

# **Analysis of critical factors affecting Nano scaling of CMOS technology**

A Thesis submitted  
by

TURAB, NAIMUL	(09-12937-1)
TANBIR, SHARIER	(09-12724-1)
KHAN, MD. NAHID HOSSAIN	(09-13552-1)
HOSSAIN, TANVIR MAHBOOB	(09-13637-1)

Under the supervision of  
  
Mr. EBAD ZAHIR  
Faculty  
Faculty of Engineering  
American International University Bangladesh (AIUB)

**Electrical and Electronic Engineering Department**  
**Faculty of Engineering**  
**American International University - Bangladesh**

**Spring Semester 2011-2012**

**August 2012**

# **Analysis of critical factors affecting Nano scaling of CMOS technology**

A Thesis submitted to the Electrical and Electronic Engineering Department of the Engineering Faculty, American International University - Bangladesh (AIUB) in partial fulfillment of the requirements for the degree of Bachelor of Science in Electrical and Electronic Engineering.

TURAB, NAIMUL	(09-12937-1)
TANBIR, SHARIER	(09-12724-1)
KHAN, MD.NAHID HOSSAIN	(09-13552-1)
HOSSAIN, TANVIR MAHBOOB	(09-13637-1)

**Electrical and Electronic Engineering Department  
Faculty of Engineering  
American International University - Bangladesh**

**Spring Semester 2011-2012**

**August 2012**

## **Declaration**

This is to certify that this project and thesis is our original work. No part of this work has been submitted elsewhere partially or fully for the award of any other degree or diploma. Any material reproduced in this thesis has been properly acknowledged.

Student's names & Signatures.

---

1. TURAB, NAIMUL  
(09-12937-1)

---

2. TANBIR, SHARIER  
(09-12724-1)

---

3. KHAN, MD.NAHID HOSSAIN  
(09-13552-3)

---

4. HOSSAIN, TANVIR MAHBOOB  
(09-13637-1)

## **Approval**

The Thesis titled “Analysis of critical factors affecting Nano scaling of CMOS technology” has been submitted to the following respected members of the Board of Examiners of the Faculty of Engineering in partial fulfillment of the requirements for the degree of Bachelor of Electrical and Electronic Engineering on August 2012 by the following students and has been accepted as satisfactory.

- 
- |                            |              |
|----------------------------|--------------|
| 1. TURAB, NAIMUL           | (09-12937-1) |
| 2. TANBIR, SHARIER         | (09-12724-1) |
| 3. KHAN, MD.NAHID HOSSAIN  | (09-13552-1) |
| 4. HOSSAIN, TANVIR MAHBOOB | (09-13637-1) |

---

Mr. Ebad Zahir  
Supervisor  
Faculty  
Electrical and Electronic Engineering Dept.  
Faculty of Engineering  
American International University-Bangladesh

---

Mr. Farhadur Arifin  
External-supervisor  
Assistant Professor  
Electrical and Electronic Engineering Dept.  
Faculty of Engineering  
American International University-Bangladesh

---

Prof. Dr. ABM Siddique Hossian  
Dean  
Faculty of Engineering  
American International University-Bangladesh

---

Dr. Carmen Z. Lamagna  
Vice Chancellor  
American International University-Bangladesh

## **Acknowledgements**

First, we would like to thank our supervisor Mr. Ebad Zahir (Faculty, Electrical & Electronic Engineering Dept, AIUB) for his continuous help, support, advice and proper guidance. His profound knowledge in this field, keen interest, patience and continuous support led to the completion of our work

We would also like to thank our external Mr. Farhadur Arifin (Assistant Professor, Electrical & Electronic Engineering Dept, AIUB) for his cooperation in this thesis.

Finally we would like to thank our honorable Dean, Prof. Dr. ABM Siddque Hossain, Faculty of Engineering & Vice Chancellor Dr. Carmen Z. Lamagna of American International University-Bangladesh (AIUB) for their support.

TURAB, NAIMUL	(09-12937-1)
TANBIR, SHARIER	(09-12724-1)
KHAN, MD.NAHID HOSSAIN	(09-13552-1)
HOSSAIN, TANVIR MAHBOOB	(09-13637-1)

# Contents

	Page No.
Declaration	(I)
Approval Sheet	(II)
Acknowledgement	(III)
Contents	(IV)
Abstract	(VI)
<b>Chapter 1 Introduction .....</b>	<b>1-10</b>
1.1 Introduction.....	1
1.2 Historical Background .....	1
1.2.1 CMOS Scaling Challenges.....	2
1.2.2 MOSFET Structure .....	3
1.2.3 Physical Structure of MOSFET .....	4
1.2.4 Doping. ....	5
1.2.5 Types of MOSFET .....	7
1.2.6 MOS Transistor operation .....	8
1.3 Related Other Work .....	10
1.4 Objective .....	10
1.5 Introduction to the thesis.....	10
<b>Chapter 2 Scaling Effect on Threshold Voltage .....</b>	<b>11-25</b>
2.1 Introduction .....	11
2.2 Effect of Gate-oxide thickness on threshold voltage .....	14
2.3 Effect of channel length reduction on threshold voltage.....	16
2.4 Effect of temperature on threshold voltage .....	19
2.5 Change in gate-material due to scaling and its effect on threshold voltage...	20
2.6 Different models of threshold voltage.....	21
2.6.1 Threshold-voltage model of MOSFET with stack High-k gate dielectric.....	21
2.6.2 Threshold voltage model for a Double-gate MOSFET .....	23
2.6.3 Threshold voltage of Tri-gate FINFET .....	24
2.7 Summary .....	25
<b>Chapter 3 Temperature .....</b>	<b>26-44</b>
3.1 Introduction .....	26
3.2 Effect of Temperature on Energy Bandgap .....	26
3.3 Effect of temperature on intrinsic carrier concentration .....	27
3.4 Effect of temperature on Carrier Concentration.....	30
3.5 Effect of temperature on Fermi level.....	32

3.6	Effect of temperature on Threshold voltage.....	33
3.7	Effect of temperature on Carrier mobility.....	36
3.8	Zero temperature Co-efficient .....	37
3.9	Effect of temperature variation on gate tunneling current at different Values of Equivalent Oxide Thickness (EOT).....	39
3.10	Effect of temperature variation on gate tunneling current at different Values of Substrate Doping (NA) .....	42
3.11	Summary .....	44
<b>Chapter 4</b>	<b>Gate-oxide thickness reduction.....</b>	<b>45-60</b>
4.1	Introduction .....	45
4.2	Background .....	45
4.3	High K-Dielectric materials .....	49
4.4	Selection of Dielectric materials.....	50
4.5	Modeling Direct Tunneling current .....	52
4.6	Dielectric breakdown .....	59
4.7	Summary .....	60
<b>Chapter 5</b>	<b>Channel Length Reduction.....</b>	<b>61-72</b>
5.1	Introduction .....	61
5.2	Channel length modulation .....	62
5.3	Drain-induced barrier lowering and punch-through .....	64
5.4	Surface scattering .....	65
5.5	Velocity saturation .....	66
5.6	Hot carrier effect .....	67
5.7	Solution .....	68
	5.7.1 Retrograde well engineering.....	68
	5.7.2 Halo Implant .....	70
5.8	Summary .....	72
<b>Chapter 6</b>	<b>Discussion and Conclusion.....</b>	<b>73-79</b>
6.1	Discussion .....	73
6.2	Suggestion for Future work .....	75
	6.2.1 Multi-gate FET's .....	75
	6.2.2 Tri-gate Transistor .....	76
	6.2.3 Four-gate Transistor .....	76
6.3	Future Techniques .....	77
	6.3.1 Novel gate-oxide materials .....	77
	6.3.2 CNTFET Transistor .....	78
6.4	Conclusion .....	79
<b>References .....</b>		<b>80-82</b>

## **Abstract**

CMOS technology is one of the most important achievements in modern engineering history. As the CMOS devices are scaled down to the nanometer scale, performance of the devices gets affected significantly. This thesis paper describes some of the effects of scaling on CMOS devices. Scaling of CMOS transistor into nanometer regime causes changes in different parameters which can affect the performance of CMOS. In this paper we have analyzed four such critical parameters- threshold voltage, temperature, gate oxide thickness and channel length. We have analyzed how scaling affects these parameters by using graphs and equations. We have used different models to help us analyze these effects. Then we have also given some solution to these problems. Finally we have suggested some technology which might be introduced in the future to deal with the scaling problems. All the information that has helped us in the analysis is collected from reputed journals and books.

# Chapter 1

## Introduction

### 1.1 Introduction

CMOS is a very important part of every electronic that we use today. The performance of the electronics depends on the number of CMOS in it. So increasing the number of CMOS will increase the performance. Miniaturization of devices is a continuous trend in the microelectronic industry. With decreasing feature size, the device cost decreases and its performance increases. This has led silicon microelectronics a transition to the sub-micrometer scale and then to the nanometer scale silicon electronics. Nanotechnology era has changed the performance capability of devices. As more CMOS can be integrated in a small area at nanometer scale the performance of device increases remarkably and the limits of microelectronics are overcome quite phenomenally. However scaling devices towards nanometer-scale dimensions is rapidly exacerbating reliability problems in large-scale integrated ICs. Power, short channel effect, variability and reliability are key challenges for CMOS integration and scaling.

### 1.2 Historical Background:

In the last few decades silicon-based microelectronic devices have revolutionized our world. The need for higher computing power at cheaper cost has fueled increasing CMOS scaling. It all started with the invention of integrated circuit in late 1950's that unveiled the possibility of using transistors in almost all kinds of electronic circuits. The next major breakthrough came with the demonstration of the first metal-oxide semiconductor field-effect transistor (MOSFET) in 1960 by Kahng and Atalla which would enable cost-effective integration of large number of transistors with interconnections on a single silicon chip. Gordon Moore noted that the number of transistor on a chip doubled every 18 to 24 months. According to his prediction modern chip can accommodate thousands of transistors in 1 mm space

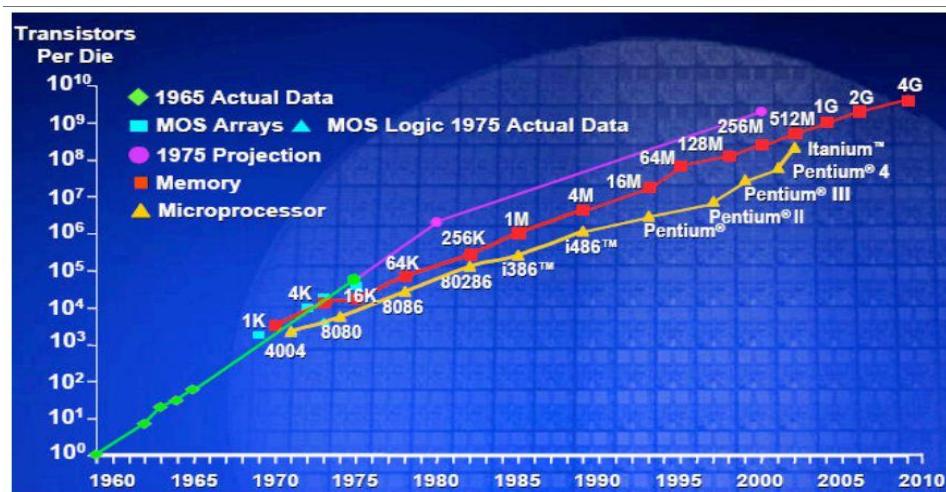


Figure 1.1: Increase in transistor number with time

### **1.2.1 CMOS scaling challenges:**

In nanotechnology based electronics field such as manufacturing microprocessors, nanoscale CMOS is the main device through which a microprocessor can be made. But scaling of CMOS in nanometer scale arises many factors which affects the circuit performance. CMOS must be scaled in such a way so that improve circuit performance and increase speed. Highest concern must be taken about the power consumption of CMOS used in nanoscale. There are two types of power consumption of CMOS – 1) dynamic power and 2) static power. Leakage current is the main factor for static power consumption. Gate oxide thickness is one of the main factors that affect directly leakage current. Aggressive reduction of gate oxide thickness cause huge amount leakage current. So, for high performance logic circuits it has significant effect on performance.

The scaling of transistor has several advantages in addition to increasing the on-chip transistor density. Delay of the logic gates decreases and the operating frequency of the transistors increases by a factor of  $1/L_g$  ( $L_g$  is transistor gate length) enabling faster operation of the circuits. For same amount of functionality, the chip area decreases by a factor of  $1/L_g^2$  which allows more dies on a single wafer resulting in large cost savings. Furthermore, since the die size is smaller, the amount of defects/die is also small leading to a higher yield in manufacturing. The active switching power per area remains constant with technology scaling allowing the circuits to run at lower power or allowing the circuits to incorporate more functionality under a fixed power constraint

Traditionally, as the channel length  $L_g$  is decreased, the oxide thickness  $T_{ox}$  is reduced commensurately to maintain strong capacitive coupling between the inversion channel and the gate terminal compared to other transistor terminals. This helps to control the short channel effects such as threshold voltage roll-off, sub-threshold slope degradation and drain induced barrier lowering (DIBL) which act together to increase the off-state transistor leakage. Scaling of  $L_g$  is also accompanied by a corresponding increasing body doping  $N_A$  and decrease in source/drain junction depth  $X_j$  to reduce the leakage path below the inversion channel. This scaling methodology has worked very well for several decades but in the recent past the conventional CMOS technology is approaching physical limits whereby further scaling is beginning to have a negative impact on the transistor performance [1].

The thin gate oxide  $T_{ox}$  loses its ideal insulating property and there is significant increase in gate current through direct quantum tunneling of carriers across the oxide. This results in an increase in the on-state leakage. The increase in body doping leads to reduction in carrier mobility due to the large vertical electric field and the enhancement in the current drive due to transistor scaling decreases. The large vertical field created at source/drain junction due to the high body doping also leads to large band-to-band tunneling current and increased off-state leakage. Increased body doping also leads to larger body and junction capacitances making the transistor switching slower. In scaled transistors, the random device-to-device variability in the form of random dopant fluctuations is also worsened due to the presence of heavy body doping. The reduced junction depths increase the source/drain resistance which reduces the current drive of the transistor. [2].

To solve these scaling issues, two different paths can be charted. The first path involves introduction of new technologies and new materials into the conventional planar bulk MOSFET to allow further scaling and boost the performance of scaled transistors.

The second path involves adoption of new transistor architectures such as ultra-thin body FETs and multi-gate FETs which inherently have superior electrostatic control over the inversion channel. In order to study the circuit level benefits and reliability of these two CMOS scaling solutions, timely understanding and modeling of the associated device physics and device behavior is needed.

### 1.2.2 MOSFET Structure:

MOSFET transistors are the core of today's integrated circuits (ICs). Originally computers used mechanical switches to solve Boolean operations. But the smaller, faster, cooler MOSFET transistors allowed computers to evolve and dominate our lives.

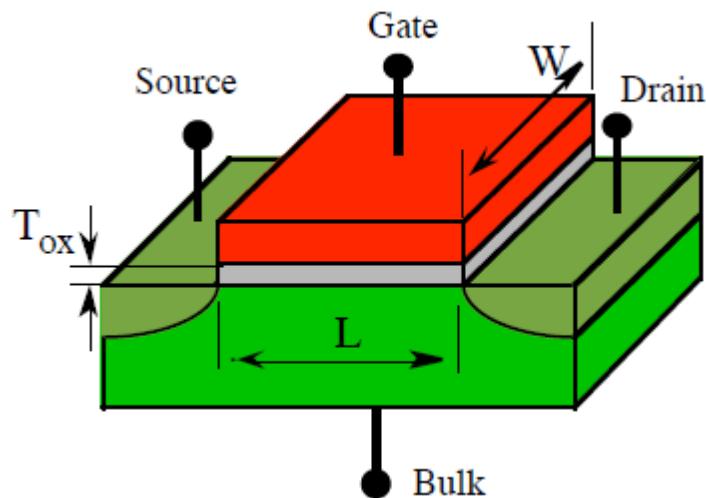


Figure 1.2: MOSFET Transistor

Figure 1.2 sketches a MOSFET transistor. The bottom rectangular block of material is the silicon substrate often referred to as the bulk. There are four electronically active regions that are marked: gate (G), source (S), and drain (D), and the bulk terminal (B) to which the gate, drain, and source voltages are typically referenced. The rectangular gate region lies on top of the bulk separated by a thin silicon oxide dielectric with thickness  $T_{ox}$ . Two other important dimensions are the transistor gate length and width. The drain and source regions are embedded in the substrate but have an opposite doping to the substrate.

There are two types of MOSFET transistors, the nMOS transistor and the pMOS transistor. The charged carriers are holes in pMOS transistors and electrons in nMOS transistors.

The gate terminal is analogous to the light switch on the wall. When the gate has a high voltage, the transistor closes like a wall switch, and the drain and source terminals are electrically connected. Just as a light switch requires a certain force to activate, the transistor gate terminal needs a certain voltage level to switch and connect the drain and source terminals. This voltage is called the transistor threshold voltage  $V_{th}$  and is a fixed voltage for nMOS and for pMOS devices in a given fabrication process.

An ideal transistor has zero ohms between the drain and source when it is in the ON-state and infinite resistance between these terminals in the OFF-state.

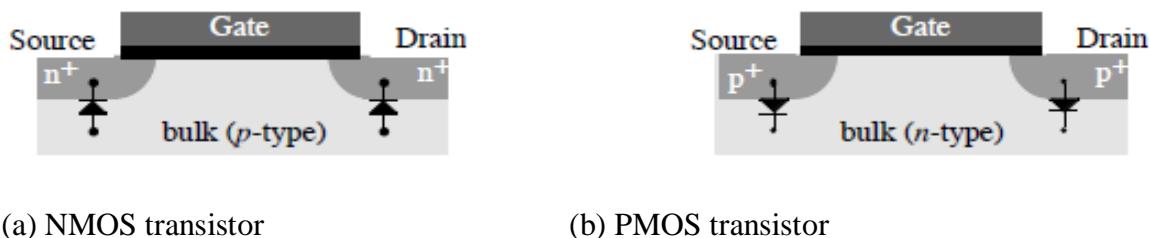
The ideal device should also switch between ON and OFF-states with a zero delay time as soon as the control variable changes state.

### 1.2.3 Physical structure of MOSFET:

MOS transistor construction begins with a lightly doped host crystalline substrate structure. nMOS substrates are doped with p-doped silicon, while the pMOS substrates use n-doped silicon.

The gate oxide ( $T_{ox}$ ) electrically isolates the gate terminal from the semiconductor crystalline structures underneath. It is made of oxidized silicon forming a non-crystalline, amorphous  $SiO_2$ . The gate oxide thickness ( $T_{ox}$ ) typically ranges from near 15 Å to 100 Å (1 Å = 1 Angstrom =  $10^{-10}$  m).  $SiO_2$  molecules are about 3.2 Å in diameter so that this vital dimension is now a few molecular layers thick. The purpose of the  $SiO_2$  dielectric is to allow an electric field generated by the gate voltage to influence the amount of charge passing between the drain and source. A thinner gate oxide allows the electric field gate to better control the device state and allows faster transistors. The thin gate oxide has sometimes been referred to as the beating heart of the transistor.

The nMOS transistor has a p-doped silicon substrate with n-doping for the drain and source. pMOS transistors have a complementary structure with an n-doped silicon bulk and p-doped drain and source regions. The gate region above the thin oxide dielectric is constructed with polysilicon in both transistors. Polysilicon is made of many small silicon crystals. The region between the drain and source just under the gate oxide is called the channel, and is where charge conduction takes place. The electronic distinction between the two transistors is that electrons are the channel current in the nMOS transistor, and holes are the channel current in the pMOS transistor. Since drain and source dopants are opposite to the substrate (bulk), they form pn-junction diodes that are either reverse or zero biased in normal operation.



(a) NMOS transistor

(b) PMOS transistor

Figure 1.3: Structure of two types of MOSFET

The distance from the drain to the source is a geometrical parameter called the channel length (L) and the lateral dimension is the transistor channel width (W). Transistor length and width are parameters set by the circuit designer and process engineer. The width to length ratio (W/L) is linearly related to the drain current capability of the transistor. A wider transistor will pass more current. The gate is the control terminal, while the source provides electron or hole charge carriers that empty into the channel and are collected by the drain.

## 1.2.4 Doping:

Pure silicon atom is surrounded by four adjacent silicon atoms. It has equal number of electrons and holes.

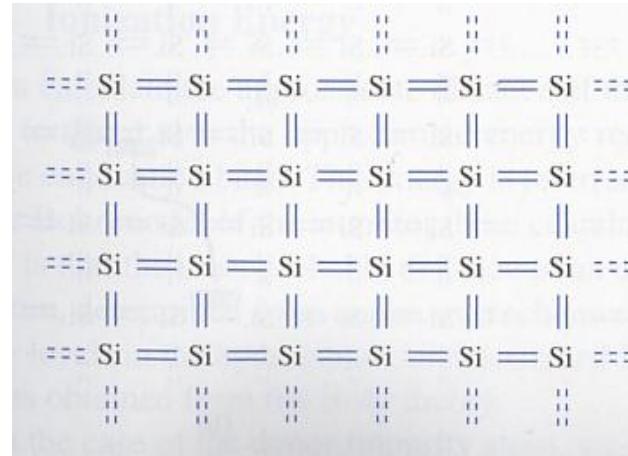


Figure 1.4: Intrinsic silicon lattice

When silicon is doped with donor atom like phosphorous it has excessive electrons.

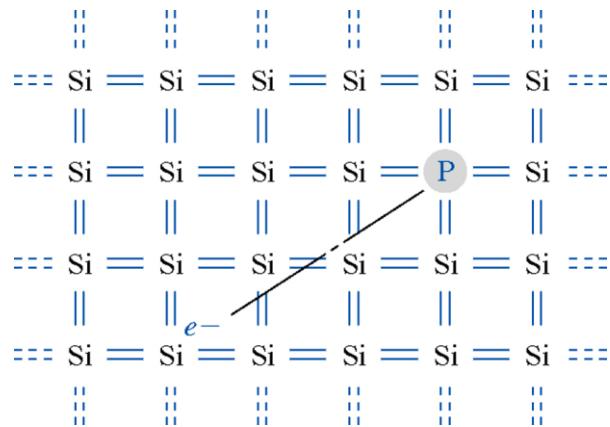


Figure 1.5: N-type: silicon lattice doped with donor impurity phosphorous

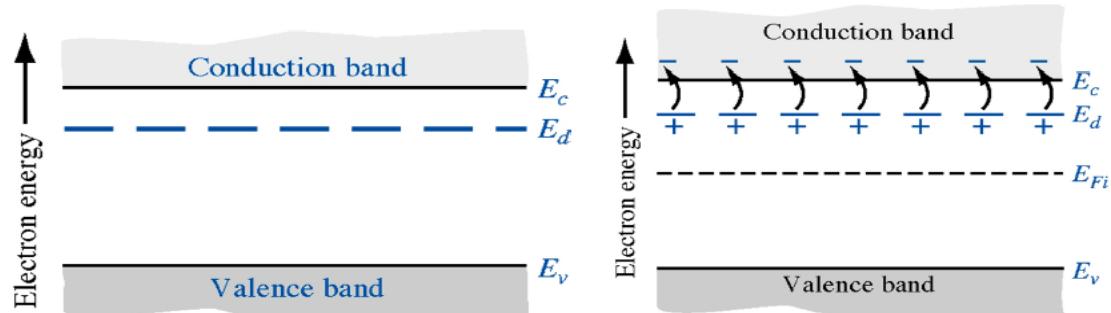


Figure 1.6: Donor electron energy level

Little energy is required to move the electrons from donor state to conduction band. Positively charged donor ions are fixed but donor electrons in the conduction band can move through the crystal.

When silicon is doped with acceptor atoms like boron it has excessive holes.

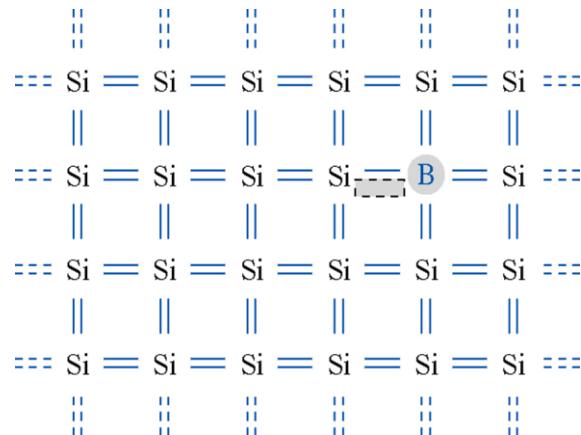


Figure 1.7: P-type: silicon lattice doped with acceptor impurity boron

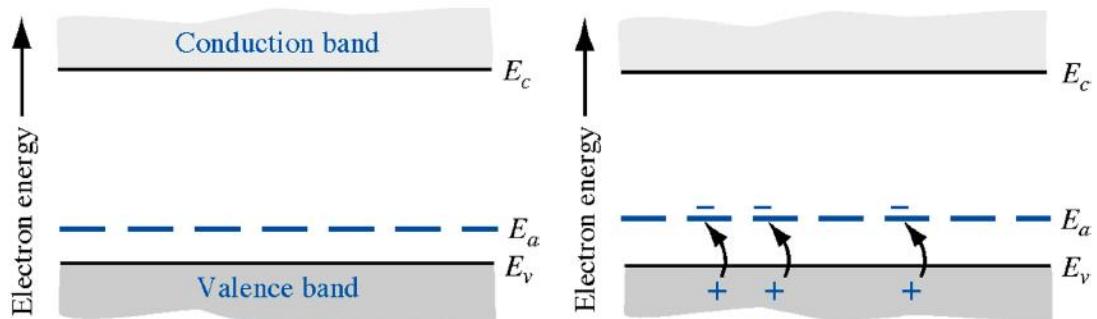


Figure 1.8: Acceptor energy level

Little energy is required to move valence electrons to acceptor levels. Negatively charged acceptor ions are fixed but holes in the valence band can move through the crystal

### 1.2.5 Types of MOSFET:

There are basically four type of MOSFET:

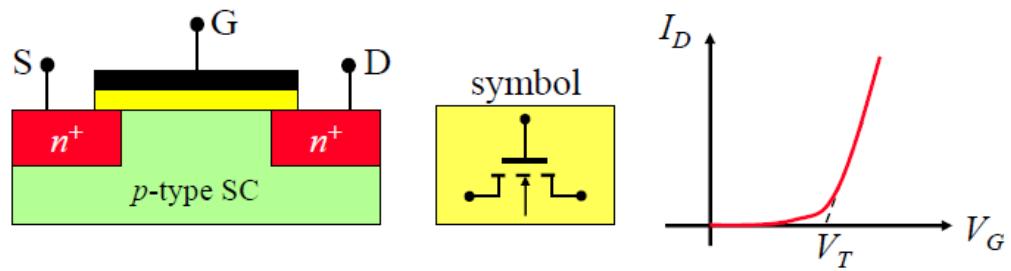


Figure 1.9: n-channel enhancement mode MOSFET

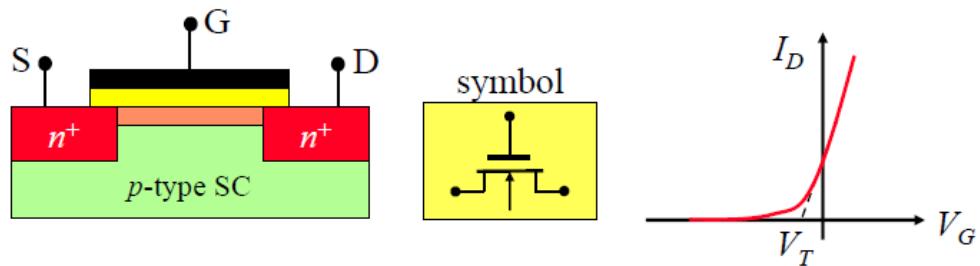


Figure 1.10: n-channel depletion mode MOSFET

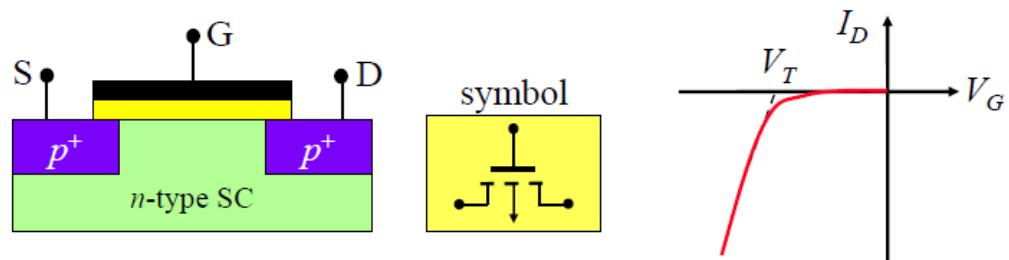


Figure 1.11: p-channel enhancement mode MOSFET

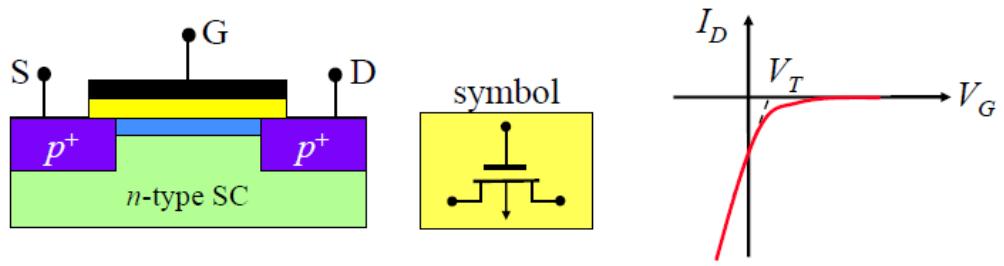


Figure 1.12: p-channel depletion mode MOSFET

### 1.2.6 MOS Transistor Operation:

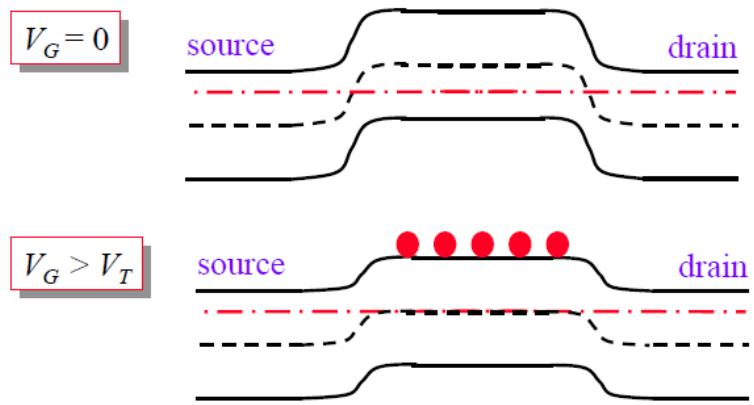


Figure 1.13: Role of gate electrode for n-channel MOSFET

Positive gate voltage reduces the potential energy barrier seen by the electrons from the source and the drain regions. It also inverts the surface and increases the conductivity of the channel.

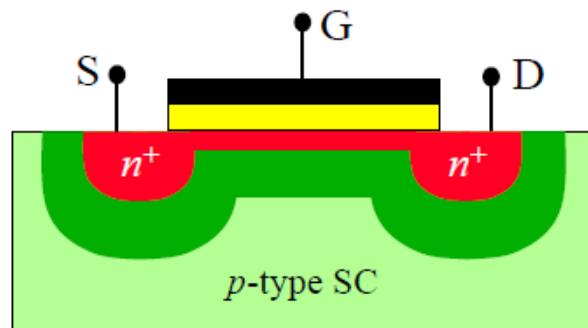


Figure 1.14:  $V_G > V_{th}$ ,  $V_D > 0$  (small)

Variation of electron density along the channel is small:  $I_D$  directly proportional to  $V_D$

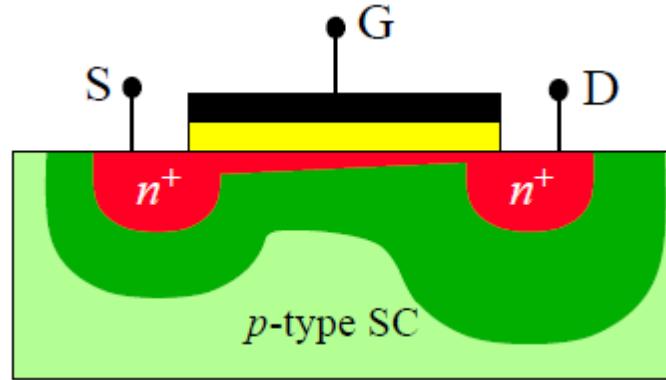


Figure 1.15:  $V_G > V_{th}$ ,  $V_D < V_G - V_{th}$

Increase in the drain current reduces due to the reduced conductivity of the channel at the drain end

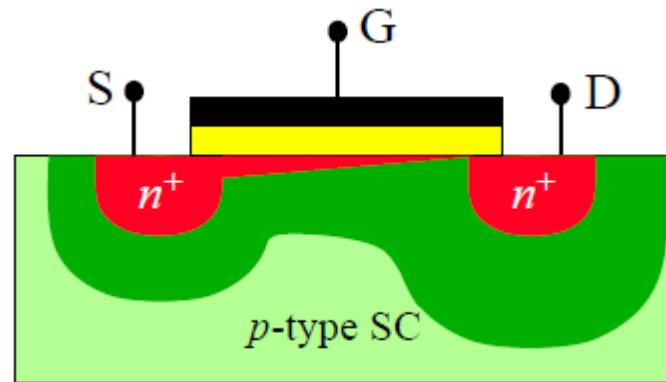


Figure 1.16:  $V_G > V_{th}$ ,  $V_D = V_G - V_{th}$

Pinch-off point. Electron density at the drain end of the channel is identically zero

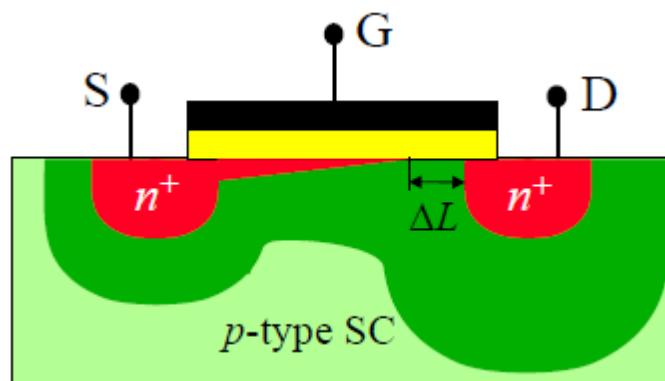


Figure 1.17:  $V_G > V_{th}$ ,  $V_D > V_G - V_{th}$

Post pinch-off characteristic. The excess drain voltage is dropped across the highly resistive pinch-off region. [3]

### **1.3 Related other work:**

Scientist are also working on FINFET, Tri-gate, CNTFET and many other structure which can replace CMOS in the future electronics. Scientists are working on different dielectric materials to replace silicon. They are doing their research on investigating the molecular defect, bandgap and on other quantum mechanical effects of CMOS technologies.

### **1.4 Objective:**

The objective of this thesis is to analyze the effect of scaling on different CMOS parameter by the help of different reputed journals and correlate the parameters with each other.

### **1.5 Introduction to the thesis:**

We have divided our thesis into six chapters. In the first chapter we have discussed about the need for scaling and the scaling challenges. From chapter 2 to chapter 5 we have analyzed the effect of scaling on four CMOS parameters- threshold voltage, temperature, gate oxide thickness and channel length. Finally in chapter 6 we have discussed about how the problems of scaling are overcome by using different technology and architecture.

## Chapter 2

### Scaling Effect on Threshold Voltage

#### 2.1 Introduction

According to Moore's law the number of transistor incorporated in a chip will approximately double every 24 months. To maintain this law the scaling of MOSFET transistor faces many obstacles. As the technology is scaled into the nanometer regime the physical parameter such as gate oxide and channel length of the MOSFET decreases which eventually affects the threshold voltage. As shown in Figure 2.1.1 the power supply and threshold voltage decreases as the MOSFET channel length is scaled down to the nanometer regime. [1]

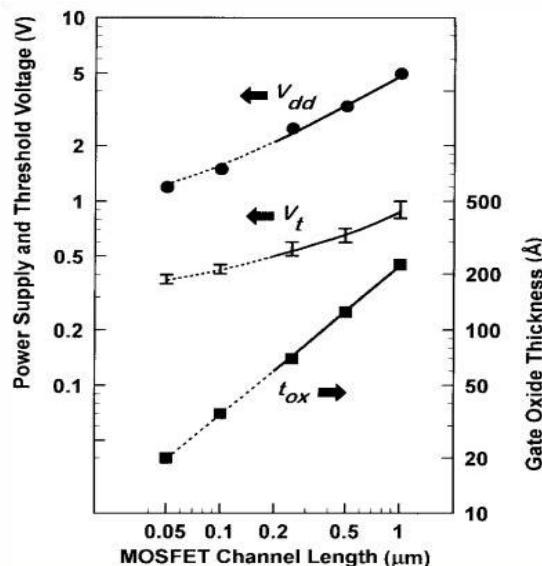


Figure 2.1.1: Power supply voltage ( $V_{dd}$ ), threshold voltage ( $V_T$ ), and gate-oxide thickness trends versus channel length for CMOS logic technologies [1]

To analyze the threshold voltage effects in nanometer the formation of threshold voltage in MOSFET structure must be understood. Threshold voltage is best explained with the help of energy band formation at the metal-oxide-silicon interface. The characteristic of p-type substrate is considered for the explanation although all of the considerations are equally valid for an n-type substrate with the proper changes in signs and symbols. Energy Band Diagram of metal-oxide-silicon interface at equilibrium condition and at threshold condition is shown in Figure 2.1.2.

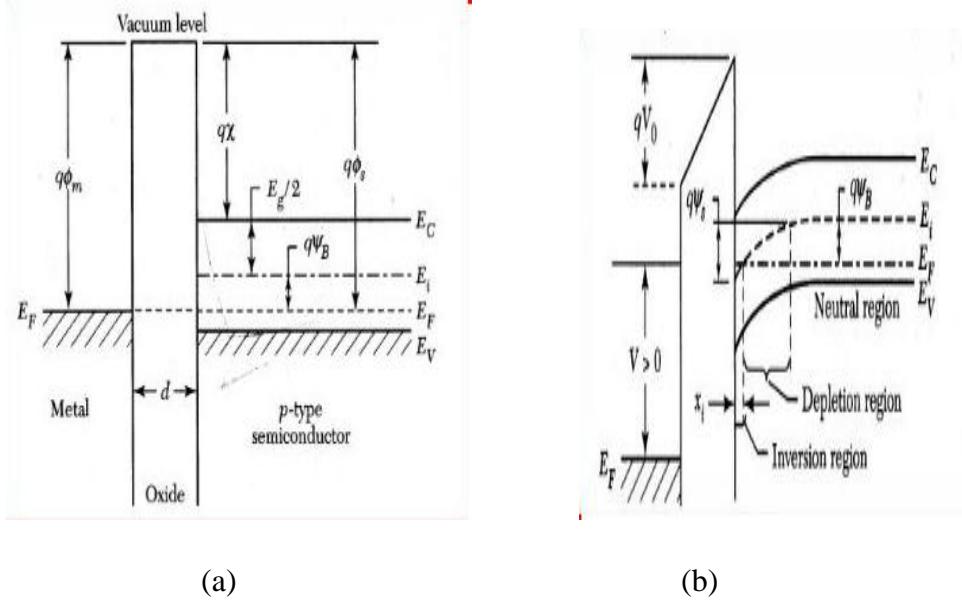


Figure 2.1.2: Energy band diagrams of metal-oxide-silicon interface of (a) Ideal MOS-C at  $V=0$ , (b) Threshold or inversion condition when  $V_g > V_T$ . [4]

The energy band diagram of ideal N-MOSFET at metal-oxide-silicon interface when  $V_g=0$  is shown in figure 2.1.2(a). This energy band is also called MOS flat band diagram as the silicon conduction and balance band level are flat at the surface of  $\text{SiO}_2$ -Si interface. The energy difference between the Fermi level and the vacuum level is called the work function ( $q\phi$ ),  $q\phi_m$  is the work function of metal and  $q\phi_s$  is the work function of semiconductor.  $E_i$  is the intrinsic Fermi level which lies at the middle of valance and conduction band.

The energy band diagram during the inversion case or the threshold condition is as shown in figure 2.1.2(b). Threshold voltage is defined as fixed positive voltage when applied at the gate terminal, the positive charge at the metal surface attracts electrons (minority carrier) at the surface of  $\text{SiO}_2$ -Si interface which causes an inversion layer thus resulting in a channel between source and drain. This phenomenon is best explained by the help of MOS energy band diagram at inversion case. When positive voltage is applied across the gate, the Fermi level of metal moves down by  $V$  compared to the semiconductor Fermi level this causes a downward bending of the energy band at the semiconductor surface. As shown in figure, the intrinsic Fermi level  $E_i$  at the surface crosses over the Fermi level so the energy difference between  $E_f$  and  $E_i$  increase. Electron concentration in the semiconductor depends exponentially on the energy difference  $E_f - E_i$ . Therefore,  $x_i$  represents the width of the inversion region where the concentrations of electrons are high. Electron concentration at the surface is equal to the substrate impurity concentration i.e.  $n_s = N_A$ . [4]

$$n_p = n_i e^{(E_f - E_i)/kT}$$

Where,  $n_p$ = electron concentration,  $n_i$ = intrinsic concentration and T= Temperature

The electrostatic potential  $\psi_S$  is called the surface potential and  $\psi_B$  is the potential difference between intrinsic Fermi level and Fermi level. As shown in the figure 2.1.2 at the inversion region,  $\psi_S > \psi_B$ .  $\psi_S$  is given by, [4]

$$\psi_S(\text{inv}) \cong 2\psi_B = \frac{2kT}{q} \ln \left( \frac{N_A}{n_i} \right)$$

Where,  $N_A$  =substrate impurity concentration

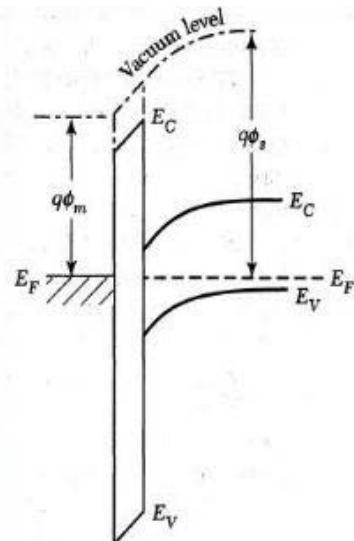


Figure 2.1.3: Ideal MOS-C at thermal equilibrium condition [4]

From figure 2.1.2(b),  $V_o$  is the potential across the gate oxide. The energy band of the oxide bends accordingly with the applied gate potential.  $V_o$  is positive if the oxide band tilts downward toward the gate and vice versa. The equation of  $V_o$  is given by, [4]

$$V_o = \varepsilon_0 d = \frac{|Q_s|d}{\varepsilon_{ox}} = \frac{|Q_s|}{C_0} \quad \text{Equation (1)}$$

Where,  $\varepsilon_0$ = electric field in the oxide,  $Q_s$  =charge per unit area,  $C_0$  =oxide capacitance per unit area and  $\varepsilon_{ox}$  =dielectric constant.

For flat-band diagram as shown in figure 2.1.2(a) the difference between work function of metal and semiconductor is zero.  $\phi_{ms} = \phi_m - \phi_s = 0$ . The energy band diagram shown in figure 2.1.3 is formed when the metal, oxide and semiconductor are attached at thermal equilibrium. Therefore to achieve a flat band diagram a voltage equal to the work function,  $\phi_{ms}$  must be applied at the gate terminal. Therefore,

$$V_{FB} = \phi_{ms}$$

Finally, after all the analysis of parameters in the energy band diagram of metal-oxide-silicon interface the overall threshold voltage equation can be given by, [4]

$$V_T = V_{FB} + 2\Psi_B + \frac{\sqrt{2\epsilon_s q N_A (2\Psi_B + V_{BS})}}{C_o} \quad \text{Equation (2)}$$

Where,  $V_T$  = Threshold voltage,  $V_{FB}$  = Flat band voltage,  $\Psi_B$  = potential difference between intrinsic Fermi level and Fermi level,  $N_A$  = substrate impurity concentration,  $C_o$  = oxide capacitance,  $\epsilon_s$  = dielectric permittivity and  $V_{BS}$  = reverse substrate-source bias.

Reverse substrate-source bias,  $V_{BS}$  is considered zero for most of the operation. The above equation provides the factors that can cause changes to threshold voltage. So by observing the changes in the factors the effect of scaling on threshold voltage can be explained. In general, substrate doping, oxide thickness, source to substrate voltage bias, gate material and surface charge density are some factors that affects the threshold voltage. However, the preceding section will only include the effect on threshold voltage due to the scaling of gate-oxide thickness and reduced channel length into the nanometer regime.

## 2.2 Effect of Gate-oxide thickness on threshold voltage

Figure 2.1.1 shows that power supply has not been decreasing at a rate proportional to the channel length. This means the field effect increases as the channel length decreases. So to control the electrostatic effects the gate-oxide thickness is reduced nearly in proportion to the channel length. Another cause of reducing gate oxide thickness is to maintain a high drive current as gate voltage is reduced. However, decrease in gate-oxide thickness affects the threshold voltage in various ways.

According to the equation of threshold voltage, as the gate-oxide width gets smaller the  $V_0$  parameter of the threshold equation gets affected,  $V_0$  is given by,

$$V_o = \frac{|Q_s|}{C_o}$$

Where,  $Q_s$ =charge per unit area,  $C_o$  =oxide capacitance per unit area

From equation (2) it can be said that  $V_T$  is inversely proportional to  $C_o$ ,

$$V_T \propto \frac{1}{C_o}$$

And  $C_o$  is given by,

$$C_o = \frac{\epsilon_{ox}}{d} \quad \text{Equation (3)}$$

Where,  $\epsilon_{ox}$  =dielectric constant and  $d$  =oxide thickness

From equation (3) it can be said that  $C_o$  is inversely proportional to  $d$ . This implies,

$$V_T \propto \frac{1}{C_o} \propto d$$

Threshold voltage is directly proportional to the gate-oxide thickness. Therefore threshold voltage decreases as the gate oxide thickness is reduced.

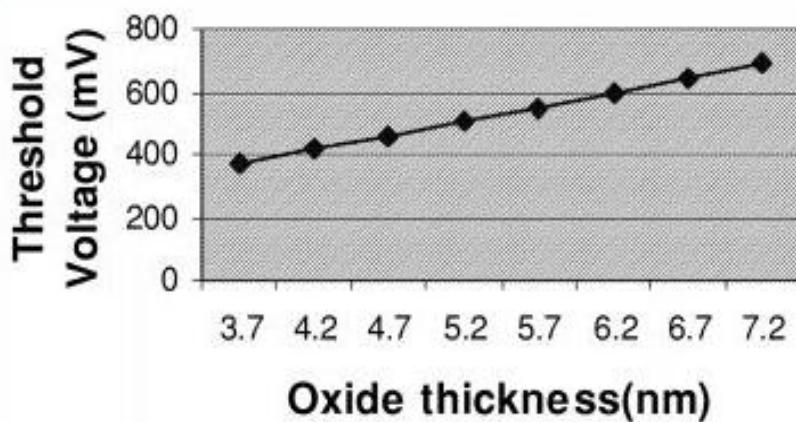


Figure 2.2.1:  $V_T$  at different oxide thicknesses. [47]

As seen from the figure 2.2.1, the threshold voltage decreases at a significant rate as the oxide thickness is scaled down.

## 2.3 Effect of channel length reduction on threshold voltage

For general MOS transistor the space charges present in the oxide –silicon interface at depletion condition are induced solely from the field created by the gate voltage. In fact the above equation of threshold voltage is independent of the horizontal field between the source and drain and it also assumes that the depletion space charge near n<sup>+</sup> source and drain region is induced by pn junction band bending. As the channel length is reduced some of the field lines originating from the source/drain terminate on charges at the channel region. Therefore the overlapping of depleted space charge between the channel regions causes a  $V_T$  roll-off phenomenon.

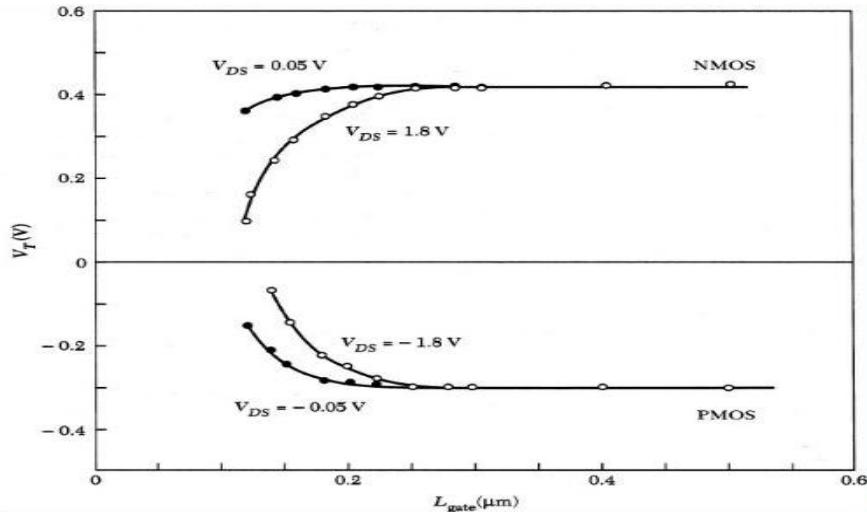


Figure 2.3.1: Threshold voltage roll-off characteristic in a  $0.15\mu\text{m}$  CMOS technology [4]

The above figure shows a  $V_T$  roll-off characteristic which can be explained by the charge sharing model. For a N-MOS operating at linear region ( $V_{DS} \leq 0.1\text{V}$ ) the depletion region at the source and drain overlaps the bulk depletion region induced by gate voltage. Since the bulk depletion charge is smaller than the general MOS model, the threshold voltage expression must be modified to account for this reduction.

$$V_{T_{SC}} = V_T - \Delta V_T$$

Where  $V_{T_{SC}}$  =Short channel threshold voltage,  $\Delta V_T$  =Threshold voltage shift

The shifted threshold voltage is derived by using the charge sharing model shown in figure 2.3.2

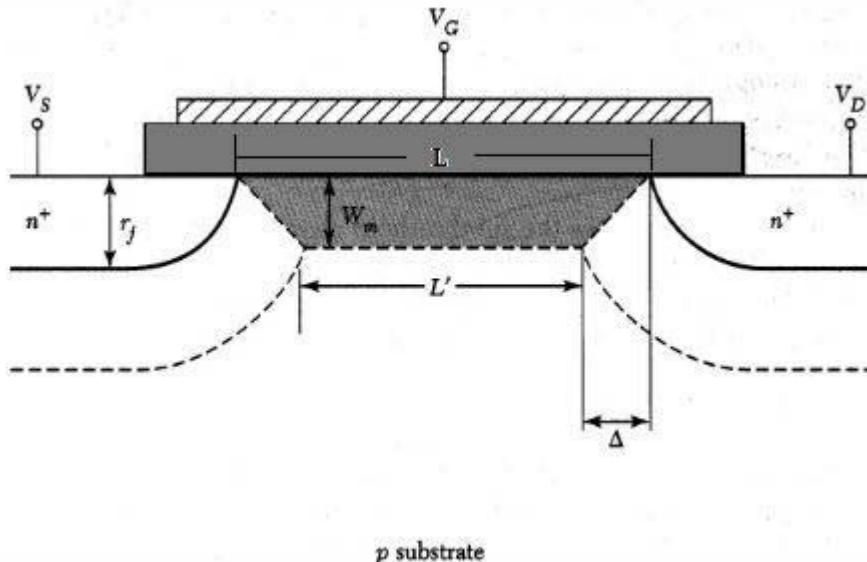


Figure 2.3.2: Schematic of the charge sharing model of N-MOS [4]

According to the figure,  $r_j$  is the source depth,  $L$  is the actual channel length,  $L'$  is the reduced channel length,  $\Delta$  is the change in channel length and  $W_m$  is the depletion width. The threshold voltage shift is due to the change in area of the depletion region beneath the gate-oxide from a rectangular shape to a trapezoidal shape.

$$\Delta V_T = -\frac{qN_A W_m r_j}{C_o L} \left( \sqrt{1 + \frac{2W_m}{r_j}} - 1 \right) [4]$$

Where,  $N_A$  =substrate doping concentration,  $C_o$  =gate oxide capacitance per unit area

For short channel length,  $\Delta$  is comparable to  $L$  so the charges needed to turn on the device are drastically reduced. From the above equation if the  $q, N_A, W_m, r_j$  and  $C_o$  are kept constant then the shift in threshold voltage is inversely proportional to the channel length,  $L$ . Therefore, the thresholds shift voltage ( $\Delta V_T$ ) increases as the channel length decreases which eventually decreases the threshold voltage ( $V_T$ ).

Drain-Induced barrier lowering also contributes to the increase in threshold voltage roll-off characteristic. For a short channel when the voltage drop between drain and source increases the depletion under the drain can lower the potential barrier from the source to channel junction. If the barrier is lowered then more electrons can be injected to the channel region and the  $V_T$  roll-off becomes larger. This effect is called the drain-induced barrier lowering (DIBL).

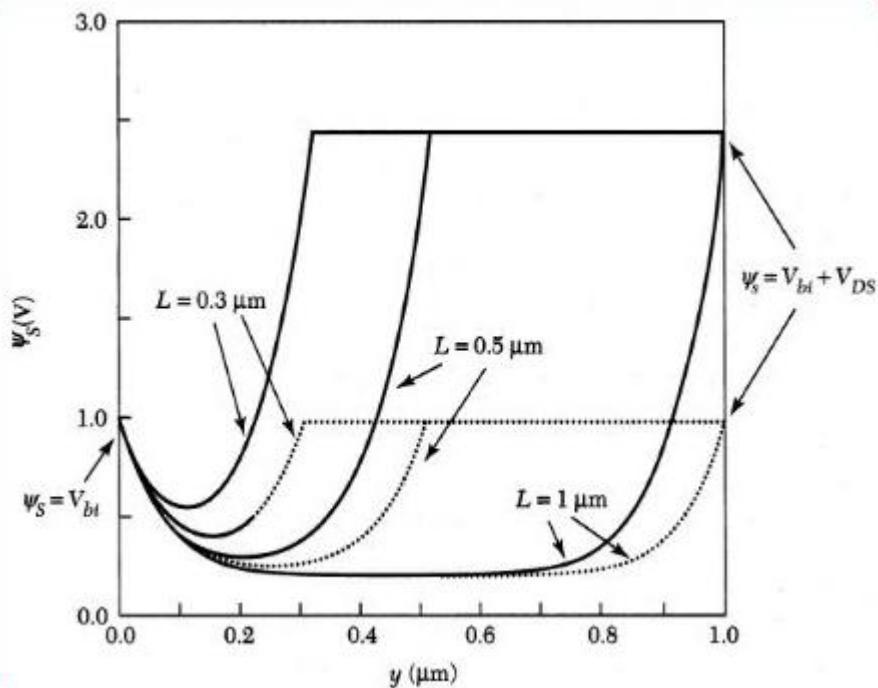


Figure 2.3.3: Calculated surface potential along the channel for n-channel MOSFET with different channel length. Oxide thickness  $d$  and substrate doping  $N_A$  are 10nm and  $10^{16}\text{cm}^{-3}$ , respectively. The substrate bias is 0V. [4]

The above figure shows the change in surface potential  $\psi_s$  as a function of different channel length. The dotted lines are for  $V_{DS} = 0$  and solid lines are for  $V_{DS} > 0$ . For a CMOS the depletion width at drain is greater than that of source when the device operates at saturation condition. For long channel the increase in depletion width does not affect the potential barrier whereas for short channel the potential barrier decreases as the device operates in saturation region. As shown in above figure for  $L=0.3\mu m$  and  $L= 0.5\mu m$  the potential barrier is less compared with  $L= 1\mu m$ .

The fields between the source and drain are too close to each other which cause substantial increase in electron injection from the source to drain. Therefore the threshold voltage decreases further with increasing drain bias in short channel MOSFET which further results in the increase of sub threshold current or leakage current.

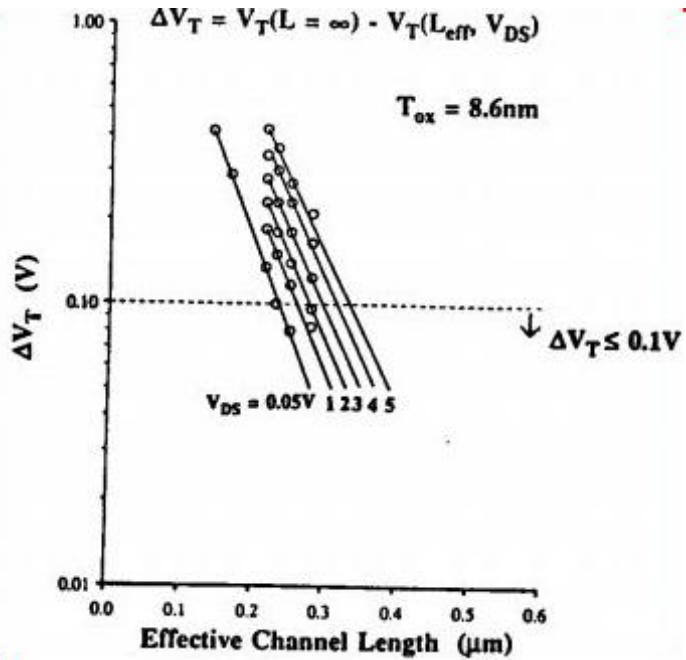


Figure 2.3.4:  $\Delta V_T$  versus effective channel lengths. [5]

## 2.4 Effect of Temperature on Threshold voltage

As shown in figure 2.4.1, threshold voltage decreases as the temperature is increased. Temperature depends mainly on background doping level and leakage current. As doping level increases the temperature increases and when scaled to nanometer the increase of leakage current also increases the temperature. The effects of temperature on intrinsic carrier concentration, Fermi level, energy band gap etc are discussed in chapter 3. Threshold voltage does not have a direct relationship with these parameters but it is not independent of the effects. Therefore taking the effects of such parameters into consideration it can be concluded that threshold voltage decreases with increase in temperature.

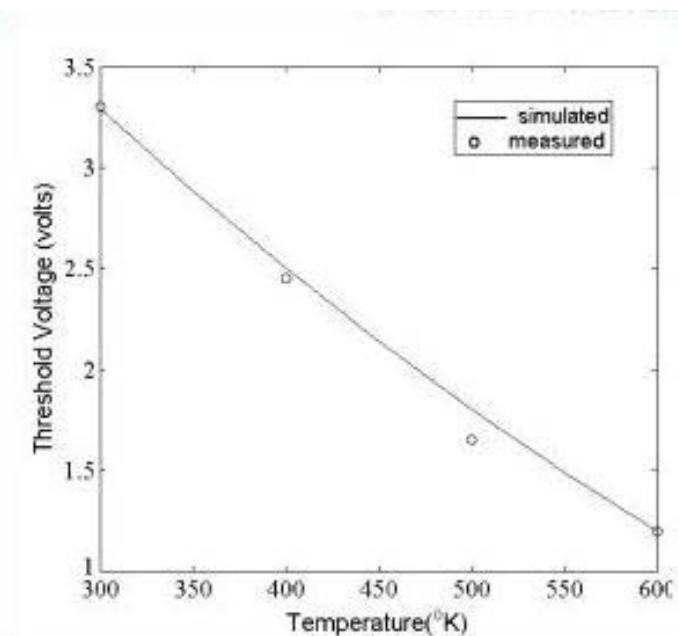


Figure 2.4.1: Threshold voltage obtained from simulation and measurement for different temperature [48]

## 2.5 Change in gate-material due to scaling and its effect on threshold voltage

The work function ( $q\phi$ ) plays a key role in determining the threshold voltage. As the gate material is changed the work function also changes respectively. Therefore change in gate material affects the threshold voltage. When metal was used as gate material, gate voltages were large; the threshold voltage could still be overcome by the applied gate voltage. However, the scaling of CMOS devices caused the gate material to change accordingly.

As transistor sizes were scaled down, the applied voltage was also brought down (as shown in figure 2.1.1) so a transistor with a high threshold voltage would become non-operational under these new conditions. Thus, poly-crystalline silicon became the gate material because its work function can be modulated by adjusting the type and level of doping. Because of the same band gap of polysilicon and underlying channel, it is easy to tune the work function to achieve low threshold voltage for both NMOS and PMOS devices. By contrast, for metal gates, tuning the work function to obtain low threshold voltage would require the use of different metals for each device type which introduces additional complexity to the fabrication process.

Polysilicon gate have been the choice for last twenty years. However, when CMOS devices are scaled further at nanometer regime the same band gap between the polysilicon gate and the channel became a significant factor. Such disadvantage has led to their replacement by metal gates again.

Therefore, when the transistors are extremely scaled down, the sub threshold voltage and tunneling current increases so the need for a higher  $V_T$  has become important again. So using metal gate would result in a higher  $V_T$  compared to polysilicon gate. After the 45nm technology of Intel, the gates are again made with metal in conjunction with high-k insulators.

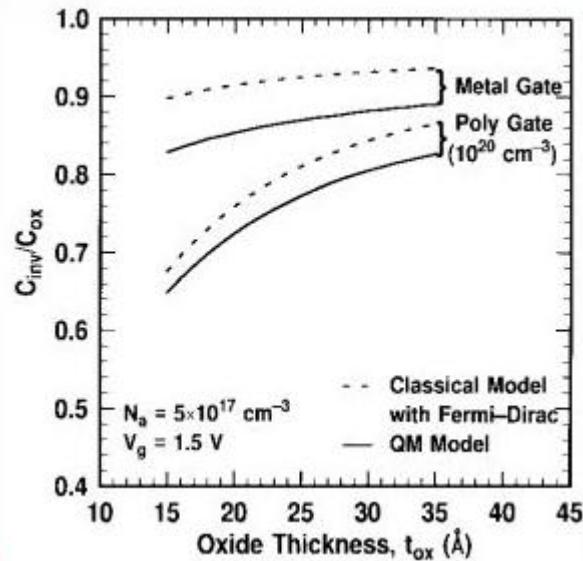


Figure 2.5.1: Calculated MOS capacitance under inversion normalized to the oxide capacitance as function of gate oxide thickness for Metal gate and Poly gate. [1]

## 2.6 Different models of threshold voltage

### 2.6.1 Threshold-voltage model of MOSFET with stack High-k gate dielectric

High-k/Si interface suffer from unacceptably high interface-state density, thus resulting in low carrier mobility. To overcome this problem a thin interlayer, e.g. SiON for Si MOS is placed beneath the High-k gate dielectric. In this section a threshold voltage model is developed for a stack High-k gate dielectric with a thin layer. Also by using a thin dielectric layer the threshold behaviors of the stack High-k dielectric can be improved.

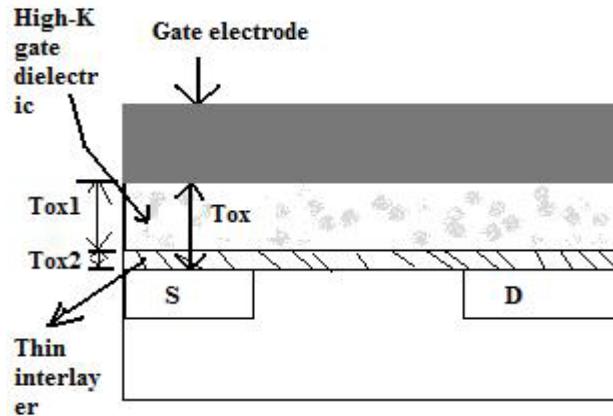


Figure 2.6.1: Schematic diagram of MOSFET with stack-high-k gate dielectric.  
[7]

The threshold voltage equation of model is as follows,

$$V_{TH} = V_{th0} - \frac{[2(V_{bi} - 2\phi_f) + V_{ds}]}{2 \cosh(\frac{L}{2l}) - 2}$$

Where,  $V_{TH}$ = threshold voltage,  $V_{th0}$ = threshold voltage of 1-D long channel MOSFET,  $L$ = channel length,  $l$ = characteristic length,  $V_{ds}$ = built-in potential of drain-substrate junction and  $V_{bi}$ = built-in potential of the drain-source voltage.

For a MOSFET the threshold roll-off occurs due to short channel effect (SCE) and fringing-field effect (FFE). From the above equation it can be said that threshold voltage is related to the characteristic length. The larger the characteristic length the stronger is the FFE.

From the graph shown in figure 2.6.2, threshold-voltage roll-off of the stack gate-dielectric MOSFET with  $k_{ox1}=60$  and  $k_{ox2}=3.9$  is closer to that of MOSFET with single  $\text{SiO}_2$  as gate dielectric. This indicates that the fringing-field effect of small-scaled MOSFET can be effectively suppressed by inserting a low-k interlayer between High-k dielectric and substrate. Therefore, a stack gate dielectric with a thin interlayer of low-k should be used to improve the threshold-voltage behaviors of MOSFET.

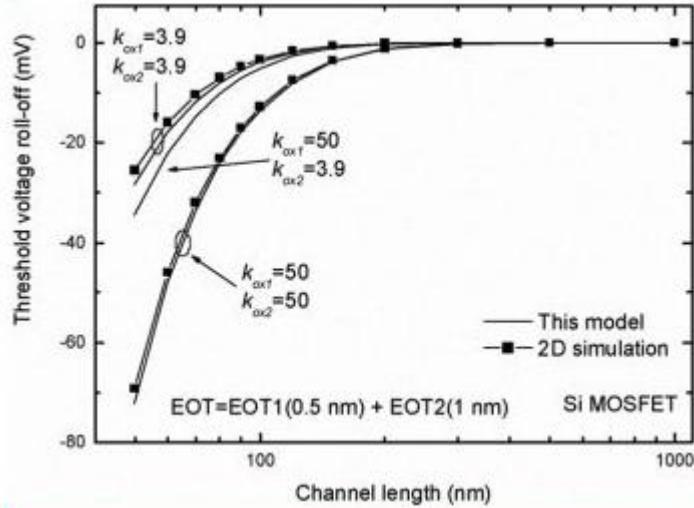


Figure 2.6.2: Threshold-voltage roll-off of Si MOSFET with single and stack gate dielectric versus channel length. [7]

### 2.6.2. Threshold voltage model for a Double-gate MOSFET

As the conventional single gate MOSFET is scaled down the short channel effect increases, leakage current increases and performance gets declined so the use of Double-gate MOSFET suppresses the short channel effect and increases the performance of device for further scaling.

Consequently, the variation in threshold voltage for double-gate MOSFET is much smaller than that of the conventional single-gate MOSFET. For further scaling of Double-gate MOSFET the threshold voltage can be controlled by using asymmetric double gate, in which two gates are of separate work-function. Another option can be by using N+ and P+ polysilicon gates. [8]

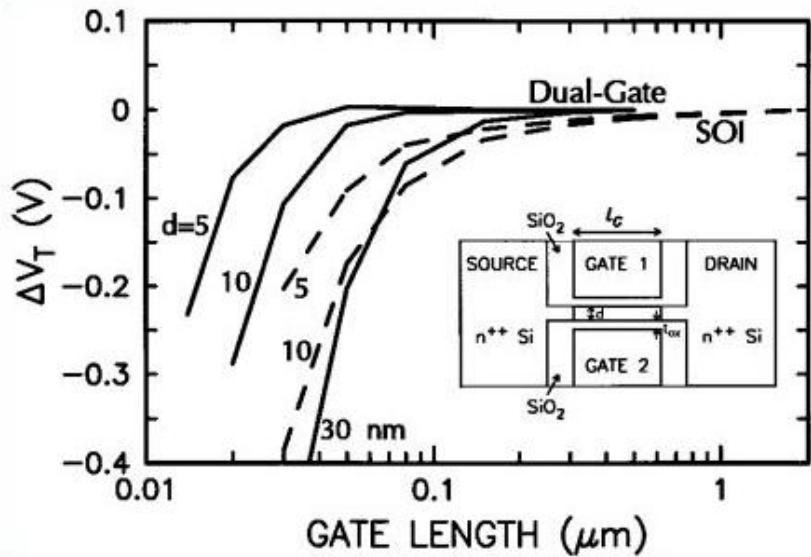


Figure 2.6.3: Simulated threshold voltage versus channel length, comparing short-channel effect of double-gate FET's (solid lines) with SOI MOSFET's (dashed Lines), where the threshold of the long channel FET's has been take as zero.[1]

### 2.6.3. Threshold voltage of Tri-gate FINFET

The threshold voltage of N-channel MOSFET is given below, [9]

$$V_{TH} = \phi_{ms} - \frac{Q_{ss}}{C_{ox}} + 2\phi_F + \frac{qN_A x_{dmax}}{C_{ox}}$$

The gate barrier is dependent on the difference of work function between metal and silicon. From the above equation it can be said that the threshold voltage is dependent on gate barrier. Therefore, the threshold voltage is higher for high barrier.

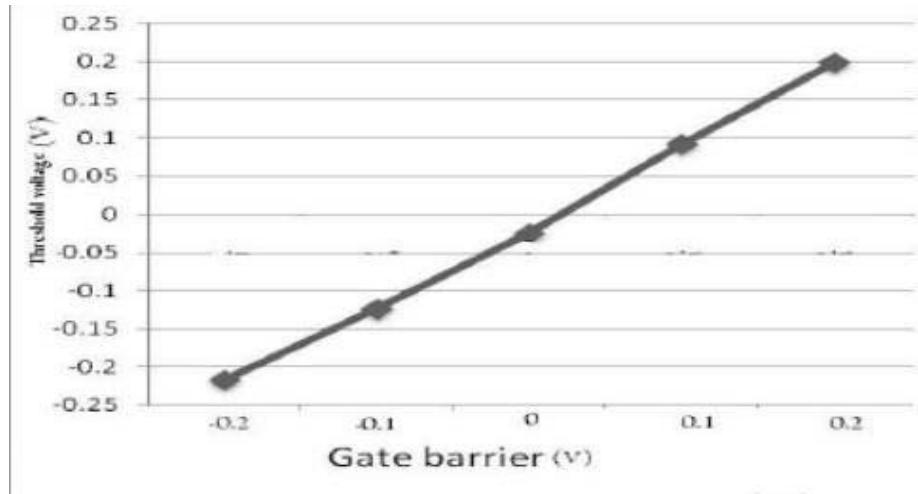


Figure 2.6.4: Variation of Threshold voltage in TG-FinFET with gate barrier. [9]

## 2.7 Summary

Scaling of MOSFET into the nanometer decreases the channel length which drastically reduces the performance of the CMOS devices. The roll-off threshold voltage due to short channel length effect increases as explained in section 2.3 is a major problem while approaching towards the nanometer scale. Therefore to keep such effect under control the gate oxide thickness must be reduced nearly in proportion to channel length. However, decreasing the gate oxide also lowers the threshold voltage as explained in section 2.2.

A smaller value of threshold voltage is desired as the device approaches the nanometer regime but as we approach into the nanometer regime the threshold voltage decreases drastically which results in the increases of leakage current. Different authors of respective IEEE journal have suggested various models with better control over threshold voltage at the nanometer regime. A threshold-voltage model of MOSFET with stack High-k gate dielectric is taken from an IEEE paper which is provided in section 2.6.1. This model is based on the work of Liu's [45] and F.Ji [46]. Figure 2.6.2 of this model shows improved threshold behaviors for stack high-k gate-dielectric MOSFET. The use of high-k as gate-dielectric reduces the tunneling current and decreases the hot carrier effect in a short channel and maintains the operating temperature of the devices at optimum level. When the devices are further scaled down such model becomes less effective, therefore new modified structures are introduced which helps in maintaining the Moore's law. Later in section 2.6.2 and 2.6.3 threshold model for Double-gate MOSFET and Tri-gate FinFET are discussed. By using such models the overall performance of CMOS devices increases significantly.

# **Chapter 3**

## **Temperature**

### **3.1 Introduction:**

To successfully design a CMOS transistor, one must have a thorough understanding of the temperature behavior of MOS transistors. Therefore, in this Chapter, we consider the temperature dependency of some basic semiconductor (Si) physical properties.

As transistors are scaled into Nano meter range, classical mechanics fails and quantum mechanics play a big role in the properties of semiconductor. All the parameter studied here are for the Nano meter range.

We will emphasize on energy band gap, intrinsic carrier concentration, extrinsic carrier concentration and Fermi potential. We will also emphasize on the MOS transistor parameters threshold voltage and carrier mobility. The equations which are generally used to describe the temperature behavior of the CMOS threshold voltage and mobility are presented. The presence of the zero temperature coefficient (ZTC) point in the transconductance characteristics of a CMOS transistor is studied and the conditions under which such a point can exist are investigated. Finally we examine the effect of variation of temperature on gate tunneling current at different values of Equivalent oxide thickness (EOT) and Substrate doping (Na) at different technology nodes.

### **3.2 Effect of Temperature on Energy Bandgap:**

Energy bandgap is the separation between the energy of the lowest conduction band and the highest valence band and is denoted by  $E_g$ . It is one of the most important parameter of semiconductor physics.

Under normal atmosphere and room temperature the value of bandgap for silicon is 1.12 eV. For highly doped materials the bandgap become smaller. Figure 1 shows variation of bandgap as a function of temperature. Bandgap approaches 1.17 eV for silicon at 0K. The temperature coefficient  $\frac{dE_g}{dT}$  is negative for silicon.  $E_g$  has a weak temperature dependence.

$E_g$  of silicon is given by [11]

$$E_g(T) = E_g(0) - \frac{\alpha T^2}{T + \beta}$$

Where  $E_g(0) = 1.170$ ,  $\alpha = 4.73 \times 10^{-4}$ ,  $\beta = 636$

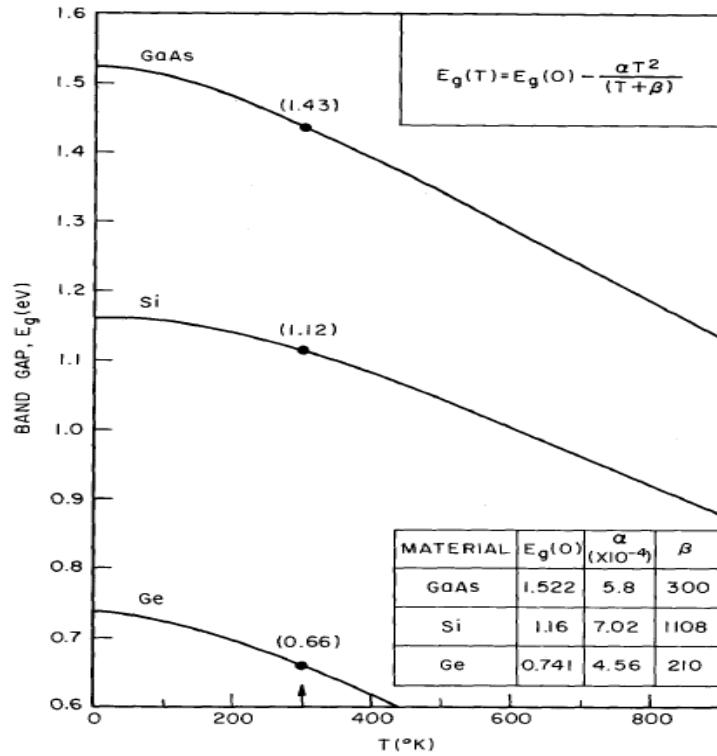


Figure 3.2.1: Energy bandgap of GaAs , Si ,Ge as a function of temperature[11].

### 3.3 Effect of Temperature on intrinsic carrier concentration:

Conductivity of a semiconductor depends on the electron and hole concentration. These quantities are determined as follows-

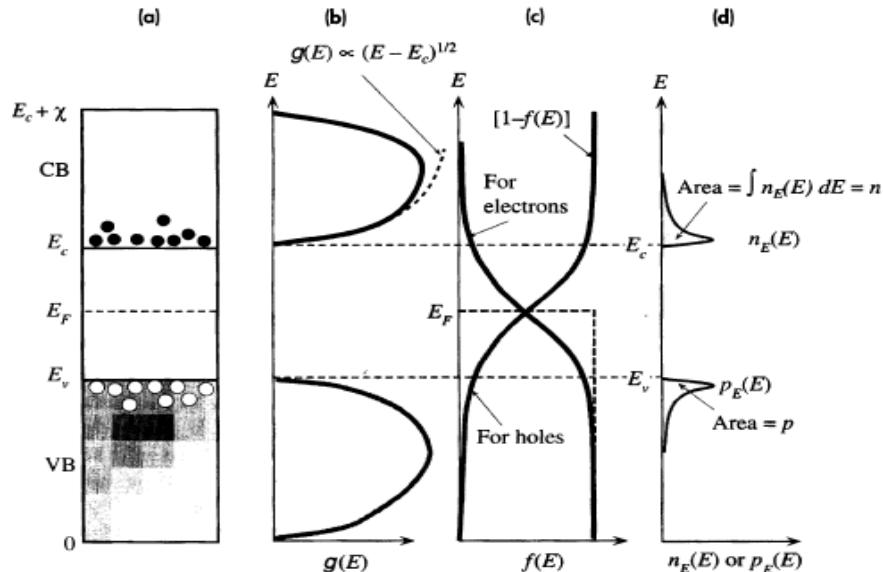


Figure 3.3.1 (a) Energy band diagram (b) Density of states (c) Fermi-Dirac probability function (d) energy density of electron in the CB [12].

Let  $g_{cb}(E)$  = density of state in conduction band (CB)

This is the number of state per unit energy per unit volume.

The probability of finding an electron in a state with energy E is given by Fermi-Dirac function  $f(E)$  [12]

$$f(E) = \frac{1}{1 + \exp\left(\frac{(E - E_F)}{KT}\right)}$$

The actual number of electrons per unit energy per unit volume  $n_e(E)$  in the CB is [12]

$$n_e(E)dE = g_{cb}(E) f(E) dE$$

Therefore electron concentration n is

$$n = \int_{E_c}^{E_c + \varkappa} n_e(E) dE = \int_{E_c}^{E_c + \varkappa} g_{cb}(E) f(E) dE$$

Assuming  $(E_c - E_F) \gg KT$ , and assuming that number of electron in CB is far less than the number of states in this band

we get

$$f(E) = \exp\left(-\frac{(E - E_F)}{KT}\right)$$

Taking upper limit to be  $E = \infty$  rather than  $E_c + \varkappa$ . (Since  $f(E)$  decays rapidly with energy so that  $g_{cb}(E) f(E) \rightarrow 0$  near the top of the band )

Since  $g_{cb}(E) f(E)$  is significant only close to  $E_c$  we get[12]

$$g_{cb}(E) \approx \frac{(\pi 8 \sqrt{2}) m_e^{3/2}}{h^3} (E - E_c)^{1/2}$$

Thus

$$n \approx \frac{(\pi 8 \sqrt{2}) m_e^{3/2}}{h^3} \int_{E_c}^{\infty} (E - E_c)^{1/2} \exp\left(-\frac{(E - E_F)}{KT}\right) dE$$

$$n = N_c \exp\left(-\frac{(E_c - E_F)}{KT}\right)$$

Where  $N_c$  is the effective density of states at the CB edge.

$$N_c = 2 \left( \frac{2\pi m_e K T}{h^2} \right)^{3/2}$$

Similarly hole concentration p is

$$p = \int_0^{E_v} p_e(E) dE = \int_0^{E_v} g_{vb}(E)[1 - f(E)] dE$$

Assuming  $E_F$  is few volts above  $E_v$

$$p = N_v \exp\left(-\frac{(E_F - E_v)}{KT}\right)$$

Where  $N_v$  is the effective density of states at the VB edge.

$$N_v = 2\left(\frac{2\pi m_h KT}{h^2}\right)^{3/2}$$

As  $n=p$ , the intrinsic concentration  $n_i$  is given by [14]

$$np = n_i^2 = N_c N_v \exp\left(-\frac{(E_c - E_v)}{KT}\right)$$

$$n_i = \sqrt{N_c N_v} \exp\left(-\frac{E_g}{KT}\right)$$

$$= 4.9 \times 10^{15} \left(\frac{m_e m_h}{m_0^2}\right)^{3/2} T^{3/2} \exp\left(-\frac{E_g}{KT}\right)$$

$$= A T^{3/2} \exp\left(-\frac{E_g}{KT}\right)$$

Here

$m_e$  = effective electron mass

$m_h$  = effective hole mass

K = Boltzmann's constant

T = absolute temperature

$E_g$  = energy of bandgap

$m_0$  = free electron mass

The concentration  $n_i$  of intrinsic electrons and holes depends on the amount of energy needed to break a bond (e.g., the energy bandgap) and on the amount of energy available (e.g., the thermal energy as characterized by the temperature)

$E_g$  can be said as the binding energy and  $KT$  as disruptive force.

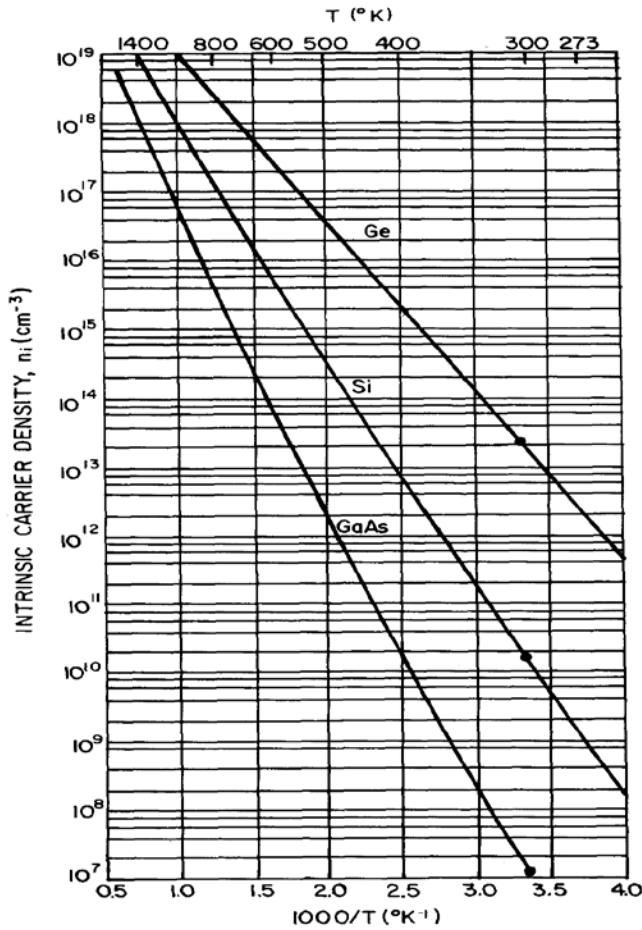


Figure 3.3.2: Intrinsic carrier concentration of Ge, Si and GaAs as a function of reciprocal of temperature [11].

### 3.4 Effect of Temperature on Carrier Concentration:

When impurity atoms are added, to preserve charge neutrality Fermi level must adjust itself. If donor impurity with concentration  $N_d(cm^{-3})$  are added to the crystal, to preserve charge neutrality total negative charges must equal total positive charges.

$$n = N_D^+ + p$$

n = electron density

p = hole density

$N_D^+$  = number of ionized donors

At low temperature all donor will not be ionized. We need to know the probability  $f_d$  ( $E_d$ ) of finding an electron in a state with energy  $E_d$  at a donor. [12]

$$f_d(E_d) = \frac{1}{1 + \frac{1}{g} \exp\left(\frac{E_d - E_F}{KT}\right)}$$

Then number of ionized donor [12]

$$\begin{aligned}
 N_D^+ &= N_d \times (\text{probability of not finding an electron at } E_d) \\
 &= N_d \times (1 - f_d(E_d)) \\
 &= \frac{N_d}{1 + 2 \exp\left(\frac{E_F - E_d}{KT}\right)}
 \end{aligned}$$

$N_d$  = donor concentration

Here  $g=2$  because electron state at donor can take an electron with spin either up or down but not both.

Similarly number of ionized acceptor [11]

$$N_A^- = \frac{N_A}{1 + g \exp\left(\frac{E_A - E_F}{KT}\right)}$$

$N_A$  = acceptor concentration

Here  $g=4$  because in silicon each acceptor impurity level can accept one hole of either spin and the impurity level is doubly degenerate as a result of the two degenerate valence bands at  $k=0$

When impurity atoms are added at relatively elevated temperature, most of the donors and acceptors are ionized. Carrier concentration  $n_c$  is [11]

$$n_c = \frac{1}{2} \left[ (N_D - N_A) + \sqrt{(N_D - N_A)^2 + 4n_i^2} \right] \quad \text{for n-type}$$

$$n_c = \frac{1}{2} \left[ (N_A - N_D) + \sqrt{(N_D - N_A)^2 + 4n_i^2} \right] \quad \text{for p-type}$$

$N_D$  = donor concentration

$N_A$  = acceptor concentration

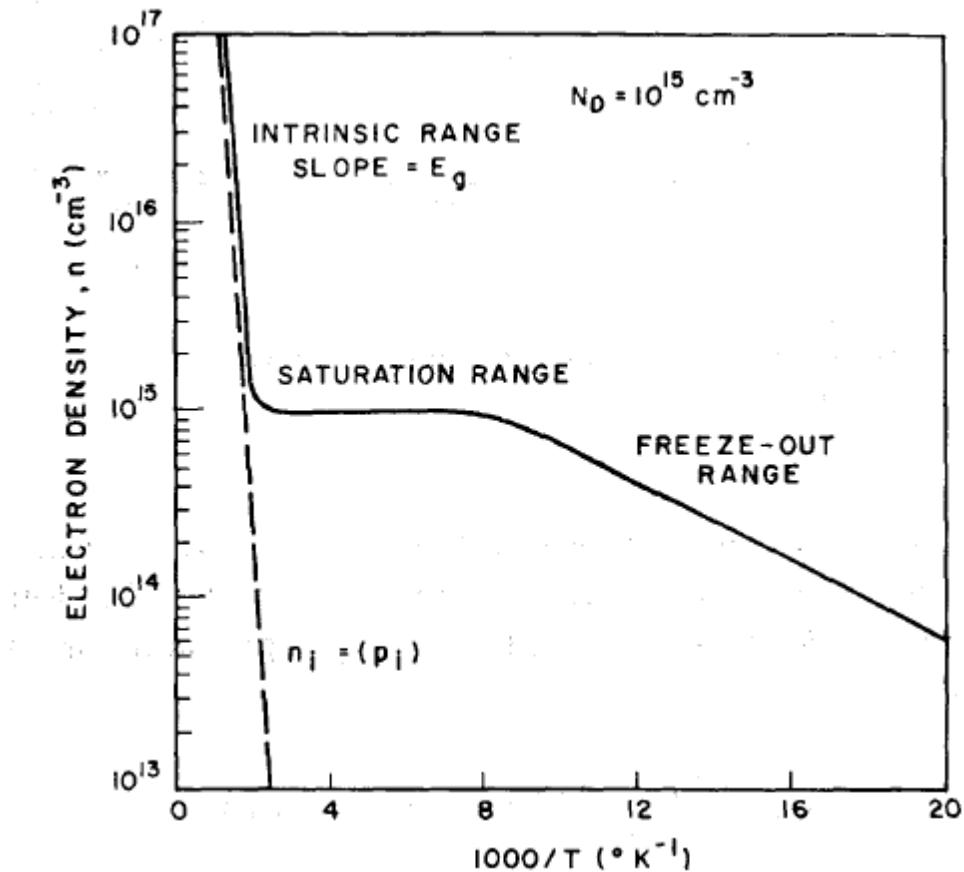


Figure 3.4.1: electron density as a function of temperature for silicon with a donor concentration of  $N_D=10^{15} \text{ cm}^{-3}$  [11].

Figure 3.4.1 shows the carrier concentration of Si as a function to temperature for donor concentration  $N_D=10^{15} \text{ cm}^{-3}$ .

At low temperature all impurities are not ionized. So carrier density is less than the donor concentration. This region is called freeze-out region.

As temperature increase all impurities get ionized and  $n_c = N_D$ . This region is called extrinsic region and have a wide range.

At higher temperature intrinsic carrier concentration  $n_i$  is increased and at temperature at which this  $n_i$  becomes equal to  $N_D$  is called intrinsic temperature. As we can see, this temperature can be delayed by increasing the impurity density.

### 3.5Effect of Temperature on Fermi level

For intrinsic semiconductor the Fermi level lies in the midway between the conduction and valence bands. In n-type semiconductor the Fermi level is closer to the conduction band. In p-type semiconductor the Fermi level is closer to the valence band.

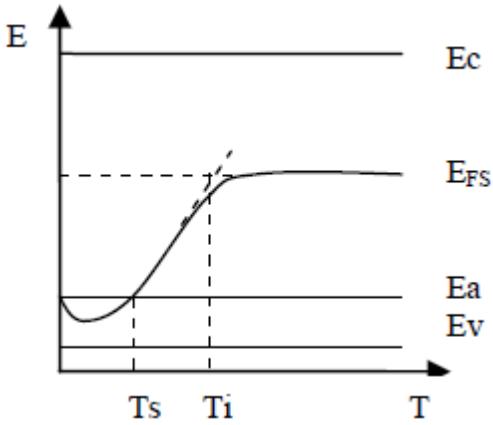


Figure 3.5.1: Variation of Fermi level with temperature in NMOS [13].

For temperature between  $T_s$  and  $T_i$  fermi level dependency on temperature is given by [13]

$$E_{FS} = - (E_g) - [KT \times \log\left(\frac{N_a}{N_v}\right)]$$

$E_{FS}$  = Fermi Level

$T_s$  is the temperature below which impurities are weakly ionized and  $T_i$  is the temperature at which silicon becomes intrinsic.

### 3.6 Effect of temperature on Threshold Voltage

The expression for threshold voltage  $V_{TH}$  of MOS transistor is [32]

$$V_{TH} = \varphi_{ms} \pm \frac{Q_{ss}}{c_{ox}} + 2\varphi_F + \Delta V_T(N_i) \pm \gamma(N_s, t_{ox}, L, W) \sqrt{2\varphi_F + V_0}$$

Where,

$\varphi_{ms}$  = metal-silicon work function difference

$Q_{ss}$  = surface-state charge density per unit area

$c_{ox}$  = gate oxide capacitance per unit area

$\varphi_F$  = Fermi potential of the substrate

$\Delta V_T(N_i)$  = threshold shift owing to a channel implant  $N_i$  with depth  $d_i$

$\gamma$  = body effect constant depend on substrate doping  $N_s$

$N_s$  = substrate doping

$t_{ox}$  = gate oxide thickness

L = channel length

W= channel width

$V_0(N_i, N_s, d_i)$  = correction term owing to the threshold shift implant

The major effect on variation of threshold voltage are Fermi level,  $\varphi_F$  and gate-semiconductor work function difference,  $\varphi_{ms}$ . We have already seen the relation between Fermi level and temperature. For a silicon gate which is doped oppositely to the substrate, the contact potential is determined by the pn product. For an n-type doped gate the gate-semiconductor work function is

$$\varphi_{ms}(T) = -\frac{KT}{q} \ln\left(\frac{N_s N_p}{n_i^2}\right) \quad (\text{NMOS})$$

$$\varphi_{ms}(T) = -\frac{KT}{q} \ln\left(\frac{N_s}{N_p}\right) \quad (\text{PMOS})$$

Where,

$N_p$  = carrier concentration in polysilicon gate

The temperature variant term in  $\varphi_{ms}$  are the intrinsic carrier concentration,  $n_i$ , and the thermal voltage,  $\frac{KT}{q}$ . Oxide capacitance and ionization of surface states have minor effect on threshold voltage which can be safely ignored.

Threshold voltage decrease approximately linearly with temperature

$$V_{TH}(T) = V_{TH}(T_0) - \alpha_{VT}(T - T_0)$$

Where

$T_0$  = reference temperature

$\alpha_{VT}$  = Temperature coefficient of the threshold voltage

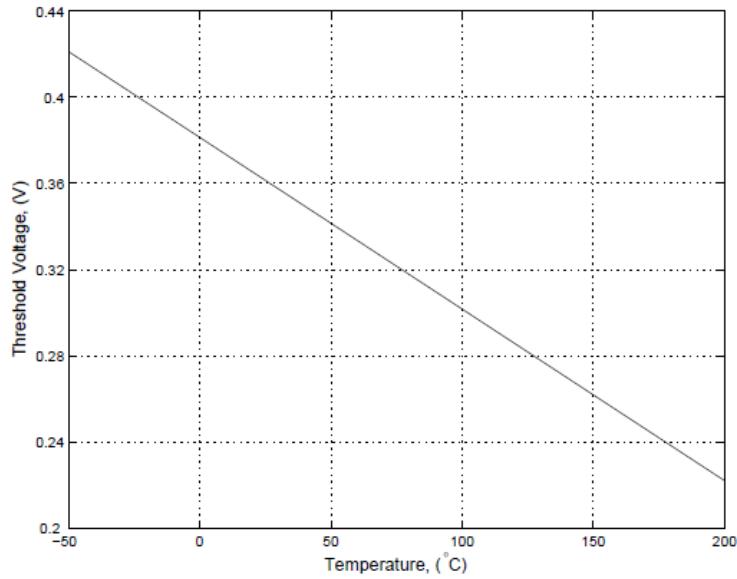


Figure3.6.1: Threshold voltage of an NMOS transistor ( $W/L=2/1$ ) as a function of temperature

Temperature coefficient of the threshold voltage is

$$\alpha_{VT} = \left| \frac{\partial V_{TH}}{\partial T} \right| = \left| \frac{\varphi_{ms}}{T} + 2 \frac{\varphi_F}{T} + \frac{\gamma(N_s, t_{ox}, L, W)}{\sqrt{2\varphi_F + V_0}} \right|$$

The length of the transistor affects the value of the threshold voltage temperature coefficient,  $\alpha_{VT}$ . For a short channel transistor, the magnitude of  $\alpha_{VT}$  is smaller due to the fact that part of the depletion charge associated with channel formation is depleted from the source and drain rather than from the gate. This effect leads to a reduction of the body effect,  $\gamma$ , resulting in the decrease of the  $\alpha_{VT}$ .

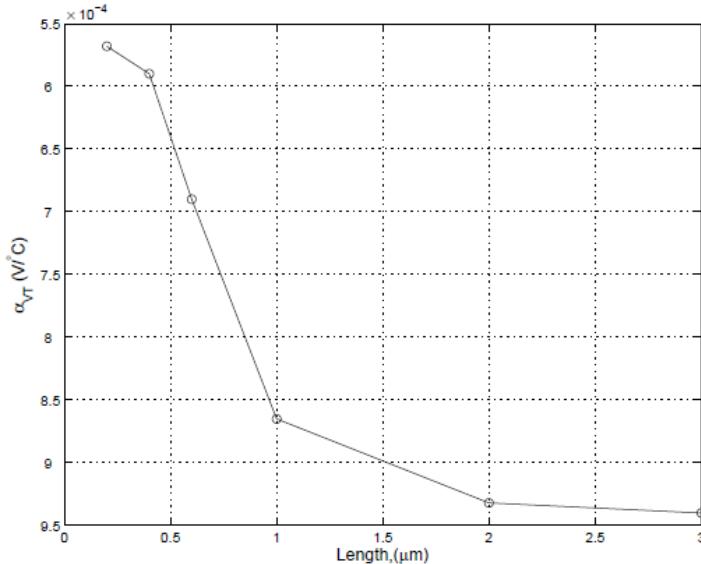


Figure3.6.2: Change of  $\alpha_{VT}$  with the length of an NMOS transistor with  $W=5$

### 3.7 Effect of Temperature on Carrier Mobility

The carrier mobility in Si is affected by two basic scattering mechanisms: lattice scattering and impurity scattering [11]. Lattice scattering results from thermal vibrations of the lattice atoms at any temperature above absolute zero. The interaction between the carriers and the lattice allow some energy to be transferred, leading to scattering. The mobility of carriers impeded by impurity scattering is [11]

$$\mu_L \sim (m^*)^{-5/2} T^{-3/2}$$

$\mu_L$  = mobility of carriers impeded by lattice scattering

$m^*$  = conductivity effective mass

Impurity scattering is caused by distortion in the lattice, due to impurities such as dopants. The mobility of carriers impeded by impurity scattering is [11]

$$\mu_I \sim (m^*)^{-1/2} N_I^{-1} T^{3/2}$$

$\mu_I$  = mobility of carriers impeded by Impurity scattering

$m^*$  = conductivity effective mass

$N_I$  = concentration of impurities

We can see that as the impurity concentration increases, the mobility decreases. Also, the mobility decreases as  $m^*$  increases. Therefore, for a given impurity concentration, the electron mobility is larger than the hole mobility.

At low temperatures, the carriers have low kinetic energy and their time of passage past ionized impurities is longer than at high temperature. Therefore, at low temperatures, impurity scattering dominates. As the temperature is increased, the carriers move faster, they remain near the impurity atom for a shorter time and therefore the collisions due to impurity scattering become less significant than the collisions with the neutral atoms of lattice. Thus, at high temperature, lattice scattering is the major scattering process and the carrier mobility will be proportional to  $T^{3/2}$ . The combined mobility, which includes the two mechanisms above, is

$$\frac{1}{\mu} = \left( \frac{1}{\mu_L} + \frac{1}{\mu_I} \right)$$

In the analysis of semiconductor devices it is often necessary to know the magnitude of the mobility for different doping densities and temperatures.

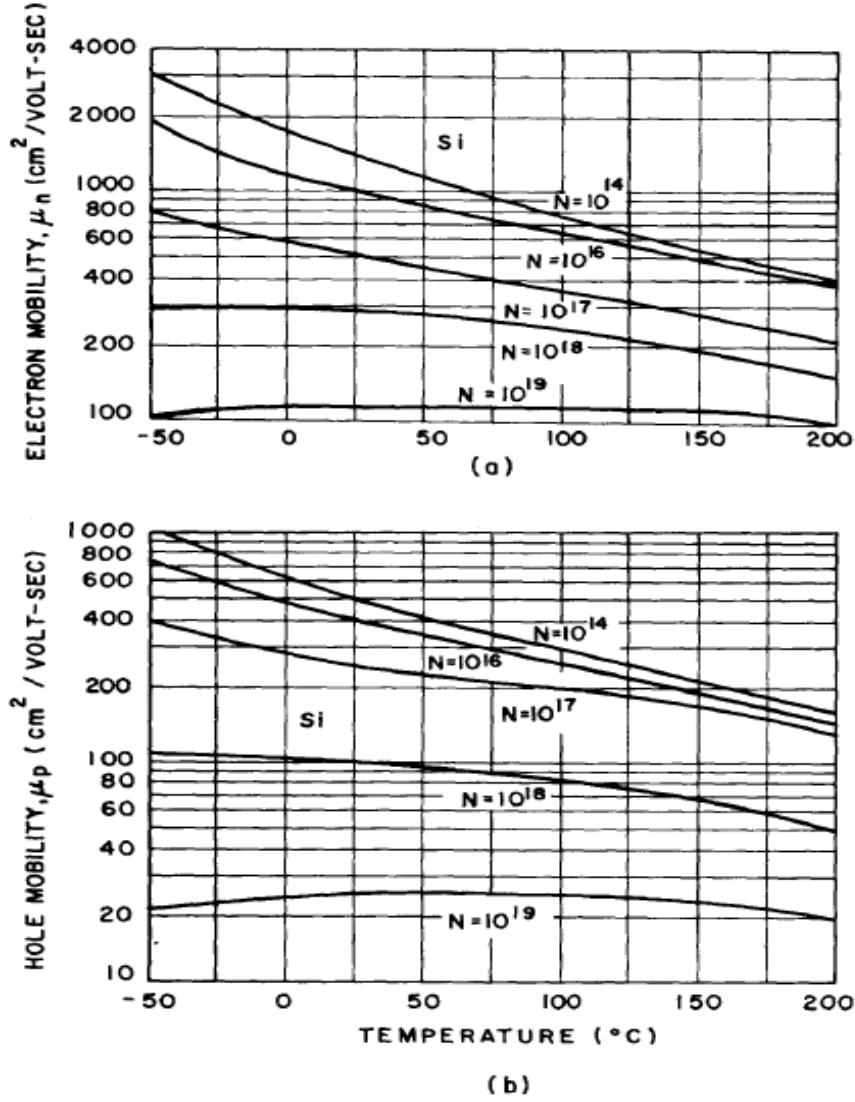


Figure 3.7.1: Mobility as a function of temperature[11].

As it can be seen, lattice scattering dominates and the mobility decreases with temperature for lightly doped samples (e.g., the sample with doping of  $10^{14}$ ). The effect of impurity scattering becomes more significant at low temperatures for heavily doped samples and the mobility increases as the temperature increases. It can also be seen that, for a given temperature, the mobility decreases as the impurity concentration increases because of enhanced impurity scattering.

### 3.8 Zero temperature coefficients:

The drain current of an NMOS transistor in the saturation region is given by<sup>[32]</sup>

$$I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2$$

$\mu_n$  = carrier mobility

$V_{TH}$  = Threshold voltage

The threshold voltage and the mobility are the main temperature dependent parameters. As the temperature increases, both the threshold voltage and the mobility decrease. But the decrease of  $V_{TH}$  and the decrease of  $\mu_n$  have opposing effects on the drain current; a lower threshold voltage tends to increase the drain current, but a lower mobility tends to decrease it. At some value of  $V_{GS}$ , this mutual compensation of mobility and threshold voltage results in a zero temperature coefficient, (ZTC), at a particular bias point of a MOS transistor.

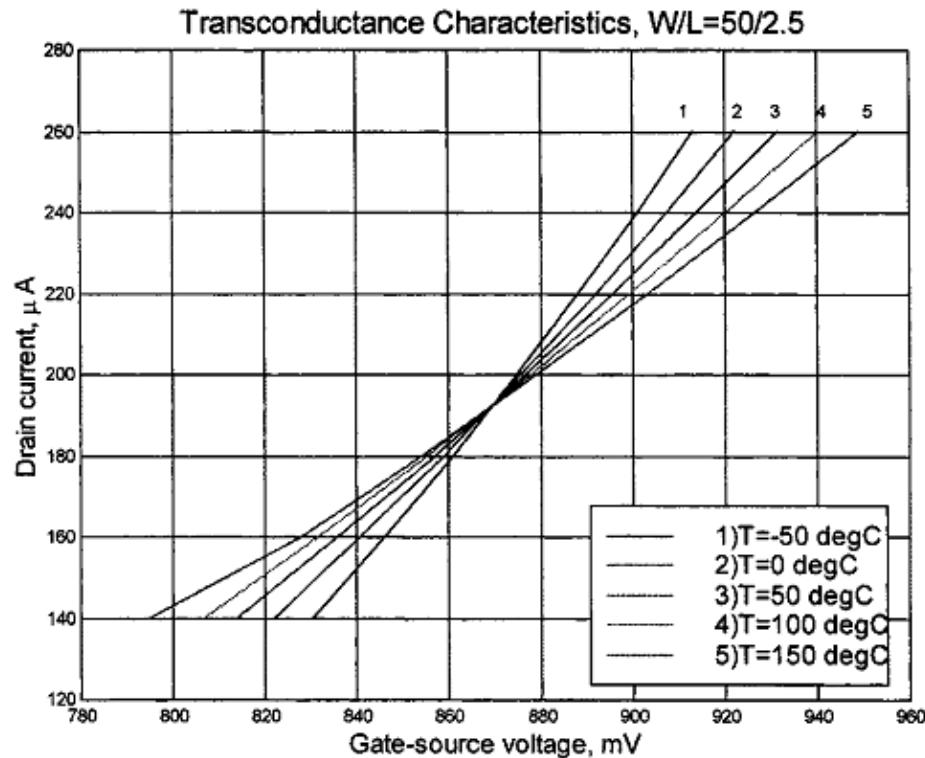


Figure 3.8.1: Transconductance characteristic for n-channel transistor[32]

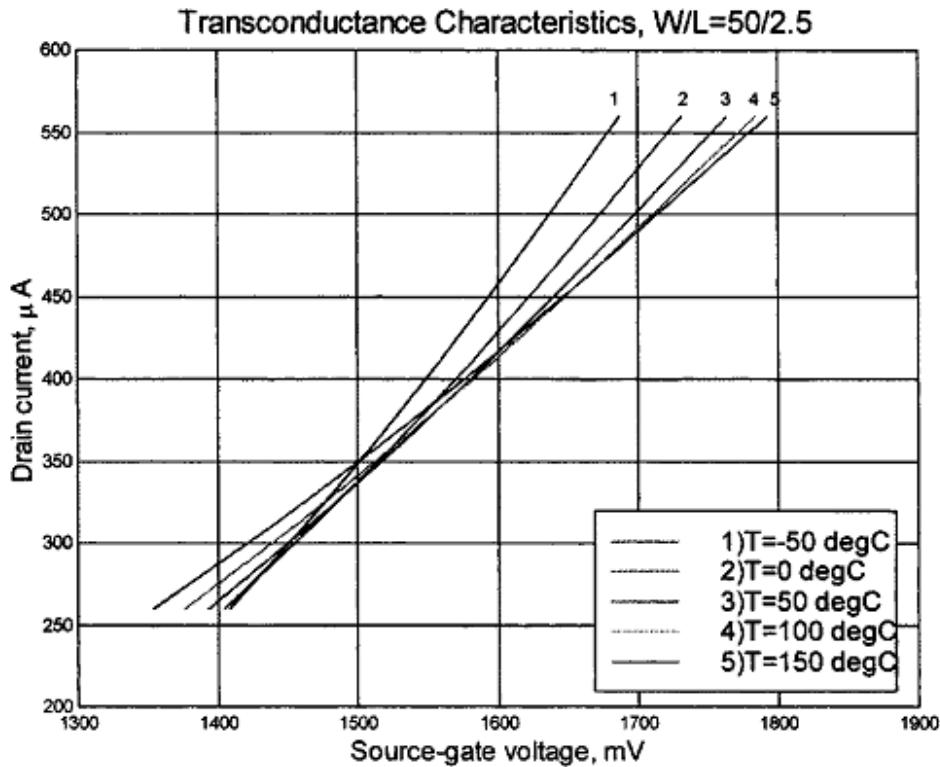


Figure 3.8.2: Transconductance characteristic for p-channel transistor[32]

It can be seen that the transconductance characteristics have a common intercept point. The temperature independent voltage,  $V_{GSF}$  is

$$V_{GSF} = V_{TH}(T_0) - \alpha_{VT}(T - T_0)$$

Temperature independent drain current,  $I_{DF}$  for this bias voltage is

$$I_{DF} = \frac{\mu_n(T_0) T_0^2 C_{ox}}{2} \left( \frac{W}{L} \right) \alpha_{VT}^2$$

### **3.9 Effect of temperature variation on gate tunneling current at different values of Equivalent Oxide Thickness (EOT):**

Total gate tunneling current includes both direct tunneling ( $J_t$ ) and Source Drain Extension (SDE) tunneling ( $J_{sd}$ ). The  $J_{\text{limit}}$  is the maximum limit of gate tunneling as specified by ITRS at each technology node.

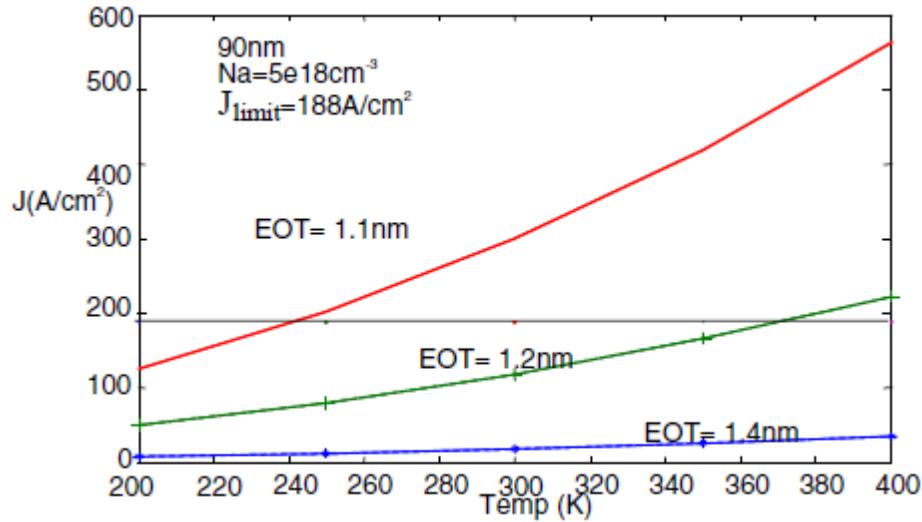


Figure 3.9.1 Variation of gate tunneling current at 90nm technology node for different values of EOT[13].

Fig 3.9.1 shows that at 90nm technology node, for  $V_{gs}=1V$ , EOT can be scaled down to 1.2nm without exceeding the  $J_{limit}$  up to 370K temperature.

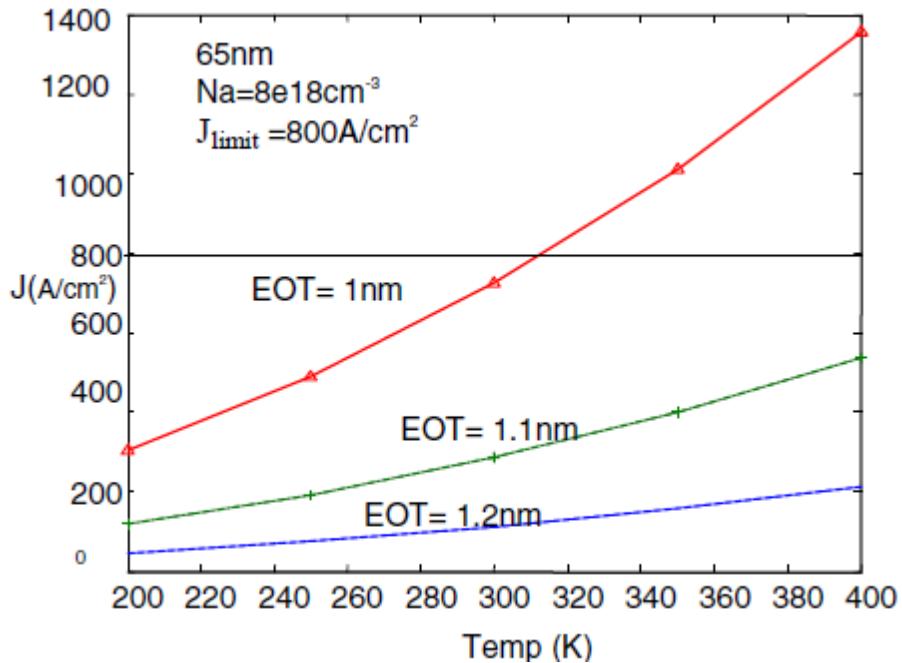


Figure 3.9.2 Variation of gate tunneling current at 65nm technology node for different values of EOT[13].

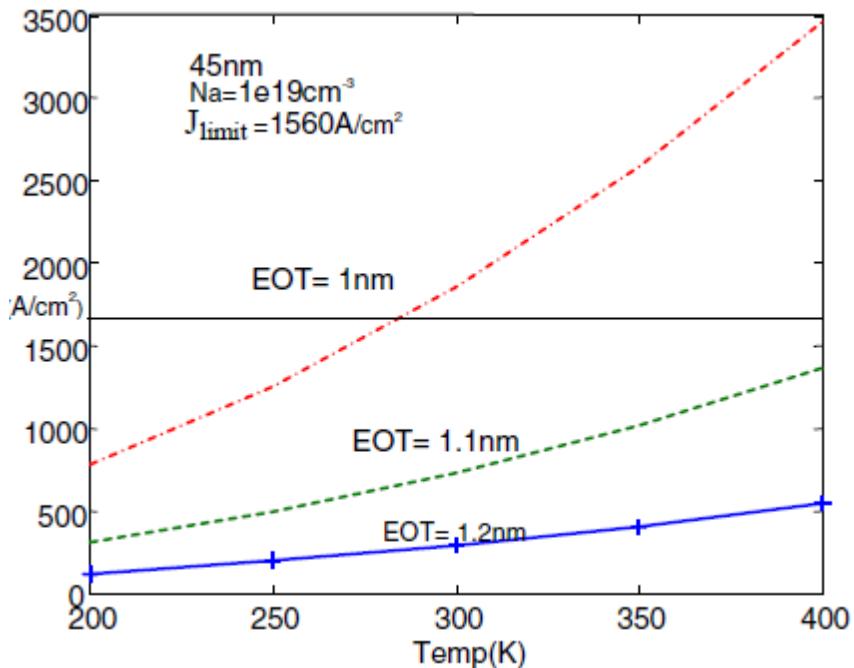


Figure 3.9.3 Variation of gate tunneling current at 45nm technology node for different values of EOT[13].

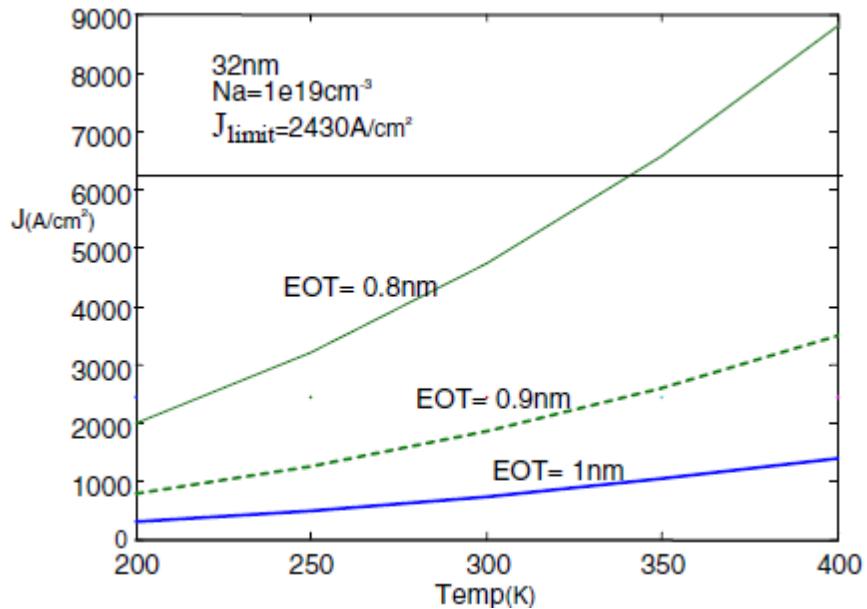


Figure 3.9.4 Variation of gate tunneling current at 32nm technology node for different values of EOT[13].

Figure 3.9.2 and Figure 3.9.3 shows safe range of EOT lies at 1.1nm for 65nm and 45nm technology nodes without exceeding  $J_{\text{limit}}$  for entire range of temperature.

Figure 3.9.4 shows that for 32nm technology node EOT can be scaled down to 0.8nm working at room temperature.

For each technology node, leakage due to gate tunneling currents is not a strong function of temperature until the EOT becomes sufficiently thin. When designing a nMOS device for use in a high temperature environment (greater than 400K) it has been shown that the EOT must be considered. For the 90 nm, 65 nm, 45 nm, and 32 nm technology nodes nMOS devices with an

EOT equal to or greater than 1.2 nm, 1.1 nm, 1.1 nm, and 0.8 nm respectively should be considered for use in high temperature environments. All the plots clearly reveal that the gate tunneling significantly increases even for 0.1nm decrease in EOT and situation worsens if operating temperature increases even moderately above room temperature. If an MOS device is designed for use in a low temperature environment (less than 240 K) then the designer has much more freedom in choosing the EOT.

### 3.10 Effect of temperature variation on gate tunneling current at different values of Substrate Doping (Na):

The gate tunneling current increases as the substrate doping level decreases because at lower substrate doping levels the formation of inversion layer requires less gate bias.

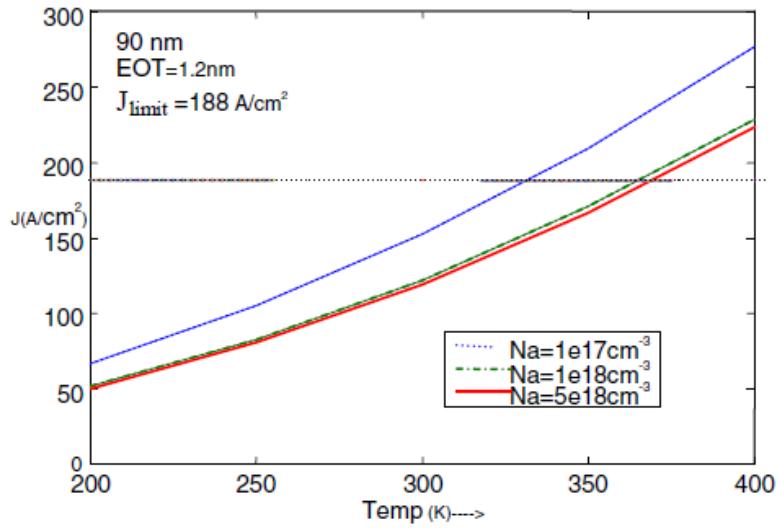


Figure 3.10.1 Variation of gate tunneling current at 90nm technology node for different values of Na [13].

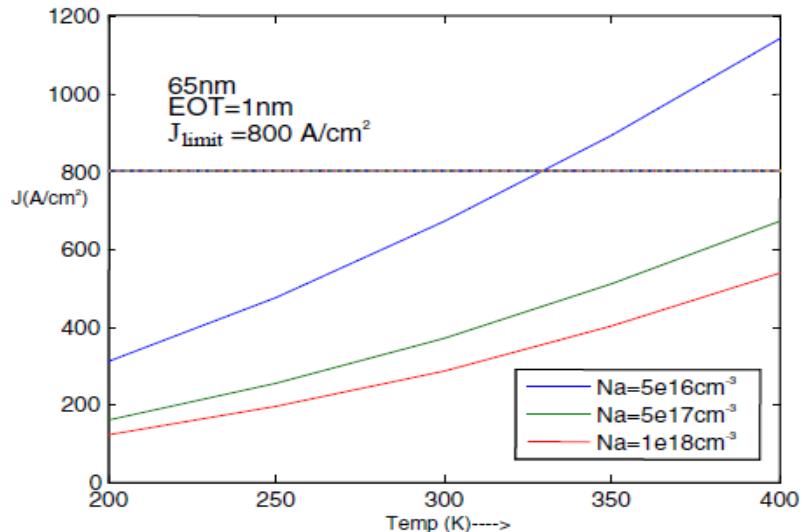


Figure 3.10.2 Variation of gate tunneling current at 65nm technology node for different values of Na [13].

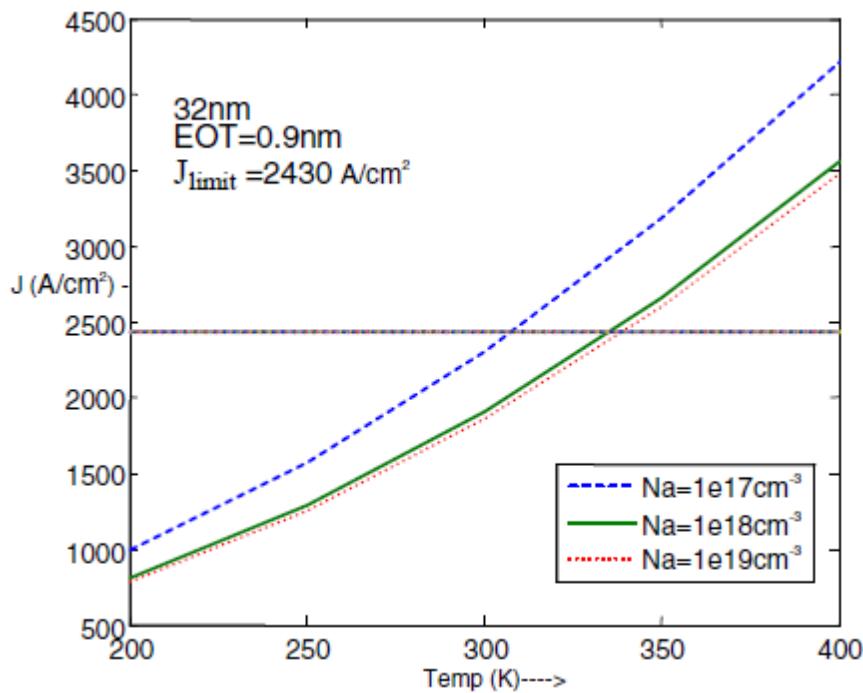
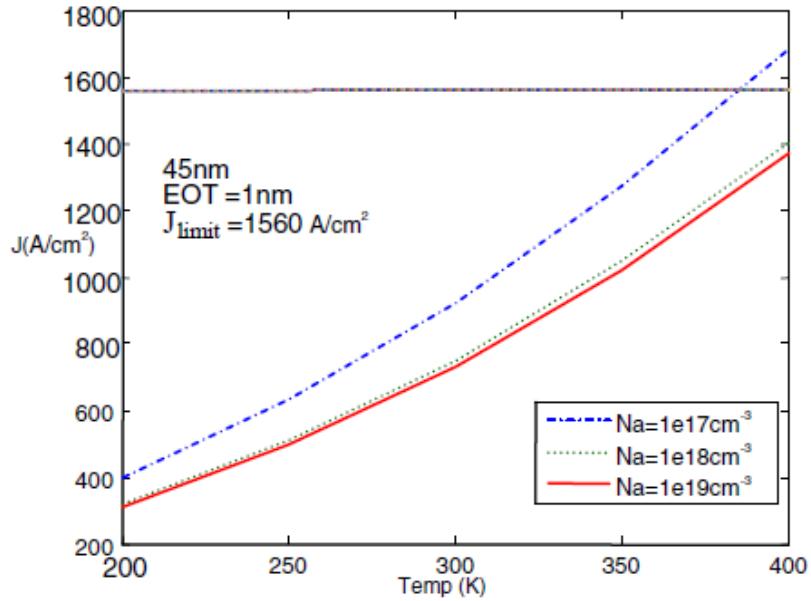


Figure 3.10.1 to Figure 3.10.4 show the dependencies of gate tunneling current on Na and temperature. It is seen that gate tunneling current varies moderately with Na. At lower technology node and at higher operating temperature Na should be controlled carefully.

### **3.11 Summary:**

In this chapter we have studied the effect of temperature on different semiconductor parameters. The first parameter was energy bandgap which S. M. Sze has shown decreases with temperature in Figure 3.2.1. Then we studied the parameter intrinsic carrier concentration. We saw that intrinsic carrier concentration increases with temperature as shown in Figure 3.3.2. Then in Figure 3.4.1 we saw the variation of carrier concentration with temperature. G. Joshi showed how fermi level varies with temperature in Figure 3.5.1. In section 3.6 we studied the effect of temperature on MOS transistor parameter Threshold voltage. Figure 3.6.1 shows how threshold voltage decreases with temperature. In Figure 3.7.1 S. M. Sze showed the decrease in mobility with temperature. In section 3.8 I. M. Filanovsky showed temperature coefficient, (ZTC) from mutual compensation of mobility and threshold voltage in Figure 3.8.1. In Figure 3.9.1 to Figure 3.9.4 G. Joshi showed the Variation of gate tunneling current for different values of EOT. He also showed the Variation of gate tunneling current for different values of Na from Figure 3.10.1 to Figure 3.10.4.

## **Chapter 4**

### **Gate-oxide thickness reduction**

#### **4.1 Introduction:**

Nanotechnology is a revolutionary improvement in the modern technological field. In last two decade this field has done enormous improvement especially in electronics field. Due to nanotechnology transistor size became smaller. In 1965 Gordon Moore noted that the number of transistor on a chip doubled every 18 to 24 months. According to his prediction modern chip can accommodate thousands of transistors in 1 mm space.

This scaling methodology has worked very well for several decades, but as silicon CMOS technology advances into the nanometer regime, fundamental and practical limits impede the traditional scaling of transistors. Gate dielectric thickness has been the single most important dimension to enable device scaling. Gate dielectric scaling increase capacitive coupling as well as on transistor drive current ( $I_{on}$ ).

Many phenomenon that affect the gate oxide thickness reduction ( $t_{ox}$ ) are dielectric scaling, equivalent gate-oxide (EOT), direct tunneling etc. This chapter demonstrates the concept of choosing right high-k dielectric material as gate-oxide and modeling of direct tunneling current through the gate-oxide.

#### **4.2 Background**

Semiconductor is type of material whose conductivity is in between the conductor and insulator. This means conductivity of semiconductors roughly in the range of  $10^{-8}$  to  $10^3$  siemens per centimeter ( $S \cdot cm^{-1}$ ) [15]. There are various types of semiconductor materials. Such as scrySTALLINE solids, amorphous and liquid semiconductors. These include hydro genated amorphous silicon and mixtures of arsenic, selenium and tellurium in a variety of proportions. But among all silicon is the most used semiconductor in the electronic industries. According to band theory semiconductor characteristics can be explained.

Structure of a material depends on the energy of its internal atoms. This energy can vary over wide range. This range is called energy band. To demonstrate the difference among semiconductor, insulator and conductor, band diagram must be drawn. Figure 4.2.1 shows the energy gap between conductor, semiconductor and insulator. Due to small energy gap between valance band and conduction band, electron can be easily excited and jump to conduction. So semiconductor sometimes acts as conductor and sometimes acts as insulator.

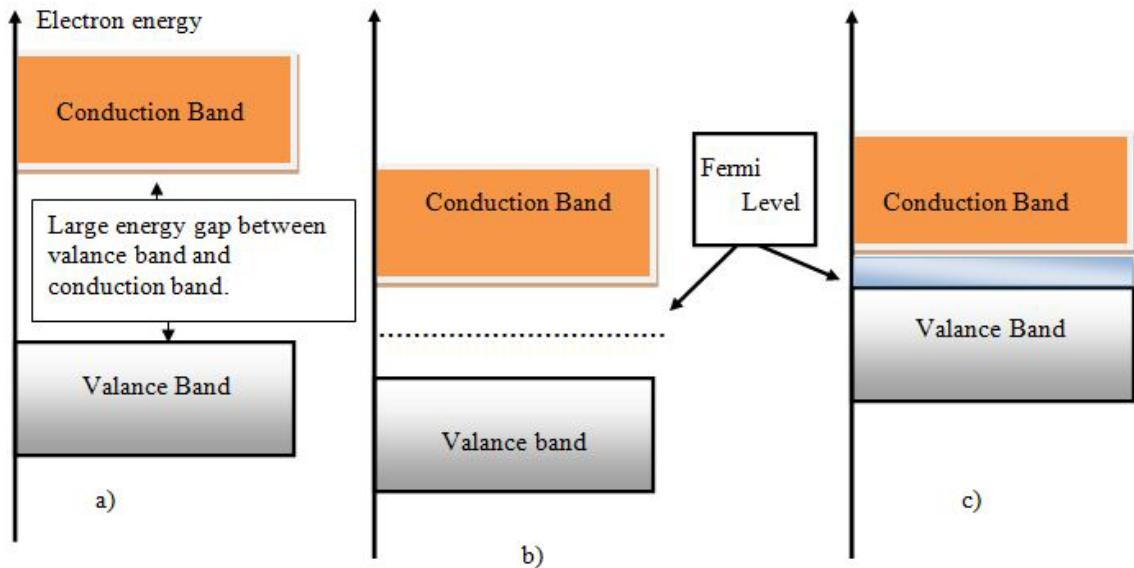


Figure 4.2.1: Energy band diagram of a) insulator, b) semiconductor and c) conductor [16].

Dielectric material is one kind of insulator. It is characterized by dipole. Dipole (Figure 4.2.2) is defined as a pair of opposite charges of magnitude  $q$  is defined as the magnitude of the charge times the distance between them and the defined direction is toward the positive charge. Usually dipoles are randomly distributed in the material. But when an electric field is applied, dipoles become polarized. As electric field increases, polarization also increases. So dielectric must be good insulator to minimize leakage current.

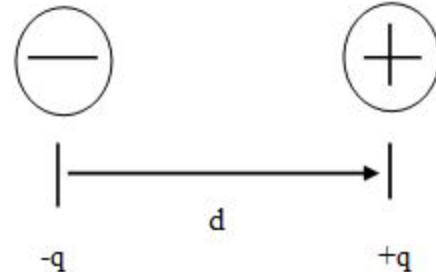


Figure 4.2.2: Electric Dipole

CMOS consist of N-MOSFET and P-MOSFET. Typical example of a CMOS is inverter. In figure 4.2.3, construction of a MOSFET has shown.

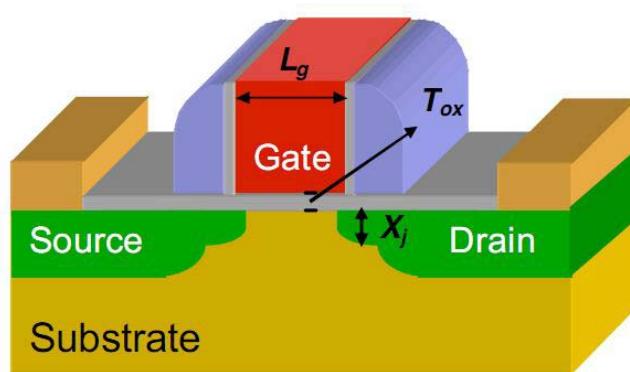


Figure 4.2.3: Schematic diagram of the bulk MOSFET.

As shown in the figure 4.1.3 that oxide or  $t_{ox}$  is very small in thickness. Usually it is in nanoscale regime. During scaling the size of transistor, gate oxide also need to be scaled. So the possibility of flowing leakage current also increases. In nanoscale regime CMOS and other transistors are switching so fast that even small amount of leakage or  $I_{off}$  can degrade the performance of the device. Modern electronic appliances such as laptops, smart phones, and tablets are using day by day and they are running on lithium-ion battery. If the microprocessor use more standby power then the battery life will be shorter. So it is important to reduce leakage current. To reduce leakage current gate-oxide thickness is an important factor.

According to the prediction of Moore, transistors are now scaling rapidly and number of transistor in a chip is increasing day by day.

CMOS contain NMOS and PMOS. MOSFET are two types. 1) Depletion type MOSFET and 2) Enhancement type MOSFET.

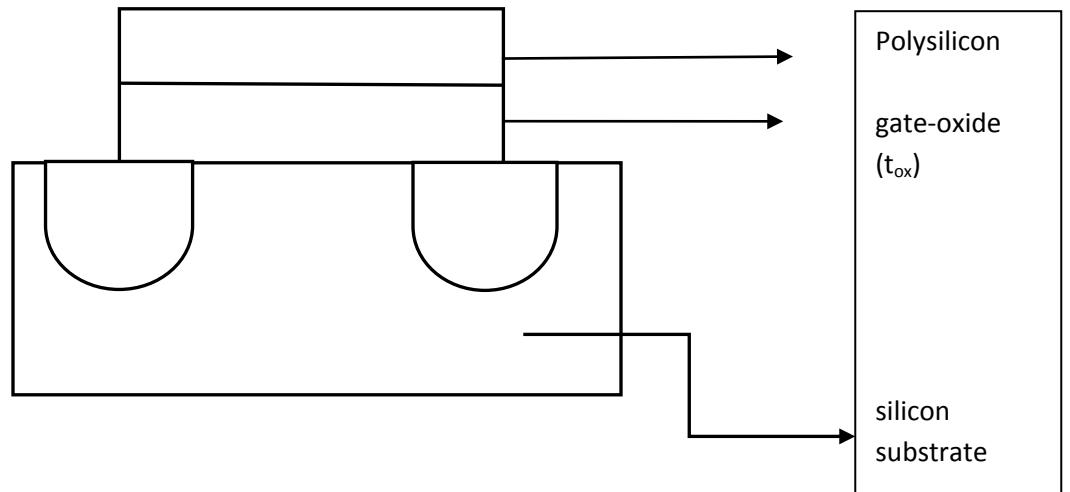


Figure 4.2.4: Structure of a typical MOSFET

There are various types of nanoscale devices. Three basic solid state devices are (1) Quantum Dots (or "artificial atoms"), (2) Resonant Tunneling Devices, and (3) Single-Electron Transistors (SET) [18].

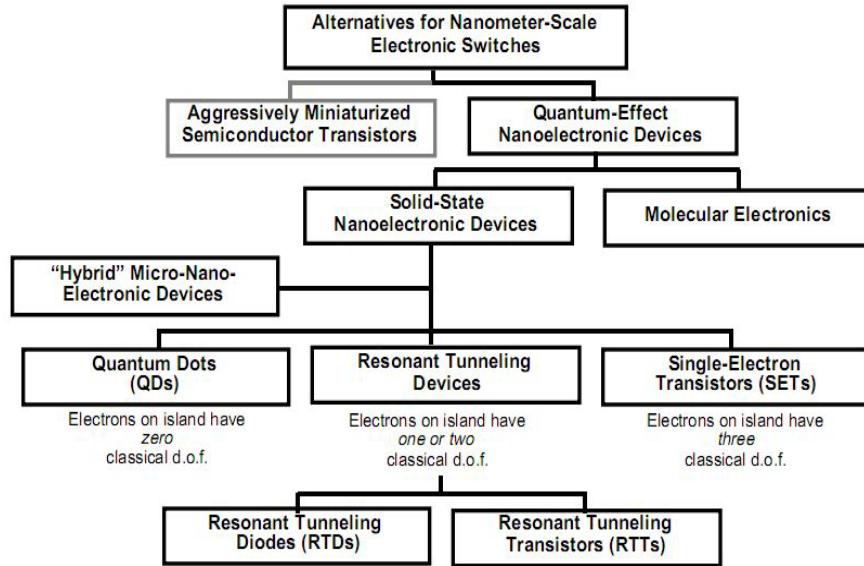


Figure 4.2.5: Overview of Nanoscale devices [18]

CMOS is in the Molecular electronics group. So, molecular electronics will be discussed here.

CMOS is made of some molecular material or semiconductor material. To analysis the effect of scaling the oxide thickness, quantum mechanical effect must be considered. During scaling of CMOS some phenomenon must be carefully observed that control leakage current like dielectric scaling, doping, tunneling of electron and equivalent gate oxide thickness (EOT).

As discussed previously when external electric field is applied to material, dipoles become polarized and capacitance form (figure 4.2.6). This capacitance is proportional to the applied electric field. In this manner dielectric material form.

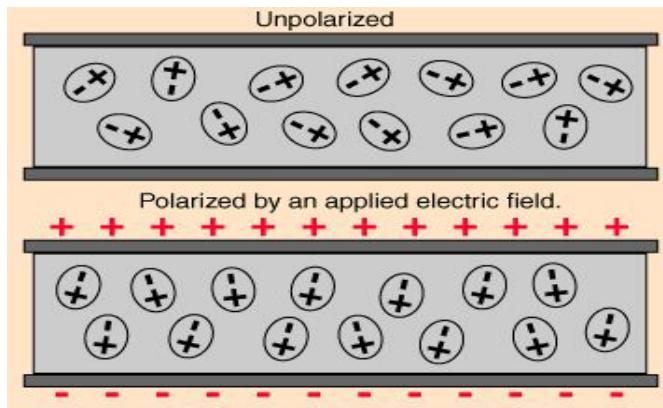


Figure 4.2.6: Formation of dielectric and capacitance. [5]

### 4.3 High-K Dielectric Materials:

In the scaling process gate-oxide thickness plays a very important role. Leakage current is the main obstacle in the scaling process of CMOS in nanoscale regime. If gate-oxide is thick then the electron can easily pass through the barrier as the height of the barrier is small and mobility of electron will be high. On the other hand, if the gate-oxide is thin then height of the barrier will be large. So electrons cannot pass through the barrier easily. Figure 4.3.7 illustrated this phenomenon. If the gate-oxide is thin we can have more control on the device.

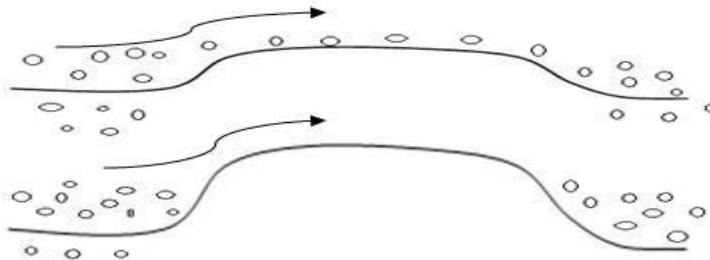


Figure 4.3.7: Barrier height of electron

Usually  $\text{SiO}_2$  was used as gate dielectric material in transistors. As transistor size was not small as today's transistor, so this material could easily be used in those transistors. But today, modern device's circuitry is so complicated that transistor must be scaled to fit into the chip and it must be faster with minimum leakage current to have low power consumption. So alternative dielectric material should be used to minimize leakage current of the circuit.

Gate-oxide thickness has scaled from 100 nm to 1.2 nm over last 35 years. Researchers are now trying to reach the scaling limit of “0.7 nm” which has only two atomic layers [19].

Figure 4.3.8 indicate the physical thickness limit of  $\text{SiO}_2$  and figure 4.3.9 shows the cross section of MOSFET.

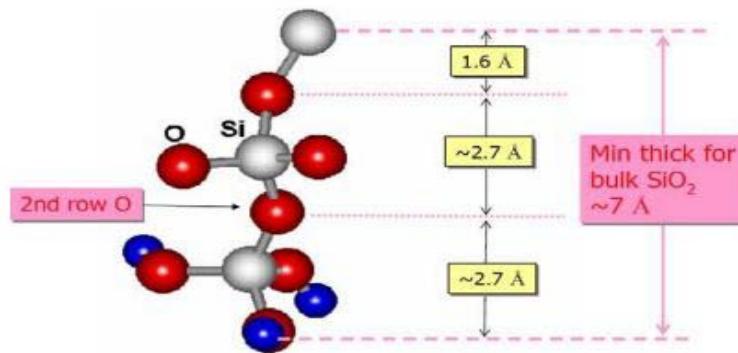


Figure 4.3.8: Bonding structure of  $\text{SiO}_2$  indicating the minimum thickness of the bulk  $\text{SiO}_2$  is about 7  $\text{\AA}$  [19]

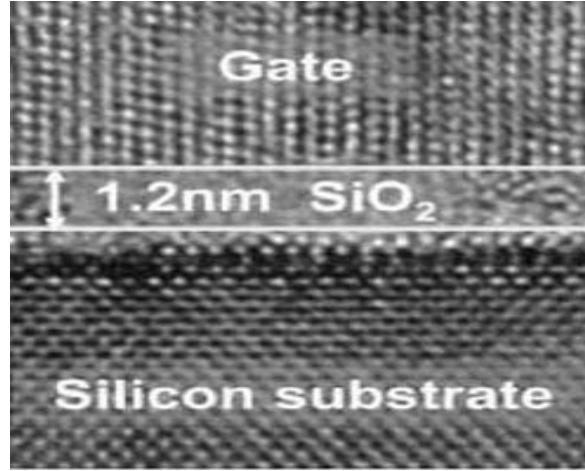


Figure 4.3.9: HRTEM image of MOSFET with SiO<sub>2</sub> as gate-oxide material [22].

Table 4.1: Various parameters of high-K dielectric [26]

MATERIAL PARAMETER	SiO <sub>2</sub>	HfO <sub>2</sub>	Si <sub>3</sub> N <sub>4</sub>	Al <sub>2</sub> O <sub>3</sub>	La <sub>2</sub> O <sub>3</sub>
Barrier Height (eV) ( $\chi$ )	3.1	1.13	2.12	2.80	2.30
Effective Mass of electrons in the dielectric $m/m_0$	0.32	0.17	0.50	0.35	0.26
Relative permittivity	3.9	24	6.9	10	27

#### 4.4 Selection of Dielectric materials:

During selection of dielectric material for gate oxide of CMOS, some criteria should be considered [21].

They are: 1) continuous scaling ability to lower EOT, 2) probability of carrier mobility when using high-k oxides, 3) change in threshold voltage 4) electronic defect in oxides.

Before the micrometer regime, metal electrode was used as gate oxide material. Polysilicon was used as gate dielectric in micrometer regime. Then again metal electrode came back in Intel's 45 nm technology due to ploy depletion problem in polysilicon [31]. During scaling of CMOS, work function decreases along with band gap. Tunneling probability also increases with decreasing physical size. So, in modern microprocessors, dielectric material such as Si<sub>3</sub>N<sub>4</sub>, Al<sub>2</sub>O<sub>3</sub>, Ta<sub>2</sub>O<sub>5</sub>, TiO<sub>2</sub>, SrTiO<sub>3</sub>, ZrO<sub>2</sub>, HfO<sub>2</sub>, HfSiO<sub>4</sub>, La<sub>2</sub>O<sub>3</sub>, Y<sub>2</sub>O<sub>3</sub>, a-LaAlO<sub>3</sub> are using as gate oxides.

Important requirement of selecting oxides are [21]:

1. Dielectric material must be chosen in such a way that it must have enough high  $k$  value so that it can be used for reasonable number of year of scaling.
2. It must be thermodynamically stable.
3. It must have stable kinetic energy.
4. It will act as good insulator.

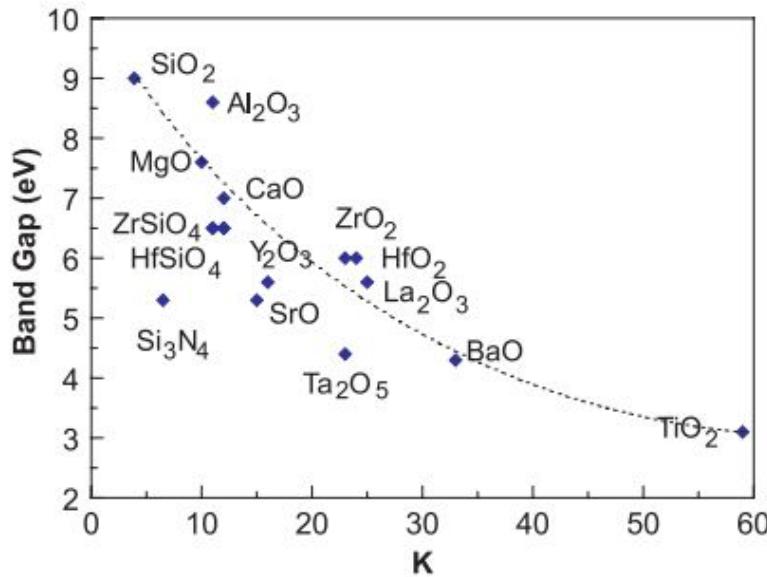


Figure 4.4.10: Dielectric constant of dielectric materials vs. band gap of various gate oxides [21].

In figure 4.4.10and table 4.2 shows comparison among various gate oxides with their advantages and disadvantages. It is seen from the figure 4.10 that value of  $K$  should be above 10 and preferably between 25-30 as higher  $k$  value will occur negative band gap. So, relatively low- $k$  valued oxide should be preferred as gate oxide. Because these material has acceptable range of conduction band energy and band offset energy.

$\text{HfO}_2$  is thermally stable than  $\text{ZrO}_2$ [18]. But stability can be properly shown by ternary phase diagram which is shown in figure 4.4.11

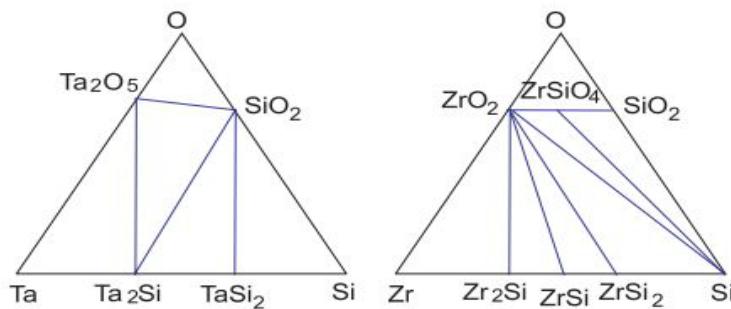


Figure 4.4.11:Ternary phase diagrams of meta stable Ta-Si-O and stable Zr-Si-O [21]

Table 4.2: Comparison of various features and characteristics of existing and potential dielectric materials [19].

Dielectric	Dielectric constant (bulk)	Bandgap (eV)	Conduction band offset (eV)	Merits	Drawbacks
Silicon dioxide ( $\text{SiO}_2$ )	3.9	8.9	3.15	Excellent Si interface, low $Q_{\text{ox}}$ and $D_{\text{it}}$	Low- $\kappa$ , EOT > 0.8 nm
Silicon nitride ( $\text{Si}_3\text{N}_4$ )	7–7.8	5.3	2.1	Good interface and bulk properties, medium $Q_{\text{ox}}$ and $D_{\text{it}}$	Low- $\kappa$ , EOT > 0.5 nm
Aluminum oxide ( $\text{Al}_2\text{O}_3$ )	9–10	8.8		$E_g$ comparable to $\text{SiO}_2$ , amorphous Good thermal stability	Medium $Q_{\text{ox}}$ and $D_{\text{it}}$ , medium $\kappa$
Tantulum pentoxide ( $\text{Ta}_2\text{O}_5$ )	25	4.4	0.36	High- $\kappa$	Unacceptable $\Delta E_C$ , not stable on Si,
Lanthana ( $\text{La}_2\text{O}_3$ )	~27	5.8	2.3	High- $\kappa$ , better thermal stability	Moisture absorption, instable with Si
Gadolinium oxide ( $\text{Gd}_2\text{O}_3$ )	~12	~5	— <sup>a</sup>	Low $D_{\text{it}}$ — <sup>a</sup>	High $Q_{\text{ox}}$ , Crystallization
Yttrium oxide ( $\text{Y}_2\text{O}_3$ )	~15	6	2.3	Large $E_g$	Low crystallization temperature, hight $D_{\text{it}}$ , silicide formation
Hafnia ( $\text{HfO}_2$ )	~20	5.6–5.7	1.3–1.5	Most suitable compared to other candidates	Crystallization, silicate and silicide formation,
Zirconia ( $\text{ZrO}_2$ )	~23	4.7–5.7	0.8–1.4	Similar to hafnia	High $Q_{\text{ox}}$ and $D_{\text{it}}$ , Marginal stable with Si, crystallization, silicide formation
Strontium titanate ( $\text{SrTiO}_3$ )	~300	3.3	–0.1	High- $\kappa$	Unacceptable $E_g$ and $\Delta E_C$ , field fringing effect

Conduction band (CB) offset is the most important factor to select high- $\kappa$  material. CB offset must be higher than 1 eV to reduce leakage current [21]. It is the energy difference between the conduction band to valance band

CB offset can be modeled by metal induced gap states (MIGS). CB offsets is low in group III and IV metal oxides.

## 4.5 Modeling Direct Tunneling Current:

Schottky barrier height (SBH) is the rectifying barrier for electrical conduction across the metal-semiconductor (MS) junction. In metal or semiconductor, SBH is the energy difference in conduction band minimum and the Fermi level. And for a p-type interface, the SBH is the difference between the valence band maximum of the semiconductor and the metal Fermi level [24].

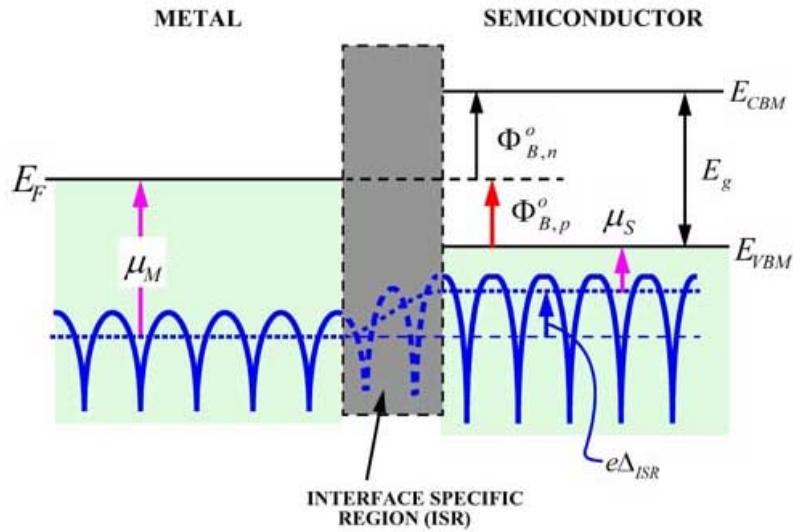


Figure 4.5.12: Distributions of crystal potential,  $-eV(r)$ , (solid blue lines) and energy bands (solid straight black lines) at a metal-semiconductor interface. The average crystal potential is drawn as dotted blue line. Band bending has been ignored in this diagram [25].

Schottky-Mott relationship,

$$\varphi_{b,n}^0 = \varphi_M - \chi_S \quad \text{--- (4.1)}$$

Where,

$\varphi_M$  = work function of the metal.

$\chi$  = electron affinity of the semiconductor

$$\Delta_{gap} = \delta_{gap} Q_{SC} / \varepsilon_0 \quad \text{--- (4.2)}$$

Where,

$$Q_{SC} = (2e\epsilon_S N_D V_{bb})$$

$$\Delta_{gap} = \chi_S + eV_{bb} + eV_N - \varphi_N \quad \text{--- (4.3)}$$

Combining two equations, it is assumed that there is no band bending with no bias applied. When a bias is applied, the band diagram starts to change. In the phenomenon has shown in the figure 4.5.13

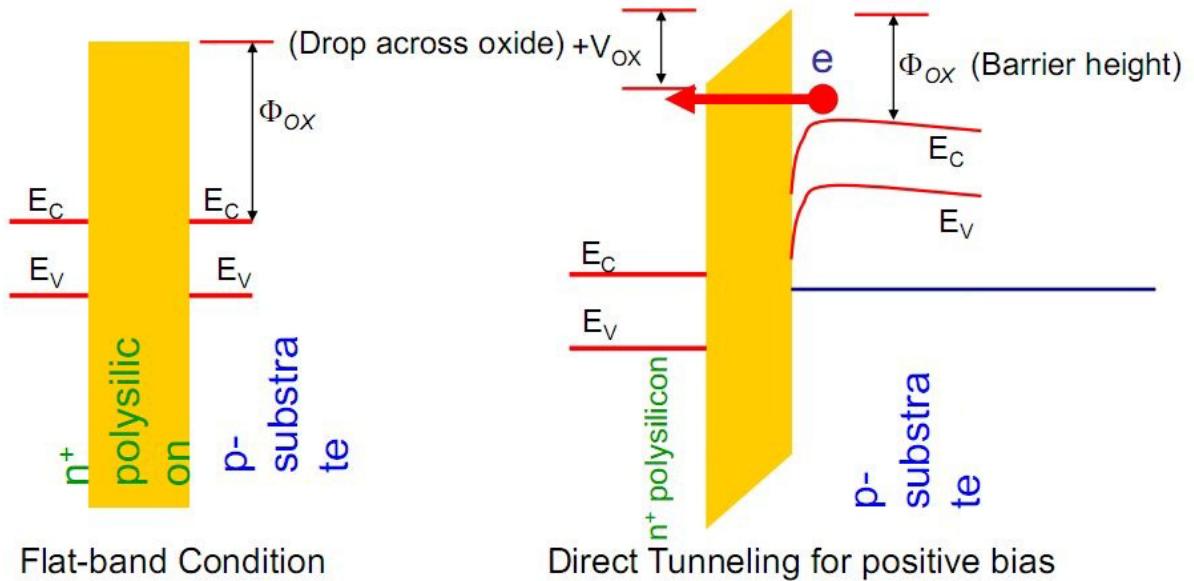


Figure 4.5.13: Band diagram of tunneling leakage current with and without applied bias [25].

This model deals with absence of a gap between the metal and the semiconductor. Due to large band gap of 9eV and large barrier height, SiO<sub>2</sub> acts as an ideal insulator under moderate bias condition [27]. According to quantum mechanical effect, when a bias is applied electron can tunnel through the Si-SiO<sub>2</sub> barrier. Tunneling current density can be characterized by Fower-Nordheim equation [27] which is given by

$$J_{fn} = AE_{ox}^2 \exp\left(\frac{-B}{E_{ox}}\right) \quad \text{--- (4.4)}$$

Where,

A is constant related to the Si-SiO<sub>2</sub> barrier height,

$\varphi_{ox}$  and B is constant related to the effective mass, m<sup>\*</sup>;

Magnitude of the applied bias  $V_{ox} = E_{ox} t_{ox}$

Tunneling current is a function of the oxide thickness  $T_{ox}$ , oxide effective mass  $m_{ox}$  and barrier height  $\varphi_{ox}$  [24]. By taking appropriate effective mass and barrier tunneling current can be predicted which is shown the figure 4.14.

Component that are needed to model direct gate tunneling current are [30]

- 1) electron tunneling from the conduction band (ECB),
- 2) electron tunneling from the valence band (EVB), and
- 3) hole tunneling from the valence band (HVB) .

Tunneling direct tunneling probability through the gate oxide can be modeled by WKB (Wentzel-Kramers-Brillouin) model. Using Schrodinger and Poisson equation this probability can also be found.

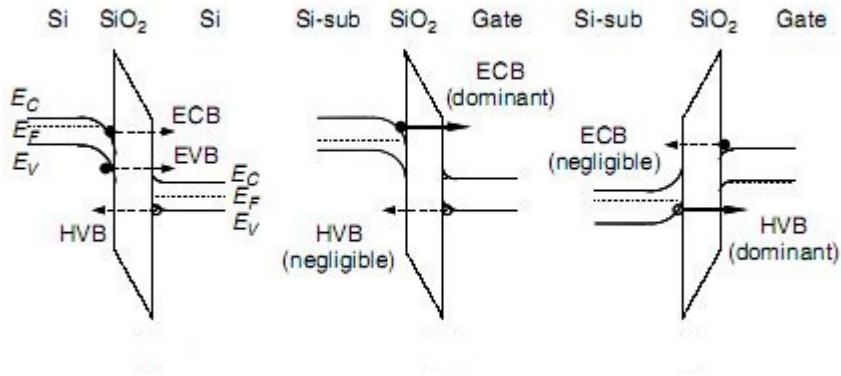


Figure 4.5.14: Direct tunneling current in gate dielectric [20].

According to WKB model, tunneling current density is given by

$$J_n = A \left( \frac{\varphi_{ox}}{V_{ox}} \right) \left( \frac{2\varphi_b}{V_{ox}} - 1 \right) E_{ox}^2 * \exp \left( \frac{-B \left[ 1 - \left( 1 - \frac{V_{ox}}{\varphi_{ox}} \right)^{3/2} \right]}{E_{ox}} \right) \quad (4.5)$$

Where,

$$A = (q^3 / 8\pi h \varphi_{ox}); \\ B = \frac{8\pi \sqrt{2m_{ox}} \varphi_{ox}^{3/2}}{3hq};$$

$$E_{ox} = \left( \frac{V_{ox}}{T_{ox}} \right);$$

$T_{ox}$  = Oxide thickness;

$m_{ox}$  = Effective mass in the oxide;

Substituting values of A and B in equation 4.4, tunneling leakage current density can be found.

$$J_n = (q^3 / 8\pi h \varphi_{ox} \epsilon_{ox}) * C(V_{ox}, T_{ox} V_g, \varphi_{ox}) * \exp \left( \frac{-8\pi \sqrt{2m_{ox}} \varphi_{ox}^{3/2} / 3hq \left[ 1 - \left( 1 - \frac{V_{ox}}{\varphi_{ox}} \right)^{3/2} \right]}{E_{ox}} \right) \quad (4.5)$$

Where exponential term represents WKB approximation of tunneling probability [20].

$$C(V_{ox}, T_{ox} V_g, \varphi_{ox}) = \exp \left[ \frac{20}{\varphi_{ox}} \left( \frac{|V_{ox}| - \varphi_b}{\varphi_{ox}} + 1 \right)^\alpha \left( 1 - \frac{|V_{ox}|}{\varphi_{ox}} \right) \right] \left( \frac{V_g}{t_{ox}} \right) N \quad (4.6)$$

$\alpha$  = fitting parameter depending on the tunneling process;

$\varphi_{ox_0}$  = Si/SiO<sub>2</sub> barrier height (3.1 eV for electron and 4.5 eV for hole)  
 $\varphi_{ox}$  = actual barrier height (3.1 eV for ECB, 4.2 eV for HVB [30] with silicon electrode )

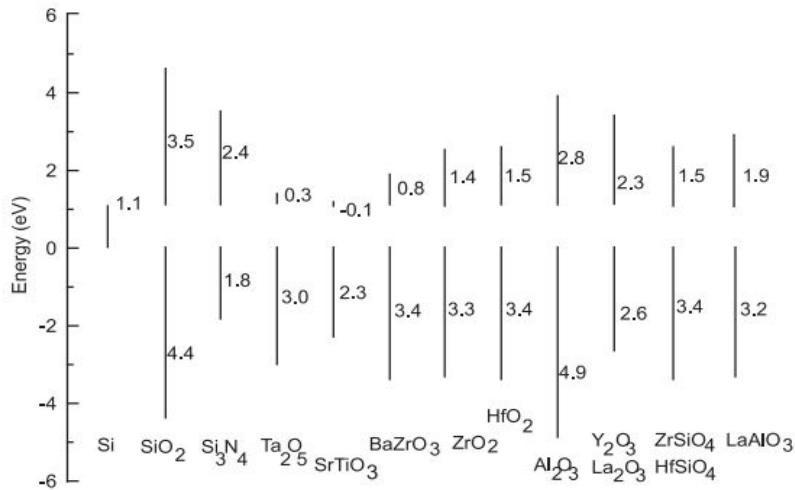


Figure 4.15: Barrier height energy of dielectric materials [21]

Curvature of the figure 4.5.14 increases as  $\alpha$  decreases. N in the equation (4.6) is related to the density of tunneling carriers. For inversion and accumulation region:

$$N = \frac{\varepsilon_{ox}}{t_{ox}} \left\{ s \ln \left[ 1 + \exp \left( \frac{V_{ge} - V_{th}}{s} \right) \right] + V_t \ln \left[ 1 + \exp \left( - \frac{V_g - V_{fb}}{V_t} \right) \right] \right\} \quad (4.7)$$

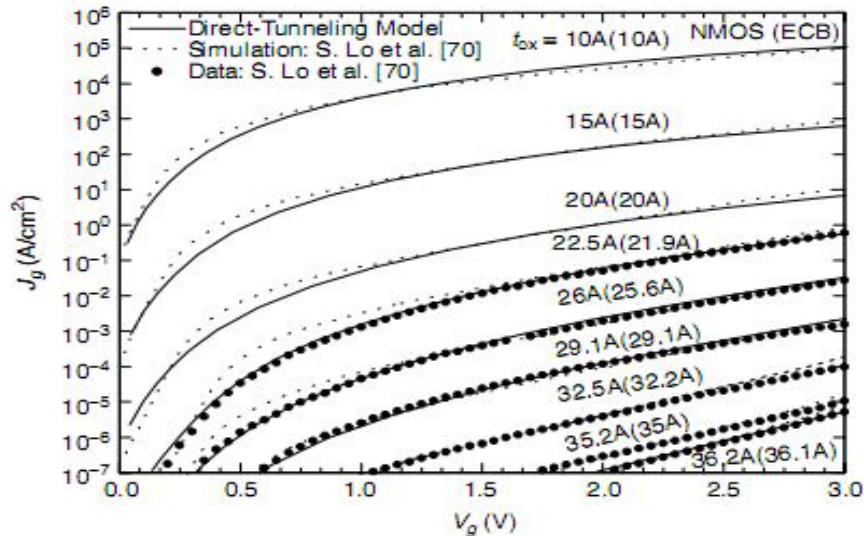


Figure 4.5.16: Direct Leakage current model of Lee and Hu with gate oxide thickness ranging from 3.6 nm to 1nm [20].

In the figure 4.5.16 solid lines are results by numerical simulation, dashed lines from closed-form model and scattered symbol are the data measured from real device. Through Jet vapor-deposition (JVD) this model is best fitted with silicon nitride [20]. Table 4.3 shows the parameter

of direct leakage current in different region.

Table 4.3: Direct Tunneling Leakage Current parameter for SiO<sub>2</sub> and JVD Silicon Gate Dielectric [20]

	SiO <sub>2</sub>		JVD Silicon Nitride	
	ECB	HVB	ECB	HVB
$\phi_b$	3.10	4.50	2.10	1.90
$\phi_{bo}$	3.10	4.50	2.10	1.90
$m_{ox}$	0.40	0.32	0.50	0.41
$\alpha$	0.6	0.4	0.4	1.0

Due to existence of electric field in the gate-oxide, the electrostatic boundary condition results band bending in the polysilicon gate near the gate oxide. Depletion region form capacitance in the channel. The equivalent capacitance is given by

$$\frac{1}{C_g} = \frac{1}{C_{gd}} + \frac{1}{C_{ox}} + \frac{1}{C_{ac}} = \frac{X_{gd}}{\epsilon_0 \epsilon_{si}} + \frac{t_{ox}}{\epsilon_0 \epsilon_{ox}} + \frac{X_{ac}}{\epsilon_0 \epsilon_{si}} \quad (4.8)$$

Where,

$$\epsilon_0 = 8.85 \times 10^{-12} \text{ F/m}$$

$$\epsilon_{si} = 11.7 \text{ (relative dielectric constant of silicon)}$$

Equivalent gate oxide (EOT) can be defined as a number used to compare performance of high-k dielectric MOS gates with performance of SiO<sub>2</sub> based MOS gates; shows thickness of SiO<sub>2</sub> gate oxide needed to obtain the same gate capacitance as the one obtained with thicker than SiO<sub>2</sub> dielectric featuring higher dielectric constant  $\epsilon$ ; e.g. EOT of 1 nm would result from the use a 10 nm thick dielectric featuring  $\epsilon=39$  ( $k$  of SiO<sub>2</sub> is 3.9) and it is given by

$$EOT = t_{ox} \frac{\epsilon_{SiO_2}}{\epsilon_{ox}} \quad (4.8)$$

$t_{ox}$  is equivalent to SiO<sub>2</sub> film with a thickness given by the EOT value. By measuring capacitance EOT can be determined.

Effective gate capacitance (CET) is given by

$$CET = \frac{\epsilon_0 \epsilon_{SiO_2}}{C_g} \approx EOT + \frac{X_{gd} + X_{ac}}{3} \quad (4.9)$$

Where,

$$X_{ac} = 6.20 \times 10^{-5} \left( \frac{V_g + V_{th}}{t_{ox}} \right)^{-0.4}$$

Channel capacitance increases because of displacement of two dimensional electron gases [21]. It is difficult to remove this capacitance.

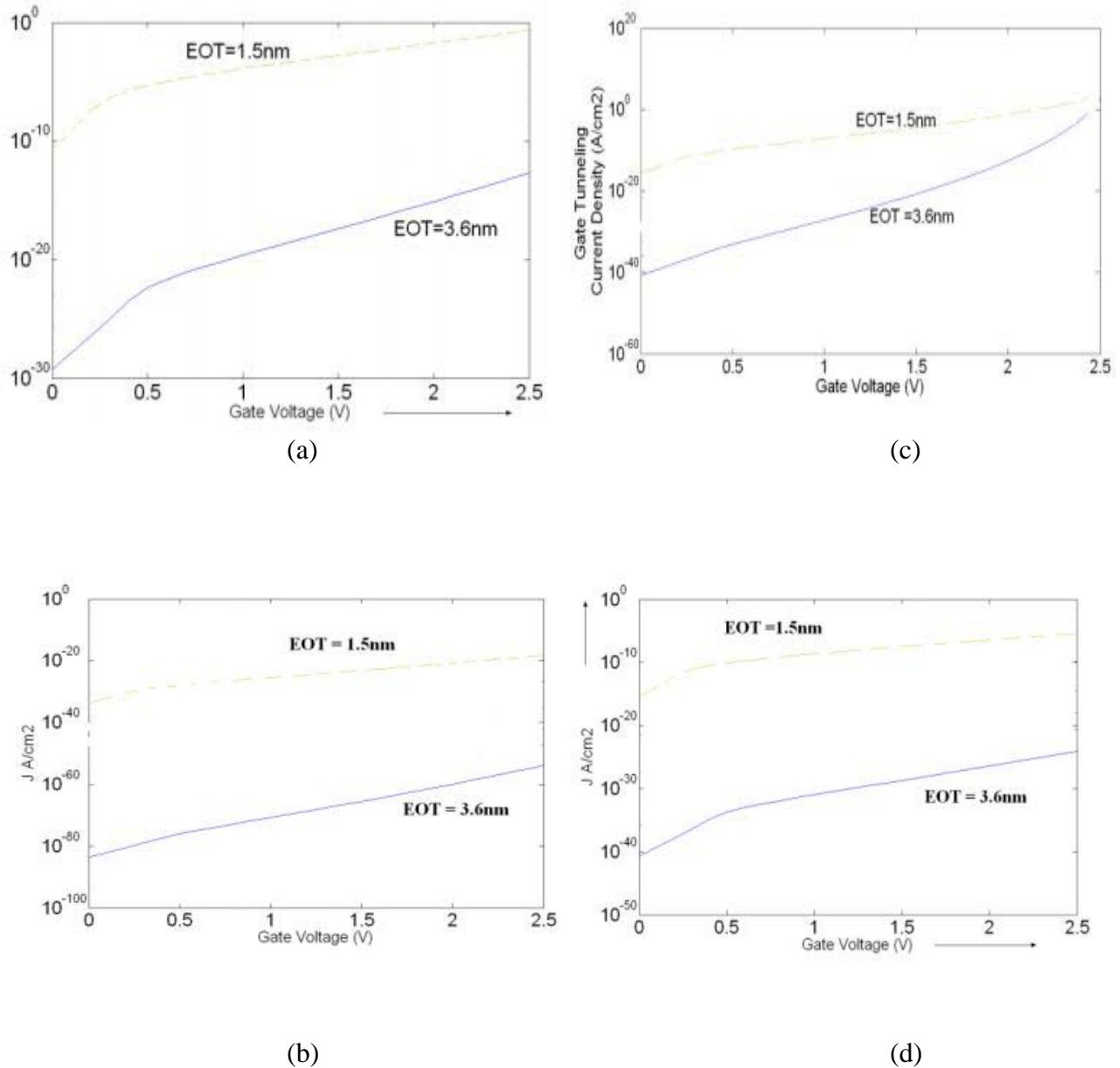


Figure 4.5.17: Gate electron tunneling current density at various gate to source voltage in case of different dielectric materials (a)  $\text{Si}_3\text{N}_4$ , (b)  $\text{Al}_2\text{O}_3$ , (c)  $\text{HfO}_2$ , (d)  $\text{La}_2\text{O}_3$  [26].

From the figure 4.5.17, it is seen that High-K gate dielectric has better tunneling current leakage handling capability compared to  $\text{SiO}_2$ .

Equation of gate capacitance is given by

$$C = \epsilon_0 K A / t \text{-----(4.10)}$$

Where,

$\epsilon_0$  = Permittivity of free space;

$K$  = relative permittivity;

$A$  = Area;

$t$  = Thickness of  $\text{SiO}_2$ .

From the equation 4.10, it is seen that as thickness or distance increases capacitance decreases. So, tunneling current decreases with an increase in distance [21].

Solution of preventing direct tunneling is to replace  $\text{SiO}_2$  with high-k dielectric material. Because high-k dielectric material has the ability to create strong barrier to prevent direct tunneling leakage current which is shown in figure 4.15.

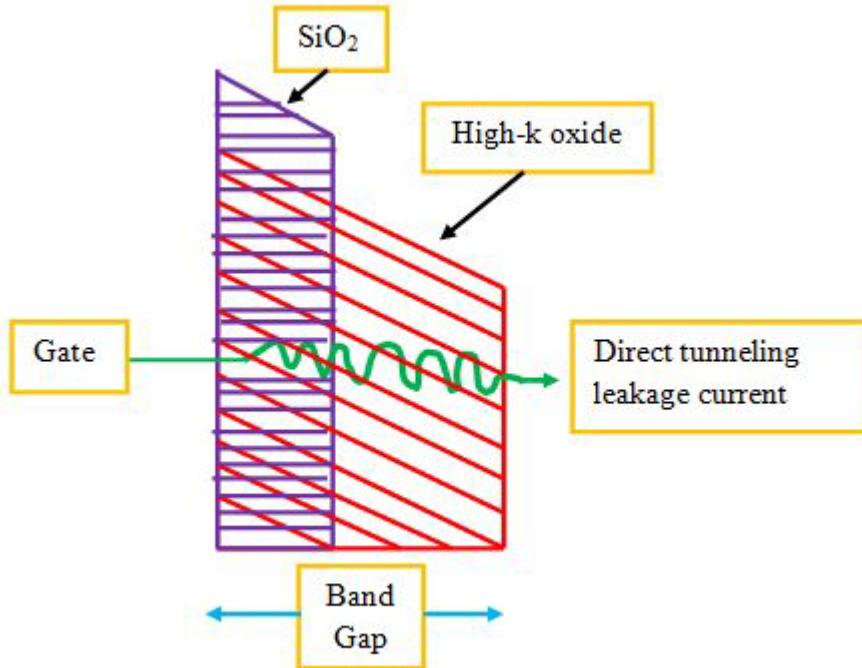


Figure 4.5.18: Comparison of  $\text{SiO}_2$  vs. other high-k dielectric material to prevent direct tunneling

#### 4.6 Dielectric Breakdown:

If local electric field is larger than applied electric field, then dielectric breakdown occur. This phenomenon occurs due to polarization effect which is directly proportional to the dielectric constant. As dielectric constant increases, dielectric break down also increases. Dielectric breakdown can be modeled by MacPherson et al [19] which is given by

$$E_{BD} = \frac{\Delta H_0}{p_0(2+k)/3} \text{----- (4.10)}$$

Where,

$\Delta H_0$  = Activation energy required for metal-ion displacement;

$p_0$  = Molecular dipole-moment component opposite to the local field which is governed by the valence state, number of active dipole and bonding component.

It is further discussed in chapter 5 section 5.6.

#### **4.7 Summary:**

In the scaling process, gate-oxide thickness is a very important factor that affects the performance of CMOS in nanoscale regime. In this chapter, different dielectric high-K material and their properties have been discussed to compare better gate dielectric. Due to scaling of CMOS, probability of high leakage current increases. Compared to  $\text{SiO}_2$ , other dielectric material can reduce the tunneling probability through the gate-oxide thickness. According to the journal of Amit Chaudhry and Jatindra Nath Roy, leakage current is the result of quantum mechanical effect. To reduce this problem, mathematical analysis has been done to model direct tunneling leakage current using WKB approximation. Different parameters have been analyzed in figures to show the amount of leakage in different gate oxide-thickness. The most important matter that have to consider during choosing high-k dielectric material is band gap. From the graph, it is seen that high-k dielectric should not be chosen with a value higher than 30. Otherwise, it will go to negative band gap region. So there must be a tradeoff between the scaling ranges of gate-oxide thickness and selection of high-k dielectric material.

## Chapter 5

# Channel Length reduction

### 5.1 Introduction:

Since the 1960's the price of one bit of semiconductor memory has dropped 100 million times and the trend is continuing. A dramatic drop has been also observed at costs of logic gates. Over the past several decades the desire to improve device performance has led to aggressive scaling of MOSFETs. The definition of a short channel is when the gate length is on the same magnitude as the depletion region of the drain and source junctions. Channel length reduction results in increased probability of direct tunneling of charge carriers through the gate-oxide. Due to this, the gate leakage current has increased at an alarming rate. It is predicted that gate leakage current increases at a rate of 500x per technology generation, while sub-threshold leakage current increases by 5x per technology generation. This would result in gate leakage becoming the dominant contributor to gate leakage current. Gate leakage is an exponential function of the electric field across the gate oxide, so gate leakage current shows an exponential dependence on the gate to source ( $V_{GS}$ ) bias. At high gate bias gate leakage current decreases with increasing drain to source bias.

Gate length is the physical length and can be accurately measured with a scanning electron microscope (SEM). Channel length in a CMOS cannot be determined precisely due to the lateral diffusion of the source and drain junctions.

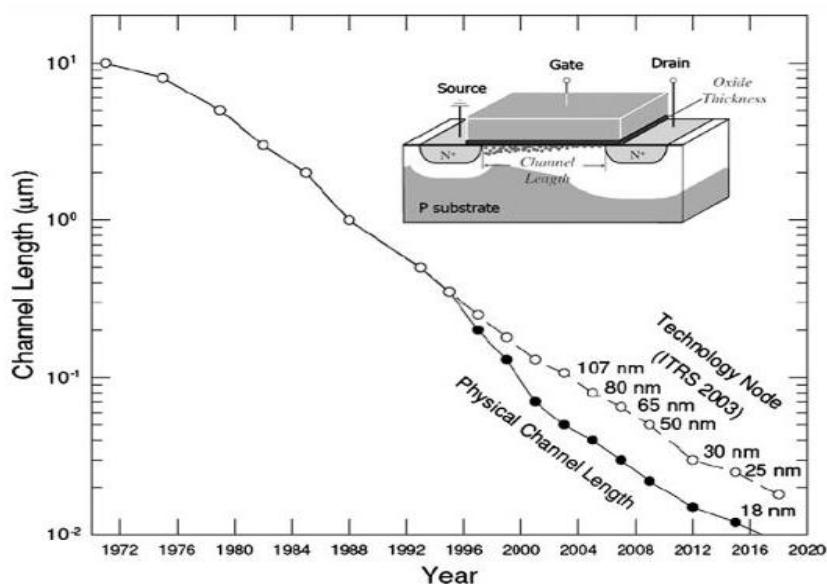


Figure 5.1.1: the trend in MOS channel length scaling and the prediction made in the 2003 International Technology Roadmap for Semiconductor (ITRS 2003). Semiconductor companies often develop the downscaled products ahead of the roadmap in recent years and the physical channel length in these products is expected to be shorter.

The inset illustrates a conceptual MOS transistor structure. The channel length and the gate oxide thickness are two of the major scaling parameters. CMOS technology dimensions are constantly being scaled to a limit where device performance will be assessed against fundamental limits. The most effective to reduce the power is by reducing the supply voltage. But due to power reduction high speed circuit operations suffer from reduced drive current. For this reason drastic reduction of gate oxide thickness as well as gate length is desirable. [19]

## 5.2 Channel length modulation:

Channel length modulation (CLM) is one of several short channel effects in MOSFET. As the length of the inverted channel region is shortened with the drain bias being increased it is called Channel length modulation. Due to this effect there is an increase in current with drain bias, as well as a reduction of output resistance. CLM effect typically increases in small devices with low doped substrates.

To fully explain CLM ‘pinch off’ of channel is introduced. The channel is formed by attraction of carriers to the gate. The current drawn through the channel is nearly a constant independent of drain voltage in saturation mode. Instead of flowing in a channel the carriers flow in a subsurface pattern made possible because the drain and the gate both control the current. As the drain voltage increases it controls the current further towards the source. So the un-inverted region expands towards the source. This shortens the length of the channel region. As resistance is proportional to length the resistance decreases as the channel length decreases. This causes an increase in current with increase in drain bias for a MOSFET operating at saturation.

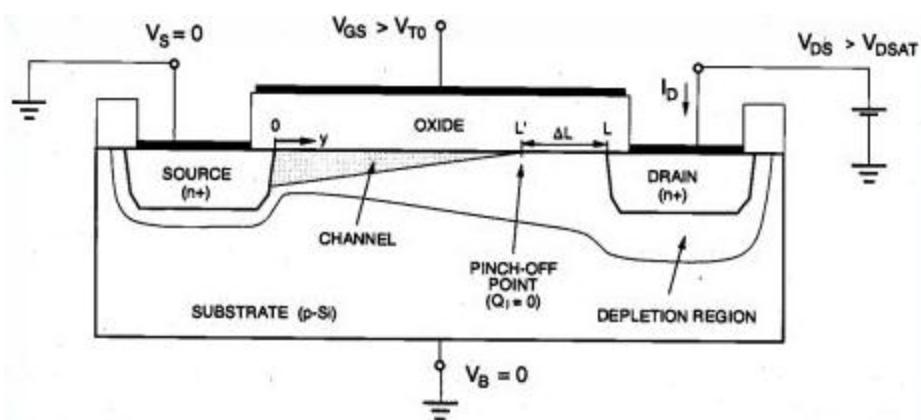


Figure 5.2.1: Channel length modulation in an N-channel MOSFET's operation in saturation mode. [40]

Channel length modulation has more effect if the source-to-drain separation is shorter, if the drain junction is deeper or the thicker the oxide insulator. Channel length modulation in active mode is described using the Shichman-Hodges model given below:

$$I_D = K'_n \frac{W}{L} (V_{GS} - V_{TH})^2 (1 + \lambda V_{DS})$$

Where,  $I_D$ =drain current,  $K'_n$  =technology parameter sometimes called the transconductance coefficient,  $\frac{W}{L}$  = MOSFET width and length,  $V_{GS}$  = gate-to-source voltage,  $V_{TH}$  = Threshold voltage,  $V_{DS}$  = drain-to-source voltage,  $\lambda$ =channel length modulation parameter.

CLM is considered so important because it decides the MOSFET output resistance. CLM is usually taken to be inversely proportional to MOSFET channel length L. The effect of CLM upon the output resistance of MOSFET varies both with the device and the applied bias. An extreme case of channel length modulation is punch through where the channel length reduces to zero.

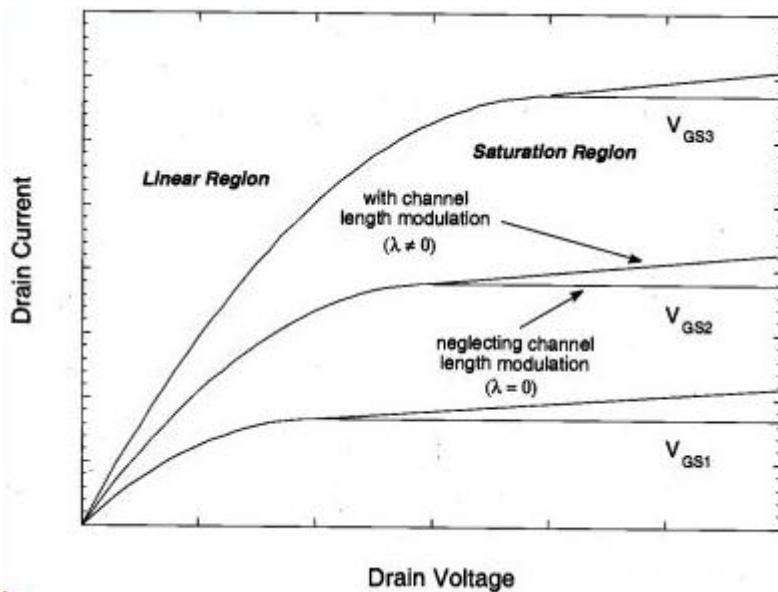


Figure 5.2.2: Current-voltage characteristics of an N-channel MOS transistor, including the channel length modulation effect. [40]

### 5.3 Drain-induced barrier lowering and punch-through

When the gate length is small, the depletion regions surrounding the drain extend to the source, so that the two depletion layer merge, punch-through occurs. Therefore, space charge region near the drain touches the source in a location below the MOS interface surface where the gate bias cannot control the potential, resulting in a leakage current between the source and drain via the space charge region.

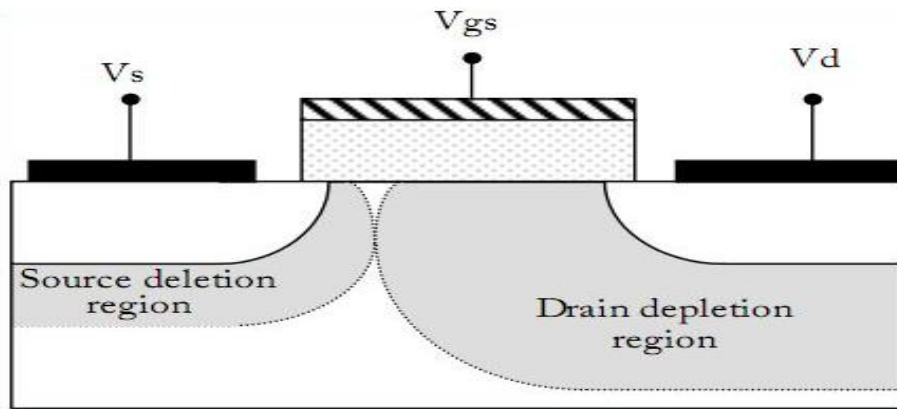


Figure 5.3.1: Schematic diagram for punch-through. [38]

The capability of a MOSFET to switch off this current path and suppress the short channel effects is a major priority in MOSFET design. Punch-through can be minimized with thinner oxides, larger substrate doping, shallower junctions, and obviously with longer channels. The current flow in the channel depends on creating and sustaining an inversion layer on the surface. If the gate bias voltage is not sufficient to invert the surface ( $V_{GS} < V_T$ ), the carriers (electrons) in the channel face a potential barrier that blocks the flow. Increasing the gate voltage reduces this potential barrier and, eventually, allows the flow of carriers under the influence of the channel electric field. In small-geometry MOSFETs, the potential barrier is controlled by both the gate-to-source voltage ( $V_{GS}$ ) and the drain-to-source voltage ( $V_{DS}$ ). If the drain voltage is increased, the potential barrier in the channel decreases, leading to drain-induced barrier lowering (DIBL). The reduction of the potential barrier eventually allows electron flow between the source and the drain, even if the gate-to-source voltage is lower than the threshold voltage. The flow of electron at this condition causes the sub-threshold current to increase. Relation of DIBL and threshold voltage was shown in chapter-2.

In figure 5.3.2 (a) it shows the energy band diagram along the semiconductor/insulator interface of a long channel device at  $V_{gs}=0$ . Figure 5.3.2(b) it shows the case at  $V_{gs}=V_{t-long}$ .

In fig.5.3.2(c) it is seen that the barrier decreases for the case of a short channel device at  $V_{gs}=0$ .

If the channel is too short,  $E_c$  will not reach the same peak value as was seen in figure 5.3.2(a) as a result of this smaller  $V_{gs}$  is needed. Again figure 5.3.2(d) shows that the barrier between conduction and balance band decreases significantly when  $V_{gs}=V_{t-short}$ .

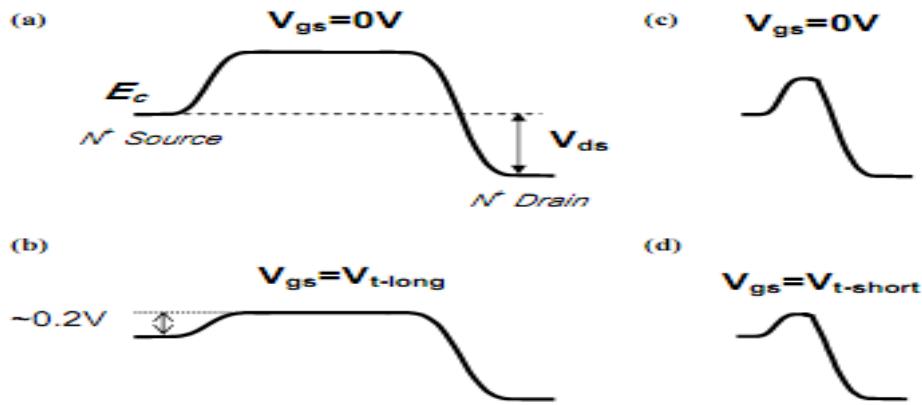


Figure 5.3.2(a)-(d): energy band diagram from source to drain when  $V_{gs}=0V$  and  $V_{gs}=V_t$ . (a)-(b) long channel; (c)-(d) short channel.

## 5.4 Surface Scattering

In MOSFET operation, carriers are confined in a small thickness near the silicon-oxide interface, or equivalently, at the surface of the silicon substrate on which the oxide is grown. Surface scattering is the collisions suffered by the electrons that are accelerated toward the interface by electric field in the x-axis direction. As the channel length becomes smaller due to the lateral extension of the depletion layer into the channel region, the longitudinal electric field component  $E_y$  increases, and the surface mobility becomes field-dependent. Since the carrier transport in a MOSFET is confined within the narrow inversion layer, the surface scattering causes reduction of the mobility and the electrons move with great difficulty parallel to the interface, so that the average surface mobility, even for small values of  $E_y$ , is about half as much as that of the bulk mobility.

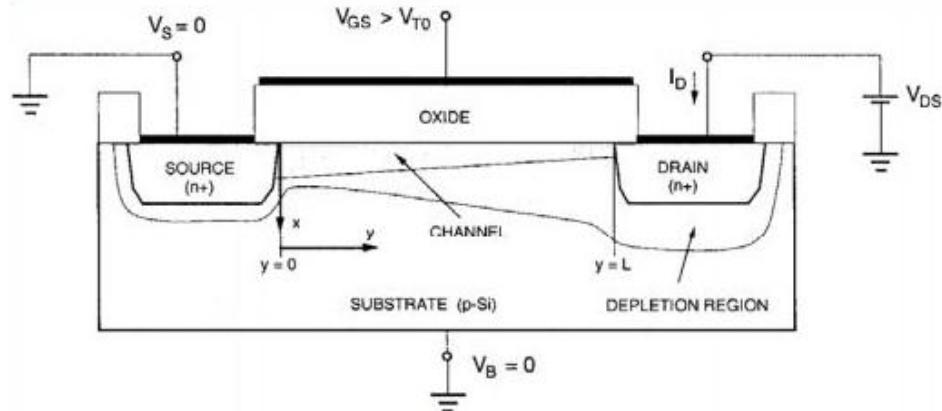


Figure 5.4.1: Cross sectional view of an n-channel transistor, operating in linear region. [40]

## 5.5 Velocity Saturation

The performance short-channel devices are also affected by velocity saturation, which reduces the transconductance in the saturation mode. At low  $E_y$ , the electron drift velocity in the channel varies linearly with the electric field intensity. However, as longitudinal electric field component  $E_y$  increases the drift velocity tends to increase more slowly, and the drift velocity approaches a saturated value. Note that the drain current is limited by velocity saturation instead of pinch-off. This occurs in short-channel devices when the dimensions are scaled without lowering the bias voltages. Therefore, it can be concluded that this velocity is a function of the field itself.

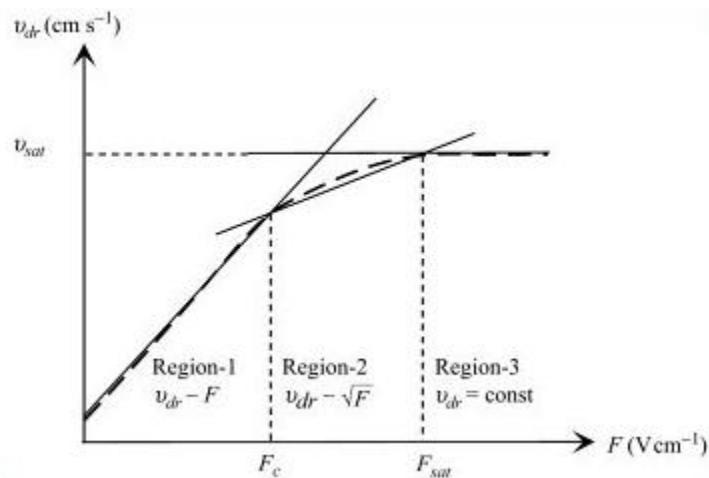


Figure 5.5.1: Drift velocity-electric field characteristics in silicon, where  $F_c$ =critical field,  $F_{sat}$ = saturation field. [39]

Figure 5.5.1 shows relation between Drift velocity ( $V_{dr}$ ) and electric field ( $F$ ), the slope of the  $v_{dr}-F$  curve represents the mobility. The curve indicates that mobility decreases with

increasing electric field, transitioning from a linear dependence of velocity on electric field (Region-1) to a limiting field-independent saturation velocity (region-3) is shown.

At region-1 which low-field region, the drift velocity is linearly related to the applied electric field. Region-2 which is intermediate field region, the hot carries effect increases. Finally at region-2 which is high field region, the carriers gain enough energy which is dissipated by scattering.

## 5.6 Hot Carrier effect

As the critical dimension of CMOS devices are scaled down to nanometer scale, the horizontal and vertical electric field in the channel region increases significantly. At this high field some of the electrons and hole acquire enough energy to cause impact ionization and are referred to as hot electron or hot carriers. This hot carrier may be injected into the gate oxide and cause permanent changes in the oxide-interface charge distribution. If hot electrons with energy greater than 3.2 eV overcome the potential barrier between Si and SiO<sub>2</sub>, they may be trapped in the oxide and give rise to a gate current. Since the energy barrier between the Si and SiO<sub>2</sub> is lower for electron injection (3.2 eV) than for hole injection (4.7 eV), hot carrier degradation is more severe for n-channel MOSFETs than for p-channel MOSFETs.

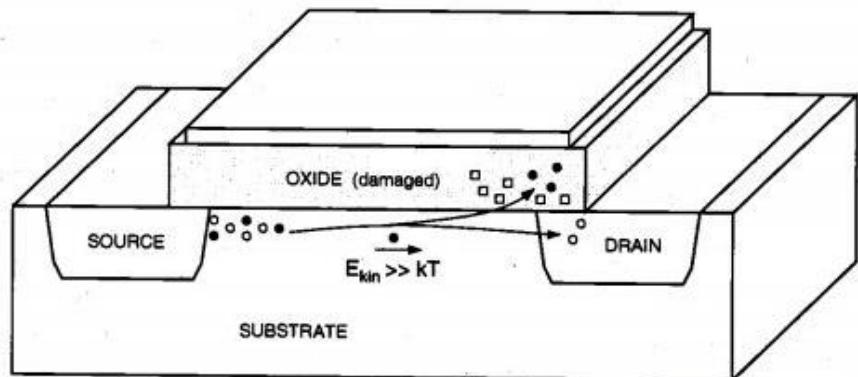


Figure 5.6.1: Hot carrier injection into the gate-oxide and resulting oxide damage. [40]

As shown in the above figure, the flow of electron from source to drain causes the channel hot-electron effect. This effect increases as the voltage between drain and source increases. The hot-carrier injection affects the transistor characteristics by causing degradation in the transconductance and shift in the threshold voltage.

## 5.7 Solution:

### Retrograde well engineering:

Conventional channels formed by implanting dopants into the substrate and diffusing them (at high temperature) to a certain depth, can be quite susceptible to short channel effects such as punch through, mobility degradation and latch up. One design that minimizes these undesirable effects is known as a retrograde profile. Such a design is achieved by using high energy ion implantation to place dopants at a desired substrate depth and then annealing at a low temperature to activate the implants. A key feature of retrograde structures is the use of slow diffusing dopants such as arsenic or antimony for p-channel devices and indium for n-channel devices. [41]

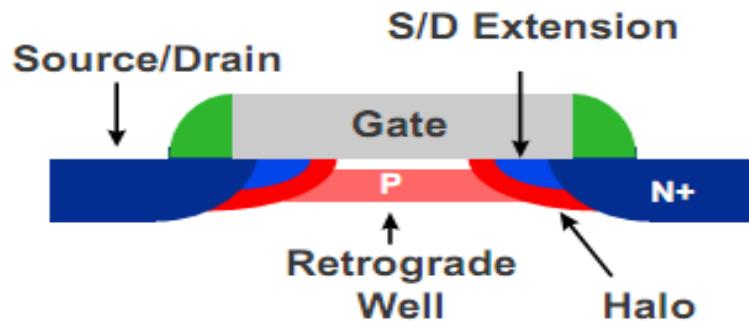


Figure 5.7.1: schematic representation of different aspects of well engineering

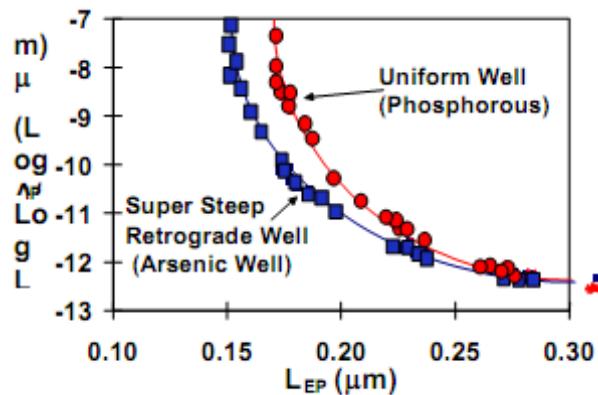


Figure 5.7.2: leakage current as a function of channel length and uniform well transistors with the same threshold voltage.

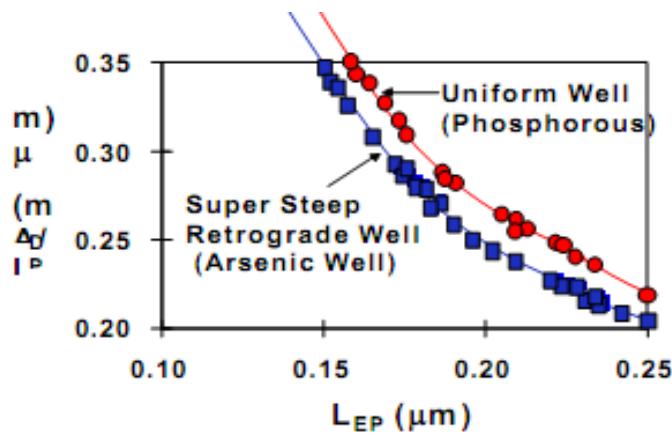


Figure 5.7.3: saturated drive current ( $I_{dsat}$ ) versus channel length and uniform well transistors.

In Figure 5.7.2 it shows improved source-to-drain for retrograde well device for sub-0.25um channel lengths meaning improved short channel effects. But in Figure 5.7.3 shows a decrease in saturated drive current for the same retrograde well device. In Figure 5.7.4 it shows families of curves for drain current versus drain voltage for retrograde well and uniform well devices. The devices have channel length of 0.15um. For devices with the same channel length, the linear drive current is approximately equal, indicating no charge in mobility for retrograde wells. However the current does saturate at a lower drain bias. [41]

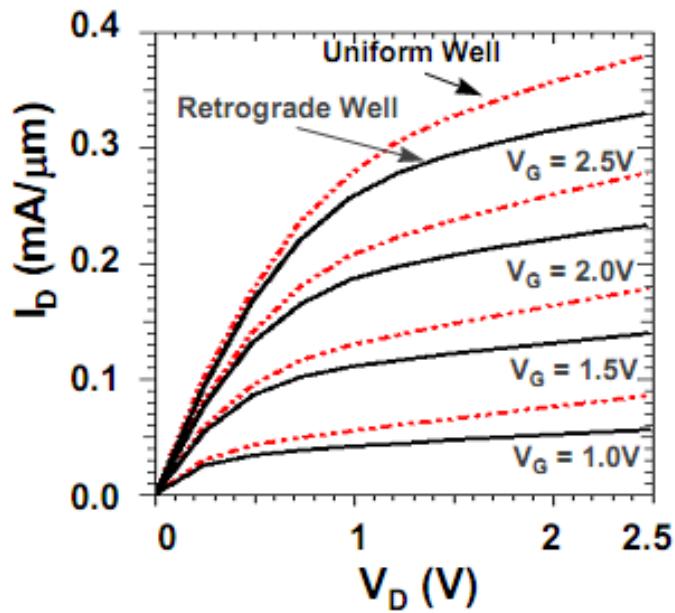


Figure 5.7.4:  $I_d$  vs.  $V_d$  characteristics for retrograde well and uniform well devices as a function of gate voltage.

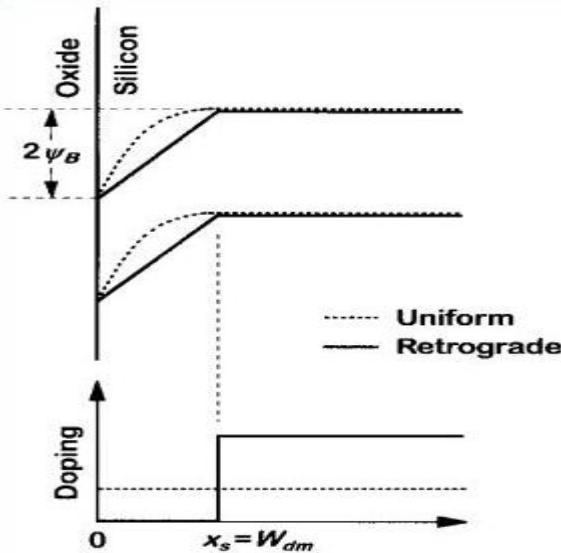


Figure 5.7.5: Band diagram at threshold condition of a uniformly doped and an extreme retrograde-doped channel. [1]

As shown in figure 5.7.5 the band bending at threshold condition of an extreme retrograde-doped channel is more improved than that of uniform doped channel. For the same gate depletion width, the depletion charge of retrograde channel is one-half of that of a uniformly doped channel. Hence this shows that the threshold voltage reduces with improve mobility and it is decoupled from the gate-controlled depletion width. [1]

### Halo Implant:

One way of reducing short channel effects by extending  $V_t$  roll-off is through halo implants. Halo implants increase the doping near the source and drain implant.

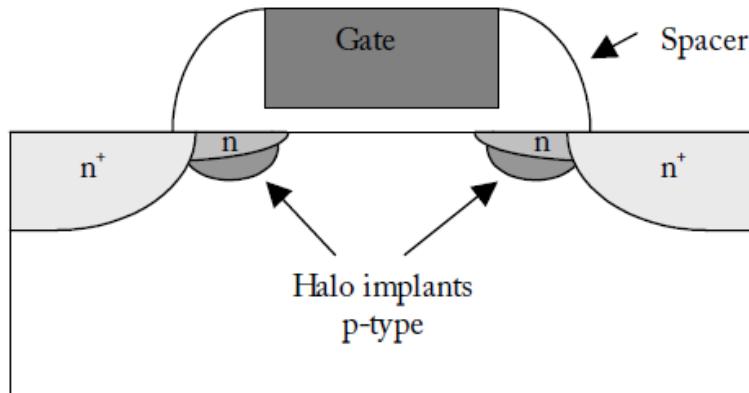


Figure 5.7.6: Halo implant for n-channel MOSFET

The implants near the source or drain can be symmetrical or asymmetrical. Halo implants do help reduce short channel effects such as DIBL, punch through, and  $V_t$  roll-off. But also halo reduces drive current in the transistor. The tradeoff between drive current and the reduction of short channel effects must be considered to maximize performance of the transistor. The implants add more of a barrier between the source drain junctions with the channel. [41]

For conventional MOSFETs the threshold voltage must not roll off more than 100 mV. According to this requirement, transistors could not be fabricated below 0.3 um without the use of halo implants. With halo implants, transistors can be scaled well below 0.2 um.

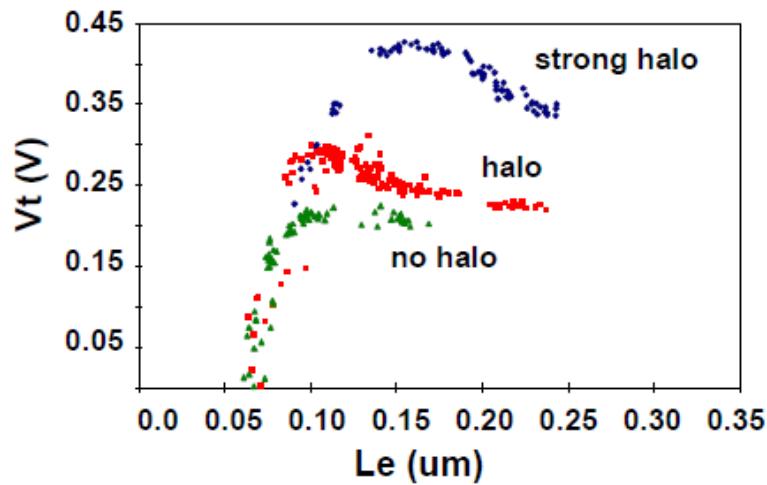


Figure 5.7.7: threshold voltage as a function of channel length for a no halo, halo and strong halo device

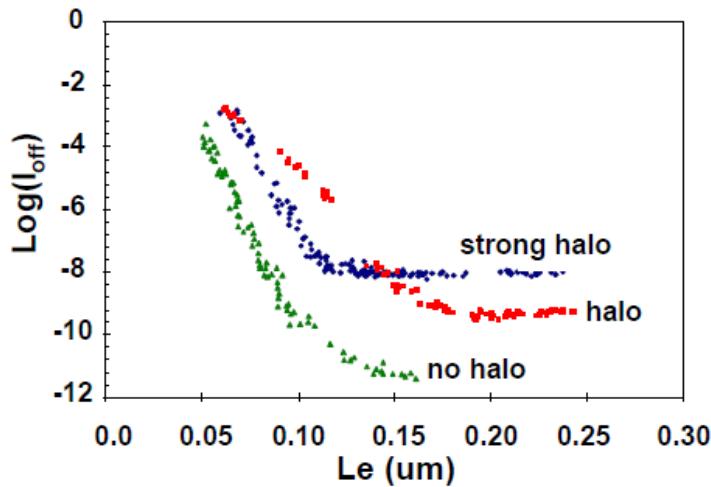


Figure 5.7.8: Off state leakage current as a function of channel length for a no halo, halo, and strong halo devices.

As shown in Figure 5.7.7 the increase in the threshold voltage as the halo implant is increased. Figure 5.7.8 after that shows the decrease in  $I_{off}$  for devices with halo implants. Also, Figure 5.7.7 shows  $V_t$  roll-up which is more dominant for the strong halo. [41]

## 5.8 Summary:

Reduced channel length plays a vital role when the devices are scaled down to the nanometer scale. Short channel effects in MOSFET are discussed in this chapter and solution to such effects are introduced at the end section of this chapter. As the channel length gets smaller the Channel length modulation effect increases. As discussed in section 5.2, channel length modulation in active mode was described by Shichman-Hodges model. The drain-induced barrier lowering is another key effect of short channel length. Due to DIBL the threshold voltage roll-off increases resulting in the increase of leakage current. DIBL effect relation to threshold voltage was covered more in chapter 2. The punch-through effect is another major cause of leakage current which increases exponentially with recessed channel length. For a constant field scaling, velocity saturation and hot-carrier effect degrades the performance of a CMOS device. Increased hot-carrier effect damages the gate oxide resulting in increased gate leakage current. The short channel effects cause leakage current which eventually increases the temperature of the device. Channel engineering is the key to overcome such short channel effects. Section 5.7 discusses some of the engineering ways and their influence on other device parameter. Retrograde well engineering is one of such channel engineering which suppresses the short channel effects effectively. Figure 5.7.5 shows that the use of retrograde well decreases the threshold voltage resulting in increased mobility. Figure 5.7.3 shows that leakage current control is better with the use of retrograde well. Channel length modulation can be reduced mainly by increasing the doping density as gate length is reduced. Most of the short channel effect such as DIBL, punch-through and  $V_T$  roll-off are decreased by using a Halo implant as shown in figure 5.7.6. However, the effects of short channel cannot be suppressed by channel engineering alone so other physical parameter such as the gate-oxide thickness must be reduced at a similar proportion with the channel length. At nanometer scale direct tunneling leakage current increases due to decrease in gate oxide thickness. The gate oxide needs to be replaced with High-k gate material and the choice of proper gate material was discussed in chapter 4. The temperature of a device mainly increases due to the increase in leakage current. Therefore, reduced leakage current keeps the device at optimum temperature. Finally, reducing the effects of short channel improves the overall performance of a CMOS device.

# **Chapter 6**

## **Discussion and Conclusion**

### **6.1 Discussion**

After more than 40 years of aggressive downsizing of MOSFETs, geometric scaling is reaching fundamental limits. New materials and novel processes will be necessary in order to extend device scaling to the last CMOS technology node of the International Technology Roadmap for Semiconductors (ITRS).

The single-gate planar bulk silicon MOSFET has served the industry very well for the last several decades. With the conventional scaling techniques approaching fundamental physical limits, new CMOS architectures are being investigated. At the same time, new materials are being incorporated into the conventional planar MOSFET to extend it beyond 45nm technology node. Scaled MOSFETs are exhibiting new device behaviors which were previously insignificant or absent.

In this thesis we have analyzed different parameter which affects the performance of CMOS. We have emphasized more on threshold voltage, temperature, gate oxide thickness and channel length. Each of these parameters is affected by each other. So, when one of this parameter is changed, it should be carefully examined how it effect the other parameters.

While analyzing the parameters the biggest problem that we had was the lack of instruments in our lab for the analysis. This type of work requires state of the art instruments which were lacking in our university. So we had to collect information from many reputed journals to complete our work.

Our target was to analyze the effects of scaling on the different parameters of CMOS and correlate the parameters with each other.

If anyone is interested in doing work on CMOS then they can use this thesis paper as a guide line to understand CMOS well and use this information to work on other parameters and architecture of CMOS as well.

Scaling of MOSFET into the nanometer regime decreases the channel length which drastically reduces the performance of the CMOS devices. The threshold voltage roll-off due to short channel length effect was explained in section 2.3 of chapter-2. To keep such effect under control the gate oxide thickness must be reduced nearly in proportion to channel length. However, decreasing the gate oxide also lowers the threshold voltage as explained in section 2.2 of chapter-2.

A smaller value of threshold voltage is desired as the device approaches the nanometer regime but as we approach into the nanometer regime the threshold voltage decreases drastically which results in the increases of leakage current. For better control over threshold voltage at the nanometer regime a model of MOSFET threshold voltage with stack High-k gate dielectric was shown. When the devices are further scaled down such models becomes less effective, therefore new modified structures are needed which are discussed in brief.

Effects of temperature on semiconductor parameters are discussed in chapter 3. Section 3.2 to section 3.5 discusses this parameter. In section 3.5 we see the relation between temperature and threshold voltage. Threshold voltage depends on Fermi level. Fermi level is effected by temperature as shown in section 3.5 and Fermi level depends on energy band gap. Figure 3.2.1 shows how the bandgap of different semiconductor are affected by temperature. Section 3.6 shows the temperature effects on mobility. In section 3.7 how mutual compensation of mobility and threshold voltage results in a zero temperature coefficient, (ZTC), is shown. Section 3.9 and 3.10 shows the Effect of temperature variation on gate tunneling current at different values of Equivalent Oxide Thickness (EOT) and substrate doping (Na). EOT is discussed in detail in chapter 4.

In chapter 4 different dielectric high-K material and their property have been discussed to compare better gate dielectric. Scaling of CMOS increases the probability of high leakage current. Compared to  $\text{SiO}_2$  other dielectric material can reduce the tunneling probability through the gate-oxide thickness. The most important parameter that has to be considered during choosing high-k dielectric material is band gap. There must be a tradeoff between the scaling ranges of gate-oxide thickness and selection of high-k dielectric material.

Reduced channel length plays a vital role when the devices are scaled down to the nanometer regime. Short channel effects in MOSFET and solution to such effects are discussed in chapter 5. In section 5.2, channel length modulation in active mode was described by Shichman-Hodges model. The drain-induced barrier lowering is another key effect of short channel length. Due to DIBL the threshold voltage roll-off increases resulting in the increase of leakage current. DIBL effect relation to threshold voltage was covered more in chapter 2. The punch-through effect is another major cause of leakage current which increases exponentially with recessed channel length. For a constant field scaling, velocity saturation and hot-carrier effect degrades the performance of a CMOS device. Increased hot-carrier effect damages the gate oxide resulting in increased gate leakage current. The short channel effects cause leakage current which eventually increases the temperature of the device. Channel engineering is the key to overcome such short channel effects. Section 5.7 discusses some of the engineering ways and their influence on other device parameter. Figure 5.7.5 shows that the use of retrograde well decreases the threshold voltage resulting in increased mobility. Figure 5.7.3 shows that leakage current control is better with the use of retrograde well.

Most of the short channel effect such as DIBL, punch-through and  $V_T$  roll-off is decreased by using a Halo implant as shown in figure 5.7.6. However, the effects of short channel cannot be suppressed by channel engineering alone so other physical parameter such as the gate-oxide thickness must be reduced at a similar proportion with the channel length. At nanometer scale direct tunneling leakage current increases due to decrease in gate oxide thickness. The gate oxide needs to be replaced with High-k gate material and the choice of proper gate material was discussed in chapter 4. The temperature of a device mainly increases due to the increase in leakage current. Therefore, reduced leakage current keeps the device at optimum temperature.

Finally in this chapter we will show how the different problems of scaling are overcome using different technology and architecture.

## 6.2 Suggestions for future work:

Every possible aspect of scaling CMOS in nanoscale regime is examined. Due to limitation of doing practical research in our university, correlation of different parameters that controls the scaling process is analyzed from different published journals. This research can also be improved by concentrating in multi gate technology. So if any students want to do further research on CMOS technology in nanoscale regime, they can do their research in multi-gate CMOS technology which is mentioned below:

### 6.2.1. Multi-Gate FETs

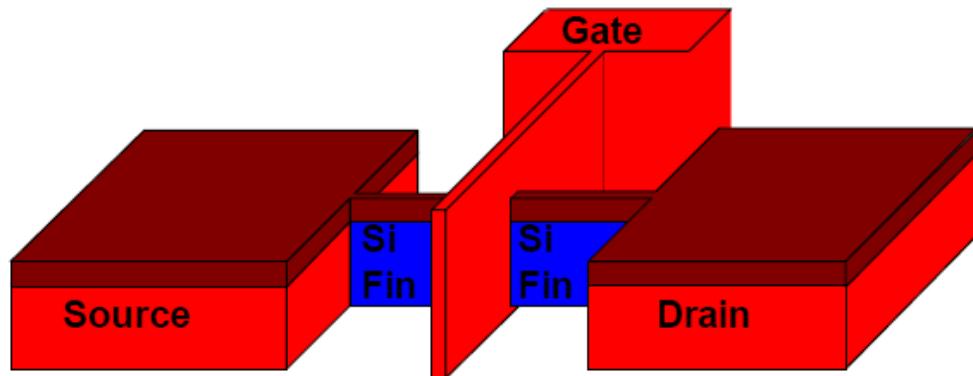


Figure 6.2.1: Multi-Gate FET

Promising alternative to extend CMOS scaling beyond 32nm node is the introduction of a new FET architecture - the Multi-gate (MG) FET [35]. MG-FETs offer stronger electrostatic control of the inversion channel through the use of multiple gates. This reduces the detrimental short channel effects (SCE) and makes the MG-FETs more scalable than the planar bulk CMOS.

The good control of SCE in MG-FETs allows the silicon body to be lightly doped compared to single-gate FETs. The reduced body doping results in lower electric field in the channel which translates to an improvement in carrier mobility, gate leakage currents and device reliability. The use of less channel doping also helps to reduce random dopant fluctuations in MG-FETs and hence decreases the variability in device performance which will be critical for some circuit designs such as SRAM arrays. The combination of light body doping and thin body yields steeper sub-threshold swing and lower junction and body capacitance. Because of these additional benefits, MG-FETs show better logic delay than the planar bulk devices.

The silicon body can be controlled by either two gates or three gates or four gates. The more the number gates the higher the electrostatic control of channel but there is a trade-off with the corresponding process complexity. The existence of several MG-FETs configurations together with the control of inversion channel by multiple gates makes modeling of MG-FETs quite challenging.

### 6.2.2 Tri-gate Transistor:

Tri-gate (3D) CMOS transistor has done revolutionary change in the technological fields of semiconductor industry. It use metal gate electrode replacing polysilicon in Intel 22 nm technology node which has a work function near to midgap [33].In traditional transistors,

electron flows through single gate where in tri-gate transistors electron flows through three gate. It is a fully depleted transistor. It can have multiple fins connected simultaneously. So, delay is reduced and device performance increases. It also reduces leakage current in aggressive way compared to other technology.

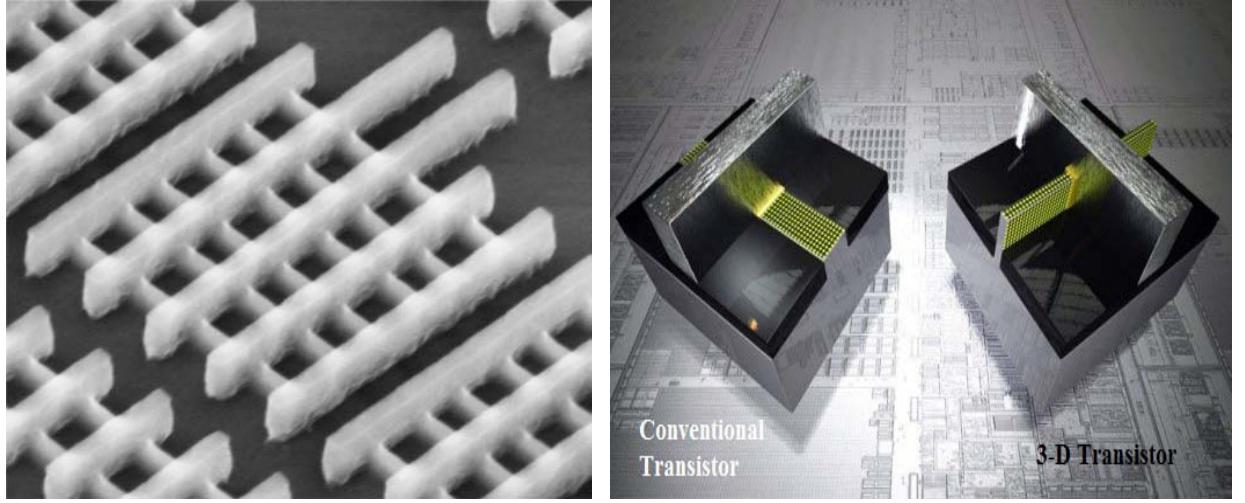


Figure 6.2.2: Tri-gate (3-D) Transistor [34].

### 6.2.3. Four-gate Transistor

Four-gate Transistor has the maximum number of gate present which can be achieved. The body of the transistor is surrounded by top MOS gate (VPG), a bottom substrate gate (VSUB), and two side junction gate (VG1&VG2). The N-type Body has two N+ contacts which act as source and drain. Compared to MOSFET, G4-gate is majority carrier device and the gate length of MOSFET is the width of G4-gate and vice versa. The drain current is modulated by top MOS gate and side junction gate

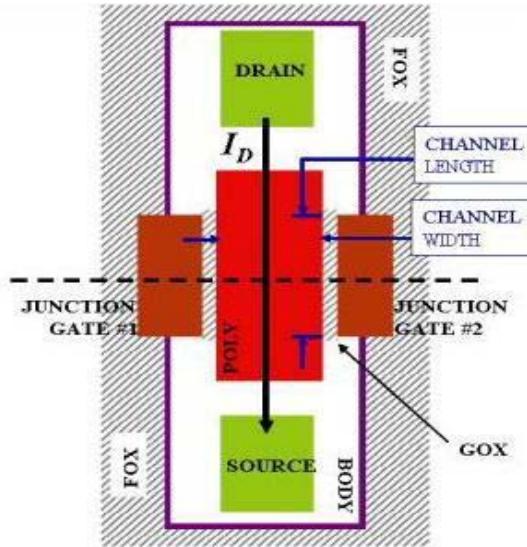


Figure 6.2.3 Top View of Four-gate FET structure [10].

The G4-gate FET uses the characteristic of MOS and JFET to control the current flow. The G4-gate FET can be used for both high and low gate voltages. It can also be used to achieve low noise and RF performance. With use of four-gate transistor the density of digital logic per transistor is expected to increase.

### 6.3 Future techniques:

The scaling of MOSFET seems to reach at the end of its limitation, and the cause of such limitation is often believed to be in the lithography and the availability of sufficiently small wavelength of lights to pattern the minimum pattern size. To maintain Moore's law of scaling, new innovation in MOSFET structure are introduced. This chapter explains some of such innovation techniques which minimized the effects of scaling as described earlier.

#### 6.3.1 Novel Gate-Oxide Materials:

Currently in research level nano-graphene ribbons (GNRs) are using in multi-layer of gate in CMOS through atomic layer deposition. This technique reduces probability of tunneling current.

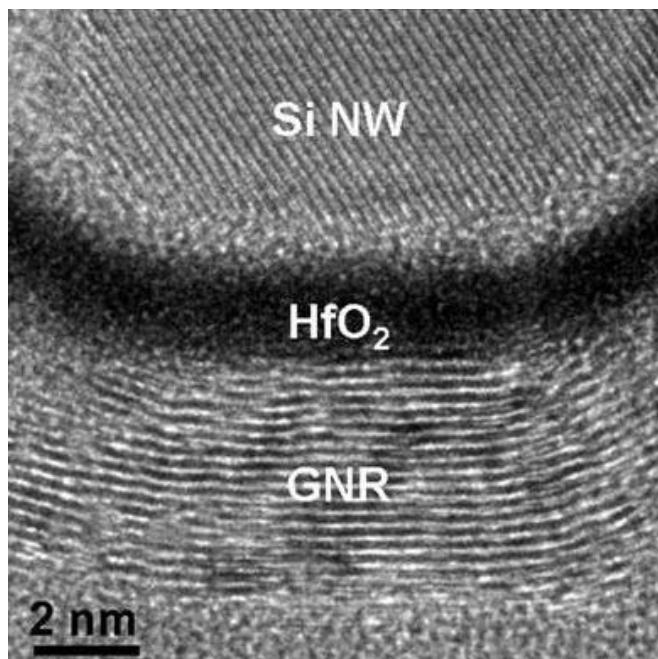


Figure 6.3.1: Multi-layer graphene device [42]

Gallium Nitride (GaN) is also using in CMOS technology as gate-oxide materials which have wide band gap (3.4 eV) [43] and flexible dielectric constant 8.9, 9.5, 10.4, (Static), 5.35 (high frequency) [44].

### 6.3.2 CNTFET Transistor:

Carbon nanotubes (CNTs) are nanometer-diameter carbon cylinders consisting of graphene sheet wrapped up to form tube. CNT has unique electrical and mechanical properties that it can be shaped to act as conductor, semiconductor and insulator depending on its structure. CNTFET is similar to CMOS FET, it has three terminals; source, drain and gate. Nanotube of CNTFET act as a conduction channel of MOSFET, the electron flow takes place inside this nanotube.

CNTFET shows significant improvement in effective PMOS channel mobility and NMOS intrinsic gate delay over Si-MOSFET [36]. CNTFET has high gate capacitance, improved channel transport and the compatibility with high-k gate dielectrics are definite advantage for CNTFETs over MOSFETs. The improved channel velocity for the CNTFET arises from the increased mobility which is also an advantage compared to the mobility degradation caused by short channel effects in MOSFETs.

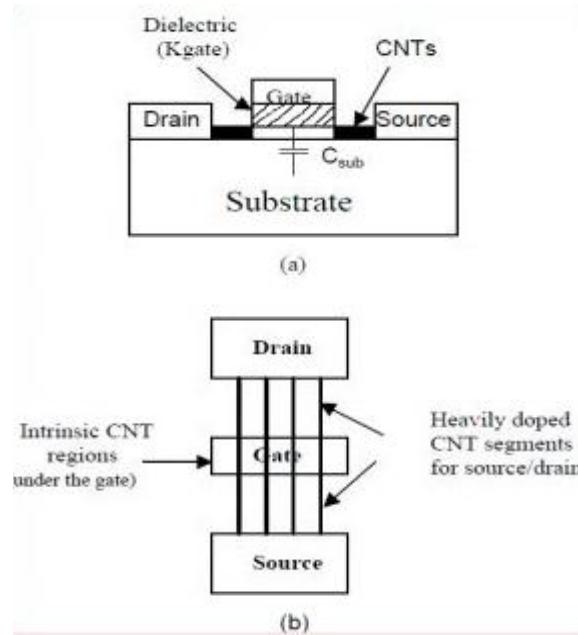


Figure 6.3.1: Schematic diagram of a carbon nanotube Transistor: (a) Cross Sectional view b) Top View [37].

As we approach into the nanometer regime, the scaling difficulties discussed in the previous chapters are significantly improved by using CNTFETs transistor for future technology. As MOSFET technology approaches saturation level CNTFETs are one of the best alternative for CMOS technology.

## **Conclusion:**

The economics behind the CMOS scaling is strong enough to keep it alive for several more generations to come. But at the same time, the conventional planar bulk silicon MOSFET continues to shrink aided by incorporation of new materials into the conventional MOSFET structure, be it the gate-dielectric or the gate electrode or the source/drain region. New compact models will continue to emerge alongside new FET architectures to enable evaluation of these new architectures. At the same time, the existing models for bulk FETs will continue to incorporate new physics to describe advanced bulk FETs. Nanoscale CMOS is an endless exciting road for both technology and model developers.

In conclusion, CMOS scaling below 100-nm channel length faces several fundamental limiting factors stemming from electron thermal energy and quantum-mechanical tunneling. Many of the potential barriers in a MOSFET that kept the standby leakage low are losing their effectiveness when scaled to lower barrier heights or thinner widths. Inevitably, both the standby power and the active power of a high-performance processor will rise. As a tradeoff, the performance gained from scaling will slow. Nevertheless, by using properly optimized doping profiles and pushing the silicon depletion width to the tunneling limit, it is likely that mainstream CMOS scaling will be extended to 20-nm channel length with Nano scaled gate oxides and voltage levels. Beyond that, cooling to low temperature might provide the additional design space needed to extend CMOS devices to 10 nm for server applications.

## References:

- [1] Y. Taur, D. A. Buchanan, W. Chen, D. J. Frank, K. E. Ismail, S. H. Lo, G. A. Sai-Halasz, R. G. Viswanathan, H. J. C. Wann, S. J. Wind, and H.-S. Wong, "CMOS Scaling into the Nanometer Regime," Proc. IEEE, vol. 85, pp. 486–504, 1997.
- [2] Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 μm MOSFET's: A 3-D "atomistic simulation study," IEEE Trans. Electron Devices, vol. 45, pp. 2505–2513, 1998.
- [3] Robert L. Boylestad, Louis Nashelsky, "Electronic Devices and circuit theory" 9<sup>th</sup> edition.
- [4] S.M.SZE, "Semiconductor Devices Physics and Technology", 2nd edition, pp- 178-213.
- [5] James E. Chung, Min-Chie Jeng, James E.Moon,Ping-Keung Ko, Chenming Hu, "Performance and reliability Design Issues for Deep-Sub micrometer MOSFET's", IEEE Transactions on Electron Devices, VOL.38, NO.3, MARCH 1991, pp-546.
- [6] <http://www2.egr.uh.edu/~smotamar/CMOS/CMOS.html>
- [7] F. Ji, J. P. Xu, J. J. chen, H. X. Xu, C. X. Li, P. T. Lai, "A Compact Threshold-Voltage Model of MOSFET with Stack High-k Gate Dielectric", IEEE,2009,pp-236,238.
- [8] Leland Chang, Stephen Tang, Tsu-Jae King, Jeffrey Bokor and Chenming Hu, "Gate Length Scaling and Threshold Voltage Control of Double-Gate MOSFETs", IEEE, 2000, pp- 31.2.2.
- [9] V. Narendar, Ramanuj Mishra, SanjeevRai, Nayana R and R.A. Mishra, "THRESHOLD VOLTAGE CONTROL SCHEMES IN FinFETS", International Journal of VLSI design & Communication Systems (VLSICS), Vol.3, No.2, April 2012, pp- 181-182.
- [10] S. Cristoloveanu, B.Blatlock, F. Allibert, B. Dufrene, M. Mojarradi, "The Four-Gate Transistor", ESSDERC, 2002, pp- 323.
- [11] S. M. Sze, "Physics of Semiconductor Devices", John Wiley and Sons,2<sup>nd</sup> edition, New Jersey, USA,pp-24,28,34,36,41
- [12] S. O. Kasap, "Principles of Electronic Materials and Devices", Tata McGraw-Hill,3<sup>rd</sup> edition . pp-381,314, 390
- [13] G. Joshi, D.N. Singh, and S. Thangjam, "Effect of Temperature Variation on Gate Tunneling Currents in Nanoscale MOSFETs", IEEE 2008
- [14] E. F. Schubert, "Physical Foundation of Solid State Devices", Rensselaer Polytechnic Institute, Troy NY, USA, 2009 edition. pp-140,
- [15] <http://www.4spe.org/plastics-encyclopedia/semiconductors>
- [16] <http://hyperphysics.phy-astr.gsu.edu/hbase/solids/band.html>
- [17] <http://hyperphysics.phy-astr.gsu.edu/hbase/electric/dielec.html#c1>
- [18] James C. Ellenbogen, Ph.D. "A Brief Overview of Nanoelectronic Devices", Nanosystems Group, The MITRE Corporation McLean, VA 22102, January1998, pp-2.
- [19] Hei Wong, Hiroshi Iwai, "On the scaling issues and high-k replacement of ultrathin gate dielectric for nanoscale MOS transistors", Japan, March 2006, pp- 1872-1881,1886.
- [20] Ban P. Wong, Anurag Mittal, Yu Cao, Greg Starr, "NANO-CMOS CIRCUIT AND PHYSICAL DESIGN", JHON WILEY & SONS, INC, PUBLICATION, USA, 2005, pp- 6-18, 27, 50-52.

- [21] J. Robertson, "High dielectric constant oxides", The European Physical Journal Applied Physics, Cambridge University Press, UK, 2004, pp-265-269, 273-278
- [22] Bradley D. Fahlman, "Materials Chemistry", Springer, 2008, pp-169-171.
- [23] Chi On Chui, "Nanoscale Germanium MOS Dielectrics—PartII: High- $\kappa$  Gate Dielectrics", IEEE Transactions on Electron Devices, VOL.53, NO.7, July 2006, pp-1509-1514
- [24] <http://academic.brooklyn.cuny.edu/physics/tung/Schottky/index.htm>
- [25] <http://academic.brooklyn.cuny.edu/physics/tung/Schottky/isr-simp.jpg>
- [26] Amit Chaudhry ,Jatindra Nath Roy, "A Quantum Mechanical Model of Gate Oxide Direct Current Density in High-k Dielectrics for Nanoscale MOS Applications", ELEKTRIKA, 13 (1),VOL. 13, NO. 1, 1-6, 2011,pp-2-4.
- [27] <http://www.siliconfareast.com/sio2si3n4.htm>
- [28] Subhra Dhar, Manisha Pattanaik, and Poolla Rajaram, "Advancement in Nanoscale CMOS Device Design En Route to Ultra-Low-Power Applications", *VLSI Design*, vol. 2011, Article ID 178516, doi:10.1155/2011/178516, India, 2011, pp-4.
- [29] Amit Chaudhry, Jatindra Nath Roy, "A Quantum Mechanical Model of Gate Oxide Direct Current Density in High-k Dielectrics for Nanoscale MOS Applications", ELEKTRIKA, 13 (1), VOL. 13, NO. 1, 1-6, 2011,pp-2-4.
- [30] Wen-Chin Lee and Chenming Hu, "Modeling CMOS Tunneling Currents Through Ultrathin Gate Oxide Due to Conduction and Valence Band Electron and Hole Tunneling", *IEEE TRANSACTIONS ON ELECTRON DEVICES*, VOL.48, NO.7, JULY2001, pp-1-4.
- [31] Chamila Siyambalapitiya, Robert H. Hyde, Kristin Kusmirek, SanjuktaBhanja, "Effectiveness of Knowledge Module on "Intel 45 nm Transistor and High-k Dielectric" into Undergraduate Semiconductor Devices Course", USA, 2012, pp-2.
- [32] I. M. Filanovsky and A. Allam, "Mutual compensation of mobility and threshold voltage temperature effects with applications in CMOS circuits," *IEEE Transactions on Circuits and Systems-I*, vol. 48, no. 7, July 2001, pp. 876-883.
- [33] Robert S. Chau, "Integrated CMOS Tri-Gate Transistors: Paving the Wayto Future Technology Generations", Intel Technology Magazine, August 2006, pp-4.
- [34] <http://newsroom.intel.com/docs/DOC-2032>
- [35] D. Hisamoto, W.-C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kua, E. Anderson, T.-J. King, J. Bokor, and C. Hu, "FinFET - A Self-Aligned Double Gate MOSFET scalable to 20nm," *IEEE Trans. Electron Devices*, vol. 47, no. 12, pp. 2320–2325, 2000.
- [36] Robert Chau, Mark Doczy, Brian Doyle, Suman Duttam Gilbert Dewey, Jack Kavalieratos, Ben Jin, Matthew Metz, AmlanMajumdar and Marko Radosavljevic, "Advanced CMOS Transistors in the Nanotechnology Era for High Performance, Low-Power Logic Applications", Components Research, logic Technology Development, Intel Corporation.

- [37] Mohini Polimetla and Rajat Mahapatra, "Analysis of Current Mirror in 32nm MOSFET and CNTFET Technologies", World Academy of Science, Engineering and Technology 60, 2011, pp-807 .
- [38] W. Lee, T. Osakama, K. Asada, and T. Sugano, "Design Methodology and Size Limitations of Submicrometer MOSFET's for DRAM Application", IEEE Transactions on Electron Devices, Vol. ED-35, 1988, pp. 1876-1884.
- [39] A.B. Bhattacharyya, "Compact MOSFET Models For VLSI Design", April 13.2009.
- [40] Sung-Mo Kang, Yusuf Leblebici, "CMOS Digital Integrated Circuits Analysis and Design", 3rd edition, pp-73,75,88,111.
- [41] S. Thompson, P. Packan, M. Bohr, MOS Scaling: Transistor Challenges for the 21st Century, Intel Technology Journal, Q398, pp. 1-19.
- [42] <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2965644/>
- [43] Long, R.D.; McIntyre, P.C. "Surface Preparation and Deposited Gate Oxides for Gallium Nitride Based Metal Oxide Semiconductor Devices. Materials" 2012, 5, 1297-1335.
- [44] <http://www.ioffe.rssi.ru/SVA/NSM/Semicond/GaN/optic.html>
- [45] Xiaoyan Liu, Jifeng Kang, Ruqi Han, and Yangyuan Wang, "threshold voltage model for MOSFETs with high-k gate dielectrics," IEEE Electron Device Letters, volume.23 (5), p.270, May 2002.
- [46] J. P Xu, F. Ji, P. T. Lai and 1.G Guan, "Influence of sidewall spacer on threshold voltage of MOSFET with high-k gate dielectric," Microelectronics Reliability, volume.48, p.181, 2008.
- [47] N. Sirisantana, L. Wei, and K. Roy, "High-performance low-power CMOS circuits using multiple channel length and multiple oxide thickness," in Proc. Int. Conf. Computer Design, 2000, pp. 227–232.
- [48] Md. Hasanuzzaman, Syed K. Islam, Leon M. Tolbert, "Effects of temperature variation (300–600 K) in MOSFET modeling in 6H–silicon carbide", Solid-State Electronics 48 (2004) 125–132, pg-132.