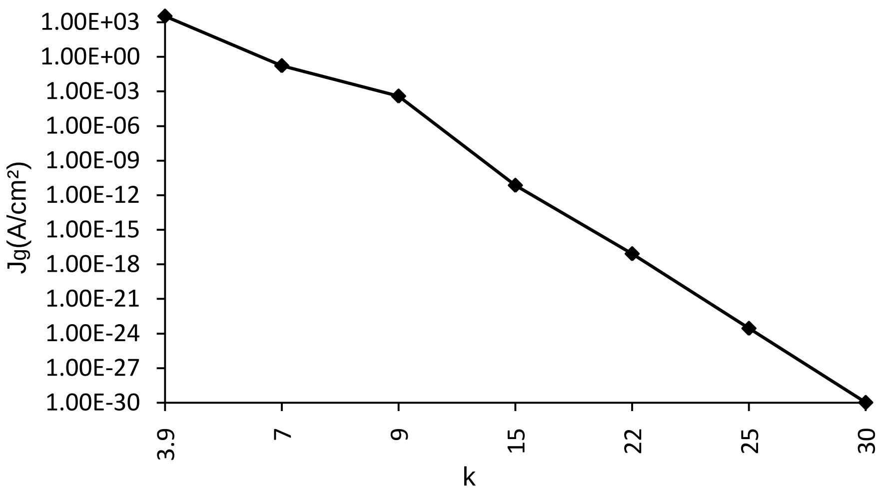

Figure 5.5:J_g versus k characteristics at V_g=2 V

The decrease in tunnelling current density with increase in the value of k is attributed to the higher physical thicknesses allowed by high-k dielectrics.

RESULTS FOR MODELLING OF DIRECT TUNNELLING GATE LEAKAGE CURRENT THROUGH HIGH-k/SiO_2 GATE STACKS

The improved model presented in chapter 4 has been tested for a 10µm N^+Poly-Si–gated NMOS having the parameters shown in Table 5.3

Table 5.3: Parameter values for 10µm N^+Poly-Si–gated NMOS

S.No	Parameter	Value
1.	Channel doping concentration	4.7e19cm^{-3}
2.	Gate doping concentration	9e19cm^{-3}
3.	EOT	2.5 nm
4.	Interfacial oxide thickness	0.1 nm

Table 5.4: Model parameters for ECB tunnelling process

S.No	Parameter	Value
1.	m_1	0.5m_o
2.	m_2	0.5m_o
3.	ϕ_{b1}	3.1 eV
4.	ϕ_{b2}	2.8 eV
5.	α	0.13

- **J_g versus V_g characteristics for Al_2O_3/SiO_2 gate stack**

The parameters used in the model for the calculation of J_g versus V_g characteristics are shown in Table 5.4.

Figure 5.6 shows the calculated as well as simulated J_g versus V_g characteristics in the inversion region. It can be seen that the simulated results match fairly with the calculated results.

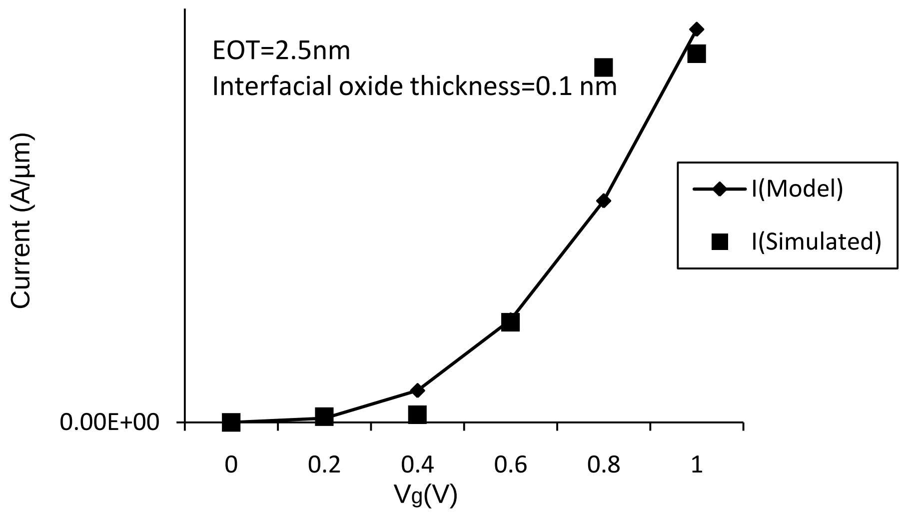

Figure 5.6: Current versus V_g characteristics

- **Impact of variation of interfacial oxide thickness on Tunnelling current density**

Figure 5.7 shows the variation of tunnelling current density with the thickness of the interfacial layer at $V_g = 1V$ and a fixed EOT of 2.5nm. It can be seen that as the interfacial layer thickness decreases, the tunnelling current density decreases. Thus, while designing a high-k gate stack from leakage point of view, the interfacial layer must be kept as thin as possible.

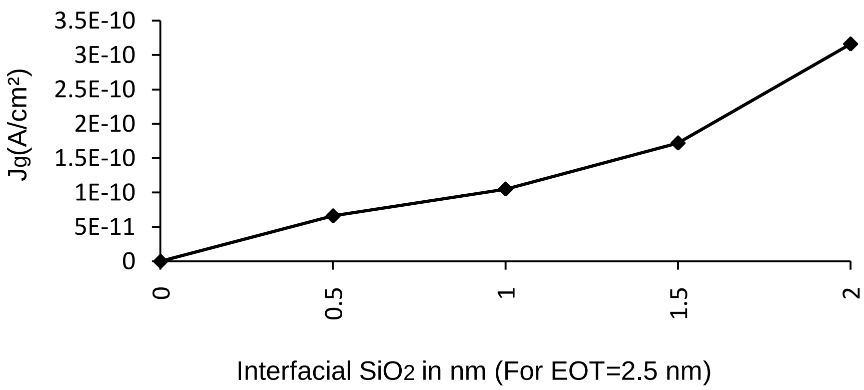

Figure 5.7:J_g versus Interfacial SiO_2 thickness at V_g=1V

- **Impact of variation of capping layer thickness on Tunnelling current density**

One method to improve the stability of high-k gate stacks especially while using poly-Si gates is the introduction of a capping layer of SiO_2 or SiON. The capping layer is introduced between the gate and the high-k dielectric to prevent any bonding to take place between the two layers as this can lead to the formation of defect sites which trap charges and lead to scattering of carriers which in turn leads to mobility degradation. Figure 5.8 shows the variation of tunnelling current density with capping layer thickness at $V_g = 1V$ and a fixed EOT of 2.5nm. It can be seen that as the thickness of the capping layer increases, the tunnelling current density increases due to the decrease in thickness of the high-k layer. Thus, while designing high-k gate stacks from leakage point of view the thickness of capping layers should be kept as thin as possible.

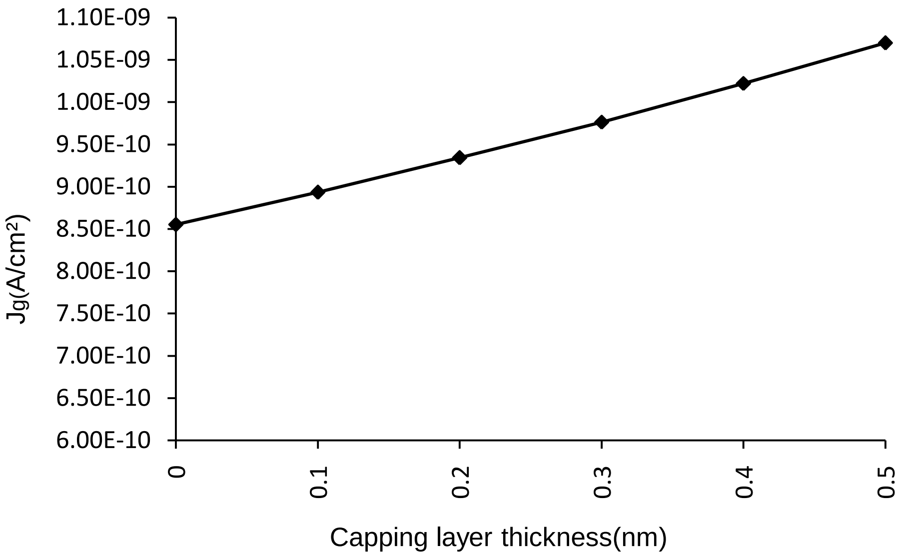

Figure 5.8:J_g versus capping layer thickness at V_g=1V

RESULTS FOR MODELLING OF DIRECT TUNNELLING GATE LEAKAGE CURRENT THROUGH HIGH-k/SiO_2 GATE STACKS USING METAL GATE

As discussed in chapter 4, there is a need to replace poly-silicon gate by metal gate. Following are the results for the same:

- **J_g versus V_g Characteristics**

  Figure 5.9 shows the J_g versus V_g Characteristics in the inversion region with Aluminium being used as the gate instead of poly-Silicon. The model parameters used for the calculation of J_g versus V_g Characteristics have been taken from Table 5.4.

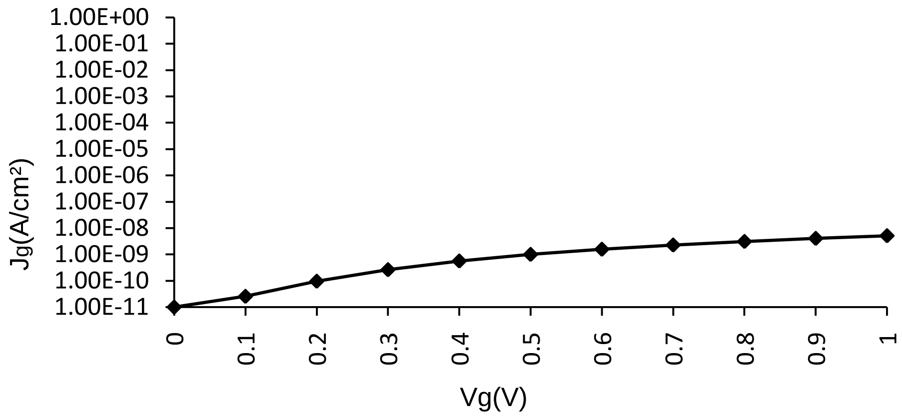

Figure 5.9: J_g versus V_g characteristics for Aluminium Gate

- **Impact of variation of Gate work function on Leakage**

Figure 5.10 shows the variation of tunnelling current density with gate work function. It can be seen that as the work function increases, the tunnelling current density decreases.

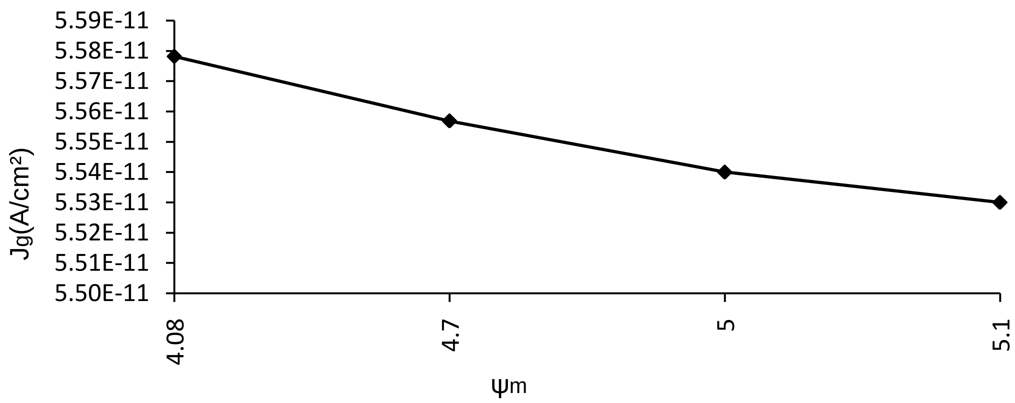

Figure 5.10: J_g versus ψ_m characteristics

<u>**COMPARISON OF TUNNELLING CURRENT DENSITIES**</u>

High-k dielectrics reduce the direct tunnelling gate leakage current as they allow for higher physical thicknesses. An interfacial layer of SiO_2 is necessary to maintain a good interface between the gate-oxide and channel and to prevent mobility degradation. High-k gate stacks reduce the direct tunnelling gate leakage current by an appreciable amount. J_g versus V_g characteristics have been plotted for a 10 µm N^+Poly-Si–gated NMOS for SiO_2 as the gate dielectric as well as for Al_2O_3/SiO_2 gate stack as shown in Figure 5.7. It can be seen that there is a decrease in the gate leakage current density by using Al_2O_3/SiO_2 gate stack by an order of magnitude of approximately 10^4.

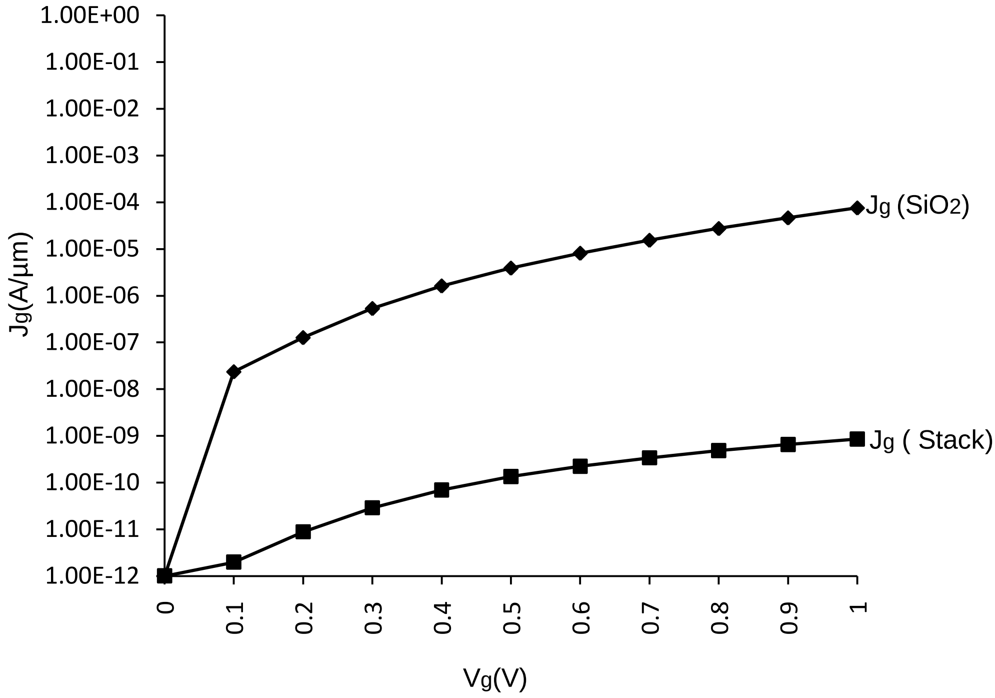

Figure 5.11: Comparison of tunnelling current densities

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

As we are heading towards higher CMOS technologies, the study of Direct tunnelling gate leakage current becomes more crucial.

- An analytical model for direct current gate leakage current through high-k/SiO_2 gate stacks has been developed. The J_g versus V_g characteristics obtained from this model have been compared with simulation results. A good agreement has been seen between the calculated and simulated characteristics. This suggests that this model is accurate and can work well when incorporated in any simulation tool.
- At a fixed value of EOT, as the interfacial oxide thickness increases, the tunnelling current density increases .This increase in current density is due to the decrease in the thickness of high-k layer with an increase in thickness of the interfacial layer. Thus, while designing any high-k gate stack from a leakage point of view, the interfacial layer thickness must be kept as thin as possible.
- At a fixed value of EOT, as the capping layer thickness increases, the tunnelling current density increases. This increase in current density is due to the decrease in thickness of high-k layer with an increase in thickness of capping layer. Thus, while designing any high-k gate stack from a leakage point of view, the capping layer thickness must be kept as thin as possible.
- It has be seen that there is a decrease in the gate leakage current density by using Al_2O_3/SiO_2 gate stack by an order of magnitude of approximately 10^4.This suggests that in order to minimize leakage, it is better to use high-k stacks rather than pure dielectrics.

FUTURE SCOPE

In this work, a semi-empirical model has been presented for the determination of direct tunnelling gate leakage current through single layer and two-layer gate stacks for polysilicon gate as well as metal gate. It can be modified further to be used for metal silicide gates. This model needs to be modified for tri-gate transistors, gate all around transistors, FinFETs and other advanced structures where many other effects such as corner effect come into picture. Quantum corrections need to be incorporated in the model in order to obtain more accurate results.

REFERENCES

[1] D.J.Frank, R.H.Dennard et al., "Device Scaling Limits of Si MOSFETs and Their Application Dependencies", *Proc. IEEE,* vol.89, pp.259-288, February 2001.

[2] www.wikipedia.org

[3] B.Davari, R.H.Dennard et al., "CMOS Scaling for High Performance and Low Power The Next Ten Years", *Proc. IEEE,* vol.83, pp.595-606, April 1995.

[4] Gordon Moore, "Cramming More Components onto Integrated Circuits", *Proc. IEEE,* vol.86, pp.82-85, January 1995.

[5] Kaushik Roy et al., "Leakage Current Mechanisms and Leakage Reduction Techniques in Deep Submicron CMOS Circuits", *Proc. IEEE*, vol. 91, pp.305-327, February 2003.

[6] Yannis Tsividis, "Operation and Modelling of The MOS TRANSISTOR", Publisher OXFORD university Press 2008.

[7] L.Risch, "Pushing CMOS Beyond the Roadmap", *Solid State Electronics*, vol.50, pp. 527-535, March 2006.

[8] Vishwas Jaju, "Silicon-On-Insulator Technology", *Proc. IEEE International. SOI Conference*, pp.23-26, *2002*.

[9] Vinay Kumar Yadav and Ashwani K. Rana ,"Impact of Channel Doping on DG MOSFET Parameters in Nano Regime-TCAD Simulation", *International Journal of Computer Applications*,Vol.37,pp.36-41, January 2012.

[10] J. Lolivier, J. Widiez et al., " Experimental Comparison Between Double Gate, Ground Plane, and Single Gate SO1 CMOSFETs, *Solid state device conference*, pp.77-80, September 2004.

[11] M.Jagadesh Kumar and M. Siva, "The Ground Plane in Buried Oxide for Controlling Short-Channel Effects in Nanoscale SOI MOSFETs", *IEEE trans. on electron devices*, vol. 55, June 2008.

[12] Anurag Chaudhry and M. Jagadesh Kumar, "Controlling Short-channel Effects in Deep Submicron SOI MOSFETs for Improved Reliability: A Review", *IEEE trans. on Device and Materials Reliability,* vol.4, pp.99-109, March 2004.

[13] Shunpei Abe,Yoshiyasu Miyazawa et al, "Suppression of DIBL in Deca-nano SOI MOSFETs by Controlling Permittivity and Thickness of BOX layers", *Proc. IEEE,* vol.95, pp.329-332, January 2009.

[14] FZ. Rahou, A. Guen-Bouazza & M. Rahou, "Electrical Characteristics Comparison Between Fully-Depleted SOI MOSFET and Partially-Depleted SOI MOSFET using SILVACO SOFTWARE", *Global Journal of Researches in Engineering Electrical and Electronics Engineering,* vol.13,pp.1-6, 2013.

[15] Tetsu Tanaka, Hiroshi Horie et al, "Analysis of P^+ Poly Si Double-Gate Thin-Film SOI MOSFETS", *Proc. IEEE* , vol.78, pp.360-364, January 1991.

[16] K.F Scheugraf, C.C King et al, "Ultra-thin Silicon Dioxide Leakage Current and Scaling Limit", *Digest. Symposium. VLSI*, pp. 18-19, 1992.

[17] Wen-Chin Lee, "Modeling CMOS Tunnelling Currents Through Ultrathin Gate Oxide due to Conduction and Valence Electron and Hole Tunnelling", *IEEE trans. on electron devices,* vol. 48, July 2001.

[18] Yee Chia Yeo, Qiang Lu et al, "Direct Tunnelling Gate Leakage Current in Transistors with Ultrathin Silicon Nitride Gate Dielectric", *IEEE Electron Device Letters,* vol.21, November 2000.

[19] Yee-Chia Yeo, Tsu-Jae King et al, "MOSFET Gate Leakage Modeling and Selection Guide for Alternative Gate Dielectrics Based on Leakage Considerations", *IEEE trans. on electron devices,* vol. 50, April 2003.

[20] J.Robertson, "High Dielectric Constant Oxides", *Eur.Phys.J.Appl.Phys.28,* pp.265-291, 2004.

[21] Mohan V.Dunga, Xuemei Xi et al, "Modeling of Direct Tunneling Current in Multi-Layer Gate Stacks", *Nanotech 2003,* vol.2, www.nsti.org.

[22] Ghadar Darbandy, Romain Ritzenthaler et al, "Analytical Modeling of Direct Tunnelling Current Through Gate Stacks for The Determination of Suitable High-K Dielectrics for Nanoscale Double-gate MOSFETs", *Semicond.Sci.Technol*.26, pp.1-8, 2011.

[23] Thomas Kauerauf, Bogdan Govoreanu et al, "Scaling CMOS: Finding the Gate Stack with The Lowest Leakage Current", *Solid State Electronics*, vol.49, pp. 695-701, June 2005.

[24] E.Merzbacher, Quantum Mechanics.New York:Wiley, 1961.

[25]Y.shi,T.P.ma et al, "Polarity Dependent Gate Tunnelling Current and Oxide Breakdown in Dual-Gate CMOSFETs", *IEEE Electron Device Lett*ers, vol.19, pp.391-393, November 1998.

[26] W.Liu et al (1999) BSIM 3v3.2 MOSFET Model-Users' Manual.