InGaSb n-MOSFET: Modeling & Performance Analysis

by

Muhammad Shaffatul Islam

A thesis submitted to the Department of Electrical and Electronic Engineering, Khulna University of Engineering & Technology (KUET) in partial fulfillment of the requirements for the degree
of
Master of Science in Electrical and Electronic Engineering

Department of Electrical and Electronic Engineering
Khulna University of Engineering & Technology
Khulna-9203, Bangladesh
December, 2015

Declaration

This is to certify that the thesis work entitled " ***InGaSb n-MOSFET: Modeling & Performance Analysis***" has been carried out by *Muhammad Shaffatul Islam* in the Department of *Electrical and Electronic Engineering*, Khulna University of Engineering & Technology, Khulna, Bangladesh. The above thesis work or any part of this work has not been submitted anywhere for the award of any degree or diploma.

Signature of Supervisor

Signature of Candidate

Approval

This is to certify that the thesis work submitted by *Muhammad Shaffatul Islam* entitled " *InGaSb n-MOSFET: Modeling & Performance Analysis*" has been approved by the board of examiners for the partial fulfillment of the requirements for the degree of *Master of Science in Electrical and Electronic Engineering* in the Department of *Electrical and Electronic Engineering*, Khulna University of Engineering & Technology, Khulna, Bangladesh in May 2015

BOARD OF EXAMINERS

____

Dr. Md. Rafiqul Islam
Professor,
Department of Electrical and Electronic Engineering,
Khulna University of Engineering & Technology,
Khulna-9203, Bangladesh.

Chairman
(Supervisor)

____

Head of the Department
Department of Electrical and Electronic Engineering,
Khulna University of Engineering & Technology,
Khulna-9203, Bangladesh.

Member

____

Dr. Ashraful Ghani Bhuiyan
Professor,
Electrical and Electronic Engineering,
Khulna University of Engineering & Technology,
Khulna-9203, Bangladesh.
.

Member

____

Dr. Md. Rafiqul Islam
Professor ,
Department of Electrical and Electronic Engineering,
Khulna University of Engineering & Technology,
Khulna-9203, Bangladesh.

Member

____

Dr. Anisul Haque
Professor,
Department of Electrical and Electronic Engineering,
East West University,
Dhaka, Bangladesh.

Member (External)

Acknowledgements

First, I express my highest gratitude to the almighty ALLAH for the completion of my thesis work and also for my graduate study times in KUET.

Then I wish to express my sincere gratitude to my supervisor Dr. Md. Rafiqul Islam, Professor, Department of Electrical and Electronic Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh, for his invaluable support and excellent guidance. He has taught me not only a lot of technical work but also broadened my way of thinking and how to approach challenges.

I would like to thank Mr. Md Nur Kutubul Alam for his valuable help from different aspect through out my graduate research years. I am also thankful to him for his extensive support in technical discussions and coding related matters.

I must thank all the faculty members of EEE department, KUET for their help to complete the research works as a full-time student.

Abstract

In recent years multigate MOSFET structures are becoming very much attractive to face the scaling-induced challenges in future electronic devices. Due to the high electron and hole mobility, antimonide-based III-V materials are highly interested in the channel of low power and high-speed all-antimonide CMOS digital logic devices.

In this thesis work InGaSb-based double gate (DG) nMOSFET architectures with DG-nMOSFET are proposed and numerically simulated. To assess and hence to compare the ballistic performance of DG-JnMOSFET and, non-equilibrium greens function method is utilized under the framework of well-known SILVACO's ATLAS device simulation package without taking into account of scattering. Wave function penetration to the oxide is taken into account in the simulation. In this study the effect of gate length on drain current and other logic figures of merit like subthreshold slope (SS), I_{ON}, $I_{OFF,}$ and I_{ON}/I_{OFF} are investigated. The results obtained for DG-nMOSFET demonstrate that there is negligible dependence on the "ON" current with the gate length. Small increase in the OFF current is found for decreasing gate length from 15 nm to 13 nm, however, the OFF current increases significantly when the gate length decreases from 13 nm. Subthreshold slope decreases with increasing gate length. The maximum subthreshold swing is evaluated 80 mV/decade for the gate length 7nm, which reduces to 70 mV/decade for the gate length 15 nm. Also the threshold voltage decreases with decreasing gate length. Effect of the gate oxides (Al_2O_3 and HfO_2) in same equivalent oxide thickness (EOT) is also studied. The shift in threshold voltage and change in I_{ON}/I_{OFF} are found strong dependence with gate oxides having different dielectric constants. Further, the effect of gate underlap is analyzed for the gate lengths 10nm and 15nm with different underlap lengths. Drain current found to vary significantly for different underlap lengths when gate length is 10nm. However, for the gate length 15 nm the variation of drain current is insignificant with underlap length. The threshold voltage is also found to have strong dependence with gate length and gate underlap length.

Contents

		PAGE
Title Page		i
Declaration		ii
Certificate of Research		iii
Acknowledgement		iv
Abstract		v
Contents		vii
List of Tables		viii
List of Figures		ix
CHAPTER 1	Introduction	1
	1.1 Motivation: Advantage of III-V Compound Semicoductors	3
	1.2 Recent Advances in III-V based devices for Logic Applications	5
	1.3 Junctionless MOSFET Architecture	7
	1.4 Scope of The Thesis	10
	1.5 Layout of The Thesis	11
CHAPTER 2	Double Gate MOSFET’s	16
	2.1 Bulk MOSFET Scaling Challanges	18
	2.2 Double Gate MOSFET’s Bands	22
	2.3 Quantum Confinement Effects in Nanoscale DG-MOSFET’s	26
CHAPTER 3	Quantum Transport Simulation of Proposed InGaSb DG-nMOSFET	31
	3.1 Quantum Transport Simulation Using Mode Space Approach	32
	3.2 Device Design & Simulation Parameters	37
CHAPTER 4	Simulation Results & Discussion	43
	4.1 Effect of Gate Length on the ballistic performance of nanoscale InGaSb DG-nMOSFET	44
	4.2 Effect of gate oxide keeping same EOT	
	4.3 Effect of gate underlap on ballistic performance	49

CHAPTER 5

Conclusion

LIST OF TABLES

Table No.	Description	Page
3.1	Parameters related to InGaSb DG-nMOSFET device simulation	38
3.2	Parameters related to InGaSb Junctionless DG-nMOSFET device simulation	40
4.2.1	Logic Figure of merits for two different oxide	53
4.4.1	Logic Figure of merits for two different gate lengths	62

LIST OF FIGURES

Figure No.	Description	Page
1.1	Power dissipation in CPU for various technologies as a function of the year of introduction.	2
1.2	Transistor operating voltage range showing energy efficient computation for near threshold voltage operation	2
1.3	Electron and hole mobility of various III-V compound semiconductors.	3
1.4	Electron mobility and hole mobility of $In_xGa_{1-x}Sb$ as a function of composition.	4
1.5	Source and drain doping of inversion-mode and junctionless transistor with short channel and ultra-short channel.	9
2.1	Moore's Law	21
2.2	Schematic illustration of the scaling of Si technology by a factor α	23
2.3	Channel length reduces the barrier for the majority of carriers to enter the channel.	24
2.1.4	Subthreshold characteristics of long- and short-channel devices at low and high drain bias [41].	21
2.1.5	Simulated IV curves for11-nm-channel-length bulk MOSFETs. Each gray curve corresponds to a different random placement of the dopants in keeping with the designed average doping profiles. The solid black curve is the geometric average of the curves, while the dashed curve is the expected IV curve based on continuum doping profiles.	22
2.2.1	Structure of (a) double-gate, (b) ground-plane (or back-gate FET, BG). The upper figures are cross-sections of the devices; the lower figures illustrate their energy band diagrams. The gate work functions of the top and bottom gates can be the same for symmetric DG FET (SDG) or different for asymmetric DG FET (ADG) [49]	22
2.2.2	Compressive strain-dependent first light hole wave function of $Ga_{0.65}In_{0.35}Sb$ QW Electric field lines originating from the drain for (a) bulk MOSFET and (b) DG MOSFET. In bulk MOSFETS, electric field lines can penetrate to the source that causes the DIBL. IN DG MOSFET, in contrast, lines are terminated on either top or bottom gate..	23
2.2.3	Simulated (using FIELDAY) threshold voltage roll-off for DG MOSFET with tox= 1.5nm and 1nm. The silicon channels are	24

	undoped. Gate material is assumed to have a mid-gap work function [50].	
2.2.4	Simulated (using FIELDAY) DIBL for double-gate MOSFET with tox= 1.5nm and 1nm. The Si channel is undoped and T= 300K [50].	24
2.2.5	Subthreshold swing obtained with Medici simulations versus channel length for two different values of silicon thickness (N_A=$10^{16}$$cm^{-3}$, tox=1.5 nm, and VDS=0.1 V) [51].	25
2.2.6	Illustration of band-to-band tunneling in a reverse biased p-n junction	26
2.3.1	Double-gate SOI MOSFET.	27
2.3.2	Energy band diagram in the transverse direction of DG-SOI MOSFET of Fig. 2.12.	27
2.3.4	Energy gap widening and the splitting of the conduction band into subbands due to carrier confinement.	28
2.3.5	(a) Simulated electron density dependence on gate voltage. (b) Threshed voltage dependence on silicon layer thickness. Gate oxide layer thickness is 7 nm. Buried oxide (BOX) layer thickness is 80nm. Channel doping concentration (N_A) is 10^{17} cm^{-3}. Temperature is 300 K. [57].	28
3.1.1	A model of DG MOSFET divided into vertical slices. The horizontal spacing between the slices is Δx.	33
3.2.1	Device structure used for ballistic performance analysis of InGaSb double gate nMOSFET. Various dimensions are given in Table-3.1 .	38
3.2.2	Device structure used for ballistic performance analysis of InGaSb Junctionless double gate nMOSFET. Various dimensions are given in Table-3.2 .	39
4.1.1	I_D-V_G characteristic's as a function of different gate length.	45
4.1.2	Conduction band profile for different gate length (a) OFF state condition (b) ON state condition.	46
4.1.3	(a) Subthreshold slope as a function of gate length (b) Threshold voltage as a function of gate length.	47
4.1.4	ON state and OFF state currents as well as I_{ON}/I_{OFF} ratio as a function of gate length (a) ON state current (b) OFF state current (c) I_{ON}/I_{OFF} ratio.	48
4.2.1	I_D-V_G characteristics of double gate MOSFET for different gate	50

	dielectric material.	
4.2.2	Conduction band profile of the double gate MOSFET (a) across the channel (b) along the channel, at V_D=0.05V and V_G=2V.	52
4.2.3.	Variation of subthreshold swing (SS) as a function of gate oxide dielectric constant.	53
4.3.1	Drain current behavior for three different underlap length for (a) gate length 10nm and (b) gate length 15nm.	55
4.3.2	(a) Threshold voltage and (b) Subthreshold slope as a function of gate underlap for two different gate length 10 and 15nm.	55
4.3.3	(a) ON current. (b) OFF current as a function of gate underlap for two different. (c) I_{ON}/I_{OFF} ratio as a function of different gate underlap for the gate length 10nm and 15nm.	57
4.3.5	Channel position dependent ON state conduction band profile for different gate underlap when gate length is (a) 10nm (b) 15nm.	57

CHAPTER 1

Introduction

1.1 Background

A transistor is essentially a "switch" which controls the flow of current between two terminals, using a "gate" that acts as a valve. The goal is to make this switch work at very high speed with very low power dissipation. Scaling the dimensions of transistors has led to exponential increase in density and switching speed, and a similar decrease in the switching energy [1]. The supply voltage of the transistor has been scaled in a similar proportion as the dimensions to reduce the power dissipation in the transistors as the number of transistors per chip increased exponentially. The processor power continued to rise even with scaling the supply voltage due to die size and fast frequency scaling [2]. As the total power dissipation in the microprocessor reached 100W (Fig. 1.1), the frequency and die scaling are stopped. MOSFETs entered a "power constrained scaling" phase where the power dissipation was limited to 100W per chip [3]. To integrate more functions or to pack more transistors into the microprocessor, the supply voltage is required to be scaled. Since the threshold voltage and the sub-threshold slope of the transistor did not scale with the supply voltage [4], the leakage power density continued to increase as the supply voltage was scaled. Scaling the supply voltage without scaling the threshold voltage would lead to reduce switching speed and slower performance, which is not desired. Innovations in the past decade (high-κ metal gate, strain and tri-gate architectures) have improved the performance of the transistors in the past decade and helped scale the supply voltage moderately (from ~1.2V for planar 65nm to ~0.9V for 22nm tri-gate transistors) [5]. Future transistor scaling will require enhancement in device electrostatics (multi-gate), channel transport enhancement (beyond strained Silicon) and reduction in parasitic (contact, junction engineering etc.). Near threshold voltage operation can improve the energy efficiency of computing as shown in Fig.1.2, and can help scale the supply voltage [6]. Introducing a new material which has a higher carrier velocity than silicon would help to aggressively scale the supply voltage to around 0.5V, while maintaining the same or better performance levels as Si at high gate overdrive. III-V semiconductors comprising of elements from group III and V of the

periodic table have extra-ordinary mobility (which translates to enhanced carrier velocities for short channel MOSFETs) compared to silicon and can help to achieve near threshold voltage operation. Figure 1-3 shows the electron and hole mobility of various III-V compound semiconductors [7]. Antimonide based semiconductors are very attractive since they have high electron as well as high hole mobility.

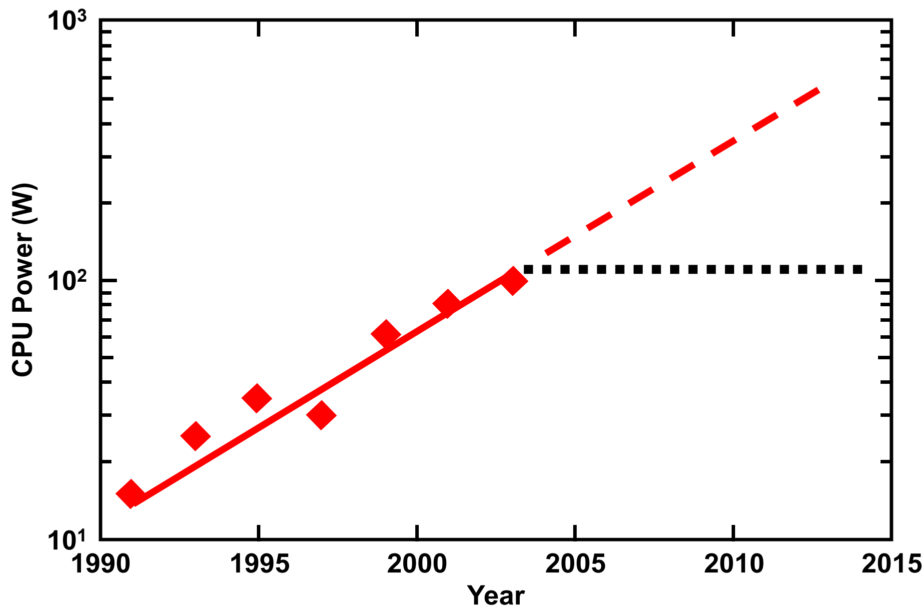


Fig 1.1: Power dissipation in CPU for various technologies as a function of the year of introduction (Source: Intel [3])

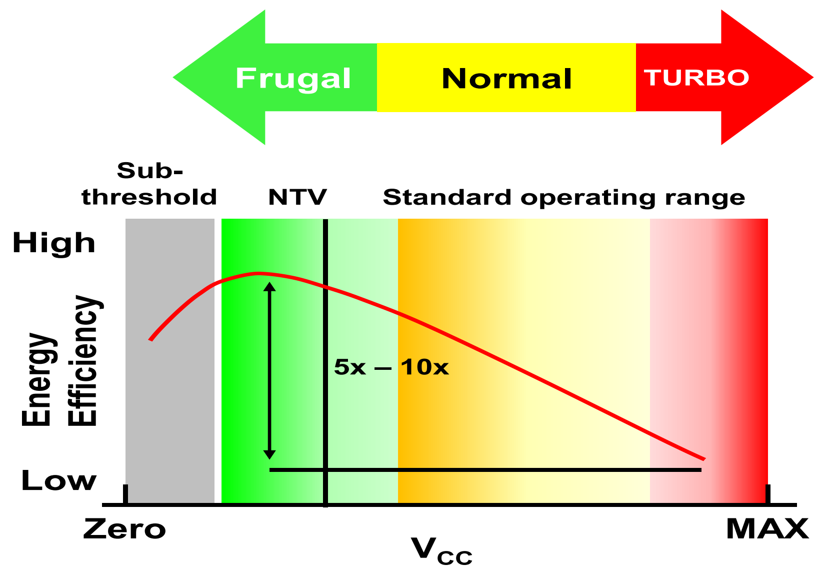


Fig 1.2: Transistor operating voltage range showing energy efficient computation for near threshold voltage operation. (Source: Intel Labs ISSCC [5])

1.2 Motivation: advantages of III-V compound semiconductors

Most of the III-V semiconductors have a direct band gap, thus making them highly suitable for optoelectronic applications. Motivation for consideration of III-V materials in traditional MOS transistor structures for digital applications is the extremely high electron mobility due to their small effective mass in the Γ valley. Several main challenges have prevented the implementation of III-V materials for future CMOS applications. What attracts attention to III-V CMOS technology has been the extraordinary electron mobility of certain III-V compounds as shown in Figure 1.3[1]. The room temperature mobility of electrons and holes in inversion layers and quantum wells is shown as a function of the actual semiconductor lattice constant. For relaxed layers under no strain, the lattice constant is its natural one. Due to their unique material properties, antimony (Sb) based III-V materials (materials comprised of elements from group III and V of periodic table) are of strong interest as channel material for low power all-antimonide complementary metal-oxide-semiconductor (CMOS) digital logic. Sb-based MOSFETs (Metal-Oxide-Semiconductor Field Effect Transistors) can operate at high speed and very low supply voltage, which promises to dramatically lower the power dissipation in future high-speed logic circuits. Antimonide based materials have high electron as well as high hole mobility [6][7].

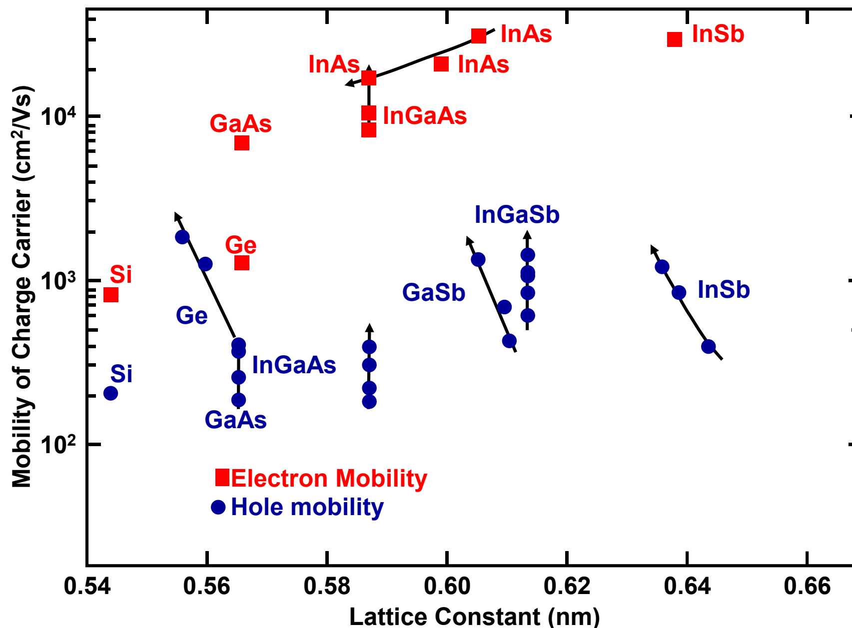


Fig 1.3: Electron and hole mobility of various III-V compound semiconductors (Source: del Alamo MIT [1]). Antimonides are very attractive since they have high electron as well as high hole mobility.

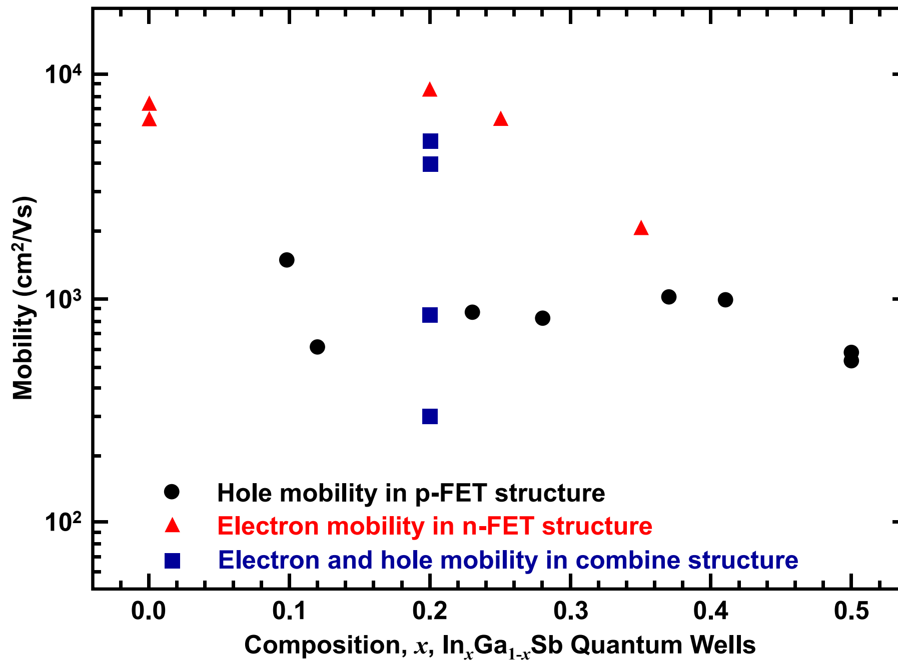


Fig 1.4: Electron mobility and hole mobility of $In_xGa_{1-x}Sb$ as a function of composition [8]

Figure 1.4 shows the electron and hole mobility of InGaSb QW heterostructure for different percentage of strain. In InGaSb channel material the sheet charge density can be achieved $10^{12}/cm^2$and the mobility is 10 times higher than Si. This can give rise to a substantial improvement in drive current for InGaSb based devices which can help scale supply voltage to about 0.5V. The injection velocity V_{inj}, to first order, is proportional to $\sqrt{\mu}$, in which μ corresponds to low field carrier mobility [9].The ballistic limit refers to the case where carriers travel from source to drain without being scattered and losing their momentum. This can only be achieved at extremely short gate length. As the device scales down, it gradually approaches to the ballistic limit, which leads to the increase of injection velocity until it saturates at extremely short channel length. Hence a study of III-V materials in short channel devices could be an excellent indicator for the performance of nanoscale MOS transistors made of III-V'semiconductors.

1.3 Recent advances in III-V based devices for logic applications

In recent years, as the continuous scaling on Si CMOS is becoming more and more difficult, pursuit of alternative materials and device structures beyond planar Si CMOS is a dynamic area in both academia and industry, which has led people to look at the possibility of III-V once again. Historically, there have been several challenges for III-V to be realized for logic applications. Breakthrough's in recent research have partially solved, or relieved some of the challenges.

The first III-V MOSFET in history, a GaAs MOSFET, was built by the Radio Corporation of America in 1965 [10]. Soon after that it was realized that finding the low-defect, thermodynamically stable insulator is the key to III-V MOSFETs. Unlike the perfect interface between SiO_2 and Si, compound semiconductors don't have an ideal native oxide to form a reliable MOSFET structure. In the case of GaAs, its oxide is a leaky and defective mixture of Ga_2O_3, As_2O_3 and As_2O_5, which causes pinning of the Fermi level and nullifies the device performance by preventing the formation of inversion layers through the bending of the surface potential. Since late 1990s, lots of research has been carried out to solve this issue and significant progress has been made. Nowadays, by using the technique of atomic-layer-deposition (ALD), fairly good interface can be formed between high k dielectrics (Al_2O_3 and HfO_2) and $In_xGa_{1-x}As$ channel owing to the "self-cleaning" effect in ALD. With this technique, 400 nm surface channel $In_{0.65}Ga_{0.35}As$ MOSFETs with record high current of more than 1A/mm have been fabricated, and the upper limit on the average interface trap density (D_{it}) was estimated to be ~ 1.7 $\times 10^{12}$/cm^2-eV [11]. In parallel with the development of surface channel III-V MOSFETs with ALD, a quantum well structure with $In_xGa_{1-x}As$ or InAs as channel layer has attracted a lot of attention in parallel. Such quantum

well III-V FETs are similar to the traditional HEMTs structure widely used in RF CMOS, but their potential for digital circuits have been studied recently intensively. Different from the surface channel MOSFETs, the In_xGa_{1-x} channel layer in the quantum well structure is sandwiched between two $In_{0.52}Al_{0.48}As$ on the top of bottom of it, which eliminates the interface issues due to the good interface quality between $In_xGa_{1-x}As$ and $In_{0.52}Al_{0.48}As$. The quantum well structure has been successfully fabricated on Si wafers with GaAs and $In_xGa_{1-x}As$ gradual buffer layers [12], and has shown very good I-V characteristics and scaling behavior [13-15, 16,17, 18], which is suitable for future post Si CMOS. There are two big issues related with the III-V quantum well structure. Firstly the metal gate is deposited on $In_{0.52}Al_{0.48}As$ upper barrier layer which acts as an insulator layer, but such metal/semiconductor forms Schottky barrier which leads to large gate leakage current under both low and high gate biases. This issue has recently been solved by Intel with a high-K ($TaSiO_x$) and InP composite layers deposited on top of the $In_{0.7}Ga_{0.3}As$ channel, which reduces the gate leakage by a factor of >1000 [19]. Another issue is more serious, the series resistance in the III-V quantum well transistors is about 2~3 times larger than that in up-to-date Si MOSFETs [20, 21]. Unless significant advances were made in reducing the series resistance, the excellent intrinsic device performance will be lost as the device will be dominated by large extrinsic resistance with scaling.

Another historical challenge for the implementation of III-V CMOS logic is to look for high mobility III-V p-type MOSFETs. Intel has recently reported high performance 40nm gate length InSb p-channel transistors using the quantum well structures with compressively strain between the InSb channel and the $Al_{0.4}In_{0.6}Sb$ barrier layers [22]. Compared to Si, such p-channel strained InSb quantum well transistors show~10 times

lower power at the same speed, or ~2 times higher speed at matched power. Strain application in III-V therefore, provides a possible solution for p-type transistors in the future.

It was estimated that if III-V were to replace Si CMOS, it would be beyond the 15nm node. At this ultra-short gate length, conventional planar structure in surface channel or quantum well MOSFETs will not be suitable due to the worse 2D effects. Double-gate or multi-gate MOSFETs will be necessary to meet the SCE (short channel effect) requirement. Though InGaAs has excellent electron mobilities , the hole mobility lags far behind and is even worse than that of silicon. The only material systems within III-V's that can offer both good electron and hole transport are antimonide compunds namely GaSb, InGaSb and InSb. Through pseudomorphic buffer engineering, the hole mobility can be further increased by introducing compressive biaxial strain. This demonstrates the important potential role of antimonide compounds. Recently, InGaSb based p-channel MOSFET's are studied experimentally for 150 µm gate length which utilizes a self-aligned process with high-k dielectric between the heterostructures and metal gate to cut down the gate leakage while preserving the high mobility in the channel. Low resistance source/drain contacts right next to the gate lead to reduction in the excess resistance and also enable devices with short pitch suitable for large scale integration. Also interface related issues are considered in this study [23]. In III-V based n-MOSFET's, III-V materials have the highest mobility/injection velocity, the major concern however is the degradation of the device performance due to low density-of-states (DOS) (low effective mass of carriers) and excessive spillover of charge from Γ to L valley at high sheet charges [24,25]. It is also known from the literature that the increase in In content in InGaAs/ InGaSb channel nMOSFET's : velocity injection and

separation between Γ to L increases while DOS decreases. Recently simulation based study of InGaSb based DG-nMOSFET has been performed considering tight binding (TB) parameters for ternary $In_xGa_{1-x}Sb$ calculated following the virtual crystal approximation (VCA) incorporating compositional disorder effect and fitted to bulk band gap of ternary compound [26-27]. 1D Poisson equation perpendicular to channel direction is coupled with TB Hamiltonian by Hartee-Fork potential in the gate stack. Dangling bonds at interface are pacified by hydrogen termination of hybridized orbitals to eliminate all the states within band gap [28]. A ballistic transport model was adopted to assess transport of electrons [29]. Also simulation was performed to investigate the device performance of III-V ultra-thin body field effect transistors with the consideration of the effects of materials, body thickness and dielectric effect based on the top-of – barrier model.

1.4 Scope of the thesis

The main objective of this thesis is to study the transfer characteristics of InGaSb DG-nMOSFET's along with the logic figure of merits with the considerations of the effects of different gate lengths, and different equivalent oxide thicknesses (EOT). The effect of underlap lengths for different EOT, gate lengths has also been studied. To investigate the transfer characteristics and related logic figure of merits, a widely used pure quantum mechanical approach is adopted for the non-equilibrium condition. Non equllibrium Green's function is used to assess the transport behavior in two dimensions under effective mass approximation. Simulation is performed for different gate lengths and the drain current behavior is explained from the conduction band diagram by the top-of-barrier model. Also logic figure of merits like subthrehold swing (SS), on current (I_{ON}), off current (I_{OFF}), on/off current ratio are studied.

1.5 Layout of the thesis

The dissertation describes the modeling and performance analysis of InGaSb DG-nMOSFET. The research work is divided into several chapters in which device design, simulation and modeling, device performance and output characteristics are discussed. In chapter 1 the advantages of the III-V semiconductor materials over other materials and also the recent advances in III-V MOSFET's for logic applications are discussed. The recent works on InGaSb based MOSFET's are also described in this chapter.

Chapter 2 starts with the transistor scaling trends and the double gate (DG) MOSFET structure. Also bulk MOSFET scaling challenges are discussed along with the advantages of the DG MOSFET over the bulk MOSFET. At the end of this chapter, quatum-mechanical confinement effects in nanoscale DG-MOSFET are discussed showing how these effects play a critical role in determining the device electrical characteristics.

Chapter 3 starts with the need of numerical simulation of semiconductor devices to study the performance of devices in quantum regime. Non equilibrium Green's function (NEGF) is widely used in nanoscale devices to measure the transport characteristic's. Usually there are two types of approach for the solution of NEGF equation, we used Mode space (MS) approach which is suitable for pure quantum -mechanical treatment of DG-nMOSFET. Mathematical formulation used in mode space (MS) approach has been discussed in this chapter. The device structure and material parameters used in this study are also discussed in this chapter.

Chapter 4 presents the result and discussion obtained from the numerical simulation of InGaSb DG-nMOSFET. The ballistic performance of the InGaSb DG-nMOSFET for two different equivalent oxide thicknesses (EOT) and logic figure of merits are explained

with respect to the gate length. Also the effects of same EOT for two different oxides on the ballistic current along with the other logic performance metrics are investigated. Finally the effects of gate underlap on the ballistic performance of the device are studied.

CHAPTER 2

2.1 Background

Field Effect Transistor (FET) invention was a breakthrough in the field of integrated circuits. The basic idea of the FET was patented by Lilienfeld in 1930 [30]. The idea was extensively explored and experimentally studied by numerous researchers at different universities and companies [31]. In 1960, the Metal Oxide Semiconductor Filed Effect Transistor (MOSFET) was practically demonstrated by Dawn Kahng and Mohamed Attalla [32]. The demonstrated transistor was p-channel MOSFET with 25 µm drawn channel length and 250 µm channel width.

Technology generation is denoted by a figure, for example 0.5 µm. The figure is called a node and refers to the minimum metal line width that can be afforded by the technology. This restricts the minimum drawn MOSFET channel length that can be fabricated using the technology. Each new technology generation node is about 0.7 smaller than the previous node. This periodic size reduction is called scaling. Smaller MOSFETs are desirable for several reasons: 1) Increasing the packing density and, as a result, the number of functions per chip. This, in turn, reduces the cost per function, 2) Smaller transistors switch faster because smaller gate capacitance leads to smaller RC delay and 3) Shorter transistors have higher transconductance that provide a higher gain-bandwidth product.

Transistor scaling was very aggressive in the past 40 years. In 1971, the technology generation node was the 10 µm. At that time, Intel® introduced its first processor (the "4004" microprocessor). The "4004" employed a 10 µm silicon-gate PMOS technology and was built of 2,300 transistors. The current technology generation node is the 45 nm. Recently, Intel® announced that production for the "Penryn" processor will be started in 2008. The "Penryn" employs a 45 nm high-k metal-gate CMOS technology with about 800 million transistors [33]. In 1965 Gordon Moore empirically predicted that the number of transistors per chip will be doubled every two years (every technology generation).

It is interesting to note that Moore's law is satisfied across the entire technology nodes. It is noted that, this agreement between technology nodes and Moore's law isn't by chance. The semiconductor industry is very keen to continue its exponential advance. It is predicted that conventional CMOS technology will come to its end at the 22 nm node [34]; scaling bulk MOSFET beyond the 45 nm technology is accompanied with great challenges. Challenges include Short-Channel Effects (SCEs), Drain Induced Barrier Lowering (DIBL), random dopant fluctuation and high leakage current. Industrial and academic efforts are now being exerted to find a replacement for conventional bulk MOSFET. There are two main directions: 1) adopting new transistor structures or 2) adopting new materials other than Silicon (Si) [34]. Searching for new device structure that circumvents the forgoing challenges while, in the same time, scalable to 22 nm, Double-Gate (DG) MOSFET is found to be the most promising candidate [35, 36, 37,39, 40, 41]. This is because, for a given gate length and oxide thickness, DG MOSFETs show better short-channel behavior, lower leakage current, higher drive current, and smaller subthreshold swing than other MOSFET structures [39]. In the same time, they have better power consumption and scaling capabilities [40].

In this chapter, an overview of the bulk MOSFET scaling challenges is given in section 2.1. In section 2.2 the DG MOSFET structure is presented and then, the advantages of the DG MOSFET over bulk MOSFET are explained. The principal Quantum Mechanical Effects encountered in DG MOSFETs are illustrated in section 2.3 showing how these effects play a critical role in determining the device electrical characteristics.

2.2 Future technology node requirements

The factors or features to be distinguished between several logic technologies options are:

a) High performance (HP):

The high performance corresponds to high complexity ICs that require high clock frequencies and at the same time can deal with high power consumption. The increase in clock frequency from one technology node to the next can be achieved at the device level by an improvement of the intrinsic switching time of a transistor of 17% per year, while maintaining the transistor off-state current to a value acceptable from a power consumption point of view. The intrinsic switching time (τ) is the time needed by a transistor supplying on-state current to make the gate of an identical transistor switching from ground to the supply voltage.

$$I = \frac{Q}{t} = \frac{CV}{\tau}$$

$$\tau = \frac{Q}{I} = \frac{CV}{I} \quad (2.1)$$

where,

C → Gate capacitance V → Supply voltage I → On-state current of the device

Therefore, the most efficient way to achieve enhanced performance is to scale the gate length of the transistor aggressively. Consequently, this will result in reduced gate capacitance while increasing the on-state current.

b) Low operating power (LOP):

The low operating power technology option mainly aims at relatively high performance mobile applications such as notebook computers. The key challenge is to increase the circuit performance while decreasing the dynamic power consumption as much as possible when the device is operating. The dynamic power consumption at the device level is a measure of power-delay product given by

$$P_\tau = \frac{CV}{I} P = \frac{CV}{I} VI$$

$$P_\tau = CV^2$$

$$P_\tau \propto V^2 \tag{2.2}$$

Therefore, the most efficient way to reduce the dynamic power consumption is thus to reduce the supply voltage as far as possible. Dynamic power consumption (given by 0.5 CV^2f) decreases.

c) Low stand-by power dissipation (LSTPD):

The low stand-by power option is used for lower performance, low-cost consumer applications such as cellular phones. For such applications, the main concern or key issue is to continue increasing circuit performance while maintaining the power consumption as low as possible when the IC is idle. This static power consumption at the transistor level is governed by the leakage current of the devices. Therefore, low stand-by power technology option requires very low transistor-off state currents as well as very low parasitic currents such as gate leakage.

2.3 Technology scaling

The lateral geometric dimensions of devices and interconnects are reduced. This reduction in size is referred to as "scaling" of the geometric dimensions of the integrated circuits(IC). The minimum feature size is smaller size of object (e.g. gate length or interconnect line width) on IC. Over the past decades, the MOSFET has continually been scaled down in size; typical MOSFET channel lengths were once several micrometers, but modern integrated circuits are incorporating MOSFETs with channel lengths of tens of nanometers. As a consequence of this minimum feature size of ICs, the numbers of transistors have increased over time. The semiconductor industry maintains a "roadmap", the ITRS, which sets the pace for MOSFET development. Historically, the difficulties with decreasing the size of the MOSFET have been associated with the semiconductor device fabrication process, the need to use very low voltages, and with poorer electrical performance necessitating circuit redesign and innovation (small MOSFETs exhibit

higher leakage currents, lower output resistance, lower transconductance, interconnect capacitance, process variations, etc.

2.4 Why MOSFET scaling?

Smaller size of MOSFETs is highly desirable for several reasons. The primary reason to make transistors smaller in size is to integrate more and more number of devices in a given chip area. This results in a chip with the same functionality in a smaller area, or chips with more functionality in the same area. Since fabrication costs for a semiconductor wafer are relatively fixed, the cost per integrated circuits is mainly related to the number of chips that can be produced per wafer. Hence, smaller ICs allow more chips per wafer, reducing the price per chip. In fact, over the past 30 years the number of transistors per chip has been doubled every 23 years once a new technology node is introduced. For example the number of MOSFETs in a microprocessor fabricated in a 45 nm technology can well be twice as many as in a 65 nm chip. This doubling of transistor density was first observed by Gordon Moore in 1965 and is commonly referred to as Moore's law.

It is also expected that smaller transistors switch faster. For example, one approach to size reduction is a scaling of the MOSFET that requires all device dimensions to reduce proportionally. The main device dimensions are the channel length, channel width, and oxide thickness. When they are scaled down by equal factors, the transistor channel resistance does not change, while gate capacitance is cut by that factor. Hence, the RC delay of the transistor scales with a similar factor. However, this has been traditionally the case for the older technologies, for the state-of-the-art MOSFETs reduction of the transistor dimensions does not necessarily translate to higher chip speed because the delay due to interconnections is more significant.

2.5 Moore's law

In the last forty-five years since 1965, the price of one bit of semiconductor memory has been dropped 100 million times. The primary engine that powered the proliferation of electronics is "miniaturization". By making the transistors and interconnects smaller, more circuits can be fabricated on each silicon wafer and therefore each circuit becomes

smaller. Miniaturization has also been responsible to the improvements in speed and power consumption in ICs.

Gordon Moore made an empirical observation in 1965 that the number of devices on a chip doubles every 18 or 24 months or so as shown in Figure 2.1 [41].

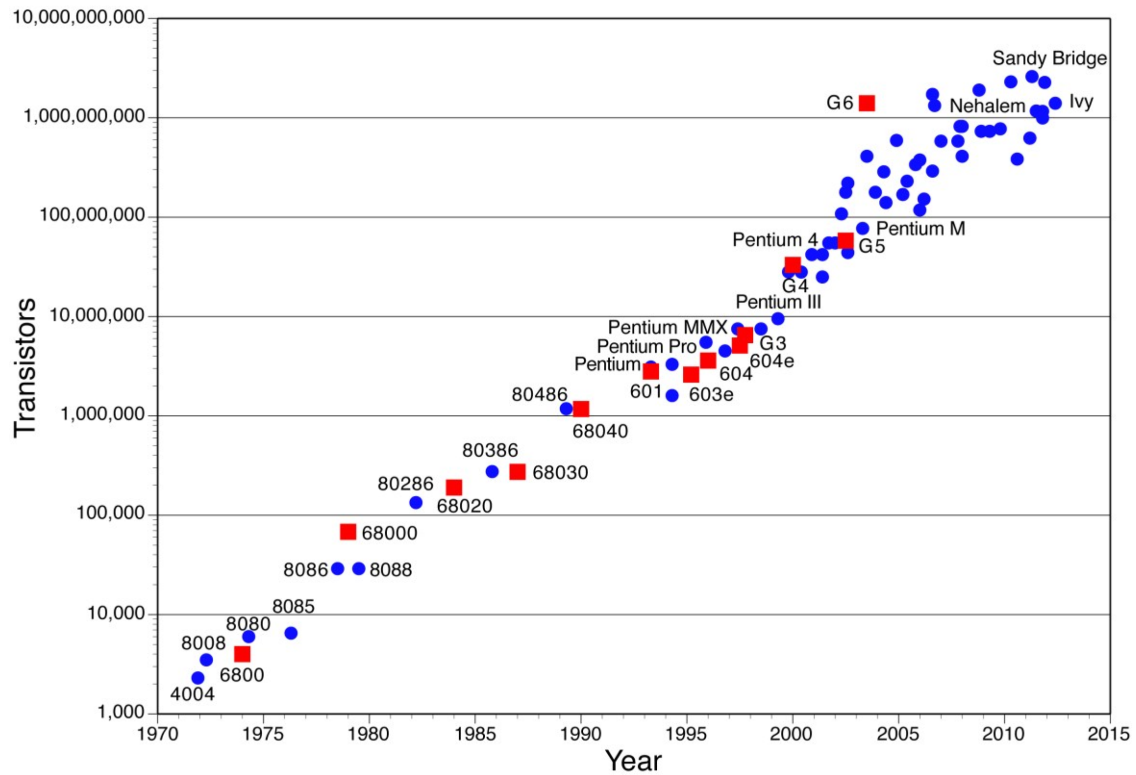


Figure 2.1: Moore’s Law

This Moore’s law is a succinct description of the rapid and persistent trend of miniaturization. Each time the minimum line width is reduced, we say that a new technology generation or technology node is introduced. Examples of technology generations is shown in Table 2.1:

Table 1.1: Improvements in Technology Node over the years

Year	2004	2006	2008	2010	2011	2013	2016	2022
Technology Node:	90 nm	65 nm	45 nm	32 nm	22 nm	16 nm	14 nm	10 nm

The numbers shown in the table refers to the minimum metal line width. Poly-Si lengths may be even smaller. At each new node, all the features in the circuit layout, such as the contact holes, are reduced in size to 70% of the previous node. Historically, a new technology node is introduced every two to three years. The main reward for introducing a new technology node is the reduction of circuit size by half. (70% of previous line width means 50% reduction in area i.e., 0.7×0.7=0.49). Since nearly twice as many circuits can be fabricated on each wafer with each new technology node, the cost per circuit is reduced significantly which drives down the cost of ICs.

It is intuitive that Moore's Law cannot be sustained forever. However, predictions of size reduction limits due to material or design constraints, or even the pace of size reduction, have proven to elude the most insightful scientists. The predicted 'limit' has been dropping at nearly the same rate as the size of the transistors. Further technology scaling requires major changes in many areas, including: 1) improved lithography techniques and non-optical exposure technologies; 2) improved transistor design to achieve higher performance with smaller dimensions; 3) migration from current bulk CMOS devices to novel materials and structures, including silicon-on-insulator, strained Si and novel dielectric materials; 4) circuit sensitivity to soft errors from radiation; 5) smaller wiring for on-chip interconnection of the circuits; 6) stable circuits; 7) more productive design automation tools; 8) denser memory cells, and 9) manageable capital costs. Metal gate and high-k gate dielectrics were introduced into production in 2007 to maintain technology scaling trends.

In addition, packaging technology needs to progress at a rate consistent with on-going CMOS technology scaling at sustainable cost/performance levels. This requires advances in I/O density, bandwidth, power distribution, and heat extraction. System architecture will also be required to maximize the performance gains achieved in advanced CMOS and packaging technologies.

2.6 MOS scaling trends

For many years , the shrinking of MOSFETs has been governed by the ideas of scaling. The basic idea is illustrated in the Figure 2.2 [42]: a large FET is scaled down by factor α to produce a smaller FET with similar behavior.

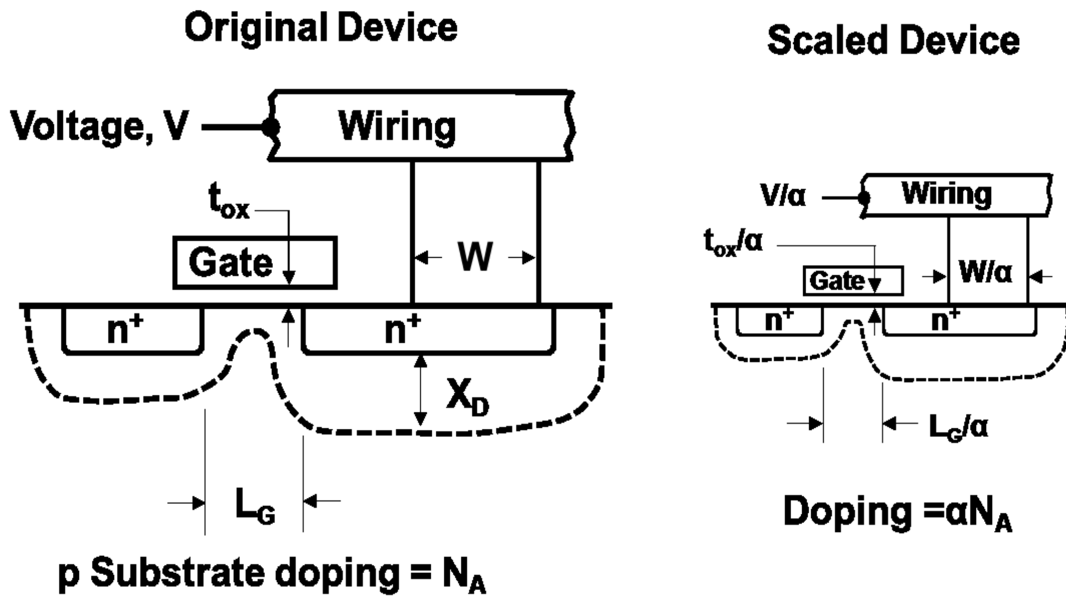


Figure 2.2 Schematic illustration of the scaling of Si technology by a factor α

Device scaling is based upon simple principles; by reducing the sizes of devices and interconnects, density, and power, the speed and performance of transistors can be improved. Device scaling mainly focuses on

a) scaling of threshold voltage with feature size
b) scaling of gate oxide thickness with feature size
c) scaling of supply voltage with feature size

With technology scaling, the MOS device channel length is reduced. When the dimensions of a MOSFET are scaled down, both the voltage level and the gate-oxide thickness are also reduced. The supply voltage V_{dd} has to be scaled down in order to keep the power consumption under control. The transistor V_{th} also had to be scaled down to maintain a high drive current and achieve performance improvement. In a given technology node, since the source-body and drain-body depletion widths are predefined based on the doping, the rate at which the barrier height increases as a function of distance from the source into the channel is constant. Therefore, when the channel length is reduced, the barrier for the majority carriers to enter the channel is also reduced as shown in Figure 2.3 [43]. As a result threshold voltage is reduced. In short channel transistor, the barrier height and therefore the threshold voltage are a strong function of the drain voltage. Figure 2.3 [43] indicates, the barrier reduces as the drain voltage is increased.

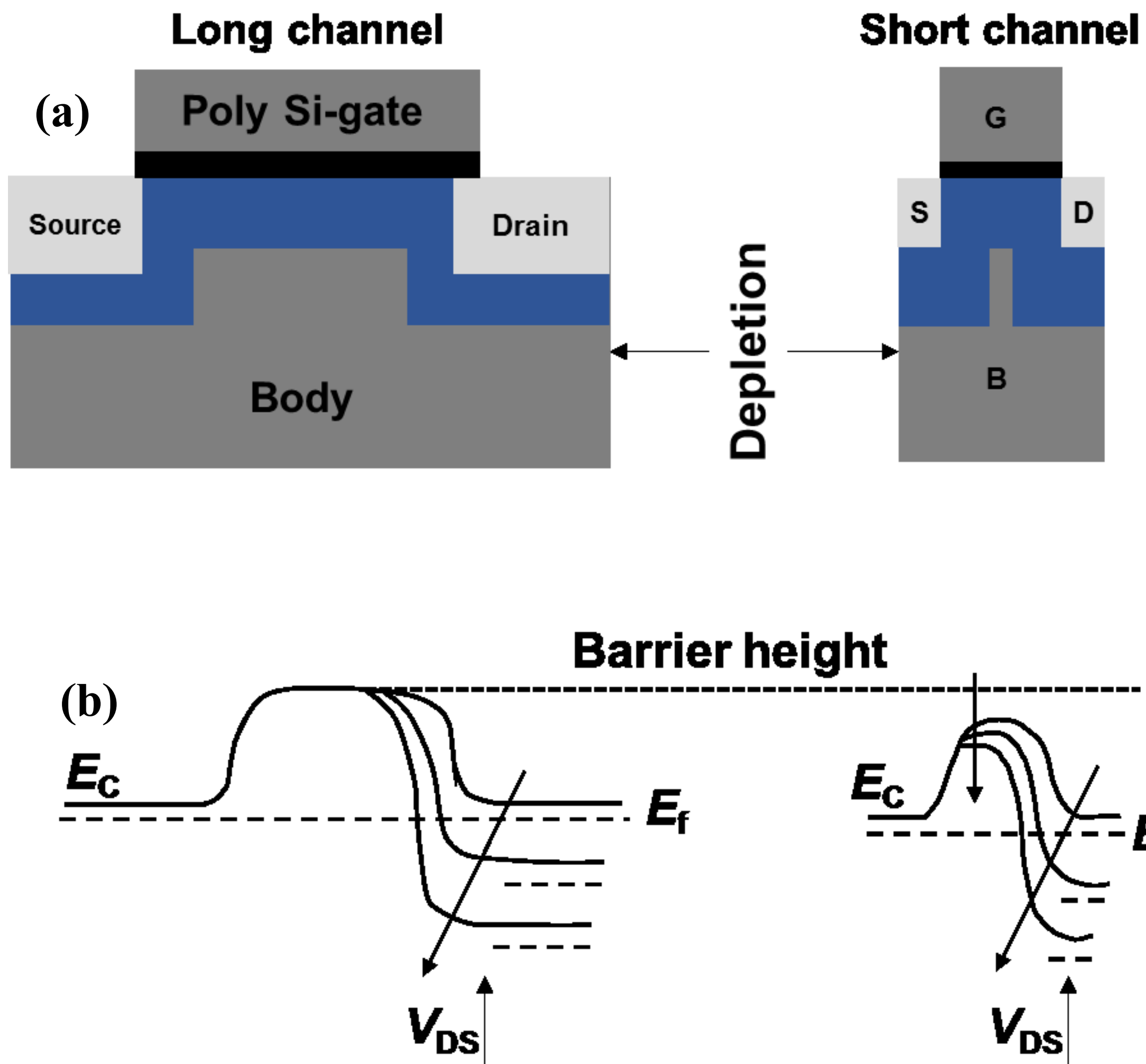


Figure 2.3 Channel length reduces the barrier for the majority of carriers to enter the channel.

Aggressive scaling of MOS technology in recent years has reduced the silicon dioxide (SiO_2) gate dielectric thickness below 20˚ A. In 90 nm, the gate oxide consists of about 5 atomic layers equivalent to 1.2 nm in thickness. The thinner the gate oxide, the higher the transistor current and consequently the switching speed. The scaling trend of power supply voltage (V_{dd}), threshold voltage (V_{th}), and gate-oxide thickness (T_{ox}) as a function of MOS channel length is shown in Figure 2.4 [44].

When V_{dd} is reduced towards shorter channel lengths, it becomes increasingly difficult to satisfy both the performance and the off current requirements. Trade-off between leakage current and circuit speed stems due to subthreshold nonscalability. For this reason and for compatibility with the standardized power supply voltage of earlier generation

systems, the general trend is that V_{dd} has not been scaled down in proportion to L and V_{th} has not been scaled down in proportion to V_{dd} as shown in Figure 2.4 [44].

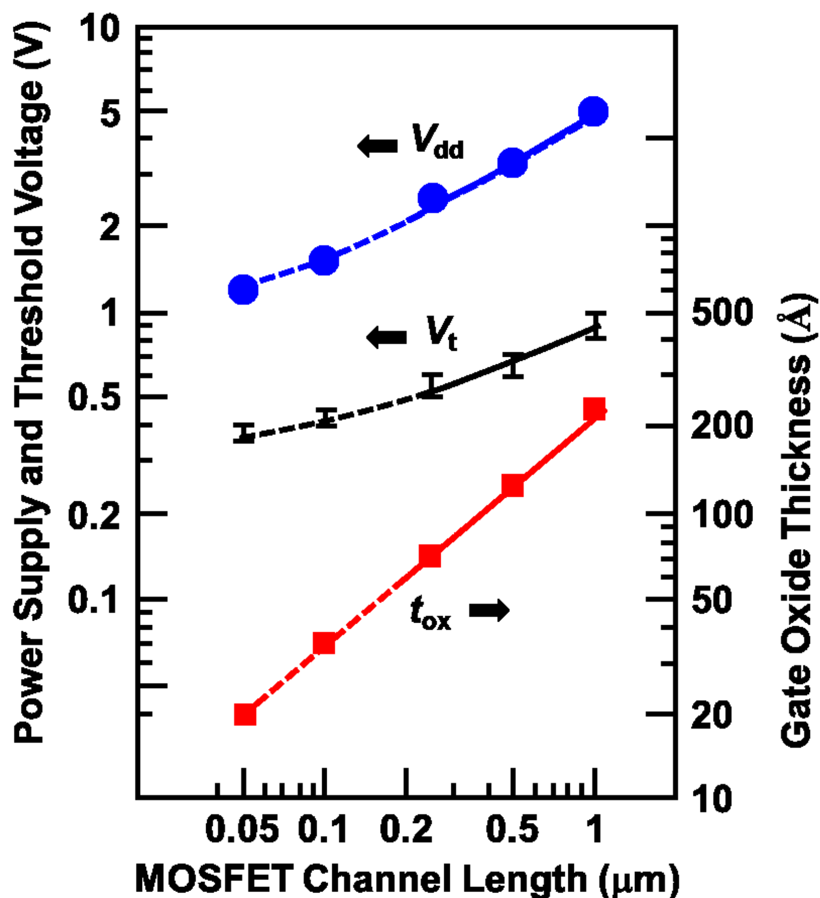


Figure 2.4 Scaling trend of power supply voltage (V_{dd}), threshold voltage (V_{th}), and gate-oxide thickness (T_{ox}) as a function of CMOS channel length.

When all of the voltages and dimensions are reduced by the scaling factor α and the doping and charge densities are increased by the same factor, the electric field configurations inside the FET remains the same as it was in the original device. This is called constant field scaling, which results in circuit speed increasing in proportion to the factor α and circuit density increasing as α_2. These scaling relations are shown in the second column of Table 2.2 along with the scaling behavior of some of the other important physical parameters

Table 2.2: Technology Scaling Rules for Three Cases

Physical Parameter	Constant Electric Field Scaling Factor	Generalized Scaling Factor	Generalized Selective Scaling Factor
Channel Length, Insulator Thickness	$1/\alpha$	$1/\alpha$	$1/\alpha_d$
Wiring Width, Channel Width	$1/\alpha$	$1/\alpha$	$1/\alpha_w$
Electric Field in Device	1	ε	ε
Voltage	$1/\alpha$	ε/α	ε/α_d
On-Current per Device	$1/\alpha$	ε/α	ε/α_w
Doping	α	$\varepsilon\alpha$	$\varepsilon\alpha_d$
Area	$1/\alpha^2$	$1/\alpha^2$	$1/\alpha_w^2$
Capacitance	$1/\alpha$	$1/\alpha$	$1/\alpha_w$
Gate Delay	$1/\alpha$	$1/\alpha$	$1/\alpha_d$
Power Dissipation	$1/\alpha^2$	ε^2/α^2	$\varepsilon^2/\alpha_w\alpha_d$
Power Density	$1/\alpha^2$	ε^2	$\varepsilon^2/\alpha_w\alpha_d$

α is the dimensional scaling parameter, ε is the electric field scaling parameter, and α_d and α_w are separate dimensional scaling parameters for the selective scaling case. α_d is applied to the device vertical dimensions and gate length, while α_w applies to the device width and the wiring.

2.7 Challenges to miniaturization of MOSFETs

Despite formidable challenges, however, many of those in the research community and industry do envision close variants of conventional microelectronic transistors becoming miniaturized into the nanometer-scale regime. For example, in ITRS, published by the Semiconductor Industry Association, projects that chips will be made from transistors with major features (gate lengths) of 70 nm in the year 2010. Individual working transistors with 40 nm gate lengths have already been demonstrated in silicon. Transistors with gate lengths as small as 25 nm have been made using Strained Silicon (S-Si). However, to provide points of reference for contrasting nanoelectronic devices

with scaled-down FETs, a few of the obstacles to FET scaling are simply enumerated below, in increasing order of their intractability.

a) High electric fields:

Due to a bias voltage being applied over very short distances, can cause avalanche breakdown by knocking large numbers of electrons out of the semiconductor at high energies, thus causing current surges and progressive damage to devices. This may remain a problem in nanoelectronic devices made from bulk semiconductors.

b) Heat dissipation:

Heat dissipation of transistors (and other switching devices), due to their necessarily limited thermodynamic efficiency, limits their density in circuits, since overheating can cause them to malfunction. This is likely to be a problem for any type of densely packed nanodevices.

c) Vanishing bulk properties and non-uniformity of doped semiconductors on small scales:

This can only be overcome either by not doping at all (accumulating electrons purely using gates, as has been demonstrated in a GaAs heterostructure) or by making the dopant atoms form a regular array. Molecular nano-electronics is one path to the latter option.

d) Shrinkage of depletion regions:

Shrinkage of depletion regions until they are too thin to prevent quantum mechanical tunneling of electrons from source to drain when the device supposedly is turned off. The function of nanoelectronic devices is not similarly impaired, because it depends on such tunneling of electrons through barriers.

e) Shrinkage and unevenness of the thin oxide layer beneath the gate:

This prevents electrons from leaking out of the gate to the drain. This leakage through thin spots in the oxide also involves electron tunneling. Long ago, MOSFETs were big and could be described via drift currents and carrier control via the gate capacitance. Now MOSFETs are small in order to increase their operation speed. Pushing the

dimensions of the gate length down influences the electrostatics of the devices. In order to preserve the electrostatic integrity of the MOSFET scaling has proceeded in a controlled way: L_g ↓ has to go together with t_{ox} ↓, N_A ↓, t_{si} ↓, V_{dd} ↓, and W ↓.

CHAPTER 3

Quantum transport simulation of proposed InGaSb DG-nMOSFET

3.1 Introduction

Numerical simulation of semiconductor devices is an important complement to analysis and experiment. Device simulation provides essential capabilities that absent in both analysis and experiment. The prediction of device behavior before fabrication is the most advantageous feature of device simulation over experimental measurement. Running a computer based simulation costs nothing in comparison with fabrication and different types of measurements. Device simulation allows a spectrum of conditions (under which fabricated device may be damaged) to be tested inexpensively. Moreover it is impossible to simulate devices which aren't yet manufacture able. Rapid device scaling pushed the dimensions of the field effect transistors to the nanometer regime. Scaling bulk MOSFET's to and beyond 32nm node technology will face fundamental limits [45] and are expected to be replaced with ultra-thin body (UTB) SOI MOSFET's and Double-Gate (DG) MOSFET's. The International Technology Roadmap for Semiconductors 2007 (ITRS) projection for the DG MOSFETs physical gate length for high performance logic is 4.5 nm for the year 2022 . For extremely scaled dimensions, the quantum effects (illustrated in chapter 1) including carrier confinement along the direction normal to the Si-SiO2interface, tunneling through the gate oxides, source to drain tunneling and quantum interference play an important role in determining the DG MOSFETs characteristics. These effects can be accurately predicted only using quantum mechanical based device simulation [46].

The non-equilibrium Green's function formalism (NEGF) provides a rigorous description of quantum transport in nanoscale devices. For ballistic transport, NEGF is equivalent to solving Schrödinger equation. The NEGF, however, provides a powerful way for treating the boundary of 2D and 3D problems. Moreover, it provides a way for treating scattering within non-ballistic devices. The rigorous description for this formalism can be found in the literature [47, 48, 49, 50, 51, 52, 53, 54, 55, 56] where it is described using an advanced language in the quantum mechanics world, namely, the second quantization language. A simpler description can be found also in the literature [57,58, 59, 60].

3.2 Quantum transport simulation using mode space approach

Quantum Transport involves solution of scrhodinger equation using open boundary conditional points where electron comes into or leaves form the region (device) under study. It can be solved in real space (RS) or mode space (MS) approach.

Computational efficiency is needed to make the self-consistent approach suitable for device design and characteristics prediction. The NEGF method has the advantage of being rigorous but the disadvantage of being heavy in computations [61]. Real-space (RS) representation is the most accurate, yet complex, representation used in the NEGF [62]. For nanoscale transistors, it is more convenient, to solve Schrödinger equation in the mode-space (MS) [63,64] where computational burden is affordable. The uncoupled mode-space (UMS) provides enormous saving in the computational burden but suffers from being valid only for ultra-thin body double-gate (DG) MOSFET [61]. In contrast, the coupled mode-space (CMS) can be used for either thin or thick bodies but it is more computationally demanding than the UMS. In this chapter we review the MS approach and examine the validity of the UMS.

In the NEGF framework, a suitable set of basis function is chosen in terms of which the operators like the Hamiltonian operator and the Green's function are represented. The MS approach is based on the assumption that the active device is decoupled from the gate contacts [63]. Decoupling is achieved by applying closed boundary condition at the insulator-metal interface. It may be also applied at the semiconductor-insulator interface if the conduction band offset between the semiconductor and the insulator is taken to be infinite. Applying the closed boundary condition at whether of the two interfaces gives rise to subbands or modes. The subbands are the eigenfunctions associated with the confinement in the gate confinement direction (y-direction in Figure 4.1.1. The model DG MOSFET is given in Figure 4.1.1 where it is divided into vertical slices with Δx spacing in the x- direction. The subbands of the structure are obtained by solving a 1D schrodinger equation in the y direction with each vertical slice, positioned at any $x=x'$, of the device along the x direction:

$$-\frac{\hbar^2}{2m^*y}\frac{\partial^{2\chi(n)}(x',y)}{\partial y^2}+E_c(x',y)^{\chi(n)}(x',y)= E^{(n)}(x')^{\chi(n)}(x',y) \qquad (3.1)$$

where $\chi^{(n)}$ (x', y) is the eigenfunction in y-direction at $x=x'$ and $E^{(n)}(x')$ represents the bottom of the subband at $x=x'$and n is the subband index.

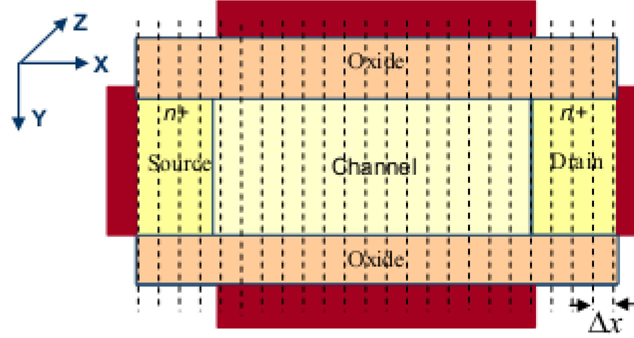


Figure 3.1.1 A model of DG MOSFET divided into vertical slices. The horizontal spacing between the slices is Δx.

At each vertical slice x', the central difference approximation for the second derivative $\chi^{(n)}$is applied :

$$\frac{\partial^2 \chi_j^{(n)}}{\partial y^2} = \frac{\chi_{j+1}^{(n)} - 2\chi_j^2 + \chi_j^{(n)}}{(\Delta y)^2} \tag{3.2}$$

where $\chi_j^{(n)} = \chi^{(n)}(x', y_j)$

Substituting eq. (3-2) into eq. (3-1), we obtain:

$$-t_y \chi_{j-1}^{(n)} + 2t_y \chi_j^{(n)} - t_y \chi_{j+1}^{(n)} + E_c \chi_j^{(n)} = E^{(n)} \chi_j^{(n)} \tag{3.3}$$

using eq. (3-3) at each node along the vertical slice x' and applying closed boundary condition at the semiconductor insulator interface , a set of linear equation are obtained and cast in the matrix form:

$$\begin{bmatrix} 2t_y + E_{c1} & -t_y & 0 & \dots & 0 \\ -t_y & 2t_y + E_{c1} & -t_y & 0 & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & 0 & -t_y & 2t_y + E_{cN-1} & -t_y \\ 0 & \dots & 0 & -t_y & 2t_y + E_{cN_y} \end{bmatrix} \begin{Bmatrix} \chi_1^{(n)} \\ \chi_1^{(n)} \\ \vdots \\ \chi_{N_y-1}^{(n)} \\ \chi_{N_y}^{(n)} \end{Bmatrix} = E^{(n)} \begin{Bmatrix} \chi_1^{(n)} \\ \chi_1^{(n)} \\ \vdots \\ \chi_{N_y-1}^{(n)} \\ \chi_{N_y}^{(n)} \end{Bmatrix} \tag{3.4}$$

Eq. (3-4) is an eigenvalue problem that can be solved forN_y different subbands. Each subband n has an eigenenergy $E^{(n)}$ and corresponding eigenvector $\chi^{(n)}$ within each slice

positioned at x' along x direction. Only the lowest N_m subbands are ignored due to their high subband energies $E^{(n)}$ and consequently low occupation probability.

The essence of the Mode space (MS) representation is the expansion of the 2D wave function $\psi(x,y)$ in terms of subbands, obtained from solving eq. 3.4, in the form:

$$\psi\,(x,y) = \sum_{n=1}^{n=N_m} \phi^{(n)}(x)\chi^{(n)}(x,y) \tag{3.5}$$

where $\phi^{(n)}(x)$ are the expansion coefficients. As indicated in the previous paragraph, only few subbands are taken into consideration and, as a result, the summation is truncated at $n= N_m$.

Substituting eq. (3-5) and multiplying by $\chi^{*(m)}(x,y)$ and integrating over y gives [65] :

$$-\frac{\hbar^2}{2m^*x}\frac{\partial^2\phi^{(m)}(x)}{\partial x^2} + [E^m(x) - E_l]\phi^m(x) = \sum_{n=1}^{n=N_m} A_{mn}\phi^{(n)}(x) \tag{3.6}$$

Where $A_{mn}(x)$ is the operator given by:

$$A_{mn}(x) = \frac{\hbar^2}{2m^*x}\left[2\int dy\chi^{*(m)}(x,y)\frac{\partial}{\partial x}\chi^{(n)}(x,y)\frac{\partial}{\partial x} + \int dy\chi^{*(m)}(x,y)\frac{\partial^2}{\partial x^2}\chi^{(n)}(x,y)\right] \tag{3.7}$$

Eq. (3-6) is the coupled mode space (CMS) transformation of the Real Space approach.

For each node m, the right hand side involves a summation over all other modes and the mth mode itself. This summation gives rise to coupling between the modes. The ultra-thin body DG-MOSFET's channel is fully depleted and the shape of the potential $E_c\,(x,\,y)$ along the y-direction varies slowly with x- position, hence the variation of χ with x can be neglected and eq (3-6) becomes

$$-\frac{\hbar^2}{2m^*x}\frac{\partial^2\phi^{(m)}(x)}{\partial x^2} + E^{(m)}(x)\phi^{(m)}(x) = E_l\phi^m(x) \tag{3.8}$$

Eq.(3-8) is called the uncoupled mode-space transformation of Real space solution of the NEGF equation. Eq. (3-8) represents the 1D transport model equation that will solved using the NEGF for each subband m.

As the total number of modes corresponds to the dimension of α_i (N_y), the size of the problem doesn't change if the non-equillibrium Green's function is expanded in the eigenmode basis. However in nanostructures with strong confinement, only few low energy modes are populated depending on the geometry of the device, the effective mass in the direction of confinement and the doping concentration. Consequently, only a reduced number of modes N_m needed to be considered, with the property $N_m<N_y$: increasing the number of modes to a value of superior to N_m must not change the carrier and current densities any more. The total mode space Green's function matrix (lesser or

retarded) $\mathbf{G_m}$ has the size$(N_xN_m) \times (N_xN_m)$ instead of $(N_xN_y) \times (N_xN_y)$ for its real space counterpart $\mathbf{G_{rs}}$. To find a transformation from G_{rs} toG_{ms}, $v^i = \Phi_1^i \ldots \Phi_N^i$matrices size of (N_yN_m) are initially created . At position x_i.v^i contains all the mode necessary to expand the real space Green's function localized there. Note that N_m can vary along the device transport axis in the MOSFET. For the formalism derivation, however N_m denotes the number of considered modes at each position. It is always much smaller then N_y. In a second step a transformation matrix $\mathbf{U}$, with size $(N_xN_y) \times (N_xN_m)$, is generated: it contains $N_x\mathbf{v^i}$ defines above.

$$\mathrm{U} = \frac{1}{\sqrt{\mathrm{norm}}}\begin{pmatrix} \mathbf{V}^1 & 0 & \ldots & 0 \\ 0 & \mathbf{V}^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \ldots & \ddots & \mathbf{V}^{\mathrm{N}x} \end{pmatrix} = \frac{\mathrm{U}}{\sqrt{norm}} \tag{3.9}$$

Where $\frac{1}{\sqrt{\mathrm{norm}}}$ is the normalization constant such that the matrix product $\mathbf{U^T.U}$equals the identity matrix $\mathbf{I}$ of size $(N_xN_m) \times (N_xN_m)$. Finally $\mathbf{U}$ relates $\mathbf{G_{rs}}$ to$\mathbf{G_{ms}}$by

$$\mathbf{G_{rs}} = \mathbf{U} \cdot \mathbf{G_{ms}} \cdot \mathbf{U^T} \tag{3.10}$$

Equation (3.9) is the matrix generalization of discretized real space Green's function $G_{i1i2j1j2}\,(E,K_Z)$ and can be expanded in coupled mode (eigenfunction) space as

$$\mathrm{G}_{i_1i_2j_1j_2}(E,K_z) = \sum_{nm} G_{i_1i_2nm}(E,K_z)\emptyset_n^{i_1}(y_{j_1})\,\emptyset_m^{i_1*}(y_{j_2}) \tag{3.11}$$

Under Ballistic condition the equation of motion of the retarded Green's function of the matrix $\mathbf{G^R}$ can be expressed by [13]

$$\left(\mathbf{E} - \mathbf{H} - \sum \mathbf{RB}\right)\mathbf{G^R} = \mathbf{I} \tag{3.12}$$

And lesser Green's function $\mathbf{G^<}$ is obtained with

$$\mathbf{G^<} = \mathbf{G^R} \sum < \mathbf{B}\,\mathbf{G^A} \tag{3.13}$$

Equation (3.11) is used to simplify the steady state equation of the motion (3.12) and (3.13) for the retarded and lesser Green's function respectively. Multiplying them on the left with $\mathbf{U^T}$and on the right side with $\mathbf{U}$ (both left hand side and right side arguments) replacing $\mathbf{G_{RS}}$ by equation (3.9), we obtain the following system of equations

$$(\mathbf{E} - \mathbf{H_{ms}} - \textstyle\sum_{\mathbf{ms}}^{\mathbf{RB}}\ )\ \mathbf{G_{ms}^R} = \mathbf{I} \tag{3.14}$$

$$\mathbf{G}_{\mathrm{ms}}^{<} = \mathbf{G_{ms}^{<R}} \textstyle\sum_{\mathbf{ms}}^{\mathbf{<B}} \mathbf{G_{ms}^{A}}.$$

Each matrix has size $(N_xN_m) \times (N_xN_m)$.The mode space version of the Hamiltonian is given by the block tridiagonal matrix $\mathbf{H_{ms}} = \mathbf{U^T.H.U}$. The block α_i in the effective mass Hamiltonian is replaced by $\boldsymbol{\alpha}_\mathbf{i}^{\mathbf{diag}} = \mathbf{v^{iT}}.\boldsymbol{\alpha}^\mathbf{i}.\mathbf{v^i}$. Since the modes $\boldsymbol{\Phi}_n^{\,i}$ do not necessarily have

the same shape all along the device transport axis,in general $\mathbf{v^{iT}.v^{i\pm1}} \neq \mathbf{I}$ (an $N_m \times N_m$identity matrix) and the $\boldsymbol{\gamma}_{\mathbf{ii\pm1}}$ blocks may be full: these represents the coupled mode effect, whose absence signifies that an equation of motion is solved for each mode independently from the others. The boundary self-energies$\sum_{\mathbf{ms}}^{\mathbf{RB}}$ and $\sum_{\mathbf{ms}}^{\mathbf{<B}}$ are directly computed in mode space and not with $\sum_{\mathbf{ms}}^{\mathbf{B}} = \mathbf{U^T}.\sum^{\mathbf{B}}.\mathbf{U}$because their real space calculation is computationally inefficient as mentioned earlier.

Solving the NEGF in coupled mode space from Eq. (3.13) presents a substantial improvement over the real space calculation from Eq. (3.11) and (3.12): the size of the linear system decreases form $(N_xN_y) \times (N_xN_y)$to $(N_xN_m) \times (N_xN_m)$. There is a gain of (N_y/N_m)in the size of the blocks building the Hamiltonian $\mathbf{H_{ms}}$, and therefore the size of the matrices that need to be inverted in the recursive algorithm. After the solution of the standard equation (3.13) is obtained carrier and current densities can be computed: Venugopal *et al* pointed out the mode space to real space transformation in Eq. (3.12) can be used for the purpose, the diagonal element of the $\mathbf{G_{rs}}^{<}$being the carrier density at each grid point. In real space method the computational costs are still very high due to the boundary condition calculations and the size of the block matrices (N_xN_y) and need to be inverted N_x times for each energy point and for each conduction band minimum . Once the system is solved the, the 2D current density vector $[J_x(x_i, y_j) J_x(x_i, y_j)]^T$ and the charge distribution $n(x_i, y_j)$ are obtained with,

$$n(x_i, y_j) = -\frac{i}{L_z \Delta_z \Delta_y} \sum_{k_z \sigma} \int \frac{dE}{2\pi} \boldsymbol{G}_{iijj}^{<}(k_z; E) \tag{3.15}$$

The coupled mode space approach, while keeping all the relevant physics, considerably simplifies the high computational burden of an real space simulation. So, the discretized real space Green's function

$$G_{i_1 i_2 j_1 j_2}(E, K_z) = \sum_{n,m} G_{i_1 i_2 nm}(E, K_z)\, \Phi_n^{i_1}(y_{j_1}) \Phi_n^{i_1}(y_{j_2}) \tag{3.16}$$

Now the insertion of the values of Eq. (3.15) into (3.14) leads to,

$$n(x_i, y_j) = -\frac{i}{L_z \Delta_x \Delta_y} \sum_{n,m} \sum_{k_z,\sigma} \times \int \frac{dE}{2\pi} G_{iinm}^{<}(k_z; E) \Phi_n^i(y_j) \Phi_m^{i*}(y_j) \tag{3.17}$$

$$J_x(x_i, y_j) = -\frac{2e}{\hbar L_z \Delta_y} \sum_{n,m} \sum_{k_z,\sigma} \times \int \frac{dE}{2\pi} \mathrm{Re}[h_{i+1ijj} G_{ii+1nm}^{<}(k_z; E) \Phi_n^i(y_j) \Phi_m^{i+1*}(y_j) \tag{3.18}$$

$$J_y(x_i, y_j) = -\frac{2e}{\hbar L_z \Delta_y} \sum_{n,m} \sum_{k_z,\sigma} \times \int \frac{dE}{2\pi} Re[h_{iij+1j} G^{<}_{iinm}(k_z; E) \Phi^{i}_{n}(y_j) \Phi_m(y_{j+1}) \quad (3.19)$$

All the correlation elements $G^{<}_{iinm}$ from the diagonal blocks of $\mathbf{G_{ms}}$are involved in the calculation of the carrier density $n(x_i, y_j)$ and the y current component $J_x(x_i, y_j)$ all the correlation elements $G^{<}_{ii+1nnm}$ from the first non-diagonal blocks $\mathbf{G_{ms}}$must be taken int account. An incomplete calculation consist in keeping only the diagonal mode space expansion coefficients $G^{<}_{iinm}$ of the non equilibrium Green's function and applying a transformation similar to Eq. (3.10). This modifies Eq. (3.17) to

$$n(x_i, y_j) = -\frac{i}{L_z \Delta_x \Delta_y} \sum_{n,m} \sum_{k_z,\sigma} \times \int \frac{dE}{2\pi} G^{<}_{iinm}(k_z; E) \, |\Phi^{i}_{n}(y_j)|^2 \quad (3.20)$$

Working with coupled mode space does not only mean only solving the systems of Eq. (3.13), but also adapting the calculation of the physical quantities to the corresponding space. The correlations $G^{<}_{iinm}$ resulting from the coupled mode effects play a non-negligible role in the calculation of all observables thus the inclusion of all the mode correlation effects ensures a correct electron density and current conservation.

3.23 Device design and simulation parameters

The device schematic which is used for the simulation in this thesis is given below.

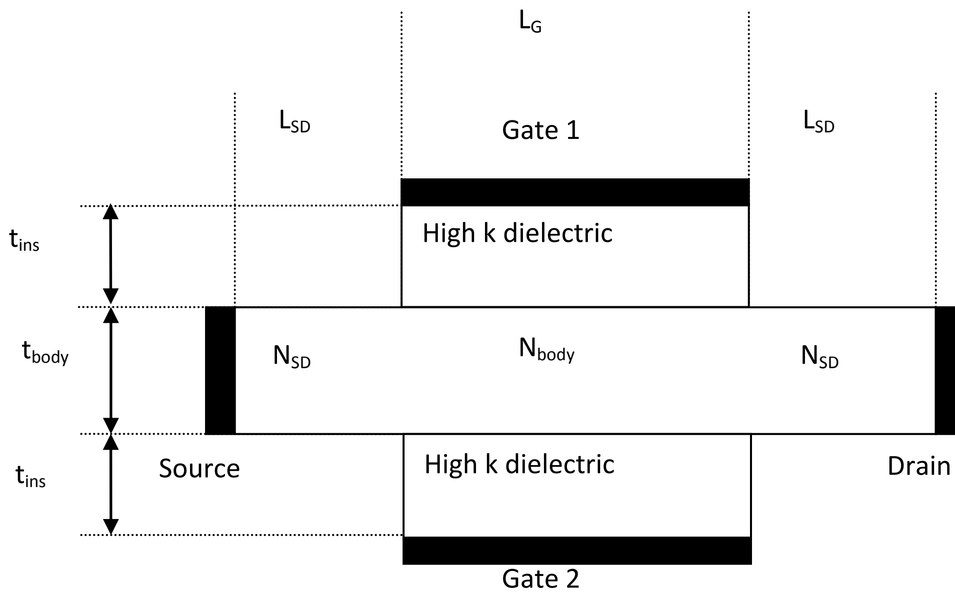


Fig. 3.2.1 Device structure used for ballistic performance analysis of InGaSb double gate nMOSFET. Various dimensions are given in Table-3.1

Table 3.1 : Parameters related to InGaSb DG-nMOSFET device simulation

InGaSbDG-nMOSFET Particulars	**Parameters**	**InGaSbDG-nMOSFET Parameter Value**
Dimension	Channel length, L_G	7nm, 10nm, 13nm,15nm
	Source and Drain length L_{SD}	5nm
	Oxide thickness, t_{ins}	1nm, EOT=1.7nm
	Body thickness, t_{body}	3nm
Doping	Source and Drain Doping, N_{SD}	10^{20} /cm^{-3}
	Body Doping , N_{Body}	0
	Junction doping profile	Step
Material Parameters	HfO_2 relative permittivity, ε_{ins}	22
	Al_2O_3 relative permittivity, ε_{ins}	9.3
	InGaSb longitudinal effective mass	0.034
	InGaSb transverse effective mass	0.034
	ΔE_C	(Al_2O_3)=2.88eV and (HfO_2) = 1.76eV
	Gate workfucntions for both gate metal (Effect of gate length)	4.7 ev
	Gate workfucntions for both gate metal (Effect of oxide material)	3.72 ev
Options	Wavefunction penetration	yes
	Valley	Gamma valley
Supply Voltage	Drain Voltage	0.5
	Gate Voltage	0-1V

CHAPTER 4

Simulation Results and Discussion

4.1 Introduction

The relentless scaling of the MOSFET's for logic node in nanometer regime faces serious short channel effects. In the present study different techniques have been proposed to find the best material and structural solution in order to have the best performance. From the material point of view, simulation and experimental results demonstrates that III-V technologies are subject to enhance the short channel effects, including larger V_{Th} roll off, larger drain induced barrier lowering and lower subthreshold slope then the silicon devices [1-4]. On the other hand, other approaches are adopted to suppress the short channel effects, among them source/drain underlap tuning is an approach to enhance the MOSFET performance. Also gate oxide scaling effect is another approach to suppress the short channel effect.

Following the theoretical analysis presented in chapter 3 the performance of the proposed InGaSb DG-nMOSFET are evaluated as a function of gate length, underlap length and also the effect of oxide on the ballistic performance is investigated of the DG-nMOSFET. To assess the performance, drain current behavior are analyzed along with other logic figure of merits like subthreshold slope(SS), threshold voltage (V_{Th}), I_{ON}, I_{OFF}, I_{ON}/I_{OFF} ratio,etc.

4.2 Ballistic performance of nanoscale InGaSb DG-nMOSFET

To meet the demand of computational power, semiconductor industries have been drastically scaling down the CMOS transistors. But the extreme reduction of channel length causes various adverse effects that results short channel effects (SCEs) and now a days it causes the approaching of Si MOSFET's to their physical limits. The possible advantages of III-V materials over Si are high ON current, minimum device delay and reduction in power supply voltage. According to ITRS roadmap [66], by 2018 III-V material will replace Si from the channel and gate length will reach to 10nm regime. Also to achieve the required control over the channel at this small length, double gate structure has been adopted. People already studied InAs materials at the channel with different dimensions [67,68,69]. For very low power application, InSb shows excellent result [11]. Also their alloy InAsSb is studied [70] and all of those show beautiful results for n-operation. But it is already established that InGaSb exhibits highest both electron and hole mobility among III-V semiconductor materials [71]. That is why for both "n" and "p" operations, InGaSb would be the best choice.

The performance of the proposed device is studied as a function of gate length considering HfO_2 as gate dielectric. The thickness of the oxide is taken into account of 10 nm and equivalent oxide thickness (EOT) is 1.77 nm. The body or channel thickness is taken as 3nm and the source and drain doping concentration is $10^{20}/cm^{-3}$ and the doping in the channel is 0. The gate metal work function is as 4.7 eV and the applied drain voltage is V_D= 0.5 V.

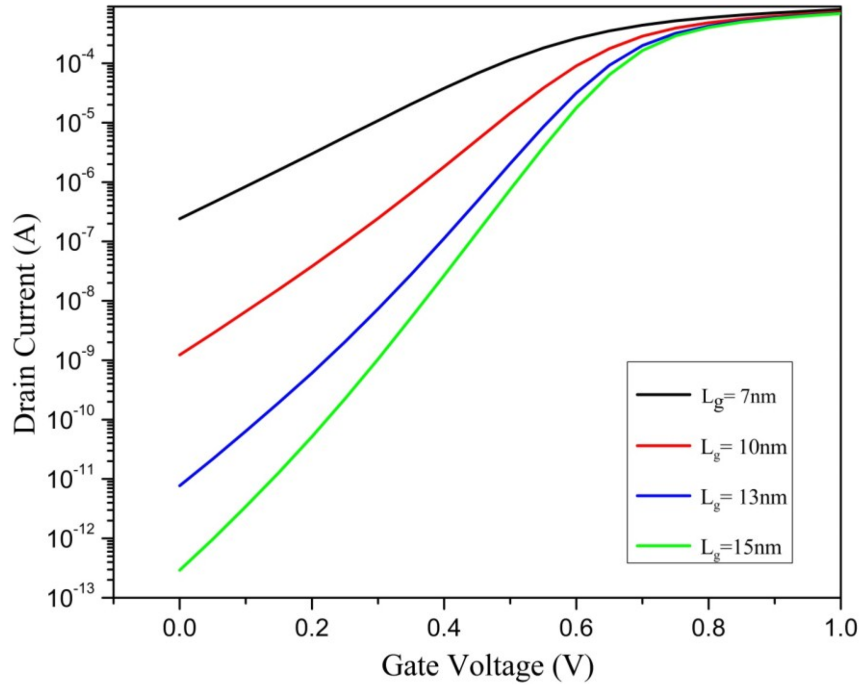


Fig. 4.1 I_D-V_G characteristics as a function of different gate lengths.

The drain current for different gate lengths are shown in the fig. 4.1. It is evident form the figure that there is negligible dependence of the "ON" current with gate length. The highest "ON" current for the gate length 7nm is evaluated 0.0008016 A/µm and and the lowest ON current is 0.000688 A/ µm the gate length 15nm. On the other hand the "OFF" current is found to have strong dependence on the gate length s. From Fig. 4.1.1 small change in the OFF current is found for the change in gate length 15 nm and 13 nm but when the gate length decreases from 13 nm the OFF current increases significantly. To understand the characteristics of the "ON" and "OFF" currents with gate length, channel position dependent energy profiles are plotted with respect to different gate length.

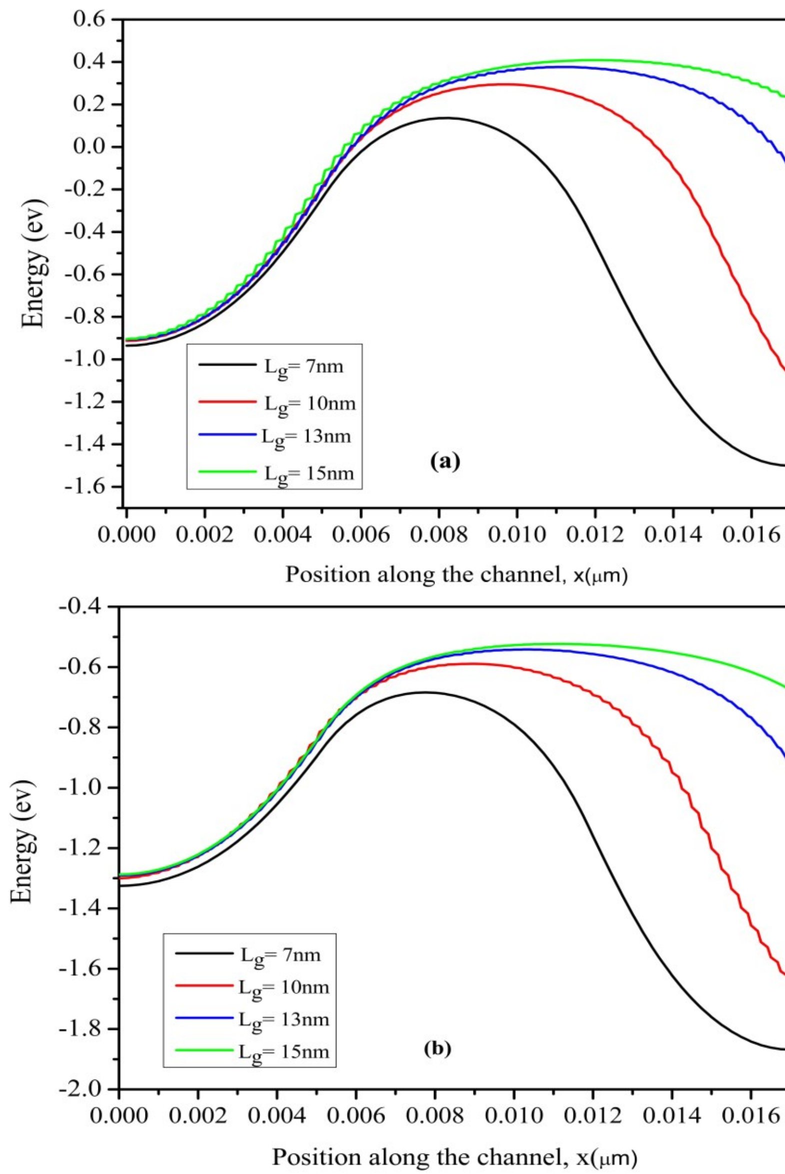
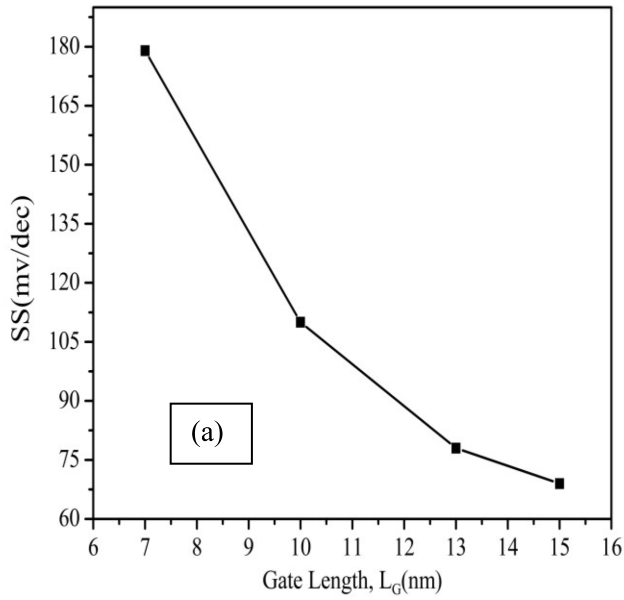


Fig.4.2 Conduction band profile for different gate length (a) OFF state condition (b) ON state condition.

The results are shown in the Fig. 4.2 (a) and 4.1.2 (b) for the "OFF" and "ON" state conditions respectively. As seen in Fig. 4.2 (a), the top of the barrier energy is almost the same for the gate lengths 10nm, 13m, and 15nm. However, for the 7nm gate length the it si found to be lower compared to other three gate lengths. The variation of the barrier heights has a strong effect on the OFF state current which clearly explains why the "OFF" state current changes with the gate length. Similarly almost the same ON current is found for the gate lengths 10nm, 13nm and 15nm due to their barrier heights shown in the Fig.4.2(b). Further a slight increase in the "ON" state current is observed for the gate length 7nm due to its energy barrier lowering compared to the gate lengths 10nm to 15nm.



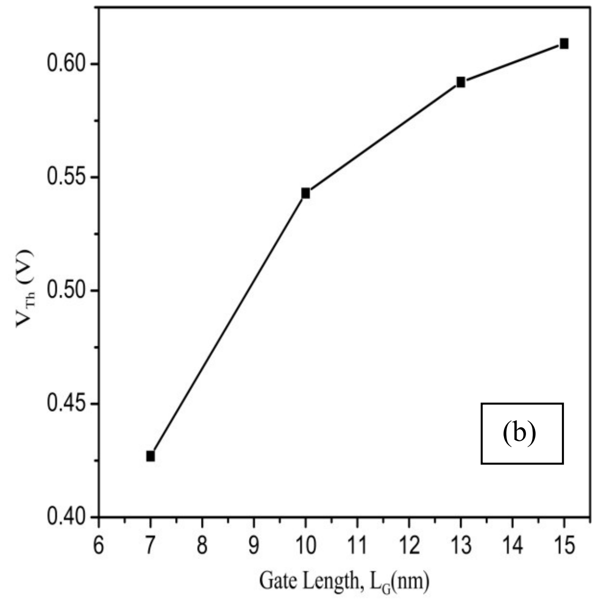
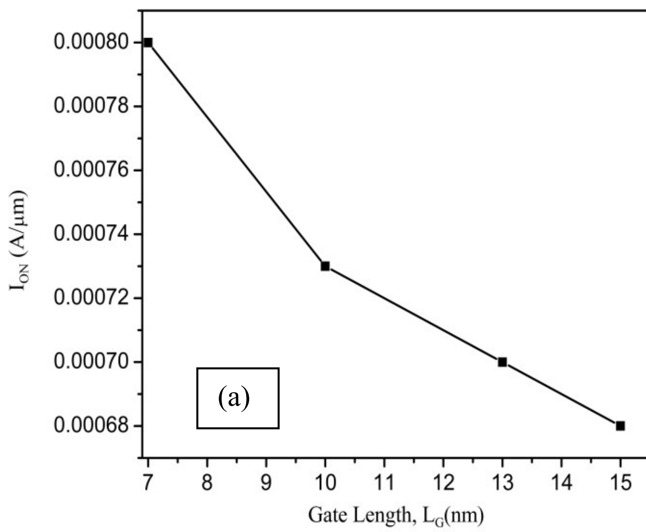
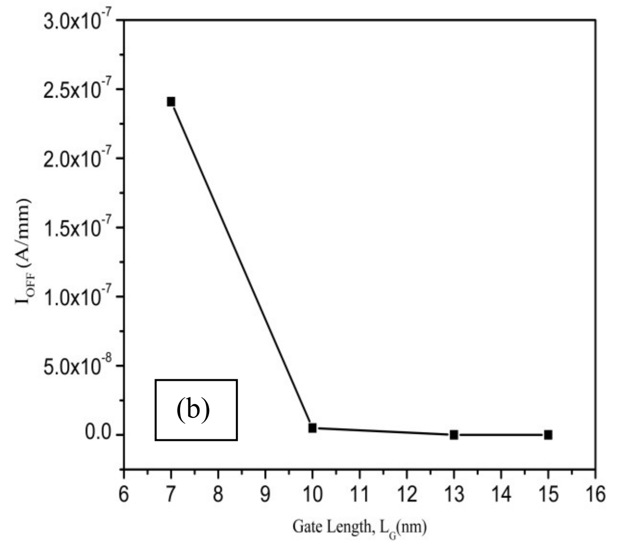


Fig. 4.3 Gate length dependent (a) Subthreshold slope (b) Threshold voltage





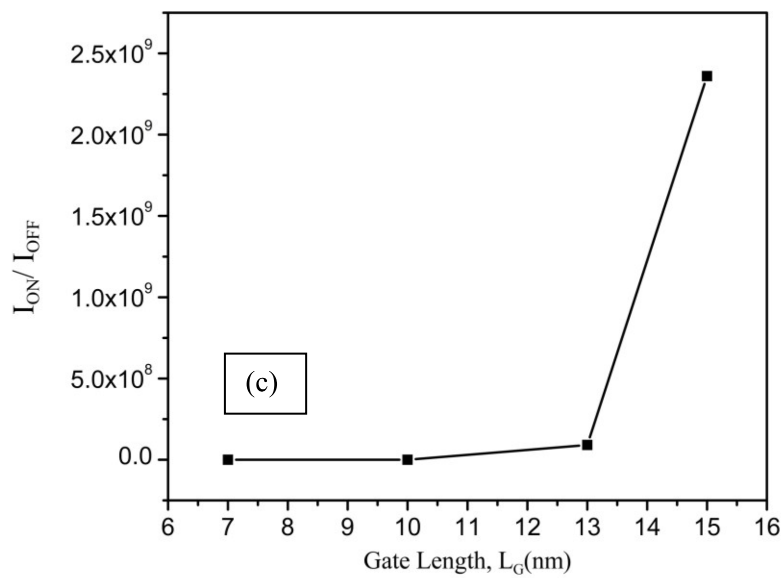


Fig. 4.4 Gate length dependent (a) ON state current (b) OFF state current and (c) I_{ON}/I_{OFF} ratio.

To explore the device performance different logic figure of merits are studied such as subthreshold slope, threshold voltage, ON current and OFF current and I_{ON}/I_{OFF} ratio. Subthreshold slope for different gate lengths is shown in the fig. 4.3 (a), where subthreshold slope is found to decrease with increasing gate length. The maximum subthreshold slope is evaluated 180 mv/decade for the gate length 7nm and it reduces to 70 mv/decade for the gate length 15nm. Threshold voltage decreases with the decrease in gate length as shown in the fig. 4.3 (b). Threshold voltage for the gate length 15nm is found 0.60V and which decreases to 0.43V for the gate length 7nm. Further the OFF current found to decrease remarkably with the increase in gate length from 7nm to 10nm. After that the OFF current almost constant with the gate length as seen in the Fig.4.4(a).The ON state current decreases from 0.008A/µm to 0.0068A/µm for the increase in gate length from 7nm to 13nm as shown in Fig. 4.4 (b).. The I_{ON}/I_{OFF} ratio is very much important to understand the device performance and found to increase with the increase in gate length in Fig. 4.4(c). The above results can be easily understood from the conduction band profiles plotted for the OFF state and ON state conditions shown in Figs. 4.2(a) and (b). In these figures the Fermi level is at 0V and the OFF state barrier energy for the gate length 7nm is slightly above the Fermi level, however, it is about 0.4V above for the gate length 15 nm. Thus, in the OFF state condition, the barrier height and width is significantly higher for the gate length 15 nm compared to the gate length 7nm. That is why the major component of the OFF state current contributed by tunneling mechanism is greatly reduced for gate length 15nm compared to the gate length 7nm, which is evident in the fig 4.4 (a). On the other hand, for the ON state condition shown in Fig. 4.2 (b) a significant change in barrier height is observed for the gate length 7nm compared to the gate lengths. The significant barrier lowering in 7nm causes the increase in

ON current. For other gate lengths the magnitude is found to be almost the same due to the energy barrier height as clearly seen in Fig.4.2 (b).

4.3 Effect of gate oxide keeping same EOT

According to ITRS roadmap III-V material may replace Si from the channel by 2018[72,66]. Among the III-V materials, InGaSb is a promising candidate for the channel of future MOSFET as it shows highest electron and hole mobilities [73]. Apart from the channel material, gate oxide of MOSFET is also a major issue. When the channel length of a device is scaled down, oxide thickness is also to be scaled to maintain the required level of C_{ox}, I_{ON} and to control the threshold voltage roll off, that is, to keep the control over the channel. Thickness of SiO_2, which is the preferred gate insulator for Si MOSFETs, is reduced over the years from 300nm for 10µm technology to 1.2nm for 65nm technology. When SiO_2 thickness is scaled down, key dielectric properties degrades and device suffers from high leakage current, dielectric breakdown, reduced channel mobility and so on [74,75]. These problems can be solved by replacing SiO_2 with a material having higher dielectric constant. High-k material can be grown physically thicker than SiO_2which has the same electrostatic effects that a thin SiO_2 of thickness equal to "Equivalent oxide thickness, EOT" has. The thick high-k dielectric offers required control over the channel with significant reduction in leakage current [75].

It is well established that to achieve the required control over the channel at small gate length targeted by ITRS, multi gate device structure is necessary. Keeping all these issues in mind, in the present research work, we studied the performance of $In_{0.3}Ga_{0.7}Sb$ double gate MOSFET for n-type operation considering two different dielectric materials Al_2O_3 and HfO_2 with same EOT of 1nm. The effects of conduction band offset and gate oxide over current-

voltage characteristics as well as related terms like subthreshold swing are investigated. Since drain current is normally calculated by quantum mechanical approaches in ballistic range, we used non equilibrium greens function (NEGF) to carry out the simulation.

In our simulation, source and drain doping concentration is taken to be $10^{20}/cm^3$, and this doping is done throughout the region marked by L_{SD} in Fig. 3.2.1. Von Neumann boundary conditions for potential are used because the ballistic carrier do not create voltage drop near source and drain contacts. Moreover, the potential in source and drain are allowed to float, because the number of carrier may reflect backward and hence the total carrier concentration in the contact can changed with bias. The body thickness of the device is 3nm. Wave function penetration in the oxide region is also taken into considaration.

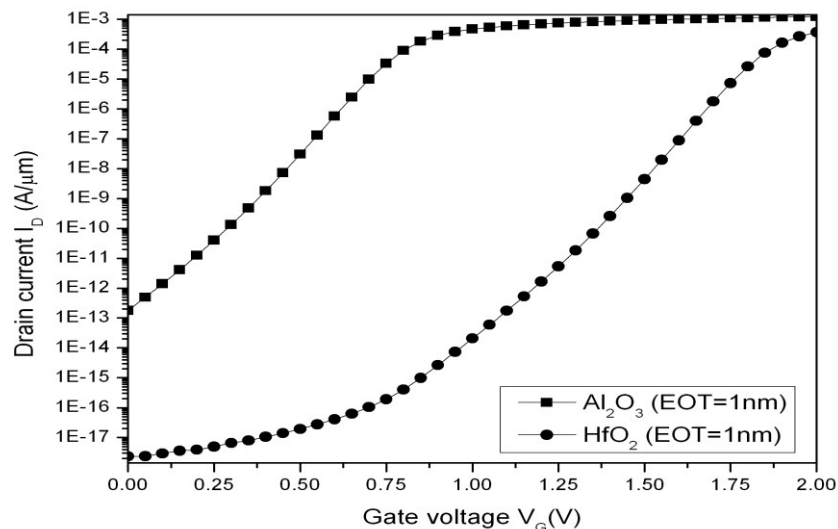
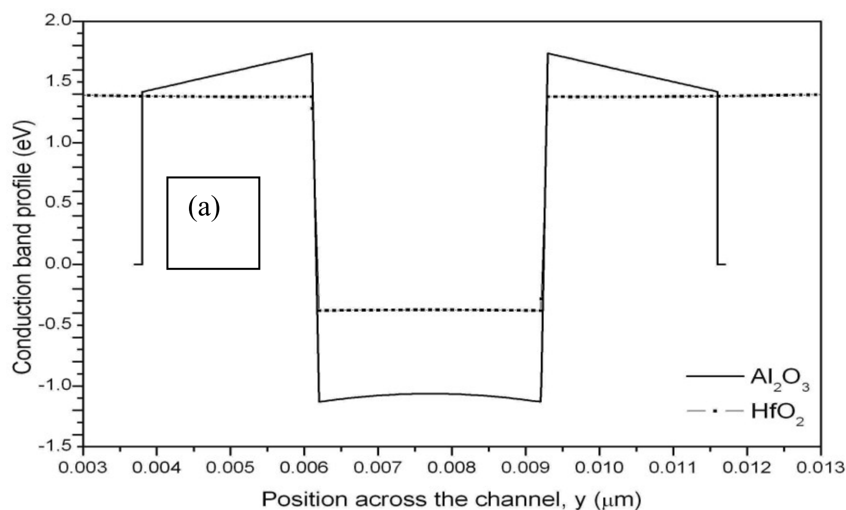


Fig. 4.5. I_D-V_G characteristics of double gate MOSFET for different gate dielectric material

Fig. 4.5 shows a comparison between the I_D-V_G characteristics plotted at V_D=0.05V for the double gate MOS device having two gate dielectrics. It is found that the drain current is higher for Al_2O_3 than that of HfO_2, although they are calculated for the same EOT. Generally it is expected that if two oxides have the same EOT, they will provide same oxide capacitance, thereby same electrostatic control over the channel and hence identical device

performance. But in case of double gate MOSFET, the deviation is found because of conduction band discontinuity (ΔE_C) between the oxide and the channel.

To understand the effect of ΔE_C, the conduction band profiles across and along the channel at V_D=0.05V and V_G=2V are estimated and plotted in Figs. 4.2.2 (a) and (b), respectively. It is clear that at very low gate bias, with keeping the gate metal work function (Φ_{metal}) fixed, lower ΔE_C implies the higher difference between the channel conduction band minima and Fermi level at the source contact, which is set at "zero". The opposite is found for high gate bias as shown in fig. 4.2.2 (a). These band positions relative to source contact fermi level implies lowest number of energy states are available for conduction if ΔE_C gets lower. Since lower number conducting states leads to lower drain current, lower conduction band offset between oxide and channel resulted lower drain current, provided that Φ_{metal} is kept fixed. For $In_{0.3}Ga_{0.7}Sb$ channel, the conduction band offset between the channel and the Al_2O_3 is 2.88eV and which is 1.76eV for HfO_2 [14]. That is why HfO_2 gives lower drain current although the results are obtained for the same EOT.



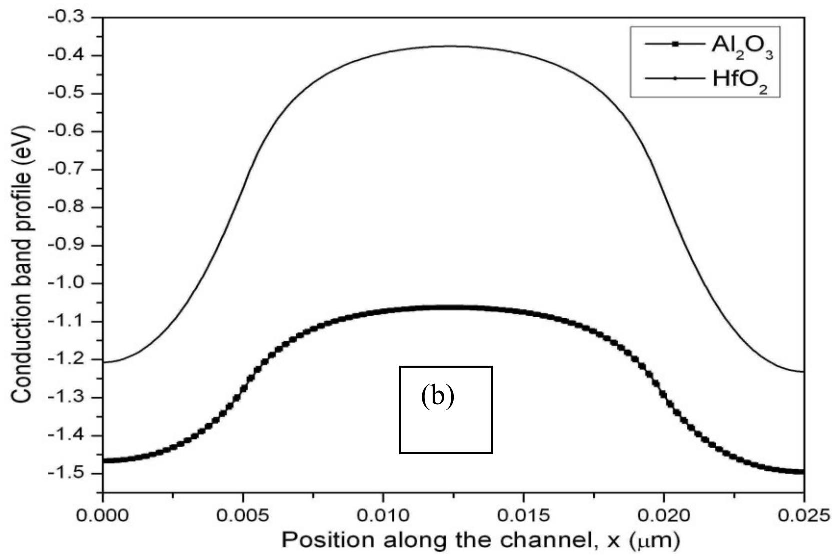


Fig. 4.6 Conduction band profile of the double gate MOSFET (a) across the channel (b) along the channel, at V_D=0.05V and V_G=2V.

It is also found in fig. 4.2.2 (b) that the conduction band profile does not start from the same energy position, for two different gate oxides. It is because in ballistic transport source and drain contacts are allowed to float to take into account any reflection of carrier during transport. Since the threshold voltage can be tuned by tuning the gate metal work function (Φ_{metal}), the I-V curve shown in Fig. 4.2.1 for the HfO_2 can be shifted to the left by reducing Φ_{metal}. If Φ_{metal} for the HfO_2 based device is chosen to be 1.12eV which is less than that of used for the Al_2O_3 device, the I-V curves will be at closest position, because $\Delta E_C(Al_2O_3)-\Delta E_C(HfO_2)=1.12$. To compare the impact of ΔE_C only, all other device parameters are kept same.

Figure-4.7 shows the change in subthreshold swing (SS) as a function of ΔE_C between the oxide and the channel. It is found that the subthreshold swing decreases with decreasing ΔE_C. Lower SS implies the sharp change of drain current with gate bias and the better electrostatic control of gate over the channel. Better control of the gate also indicates the lower influence of the drain voltage on I-V characteristic and hence suppressed SCE, like drain induced barrier lowering. Theoretically the lowest possible SS at room temperature is 60mV/dec

which means the best electrostaic control of MOSFET. In practice, SS bellow 90mV/dec is acceptable for logic application.

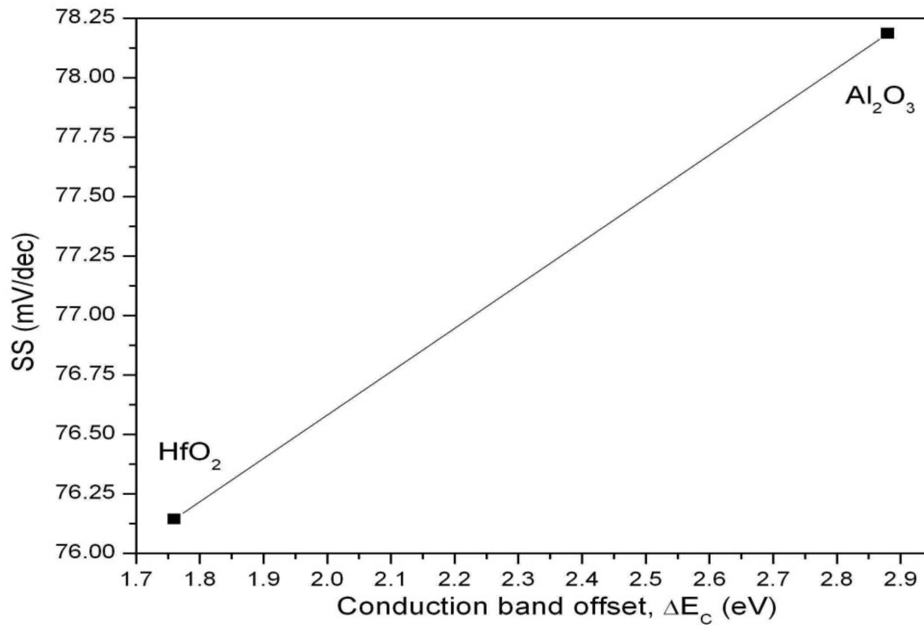


Fig. 4. 7 Variation of subthreshold swing(SS) as a function of ΔE_C

Table 4.2.1 Comparison Logic Figure of merits for two different oxide

	SS (mv/dec)	V_{Th} (V)	I_{ON} (A/µm)	I_{OFF} (A/µm)	I_{ON}/I_{OFF}
HfO_2	77.5679	1.3856	0.0020644	5.35473E-18	3.85528E+14
Al_2O_3	81.849	0.608488	0.0191982	9.26625E-13	2.07184E+10

4.4 Effect of gate underlap on ballistic performance

Performance improvement is the main objective of device scaling in last few decades. In recent years, ultralow- power consumption has become the primary design criteria for the applications of cellular phones, personal digital assistants, medical electronics, and other portable computing devices. Subthreshold circuits, which use a supply voltage (V_{dd}) lower than the threshold voltage (V_{th}) of transistor can be used for the applications required ultralow-power with low to medium (ten to hundreds of megahertz) frequency of operation. Double-gate MOS (DGMOS) transistors are suitable for subthreshold operation due to their near ideal subthreshold slope and negligible junction capacitance . Furthermore, unlike superthreshold operation, the intrinsic capacitance of DGMOS operated in the subthreshold region is also negligible and is very weakly dependent on the channel length . Hence, the effective gate capacitance in the subthreshold region is dominated by the parasitic capacitances. So it is already well established that tuning of gate underlap changes the parasitic capacitance and thereby changes the device figure of merits. Here we study the effect of gate underlap on the performance of InGaSb DG- MOSFET to assess the characteristics of drain current, subthreshold slope, ON and OFF currents.

Drain current characteristics for three different underlap are shown for two gate lengths in Fig. 4.8 It is found that the drain current depends on the underlap length as well as gate length. The value of drain current varies from 0.0028 A/µm to 0.0024 A/µm for the variation of the gate underlap from 0 nm to 4nm when gate length is 10nm.

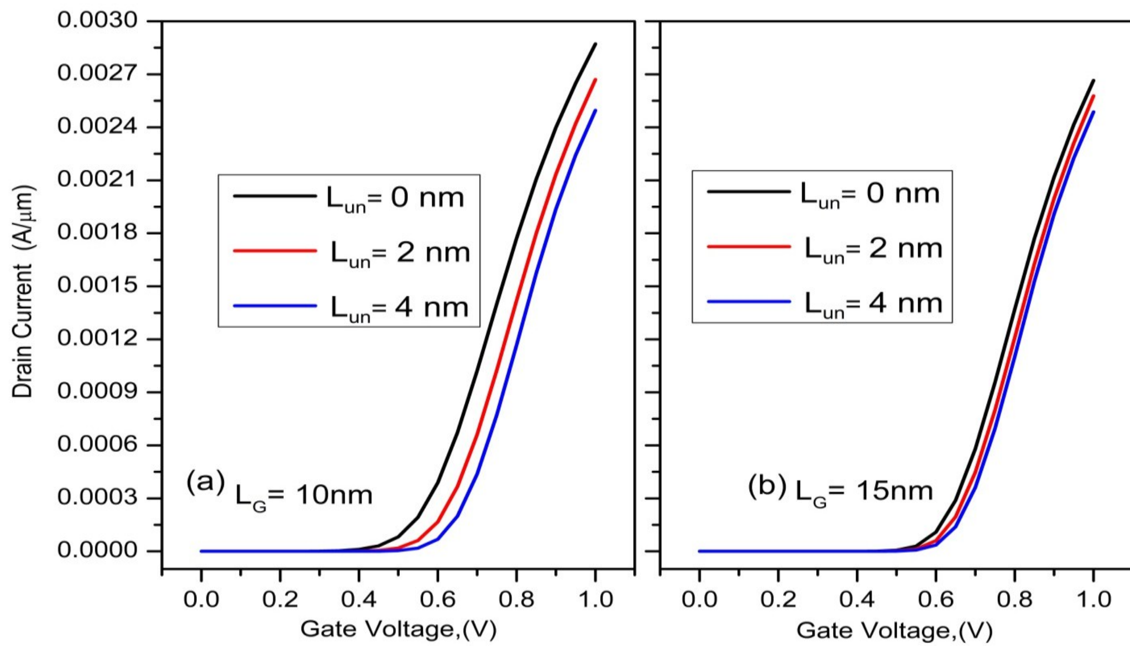


Fig. 4.8 Drain current behavior for three different underlap length (a) for gate length 10nm and (b) for gate length 15nm.

On the other hand, the variation of drain current is found to insignificant for the same range of gate underlap when the gate length is 15nm . To understand the threshold voltage with underlap length Fig. 4.3.2 (a) is plotted

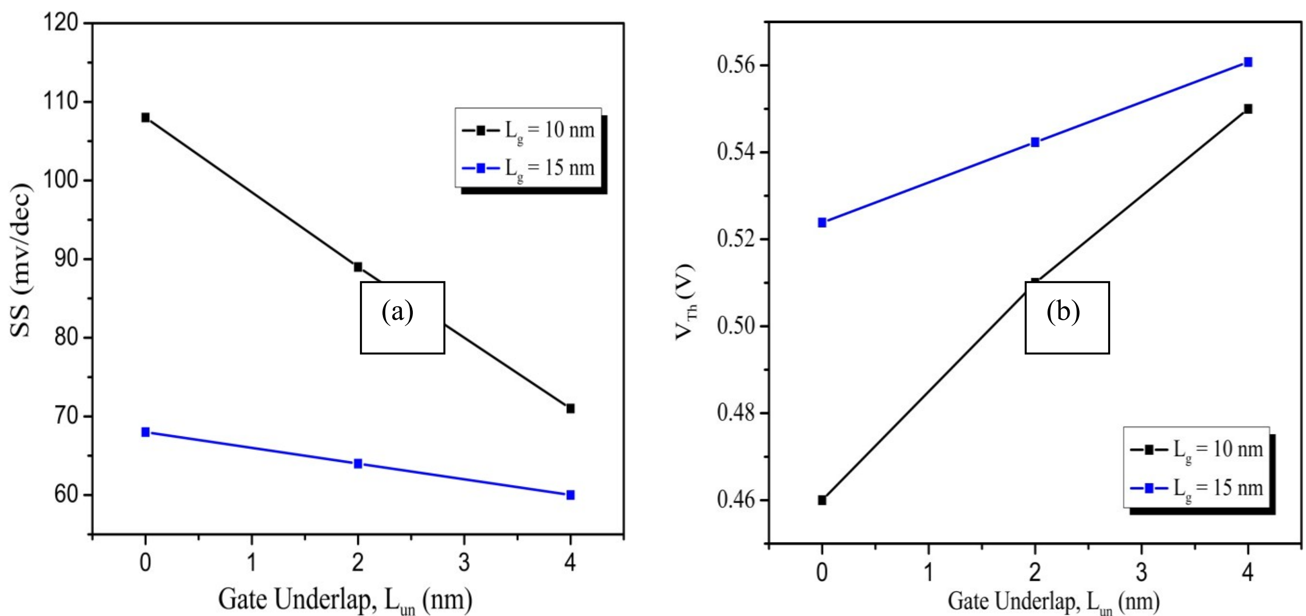


Fig. 4.9 (a) Threshold voltage and (b) Subthreshold slope as a function of gate underlap for two different gate length 10 and 15nm.

From Fig. 4.9 (a) it is evident that the change in threshold voltage for the gate length 10nm is significantly higher compared to the gate length 15nm for different underlap length. For the gate length 10nm, threshold voltage varies from0.46V to 0.54V for the variation of underlap length 0 to 4nm. On the other hand, the same threshold voltage is found to vary form 0.52V to 0.56V for the gate length 15nm. Subthreshold behavior for the DG-nMOSFET has been shown in the Fig.4.9 (b). Subthreshold slope (SS) changes significantly for the 10 nm gate length for three different gate underlap lengths. The results are found to vary from 110 mV/decade to 70 mV/decade. For the gate length 15nm the changes in SS is small and ideal SS has been achieved for the gate underlap length 4nm.

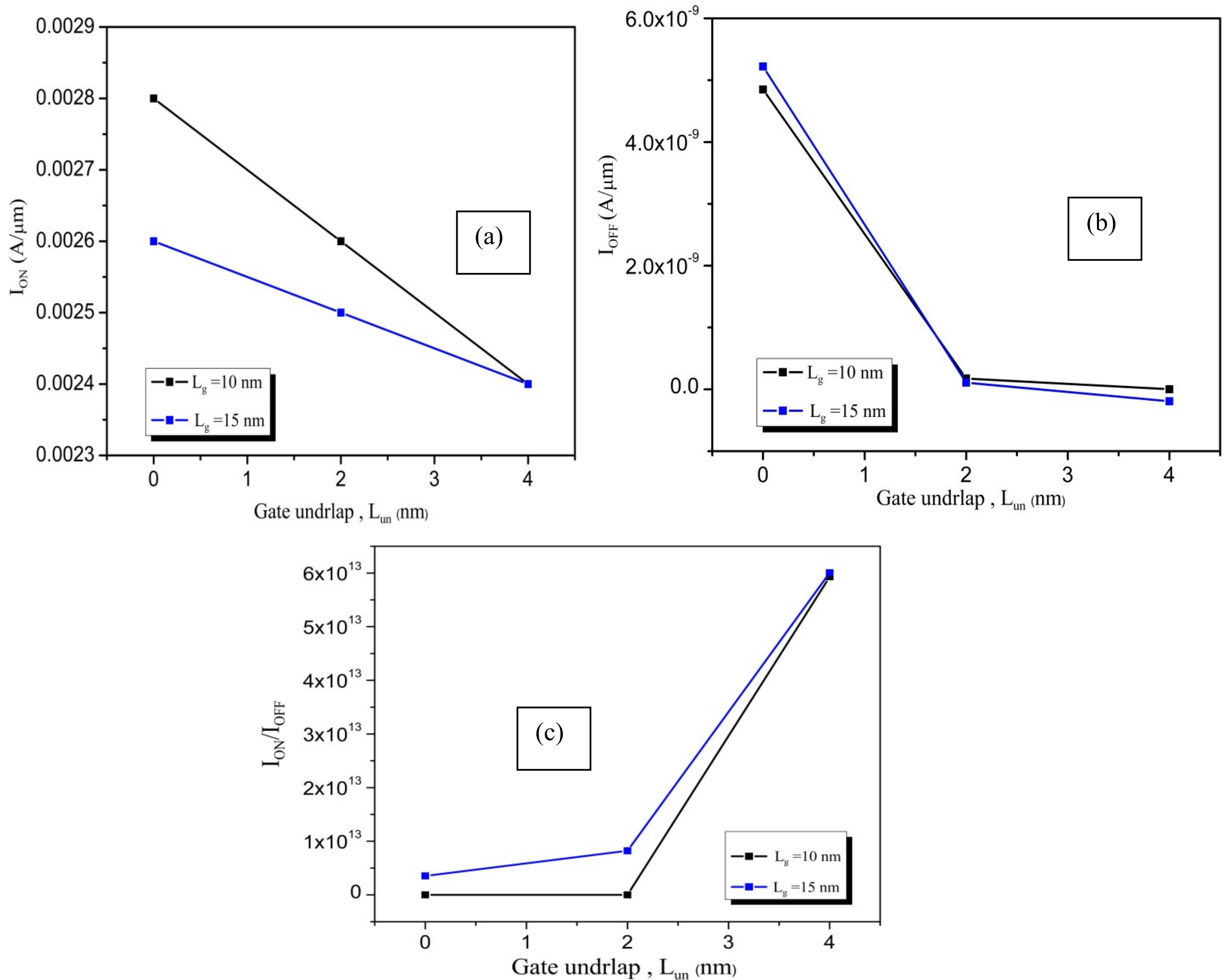


Fig. 4.10 Gate length dependent (a) ON current. (b) OFF current and (c) I_{ON}/I_{OFF} ratio characteristics for different gate underlap length

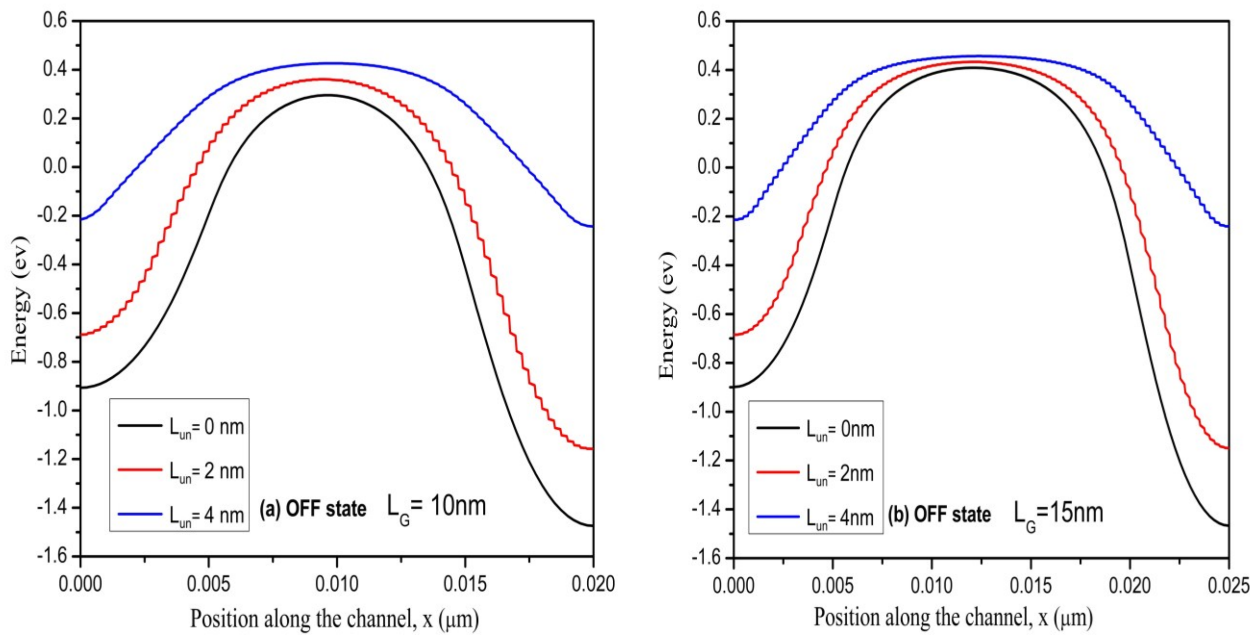


Fig. 4.11 Channel position dependent OFF state conduction band profile obtained for gate lengths gate length (a) 10nm and (b) 15nm as a function different gate underlap lengths.

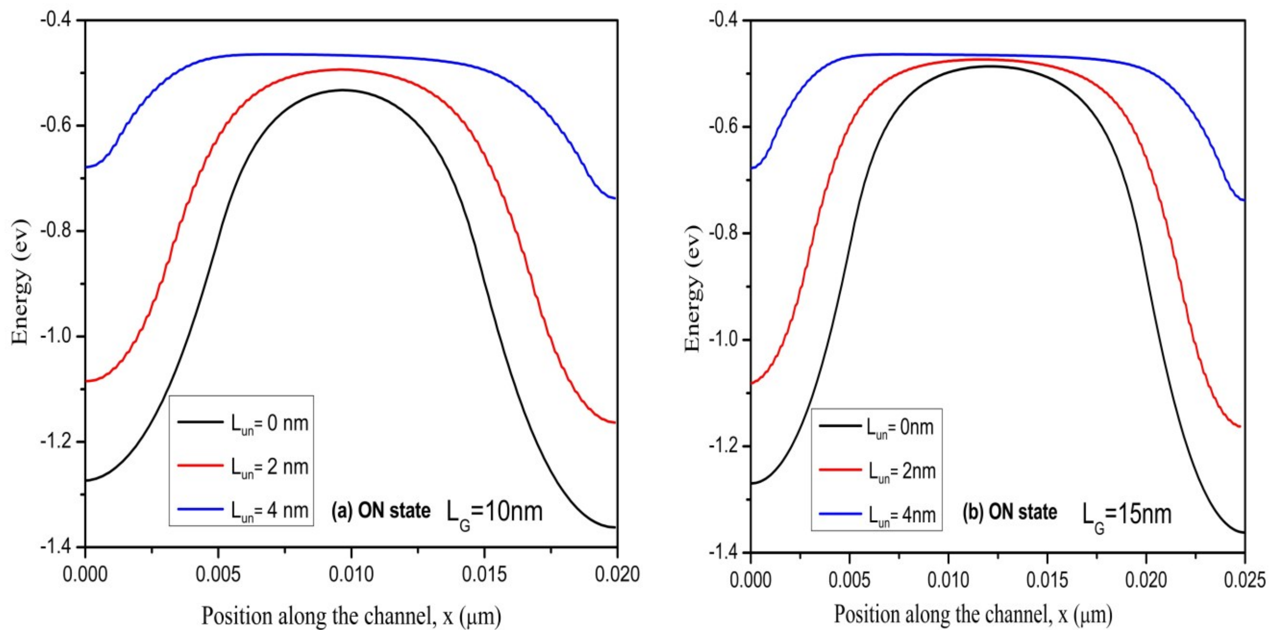


Fig. 4.12 Channel position dependent ON state conduction band profile obtained for gate lengths (a) 10nm and (b) 15nm as a function of different gate underlap lengths.

Other figure of merits like ON current , OFF current and I_{ON}/I_{OFF} ratio characteristics are studied with respect to gate underlap length as a function of different gate lengths. The results are shown in the Fig. 4.10 (a). It is evident from the Fig.4.10(a) that the highest ON current 0.0028 A/µm is obtained for the gate length 10nm when gate underlap length is 0nm. On the other hand for the gate length 15nm, the ON current is evaluated to 0.0026 A/ µm for the same gate underlap. For both the gate length the ON current is found to decrease and becomes 0.0024 A/ µm for increasing the gate underlap length 4nm.The OFF current behavior is almost the same for the gate lengths 10nm and 15nm with respect to gate underlap length as seen in Fig. 4.10 (b). Figure. 4.10 (c) shows the I_{ON}/I_{OFF} ratio, and for both of the gate lengths, the ratio found to increase with increasing gate underlap length. From the above results it is clear that the short channel effect can be controlled by gate underlap engineering. To explain the results presented in Fig. 4.10, channel position dependent conduction band profiles are plotted in Fig. 4.11 and Fig. 4.12. The figures are plotted in the same scale for the comparison. From Fig. 4.11 (a) and (b) and Fig. 4.12 (a) and (b) it is found that the barrier height modulation due to change in gate underlap influences the OFF current as well as ON current behavior and also the SS and V_{Th}. The barrier height modulation causes the increase in effective channel length . For the large underlap length ON cureent is inversely proportional to the channel length and reduces linearly with the increase in gate underlap length

CHAPTER 5

Concluding Remarks

5.1 Conclusion

In the present research, simulation of quantum transport in InGaSb DG-nMOSFET has been performed using non equilibrium Green's function (NEGF) approach in effective mass approximation. The different structural variants of the device is simulated in mode space approach of NEGF formalism. Mode space approach is suitable for the ultra-thin body DG-MOSFET's and computational burden associated with this approach is affordable compared to Real space approach. In this study the simulation is performed as a function of gate length, underlap length and effect of oxide thickness.

For the InGaSb DG-nMOSFET drain current characteristics and different logic figure of merits are analyzed for different gate length. As gate length decreases the OFF state current is increased along with the increase in the ON current and the maximum ON current is obtained for the L_G= 7nm. From the result it is evident that ON currents for different gate lengths are almost the same order but the OFF currents significantly reduce with the increasing gate length. For the variation of gate length from 7nm to 15 nm, on current reduces from 0.00080 A/µm to 0.00068 A/ µm. The scaling behavior of subthreshold slope has also been observed from the simulation. Subthreshold slope decreases and I_{ON}/I_{OFF} ratio increases with increasing gate length. Device performance degrades with the shorter gate due to short channel effects. The physics behind these phenomena can be understood from the conduction band profile in the OFF state and ON state. It is evident from the conduction band profile that, the OFF state

barrier is significantly higher for the gate lengths 15 nm, 13nm and 10nm compared to 7nm. Therefore the tunneling current (major component of OFF state current) for the gate length 15 nm, 13nm and 10 nm is greatly reduced compared to the gate length 7nm, while in the ON state condition there is a significant difference between the ON current obtained for the gate length 15nm, 13nm, 10nm and 7nm. This implied that 7nm gate length gives higher ON current then other three different gate lengths.

. Impact of gate underlap for DG-MOSFET has also been studied in this study. Three different gate underlap lengths are applied and drain current characteristics along with the logic figure of merits are determined. It is evident from the result that increasing the gate underlap length decreases the subthreshold slope that causes to increase threshold voltage. On the other hand ON current and OFF current show similar behavior for higher unnderlap length.

It is well established that to achieve the required control over the channel at the small length targeted by ITRS, multi gate device structure is necessary. we analyzed the influence of ΔE_C between the gate oxide and the channel on the performance of $In_{0.3}Ga_{0.7}Sb$ nanoscale double gate MOSFET. We took 15nm channel with Al_2O_3 and HfO_2 gate oxides having 1nm of equivalent oxide thickness and simulated the performance of the device by NEGF method. . We found that the drain current increases with increase in ΔE_C if the work function of the gate remains unchanged. Also the threshold voltage (V_T) decreases and SS increases with increasing ΔE_C. SS of the device under study goes bellow 90mV/dec for both case and it is a rule of thumb that SS bellow 90mv/dec is acceptable for logic application. Although we get lower drain current for HfO_2 because of its lower ΔE_C with InGaSb, it can be controlled by

tuning the gate metal work function. That is why the overall performance of HfO_2 is better since it gives lower SS

5.2 Suggestion for Future Work

Short channel effect is the most prominent factor as gate length decreases. In this work we have considered the smallest gate length 7nm and at this length the performance we reported here is not properly applicable for the logic circuits. The International Technology Roadmap for Semiconductors (ITRS) predicts the MOSFET channel length to be less then 8nm in ten years. It is reported in the literature that heavier conduction band effective mass can achived by channel orientation engineering and which will help to downward scaling of the gate length. Also different structural modification can mitigate the short channel effect in the DGMOSFET's like dual material gate , tri material gate and applying this approach can be studied further.

References

[1]. J. A. del Alamo, "Nanometer-scale electronics with III–V compound semiconductors," *Nature*, vol.479, no.12, pp. 317–323, Nov. 2011.

[2]. M. Alon Horowitz, E. Patil, D. Naffziger, S. Rajesh Kumar, K. Bernstein "Scaling, power, and the future of CMOS," *IEDM Technical Digest*, pp.7-15, Dec. 2005.

[3]. T. Ghani, "Challenges and Innovations inNano-CMOS Transistor Scaling, *http://microlab.berkeley.edu/text/seminars/slides/TahirGhani.pdf*. Nov. 2009.

[4]. Bohr M, "The evolution of scaling from the homogeneous era to the heterogeneous era," *International Electron Devices Meeting (IEDM),* pp.1.1.1-1.1.6, Dec. 2011.

[5]. Intel Labs at the International Solid-State Circuits Conference (ISSCC) 2012, *http://download.intel.com/newsroom/kits/isscc/2012/pdfs/ISSCC-IL-Press-Overview.pdf*.

[6]. A. Nainani, T. Irisawa , Ze Yuan, Yun Sun, T. Krishnomohan, M. Reason, B.R. Bennett, J.B. Boos, M.G. Ancona, Y. Nishi, K.C. Saraswat,"Development of high-kdielectric for antimonides and a sub 350°C III–V pMOSFET outperforming Germanium," *International Electron Devices Meeting*, pp.6.4.1-6.4.4, Dec. 2010.

[7].B.R. Bennett, N. A. Papanicolaou, M. G. Ancona, J. G. Champlin, Park D. Kruppa, W. Weaver, B.D. Bass, B.V Shanabrook, "Sb-based n- and p-channel HFETs for high-speed, low-power applications," *Device Research Conference,* , pp.159-162, June 2009.

[8]. Brian R. Bennett, Mario G. Ancona, James G. Champlin, Nicolas A. Papanicolau, J. Brad Boss " Demonstration of high-mobility electron and hole transport in a single InGaSb well for complementary circuits," *Journal of Crystal Growth,* vol. 312, pp. 37-40, 2009.

[9].M.Yokoyama,S.H.Kim,R.Zhang,N.Taoka,Y.Urabe,T.Maeda,H.Takagi,T.Yasuda,H.Yamada, O.Ichikawa,N.Fukuhara,M.Hata,M.Sugiyama,Y.Nakano,M.TakenakaandS.Takagi,"CMOS integration of InGaAs nMOSFET sand Ge pMOSFETs with self---aligned Ni---based metal S/D using direct wafer bonding," *Symposium on VLSI Technology* ,pp.60-61,June2011.

[10]. H. Becke, R. Hall, and J. White, "Gallium Arsenide MOS Transistors," *Solid-State Electronics*, vol. 8, pp. 813-816, 1965.

[11]. Y. Xuan, Y. Q. Wu, and P. D. Ye, "High-performance inversion-type enhancement-mode InGaAs MOSFET with maximum drain current exceeding 1A/mm," *IEEE Electron Device Letters*, vol. 29, pp. 294-296, 2008.

[12]. M. K. Hudait, G. Dewey, S. Datta, J. M. Fastenau, J. Kavalieros, W. K. Liu, D. Lubyshev, R. Pillarisetty, W. Rachmady, M. Radosavljevic, T. Rakshit, and R. Chau, "Heterogeneous integration of enhancement mode $In_{0.7}$ Ga $_{0.3}$As quantum well transistor on silicon substrate using thin (<= 2 gm) composite buffer architecture for high-speed and low-voltage (0.5V) logic applications," *International Electron Devices Meeting,* pp. 625-628, 2007.

[13] D. H. Kim and J. A. del Alamo, "Scaling behavior of $In_{0.7}Ga_{0.3}As$ HEMTs for logic," *International Electron Devices Meeting,* pp. 587-590, 2006.

[14] D. H. Kim and J. A. del Alamo, "Logic performance of 40 nm InAs HEMT's," *International International Electron Devices Meeting,* pp. 629-632, 2007.

[15] D. H. Kim and J. A. del Alamo, "30 nm E-mode InAs PHEMTs for THz and Future Logic Applications," *International Electron Devices Meeting*, pp. 719-722, 2008.

[16] G. Dewey, R. Kotlyar, R. Pillarisetty, M. Radosavljevic, T. Rakshit, H. Then, and R. Chau, "Logic Performance Evaluation and Transport Physics of Schottky-Gate III-V Compound Semiconductor Quantum Well Field Effect Transistors for low power Supply Voltages (V-CC) Ranging from 0.5V to 1.0V," *International Electron Devices Meeting*, pp. 452-455. 2009.

[17] D. H. Kim and J. A. del Alamo, "Scalability of sub-100 nm thin-channel InAs PHEMT's," *International Conference on Indium Phosphide & Related Materials* pp. 132-135, 2009.

[18] N. Waldron, D. H. Kim, and J. A. del Alamo, "A Self-Aligned InGaAs HEMT Architecture for Logic Applications," *IEEE Transactions on Electron Devices*, vol. 57, pp. 297-304, 2010.

[19] M. Radosavljevic, B. Chu-Kung, S. Corcoran, G. Dewey, M. K. Hudait, J. M. Fastenau, J. Kavalieros, W. K. Liu, D. Lubyshev, M. Metz, K. Millard, N.Mukherjee, W. Rachmady, U. Shah, and R. Chau, "Advanced High-K Gate Dielectric for High-Performance Short-Channel $In_{0.7}Ga_{0.3}As$ Quantum Well Field Effect Transistors on Silicon Substrate for Low Power Logic Applications," *International Electron Devices Meeting*, pp. 292-295, 2009.

[20] G. Dewey, M. K. Hudait, K. Lee, R. Pillarisetty, W. Rachmady, M. Radosavljevic, T. Rakshit, and R. Chau, "Carrier transport in high-mobility III-V quantum-well transistors and performance impact for high-speed low-power logic applications," *IEEE Electron Device Letters*, vol. 29, pp. 1094-1097, 2008.

[21]. D. H. Kim and J. A. del Alamo, "Scalability of sub-100 nm thin-channel InAs PHEMT's," *International Conference on Indium Phosphide & Related Material,* pp. 132-135, 2009.

[22] M. Radosavljevic, T. Ashley, A. Andreev, S. D. Coomber, G. Dewey, and M. T. Emeny, "High-Performance 40nm Gate Length InSb P-Channel Compressively Strained Quantum Well Field Effect Transistors for Low-Power (V_{cc} =0.5V) Logic Applications," *International Electron Devices Meeting*, pp.727-730, 2008.

[23]. Annesh Nainani, “High Performance III-V p-MOSFET” Ph. D Dissertation, Stanford University, California, May 2011.

[24]. S. Takagi, T. Irisawa, T. Tezuka, T. Numata, S. Nakaharai, N. Hirashita, Y. Moriyama, K. Usuda, E. Toyodas, D. Dissanayake, M. Schichijo, R. Nakane, S. Sugahara, M. Takenaka, N. Sugiyama, “ Carrier transport enhanced channel CMOS for improved power consumption and performance,” *IEEE Transaction on Electron Devices,* Vol.52, pp.21, 2008.

[25]. A. Rahman, G. Klimeck, M.S.Lundstrom,“Novel channel materials for ballistic nanoscale MOSFET bandstructure effects,” *International Electron Devices Meeting*, pp.604, 2005.

[26]. J.M. Jancu, R. Scholz, F. Beltram, F. Bassani, “Empirical spds* tight binding calculation for cubic semiconductors: General method and material parameters,” *Physical Review B*, vol. 57, pp. 6493, 1998.

[27]. S. J. Lee, H. S. Chung, K. Nahm, C. K. Kim, “Band structure of ternary compound semiconductors using a modified tight-binding method,” *Physical Review B*, vol. 42, pp. 1452, 1990.

[28] S. Lee, F. Oyafuso, P. von Allmen, G. Klimeck, “Boundary conditions for the electronic structure of finite-extent embedded semiconductor nanostructures,” *Physical Review B*, vol. 69, pp.045216, 2004.

[29]A. Rahman, G. Jing, S. Datta, M. S. Lundstrom, “Theory of ballistic nano transistors,” *IEEE Transactions on Electron Devices*, vol. 50, pp.1853, 2003.

[30] J.E. Lilienfeld, “Method and apparatus for controlling electric currents,” U.S. Patent 1745175, 1930.

[31] Bo Lojek, History of Semiconductor Engineering, Springer-Verlag Berlin Heidelberg 2007.

[32] D.Kang and M. M. Atalla, “ Silicon-silicon dioxide field induced surface devices,” IRE Solid-State Device Research Conference., Pittsburgh, PA, June, 1960.

[33].Mrk Bohr, Kaizad Mistry and Steve Smith, Intel Demonstrates High-K Metal Gate Transistor Breakthrough on 45nm Microprocessors. Available: well.ee.nthu.edu.tw/ycking/course/Indus07/20070418_Intel_HighK.pdf

[34]. Thomas Skotinicki, James A. Hutchby, Tsua Jae King, H S Philip Wong and Fredric Boeuf, “ The End of CMOS Scaling : Toward the introduction of new material and structural changes to improve MOSFET performance,” *IEEE Circuits & Devices Magazine,* vol. 21 , pp16-26. 2005.

[35]. Bansa. A, Paul. B.C. and Roy K., “Modeling and Optimization of fringe capacitance of nanoscale DGMOS devices,” *IEEE Transaction on Electron Devices,* vol.52, pp.256-262, 2005.

[36]. K. Kim, K. K. Das, R. V. Joshi and C. T. Chuang," Leakage Power Analysis of 25-nm Double Gate CMOS Devices and Circuits," *IEEE Transaction on Electron Devices,* vol.52, , pp. 980-986, 2005.

[37]. K. Lee, C.H. Lee, D. Park and Y. K. Choi ,"A Study of Negative Bias Temperature Instability of SOI and Body Tied FinFET's", *IEEE Electron Devices Letters,* Vol. 26, pp.326-328, May 2005.

[38]. X. Wu, P.C.H. Chan, and M. Chan," Impacts of Nonrectangular Fin Cross Section On the Electrical Characteristics of FinFET," *IEEE Transaction on Electron Devices,* Vol. 52, pp.63-68,2005.

[39]. S. Xiong and J. Boker, "Structural Optimization of SUTBDG Devices for Low Power Applications" *IEEE Transaction on Electron Devices*, Vol. 52, No.3, pp.360-366, 2005.

[40]. C. Yin and P.C.H. Chan, " Investigation of the Source/Drain Asymmetric Effects due to Gate Misalignment in Planar Double-Gate MOSFET's," *IEEE Transaction on Electron Devices,* Vol. 52, pp. 85-90. 2005.

[41]. C. C. Hu, Modern Semiconductor Devices for Integrated Circuits. Pearson , 2009.

[42]. D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H.-S. P. Wong, "Device scalinglimits of si mosfets and their application dependencies," *Proceedings of the IEEE*, vol. 89, no. 3, pp.259–288, 2001.

[43]. S. Dhar, M. Pattanaik, and P. Rajaram, "Advancement in nanoscale cmos device design en route to ultra-low-power applications," *VLSI Design*, vol. 2011, p. 2, 2011.

[44]. 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. Wann, S. J. Wind et al , "Cmos scaling into the nanometer regime," *Proceedings of the IEEE*, vol. 85, no. 4, pp. 486–504, 1997.

[45]. David J. Frank, Robert H. Dennard, Edward Nowack and Paul M, "Device Scaling Limits of Si MOSFET's and their application dependencies" *Proceedings of the IEEE*, Vol.89, 2001.

[46]. V. Sverdlov, H. Kosina and S. Selberhrr, "Modeling current transport in ultra scaled field effect transistors," *Microelectronics Reliability*, Vol. 47, pp.11-19, 2007.

[47]. P.C. Martin and J . Schwinger, "Theory of many particle system" Physical Review, vol. 115, pp.1342, 1959.

[48]. Jadanoff, L.P and Baym, G. "Quantum Statistical Mechanics" Frontiers in Physics Lecture Note Series, Benjamin/Cummings, 1962.

[49]. L.V Kyldysh, “Diagram technique for non-equilibrium processes” *Journal of Experimental and Theoretical Physics* , vol.47, pp. 1515,1965.

[50] D.C Langreth, “In Linear and Non- Linear Electron Transport in Solids” *NATO Advanced Study Institute Series B*, Vol.17, p.3 Plenum, New York, 1976.

[51]. P. Danielewicz, “Quantum Theory of non-equilibrium processes” *Annals of Physics*. Vol.152,pp.239-304,1984.

[52]. J. Rammer, and H. Smith, “Quantum field-theoretical methods in transport theory of metals” *Rev. Mod. Phys*. Vol.58.pp.323, 1986.

[53]. G. D. Mahan, “Quantum transport equation for electric and magnetic fields,” *Physics Reports*. Vol.145, pp.251,1987.

[54]. Furrukh S. Khan, John H. Davies, and John W. Wilkins. “Quantum transport equations for high electric field” *Phys. Rev. B*, vol.36,pp. 2578, 1987.

[55]. Antti-Pekka Jauho, Ned S. Wingreen, and Yigal Meir, “Time-dependent transport in interacting and non-interacting resonant tunneling systems” *Phys. Rev. B*, vol.50, pp. 5528, 1994.

[56]. H. Haung and A.P. Jauho, Quantum Kinetics in Transport and Optics of Semiconductors, Springer-Verlag Berlin Hidelbarg, 1996.

[57]. Supriyo Datta, Electronic Trasnport in Mesoscopic systems, Cambridge University Press, Cambridge, UK, 1995.

[57]. S. Datta, “Nanoscale device modeling the Green’s function method,” *Superlattices and Microstructures*, Vol.28, pp. 253-278, 2000.

[59]. S. Datta, Quantum Transport Atom to Transistor, Cambridge University Press, 2005.

[60]. M. Anantaram, M. Lundstorm, D. Nikonov, “Modeling of Nnanoscale Devices,” *Proceedings of the IEEE,* vol. 96, pp. 1511-1550, 2008.

[61]. Z.Ren, R. Venugopal, S.Goasguen, S. Datta, and M.S. Lundstorm, “NanoMOS 2.5: A Two-dimensional simulator for quantum transport in Double-gate MOSFET’s,” *IEEE Transaction on Electron Devices*, Vol.50, pp. 194-1925, Sep. 2003.

[62]. R.Venugopal, Z. Ren, S. Datta, and M. Lundstorm, “Simulating Quantum Transport in nanoscale MOSFET’s : real versus mode space approaches,” *Journal of Applied Physics*, Vol. 92, pp.3730-3739,2002.

[63]. Z. Ren, “Nanoscale MOSFET’s : physics, simulation and design” Ph. D Dissertation, Perdue University, West Lafayette, October 2001.

[64]. J. Wang, “Device physics and simulation of silicon nanowire transistors” Ph. D Dissertation, Perdue University, West Lafayette, August 2005.

[65]. D. Ferry and S. Goodnick, Transport in Nanostructures, Cambridge University Press, Cambridge, UK, 1997.

[66]. The international technology roadmap for semiconductiors, 2012 [availble] online:http://www.itrs.net/Links/2012ITRS/Home2012.htm

[67]. T. Tanaka, K. Tomioka, S. Hara, J. Motohisa, E. Sano, and T. Fukui, “Vertical Surrounding Gate Transistors Using Single InAs Nanowires Grown on Si Substrates”, *Applied Physics Express*, vol. 30,pp. 025003, 2010.

[68]. Q.T. Do, K. Blekker, I. Reglin, W. Prost, and E. J. Tegude, “High Transconductance MISFET With a Single InAs Nanowire Cannel”, *IEEE Electron Devices Letters* , vol. 28, pp.682-684, 2007.

[69]. M. Egard, S. Johannson, A.-C. Johansson, K. M. Persson, A. W. Dey, B. M. Borg, C. Thelander, L.-E. Wernersson, and E. Lind, “Vertical InAs Nanowire Wrap Gate Transistors with ft > 7 GHz and fmax > 20GHz”, *Nano Letters*, vol. 10, pp.809, 2010.

[70]. Ashkar Ali, Himanshu Madan, Ashish Agrawal, Israel Ramirez, Rajiv Misra, J. Brad Boos, Brian R. Bennett, Jeff Lindemuth , and Suman Datta, “Enhancement-Mode Antimonide Quantum-WellMOSFETs With High Electron Mobility and Gigahertz Small-Signal Switching Performance” *IEEE Electron Devices Letters*, vol. 32, pp.1689-1691, 2011.

[71]. Brian R. Bennett, Richard Magno, J. BradBoos, Walter Kruppa, Mario G. Ancona, “Antimonide-based compound semiconductors for electronic devices: A review”, *Solid State Electronics*, Volume 49, Pages 1875-1895, 2005.

[72]. Annesh Nainani, Ze Yuan, Tejas Krishnmohon, Brian R Bennett, J. Brad Boss, Mathew Reason, Mario G Ancona, Yoshio Nishi and Krishna C. Saraswat “$In_xGa_{1-x}Sb$ channel p-metal–oxide semiconductor filed effect transistors: Effect of strain and heterostructure design ” *Journal of Applied Physics* 110,014503,2011.

[73]. Aneesh Nainani Toshifumi Irisawa, Ze Yuan, Yun Sun, Tejas Krishnamohan, Matthew Reason, B rian R. Bennett, J. Brad Boos, Mario G. Ancona, Yoshio Nishi, Krishna C. Saraswat “Development of high-k dielectric for antimonides and a sub 350°C III–V pMOSFET outperforming Germanium” *International Electron Devices Meeting (IEDM)*, Page(s): 6.4.1-6.4.4, 2010.

[74]. G. Ribes, J. Mitard, M. Denais, S. Bruyere, F. Monsieur, C. Parthasarathy, E. Vincent, and G. Ghibaudo, “Review on High- k Dielectrics Reliability Issues”, *IEEE Transactions on Device and Materials Reliability*, vol. 5, pp.5-19, 2005.

[75]. E. P. Gusev, V. Narayan, M. M. Frank, "Advanced high-k dielectric stacks with polySi and metal gates: Recent progress and current challenges", *IBM Journal of Research and Development*, Vol.50,pp.387-410, 2006.