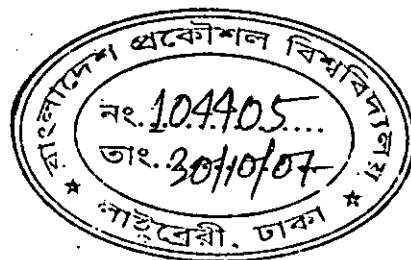


Quantum Mechanical Simulation of Nanoscale Double Gate MOSFET

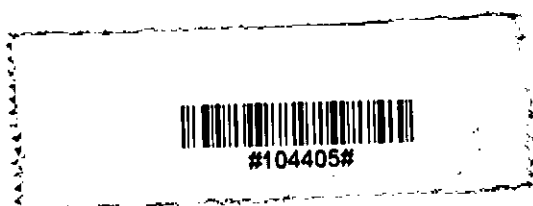


A thesis submitted to the Department of Electrical and Electronic Engineering
of
Bangladesh University of Engineering and Technology
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING

by
Md. Golam Rabbani

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

October, 2007



Declaration

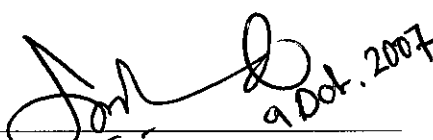
I hereby declare that this thesis has been done by me and it or any part of it has not been submitted elsewhere for the award of any degree or diploma.


Md. Golam Rabbani
09-10-07

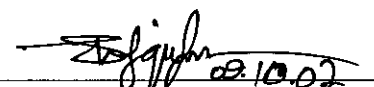
(Md. Golam Rabbani)

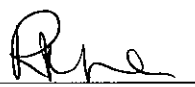
The thesis entitled “Quantum Mechanical Simulation of Nanoscale Double Gate MOSFET” submitted by Md. Golam Rabbani, Roll No.: 100506261P, Session: October, 2005 has been accepted as satisfactory in partial fulfillment of the requirements for the degree of ‘MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING’ on October 9, 2007.

BOARD OF EXAMINERS

1.  9 Oct. 2007

(Dr. Quazi Deen Mohd Khosru)
Chairman
(Supervisor)
Professor
Department of Electrical and Electronic Engineering, BUET
Dhaka-1000, Bangladesh
2. 

(Dr. S. Shahnawaz Ahmed)
Member
(Ex-officio)
Professor and Head
Department of Electrical and Electronic Engineering, BUET
Dhaka-1000, Bangladesh
3.  09.10.07

(Dr. Md. Shafiqul Islam)
Member
Professor
Department of Electrical and Electronic Engineering, BUET
Dhaka-1000, Bangladesh
4. 

(Dr. M. Rezwan Khan)
Member
(External)
Professor and Vice-Chancellor
United International University (UIU)
Dhanmondi, Dhaka, Bangladesh

To My Family

Acknowledgement

I would like to express my sincere gratitude to my advisor, Professor Quazi Deen Mohd Khosru for his generous help throughout my M.Sc. study in EEE, BUET. Professor Khosru is an admirable academic professional. He taught me not only the basics of Semiconductor Devices but also the advanced literature necessary for a quality thesis. He impressed me very much by his responsibility as well as his congenial attitude in dealing with students. He always provided timely and warm encouragement and support in difficult times. I especially thank him for his prompt reading and careful critique of my thesis. Throughout my life I will benefit from the experience and knowledge I gained working with Professor Khosru.

This thesis benefited from discussions with several of my colleagues in EEE, BUET and I am in their debt. I especially appreciate Md. Kawsar Alam for his stimulating discussions and enthusiasm about my thesis and Sishir Bhowmic for his help in the write-up and discussion on finite difference method. I heartily thank Professor Shohidul Hasan and Professor S. P. Majumder, Head of EEE department, for simulation facilities respectively in Robert Noyce Simulation Lab (RNSL) and EEE Department Computer Lab.

Finally, I owe more than words can describe to my family. This thesis is dedicated to them. But truly acknowledging them is certainly beyond my knowledge in English and it would be equally impossible even if I could do that in my native language.

Contents

	Page
Acknowledgement	v
List of Tables	viii
List of Figures	ix
Abstract	xi
1 Introduction	1
1.1 Overview of the Problem.....	1
1.2 Scaling Devices to Their Limits.....	2
1.3 Overview of the Thesis.....	5
2 2D Quantum Simulation of Nanoscale Double Gate MOSFET	7
2.1 Solution of 2D Schrödinger Equation.....	8
2.1.1 The coupled mode space (CMS) approach.....	8
2.1.2 The uncoupled mode space (UMS) approach.....	12
2.1.3 A fast uncoupled mode space (FUMS) approach.....	13
2.1.4 The nonequilibrium Green's function (NEGF).....	14
2.1.5 Calculation of charge density and current.....	16
2.1.6 Solution of 1D Schrödinger Equation.....	18
2.2 Solution of 2D Poisson Equation.....	22
2.2.1 Domain Discretization.....	24
2.2.2 Elemental Interpolation Function.....	25
2.2.3 Formulation of the System of Equations.....	27
2.2.4 Assembly to form the System of Equations.....	29

2.2.5	Imposition of Actual Boundary Condition.....	30
2.2.6	Solution of the System of Equations.....	31
3	Simulation Results and Discussions	32
3.1	Eigen Energies and Eigenfunctions (Wavefunctions).....	33
3.2	Electron Density Profile.....	38
3.3	3D Conduction Band Edge Profile.....	42
3.4	Drain-to-Source Current.....	44
3.5	Validity of FUMS approach.....	46
3.6	Comparison between femnet and nanoMOS simulation results.....	47
4	Conclusion	51
4.1	Summary.....	51
4.2	Future Work.....	52
	Bibliography	54
	Appendix A	59

List of Tables

3.1: Important physical parameters values used in the simulation.....	32
---	----

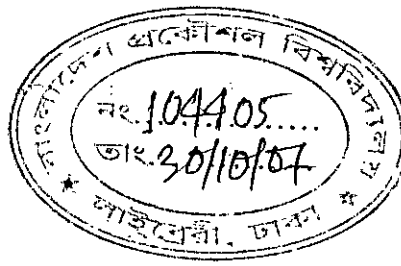
List of Figures

2.1	Symmetrical double gate MOSFET structure studied in this work. The source/drain doping is around $1 \times 10^{20} \text{ cm}^{-3}$, the channel is intrinsic and the transistor is assumed to be wide (y-dimension is treated as very large and only an x-z plane is considered in the simulation).....	8
2.2	Self-energy concept. (a) The interaction of a device with a reservoir can be represented by a self-energy matrix Σ . (b) Self-energy of the device itself (Σ_S) along with the self-energies of source (Σ_1) and drain (Σ_2).....	14
2.3	One-dimensional domain subdivided into linear elements. (a) Element and global node numbers. (b) Linear element with local node numbers.....	19
2.4	One dimensional linear interpolation functions.....	20
2.5	Domain discretization of the 2D DG MOSFET structure. Global node numbers and the element numbers are also shown.....	25
2.6	Linear interpolation functions for a triangular element. (a) N_1^e , (b) N_2^e , (c) N_3^e . The planar surfaces of the functions are shaded.....	26
2.7	Actual boundary conditions.....	30
3.1	Eigen energy levels along the transport direction (unbiased condition).....	33
3.2	Eigen energy levels along the transport direction (biased condition). From source to drain the eigen energy levels fall although there is an increase near the source-channel contact.....	34
3.3	The first (groundstate) eigen energy variation along the channel with V_g . V_d is (clock wise from top-left corner) 0.0V, 0.1V, 0.3V, and 0.5V.....	35
3.4	The first (groundstate) eigen energy variation along the channel with V_d . V_g is (clock wise from top left corner) 0.0V, 0.1V, 0.3V, and 0.5V.....	36
3.5	The unprimed and primed eigen functions along y-direction at the source or drain end. Curves without symbol are for unprimed valley while those with symbol are for primed valleys. Solid line is for ground state and dashed line is for the next higher state.....	37

3.6	The unprimed and primed eigen functions along y-direction at the middle of the channel. Curves without symbol are for unprimed valley while those with symbol are for primed valleys. Solid line is for ground state.....	37
3.7	2D electron density of different subbands along the channel (unbiased condition).....	38
3.8	2D electron density of different subbands along the channel (biased condition).....	39
3.9	2D electron density variation along the channel with V_g . Upper part for $V_d=0.0V$ and lower part for $V_d=0.5V$	40
3.10	2D electron density variation along the channel with V_d . Upper part for $V_g=0.0V$ and lower part for $V_g=0.5V$	40
3.11	Variation of 3D electron density with silicon thickness, t_{si} . (a), (b), and (c) show the 3D density for t_{si} values 1.5, 3.0, and 5.0 nm, respectively; (d) shows the ground state eigenfunction along a vertical slice.....	41
3.12	Conduction band profile; (a) 3D, $V_g=0.3V$, $V_d=0V$, (b) 3D, $V_g=0.3V$, $V_d=0.3V$, (c) 1D, $V_g=0.3V$, $V_d=0V$, (d) 1D, $V_g=0.3V$, $V_d=0.3V$	42
3.13	3D conduction band profile at different V_g . $V_d=0.5V$	43
3.14	3D conduction band profile at different V_d . $V_g=0.5V$	43
3.15	I_{DS} - V_{DS} characteristic of a DG MOSFET with $L_g=10.0$ nm and $t_{si}=3.0$ nm.....	44
3.16	I_{DS} variation with silicon thickness (t_{si}). $V_{GS}=0.5V$	45
3.17	I_{DS} vs. V_{DS} characteristic dependence on maximum number of subbands considered in the calculation.....	45
3.18	Comparison of subband levels and their populations produced by FUMS and UMS approach. Solid line represents results of FUMS and symbol represents results of UMS.....	46
3.19	Comparison of I_{DS} vs. V_{DS} profiles produced by FUMS and UMS approach. Solid line represents results of FUMS and symbol represents results of UMS.....	47
3.20	Comparison between eigen energy levels produced by femnet (blue) and nanomos (red). (a) unprimed subbands, (b) primed subbands.....	48
3.21	Comparison between 2D subband populations produced by femnet (blue) and nanomos (red). (a) unprimed subbands, (b) primed subbands.....	49
3.22	Comparison between drain currents produced by femnet (solid filled circle in blue) and nanomos (solid open circle in red).....	50

Abstract

This thesis discusses device physics and simulation issues of symmetrical nanoscale double-gate transistors at the quantum level. The purpose of the thesis is threefold 1) an understanding of the basic physics of symmetrical double-gate MOSFETs (or any nanoscale device for that matter), 2) implementation of appropriate physics and methodology in nanoscale device modeling, 3) development of a new TCAD (technology computer-aided design) tool for quantum level device simulation. We concentrate on the technical issues by investigating a double-gate structure, which has been widely accepted as the ideal device structure for ultimate CMOS scaling in the future. After a brief look at the basic physics of symmetrical double-gate MOSFET, we focus on quantum effects and non-equilibrium ballistic transport in extremely scaled transistors (in contrast to quasi-equilibrium, scattering-dominant transport in long channel devices), where a non-equilibrium Green's function (NEGF) formalism has been used to deal with the quantum transport problem. The self-consistent solution of Poisson's equation and NEGF in nanoscale device modeling is shown to be very promising and general. Results have been compared with those of a standard simulator and it has been found that the developed simulator produces similar results but requires much less time because of the use of our fast uncoupled mode space (FUMS) approach. Finite element method (FEM) has been used in discretizing the relevant PDEs. FEM can handle complex geometry more easily than the usually used finite difference (FDM) method. FEM along with FUMS makes nanoscale device simulation efficient and flexible.



Introduction

1.1. Overview of the Problem

CMOS technology has been proven as one of the most important achievements in modern engineering history. For decades, progress in device scaling has followed an exponential curve: device density on a microprocessor doubles every three years. This is known as the Moore's law [1]. The minimum dimension size of a single device for present day technology is less than 100 nm in gate-length. Continued success in device scaling is necessary for further development of the semiconductor industry in the years to come. A group of leading companies publishes their projections for the next decade in the recent International Technology Roadmap for Semiconductors (ITRS-05) [2]. The roadmap projects a device gate-length down to ~30 nm around 2014 [2]. Scaling beyond 30 nm, however, can be much more difficult and different for it is quite close to the fundamental limit of semiconductor physics. Even if lithographic and etching techniques can provide the necessary dimensions, bulk CMOS will run into a number of short channel effects associated with transistor scaling. The short channel effect (SCE) is characterized by threshold voltage V_{TH} roll-off, drain induced barrier lowering (DIBL) and sub-threshold swing. As the gate length of a MOSFET is scaled with all other device parameters held constant, sub-threshold swing increases and V_{TH} decreases, which degrade MOSFET performance. The ratio of on current to off current (I_{ON}/I_{OFF}) is reduced, giving designers a tradeoff between circuit speed and static power dissipation. Without doubt, we are facing numerous challenges, both practically and theoretically. Device simulation requires new theory and approaches to understand device physics and to design devices at the sub-30nm-scale. Efforts have been put forth in recent years [3-7], but much more is needed.

The principle objectives of this thesis are: 1) to understand the basic physics of symmetrical double-gate MOSFETs, 2) to implement appropriate physics and methodology in nanoscale device modeling, and 3) to develop a new TCAD (technology

computer aided design) tool for quantum level device simulation. We address the technical issues by investigating a double-gate structure, which more and more research evidence indicates to be the ideal device structure for ultimate CMOS scaling [4]. After a brief look at the basic physics of symmetrical double-gate MOSFET, we focus on quantum effects and non-equilibrium, near-ballistic transport in extremely scaled transistors (in contrast to quasi-equilibrium, scattering-dominant transport in long channel devices), where a non-equilibrium Green's functions formalism (NEGF) has been used to deal with the quantum transport problem [8-10]. In the remaining parts of this chapter, we will give a quick review of why the double gate structure is preferred for future device scaling. Comprehensive discussions on device scaling can be found in various literatures [3, 11-12] and detailed descriptions of NEGF framework can be found in [10, 13]. In the next section we give a quick review of why the double gate structure is preferred for future device scaling.

1.2. Scaling Devices to Limits

There are two primary device structures that have been widely studied and used in CMOS technology. One is the bulk structure, where a transistor is directly fabricated on the semiconductor substrate. The other one is called SOI (silicon-on-insulator), where a transistor is built on a thin silicon layer, which is separated from the substrate by a layer of insulator. The bulk structure is relatively simple from a device process point-of-view, and it is still the standard structure in almost all CMOS based products. For device scaling, we basically try to balance two things: device functionality and device reliability. Both of them have to be maintained at a smaller dimensional size. To accomplish this, we need to suppress any dimension related effects or short channel effects (SCEs) as much as possible. SCEs include threshold voltage (V_{TH}) variations versus channel length, typically V_{TH} roll-off at shorter channel lengths. This effect is usually accompanied by degraded subthreshold swing (S), which causes difficulty in turning off a device. SCEs also include the drain-induced barrier lowering (DIBL) effect. DIBL results in a drain voltage dependent V_{TH} , which complicates CMOS design at a circuit level. As a transistor scales, reliability concerns become more pronounced. Unwanted leakage currents can

make the device fail to function properly. Primarily, there are two kinds of leakages, gate tunneling current and junction tunneling current. Both of them result from extremely scaled dimensions and high electric fields. According to device scaling physics, increasing channel doping concentration (N_B) can effectively suppress SCEs. Frank *et al.* published their work quantifying the dependence of the scale length on N_B [14]. To a first order approximation, their theory gives the following equation,

$$\Lambda = W_{dm} + (\epsilon_{Si}/\epsilon_I)T_I \quad (1.1)$$

where Λ is the scale length, W_{dm} is the maximum channel depletion depth, T_I is the insulator thickness, and ϵ_{Si}/ϵ_I is the ratio of dielectric constants of silicon and the insulator. W_{dm} can be directly related to N_B (see for example [15]). Depending on the complexity of the channel doping profile, this theory predicts that the minimum design length L_G lies between Λ to 2Λ . It is quite clear in Eq. (1.1) that high N_B results in reduced W_{dm} , therefore a shorter scale length Λ . Of course, thinner T_I or higher ϵ_I also helps device scaling.

Device scaling has come a long way. In the early days, L_G was relatively long, a low uniform N_B could be used providing satisfactory immunity of SCEs. A low N_B gives a small body effect coefficient, which improves the subthreshold swing. [15]. As the channel length decreases, a retrograde or ground plane doping profile can be introduced [16-17]. This doping profile has a low doping region near the Si/Oxide interface, but a high doping region underneath. The top region provides better body effect, while the bottom region suppresses SCEs. To achieve even shorter channel lengths, a ground plane profile is not enough, a more complicated doping profile has to be added, namely the super halo [18]. In this case, high gradient halo dopings are formed next to the source/drain junction region. These heavily doped regions can effectively protect the source end of the channel region from the influence due to the electric fields from the drain diffusion region.

As the channel length varies around the nominal L_G , a shorter length causes the halo regions to merge, ending up with higher N_B , which resists V_{TH} rolloff. By using the ground plane and halo doping profiles, simulations show that the bulk structure can be scaled down to ~ 25 nm regime [18]. Beyond that, device scaling of the bulk structure is limited by severe degradation of junction leakage which is caused by the high built-in fields, and can not be avoided in the wake of the super halo engineering. Partially depleted SOI MOSFETs scale in a very similar manner as the bulk devices do. The buried oxide layer in a SOI device can provide superior electric isolation between the active device region and the substrate region. This property is considered a big improvement over bulk devices. Body isolation, however, also results in charge buildup (majority carriers) within the body region, which gives rise to the unwanted floating body effect (FBE) [19]. A fully depleted SOI MOSFET can help relieve the FBE, but a fully depleted single gate SOI MOSFET is not considered a desired structure for scaling. A single gate SOI device typically has a thick buried oxide layer, which can not terminate any electric lines from the drain end, leaving the source vulnerable to the influence of the drain [17, 20]. All recent studies indicate that the ultra-thin body double gate (DG) SOI MOSFET is the ideal device structure for ultimate scaling [4, 21-22]. In an ultra-thin body DG MOSFET, the second gate electrode can significantly suppress the SCEs. Referring to Eq. 1.1, and noting that W_{dm} can be approximated by $T_{Si}/2$ (T_{Si} is the silicon body thickness), when T_{Si} is scaled to nanometer thicknesses (close to T_f), clearly the scale length will downsize into the nanometer regime. It should be also noticed that high body doping is not needed here, so the band-to-band tunneling junction leakage is no longer a big concern. Moreover, the use of ultra-thin bodies will result in reduced metallurgical junction perimeter, therefore low junction capacitance. The bodies are typically lightly doped, giving other advantages: 1) there is barely room for the FBE to come into play, 2) the V_{TH} variation due to dopant fluctuations can be eliminated, 3) close-to-ideal subthreshold swing (60 mV/dec) can be achieved, 4) severe mobility degradation due to ion scattering might be avoided. From a technical point of view, DG MOSFETs are difficult to build. Gate self-alignment is hard to achieve. A misaligned gate will cause high overlap capacitances on one side of the gate, and large underlap

junction resistances on the other side of the gate. Recent works show that clever process designs can help get rid of gate misalignment [23-24]. Extension region resistances pose another concern in DG MOSFET design. Due to the use of ultra-thin bodies, these resistances can be very high, limiting device performance. The proposed solution is to use fanned out source/drain regions as close as possible to the channel region [25]. The use of ultra-thin bodies also leaves limited room for adjusting V_{TH} with body doping. Gate stack engineering has to be done to obtain an appropriate V_{TH} , either by employing new contact materials with desirable workfunctions, or maintaining an offset voltage between the two gate electrodes to mimic a different workfunction [24-26]. Quantum effects (subband splitting) can become significant as the confinement of carriers becomes stronger within ultra-thin bodies, translating to sensitivity of V_{TH} to the body thickness. This fundamental physics effect poses an additional difficulty to control V_{TH} in ultra-thin bodies. (It is worthwhile to point out that this subband splitting effect will increase the band gap between lowest electron subband and highest hole subband, which may considerably suppress the band-to-band tunneling leakage in ultra-thin silicon bodies.) Despite the existence of numerous difficulties, the excellent scaling capability demonstrated by the ultra-thin body DG structure can never be underestimated. For this reason, therefore, this thesis will be concentrated on a study of ultra-thin body DG MOSFETs.

1.3. Overview of the Thesis

The thesis is organized as follows. In Chapter 2, we describe the numerical techniques used in developing a 2D simulator for nanoscale double-gate MOSFETs. First, we implement a quantum ballistic transport model by employing fast uncoupled mode space (FUMS) approach in the nonequilibrium Green's function (NEGF) formalism. Then the 2D Poisson equation is solved using the Gummel iteration. All the relevant equations (1D Schrödinger equation, 1D NEGF and 2D Poisson equation) are discretized using the finite element method (FEM).

In Chapter 3, we present simulation results of our work. Different properties of the nanoscale DG MOSFET are demonstrated through relevant plots. We provide eigen energy and eigen-function profiles, electron density profiles, conduction band edge plots, and drain current profiles taking into account bias, thickness and maximum number of subbands dependence. We also demonstrate validity of the EUMS approach by comparison with the uncoupled mode space (UMS) approach. Finally, results of our simulation process are compared with those from a standard simulator.

Chapter 4 summarizes the conclusions of this research, and lists a few potential directions for future work. A sample 2D mesh generator code developed for the simulation and a flow chart of the simulation process are given in Appendix A.

2D Simulation of Nanoscale Double Gate MOSFET

Recently Taur *et al.* [29-31] has developed analytical solutions of charge, potential and drain-current for symmetric undoped double-gate MOSFETs. Quantum effects have been neglected in the development of those analytical formulations. But as transistor dimensions are scaled down, quantum effects, which affect the threshold voltage (confinement), gate capacitance (charge centroid shift), off-current (source barrier tunneling) and gate leakage begin to manifest themselves and semiclassical methods or the analytical models that disregard these effects become inadequate in capturing the physics of ballistic transport. The Nonequilibrium Green's Function (NEGF) approach [35-36] provides a general simulation framework for quantum mechanical simulation (with or without scattering) and it will be applied to model electron transport in this work.

In this chapter, we present the two-dimensional (2D) quantum-mechanical simulation of nanoscale double gate (DG) MOSFETs in the finite element method (FEM domain) approach. Effective mass approximation (whose validity for nanoscale device simulation has been established in [37]) is used and we determine the upper performance limit of the devices so that scattering is disregarded. For a method for incorporation of scattering into the simulation refer to [38]. For the solution of 2D Schrödinger equation, instead of the usually used real-space approach, we use the mode space approach [39] which has been shown to be very efficient for the device geometry simulated in this work. The geometry also permits us to decouple the modes thereby greatly reducing the simulation cost. As the silicon channel thickness is less than 5 nm, strong quantum confinement is present so that only a few lower subbands are occupied and need to be treated. We further show that due to the small device size there exists a strong co-relation between the potentials at different points of the device and an averaging approach can be utilized to speed up the simulation without sacrificing accuracy.

2.1 Solution of 2D Schrödinger Equation

In this section we show how the 2D problem can be rendered to a number of 1D problems through the use of the novel mode space [39] (subband decomposition) approach. Utilizing the convenient structural symmetry of DG geometry, the modes can be uncoupled and only a few are needed consideration, which greatly improves simulation efficiency. The simulation can be further speeded up through the use of an averaging approach called the fast uncoupled mode space (FUMS) approach.

2.1.1 The coupled mode space (CMS) approach

The double gate MOSFET geometry that we use is shown in Fig. 2.1. We assume an intrinsic device. The device is wide in the y -direction and no confining potential exists in this direction, so the wave functions in this direction are plane waves and the 2D simulation is performed in the x - z plane as shown in Fig. 2.1.

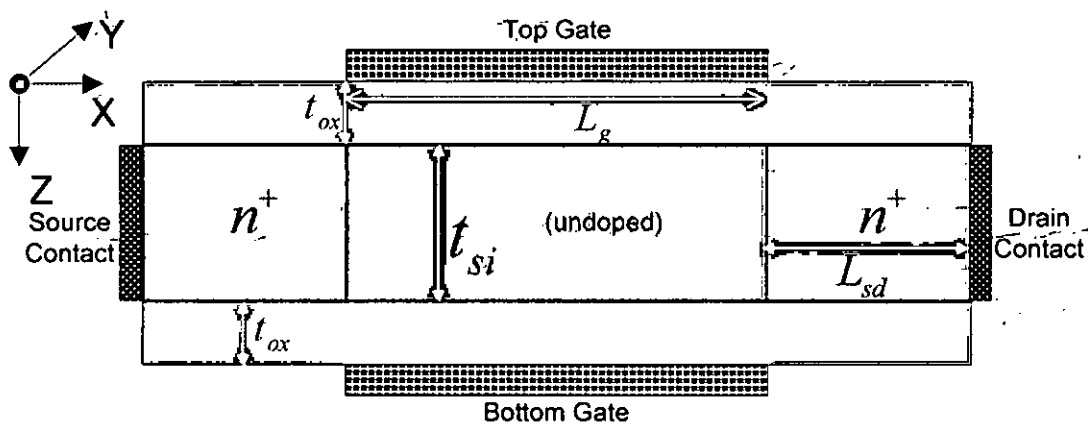


Fig. 2.1 Symmetrical double-gate MOSFET structure studied in this work. The source/drain doping is around $1 \times 10^{20} \text{ cm}^{-3}$, the channel is intrinsic and the transistor is assumed to be wide (y -dimension is treated as very large and only an x - z plane is considered in the simulation).

Thus the total wave function can be written as $\psi(x, y, z) = \Phi(x, z)e^{ik_y y} / \sqrt{W}$ where W is the width of the device (in y -direction) and $\Phi(x, z)$ is obtained from

$$-\frac{\hbar^2}{2m_x^*} \frac{\partial^2 \Phi(x, z)}{\partial x^2} - \frac{\hbar^2}{2m_z^*} \frac{\partial^2 \Phi(x, z)}{\partial z^2} + E_C(x, z)\Phi(x, z) = (E - E_{k_y})\Phi(x, z) = E_l \Phi(x, z) \quad (2.1)$$

where m_x^* , m_z^* are the electron effective masses in the x -, and z -directions respectively,

E is the assumed energy, $E_l = E - E_{k_y}$ is the longitudinal energy, and $E_{k_y} = \frac{\hbar^2 k_y^2}{2m_y^*}$.

We can directly discretize Eq. (2.1) in real space but it imposes a heavy computational burden [39]. For MOSFETs with a thin body (<5nm), quantum confinement in the z -direction introduces subbands, and only a few of the lowest subbands are necessary as population of the higher subbands are negligible. Accordingly, we can expand $\Phi(x, z)$ in the subband eigenfunction space,

$$\Phi(x, z) = \sum_n \varphi^n(x) \xi^n(z; x) \quad (2.2)$$

where n indicates the subband number, $\varphi^n(x)$ are expansion coefficients, $\xi^n(z; x = x_0)$ is the n th eigenfunction of the following 1D Schrödinger equation at the slice $x = x_0$ of the DG MOSFET,

$$-\frac{\hbar^2}{2m_z^*} \frac{\partial^2 \xi^n(z; x_0)}{\partial z^2} + E_C(z; x_0) \xi^n(z; x_0) = E_{sub}^n(x_0) \xi^n(z; x_0) \quad (2.3)$$

where $E_C(z; x_0)$ is the conduction band profile along z -direction at slice $x = x_0$, $E_{sub}^n(x_0)$ is the n th subband energy level due to $E_C(z; x_0)$. According to the property of eigenfunctions, $\xi^n(z; x)$ satisfies the following equation for any x

$$\int_z \xi^m(z; x) \xi^n(z; x) dz = \delta_{m,n} \quad (2.4)$$

where $\delta_{m,n}$ is the Kronecker delta function. Inserting Eq. (2.2) into Eq. (2.1) and making use of Eq. (2.3) we get;

$$\begin{aligned} & -\frac{\hbar^2}{2m_x^*} \frac{\partial^2}{\partial x^2} \left(\sum_n \varphi^n(x) \xi^n(z; x) \right) - \frac{\hbar^2}{2m_z^*} \frac{\partial^2}{\partial z^2} \left(\sum_n \varphi^n(x) \xi^n(z; x) \right) + E_c(z; x) \left(\sum_n \varphi^n(x) \xi^n(z; x) \right) \\ & = E_l \left(\sum_n \varphi^n(x) \xi^n(z; x) \right) \\ \text{or, } & -\frac{\hbar^2}{2m_x^*} \frac{\partial^2}{\partial x^2} \left(\sum_n \varphi^n(x) \xi^n(z; x) \right) + \sum_n \left(-\frac{\hbar^2}{2m_z^*} \frac{\partial^2}{\partial z^2} \xi^n(z; x) + E_c(z; x) \xi^n(z; x) \right) \varphi^n(x) \\ & = E_l \left(\sum_n \varphi^n(x) \xi^n(z; x) \right) \\ \text{or, } & -\frac{\hbar^2}{2m_x^*} \frac{\partial^2}{\partial x^2} \left(\sum_n \varphi^n(x) \xi^n(z; x) \right) + \sum_n \varphi^n(x) E_{sub}^n(x) \xi^n(z; x) = E_l \left(\sum_n \varphi^n(x) \xi^n(z; x) \right) \end{aligned} \quad (2.5)$$

Now we multiply on both sides by $\xi^m(z; x)$ and do an integral in z -direction. According to Eq. (2.4), we obtain the following 1D-Schrödinger equation:

$$\begin{aligned} & -\frac{\hbar^2}{2m_x^*} \left(\sum_n a_{mn}(x) \right) \frac{\partial^2}{\partial x^2} \varphi^m(x) - \frac{\hbar^2}{2m_x^*} \sum_n c_{mn}(x) \varphi^m(x) - \frac{\hbar^2}{m_x^*} \sum_n b_{mn}(x) \frac{\partial}{\partial x} \varphi^n(x) \\ & + E_{sub}^m(x) \varphi^m(x) = E_l \varphi^m(x) \end{aligned} \quad (2.6)$$

where

$$a_{mn}(x) = \int_z \xi^m(z; x) \xi^n(z; x) dz = \delta_{mn} \quad (2.7a)$$

$$b_{mn}(x) = \int_z \xi^m(z; x) \frac{\partial}{\partial x} \xi^n(z; x) dz \quad (2.7b)$$

$$\text{and } c_{mn}(x) = \int_z \xi^m(z; x) \frac{\partial^2}{\partial x^2} \xi^n(z; x) dz \quad (2.7c)$$

Eq. (2.6) is the basic equation for the coupled mode-space (CMS) approach. From the derivation above, it is clear that the CMS formalism is mathematically equivalent to the real space calculation if all the modes (i.e., $m, n = 1, 2, 3, \dots, N_z$, where N_z is the number of nodes in the z -direction) are included. In practice, due to strong quantum confinement usually only a few of the lowest subbands (i.e., $m, n = 1, 2, \dots, M; M \ll N_z$) are occupied and need to be included in the calculation (which means that if we increase the mode number M , the device characteristics such as the electron density profile and terminal currents will not change any more). Thus with the first M subbands considered, Eq. (2.6) represents an equation group that contains M equations, each representing a selected mode. We can write down these M equations in a matrix format

$$H \begin{bmatrix} \varphi^1(x) \\ \varphi^2(x) \\ \dots \\ \dots \\ \varphi^M(x) \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & \dots & h_{1M} \\ h_{21} & h_{22} & h_{23} & \dots & h_{2M} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ h_{M1} & h_{M2} & h_{M3} & \dots & h_{MM} \end{bmatrix} \begin{bmatrix} \varphi^1(x) \\ \varphi^2(x) \\ \dots \\ \dots \\ \varphi^M(x) \end{bmatrix} = E_I \begin{bmatrix} \varphi^1(x) \\ \varphi^2(x) \\ \dots \\ \dots \\ \varphi^M(x) \end{bmatrix} \quad (2.8)$$

where,

$$h_{mn} = \delta_{m,n} \left[-\frac{\hbar^2}{2m_x^*} \frac{\partial^2}{\partial x^2} + E_{sub}^m(x) \right] - \frac{\hbar^2}{2m_x^*} c_{mn}(x) - \frac{\hbar^2}{m_x^*} b_{mn}(x) \frac{\partial}{\partial x} \quad (2.9)$$

$(m, n = 1, 2, \dots, M)$

By using the coupled mode space (CMS) approach, the size of the device Hamiltonian H has been reduced to $(M \times N_x) \times (M \times N_x)$ (N_x is the number of nodes in x -direction, and the mode number M that we need is normally less than 3 for the DG MOSFET structure we simulate), which is much smaller than that required in the real space representation, $(N_z \times N_x) \times (N_z \times N_x)$ (N_z is ~ 40 for the device structures simulated in this work).

2.1.2 The uncoupled mode space (UMS) approach

For the DG MOSFET, we assume that the shape of the Si body is uniform along the x direction. As a result, the confinement potential profile (in the z direction) varies very slowly along the channel direction. For instance, the conduction band edge $E_C(z;x)$ takes the same shape but different values at different x . For this reason, the eigenfunctions $\xi^m(z;x)$ are approximately the same along the channel although the eigenvalues $E_{sub}^m(x)$ are different. So we assume

$$\xi^m(z;x) = \xi^m(z) \quad (2.10)$$

$$\text{or } \frac{\partial}{\partial x} \xi^m(z;x) = 0 \quad (m=1,2,\dots,M), \quad (2.11)$$

which when substituted in Eq. (2.7) results in $b_{mn}(x) = 0$ and $c_{mn}(x) = 0$ ($m, n=1,2,\dots,M$). Using the results in Eq. (2.9), we obtain $h_{mn} = 0$ ($m \neq n$ and $m, n=1,2,\dots,M$), which means that the coupling between the modes is negligible (all the modes are uncoupled). Thus the device Hamiltonian H becomes a block-diagonal matrix,

$$H = \begin{bmatrix} h_{11} & 0 & \dots & \dots & 0 \\ 0 & h_{22} & 0 & \dots & \dots \\ \dots & 0 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & h_{MM} \end{bmatrix} \quad (2.12)$$

$$\text{where, } h_{mn} = \delta_{m,n} \left[-\frac{\hbar^2}{2m_x^*} \frac{\partial^2}{\partial x^2} + E_{sub}^m(x) \right] \quad (2.13)$$

2.1.3 A fast uncoupled mode space (FUMS) approach

As mentioned earlier, for UMS approach, we need to solve N_x 1D Schrödinger equations in a self-consistent loop to obtain the electron subbands and eigenfunctions. This part of the simulation takes most of the simulation time. To increase the simulation efficiency, we introduce an FUMS approach, which involves only one 1D Schrödinger equation in the z -direction and still provides excellent computational accuracy as compared to UMS approach.

Recall the assumption made in Sec. 2.1.2 that the eigenfunctions $\xi^m(z; x)$ are almost invariant along the x direction, $\xi^m(z; x) = \xi^m(z)$. Now we suppose that the average wave functions $\overline{\xi^m(z)}$ are the eigenfunctions of the following 1D Schrödinger equation.

$$\left[-\frac{\hbar^2}{2m_x^*} \frac{\partial}{\partial z^2} + \overline{E_C(z)} \right] \overline{\xi^m(z)} = \overline{E_{sub}^m} \overline{\xi^m(z)} \quad (2.14)$$

Here the average conduction band edge $\overline{E_C(z)}$ is obtained as.

$$\overline{E_C(z)} = \frac{1}{L_x} \int_x E_C(x, z) dx \quad (2.15)$$

where L_x is the total length of the simulated DG MOSFET (including the source/drain extensions). After computing the eigenvalues $\overline{E_{sub}^m}$ and eigenfunctions $\overline{\xi^m(z)}$ of this Schrodinger equation, we use the first-order stationery perturbation theory to obtain the subband profile as [40]

$$E_{sub}^m(x) = \overline{E_{sub}^m} + \int_x E_C(x; z) \left| \overline{\xi^m(z)} \right|^2 dz - \int_z \overline{E_C(z)} \left| \overline{\xi^m(z)} \right|^2 dz \quad (2.16)$$

The use of FUMS approach highly improves the efficiency of the simulation and makes it a practical model for extensive device simulation and design.

2.1.4 The nonequilibrium Green's function-(NEGF)

After the device Hamiltonian H is obtained, we can calculate the electron density and current using the non-equilibrium Green function (NEGF) approach [35, 10]. NEGF is a very general formalism in that it is equally applicable for DG-MOSFET, nanowire, molecular transistor, etc. Here we present the relevant equations for our particular case. The retarded Green's function for mode m is defined as [10]

$$G^m(E) = [E_i S - h_{mm} - \Sigma_S^m(E) - \Sigma_1^m(E) - \Sigma_2^m(E)]^{-1} \quad (2.17)$$

where m represents mode/subband number, E_i is the longitudinal energy, S is a discretization method dependent matrix defined below in Eq. (2.18), h_{mm} is defined by Eq. (2.13), Σ_S^m is called the self energy (potential) that accounts for scattering inside the device itself (for the ballistic case, which is assumed in this work, it is a null matrix.), Σ_1^m (Σ_2^m) is the self energy (potential) caused by the coupling between the device and the source (drain) reservoir or contact, as shown below in Fig. 2.2.

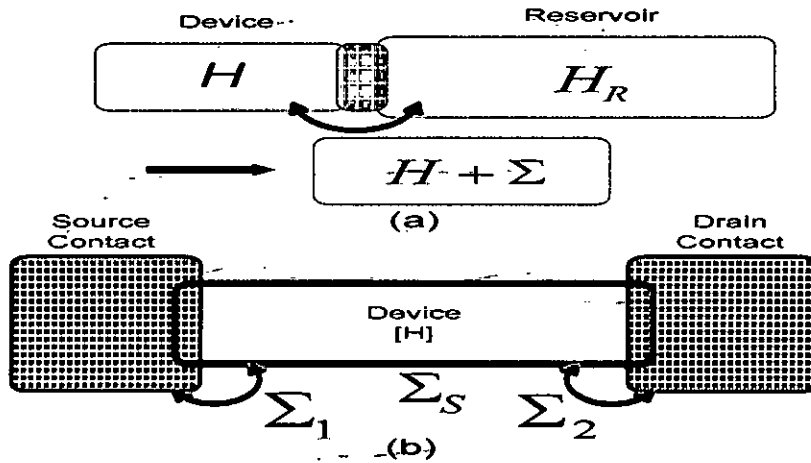


Fig. 2.2 Self-energy concept. (a) The interaction of a device with a reservoir can be represented by a self-energy matrix Σ . (b) Self-energy of the device itself (Σ_S) along with the self-energies of source (Σ_1) and drain (Σ_2) reservoirs. (From [35])

If we discretize the equations by the finite element method (FEM), then S is an $N_x \times N_x$ matrix given by [40]

$$S = \begin{bmatrix} a/3 & a/6 & 0 & \dots & \dots & \dots \\ a/6 & 2a/3 & a/6 & \dots & \dots & \dots \\ 0 & a/6 & 2a/3 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & 2a/3 & a/6 \\ 0 & \dots & \dots & 0 & a/6 & a/3 \end{bmatrix} \quad (2.18)$$

where a is the discretization grid spacing along the x-direction. The self-energies, Σ_1^m and Σ_2^m , are defined as [40]

$$\Sigma_1^m = \begin{bmatrix} -jk_{m,1}at & 0 & \dots & \dots & 0 \\ 0 & 0 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & 0 \end{bmatrix} \quad (2.19)$$

and

$$\Sigma_2^m = \begin{bmatrix} 0 & \dots & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & 0 & 0 \\ 0 & \dots & \dots & 0 & -jk_{m,N_x}at \end{bmatrix} \quad (2.20)$$

$$(m=1, 2, \dots, M \text{ and } p, q=1, 2, \dots, MN_x)$$

where $t = \hbar^2 / 2m_x^* a^2$ and $k_{m,1} [k_{m,N_x}]$ is given by $E_l = E_{sub}^m(x=0) + 2t(1 - \cos k_{m,1}a)$
 $[E_l = E_{sub}^m(x = (N_x - 1)a) + 2t(1 - \cos k_{m,N_x}a)]$.

2.1.5 Calculation of charge density and current

After the retarded Green's function $G(E)$ at a given energy E has been calculated using Eq. (2.17), the spectral density functions due to the source/drain contacts can be obtained as [40] (mode number m has not been shown for clarity)

$$A_1(E) = G(E)\Gamma_1(E)G^\dagger(E), \quad A_2(E) = G(E)\Gamma_2(E)G^\dagger(E), \quad (2.21)$$

where '+' means transpose conjugate and $\Gamma_1(E) \equiv j[\Sigma_1(E) - \Sigma_1^\dagger(E)]$ and $\Gamma_2(E) \equiv j[\Sigma_2(E) - \Sigma_2^\dagger(E)]$. Γ_1 and Γ_2 are known as broadening functions [35], which determine electron exchange rates between the active device region and the source/drain reservoirs at energy E . Although the device itself may be in a non-equilibrium state, electrons are injected from the equilibrium source/drain reservoirs. Note that the spectral functions are N_X by N_X matrices and the diagonal entries represent the local density-of-states (LDOS) at each node. The source related spectral function is filled up according to the Fermi energy in the source contact, while the drain related spectral function is filled up according to the Fermi energy in the drain contact.

The 2D electron density matrix is obtained as [41]

$$n(E_l) = \frac{1}{2\pi a} \int_0^{+\infty} \frac{2}{\pi\hbar} \sqrt{\frac{m_y^*}{2E_{k_j}}} [f(\mu_1 - E_l - E_{k_j})A_1 + f(\mu_2 - E_l - E_{k_j})A_2] dE_{k_j} \quad (2.22a)$$

where f is the Fermi-Dirac function (for an analytical expression, see [42]), μ_1 (μ_2) is source (drain) contact chemical potential, and $\frac{2}{\mu\hbar} \sqrt{\frac{m_y^*}{2E_{k_j}}}$ represents the transverse mode state density (including spin degeneracy). Since the spectral functions depend on the longitudinal energy only, they can be moved out of the integration sign. Therefore, Eq. (2.22a) reduces to [41]

$$n(E_l) = \frac{1}{\hbar a} \sqrt{\frac{m_y^* k_B T}{2\pi^3}} [\mathfrak{F}_{-1/2}(\mu_1 - E_l) A_1 + \mathfrak{F}_{-1/2}(\mu_2 - E_l) A_2], \quad (2.22b)$$

where the Fermi-Dirac integral of $\mathfrak{F}_{-1/2}$ accounts for all transverse mode contributions (see [42] for analytical approximation for $\mathfrak{F}_{-1/2}$, and also note that all quantities appearing as arguments of Fermi-Dirac integrals are normalized to $k_B T$). To obtain the total 2D electron density, we need to integrate Eq. (2.22b) over E_l . We also need to sum contributions from all the conduction band valleys and their subbands. Finally, we can get the 3D electron density by multiplying the corresponding distribution function $|\xi^n(z; x)|^2$ to the 2D density matrix at each longitudinal lattice node. The 3D electron density is fed back to the Poisson equation solver for the self-consistent solution.

Once self-consistency is achieved, the terminal current can be expressed as a function of the transmission coefficient [10]. The transmission coefficient from the source contact to the drain contact is defined in terms of the Green's function as [10]

$$T_{SD}(E) = \text{trace} [\Gamma_1(E) G(E) \Gamma_2(E) G^+(E)] \quad (2.23)$$

where *trace* means sum of the diagonal elements of a square matrix and Γ_1 , $[\Gamma_2]$ is the source (drain) broadening function defined above. It is straightforward to write the transmitted current as

$$I(E_l) = \frac{q}{h} \int_0^{+\infty} \frac{2}{\pi \hbar} \sqrt{\frac{m_y^*}{2E_{k_j}}} [f(\mu_S - E_l - E_{k_j}) A_S - f(\mu_D - E_l - E_{k_j}) A_D] T_{SD}(E_l, E_{k_j}) dE_{k_j}, \quad (2.24a)$$

where the 2 in the numerator is for spin degeneracy. Note that T_{SD} is independent of transverse energy E_{k_j} and can therefore be moved out of the integration sign. Eq. (2.24a) then reduces to

$$I(E_l) = \frac{q}{\hbar^2} \sqrt{\frac{m_y^* k_B T}{2\pi^3}} [\mathfrak{F}_{-1/2}(\mu_S - E_l) - \mathfrak{F}_{-1/2}(\mu_D - E_l)] T_{SD}(E_l) \quad (2.24b)$$

The total current is obtained from integrating over E_l and summing over all valleys and subbands.

2.1.6 Solution of 1D Schrödinger Equation

In this subsection we show the discretization of 1D Schrödinger equation using the finite element technique. Consider the 1D Schrödinger equation

$$\left[-\frac{\hbar^2}{2m^*} \frac{d^2}{dx^2} + E_C(x) \right] \psi(x) = E\psi(x) \quad (2.25)$$

where m^* is the electron effective mass, $E_C(x)$ is the conduction band edge (with the confining potential included) and E is the energy, with the boundary conditions

$$\psi|_{x=0} = 0 \text{ and } \psi|_{x=L_x} = 0 \quad (2.26)$$

(For open boundary conditions with self-energies, i.e., for Green's function see [40]). In FEM, Eq. (2.25) can be solved along with its associated boundary conditions by solving the equivalent variational problem defined by [44]

$$\begin{cases} \delta F(\psi) = 0 \\ \psi|_{x=0} = 0 \\ \psi|_{x=L_x} = 0 \end{cases} \quad (2.27)$$

where

$$F(\psi) = \frac{1}{2} \int_0^{L_x} \left[\frac{\hbar^2}{2m^*} \left(\frac{d\psi}{dx} \right)^2 + E_C \psi^2 \right] dx - \frac{1}{2} E \int_0^{L_x} \psi^2 dx \quad (2.28)$$

The different steps of the 1D FEM are explained below:

(i) Discretization

This is the first step of FEM. Here we divide the solution domain $(0, L_x)$ into small subdomains, which in this case will be short line segments as shown in Fig. 2.3.

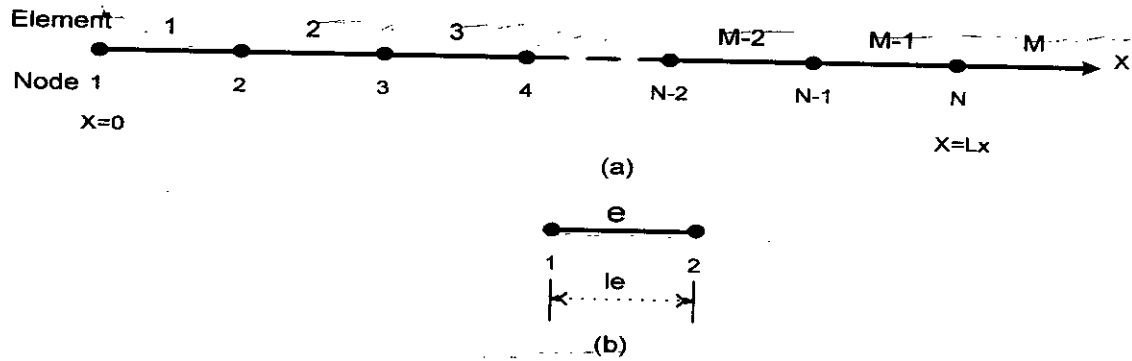


Fig. 2.3 One-dimensional domain subdivided into linear elements. (a) Element and global node numbers. (b) Linear element with local node numbers.

Let l^e ($e=1,2,\dots,M$) denote the length of the e th element M denote the total number of elements and x_i ($i=1,2,\dots,N$) denote the position of the i th node with $x_1=0$ and $x_N=L_x$. In the following, the superscript e is used to denote the quantity with a local number as its subscript, while for all other quantities the subscript is a global number. In this simple 1D case, the local and global systems are related by

$$x_1^e = x_e \text{ and } x_2^e = x_{e+1} \text{ for } e=1,2,\dots,M. \quad (2.29)$$

(ii) Selection of interpolation function:

We select linear interpolation functions for simplicity because they are adequate for the most cases. In this case, $\psi(x)$ within the e th element may be approximated as

$$\Psi^e(x) = a^e + b^e x \quad (2.30)$$

where a^e and b^e are constants to be determined. Enforcing Eq. (2.30) at the two ends of element e , we obtain

$$\psi^e(x) = \sum_{j=1}^2 N_j^e(x) \psi_j^e \quad (2.31)$$

where N_1^e and N_2^e are the interpolation or basis functions given by $N_1^e(x) = \frac{x_2^e - x}{l^e}$ and

$N_2^e(x) = \frac{x - x_1^e}{l^e}$ with $l^e = x_2^e - x_1^e$. Obviously, $N_j^e(x_i^e) = \delta_{ij}$ as shown in Fig. 2.4.

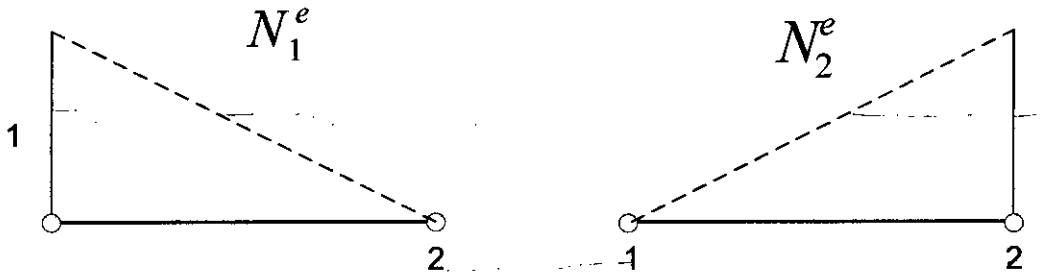


Fig. 2.4 One dimensional linear interpolation functions

(iii) Formulation of the system of equations

Derivation of Elemental Equations

For the formulation we use Ritz method. Assuming homogenous Neumann boundary condition for the time being Eq. (2.25) can be written as

$$F(\psi) = \sum_{e=1}^M F^e(\psi^e) \quad (2.32)$$

where

$$F^e(\psi^e) = \frac{1}{2} \int_{x_1^e}^{x_2^e} \left[\frac{\hbar^2}{2m^*} \left(\frac{d\psi^e}{dx} \right)^2 + E_C(\psi^e)^2 \right] dx - \frac{1}{2} E \int_{x_1^e}^{x_2^e} (\psi^e)^2 dx \quad (2.33)$$

Substituting Eq. (2.33) into Eq. (2.32) and then taking the derivative of F^e with respect to ψ^e , we obtain

$$\frac{\partial F^e}{\partial \psi_i^e} = \sum_{j=1}^2 \psi_j^e \int_{x_1^e}^{x_2^e} \left(\frac{\hbar^2}{2m^*} \frac{dN_i^e}{dx} \frac{dN_j^e}{dx} + E_C N_i^e N_j^e - E N_i^e N_j^e \right) dx \quad (2.34)$$

which in matrix form can be written as

$$\left\{ \frac{\partial F^e}{\partial \psi^e} \right\} = [K^e] \{\psi^e\} - E [S^e] \{\psi^e\} \quad (2.35)$$

with

$$K_{ij}^e = \int_{x_1^e}^{x_2^e} \left(\frac{\hbar^2}{2m^*} \frac{dN_i^e}{dx} \frac{dN_j^e}{dx} + E_C N_i^e N_j^e \right) dx \quad (2.36)$$

and

$$S_{ij}^e = \int_{x_1^e}^{x_2^e} (N_i^e N_j^e) dx \quad (2.37)$$

We note that $[K^e]$ and $[S^e]$ are symmetric and, if m^* and E_C are constant or can be approximated by constants within each element, their matrix elements can be evaluated analytically [44]. The result is

$$K_{11}^e = K_{22}^e = \frac{\hbar^2}{2m^*} + E_C \frac{l^e}{3} \quad (2.38a)$$

$$K_{12}^e = K_{21}^e = -\frac{\hbar^2}{2m^*} + E_C \frac{l^e}{6} \quad (2.38b)$$

$$S_{11}^e = S_{22}^e = \frac{l^e}{3} \quad \text{and} \quad S_{12}^e = S_{21}^e = \frac{l^e}{6} \quad (2.39)$$

Assembly to Form the System-of Equations

We will not discuss the assembly procedure for 1D as it is similar to and yet simpler than the 2D case discussed in the next section for 2D FEM formulation of Poisson equation. After the assembly is complete, we get

$$[K] = \sum_{e=1}^M [\overline{K}^e] \text{ and } [S] = \sum_{e=1}^M [\overline{S}^e] \quad (2.40)$$

where bar denotes that the corresponding matrix has been augmented by zero filling so that it is the same size as K or S .

Finally, standard procedures [44] can be used to find the eigenvalues and associated eigenfunctions.

2.2 Solution of 2D Poisson Equation

For self-consistency of the solution, Schrödinger equation has to be solved iteratively with the Poisson equation. In this section, we show solution technique for the 2D Poisson equation in the x - z plane shown in Fig. 2.1. The Poisson equation in this case is

$$-\frac{\partial}{\partial x} \left(\epsilon_x \frac{\partial \phi}{\partial x} \right) - \frac{\partial}{\partial z} \left(\epsilon_z \frac{\partial \phi}{\partial z} \right) = \frac{q}{\epsilon_0} [n(x, z) - N_D(x, z)] \quad (2.41)$$

where $\phi(x, z)$ is the potential, $n(x, z)$ is the electron density, $N_D(x, z)$ is the ionized donor density (complete ionization is assumed) at position (x, z) , q is the magnitude of electronic charge, ϵ_0 is absolute permittivity, ϵ_x and ϵ_z are the relative permittivities in the x and z directions, respectively.

We can solve Eq. (2.41) using FDM or FEM approach. However, when solving a coupled set of equations (the Poisson equation and transport equation), direct use of Eq. (2.41) leads to slow convergence [45]. Fortunately, there is a better solution algorithm for solving the Poisson equation [45]. This algorithm can provide more efficient convergence

in the iteration loop of the Poisson and transport equation. This algorithm involves performing a variable change to n , namely expressing n in terms of the potential and a quasi-Fermi energy, F_n . The quasi-Fermi potential energy is computed based on the old potential as

$$F_n = -q\phi_{old} + k_B T \cdot \mathfrak{F}_{1/2}^{-1} \left(\frac{n}{N_C} \right) \quad (2.42)$$

where $\mathfrak{F}_{1/2}^{-1}$ stands for the inverse Fermi-Dirac integral of order 1/2 and N_C the effective density of states in the conduction band (a normalization factor). Analytical approximation for $\mathfrak{F}_{1/2}^{-1}$ can be found in [46].) The electron density term in Eq. (2.41) now becomes

$$n = N_C \mathfrak{F}_{1/2} \left[\frac{F_n + q\phi}{k_B T} \right] \quad (2.43)$$

With this variable change, Eq. (2.41) now represents a set of nonlinear equations for the potential. The advantage of this approach is that it builds a negative feedback into the iterative process [45]. If the potential increases (conduction band decreases) during the Poisson solution, the subsequent transport solution will increase $n(x, z)$ as carriers flow to regions of lower energy. This coupling is built into the Poisson equation when Eq. (2.42) is used. This approach has proven effective in previous quantum and semiclassical transport simulations [47–49] and proved similarly effective in this work.

We now discretize the Poisson equation using the FEM approach and for the time being we rewrite Eq. (2.41) as

$$-\frac{\partial}{\partial x} \left(\epsilon_x \frac{\partial \phi}{\partial x} \right) - \frac{\partial}{\partial z} \left(\epsilon_z \frac{\partial \phi}{\partial z} \right) = f \quad (2.44)$$

with

$$f = \frac{q}{\epsilon_0} [n(x, z) - N_D(x, z)] \quad (2.45)$$

and use homogeneous Neumann boundary condition. Real boundary conditions will be imposed after the assembly of the system of equations, as was done for the 1D Schrödinger equation. The variational problem equivalent to the boundary-value problem in Eq. (2.44) is given by [44]

$$\delta F(\phi) = 0 \quad (2.46)$$

where

$$F(\phi) = \frac{1}{2} \iint_{\Omega} \left[\epsilon_x \left(\frac{\partial \phi}{\partial x} \right)^2 + \epsilon_z \left(\frac{\partial \phi}{\partial z} \right)^2 \right] d\Omega - \iint_{\Omega} f\phi d\Omega \quad (2.47)$$

2.2.1 Domain Discretization

This is the first step in the FEM approach. The two-dimensional DG-MOSFET structure is subdivided into a number of small triangular elements as shown in Fig. 2.5. There exist other methods [44] for the discretization of 2D region using linear triangular elements, but they are not suited for this case because of our requirement that we need to solve the 1D Schrödinger in the vertical slices (z -direction) and the NEGF in the x -direction. So the grid points in the in both the dimensions should on straight lines (grid spacing may non-uniform). The nodes and the triangular elements are numbered as shown in the figure. The discretization provides us with the following information:

Nn: number of (global) nodes

Ne: number of (triangular) elements

x : x-coordinate of the nodes ($N_e \times 1$)

z : z-coordinate of the nodes ($N_e \times 1$)

te: connectivity matrix ($3 \times N_e$)

Nb: no. of boundary points (at the two gate contacts)

td: array of the global numbers of boundary nodes ($N_b \times 1$)

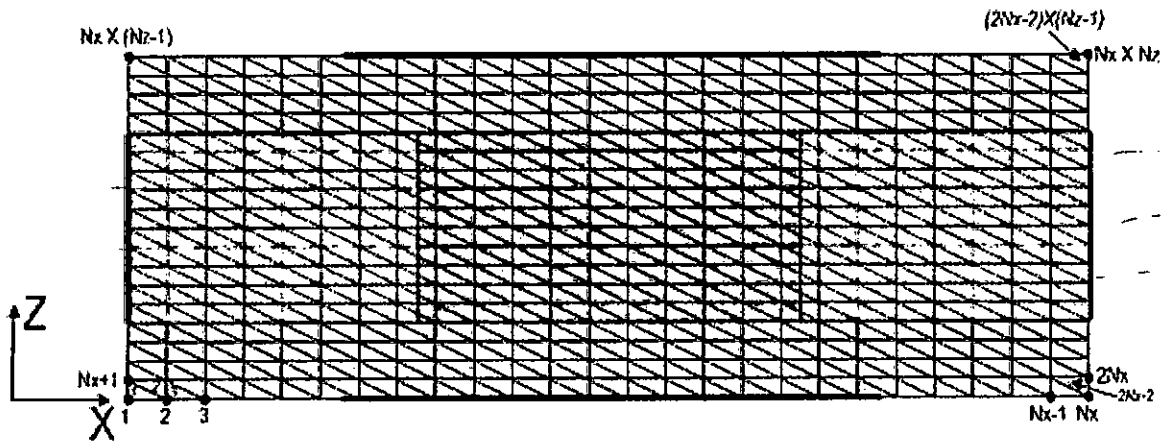


Fig.2.5 Domain discretization of the 2D DG MOSFET structure. Global node numbers and the element numbers are also shown.

All the quantities are self-explanatory, except te . Any node in the domain has a local number in the triangle it is a vertex of (and, of course, the global number). te connects a local number to its associated global number. te has N_e rows, each for a triangle. In a single row, a column contains the global number of the node with the column number as its local number.

2.2.2 Elemental Interpolation-Function

After the discretization is complete, we need to approximate the unknown function ϕ within each element through the use of interpolation functions. For the linear triangular elements, the unknown function ϕ within each element is approximated as

$$\phi^e(x, z) = a^e + b^e x + c^e z \quad (2.48)$$

where a^e , b^e , and c^e are constant coefficients to be determined and e is the element number. The interpolation function for a node has a value of one at its associated node and it linearly goes to zero as we move towards the other two nodes, as shown in Fig. 2.6 below.

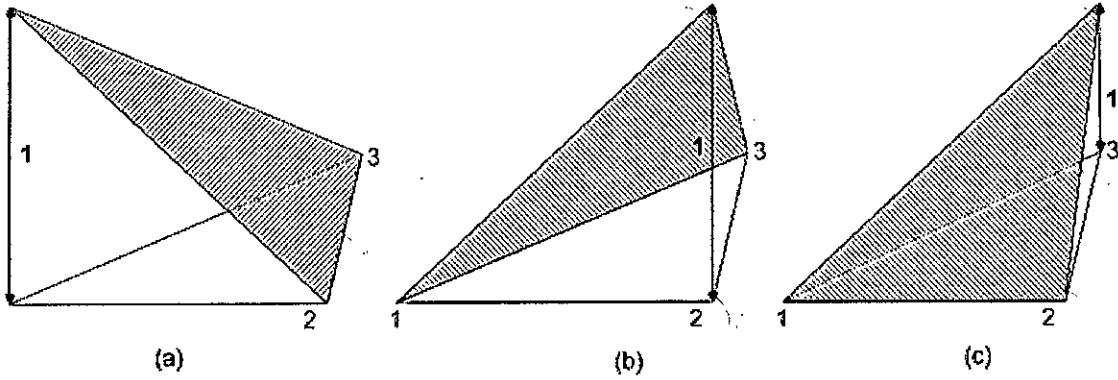


Fig. 2.6 Linear interpolation functions for a triangular element. (a) N_1^e , (b) N_2^e , (c) N_3^e . The planar surfaces of the functions are shaded.

Enforcing Eq. (2.48) at the three nodes, we obtain

$$\phi_1^e(x, z) = a^e + b^e x_1^e + c^e z_1^e$$

$$\phi_2^e(x, z) = a^e + b^e x_2^e + c^e z_2^e$$

$$\phi_3^e(x, z) = a^e + b^e x_3^e + c^e z_3^e$$

where ϕ_j^e is the value of ϕ^e at node j and x_j^e and z_j^e ($j=1,2,3$) denote the co-ordinate values of the j th node in the e th element. Solving for the constants a^e , b^e , and c^e in terms of ϕ_j^e , and substituting them back into Eq. (2.48) yields

$$\phi^e(x, z) = \sum_{j=1}^3 N_j^e(x, z) \phi_j^e \quad (2.49)$$

where $N_j^e(x, z)$ are the interpolation or expansion functions given by

$$N_j^e(x, z) = \frac{1}{2\Delta^e} (a_j^e + b_j^e x + c_j^e z) \quad j=1,2,3 \quad (2.50)$$

in which

$$\begin{aligned} a_1^e &= x_2^e z_3^e - z_2^e x_3^e; & b_1^e &= z_2^e - z_3^e; & c_1^e &= x_3^e - x_2^e \\ a_2^e &= x_3^e z_1^e - z_3^e x_1^e; & b_2^e &= z_3^e - z_1^e; & c_2^e &= x_1^e - x_3^e \\ a_3^e &= x_1^e z_2^e - z_1^e x_2^e; & b_3^e &= z_1^e - z_2^e; & c_3^e &= x_2^e - x_1^e \end{aligned}$$

and

$$\Delta^e = \frac{1}{2} \begin{vmatrix} 1 & x_1^e & z_1^e \\ 1 & x_2^e & z_2^e \\ 1 & x_3^e & z_3^e \end{vmatrix} = \frac{1}{2} (b_1^e c_2^e - b_2^e c_1^e)$$

= area of the e th element.

It can be seen that the interpolation functions have the property

$$N_i^e(x_j^e, z_j^e) = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad (2.51)$$

As a result, at node i , ϕ^e reduces to its nodal value ϕ_i^e . Another point to note is that $N_j^e(x, z)$ vanishes when the observation point (x, z) is on the element side opposite to the j th node. Therefore, the value of ϕ^e at an element side is not related to the value of ϕ at the opposite node but rather it is determined by the values at the two end points of its associated side. This important feature guarantees the continuity of the solution across the element sides.

2.2.3 Formulation of the System of Equations

With the expansion of ϕ given in Eq. (2.49), we use Ritz method to formulate the system of equations. For simplicity, let us first consider the homogeneous Neumann boundary condition. In this case, the functional can be written as

$$F(\phi) = \sum_{e=1}^M F^e(\phi^e) \quad (2.52)$$

where M denotes the total number of elements and F^e is the subfunctional given by

$$F^e(\phi^e) = \frac{1}{2} \iint_{\Omega^e} \left[\varepsilon_x \left(\frac{\partial \phi^e}{\partial x} \right)^2 + \varepsilon_z \left(\frac{\partial \phi^e}{\partial z} \right)^2 \right] d\Omega - \iint_{\Omega^e} f \phi^e d\Omega \quad (2.53)$$

with ϕ^e denoting the domain of the e th element. Introducing the expression in Eq. (2.49) for ϕ^e and differentiating F^e with respect to ϕ^e yields

$$\frac{\partial F^e}{\partial \phi_i^e} = \sum_{j=1}^3 \phi_j^e \iint_{\Omega^e} \left(\varepsilon_x \frac{\partial N_i^e}{\partial x} \frac{\partial N_j^e}{\partial x} + \varepsilon_z \frac{\partial N_i^e}{\partial z} \frac{\partial N_j^e}{\partial z} \right) d\Omega - \iint_{\Omega^e} f N_i^e d\Omega \quad i=1,2,3 \quad (2.54)$$

or

$$\left\{ \frac{\partial F^e}{\partial \phi^e} \right\} = [K^e] \{\phi^e\} - \{b^e\} \quad (2.55)$$

where

$$\left\{ \frac{\partial F^e}{\partial \phi^e} \right\} = \left[\frac{\partial F^e}{\partial \phi_1^e}, \frac{\partial F^e}{\partial \phi_2^e}, \frac{\partial F^e}{\partial \phi_3^e} \right]^T; \quad \{\phi^e\} = [\phi_1^e, \phi_2^e, \phi_3^e]^T$$

The elements of the matrix $[K^e]$ are determined by

$$K_{ij}^e = \iint_{\Omega^e} \left(\varepsilon_x \frac{\partial N_i^e}{\partial x} \frac{\partial N_j^e}{\partial x} + \varepsilon_z \frac{\partial N_i^e}{\partial z} \frac{\partial N_j^e}{\partial z} \right) dx dz \quad i, j = 1, 2, 3 \quad (2.56)$$

while the elements of the vector $\{b^e\}$ by

$$b_i^e = \iint_{\Omega^e} f N_i^e dx dz \quad i = 1, 2, 3 \quad (2.57)$$

It is evident that $[K^e]$ is a symmetric matrix. If ε_x , ε_z and f are constant within each element and equal to ε_x^e , ε_z^e , and f^e , respectively, then

$$K_{ij}^e = \frac{1}{4\Delta^e} (\varepsilon_x^e b_i^e b_j^e + \varepsilon_z^e c_i^e c_j^e) \quad (2.58)$$

$$b_i^e = \frac{\Delta^e}{3} f^e \quad (2.59)$$

2.2.4 Assembly to form the System of Equations

With the elemental Eq. (2.53), we can assemble all M elements, and then impose the stationarity requirement on F to find the system of equations

$$\left\{ \frac{\partial F}{\partial \phi} \right\} = \sum_{e=1}^M \left\{ \frac{\partial F^e}{\partial \phi^e} \right\} = \sum_{e=1}^M ([K^e] \{\phi^e\} - \{b^e\}) = \{0\} \quad (2.60)$$

which can be written compactly as

$$[K] \{\phi\} = \{b\} \quad (2.61)$$

where $[K]$ is assembled from $[K^e]$, and similarly, $\{b\}$ is assembled from $\{b^e\}$:

$$[K] = \sum_{e=1}^M [K^e], \quad \{b\} = \sum_{e=1}^M \{b^e\}. \quad (2.62)$$

Given the connectivity matrix, te , we need to add K_{ij}^e to $K_{te(i,e), te(j,e)}$ and b_i^e to $b_{te(i,e)}$ to find the augmented parameters.

2.2.5 Imposition of Actual Boundary Condition

Up to this point, we assumed Neumann or zero field boundaries on all the sides of the simulated 2D surface. Now before the system of equations is ready to solve, we need to impose the actual boundary conditions of the device. The only modification needed for this is to impose the Dirichlet boundary condition on the Nb boundary points at the two gate contacts. The boundary conditions are shown in Fig. 2.7.

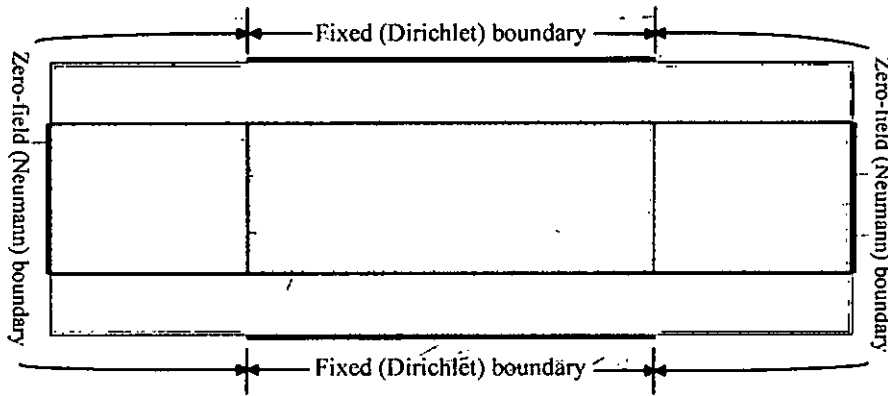


Fig. 2.7 Actual boundary conditions

The reason for using the zero field boundary condition instead of the usually-used Dirichlet boundary condition on the source contact and drain contact is given in [49] where it is shown that in the ballistic simulation if Dirichlet boundary condition is used in these two contacts then the electron density profile becomes sharply peaked at the two ends. To ensure a smooth density at the source and drain contacts the zero field boundary are more appropriate [49]. To impose the Dirichlet boundary condition at the two gate contacts we set

$$b_{td(i)} = p(i), \quad K_{td(i),td(i)} = 1, \quad K_{td(i),j} = 0 \quad \text{for } j \neq td(i) \quad (2.63a)$$

and

$$b_j \leftarrow b_j - K_{j,td(i)} p(i), \quad K_{j,td(i)} = 0 \quad \text{for } j \neq td(i) \quad (2.63b)$$

for $i = 1, 2, 3, \dots, Nd$.

2.2.6 Solution of the System of Equations

After the actual boundary conditions have been imposed Eq. (2.61) is ready to be solved. Standard routines [44] are available for solution of Eq. (2.61).

Results and Discussions

We presented the self-consistent simulation method using FUMS approach in Chapter 2. The results of the calculation for silicon symmetric double-gate nMOSFETs along with relevant discussions are presented in this Chapter. We assume $\langle 100 \rangle$ orientation for silicon and, as stated in the previous Chapter, effective mass approximation has been considered valid. Effective mass approximation has been found accurate in describing the quantization effects of electrons in a MOS inversion layer [50]. The values of different parameters used in the calculation are given in Table 3.1.

Parameter	Value
Electronic charge, q	1.602×10^{-19} C
Planck constant, h	6.626×10^{-34} J-s
Reduced Planck constant, \hbar	1.055×10^{-34} J-s
Free electron mass, m	9.11×10^{-31} kg
Longitudinal effective mass, m_l	0.91
Transverse effective mass, m_t	0.19
Oxide effective mass, m_{ox}	0.5
Permittivity of free space, ϵ_0	8.854×10^{-12} F/m
Relative permittivity of silicon, ϵ_{si}	11.7
Relative permittivity of silicon dioxide, ϵ_{ox}	3.9
Metal workfunction, w_m	4.25 eV
Work function difference between silicon and oxide, χ	3.34 eV
Boltzmann constant, k	1.38×10^{-23} J/K
Effective density of states at conduction band edge, N_C	2.8×10^{25} m ⁻³

Table 3.1: Important physical parameters values used in the simulation

3.1 Eigen-Energies and Eigenfunctions

The schematic diagram of the simulated device is as that is shown in Fig. 2.1. Due to the confinement, the carriers are quantized in the vertical direction and subbands are formed at different energies. In this section, the various eigen energies and the corresponding eigen functions are shown at different conditions. Unless specified otherwise, all the simulations have been done at room temperature (300-K) with n-type source and drain regions doped at 10^{20} cm^{-3} . The channel is intrinsic, and the source/drain-channel junction is abrupt. No gate-to-S/D overlap is assumed. The oxide thickness is 1.5 nm for both top and bottom gates.

Fig. 3.1 shows the subband energies at different vertical slices of the device in the transport or x-direction ($V_g=V_d=0.0\text{V}$). Electron penetration into the oxide is necessary in the nanoscale thickness levels involved and has duly been considered in the calculations, although it is trivial to do the simulation without considering the penetration.

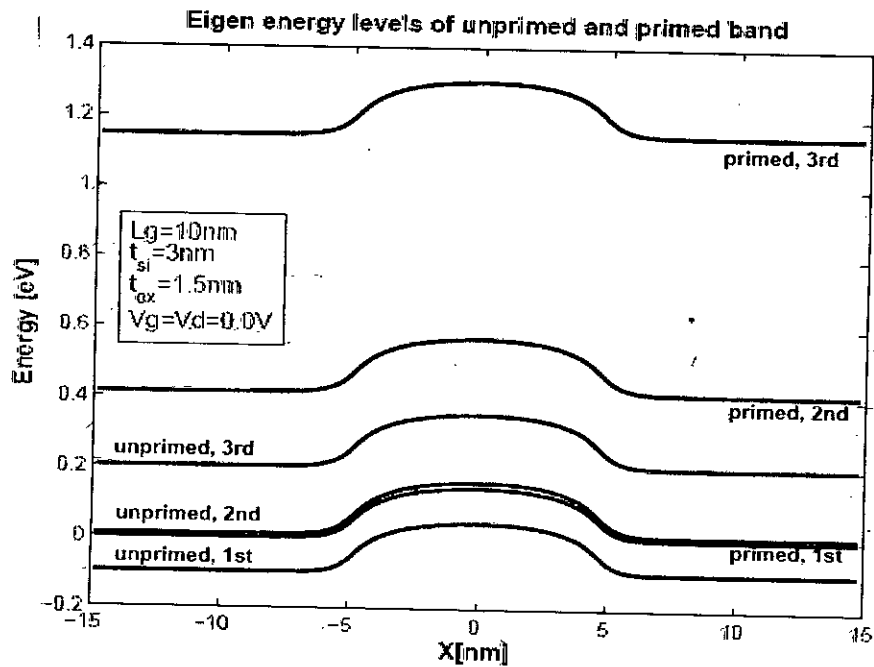


Fig. 3.1 Eigen energy levels along the transport direction (unbiased condition)

As expected at this unbiased condition ($V_g=V_d=0.0V$), the eigen energy levels are symmetric about the middle of the transport direction, and they are higher in the undoped silicon body region. The different simulation parameters are given inside the figure. It is evident from the figure that the unprimed energy level (without symbol) is at a lower energy compared to the corresponding primed level. The reason is that the unprimed valley has higher effective mass in the density of states. As the level number increases so does the energy involved. Now the lower the energy involved the higher the carrier population so higher band will have lower electron concentration. In Fig. 3.1, the third primed energy level is at very high energy levels compared to the other lower levels so that it will contain negligible carrier population compared to the other lower levels and can thus be totally disregarded without affecting the accuracy to any perceptible level but simulation time will be lower.

We will now consider the eigen energy levels at a biased condition: $V_g=V_d=0.3V$, which is typical of the device considered in this work. Note that the energy levels are considerably constant in the source ($x=-5.0\text{ nm to }-15.0\text{ nm}$) and the drain ($x= 5.0\text{ nm to }15.0\text{ nm}$) regions because the electron concentrations are flat in these regions. As we progress along the channel eigen energy levels fall down until they are flat in the drain region. The source to-channel barrier is lower due to V_g and due to V_d the bands bend down as we move towards drain.

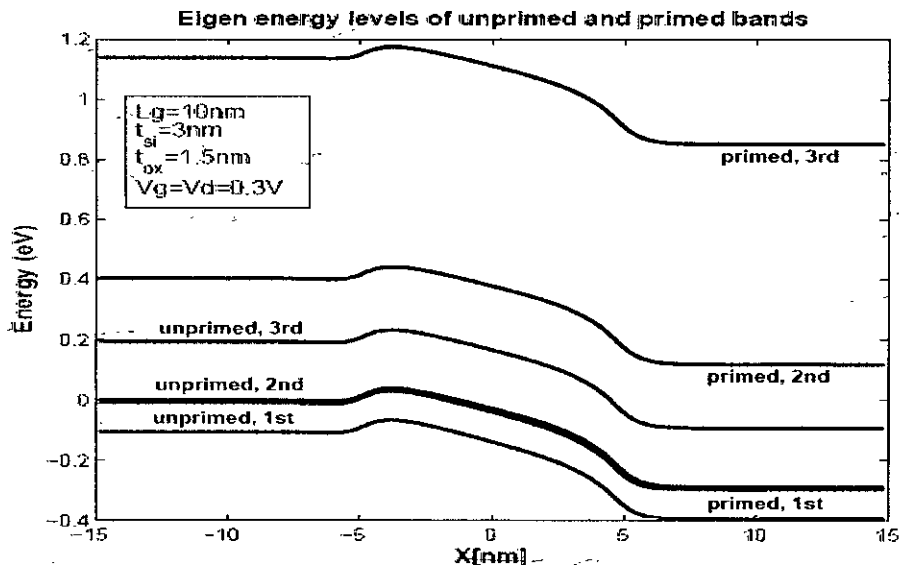


Fig. 3.2 Eigen energy levels along the transport direction (biased condition). From source to drain the eigen energy levels fall although there is an increase near the source-channel contact.

Let us now concentrate on the (unprimed) ground state eigen energy along the channel. Fig. 3.3 shows variation of the ground eigen state along the channel with V_g at a fixed V_d . For a particular V_d , as V_g increases it lowers down the barrier in the channel region so the eigen energy decreases but the eigen energy in the source and drain region remains constant irrespective of the value of V_d .

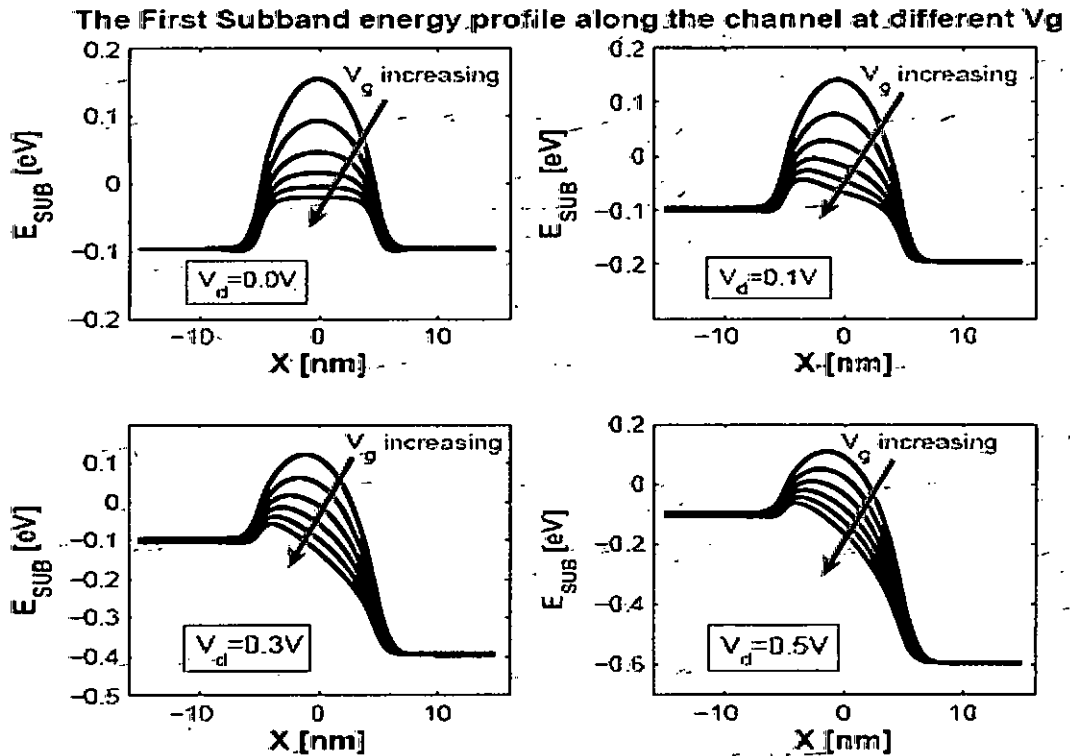


Fig. 3.3 The first (groundstate) eigen energy variation along the channel with V_g . V_d is (clock wise from top left corner) 0.0V, 0.1V, 0.3V, and 0.5V

Fig. 3.4 shows variation of the ground eigen state along the channel with V_d at a fixed V_g . For a particular V_g , as V_d increases it lowers down the barrier in the drain side of channel so the eigen energy decreases in this region. Also clear from the figure is that the eigen energy decreases in the drain region as drain voltage pulls down E_c in drain region but the eigen energy in the source remains constant irrespective of the value of V_d .

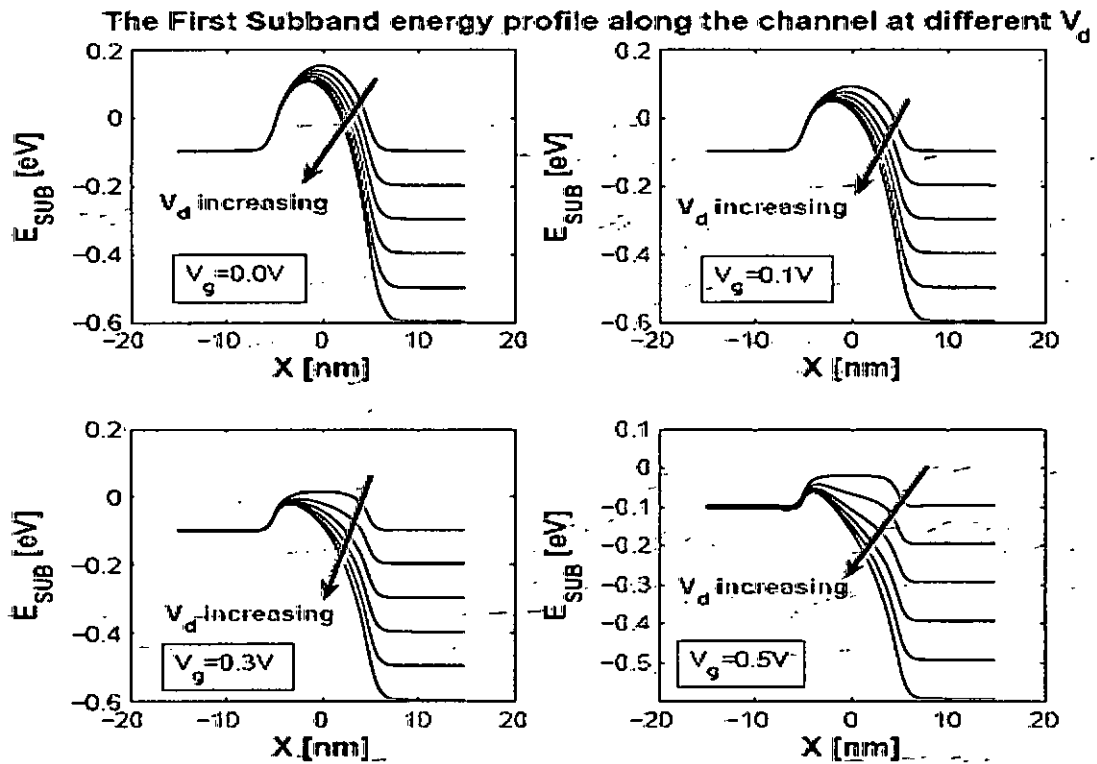


Fig. 3.4 The first (groundstate) eigen energy variation along the channel with V_d . V_g is (clock wise from top left corner) 0.0V, 0.1V, 0.3V, and 0.5V

The squared eigen functions or wavefunctions along a vertical slice (y -direction) at the middle of the channel ($x=0$) of the device at $V_d=V_g=0.0V$ are shown in Fig. 3.5. The area under any function is 1, but the unprimed wavefunctions have higher maxima. Due to the silicon-to-oxide energy barrier, the eigen functions are zero in the oxide regions except near the oxide-silicon interface where because of penetration of electron the wavefunctions are nonzero. As stated earlier, the incorporation of penetration effect in the calculation is necessary for the thickness of oxide involved. We can also clearly see that the penetration effect for primed valley is higher due to the lower density of state mass. Fig 3.6 shows the same thing except that in this case $V_g=V_d=0.3V$. As can be seen from data cursors of the two figures there is a slight drop in amplitude of different wavefunctions as the gate and drain voltage is increased.

Electron wavefunctions of the subbands along the vertical direction

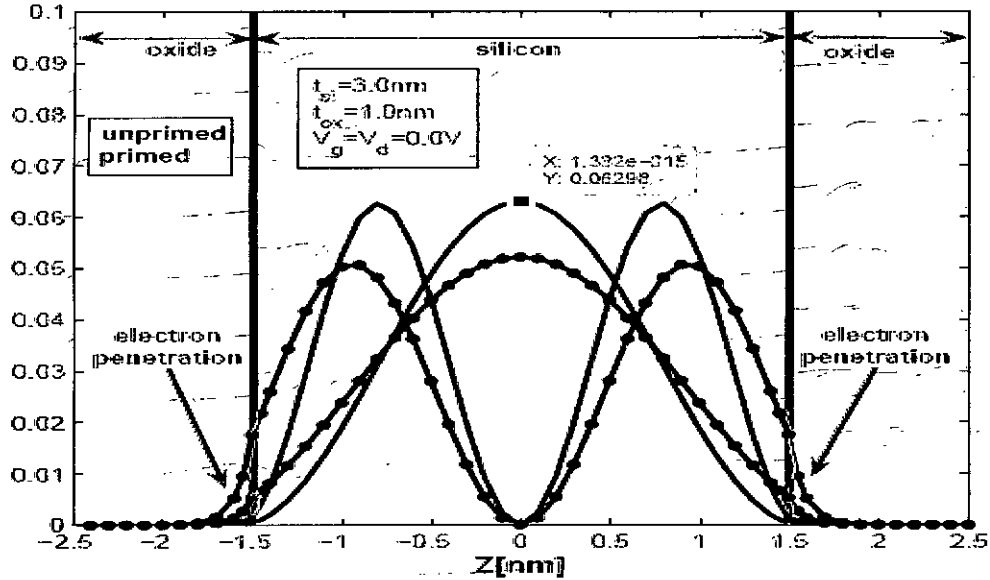


Fig. 3.5 The unprimed and primed eigen functions along z -direction at the middle of the channel. Curves without symbol are for unprimed valley while those with symbol are for primed valleys. $V_g = V_d = 0.0\text{ V}$.

Electron wavefunctions of the subbands along the vertical direction

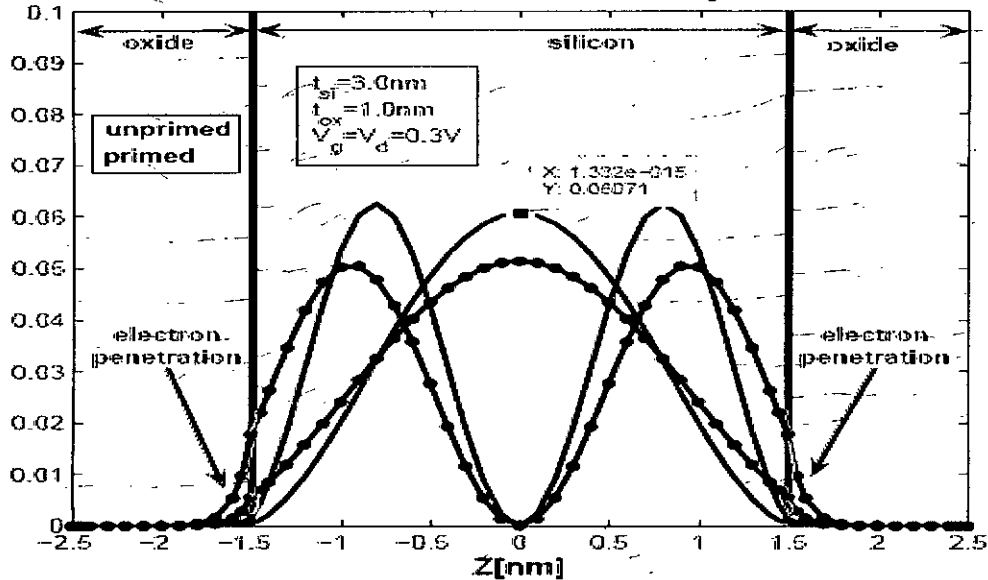


Fig. 3.6 The unprimed and primed eigen functions along z -direction at the middle of the channel. Curves without symbol are for unprimed valley while those with symbol are for primed valleys. $V_g = V_d = 0.3\text{ V}$.

3.2 Electron Density Profile

In this section, we present simulation results for electron density in the device for different stated conditions. In Fig. 3.7, 2D electron density of different subbands along the channel is shown. In this unbiased condition, the density profile should be symmetric about the middle of the channel and this is supported by Fig. 3.11. If we consider the comparative population, then contribution of three low concentration bands may be neglected in the self-consistent calculation.

For the typical biased condition ($V_g=V_d=0.5V$), the concentration profile is plotted in Fig. 3.8. Similar to the unbiased case, the contribution of three low concentration bands may be neglected in this condition too. It is expected that the density profile will not be symmetrical, but concentration at the drain side of the channel is lower due to the presence of a source to drain voltage. This is a direct result of the curves in Fig. 3.3 and Fig. 3.4.

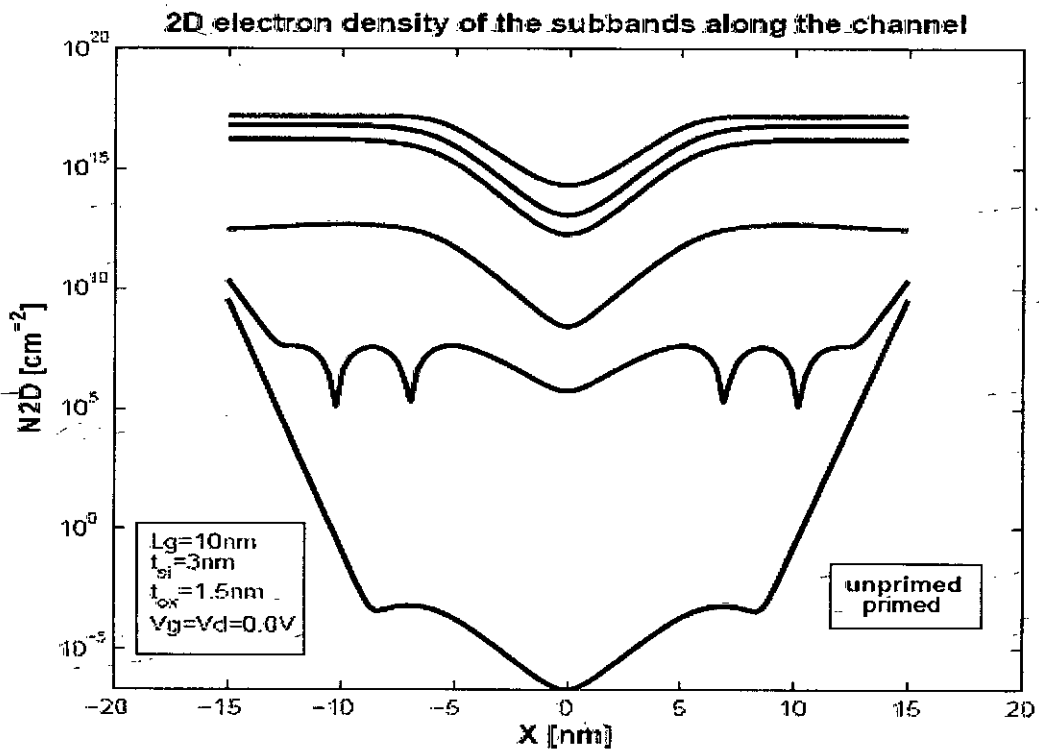


Fig. 3.7 2D electron density of different subbands along the channel (unbiased condition)

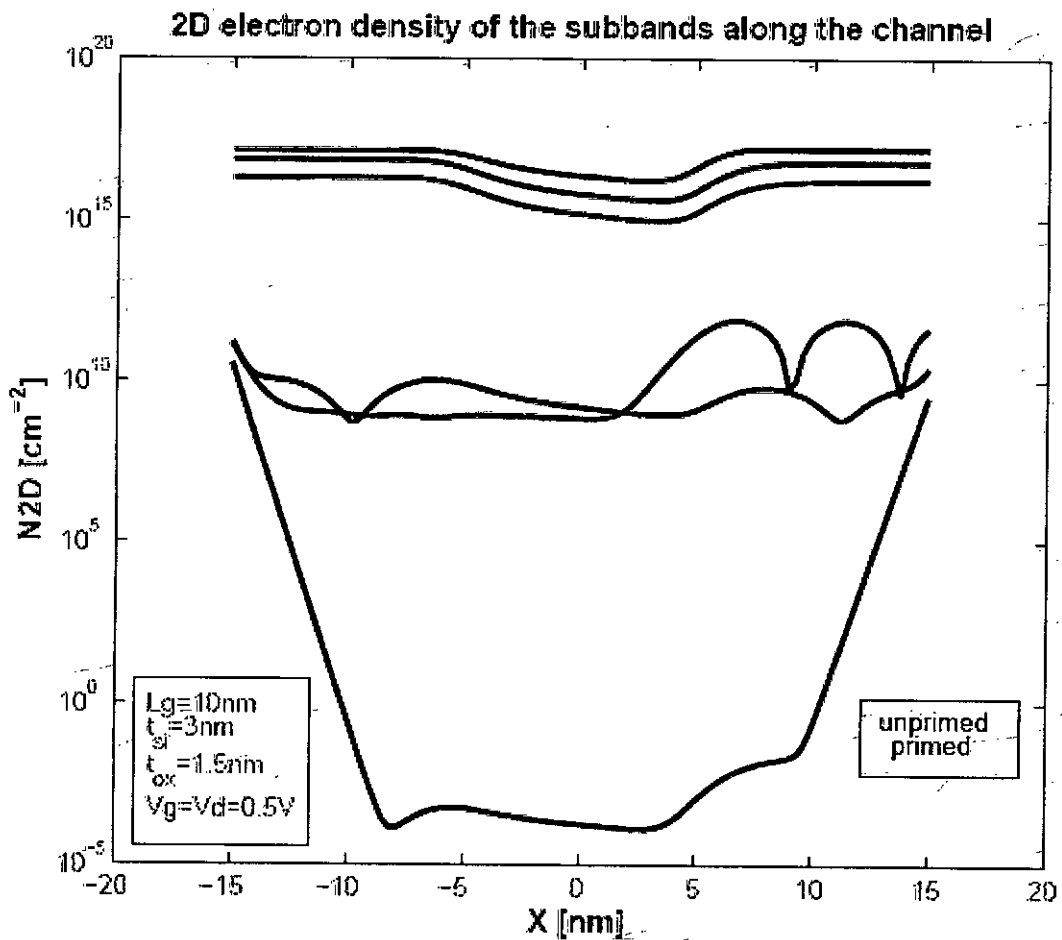


Fig. 3.8 2D electron density of different subbands along the channel (biased condition)

Then Fig. 3.9 shows variation of 2D electron concentration with V_g at two different drain voltages; upper part ($V_d=0.0V$) and lower part ($0.5V$). As V_g increases, the electron density in the channel increases. In the case of $V_d=0.5V$, due to lower channel to gate voltage, concentration is lower in the drain side compared to the source side.

Next variation of 2D charge density with drain voltage is shown for the cases of $V_g=0.0V$ and $V_g=0.5V$ is depicted in Fig. 3.10. As V_d increases, concentration at the drain side of the channel decreases, but concentration at the source remains constant. For higher V_g concentration in the channel is higher as can be seen from a comparison of the upper ($V_g=0.0V$) and lower ($V_g=0.5V$) part of Fig. 3.10.

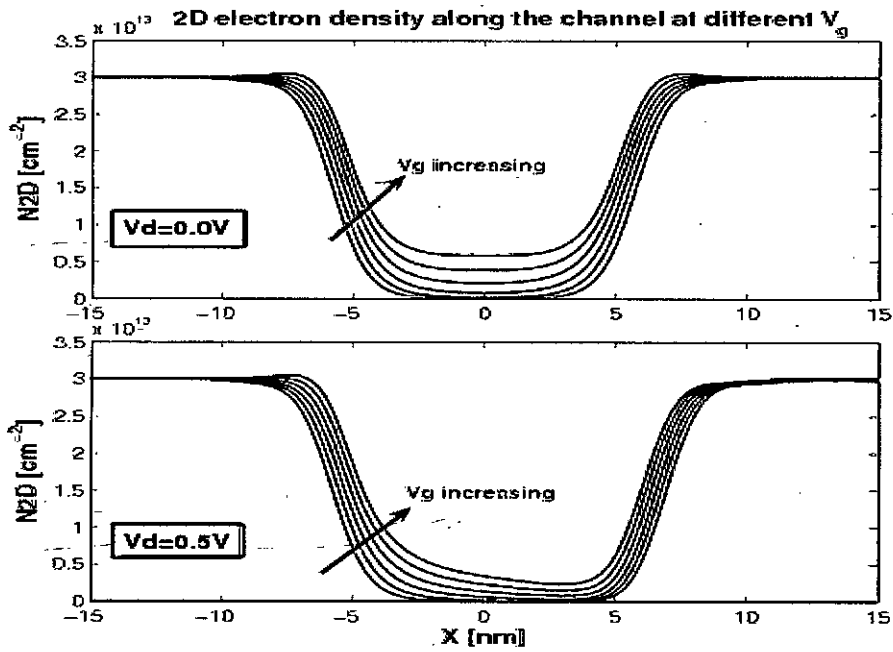


Fig. 3.9 2D electron density variation along the channel with V_g . Upper part for $V_d=0.0V$ and lower part for $V_d=0.5V$.

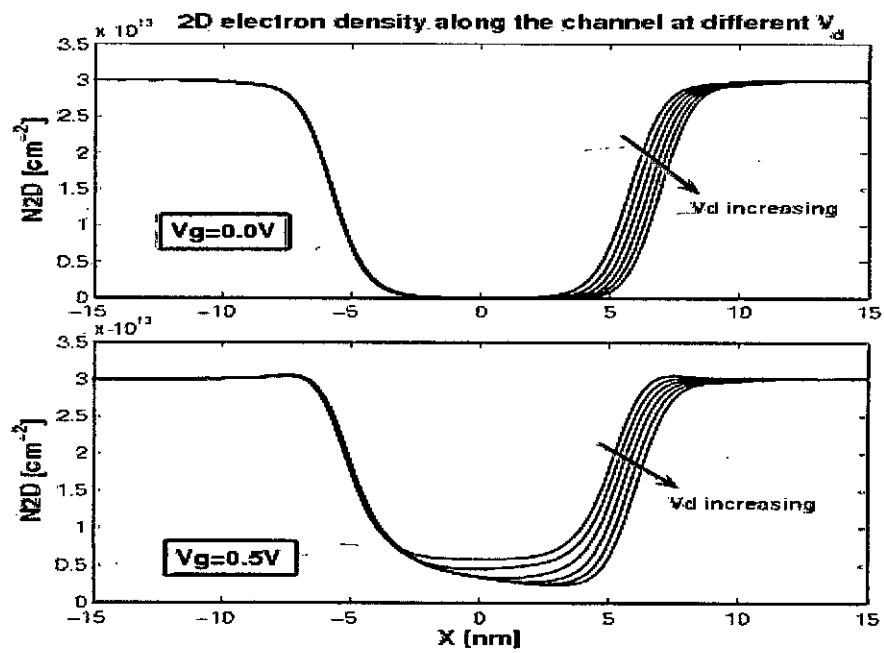


Fig. 3.10 2D electron density variation along the channel with V_d . Upper part for $V_g=0.0V$ and lower part for $V_g=0.5V$.

Finally the 3D electron density variation along with the ground-state wavefunction is shown below for three different device thicknesses in Fig. 3.11. Here, $V_g=V_d=0.5V$. Fig. 3.11(a) shows 3D for $t_{si}=1.5$ nm while Fig. 3.11(b) shows the same for $t_{si}=3.0$ nm. As the thickness increases the density maximum decreases. If increase in t_{si} is continued then the 3D density profile develops a valley in the source and drain regions as depicted in Fig. 3.16(c). Thus at sufficiently large t_{si} there will be two independent channels close to the two oxide-silicon interfaces. But as the thickness is reduced the valley vanishes as the two channels combine.

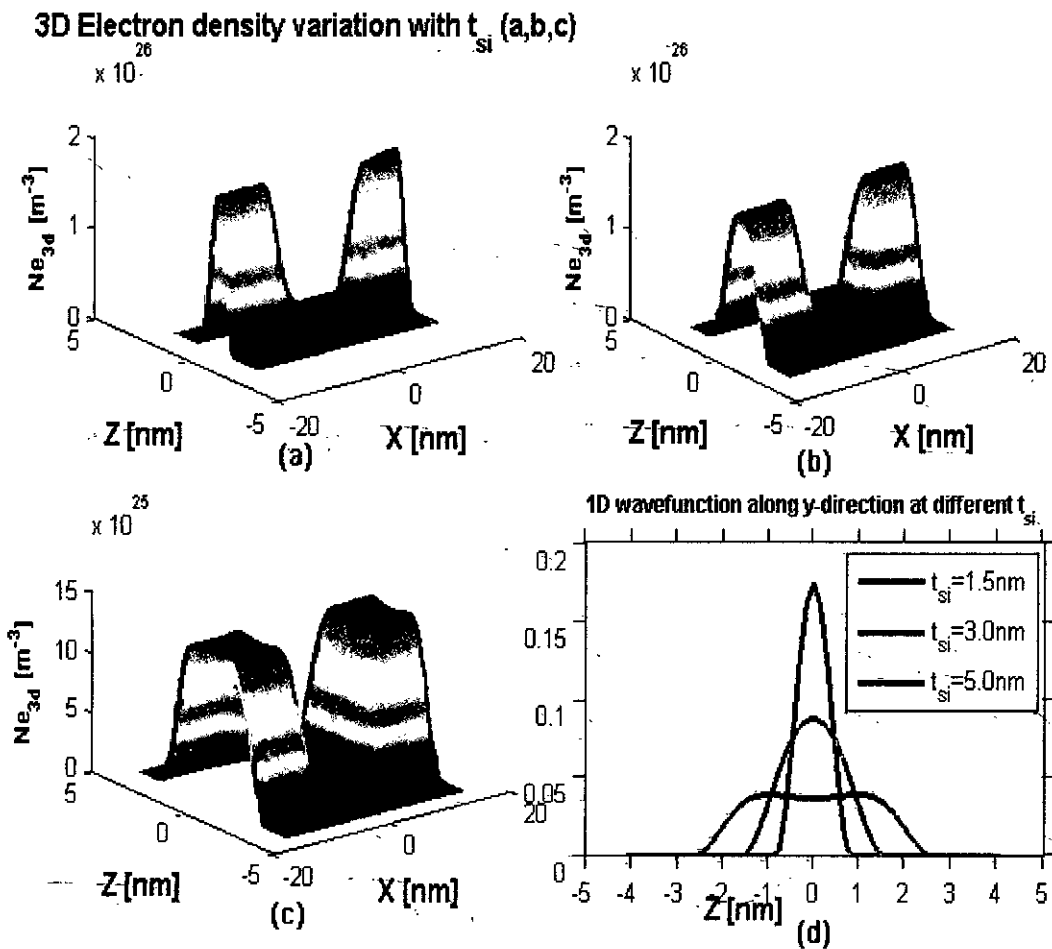


Fig. 3.11 Variation of 3D electron density with silicon thickness, t_{si} . (a), (b), and (c) show the 3D density for t_{si} values 1.5, 3.0, and 5.0 nm, respectively; while (d) shows the ground state eigenfunction along a vertical slice in the middle of the source region in the x-direction.

3.3 3D Conduction Band Edge Profile

In this section the 3D conduction band edge (E_C) profile is pictured for different biasing conditions and silicon thickness. Fig. 3.12 plots the 3D and 1D conduction band edge in the device with silicon thickness $t_{si}=3.0$ nm. Note that the barrier from silicon to oxide is not considered when plotting the data. Fig. 3.12(a) shows the profile for $V_g=0.3V$, $V_d=0.0V$ while Fig. 3.12(b) shows the profile at $V_g=V_d=0.3V$. The self-consistent E_C profile in (a) is symmetric along the lines $x=0$ and $y=0$. E_C is higher at the channel region as the silicon in this region is undoped while it is lower in the doped source and drain regions. For the $V_d=0.3V$ case (b), when moving from source towards the drain through the channel, a small barrier is encountered at the source and channel junction (electrons from the source are reflected by this small barrier). After this barrier, E_C decreases due to V_d . Clearly, electrons from drain face a very high barrier towards the channel and are mostly reflected back. The conduction band is almost flat in the two oxide regions. (c) and (d) show z-directed 1D E_C for conditions (a) and (b), respectively. In the silicon film E_C changes shape very slowly as assumed for UMS/FUMS approach.

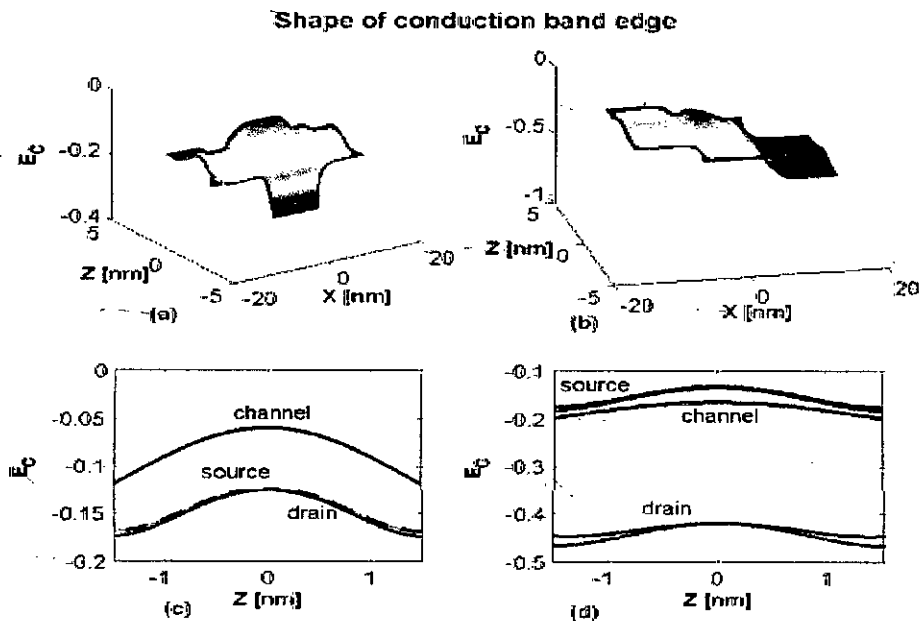


Fig. 3.12 Conduction band profile; (a) 3D, $V_g=0.3V$, $V_d=0V$, (b) 3D, $V_g=0.3V$, $V_d=0.3V$, (c) 1D, $V_g=0.3V$, $V_d=0V$, (d) 1D, $V_g=0.3V$, $V_d=0.3V$

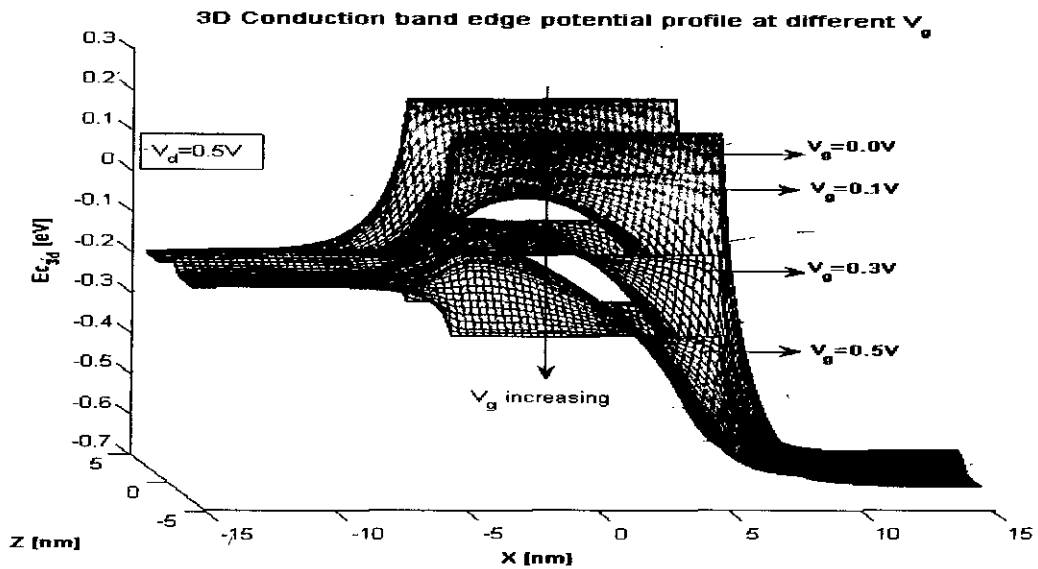


Fig. 3.13 3D conduction band profile at different V_g . $V_d=0.5V$

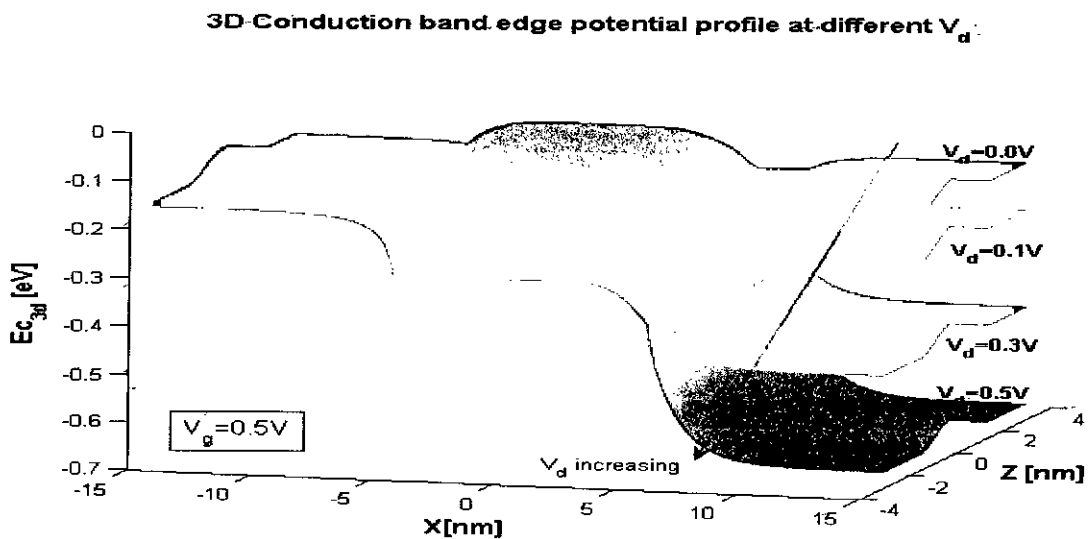


Fig. 3.14 3D conduction band profile at different V_d . $V_g=0.5V$

Variation of E_C with V_g and V_d are shown in Fig. 3.13 and Fig. 3.14, respectively. As V_g is increased with V_d fixed (0.5V), E_C is almost constant in the source and drain region, but it drops down in the channel region (Fig. 3.13). As shown in Fig. 3.14, with increase in V_d for a constant V_g (0.5V), E_C drops down in the drain region (due to increasing V_d); but it is constant in the source and the source side of the channel.

3.4 Drain Current

The transport characteristic of the double gate (DG) MOSFET is discussed in this section. The general I_{DS} - V_{DS} characteristic of a DG MOSFET with $L_g=10.0$ nm and $t_{si}=3.0$ nm is shown in the Fig. 3.15. As the drain voltage increases for a fixed gate voltage the drain to source current increases. For lower V_{DS} the increase is high, but as V_{DS} is increased above 0.2-0.3V the current starts to saturate. For a certain V_{DS} value the current is higher for larger V_{GS} . Saturation is slow at larger V_{GS} compared to lower V_{GS} .

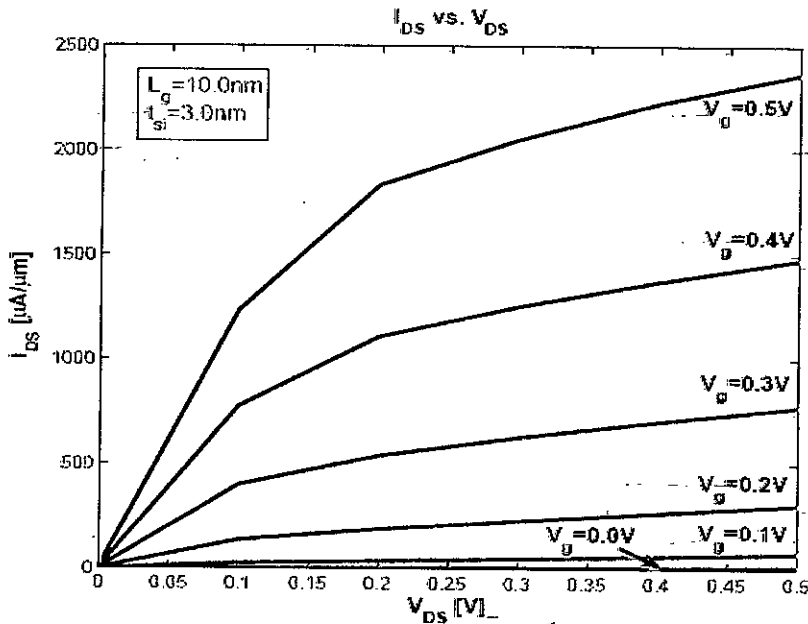


Fig. 3.15 I_{DS} - V_{DS} characteristic of a DG MOSFET with $L_g=10.0$ nm and $t_{si}=3.0$ nm. Current values are per $1\text{-}\mu\text{m}$ width of the device.

The second figure in this section (Fig. 3.16) illustrates the effect of silicon thickness on the I_{DS} - V_{DS} characteristics. As the silicon thickness is increased the drain current increases for fixed V_{DS} and V_{GS} . As t_{si} is increased the channel area increases because the whole of the silicon acts as channel, so larger thickness gives more current. Of course, this is only true for thickness level less than 5 nm. Beyond 5 nm two distinct channel develops and UMS/FUMS is not applicable for $t_{si} > 5$ nm.

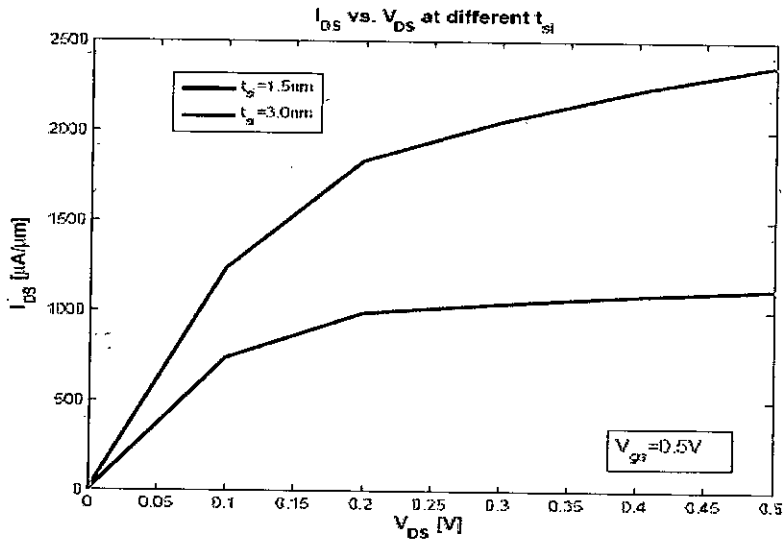


Fig-3.16 I_{DS} (per μm width) variation with silicon thickness (t_{si}). $V_{GS}=0.5V$

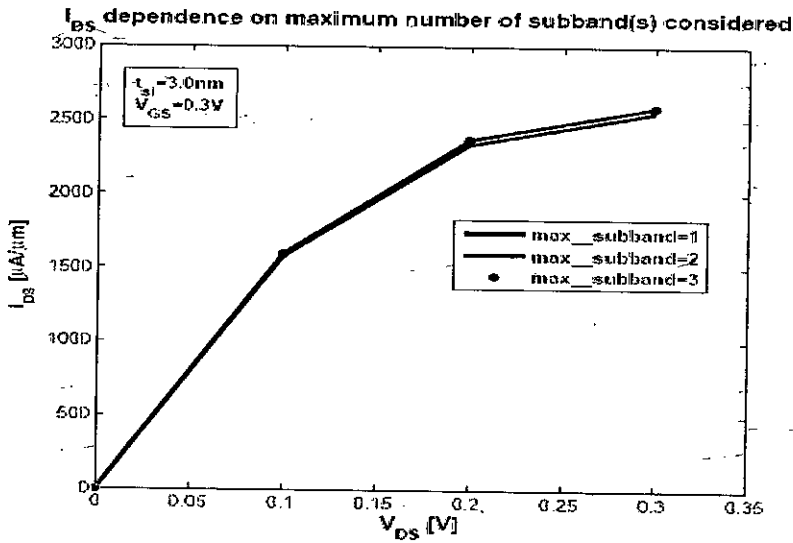


Fig. 3.17 I_{DS} (per μm width) vs. V_{DS} characteristic dependence on maximum number of subbands considered in the calculation

Above the effect of the number of subbands considered on the drain current is shown in Fig. 3.17. When 2 or 3 subbands are considered there is essentially no difference in the drain current and a single subband also gives current values close to when 2 or 3 subbands are considered. Usually 2 subbands are sufficient for the thicknesses involved in this thesis.

3.5 Validity of FUMS approach

In this section we demonstrate the validity of our FUMS approach by comparing its results with those of UMS approach. Fig. 3.18 shows comparison of subband levels and their populations produced by FUMS and UMS approach. The two approaches produce essentially the same results. But the advantage of the FUMS approach is that it requires much less time. Similar result is expected for I_{DS} versus V_{DS} characteristics as is supported by Fig. 3.19.

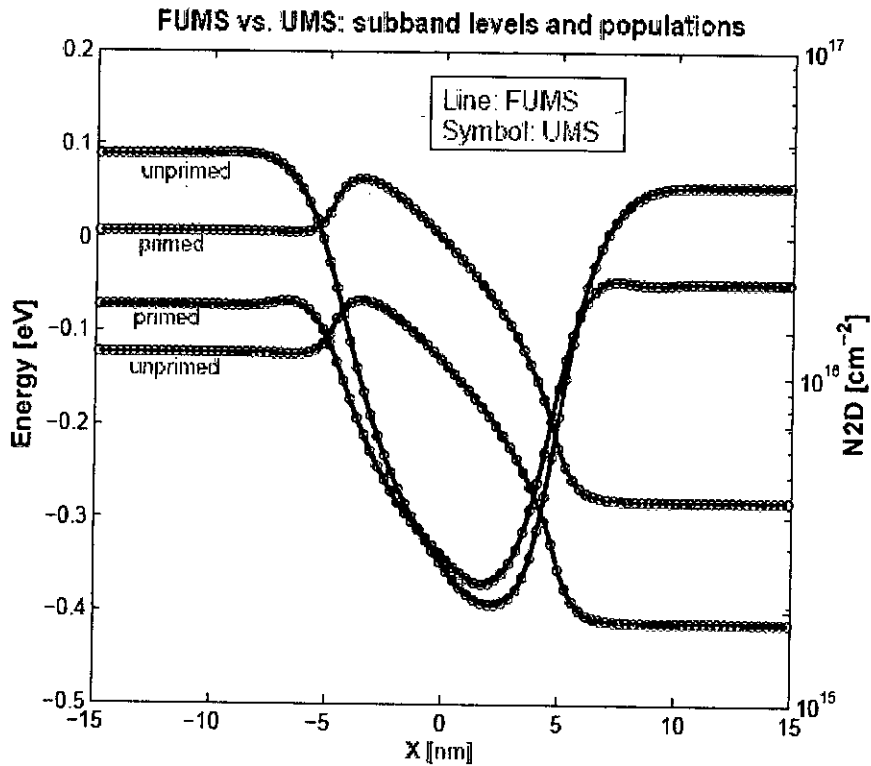


Fig. 3.18 Comparison of subband levels and their populations produced by FUMS and UMS approach. Solid line represents results of FUMS and symbol represents results of UMS.

Thus extensive simulation necessary to exploring novel devices can be made practical by our FUMS approach with FEM taking the responsibility of handling complex geometry.

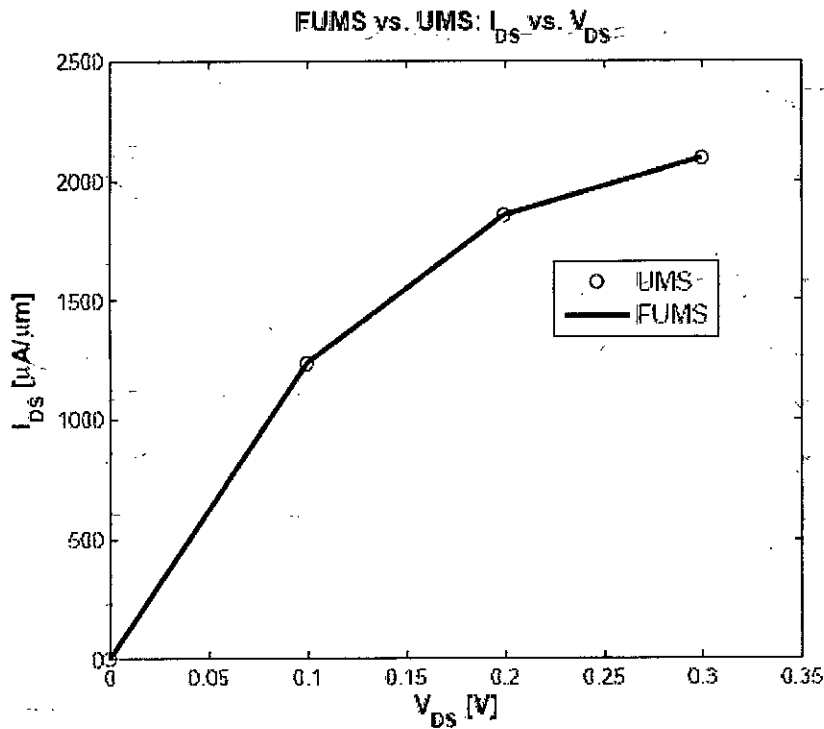


Fig. 3.19 Comparison of I_{DS} (per μm width) vs. V_{DS} profiles produced by FUMS and UMS approach. Solid line represents results of FUMS and symbol represents results of UMS.

3.6 Comparison between femnet and nanomos simulation results

In this section we show the validity of our simulation method by comparing the simulation outcomes of this work, called **femnet** (finite element method nanoelectronic transistor simulator), with similar results of the famous **nanomos** [45] in the nanohub [51]. The theory behind the two simulators is similar but not exactly; **nanomos** uses UMS approach for solving the 1D Schrödinger equation in the vertical direction (z-direction) while femnet uses the novel FUMS approach to reduce the time required to solve the 1D Schrödinger equation. Of course, the two simulators differ greatly in their PDE discretization; **nanomos** uses the simple FDM approach and femnet uses the more appropriate FEM approach. FEM allows easy handling of complex geometry and we can concentrate on a region where any variable can change rapidly by using dense mesh there. Still, the simulators produce similar results as shown in Figs. 3.20-3.22.

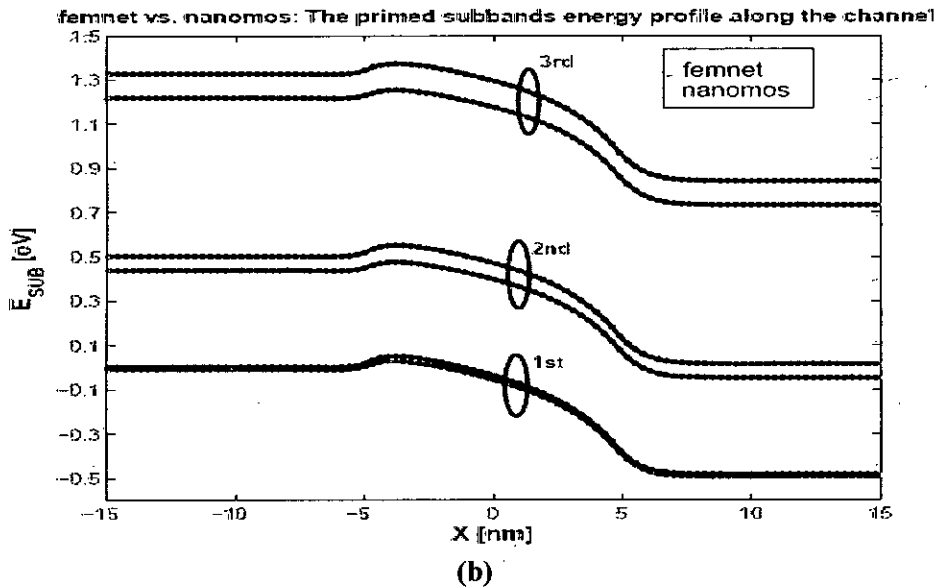
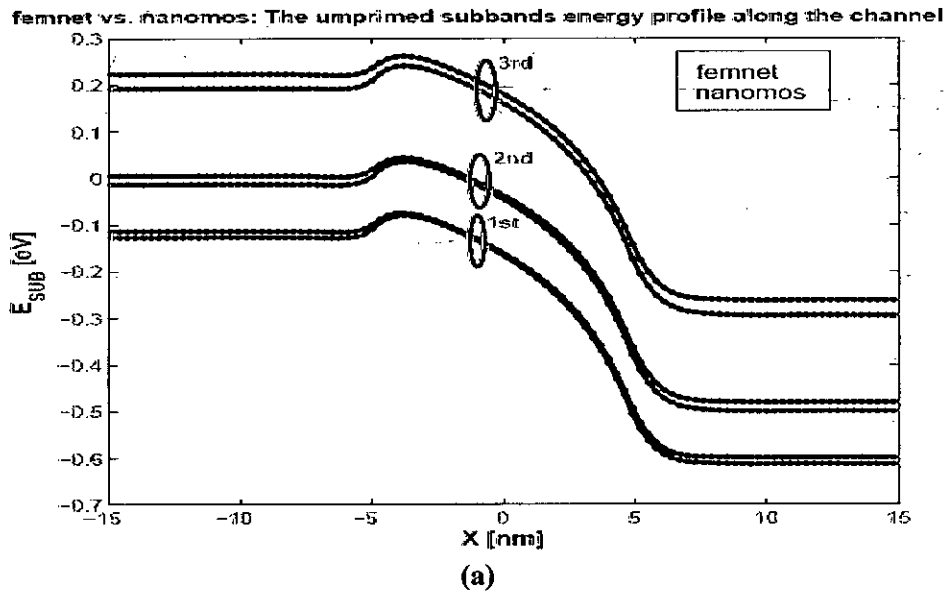
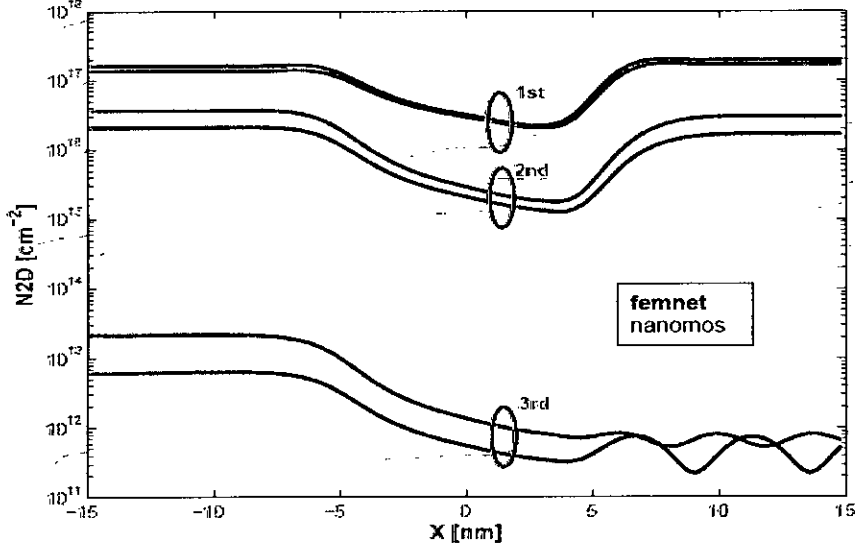


Fig. 3.20 Comparison between eigen energy levels produced by femnet (blue) and **nanomos** (red). (a) unprimed subbands, (b) primed subbands.

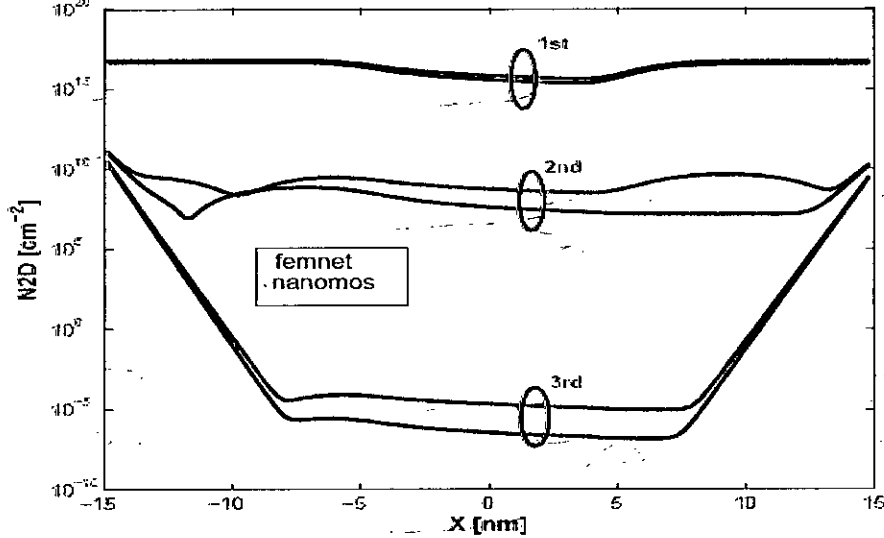
Fig. 3.20 plots the different subband energy levels in the DGMOSFET. The lower levels are very close but the higher levels have greater deviations. For femnet, the source to channel barrier is smaller so greater drain current is expected from FEM simulation.

femnet vs. nanomos: 2D electron density of the unprimed subbands along the channel



(a)

femnet vs. nanomos: 2D electron density of the primed subbands along the channel



(b)

Fig. 3.21 Comparison between 2D subband populations produced by femnet (blue) and nanomos (red). (a) unprimed subbands, (b) primed subbands.

Here in Fig. 3.21, we plot the subband population of different subbands both for femnet and nanomos. The lowest subband population is very close but as the energy increases the difference becomes larger, still general trend is similar.

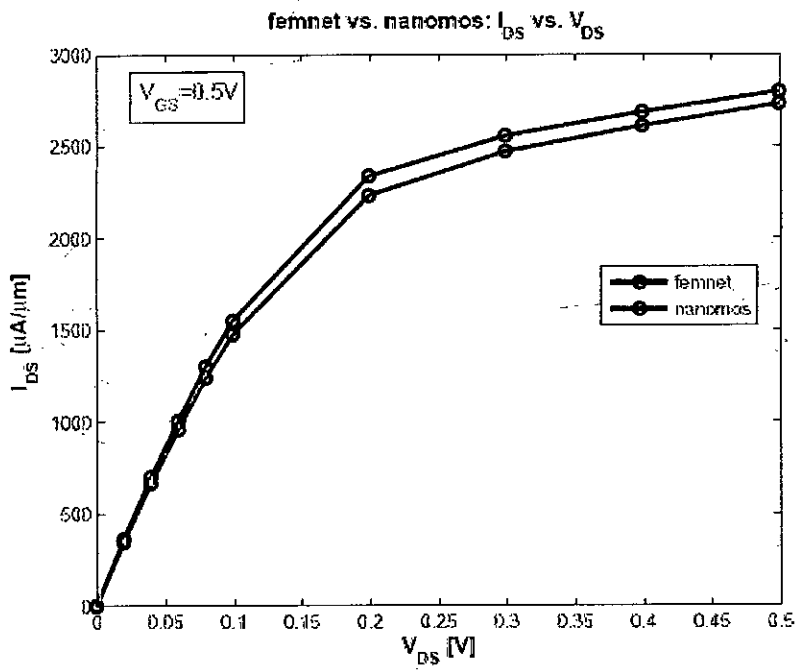


Fig. 3.22 Comparison between drain currents produced by femnet (solid filled circle in blue) and nanomos (solid open circle in red)

We compare the I_{DS} -vs. V_{DS} characteristic for the two simulation approach in Fig. 3.22. As stated earlier, femnet shows larger currents for same V_{GS} although the general behavior is again confirmed to be in agreement.

Conclusion

A modified and computationally efficient simulation approach for DG MOSFETs was presented in the previous Chapter. This Chapter summarizes the work along with directions for further study.

4.1 Summary

This thesis addressed device physics, modeling and design issues of nanoscale double gate MOSFETs at the quantum levels. A double gate MOSFET device with extremely scaled channel lengths (less than 20 nm) and body thicknesses (less than 5 nm) was considered in this study. To accomplish the objectives, a simulation tool was prepared using MATLAB. The fundamental physics equations that were solved include the Poisson equation, which dictates the electrostatics in the devices, and the Schrödinger equation, which describes the transport and distribution of carriers in the devices.

In Chapter 3, we prepared a 2D simulator, termed **femnet**, for nanoscale double-gate MOSFETs. The program solves open-boundary transport problems using non-equilibrium Green's function (NEGF) formalism using fast uncoupled mode space (FUMS) approach. Both NEGF and FEM were employed in the simulation of nanoscale DG MOSFETs. The ballistic DG MOSFETs simulation results of this work were compared with results of a standard simulation tool [45] and the appropriateness of our method and hence the use of FEM was justified. Due to the use of FUMS approach the simulation process is speeded up, but results are essentially similar to those obtained from the use of UMS approach.

Important conclusions are: i) MOSFETs essentially operate as classical devices until the channel length shrinks below 10 nm, when quantum tunneling through the channel barrier becomes significant, limiting device scaling, ii) solving the Green's function in a mode space representation can greatly reduce the size of the problem and provides good accuracy as

compared to full 2D spatial discretization, iii) fast uncoupled mode space approach with FEM makes the method computationally affordable at the PC level.

4.2 Future Work

Further extensions in this study may include:

- 1) The NEGF approach can be used to simulate gate oxide leakage characteristics of 1D MOS structures. Currently, the widely used method in modeling gate leakage is the WKB approximation [52-53]. In this method quantum transmissions through the oxide layers are evaluated approximately in a post-process operation. The electrostatic profile is obtained separately, without considering the effects on the charge distribution of the tunneling current. In addition, incident electrons are represented by plane-waves, so the effects of 2D carriers in the inversion layer on the gate tunneling may not be properly assessed. The NEGF method, however, can exactly solve the transport problem, when coupled to the Poisson equation, providing a way of self-consistently assessing the gate leakage and electrostatics profile. This method also allows us to examine the energy spectrum of the leakage current, which distinguishes the contributions from the 2D discrete states and 3D continuous states.
- 2) Different scattering models are available and scatter can be included in the simulation as it will give the method a real life touch. One method to study the dissipative transport can be the use of the Büttiker probe-based scattering models where scattering centers are treated as reservoirs that change the energy or momentum of the carriers and not the total number of carriers in the system. It has been used in the nanoMOS [45], and recently, in a nanowire simulation [54]. Each scattering center is modeled through a perturbation strength characterized by a position dependent self-energy, which can be mapped onto an equivalent mobility.
- 3) Strained silicon MOSFETs are being studied in recent years. The methods presented in this work can be extended for the study of both uniaxially and biaxially strained DG MOSFETs.

In this case, the simple effective mass approximation might not be true because the electron and hole effective masses change with strain as does the band alignment.

The NEGF approach is a very powerful mathematical tool for addressing how a quantum-state evolves temporally under a varieties of interactions within any tiny system (or quantum level device). As device scaling continues, novel structures/designs must eventually take over the role currently being played by semiconductor-based transistors. Some recent works have brought carbon-tube and molecule-cluster based device structures into focus [55-58]. These new areas provide us plenty of opportunities to apply the NEGF approach at theoretical research levels. In principle, the NEGF formalism not only enables us to understand the microscopic phenomena, but also enable us to exploit the hidden potential. Moreover, it should be noted that this approach is applicable in all non-equilibrium systems beyond electron devices [9, 59].

Bibliography

- [1]. G.E. Moore, "Progress in digital integrated electronics," *IEDM-Tech. Digest*, pp. 11-13, 1975.
- [2]. *International Technology Roadmap for Semiconductors*, Semiconductor Industry Association, 2005.
- [3]. Y. Taur, D. Buchanan, W. Chen, D. Frank, K. Ismail, S.-H. Lo, G. Sai-Halasz, R. Viswanathan, H.-J. C. Wann, S. Wind and H.-S. Wong, "CMOS scaling into the nanometer regime," *Proc. IEEE*, **85**, pp. 468-504, 1997.
- [4]. H.-S. Wong, D. Frank and P. Solomon, "Device Design Considerations for Double-Gate, Ground-Plane, and Single-Gated Ultra-Thin SOI MOSFET's at the 25-nm Channel Length Generation," *IEDM Tech. Digest*, pp. 407-410, 1998.
- [5]. S. A. Mahmood, "Analysis of Double Gate MOSFET through a Non-Charge Sheet Model", M. Sc. Thesis, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, 2006.
- [6]. K. Banoo, *Direct Solution of the Boltzmann Transport Equation in Nanoscale Si Device*, Ph.D. dissertation, Purdue University, West Lafayette, IN, 2000.
- [7]. D.K. Ferry, R. Akis, D. Vasileska, "Quantum effects in MOSFETs: use of an effective potential in 3D-Monte Carlo simulation of ultra-short channel devices," *IEDM Tech. Digest*, pp. 287-290, 2000.
- [8]. L.V. Keldysh, "Quantum transport equations for high electric fields," *Sov. Phys.-JETP* **20**, pp. 1018, 1965.
- [9]. J. Rammer and H. Smith, "Quantum field-theoretical method in transport theory of metals," *Rev. Mod. Phys.* **62**, pp. 323-359, 1986.
- [10]. S. Datta, *Electronic Transport in Mesoscopic Systems*, Cambridge University Press, Cambridge, UK, 1997.
- [11]. B. Davari, R.H. Dennard and G.G. Shahidi, "CMOS scaling for highperformance and low-power-the next ten years," *Proc. IEEE*, **89**, pp. 595-606, 1995.
- [12]. H.-S. Wong, D. Frank and P. Solomon, C. H.-J. Wann and J. Welser, "Nanoscale CMOS," *Proc. IEEE*, **87**, pp. 537-570, 1999.

- [13]. A.A. Abrikosov, L.P. Gorkov and I.E. Dzyaloshinski, *Quantum Field Theoretical Methods in Statistical Physics*, 2nd ed., Pergamon, New York, 1965.
- [14]. D.J. Frank, Y. Taur and H.-S. P. Wong, "Generalized Scale Length for Two-Dimensional Effects in MOSFETs," *IEEE Electron Dev. Lett.*, **10**, pp. 385-387, 1998.
- [15]. Y. Taur and T. Ning, *Fundamentals of VLSI Devices*, Cambridge University Press, Cambridge, UK, 1998.
- [16]. Y.-C. Sun, Y. Taur, R.H. Dennard and S.P. Klepner, "Submicron-Channel CMOS for Low-Temperature Operation," *IEEE Trans. Electron Devices*, **34**, pp. 19-27, 1987.
- [17]. S.J. Wind, D.J. Frank, and H.-S. Wong, "Scaling silicon MOS devices to their limits," *Microelectronic Engineering*, **32**, pp. 271-282, 1996.
- [18]. Y. Taur, C.H. Wann and D. Frank, "25 nm CMOS Design Considerations," *IEDM Tech. Digest*, pp. 789-792, 1998.
- [19]. A. Wei, M.J. Sherony and D.A. Antoniadis, "Effect of Floating-Body Charge on SOI MOSFET Design," *IEEE Trans. Electron Dev.*, **45**, pp. 430-438, 1998.
- [20]. L.T. Su, J.B. Jacobs, J.E. Chung and D.A. Antoniadis, "Deep-Submicrometer Channel Design in Silicon-on-Insulator (SOI) MOSFET's," *IEDM Tech. Digest*, pp. 183-186, 1994.
- [21]. Ren, S. Bourland, S. Lee, J. Denton, M.S. Lundstrom and R. Bashir, "Ultra-thin Body SOI by Controlled Oxidation of Thin Si Membranes," presented at *IEEE Silicon Nanoelectronics Workshop*, Honolulu, Hawaii, June 11-12, 2000.
- [22]. L. Chang, S. Tang, T. King, J. Bokor and C. Hu, "Gate Length Scaling and Threshold Voltage Control of Double-Gate MOSFETs," *IEDM Tech. Digest*, pp. 719-722, 2000.
- [23]. G. Neudeck, T.-C. Su and J. Denton, "Novel Silicon Epitaxy for Advanced MOSFET Devices," *IEDM Tech. Digest*, pp. 169-172, 2000.
- [24]. M.-K. Jeong, E.C. Jones, T. Kanarsky, Z. Ren, O. Dokumaci, R.A. Roy, L. Shi, T. Furukawa, Y. Taur, R.J. Miller, H-S Wong, "Deep-Submicrometer Channel Design in Silicon-on-Insulator (SOI)-MOSFET's," to appear in *IEDM*, 2001.
- [25]. D.J. Frank, S. Laux and M. Fischetti, "Monte Carlo Simulation of a 30 nm Dual-Gate MOSFET: How Short Can Si Go?" *IEDM Tech. Digest*, pp. 553-556, 1992.

- [26]. Z. Ren, R. Venugopal, S. Datta and M.S. Lundstrom, "Examination of Design and Manufacturing Issues in a 10 nm Double Gate MOSFET using Nonequilibrium Green's Function Simulation," in *IEDM-Tech. Digest*, pp. 107-110, 2001.
- [27]. M.S. Lundstrom, *Fundamentals of Carrier Transport*, 2nd ed., Cambridge University Press, Cambridge, UK, 2000.
- [28]. F. Assad, Z. Ren, D. Vasileska, S. Datta, and M.S. Lundstrom, "On the performance limits for Si MOSFET's: A theoretical study," *IEEE Trans. Electron Dev.*, **47**, pp. 232-240, 2000.
- [29]. Y. Taur *et al.*, "An analytical solution to a double-gate MOSFET with undoped body," *IEEE Electron Device Lett.*, vol. 21, pp. 245-247, July 2000.
- [30]. Y. Taur, "Analytic solutions of charge and capacitance in symmetric and asymmetric double-gate MOSFETs," *IEEE Trans. Electron Devices*, vol. 48, p. 2861, Dec. 2001.
- [31]. Y. Taur *et al.*, "A continuous, analytic drain-current model for DG-MOSFETs," *IEEE Electron Device Lett.*, vol. 25, pp. 107-109, Feb. 2004.
- [32]. J. R. Hauser and M. A. Littlejohn, "Approximations for accumulation and inversion space-charge layers in semiconductors," *Solid-State Electronics*, vol. 11, p. 667, 1968.
- [33]. M. Shoji and S. Horiguchi, "Electronic structures and phonon limited electron mobility of double-gate silicon-on-insulator silicon inversion layers," *J. Appl. Phys.*, vol. 85, pp. 2722-2731, 1999.
- [34]. H. C. Pao and C. T. Sah, "Effects of diffusion current on characteristics of metal-oxide (insulator)-semiconductor transistors," *Solid-State Electron.*, vol. 9, p. 927, 1966.
- [35]. S. Datta, "Nanoscale device modeling: the green's function method," *Superlattices and Microstructures*, vol. 28, No. 4, pp. 253-278, 2000.
- [36]. M. P. Anantram, M. S. Lundstrom, and D. E. Nikonov, "Modeling of Nanoscale Devices", arXiv:cond-mat/0610247v2, Feb. 2007.
- [37]. S. Horiguchi, *Physica B* 227, 336, 1996.
- [38]. R. Venugopal, M. Paulsson, S. Goasguen, S. Datta, and M. Lundstrom, "A simple quantum mechanical treatment of scattering in nanoscale transistors," *J. Appl. Phys.*, vol. 93, pp. 5613-5623, 2003.

- [39]. R. Venugopal, Z. Ren, D. Jovanovic, S. Datta, and M. Lundstrom, "Simulating quantum transport in nanoscale mosfets: Real vs. mode space approaches," *J. Appl. Phys.*, vol. 92, pp. 3730-3739, 2002.
- [40]. E. Polizzi and S. Datta, Proceedings of the 2003 third IEEE Conference on Nanotechnology, 12-14 August 2003, p. 40.
- [41]. R. Venugopal, *PhD Thesis, Purdue University, USA, 2003.*
- [42]. P. V. Halen and D. Pulfrey, "Accurate, short-series approximations to fermi-dirac integrals of order $-1/2$, $1/2$, 1 , $3/2$, 2 , $5/2$, 3 , and $7/2$ ", *J. Appl. Phys.*, vol. 57, pp. 5271-5274, 1985.
- [43]. E. Polizzi and N. B. Abdallah, *Physical Review B* 66, 245301, 2002.
- [44]. J. Jin, "The Finite Element Method in Electromagnetics", John Wiley & Sons, Inc., New York, USA, 1993.
- [45]. Z. Ren, R. Venugopal, S. Goasguen, S. Datta and M. Lundstrom, "nanoMOS 2.5: A two dimensional simulator for quantum transport in double-gate MOSFETs", *IEEE Trans. Elec. Devices*, vol. 50, p. 1914, 2003.
- [46]. W.B. Joyce and R.W. Dixon, "Analytic approximations for the Fermi energy of an ideal Fermi gas," *Appl. Phys. Lett.*, 31, pp. 354-356, 1977.
- [47]. M. Cahay, M. McLennan, S. Datta and M. Lundstrom, "Importance of space charge effects in resonant tunneling devices", *Appl. Phys. Lett.* vol. 50, p. 612, 1987.
- [48]. F. Venturi, R. K. Smith, E. C. Sangiorgi, M. R. Pinto, and B. Ricco, "A general purpose device simulator coupling Poisson and Monte Carlo transport with applications to deep submicron MOSFETs," *IEEE Trans. Computer-Aided Design*, vol. 8, p. 360, 1989.
- [49]. Z. Ren, *PhD Thesis, Purdue University, USA, 2001.*
- [50]. S. Jallepalli, J. Bude, W.-K. Shih, M. R. Pinto, C. M. Maziar and A. F. Tasch, "Electron and hole quantization and their impact on deep submicron silicon p- and n-MOSFET characteristics," *IEEE Trans. Electron Devices*, vol. 44, pp. 297-302, 1997.
- [51]. <http://www.nanohub.org/nanomes>
- [52]. J. Cai and C.-T. Sah, "Gate tunneling currents in ultrathin oxide metal-oxide silicon transistors," *J. Appl. Phys.*, 89, pp. 2272-2285, 2001.

- [53]. E.M.Vogel, K.Z. Ahmed, B. Hornung, W.K. Henson, P.K. McLarty, G. Lucovsky, J.R. Hauser and J.J. Wortman, "Modeled Tunnel Currents for High Dielectric Constant Dielectrics," *IEEE Trans. Electron Devices*, **45**, pp. 1350-1354, 1998.
- [54]. J. Wang, M. Lundstrom and E. Polizzi, "A three-dimensional quantum simulation of silicon nanowire transistors with the effective-mass approximation," *J. Appl. Phys.*, **96**, pp. 2192-2203, 2004.
- [55]. N.C. Greenham and R.H. Friend, "Semiconductor Device Physics of Conjugated Polymers", *Solid State Physics* **49**, ed. H. Ehrenreich and F. Spaepen, Academic Press, New York, 1995.
- [56]. C. Dekker, "Carbon Nanotubes as Molecular Quantum Wires," *Phys. Today* **52**, pp. 22-28, May, 1999.
- [57]. M.A. Reed, "Molecular-scale electronics," *Proc. IEEE*, **87**, pp. 652-658, 1999.
- [58]. J. Taylor, H. Guo and J. Wang, "Ab initio modeling of quantum transport properties of molecular electronic devices," *Phys. Rev. B*, **63**, 245407, pp.1-13, 2001.
- [59]. D.C. Langreth and J.W. Wilkins, "Theory of Spin Resonance in Dilute Magnetic Alloys," *Phys. Rev. B*, **6**, pp.3189-3227, 1972.

Appendix A

A.1 2D mesh generation using MATLAB

```
function mesh_dgmos2d
%mesh_dgmos2d - 2d mesh generator function for rectangular DG MOSFET
%This function is used in FEM solution of 2D Poisson equation
%
%Related parameters:
%-----
%Lsd: drain_or_source length
%Lg_top : gate length
%t_top: top oxide thickness
%t_bot: bottom oxide thickness
%t_si: silicon thickness
%dx: x-grid spacing
%dy: y-grid spacing
%ns: no. of smaller grids
%
%Nn: number of nodes
%Ne: number of (triangular) elements
%x : x-coordinate of nodes (next)
%y : y-coordinate of nodes (next)
%te: connectivity matrix (3xne)
%Nb: no. of boundary points
%td: array of global number of Nb's
%-----

global Lsd Lg_top Lg_bot t_top t_bot t_si dx dy ns dys
global Lsda Lg_topa Lg_bota t_topa t_bota t_sia
global Nn Ne Nx Ny Nb x y te td_yrow
global eps_top eps_bot eps_si alpx alpy
%-----

%Set up Nn, Ne, x, y:
%-----
dys=dy/2;
Lsd=round(Lsd/dx)*dx;
Lg_top=round(Lg_top/dx)*dx;
Lg_bot=round(Lg_bot/dx)*dx;
Lx=round((2*Lsd+Lg_top)/dx)*dx;
t_top=round(t_top/dys)*dys;
t_bot=round(t_bot/dys)*dys;
t_si=round(t_si/dys)*dys;
Ly=round((t_top+t_si+t_bot)/dy)*dy;
%
Lsda=round(Lsd/dx);
Lg_topa=round(Lg_top/dx);
Lg_bota=round(Lg_bot/dx);
Nx=round(Lx/dx)+1
t_topa=2*ns+round((t_top-2*ns*dys)/dy);
t_bota=2*ns+round((t_bot-2*ns*dys)/dy);
```

```

t_sia=2*ns+round((t_si-2*ns*dys)/dy);
Ny=6*ns+round((Ly-6*ns*dys)/dy)+1
Nn=Nx*Ny
Ne=2*(Nx-1)*(Ny-1);
x=zeros(Nn,1);
y=zeros(Nn,1);
for i=1:Nx
    xcol(i)=(i-1)*dx;
end
x= repmat(xcol',Ny,1);
for j=1:Ny
    if j<=ns+1
        yrow(j)=(j-1)*dys;
    elseif j>=(t_topa+1-ns+1)&j<=(t_topa+1+ns)
        yrow(j)=yrow(j-1)+dys;
    elseif j>=(t_topa+t_sia+1-ns+1)&j<=(t_topa+t_sia+1+ns)
        yrow(j)=yrow(j-1)+dys;
    elseif j>=(Ny-ns+1)&j<=Ny
        yrow(j)=yrow(j-1)+dys;
    else
        yrow(j)=yrow(j-1)+dy;
    end
end
yrow=yrow';
yd=repmat(yrow,1,Nx);
y=reshape(yd',Nn,1);
%-----

%Set up te:
%-----
te=zeros(3,Ne);
ve1=[1;2;Nx+1];
ve2=[2;Nx+2;Nx+1];
for jj=1:(Ny-1)
    for ii=1:2:(2*(Nx-1))
        ee=(jj-1)*(2*(Nx-1))+ii;
        if ii==1
            te(:,ee)=ve1;
            te(:,ee+1)=ve2;
        else
            te(:,ee)=te(:,ee-2)+1;
            te(:,ee+1)=te(:,ee-1)+1;
        end
    end
    ve1=ve1+Nx;
    ve2=ve2+Nx;
end
%-----

%Set up Nb,td:
%-----
Nb=2*(Lg_topa+1);
td_top=[Lsda+1:Lsda+Lg_topa+1]';
(Nn-(Lsda+Lg_topa+1)+1)';
td_bot=[(Nn-(Lsda+Lg_topa+1)+1):(Nn-(Lsda+1)+1)]';
td=[td_top;td_bot];

```

```

%-----
%Set up alpx and alpy:
%-----
alpx=eps_si*ones(Ne,1);
alpy=eps_si*ones(Ne,1);
alpx(1:(Nx-1)*2*t_topa)=eps_top;
alpx((Nx-1)*2*(t_topa+t_sia):Ne)=eps_bot;
alpy=alpx;
%-----
%-----

```

A.2 Flow chart of the self-consistent simulation

