DEVICE PHYSICS AND SIMULATION OF SILICON NANOWIRE TRANSISTORS

A Thesis

Submitted to the Faculty

of

Purdue University

by

Jing Wang

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2005

To Hua

ACKNOWLEDGMENTS

I would like to express my heartiest gratitude to my thesis advisor, Professor Mark S. Lundstrom, for his invaluable advices and consecutive supports for this research. Professor Lundstrom is not only an outstanding advisor but also a distinguished research leader. He has brought me to the cutting-edge research topic and provided me excellent academic training and precious ideas, which make the whole work possible. His strict attitude in research and teaching, his generosity and considerateness to students and colleagues, and his deep devotion to career have shown me the personalities that a real scientist should possess. Working with Professor Lundstrom is a priceless experience that has turned me from a young student into a professional in this field.

I am also indebted to Professor Supriyo Datta, a well-known device physicist and an extraordinary teacher. His insightful lectures have endowed me with the knowledge of quantum transport and a deep understanding of device physics. This research has greatly benefited from his valuable guidance and suggestions. It is my pleasure to acknowledge Professors Ronald Reifenberger and Kaushik Roy for serving on my Doctoral Advisory Committee and providing valuable comments and suggestions on this work.

I would like to give my sincere appreciations to my colleague, Dr. Eric Polizzi, who has taught me the numerical techniques and offered his help on building the three-dimensional simulation capability. I also acknowledge Professor Gerhard Klimeck and Anisur Rahamn for educating me on the tight binding approach for atomistic bandstructure calculations. My thanks are also given to Dr. Avik Ghosh and Geng-Chiau Liang for their contributions and/or useful suggestions for the atomistic simulations of silicon nanowire transistors, to Dr. Sebastien Goasguen for his assistance in parallel programming, and to Min Chin Chai and Joseph Taylor for their helps on visualization. In addition, I would like to thank Professor Muhammad Ashraful Alam, Dr. Zhibin Ren,

Dr. Ramesh Venugopal, Dr. Jung-Hoon Rhew, Dr. Jing Guo, Sayed Hasan, Jason Besset, Dr. Diego Kienle, Dr. Mani Vaidyanathan, Neophytos Neophytou, Siyu Koswatta, Dr. Titash Rakshit, and other group members for the fruitful discussions with them.

I wish to thank Jin Zhang for helping me understand the stochastic nature of Si/SiO_2 interface roughness and Dr. Paul Solomon at IBM T. J. Watson Research Center for his valuable suggestions and comments on the EQ approach described in Chapter 4.

I would also like to thank the Semiconductor Research Corporation (SRC), the MARCO focus center on Materials, Structures and Devices (MSD), and the NSF Network for Computational Nanotechnology (NCN) for financially supporting this work.

Finally, I want to express my appreciations to my beloved parents, grandparents and wife. However, thoroughly expressing my thankfulness to them is certainly beyond my ability in English writing, and it would either be nearly impossible even if I could do it in my native language.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

ABSTRACT

Wang, Jing. Ph.D., Purdue University, August, 2005. Device Physics and Simulation of Silicon Nanowire Transistors. Major Professor: Mark S. Lundstrom.

As the conventional silicon metal-oxide-semiconductor field-effect transistor (MOSFET) approaches its scaling limits, many novel device structures are being extensively explored. Among them, the silicon nanowire transistor (SNWT) has attracted broad attention from both the semiconductor industry and academia. To understand device physics in depth and to assess the performance limits of SNWTs, simulation is becoming increasingly important. The objectives of this thesis are: 1) to theoretically explore the essential physics of SNWTs (e.g., electrostatics, transport and bandstructure) by performing computer-based simulations, and 2) to assess the performance limits and scaling potentials of SNWTs and to address the SNWT design issues. A full three-dimensional, self-consistent, ballistic SNWT simulator has been developed based on the effective-mass approximation with which we have evaluated the upper performance limits of SNWTs with various cross-sections (i.e., triangular, rectangular and cylindrical). The results show that SNWTs provide better scaling capability than planar MOSFETs. A microscopic, quantum treatment of surface roughness scattering (SRS) in SNWTs has also been accomplished, and it shows that SRS is less important in SNWTs with small diameters than in planar MOSFETs. Finally, bandstructure effects in SNWTs with small diameters have been examined by using an empirical tight binding model, and a channel orientation optimization has been done for both silicon and germanium nanowire field-effect transistors.

1. INTRODUCTION

1.1 Background – The Emerging of Silicon Nanowire Transistors

Integrated Circuit (IC) technology has been regarded as one of the most important inventions in engineering history. The tremendous progress in IC technology in the past four decades has become the driving power of the Information Technology (IT) revolution, which has marvelously changed our lives and the whole world. The secret of the miracle in IC technology is actually simple: scaling down the dimension of each transistor, the basic element of integrated circuits, and increasing the total number of transistors in one IC chip. The device scaling has been successfully predicted by Moore's law [1] – the number of transistors on one IC chip has quadrupled every three years and the feature size of each transistor has shrunk to half of its original value at the same time. To date, microprocessors with >100 million transistors have been realized, and the corresponding metal-oxide-semiconductor field-effect transistor (MOSFET) gate lengths in modern IC chips have entered the sub-100nm regime.

Continued success in device scaling is necessary for maintaining the successive improvements in IC technology. As the MOSFET gate length enters the nanometer regime, however, short channel effects (SCEs) [2], such as threshold voltage (V_T) rolloff and drain-induced-barrier-lowering (DIBL), become increasingly significant, which limits the scaling capability of *planar* bulk or silicon-on-insulator (SOI) MOSFETs. At the same time, the relatively low carrier mobility in silicon (compared with other semiconductors) may also degrade the MOSFET device performance (e.g., ON-current and intrinsic device delay). For these reasons, various novel device structures and materials – silicon nanowire transistors [3] [4], carbon nanotube FETs [5] [6], new channel materials (e.g., strained silicon, pure germanium) [7] [8], molecular transistors

[9], et al. – are being extensively explored. Among all these promising post-CMOS structures, the silicon nanowire transistor (SNWT) has its unique advantage – the SNWT is based on silicon, a material that the semiconductor industry has been working on for over thirty years; it would be really attractive to stay on silicon and also achieve good device metrics that nanoelectronics provides. As a result, the silicon nanowire transistor has obtained broad attention from both the semiconductor industry [10] [11] and academia [3] [12] [13]. According to the fabrication technology, recently reported SNWTs can be categorized into two groups:

1) The first-type SNWTs can be viewed as 'narrow-channel' SOI MOSFETs realized by using a 'top-down' approach [10] [11] [12] [13] [14] [15]. Different from planar SOI FETs, the channel (Si body) widths of SNWTs are lithography-defined and comparable to the Si body thicknesses, so the gate stacks are allowed to wrap around the wire channels to realize multi-gate or gate-all-around FETs, which offer better gate control than planar MOSFETs [4] [13] [16]. In current experimental SNWT structures [10] [11] [12] [13] [14] [15], the wire dimensions (i.e., Si body thickness and width) range from 10nm to 100nm. At the scaling limit, where the device gate length is probably shorter than 10nm [17], this dimension has to be scaled down to the sub-10nm regime to maintain good electrostatic integrity. [4] To do this, very-high-resolution lithography (e.g., <5nm) is required to define the nanowire widths. Therefore, the ultimate scaling of the top-down SNWTs could be limited by the highest resolution of lithography that can be achieved in practice. It should also be noted that the minimum lithography-defined length in the circuits based on the top-down SNWTs should be the SNWT channel (Si body) width instead of the transistor gate length.
2) To avoid very-high-resolution lithography in the SNWT fabrication, a number of experimental groups [3] [18] [19] [20] [21] [22] [23] are trying to synthesize semiconductor (e.g., Si, Ge, GaN) nanowires by using 'bottom-up' approaches, such as the Vapor-Liquid-Solid (VLS) growth technique [19] [22] [23]. With this technology, single-crystal Si nanowires with a diameter as

small as 2-3nm have been achieved. [19] [20] Based on these bottom-up nanowires, various types of devices and circuit components have been experimentally demonstrated, e.g., field-effect transistors (FETs) [3] [18] [24], nanowire heterojunctions [22] [23] [25], logic gates [26], memory [27], decoders [28], bipolar transistors [29], thin-film transistors [30], light emitting diodes (LEDs) [31], lasers [32] [33], photodetectors [34], and nanosensors [35]. For the FET application, in particular, the bottom-up technique offers a possible, low-cost solution to achieve nanowires with ultra-small diameters and relatively smooth interfaces, which are essentially important for scaling the transistor gate length below 10nm.

In brief, the rapid progress in nanofabrication technology has shed light on the potential use of silicon nanowire transistors in future electronics. Consequently, understanding device physics of SNWTs and developing TCAD (Technology Computer Aided Design) tools for SNWT design become increasingly important. The *principle objective* of the thesis is to theoretically explore device physics of silicon nanowire transistors by doing computer-based numerical simulations. With the simulation tools we develop, we will subsequently assess the ultimate performance limits of SNWTs and address the important issues in SNWT device design.

1.2 Overview of the Methodology

In this section, we briefly discuss the simulation techniques we use in this work. The details of the methodology will be discussed in the following chapters.

1.2.1 The self-consistent simulation scheme

Simulation of electron devices normally involves a self-consistent simulation scheme between the electrostatic potential and the charge distribution inside the devices. When a device is coupled to the contacts (electrodes), some charge is transferred into or out of the device (e.g., for the source/drain contacts), or some electric field lines penetrate

into or inject out of the device (e.g., for the gate contact); both effects will result a self-consistent potential $U_{SC}\left(\vec{r}\right)$ (see Fig. 1.1). This potential is called 'self-consistent' because changes in $U_{SC}\left(\vec{r}\right)$ alter the charge density $\rho\left(\vec{r}\right)$ inside the device, which in turn modifies the potential $U_{SC}\left(\vec{r}\right)$ until both the charge density and the potential attain consistent values. To correctly model this process, we need to solve two major equations in our simulations. The first one is Poisson equation [36],

$$\vec{\nabla}\left(\varepsilon\left(\vec{r}\right)\vec{\nabla}U_{SC}\left(\vec{r}\right)\right)=-\rho\left(\vec{r}\right), \tag{1.1}$$

which determines the self-consistent potential $U_{SC}\left(\vec{r}\right)$ for a given charge density $\rho\left(\vec{r}\right)$ (see Fig. 1.1). Here the dielectric coefficient $\varepsilon\left(\vec{r}\right)$ is, in general, position-dependent due to the material transition from one simulated region to another (e.g., from the Si body to SiO_2 layers).

The second one is the transport equation that is solved to obtain the electron (carrier) density $n\left(\vec{r}\right)$ inside the device for a given $U_{SC}\left(\vec{r}\right)$. In the semiclassical context, carrier transport is described by the Boltzmann Transport Equation (BTE) [36]. In BTE, it is assumed that the motion of a single particle (e.g., electron) obeys Newton's second law while the collective behavior of the particle system (e.g., a collection of electrons) is described by statistical mechanics. [36] This assumption works quite well when the device size is relatively large (i.e., much larger than the de Broglie wavelength of electrons). [36] In the nanometer regime, however, the wave-like behavior of electrons becomes substantially significant, so the semiclassical transport equation may not be valid anymore. As a result, a full quantum mechanical transport model, such as the non-equilibrium Green's function (NEGF) approach [37] [38], is necessary.

1.2.2 The NEGF formalism

The NEGF approach is a widely used method for nanoscale device simulation. [39] [40] [41] [42] A tutorial description of this formalism is available in [38]. Here we

simply review the basic concepts and procedure that we need for the SNWT simulation. Within the NEGF formalism, the device is represented by a Hamiltonian, *H*, which is coupled to two infinite reservoirs, the source (S) and drain (D). The S/D reservoirs are characterized by their respective Fermi levels, μ_S and μ_D, which are determined by the applied voltage biases. Coupling between the active device and the S/D contacts can be described by introducing the self-energy matrices, Σ_S and Σ_D. Incoherent carrier transport (due to scattering) inside the device can also be captured by the self-energy (Σ_{Scat}) method. Once *H*, Σ_S, Σ_D, μ_S, μ_D and Σ_{Scat} are obtained, the electron density matrix and transmission coefficient at a given energy can be evaluated [38], and then the electron density and terminal currents are computed by doing numerical integrations over the energy space (see Fig. 1.1). (To be concise in this section, we refer readers to [37] and [38] for the details of the NEGF formalism.)

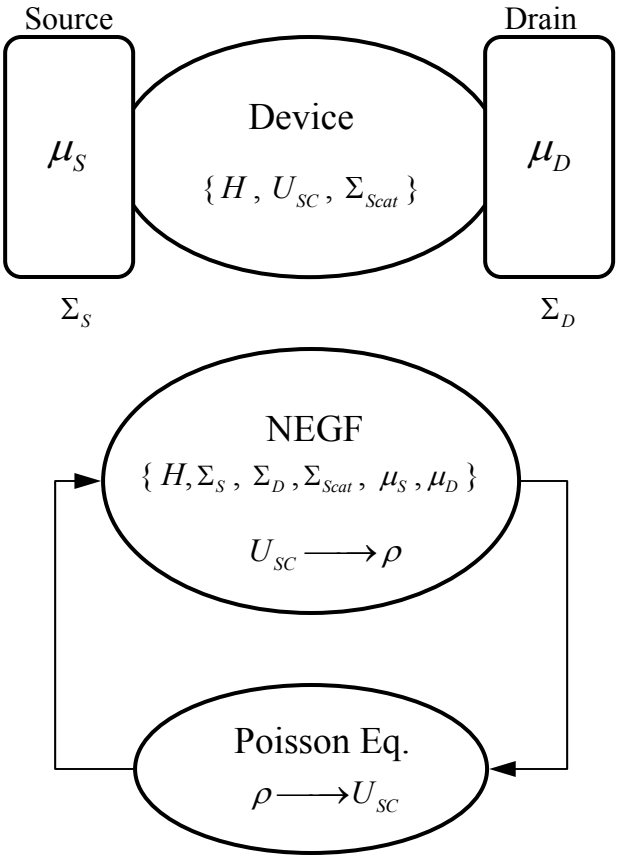


Fig. 1.1 Schematic Structure of a device coupled to the source/drain contacts and the self-consistent scheme that is used in the SNWT simulation.

To discretize the operators, H, Σ_S , Σ_D , and Σ_{Scat} , the effective-mass (EM) approximation [43] [44] is adopted. With the EM approximation, the Si and SiO_2 regions in SNWTs are treated as continuous materials, which allows a real space discretization of the operators based on the Finite Difference Method (FDM) or the Finite Element Method (FEM). [44] [45] [46] [47] To reduce the huge computational complexity in the three-dimensional (3D) real space simulation, we utilize a coupled/uncoupled mode-space representation, which greatly minimizes the size of the operator matrices (i.e., H, Σ_S , Σ_D , Σ_{Scat} , et al.) while keeping very high simulation accuracy. [44] [45] [47] This allows us to generate a practical, full 3D, quantum mechanical simulator that can be used for SNWT simulation and design. The details of this quantum mechanical SNWT simulator will be discussed in Chapter 3.

Within the NEGF formalism, the self-consistent procedure for a quantum mechanical simulation of SNWTs consists of the following steps:

1) Given a particular SNWT structure, we first select an appropriate method (e.g., FDM or FEM) to discretize all the operators.
2) To start the self-consistent loop, we need to give a guess value of $U_{SC}\left(\vec{r}\right)$. This guess value may be obtained from a semiclassical simulation (e.g., a ballistic solution of BTE [48]).
3) For a given $U_{SC}\left(\vec{r}\right)$, we can write down the device Hamiltonian, H, and then calculate the contact self-energy matrices, Σ_S and Σ_D , and the self-energy matrix, Σ_{Scat} , that represents incoherent carrier transport in SNWTs.
4) With all the information obtained in the steps above, the retarded Green's function is evaluated and then the density matrix is computed.
5) Knowing the electron density inside the SNWT, a 3D Poisson equation is solved for the self-consistent potential, $U_{SC}\left(\vec{r}\right)$.
6) Steps 3) – 5) are iterated until both the computed self-consistent potential and the electron density converge.

7) With the converged self-consistent potential and density matrix, the terminal currents of SNWTs are calculated.

1.2.3 Atomistic simulations of SNWTs

Due to the two-dimensional (2D) quantum confinement, the bulk crystal symmetry is not preserved in silicon nanowires any more. For this reason, atomistic bandstructure effects are expected to be important in nanowires with small diameters. In this thesis, we also explore the impact of bandstructure effects on SNWT device characteristics by performing atomistic simulations. Instead of doing a full 3D, atomistic simulation within the NEGF formalism, which is discouraged by its huge computational burden, we perform the atomistic simulations in the following way. First, we calculate the energy dispersion (*E-k*) relations of silicon nanowires by using a nearest-neighbor $sp^3d^5s^*$ tight binding approach [49] [50] [51]. Based on the computed *E-k* relations, the current-voltage characteristics of the corresponding SNWTs are then evaluated with a semi-numerical, ballistic FET model, named 'FETToy' [52] [53]. (The FETToy model will be introduced in Chapter 2.) By doing this, the bandstructure effects on SNWTs with arbitrary wire orientation and cross-sectional shapes can be investigated. The details of these atomistic simulations of SNWTs will be found in Chapter 6.

1.3 Outline of the Thesis

This thesis is divided into the following chapters:

1) Before we perform the detailed quantum mechanical simulation of SNWTs, a simple, analytical theory of ballistic SNWTs is presented in Chapter 2. This model is derived by modifying an analytical approach ('FETToy') proposed by A. Rahman et al. [52] for ballistic planar MOSFETs and then extended by J. Wang et al. [54] for ballistic high electron mobility transistors (HEMTs). The model treats 3D electrostatics, captures the finite semiconductor capacitance effect (or the so called 'quantum capacitance limit' [55] [56]) and can be used

to include detailed, numerical bandstructures of SNWTs (see Chapter 6). The results of this chapter illustrate the essential physics and peculiarities of one-dimensional (1D) nanowire FETs.

2) In Chapter 3, we present a self-consistent, full 3D quantum simulation of SNWTs based on the effective-mass approximation [43] [44]. First, we introduce the coupled/uncoupled mode space approach, which significantly reduces the computational expense while maintaining great accuracy as compared with the full 3D real space representation. Then we present the simulation results of ballistic SNWTs with various cross-sections (e.g., triangular, rectangular and cylindrical). Within the NEGF framework shown in this chapter, scattering in SNWTs can be phenomenologically treated by a simple model, so called the 'Büttiker probes' [57], which was previously adopted in MOSFET simulations [42]. The details of this method will be shown in Appendix.
3) Chapter 4 discusses the performance limits and scaling potential of ballistic SNWTs. It consists of three different topics. Sec. 4.1 shows a comparison between the upper performance limit of SNWTs with that of the planar double-gate MOSFET. [4] In Sec. 4.2, we propose a general approach to compare planar vs. non-planar (nanowire) FETs with the consideration of both Electrostatic integrity (gate control) and Quantum confinement (so called the 'EQ approach'). [58] Sec. 4.3 introduces a conceptual study of the channel material optimization for both planar MOSFETs and nanowire FETs based on the effective-mass approximation. [59]
4) In Chapter 5, we perform a microscopic simulation of surface roughness scattering (SRS) in SNWTs. [60] The transport model we use is the coupled mode space approach that has been introduced in Chapter 3. By using the 3D finite element method, the microscopic structure of the Si/SiO_2 interface roughness [61] [62] is directly implemented. The results show that SRS behaves quite differently in a SNWT as compared with a conventional planar

MOSFET, and SRS in a 1D SNWT is not as significant as that in a 2D MOSFET due to the reduction of density-of-states in the SNWT channel.

5) Chapter 6 discusses atomistic simulations of Si and Ge nanowire FETs. [63] [64] Based on a nearest-neighbor $sp^3d^5s^*$ tight binding approach [49] [50] [51], we develop a simulator that can calculate *E-k* relations for unrelaxed Si and Ge nanowires with arbitrary wire orientations and cross-sectional shapes. With the calculated *E-k* relations, current-voltage characteristics of various Si and Ge nanowire FETs are computed by using the FETToy model, introduced in Chapter 2. The impact of bandstructure effects on SNWT performance is investigated and the channel orientation optimization for Si and Ge nanowire FETs is performed. Finally, the validity of the widely-used parabolic effective-mass approximation for current-voltage calculation of n-type SNWTs is examined.
6) Chapter 7 summarizes the thesis and proposes possible directions for future research.

2. ANALYTICAL THEORY OF BALLISTIC NANOWIRE TRANSISTORS

This chapter describes an analytical theory of ballistic silicon nanowire transistors (SNWTs). The model is derived by modifying an analytical approach proposed by A. Rahman et al. for ballistic planar MOSFETs [52] and extended by J. Wang et al. [54] for ballistic high electron mobility transistors (HEMTs). Although detailed numerical simulations of SNWTs will be performed in the following chapters, a simple analytical theory still has its special uses and importance:

1) The analytical theory may explicitly demonstrate the dependence of device metrics (e.g., ON-current, transconductance) on device parameters (e.g., oxide capacitance, electron effective-mass) and applied voltage biases, which helps to understand the essential physics of one-dimensional (1D) nanowire FETs and to interpret numerical simulation results.
2) The simple model is highly efficient compared with three-dimensional (3D) numerical simulators, so it can be used to obtain a quick estimation of the ballistic performance limit for a given SNWT structure.
3) Although the calculations in this chapter are all based on the effective-mass approximation [43] [44] (an ellipsoidal parabolic energy band is assumed), this analytical approach can also be extended to compute the ballistic currents with arbitrary *E-k* diagrams [52]. Therefore, it will play an important role when evaluating the bandstructure effects on SNWT performance (Chapter 6).

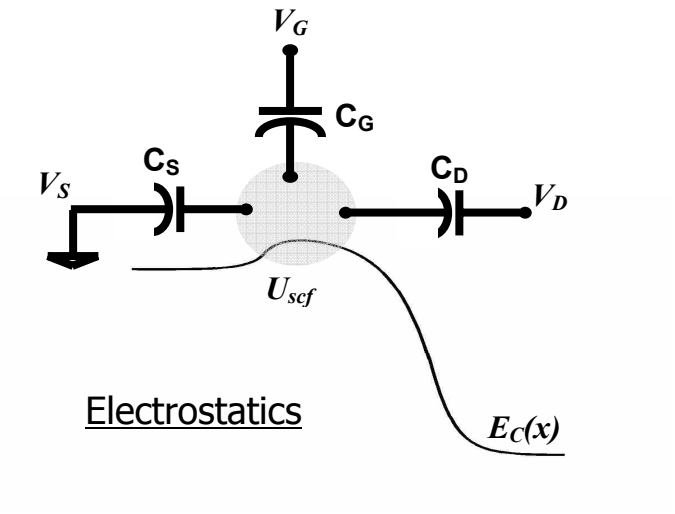
This chapter is divided into the following sections: Sec. 2.1 describes the methodology and basic assumptions, Sec. 2.2 shows the simulation results and discusses the peculiarities of 1D nanowire FETs, and Sec. 2.3 summarizes the chapter.

2.1 Methodology

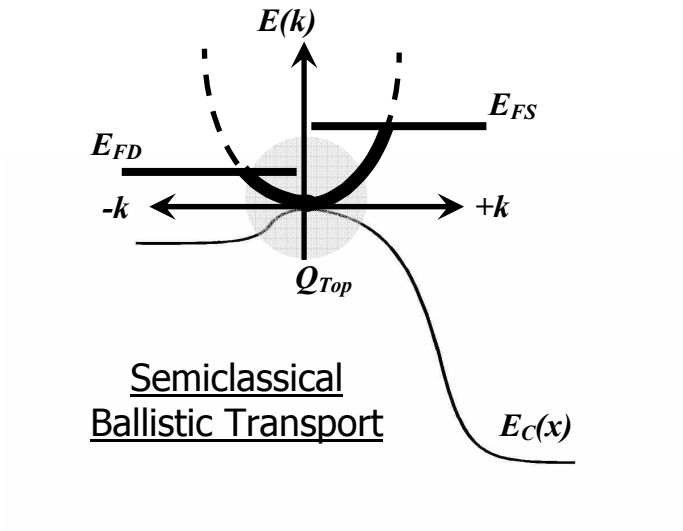
In [52], the authors proposed a general ballistic FET model (named 'FETToy' [53]) that correctly captures quantum confinement, two-dimensional (2D) electrostatics, and bias-charge self-consistency in ballistic FETs. It generalizes Natori's model [65] by treating 2D electrostatics and by properly treating the 1D electrostatics – even in the quantum capacitance limit, where the gate insulator capacitance is much greater than the semiconductor (or quantum) capacitance [55] [56]. Fig. 2.1 summarizes the essential aspects of this model. It consists of three capacitors, C_G, C_S, and C_D, which describe the electrostatic couplings between the top of the barrier and the gate, the source and the drain, respectively. [52] The potential at the top of the barrier is obtained as

$$U_{scf} = \left(\frac{C_G}{C_G + C_D + C_S}\right)V_G + \left(\frac{C_D}{C_G + C_D + C_S}\right)V_D + \left(\frac{C_S}{C_G + C_D + C_S}\right)V_S + \frac{Q_{Top}}{\left(C_G + C_D + C_S\right)}, \tag{2.1}$$

where V_G, V_S, and V_D are the applied biases at the gate, the source and the drain, respectively, and Q_{Top} is the mobile charge at the top of the barrier, which is determined by U_{scf}, the source and drain Fermi levels (E_{FS} and E_{FD}) and the *E-k* relation for the channel material. To be specific, the group velocity of each state is calculated from the tabulated *E-k* data of the channel materials, and the carrier density is then evaluated by assuming that the states with a positive (negative) group velocity are in equilibrium with the source (drain) reservoir. (For channel materials with a parabolic energy band, the $+k$ states always obtain a positive group velocity, while the $-k$ states acquire a negative one.) The process of computing the potential at the top of the barrier from Eq. (2.1), then updating the mobile charge there continues until convergence is achieved after which the drain current is readily evaluated from the known populations of all the states in the energy band. For a detailed discussion of the model and the equations for the calculation of the mobile charge and electronic current for a planar double-gate MOSFET, please refer to [52].



(a)



(b)

Fig. 2.1 Illustration of the essential features of the analytical ballistic model. (a) The 3-capacitor model to treat 2D (3D) electrostatics in nanoscale FETs. (b) The semiclassical ballistic transport model to calculate the electron current and the mobile (electron) charge at the top of the barrier.

In this work, we modify this model to simulate ballistic nanowire transistors. Compared with a planar double-gate MOSFET, a nanowire FET has two major differences that need to be considered in this simulation:

1) For a planar double-gate MOSFET, the gate oxide capacitance, C_G , is analytically obtained as

$$C_G = \frac{2\kappa\varepsilon_0}{T_{ox}}, \tag{2.2}$$

where κ is the oxide dielectric constant, ε_0 is the permittivity of vacuum and T_{ox} is the oxide thickness. For a nanowire FET, however, the gate oxide capacitance does not have an analytical expression in general, so it should be numerically computed by solving a 2D Poisson equation at the cross-section of the SNWT. In this section, we assume a coaxial gate geometry, for which the gate oxide capacitance, C_G , can be analytically obtained as [66]

$$C_G = \frac{2\pi\kappa\varepsilon_0}{\ln\left(\frac{2T_{ox}+T_{Si}}{T_{Si}}\right)}, \tag{2.3}$$

where T_{Si} is the diameter of the silicon body.

2) Due to the 1D *E-k* relation for a nanowire FET, the equations for the SNWT charge density and current are different from those for 2D planar MOSFETs. [52] To be specific, the mobile charge at the top of the barrier is obtained as

$$Q_{Top} = -q\left(n^+ + n^-\right) = -\frac{qN_{1D}}{2}\Im_{-1/2}\left(\eta_F\right) - \frac{qN_{1D}}{2}\Im_{-1/2}\left(\eta_F - U_D\right), \tag{2.4}$$

where q is the electron charge and

$$N_{1D} = M\sqrt{\frac{2k_BTm_x^*}{\pi\hbar^2}}, \tag{2.5}$$

here m_x^* is the electron effective-mass in the transport direction, $\hbar$ is Plank constant, k_B is Boltzmann constant, T is the ambient temperature and M is the valley degeneracy (for a cylindrical SNWT with a $[100]$ oriented channel, $M = 4$ and $m_x^* = 0.19m_e$, where m_e is the free electron mass). η_F in Eq. (2.4) is defined as

$$\eta_F = \frac{\mu_S - \left(\varepsilon(0) - qU_{scf}\right)}{k_B T}, \tag{2.6}$$

where μ_S is the source Fermi level and $\varepsilon(0)$ represents the lowest subband level at the top of the barrier when $U_{scf} = 0$. U_D in Eq. (2.4) is defined by the following equation

$$U_D = \frac{V_{DS}}{k_B T}, \tag{2.7}$$

where $V_{DS} = V_D - V_S$ is the applied drain bias. The function $\Im_j(\eta)$ is so called Fermi integral [36], which is defined as

$$\Im_j(\eta) = \frac{1}{\Gamma(j+1)} \int_0^\infty \frac{x^j dx}{1 + \exp(x - \eta)}. \tag{2.8}$$

Similarly, the electron current for a nanowire FET can be analytically expressed as

$$I = I^+ - I^- = M \frac{qk_B T}{\pi\hbar} \left[\Im_0(\eta_F) - \Im_0(\eta_F - U_D) \right] = M \frac{qk_B T}{\pi\hbar} \ln\left(\frac{1 + e^{\eta_F}}{1 + e^{\eta_F - U_D}} \right). \tag{2.9}$$

The self-consistent simulation scheme of this simple analytical model is as follows. We start with a guess solution of U_{scf}, and the mobile charge, Q_{Top}, is thus evaluated from Eq. (2.4). Then the computed Q_{Top} is fed back into Eq. (2.1) to obtain an updated U_{scf}. This process is iterated until a converged U_{scf} is achieved. After that, the electron current of the device is evaluated from Eq. (2.9).

Finally, let us briefly discuss several basic assumptions we have made in this simple analytical approach:

1) We assume a coaxial gate geometry and neglect the quantum confinement so that an analytical expression of the gate oxide capacitance is obtained. In general, to evaluate the gate oxide capacitance for an arbitrary SNWT structure with the consideration of quantum confinement, which keeps the electron charge centroid somehow away from the Si/SiO_2 interfaces, a 2D Poisson equation needs to be numerically solved together with a 2D Schrödinger equation. (To be concise, it is not shown here.)

2) In this section, only one subband at each valley is included when evaluating electron charge and current. This is a sound approximation when the Si body thickness (diameter) of the nanowire is sufficiently small (e.g., <2nm). For nanowire FETs with thicker bodies, multiple subbands should be considered. As described in [54], this simple analytical model can be easily extended to include multi-subbands in the calculation of mobile charge density and terminal currents.
3) As mentioned earlier, this simple analytical model is based on a semiclassical ballistic transport model, in which the quantum mechanical tunneling from the source to the drain is not considered. According to [17], source-to-drain tunneling may not be important when the channel length of the FET is >8nm, especially for the ON-state. Therefore, the semiclassical ballistic transport model is well acceptable in this simple analytical simulation.

2.2 Simulation Results and Essential Physics of 1D Nanowire FETs

In this section, we simulate an idealized, ballistic nanowire FET by using the simple analytical approach described in Sec. 2.1. A coaxial gate geometry configuration is assumed and the channel is $[100]$ oriented, so the valley degeneracy is $M = 4$ (i.e., the four unprimed valleys, $[010]$, $[0\bar{1}0]$, $[001]$, and $[00\bar{1}]$, are degenerate) and the longitudinal effective-mass is $m_x^* = 0.19 m_e$. A hypothetically thin silicon body, $T_{Si} = 1\text{nm}$, is selected to guarantee that only the lowest subband at each valley is occupied. Moreover, we assume the oxide layer thickness is $T_{ox} = 1\text{nm}$ and two oxide dielectric constants, κ, are used in the simulation: 1) $\kappa = 3.9$ is for a SiO_2 layer and 2) $\kappa = 80$ is for a hypothetical high-K insulator (e.g., ZrO_2), which is selected to illustrate the full-degenerate and quantum capacitance effects in nanowire FETs. To capture the 3D electrostatics in the simulated nanowire FETs, we assume that

$$\alpha_G = \frac{C_G}{C_G + C_D + C_S} = 0.88, \tag{2.10}$$

and

$$\alpha_D = \frac{C_D}{C_G + C_D + C_S} = 0.035 . \tag{2.11}$$

A) *I-V Characteristics:*

Figure 2.2 plots the (a) $\log(I_{DS})$ vs. V_{GS} transfer characteristics and (b) I_{DS} vs. V_{DS} characteristics of the simulated SNWT with a SiO_2 layer ($\kappa = 3.9$). Fig. 2.3 is the same plot for the SNWT with a high-K insulator layer ($\kappa = 80$). It is shown that for $\kappa = 3.9$, the channel conductance increases with gate voltage, while for $\kappa = 80$, it saturates when the gate voltage is sufficiently high (this effect is also clearly illustrated in Fig. 2.3a – the current saturates under low drain bias when the gate voltage is larger than 0.5V). To explain this, we first need to obtain an expression for the nanowire FET current under low drain bias. From Eq. (2.9), under low drain bias ($U_D \ll 1$ or $U_D \ll |\eta_F|$),

$$I = M \frac{qk_BT}{\pi\hbar} \frac{\partial \Im_0(\eta_F)}{\partial \eta_F} U_D = M \frac{qk_BT}{\pi\hbar} \frac{V_{DS}/(k_BT/q)}{1+e^{-\eta_F}} = M \frac{2q^2}{h} \frac{V_{DS}}{1+e^{-\eta_F}} . \tag{2.12}$$

So the channel conductance [2] of a nanowire FET is obtained as

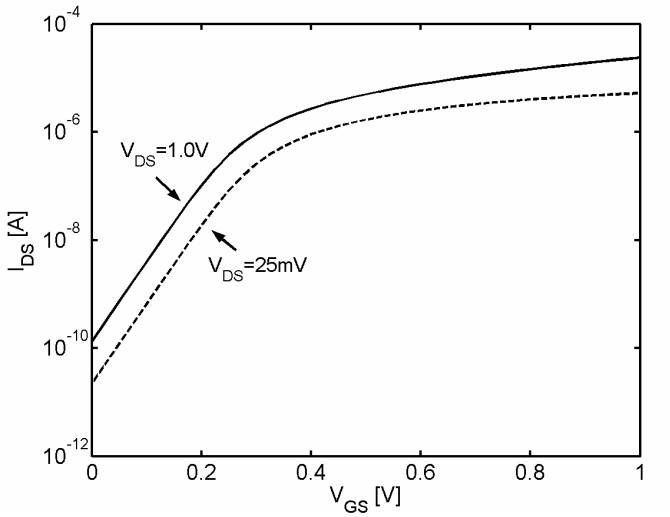
$$g_d = M \frac{2q^2}{h} \frac{1}{1+e^{-\eta_F}} . \tag{2.13a}$$

At the non-degenerate condition, $\eta_F < 0$ and $|\eta_F| \gg 1$ (i.e., the source Fermi level is well below the top of the barrier), $g_d \simeq M(2q^2/h)e^{\eta_F}$. As the gate voltage increases, η_F is raised and the channel conductance increases. But once the full-degenerate condition, $\eta_F \gg 1$ (i.e., the source Fermi level is well above the top of the barrier), is satisfied, the channel conductance saturates to a fixed value (independent of any device parameters),

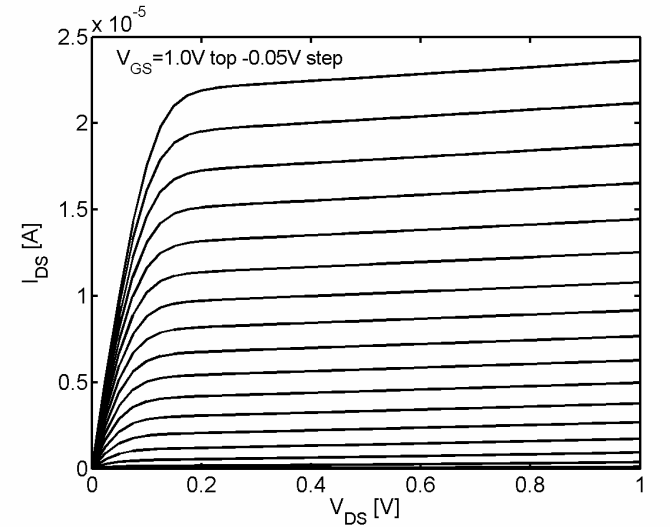
$$g_d = M\left(2q^2/h\right), \quad \text{(full-degenerate)} \tag{2.13b}$$

which is so called quantum conductance [37]. For the SNWT with a high-K insulator layer ($\kappa = 80$), due to the large gate capacitance, the top of the barrier can be efficiently lowered by increasing the gate voltage. Consequently, the device easily enters the full-degenerate regime when the gate bias is large than 0.5V. For the SNWT with a SiO_2

layer ($\kappa = 3.9$), however, the gate capacitance is significantly lower than that for $\kappa = 80$ and the full-degenerate condition can not be achieved at the bias range (0V-1V) used in this simulation.

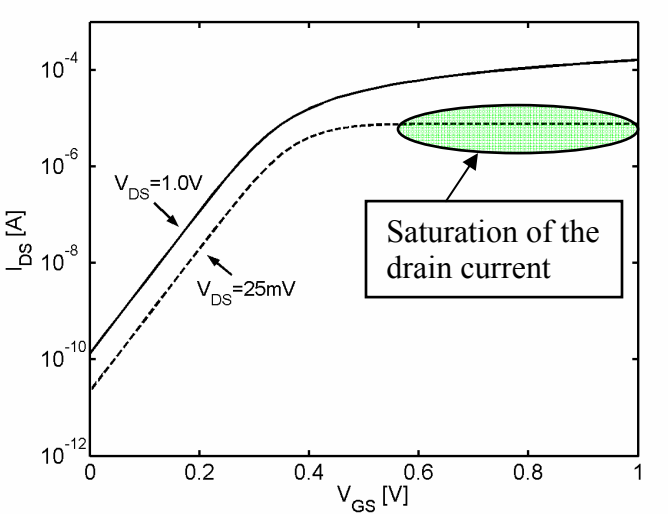


(a)

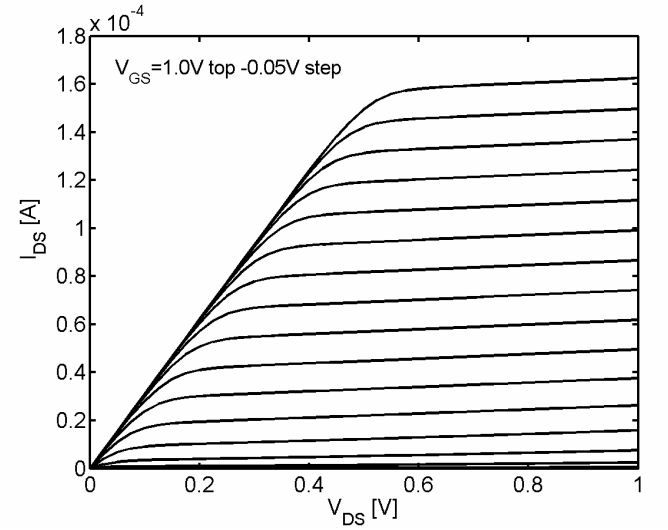


(b)

Fig. 2.2 (a) $\log(I_{DS})$ vs. V_{GS} transfer characteristics and (b) I_{DS} vs. V_{DS} characteristics of the simulated SNWT with a SiO_2 insulator layer ($\kappa = 3.9$).



(a)



(b)

Fig. 2.3 (a) $\log(I_{DS})$ vs. V_{GS} transfer characteristics and (b) I_{DS} vs. V_{DS} characteristics of the simulated SNWT with a high-K insulator layer ($\kappa = 80$).

B) *Quantum Capacitance:*

It is clear from Eq. (2.4) that the mobile charge, Q_{Top}, depends on the potential at the top of the barrier, U_{scf}. To describe this relation, a non-linear, quantum or semiconductor capacitance can be defined as [52] [55] [56]

$$C_Q \equiv \left| \frac{\partial Q_{Top}}{\partial U_{scf}} \right| . \tag{2.14}$$

Under high drain bias (e.g., ON-state), $Q_{Top} = -qn^+ = \left(-qN_{1D}/2\right)\Im_{-1/2}\left(\eta_F\right)$, so

$$C_Q = \left| \frac{\partial Q_{Top}}{\partial U_{scf}} \right| = \left| \frac{\partial Q_{Top}}{\partial \eta_F} \frac{\partial \eta_F}{\partial U_{scf}} \right| = \frac{q^2 N_{1D}}{2k_B T} \frac{\partial \Im_{-1/2}\left(\eta_F\right)}{\partial \eta_F} . \tag{2.15}$$

In general, the quantum capacitance, C_Q, needs to be evaluated numerically from Eq. (2.15). At the non-degenerate ($\eta_F < 0$ and $|\eta_F| \gg 1$) or full-degenerate ($\eta_F \gg 1$) limits, however, an analytical relation between C_Q and η_F can be obtained as follows. According to the properties of the Fermi integrals [36],

$$\Im_j\left(\eta\right) \simeq \exp\left(\eta\right), \text{ when } \eta < 0 \text{ and } |\eta| \gg 1, \tag{2.16a}$$

and

$$\Im_j\left(\eta\right) \simeq \frac{\eta^{j+1}}{\Gamma\left(j+2\right)}, \text{ when } \eta \gg 1, \tag{2.16b}$$

the quantum capacitance, C_Q, is obtained as

$$C_Q = \frac{q^2 N_{1D}}{2k_B T} \frac{\partial \exp\left(\eta_F\right)}{\partial \eta_F} = M\sqrt{\frac{q^4 m_x^*}{2\pi\hbar^2 k_B T}} \cdot \exp\left(\eta_F\right), \text{ (non-degenerate)} \tag{2.17a}$$

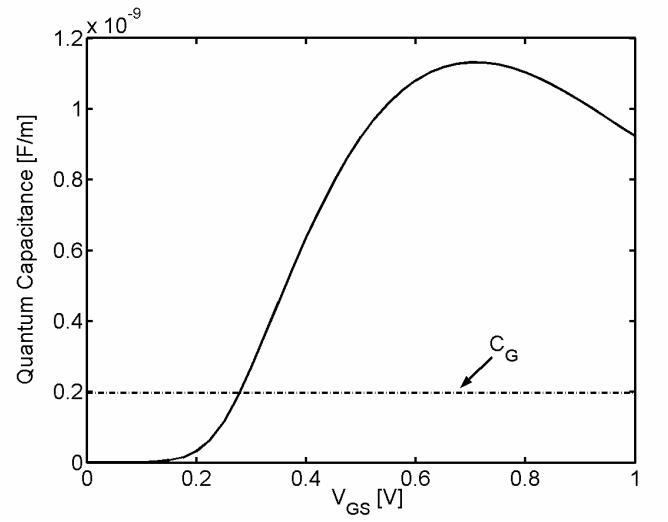
or

$$C_Q = \frac{q^2 N_{1D}}{2k_B T \Gamma\left(3/2\right)} \frac{\partial \eta_F^{\,1/2}}{\partial \eta_F} = M\sqrt{\frac{2q^4 m_x^*}{\pi^2\hbar^2 k_B T}} \cdot \eta_F^{\,-1/2} = M\sqrt{\frac{2q^4 m_x^*}{\pi^2\hbar^2\left[\mu_S - \left(\varepsilon\left(0\right) - qU_{scf}\right)\right]}},$$

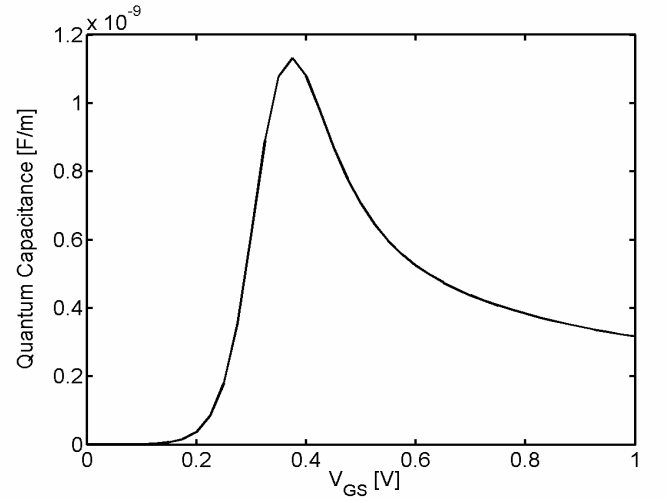
(full-degenerate). (2.17b)

It is implied from Eq. (2.17) that the quantum capacitance increases with the gate voltage under low gate bias (non-degenerate) while it decreases with the gate voltage under high gate bias (full-degenerate). This effect is clearly illustrated in Fig. 2.4, the C_Q vs. V_{GS} plots for the simulated SNWT with (a) a SiO_2 layer and (b) a high-K ($\kappa = 80$) insulator layer. As discussed in [52] [55] [56], when the gate insulator capacitance is significantly larger than the quantum capacitance (i.e., $C_Q / C_G \to 0$), the FET works at the quantum capacitance limit (QCL) – the potential at the top of the barrier is insensitive

to the local electron charge, Q_{Top}, and is solely determined by the applied voltage biases, V_G, V_D and V_S. From Fig. 2.4a, we found that $C_G < C_Q$ when $V_{GS} > 0.28\text{V}$, so the QCL is not achieved at the ON-state for the SiO_2 layer. For the high-K ($\kappa = 80$) insulator layer, $C_G = 4.05 \bullet 10^{-9}\,\text{F/m} \gg C_Q$ under the applied gate voltages ($0\text{V} < V_{GS} < 1\text{V}$), so the device is working close to the QCL.



(a)



(b)

Fig. 2.4 C_Q vs. V_{GS} plots for the simulated SNWT with (a) a SiO_2 layer and (b) a high-K ($\kappa = 80$) insulator layer.

There are several interesting phenomena that happen at the QCL. First, let us derive an expression for the transconductance of a 1D SNWT. Transconductance, g_m, for a FET is defined as [2]

$$g_m = \frac{\partial I}{\partial V_{GS}}\Big|_{\text{high } V_{DS}}. \tag{2.18}$$

For a SNWT, according to Eqs. (2.9) and (2.6),

$$g_m = \frac{\partial I}{\partial V_{GS}}\Big|_{U_D>>1} = M\frac{qk_BT}{\pi\hbar}\frac{\partial \Im_0(\eta_F)}{\partial V_{GS}} = M\frac{qk_BT}{\pi\hbar}\frac{\partial \Im_0(\eta_F)}{\partial \eta_F}\frac{\partial \eta_F}{\partial V_{GS}} = M\frac{q^2}{\pi\hbar}\frac{e^{\eta_F}}{1+e^{\eta_F}}\frac{\partial U_{scf}}{\partial V_{GS}}. \tag{2.19}$$

At the QCL, the potential at the top of the barrier, U_{scf}, is solely determined by the applied voltage biases, V_G, V_D and V_S, so Eq. (2.1) is simplified as (assuming $C_Q/C_G = 0$)

$$U_{scf} = \left(\frac{C_G}{C_G + C_D + C_S}\right)V_G + \left(\frac{C_D}{C_G + C_D + C_S}\right)V_D + \left(\frac{C_S}{C_G + C_D + C_S}\right)V_S.$$

In our simulation, we assume that the source is always grounded, so $V_S = 0$ and

$$U_{scf} = \left(\frac{C_G}{C_G + C_D + C_S}\right)V_{GS} + \left(\frac{C_D}{C_G + C_D + C_S}\right)V_{DS} = \alpha_G V_{GS} + \alpha_D V_{DS}. \tag{2.20}$$

Thus,

$$\frac{\partial U_{scf}}{\partial V_{GS}} = \alpha_G. \text{ (at the QCL, assuming } C_Q/C_G = 0). \tag{2.21}$$

Inserting Eq. (2.21) into Eq. (2.19), we obtain the expression for the transconductance of a SNWT at the QCL as

$$g_m = M\frac{q^2}{\pi\hbar}\frac{e^{\eta_F}}{1+e^{\eta_F}}\alpha_G = M\frac{2q^2}{h}\frac{\alpha_G}{1+e^{-\eta_F}}. \tag{2.22a}$$

At the full degenerate limit, $\eta_F \gg 1$, Eq. (2.22a) can be simplified as

$$g_m = \alpha_G M\frac{2q^2}{h}. \text{ (full degenerate)} \tag{2.22b}$$

Recall Eq. (2.13b), the expression for the channel conductance of a SNWT at the full-degenerate limit is

$$g_d = M\frac{2q^2}{h}. \tag{2.13b}$$

Interestingly, an analytical relation between g_m and g_d at the QCL and the full-degenerate limit is obtained as

$$g_m = \alpha_G g_d. \text{ (full degenerate)} \tag{2.23}$$

Fig. 2.5 plots the g_d vs. V_{GS} (solid) and g_m vs. V_{GS} (dashed) curves for the simulated SNWT with a high-K ($\kappa = 80$) insulator layer. It is clearly shown that when $V_{GS} > 0.6\text{V}$, the device enters the full-degenerate regime and the channel conductance, g_d, saturates at the quantum conductance value $M\left(2q^2/h\right) = 4\times 7.72\bullet 10^{-5} = 3.1\bullet 10^{-4}\text{S}$, while the transconductance, g_m, approaches a value of $2.56\bullet 10^{-4}\text{S}$, which is slightly (~5%) lower than the value $\alpha_G M\left(2q^2/h\right) = 0.88\times 3.1\bullet 10^{-4} = 2.7\bullet 10^{-4}\text{S}$ predicted by Eq. (2.22b). The reason is that Eq. (2.22b) is derived assuming $C_Q/C_G = 0$. For the device structure we simulate, C_Q/C_G is not strictly equal to zero but a small value of $C_Q/C_G = 3.16\bullet 10^{-10}/4.05\bullet 10^{-9} = 0.078$ at the ON-state, which causes the small discrepancy between the numerical results and the analytical expression.

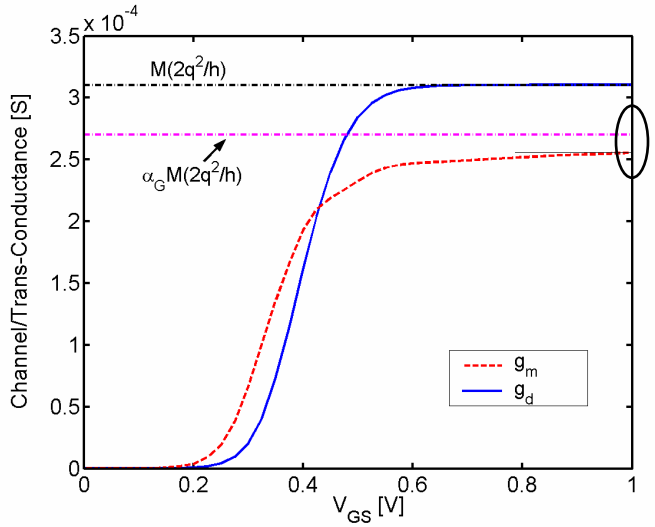
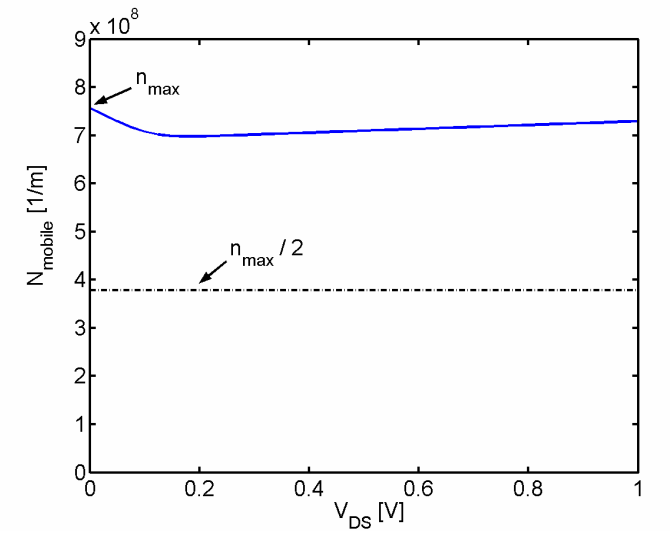
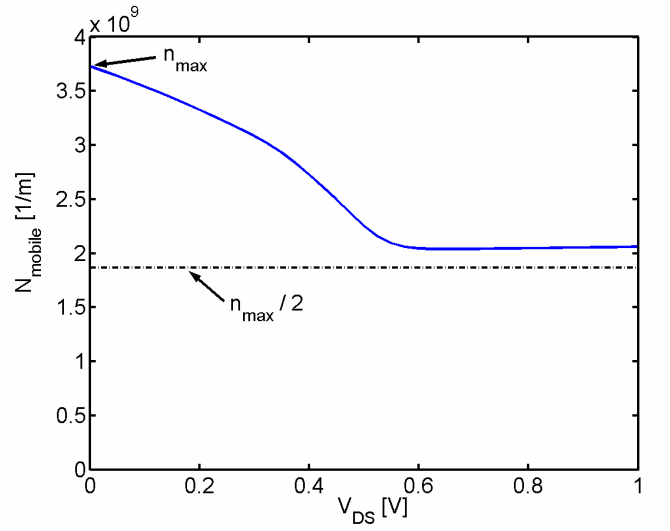


Fig. 2.5 The g_d vs. V_{GS} (solid) and g_m vs. V_{GS} (dashed) curves for the simulated SNWT with a high-K ($\kappa = 80$) insulator layer.

Another good way to show the features at the QCL is to plot the mobile electron density at the top of the barrier, $N_{mobile} = -Q_{Top}/q$, vs. the drain bias, V_{DS}. As we know, for a conventional MOSFET, the gate insulator capacitance is much smaller than the semiconductor capacitance at the ON-state, so N_{mobile} is controlled by the gate voltage [2] [65] and is independent of the drain bias if we neglect the drain induced barrier lowering (DIBL) effect [2]. At the QCL, however, the potential at the top of the barrier is fixed by the applied voltage biases and the increasing drain bias vacates the –*k* states (n^-) while leaving the +*k* states (n^+) unchanged. Therefore, the mobile electron density, N_{mobile}, at the QCL rapidly decreases from the equilibrium value, $n_{\max} = n\big|_{V_{DS}=0}$, to $n_{\max}/2$, as the drain bias increases from zero. When the drain bias exceeds $\left[\mu_S - \left(\varepsilon(0) - qU_{scf}\right)\right]/q$, all the –*k* states ($n^-$) have been nearly depleted, so the increasing V_{DS} will not significantly reduce N_{mobile} any more and N_{mobile} starts to saturate at a fixed value, $n_{\max}/2$. It should also be noted that if DIBL is considered, the N_{mobile} will increase with V_{DS} under high drain bias since increasing V_{DS} lowers the top of the barrier and consequently increases the degeneracy factor, η_F. Fig. 2.6 plots the N_{mobile} vs. V_{DS} curves (solid) for the simulated SNWTs with (a) a SiO_2 layer and (b) a high-K ($\kappa = 80$) insulator layer. The former device is working well below the QCL while the latter one is close to the QCL (see Fig. 2.4). Differences between the two plots clearly illustrate the quantum capacitance effect discussed above. This effect become more and more important for the device characteristics of nanoscale FETs [52] [56] as the gate capacitance continues to be raised by scaling down the oxide thickness and/or adopting high-K materials.



(a) $\kappa = 3.9$



(b) $\kappa = 80$

Fig. 2.6 The N_{mobile} vs. V_{DS} curves (solid) for the simulated SNWTs with (a) a SiO_2 layer ($\kappa = 3.9$) and (b) a high-K ($\kappa = 80$) insulator layer. The former device is working well below the QCL while the latter one is close to the QCL. For $\kappa = 80$, the discrepancy between the computed N_{mobile} under high V_{DS} and the theoretically predicted value, $n_{\max}/2$, is caused by both DIBL and the fact that the ratio C_Q/C_G is not strictly equal to zero but a small value (~0.1) under high V_{DS}.

C) *Injection Velocity:*

It is also interesting to explore the average injection velocity, υ_{inj}, under high drain bias. For a SNWT, υ_{inj} is obtained as

$$\upsilon_{inj} = \frac{I^+}{qn^+} = \frac{M\frac{qk_BT}{\pi\hbar}\Im_0(\eta_F)}{q\frac{M}{2}\sqrt{\frac{2k_BTm_x^*}{\pi\hbar^2}}\cdot\Im_{-1/2}(\eta_F)} = \sqrt{\frac{2k_BT}{\pi m_x^*}}\frac{\Im_0(\eta_F)}{\Im_{-1/2}(\eta_F)}. \quad (2.24)$$

According to the properties of the Fermi integrals (see Eq. (2.16)),

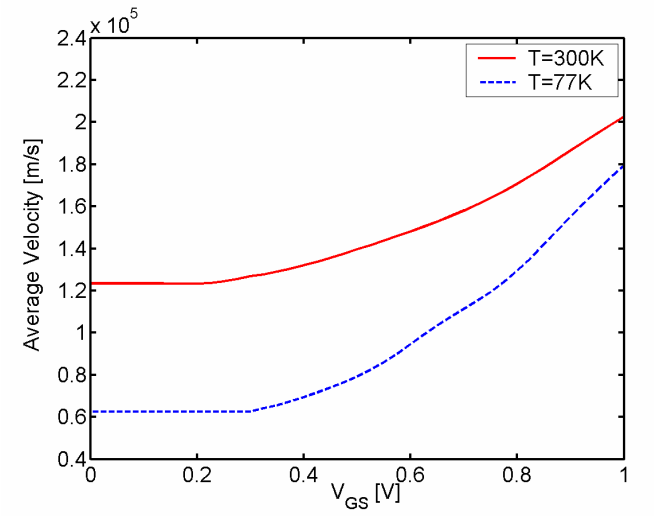
$$\upsilon_{inj} = \sqrt{\frac{2k_BT}{\pi m_x^*}}\frac{e^{\eta_F}}{e^{\eta_F}} = \sqrt{\frac{2k_BT}{\pi m_x^*}} = \upsilon_T \quad \text{(non-degenerate)}, \quad (2.25a)$$

where υ_T is the uni-directional thermionic velocity [36] [52] [67], and

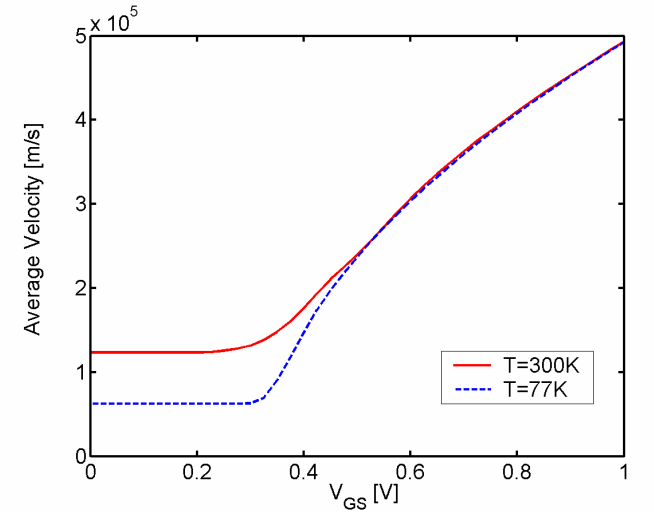
$$\upsilon_{inj} = \sqrt{\frac{2k_BT}{\pi m_x^*}}\frac{\eta_F}{\Gamma(2)}\frac{\Gamma(3/2)}{\eta_F^{1/2}} = \frac{1}{2}\sqrt{\frac{2\left[\mu_S - \left(\varepsilon(0)+U_{TOP}\right)\right]}{m_x^*}} = \frac{\upsilon_F}{2} \quad \text{(full-degenerate)}, \quad (2.25b)$$

where υ_F is the Fermi velocity [36] [67].

In Fig. 2.7, we plot the high-V_{DS} injection velocity, υ_{inj}, vs. gate voltage, V_{GS}, curves for the simulated SNWTs with (a) a SiO_2 layer ($\kappa = 3.9$) and (b) a high-K ($\kappa = 80$) insulator layer. Two ambient temperatures, 300K (solid) and 77K (dashed), are considered. It is clearly shown that under low gate bias (non-degenerate), υ_{inj} is equal to υ_T and independent of the gate bias. Under high V_{GS}, the carrier degeneracy makes υ_{inj} monotonically increase with the gate bias for both $\kappa = 3.9$ and $\kappa = 80$. For $\kappa = 80$, when $V_{GS} > 0.5V$, the device enters the full-degenerate regime, so $\upsilon_{inj} = \upsilon_F/2$ and is independent of the temperature. Consequently, the two curves (solid and dashed) in Fig. 2.7b lie on top of each other when $0.5\text{V} < V_{GS} < 1.0\text{V}$.



(a) $\kappa = 3.9$



(b) $\kappa = 80$

Fig. 2.7 The high V_{DS} injection velocity, υ_{inj}, vs. gate voltage, V_{GS}, curves for the simulated SNWTs with (a) a SiO_2 layer ($\kappa = 3.9$) and (b) a high-K ($\kappa = 80$) insulator layer. Two ambient temperatures, 300K (solid) and 77K (dashed), are selected.

2.3 Summary

In this chapter, we presented a simple, analytical theory of ballistic nanowire FETs. The model was derived by modifying an analytical approach that was previously used for ballistic planar MOSFETs [52]. A coaxial SNWT was simulated using this

model and the results illustrate the essential physics and peculiarities of 1D nanowire FETs, such as the saturation of channel conductance at the full-degenerate limit and the saturation of transconductance at the quantum capacitance limit and the full-degenerate limit. As mentioned in [52], detailed bandstructure information can also be implemented into this analytical model, so this approach provides an opportunity to investigate the bandstructure effects on the ballistic limits of SNWTs (see Chapter 6).

3. A THREE-DIMENSIONAL QUANTUM SIMULATION OF BALLISTIC SILICON NANOWIRE TRANSISTORS WITH THE EFFECTIVE-MASS APPROXIMATION

In this chapter, we present a three-dimensional (3D), quantum mechanical simulation approach to treat silicon nanowire transistors (SNWTs) within the effective-mass approximation [43] [44]. Ballistic transport is assumed, which gives the upper performance limit of the devices. The use of a mode space approach (either coupled or uncoupled) [41] [45] [47] produces high computational efficiency that makes our 3D quantum simulator practical for extensive device simulation and design. Scattering in SNWTs can also be treated by a simple model that uses so-called Büttiker probes [57], which was previously used in MOSFET simulations [42]. The detailed of this method will be shown in Appendix. With the use of this simulator, the ballistic performance limits of SNWTs with various cross-sections can be evaluated (Chapter 4) and a microscopic treatment of surface roughness scattering in SNWTs becomes feasible (Chapter 5).

3.1 Introduction

Different from a planar MOSFET, which has a uniform charge and potential profile in the transverse direction (normal to both the gate and the source-to-drain direction), a SNWT has a 3D distribution of electron density and electrostatic potential. As a result, a 3D simulator is required for the simulation of SNWTs. In this chapter, we propose a 3D self-consistent, quantum simulation of SNWTs based on the effective-mass approximation [43] [44]. The calculation involves a self-consistent solution of a 3D Poisson equation and a 3D Schrödinger equation with open boundary conditions. Using

the finite element method (FEM), we solve the 3D Poisson equation rigorously to obtain the electrostatic potential. [45] At the same time, we solve the 3D Schrödinger equation by a (coupled/uncoupled) mode space approach, which provides both computational efficiency and high accuracy as compared with direct real space calculations [41] [45] [47]. Since the (coupled/uncoupled) mode space approach treats quantum confinement and transport separately, the procedure of the calculation is as follows:

Step 1: Solve the 3D Poisson equation for the electrostatic potential;

Step 2: Solve a two-dimensional (2D) Schrödinger equation with a closed boundary condition at each slice (cross-section) of the nanowire transistor (see Fig. 3.1) to obtain the electron subbands (along the nanowire) and the corresponding eigenfunctions;

Step 3: Solve (coupled/uncoupled) one-dimensional (1D) transport equations by the nonequilibrium Green's function (NEGF) approach [37] [38] for the electron charge density;

Step 4: Go back to Step 1 to calculate the electrostatic potential. If it converges, then calculate the electron current by the NEGF approach (as in Step 3) and output the results. Otherwise continue Steps 2 and 3.

Different transport models (in Step 3) can be implemented into our simulator. In this chapter, we will focus on the ballistic NEGF model, which gives the upper performance limit of SNWTs. In Appendix, we will introduce a dissipative NEGF model with a simple treatment of scattering with the Büttiker probes [42] [57], which offers a phenomenological way to capture scattering in the quantum mechanical framework.

A rigorous treatment of scattering and a detailed calculation of bandstructures are very important to understand physics in Si nanowires in detail. However, the huge computational cost involved in such a rigorous model can prevent it from being used for extensive device simulation and design. As we will show, the use of the effective-mass approximation and the simple treatment of scattering with the Büttiker probes (see Appendix) greatly reduces the computational complexity while still capturing the essential device physics of SNWTs (i.e., 3D electrostatics, quantum confinement, source-to-drain tunneling and scattering, etc), so the method we discuss in this chapter (and

Appendix) can be used as a practical 3D quantum approach for device study and design of SNWTs. [4] This chapter is divided into the following sections: Sec. 3.2 describes our methodology for ballistic SNWTs and provides the basic equations, Sec. 3.3 discusses the simulation results for ballistic SNWTs with arbitrary cross-sections (e.g., triangular, rectangular and cylindrical), and Sec. 3.4 summarizes key findings.

3.2 Theory

Figure 3.1 shows a schematic structure of the Si nanowire transistors simulated in this chapter. This intrinsic device structure is connected to two infinite reservoirs, the source (S) and the drain (D), so the S/D extension regions are terminated using open boundary conditions. As shown in Fig. 3.1b, a uniform grid with a grid spacing of a is used along the channel (x) direction. In the y-z plane (the cross-section of the SNWT), a 2D finite element mesh with triangular elements is generated by Easymesh-1.4 [68], which allows us to treat nanowires with arbitrary cross-sections (e.g., triangular, rectangular and cylindrical). By doing this, a 3D finite element mesh with prism elements is constructed. When solving the Poisson equation, the 3D Laplacian is directly discretized by the FEM approach. The obtained linear system is solved using a Preconditioned Conjugate Gradient method with Incomplete Cholesky factorization [45]. More details about the numerical techniques can be found in [45].

As mentioned earlier, we solve the 3D Schrödinger equation by the mode space approach [41] [45] [47], which is based on an expansion of the active device Hamiltonian in the subband eigenfunction space. As a result, we need to solve a 2D Schrödinger equation by the FEM at each slice of the SNWT to obtain the subband eigenenergy levels and eigenfunctions (modes). After that, the original 3D device Hamiltonian is transformed into a 1D Hamiltonian in the x direction, which can be used to calculate electron density and current within the NEGF formalism. In this section, we will first give an overview of the coupled mode space (CMS) approach for the SNWT simulation (Subsection 3.2.1), which is mathematically equivalent to a direct real space solution if adequate modes are included (to be discussed later) [41] [45] [47]. Then we will

introduce the uncoupled mode space (UMS) approach (Subsection 3.2.2) and a fast uncoupled mode space (FUMS) approach (Subsection 3.2.3), which are a simplification of the CMS approach to provide high computational efficiency. The simulation results in Sec. 3.3 illustrate that the UMS and FUMS approaches show excellent agreement with the CMS approach for SNWT simulations.

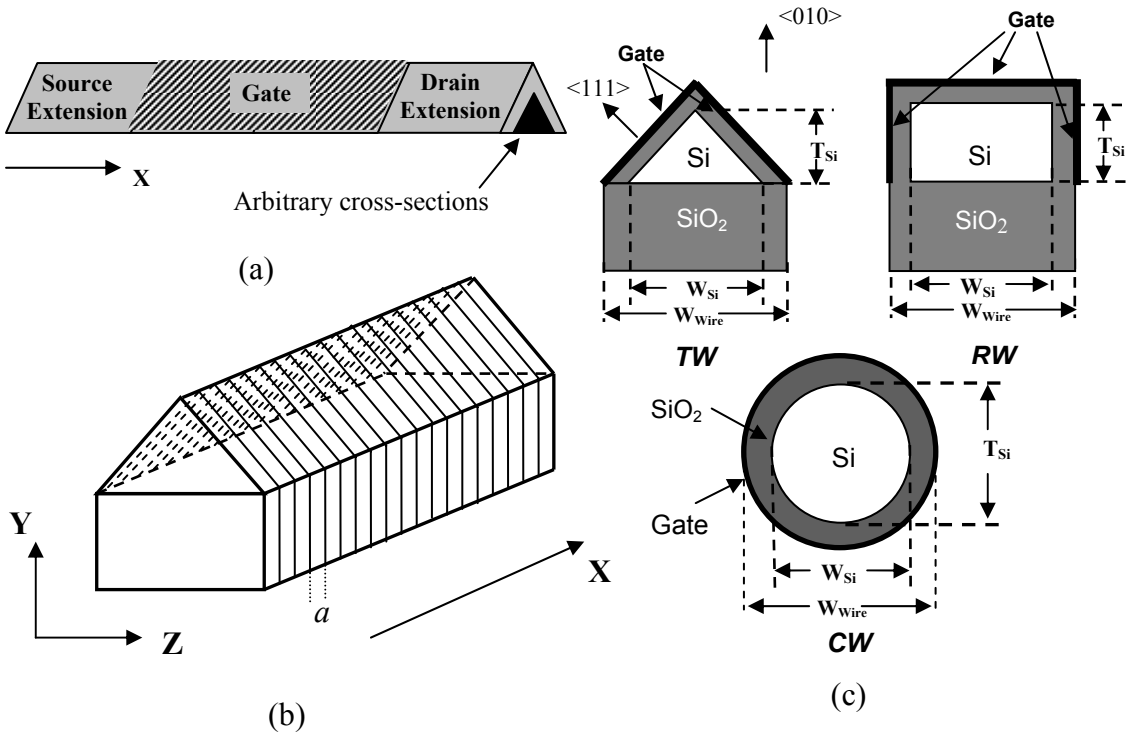


Fig. 3.1 The simulated SNWT structures in this work. (a) A schematic graph of an intrinsic SNWT with arbitrary cross-sections (for clarity, the SiO_2 substrate is not shown here). (b) The gird used in the simulation of SNWTs. (c) The cross-sections of the simulated triangular wire (TW), rectangular wire (RW) and cylindrical wire (CW) FETs. T_{Si} is the silicon body thickness, W_{Si} is the silicon body width and W_{Wire} is the wire width. For the TW, the direction normal to each gate is <111>, so the channel is <101> oriented. In contrast, for the channel of the RW, both <101> and <100> orientations are possible. For the CW, we assume the channel to be <100> oriented.

3.2.1 The coupled mode space approach

In this part of the work, we will briefly review the coupled mode space approach and list basic equations for our particular case of interest.

In the 3D domain, the full stationery Schrödinger equation is given by

$$H_{3D}\Psi(x,y,z)=E\Psi(x,y,z), \tag{3.1}$$

where H_{3D} is the 3D device Hamiltonian. Assuming an ellipsoidal parabolic energy band with a diagonal effective-mass tensor (for the case that the effective-mass tensor includes non-zero off-diagonal elements, please refer to [69]), H_{3D} is defined as

$$H_{3D}=-\frac{\hbar^2}{2m_x^*(y,z)}\frac{\partial^2}{\partial x^2}-\frac{\hbar^2}{2}\frac{\partial}{\partial y}\left(\frac{1}{m_y^*(y,z)}\frac{\partial}{\partial y}\right)-\frac{\hbar^2}{2}\frac{\partial}{\partial z}\left(\frac{1}{m_z^*(y,z)}\frac{\partial}{\partial z}\right)+U(x,y,z), \tag{3.2}$$

here m_x^*, m_y^* and m_z^* are the electron effective-mass in the x, y, and z directions, respectively, and $U(x,y,z)$ is the electron conduction band-edge profile in the active device. We note that the effective-mass varies in the y and z directions due to the transition between the Si body and the SiO_2 layer. Now let us expand the 3D electron wavefunction in the subband eigenfunction space,

$$\Psi(x,y,z)=\sum_n \varphi^n(x)\cdot\xi^n(y,z;x), \tag{3.3}$$

where $\xi^n(y,z;x=x_0)$ is the *nth* eigenfunction of the following 2D Schrödinger equation at the slice ($x=x_0$) of the SNWT,

$$\left[-\frac{\hbar^2}{2}\frac{\partial}{\partial y}\left(\frac{1}{m_y^*(y,z)}\frac{\partial}{\partial y}\right)-\frac{\hbar^2}{2}\frac{\partial}{\partial z}\left(\frac{1}{m_z^*(y,z)}\frac{\partial}{\partial z}\right)+U(x_0,y,z)\right]\xi^n(y,z;x_0)=E_{sub}^n(x_0)\xi^n(y,z;x_0), \tag{3.4}$$

here $E_{sub}^n(x_0)$ is the *nth* subband energy level at $x=x_0$. According to the property of eigenfunctions, $\xi^n(y,z;x)$ satisfies the following equation for any x,

$$\oint_{y,z}\xi^m(y,z;x)\xi^n(y,z;x)dydz=\delta_{m,n}, \tag{3.5}$$

where $\delta_{m,n}$ is the Kronecker delta function [36].

Inserting Eqs. (3.2) and (3.3) into Eq. (3.1) and using the relation described by Eq. (3.4), we obtain

$$-\frac{\hbar^2}{2m_x^*(y,z)}\frac{\partial^2}{\partial x^2}\left(\sum_n \varphi^n(x)\cdot\xi^n(y,z;x)\right)+\sum_n \varphi^n(x)\cdot E_{sub}^n(x)\xi^n(y,z;x)=E\sum_n \varphi^n(x)\cdot\xi^n(y,z;x). \tag{3.6}$$

Now we multiply by $\xi^m(y,z;x)$ on both sides and do an integral within the y-z plane. According to Eq. (3.5), we obtain the following 1D coupled Schrödinger equation

$$-\frac{\hbar^2}{2}\left(\sum_{n=1}^{\infty}a_{mn}(x)\right)\frac{\partial^2\varphi^m(x)}{\partial x^2}-\frac{\hbar^2}{2}\sum_n c_{mn}(x)\varphi^n(x)-\hbar^2\sum_n b_{mn}(x)\frac{\partial\varphi^n(x)}{\partial x}+E_{sub}^m(x)\varphi^m(x)=E\varphi^m(x), \tag{3.7}$$

where

$$a_{mn}(x)=\oint_{y,z}\frac{1}{m_x^*(y,z)}\xi^m(y,z;x)\xi^n(y,z;x)dydz, \tag{3.8a}$$

$$b_{mn}(x)=\oint_{y,z}\frac{1}{m_x^*(y,z)}\xi^m(y,z;x)\frac{\partial}{\partial x}\xi^n(y,z;x)dydz, \tag{3.8b}$$

and

$$c_{mn}(x)=\oint_{y,z}\frac{1}{m_x^*(y,z)}\xi^m(y,z;x)\frac{\partial^2}{\partial x^2}\xi^n(y,z;x)dydz. \tag{3.8c}$$

The basic equation for the CMS approach is Eq. (3.7). In our simulation, since the electron wavefunction is mainly located in the silicon, we can neglect a_{mn} if $m\neq n$ ($a_{mm}\gg a_{mn}$) [47] and simplify Eq. (3.7) as

$$-\frac{\hbar^2}{2}a_{mm}(x)\frac{\partial^2}{\partial x^2}\varphi^m(x)-\frac{\hbar^2}{2}\sum_n c_{mn}(x)\varphi^n(x)-\hbar^2\sum_n b_{mn}(x)\frac{\partial}{\partial x}\varphi^n(x)+E_{sub}^m(x)\varphi^m(x)=E\varphi^m(x). \tag{3.9}$$

From the derivation above, it is clear that the CMS formalism, Eqs. (3.7) and (3.8), is mathematically equivalent to the real space calculation if all the modes (i.e., $m,n=1,...,N_{YZ}$, where N_{YZ} is the number of nodes in the y-z plane) are included. In practice, due to strong quantum confinement in SNWTs, usually only a few of the lowest

subbands (i.e., $m,n=1,...,M$, $M \ll N_{YZ}$) 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 (i.e., *m,n*=1,...,*M*), Eq. (3.9) represents an equation group that contains *M* equations, each representing a selected mode. We can write down these *M* equations in a matrix format

$$H\begin{bmatrix}\varphi^1(x)\\ \varphi^2(x)\\ \dots\\ \dots\\ \varphi^M(x)\end{bmatrix}=\begin{bmatrix}h_{11} & h_{12} & h_{13} & \dots & h_{1M}\\ h_{21} & h_{22} & h_{23} & \dots & h_{2M}\\ \dots & \dots & \dots & \dots & \dots\\ \dots & \dots & \dots & \dots & \dots\\ h_{M1} & h_{M2} & h_{M3} & \dots & h_{MM}\end{bmatrix}\begin{bmatrix}\varphi^1(x)\\ \varphi^2(x)\\ \dots\\ \dots\\ \varphi^M(x)\end{bmatrix}=E\begin{bmatrix}\varphi^1(x)\\ \varphi^2(x)\\ \dots\\ \dots\\ \varphi^M(x)\end{bmatrix}, \quad (3.10)$$

where

$$h_{mn}=\delta_{m,n}\left[-\frac{\hbar^2}{2}a_{mm}(x)\frac{\partial^2}{\partial x^2}+E^m_{sub}(x)\right]-\frac{\hbar^2}{2}c_{mn}(x)-\hbar^2 b_{mn}(x)\frac{\partial}{\partial x}, \ (m,n=1,2,\dots,M). \quad (3.11)$$

By using the coupled mode space approach, the size of the device Hamiltonian, *H*, has been reduced to $M\cdot N_X \times M\cdot N_X$ (N_X is the number of nodes in the x direction, and the mode number *M* we need is normally less than five for the SNWT structures we simulate), which is much smaller than that in the real space representation, $N_{YZ}\cdot N_X \times N_{YZ}\cdot N_X$ (N_{YZ} is ~1,000 for the device structures simulated in this chapter).

After the device Hamiltonian *H* is obtained, we can calculate the electron density and current using the NEGF approach. The NEGF approach, a widely used method for the simulation of nanoscale electronic devices, has been discussed in [37] [38]. Here we list the relevant equations for our particular case.

The retarded Green's function of the active device is defined as [37]

$$G(E)=\left[ES-H-\Sigma_S(E)-\Sigma_1(E)-\Sigma_2(E)\right]^{-1}, \quad (3.12)$$

where the device Hamiltonian, *H*, is defined by Eq. (3.10), Σ_S is the self-energy that accounts for the scattering inside the device (in the ballistic limit, it is equal to zero), Σ_1 (Σ_2) is the self-energy caused by the coupling between the device and the source (drain) reservoir. If we discretize the equations by the 1D (in the x direction) finite difference

method (FDM), the matrix S in Eq. (3.12) is equal to an $M \cdot N_X \times M \cdot N_X$ identity matrix. The self-energies, Σ_1 and Σ_2, are defined as [37]

$$\Sigma_1[p,q] = -t_{m,1} \exp\left(jk_{m,1}a\right)\delta_{p,(m-1)N_X+1}\delta_{q,(m-1)N_X+1}, \ (j=\sqrt{-1}) \ \text{(FDM)}, \quad (3.13)$$

$$\Sigma_2[p,q] = -t_{m,N_X} \exp\left(jk_{m,N_X}a\right)\delta_{p,mN_X}\delta_{q,mN_X}, \ (m=1,2,\ldots, M, \ p,q=1,2,\ldots,MN_X) \ \text{(FDM)}, \quad (3.14)$$

where $t_{m,1} = \left(\hbar^2/2a^2\right)a_{mm}(x)\big|_{x=0}$ and $t_{m,N_X} = \left(\hbar^2/2a^2\right)a_{mm}(x)\big|_{x=(N_X-1)a}$ ($a_{mm}(x)$ is defined by Eq. (3.8a)), and $k_{m,1}$ and k_{m,N_X} are determined by $E = E_{sub}^m(0) + 2t_{m,1}\left(1-\cos k_{m,1}a\right)$ and $E = E_{sub}^m\left[\left(N_X-1\right)a\right] + 2t_{m,N_X}\left(1-\cos k_{m,N_X}a\right)$, respectively.

If we discretize the equations by the 1D (in the x direction) FEM, the matrix S in Eq. (3.12) becomes a $M \cdot N_X \times M \cdot N_X$ block diagonal matrix

$$S = \begin{bmatrix} S_0 & 0 & \cdots & \cdots & 0 \\ 0 & S_0 & 0 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & S_0 \end{bmatrix} \ \text{(FEM)}, \quad (3.15)$$

where S_0 is a $N_X \times N_X$ matrix [46]

$$S_0 = \begin{bmatrix} a/3 & a/6 & 0 & \cdots & \cdots & 0 \\ a/6 & 2a/3 & a/6 & \ddots & \ddots & \vdots \\ 0 & a/6 & 2a/3 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 2a/3 & a/6 \\ 0 & \cdots & \cdots & 0 & a/6 & a/3 \end{bmatrix} \ \text{(FEM)}. \quad (3.16)$$

The self-energies, Σ_1 and Σ_2, are defined as [46] [47]

$$\Sigma_1[p,q] = -jk_{m,1}at_{m,1}\delta_{p,(m-1)N_X+1}\delta_{q,(m-1)N_X+1} \ \text{(FEM)}, \quad (3.17)$$

$$\Sigma_2[p,q] = -jk_{m,N_X}at_{m,N_X}\delta_{p,mN_X}\delta_{q,mN_X}, \ (m=1,2,\ldots, M \text{ and } p,q=1,2,\ldots,MN_X) \ \text{(FEM)}. \quad (3.18)$$

By inserting Eqs. (3.13)-(3.14) or (3.15)-(3.18) into Eq. (3.12), we can evaluate the retarded Green's function, *G(E)*, at a given energy *E*. Then the spectral density functions due to the S/D contacts can be obtained as [37]

$$A_1(E) = G(E)\Gamma_1(E)G^{\dagger}(E) \text{ and } A_2(E) = G(E)\Gamma_2(E)G^{\dagger}(E), \quad (3.19)$$

where $\Gamma_1(E) \equiv j\left[\Sigma_1(E) - \Sigma_1^{\dagger}(E)\right]$ and $\Gamma_2(E) \equiv j\left[\Sigma_2(E) - \Sigma_2^{\dagger}(E)\right]$, which determine the electron exchange rates between the active device region and the S/D reservoirs at energy *E*. In this coupled mode space, the diagonal elements of the spectral function matrices represent the local density of states (LDOS) in the device for each mode. We define the LDOS for mode *m* as D_1^m (due to the source) and D_2^m (due to the drain). Here D_1^m and D_2^m are both $N_X \times 1$ vectors obtained as

$$D_1^m[p] = \frac{1}{\pi a} A_1\left[(m-1)N_X + p, (m-1)N_X + p\right], (p=1,2,...,N_X), \quad (3.20)$$

$$D_2^m[p] = \frac{1}{\pi a} A_2\left[(m-1)N_X + p, (m-1)N_X + p\right], (p=1,2,\ldots,N_X). \quad (3.21)$$

Then the 1D electron density (in m^{-1}) for mode *m* can be calculated by

$$n_{1D}^m = \int_{-\infty}^{+\infty}\left[D_1^m f(\mu_S, E) + D_2^m f(\mu_D, E)\right]dE, \quad (3.22)$$

where *f* is the Fermi-Dirac statistics function [36], and μ_S (μ_D) is the source (drain) Fermi level, which is determined by the applied bias. The electron density obtained by Eq. (3.22) is a 1D distribution (along the x direction). To obtain a 3D electron density, we need to couple Eq. (3.22) with the quantum confinement wavefunction for mode *m*,

$$n_{3D}^m(x,y,z) = n_{1D}^m(x)\left|\xi^m(y,z;x)\right|^2. \quad (3.23)$$

The total 3D electron density needs to be evaluated by summing the contributions from all the subbands in each conduction-band valley. Then this 3D electron density is fed back to the Poisson solver for the self-consistent calculations. Once self-consistency is achieved, the electron current is computed by

$$I_{SD} = \frac{q}{\pi\hbar}\int_{-\infty}^{+\infty} T(E)\left[f(\mu_S, E) - f(\mu_D, E)\right]dE, \quad (3.24)$$

where the transmission coefficient, *T(E)*, can be evaluated as [37]

$$T(E) = trace\left[\Gamma_1(E)G(E)\Gamma_2(E)G^{\dagger}(E)\right]. \tag{3.25}$$

To obtain the total electron current, we also need to add up current components in all the conduction-band valleys.

3.2.2 The uncoupled mode space approach

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

$$\xi^m(y,z;x) = \overline{\xi^m}(y,z) \tag{3.26}$$

or $$\frac{\partial}{\partial x}\xi^m(y,z;x) = 0\,,\ (m=1,2,...,M), \tag{3.27}$$

which infers

$$a_{mm}(x) = \overline{a_{mm}} = \oint_{y,z} \frac{1}{m_x^*(y,z)}\left|\overline{\xi^m}(y,z)\right|^2 dydz\,, \tag{3.28a}$$

$$b_{mn}(x) = 0 \text{ and } c_{mn}(x) = 0\,,\ (m,n=1,2,\ldots,M). \tag{3.28b}$$

Inserting Eq. (3.28b) into Eq. (3.11), we obtain $h_{mn} = 0$ ($m \neq n$ and $m,n=1,2,\ldots,M$), 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

$$H = \begin{bmatrix} h_{11} & 0 & \cdots & \cdots & 0 \\ 0 & h_{22} & 0 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & h_{MM} \end{bmatrix}. \tag{3.29}$$

Since all the input matrices at the RHS of Eq. (3.12) are either diagonal or block-diagonal, the retarded Green's function *G(E)* is block-diagonal,

$$G(E)=\begin{bmatrix} G^1(E) & 0 & \cdots & \cdots & 0 \\ 0 & G^2(E) & 0 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & G^M(E) \end{bmatrix}, \tag{3.30}$$

where $G^m(E)$ (m=1,2,…,M) is the Green's function for mode m and is obtained as

$$G^m(E)=\left[ES^m-h_{mm}-\Sigma_S^m(E)-\Sigma_1^m(E)-\Sigma_2^m(E)\right]^{-1}, \tag{3.31}$$

here S^m, Σ_S^m, Σ_1^m and Σ_2^m are all $N_X \times N_X$ matrices and defined as

$$S^m[p,q]=S\left[(m-1)N_X+p,(m-1)N_X+q\right], (p,q=1,2,\ldots,N_X), \tag{3.32}$$

$$\Sigma_S^m[p,q]=\Sigma_S\left[(m-1)N_X+p,(m-1)N_X+q\right], (p,q=1,2,\ldots,N_X), \tag{3.33}$$

$$\Sigma_1^m[p,q]=\Sigma_1\left[(m-1)N_X+p,(m-1)N_X+q\right], (p,q=1,2,\ldots,N_X), \tag{3.34}$$

and

$$\Sigma_2^m[p,q]=\Sigma_2\left[(m-1)N_X+p,(m-1)N_X+q\right], (p,q=1,2,\ldots,N_X). \tag{3.35}$$

Knowing the retarded Green's function, the spectral density functions due to the S/D contacts for each mode m can be obtained as [37]

$$A_1^m(E)=G^m(E)\Gamma_1^m(E)G^{m\dagger}(E) \text{ and } A_2^m(E)=G^m(E)\Gamma_2^m(E)G^{m\dagger}(E), \tag{3.36}$$

where $\Gamma_1^m(E)\equiv i\left[\Sigma_1^m(E)-\Sigma_1^{m\dagger}(E)\right]$ and $\Gamma_2^m(E)\equiv i\left[\Sigma_2^m(E)-\Sigma_2^{m\dagger}(E)\right]$. The LDOS for mode m, D_1^m (due to the source) and D_2^m (due to the drain), can then be evaluated by

$$D_1^m[p]=\frac{1}{\pi a}A_1^m[p,p] \text{ and } D_2^m[p]=\frac{1}{\pi a}A_2^m[p,p], (p=1,2,\ldots,N_X). \tag{3.37}$$

After that, the electron charge density is computed by Eqs. (3.22) and (3.23). For the calculation of electron current, the total transmission coefficient can be written as a summation of the transmission coefficient $T^m(E)$ for each mode m,

$$T(E)=\sum_{m=1}^{M}T^m(E), \tag{3.38}$$

where $T^m(E)$ is obtained as [37]

$$T^m(E) = trace\left[\Gamma_1^m(E) G^m(E) \Gamma_2^m(E) G^{m\dagger}(E)\right]. \quad (3.39)$$

Finally, Eq. (3.38) is inserted into Eq. (3.24) to compute the electron current for the SNWT.

As we will show in Sec. 3.3, this UMS approach shows excellent agreement with the CMS approach while maintaining higher computational efficiency. (The validity of the UMS approach for planar MOSFET simulation has been established by Venugopal et al. [41] by doing a careful study of the UMS approach vs. 2D real space approach.)

3.2.3 A fast uncoupled mode space approach

As described earlier, for both CMS and UMS approaches, we need to solve N_X 2D Schrödinger equations, shown in Eq. (3.4), in a self-consistent loop to obtain the electron subbands and eigenfunctions. For the device structures simulated in this chapter, this part of simulation usually takes more than 90% of the computational complexity, which makes parallel programming necessary. To increase the efficiency of our simulator and to make it executable on a single processor, we introduce a fast uncoupled mode space approach [45] [47], which only involves one 2D Schrödinger equation problem in a self-consistent loop and still provides excellent computational accuracy as compared with the CMS and UMS approaches. (The transport part of calculation in FUMS is the same as that in UMS.)

Recall the assumption made in Subsection 3.2.2 that the eigenfunctions $\xi^m(y,z;x)$ are invariant along the x direction, $\xi^m(y,z;x) = \overline{\xi^m}(y,z)$. Now we suppose that the average wavefunctions $\overline{\xi^m}(y,z)$ are the eigenfunctions of the following 2D Schrödinger equation

$$\left[-\frac{\hbar^2}{2}\frac{\partial}{\partial y}\left(\frac{1}{m_y^*(y,z)}\frac{\partial}{\partial y}\right) - \frac{\hbar^2}{2}\frac{\partial}{\partial z}\left(\frac{1}{m_z^*(y,z)}\frac{\partial}{\partial z}\right) + \overline{U}(y,z)\right]\overline{\xi^m}(y,z) = \overline{E_{sub}^m}\cdot\overline{\xi^m}(y,z). \quad (3.40)$$

Here the average conduction band-edge $\overline{U}(y,z)$ is obtained as

$$\overline{U}(y,z)=\frac{1}{L_X}\int_0^{L_X} U(x,y,z)\,dx\,, \tag{3.41}$$

where L_X is the total length of the simulated SNWT, including the S/D extensions. After computing the eigenvalues $\overline{E_{sub}^m}$ and eigenfunctions $\overline{\xi^m}(y,z)$ of this Schrödinger equation, we use the first order stationery perturbation theory to obtain the subband profile as [45] [47]

$$E_{sub}^m(x)=\overline{E_{sub}^m}+\oint_{y,z} U(x,y,z)\left|\overline{\xi^m}(y,z)\right|^2 dydz-\oint_{y,z}\overline{U}(y,z)\left|\overline{\xi^m}(y,z)\right|^2 dydz\,. \tag{3.42}$$

So far the subbands $E_{sub}^m(x)$ and the corresponding eigenfunctions $\xi^m(y,z;x)$ have been obtained approximately by only solving one 2D Schrödinger equation. The simulation results in Sec. 3.3 will show that this FUMS approach has great accuracy for the calculation of both internal characteristics (e.g., the subband profiles) and terminal currents. The use of the FUMS approach highly improves the efficiency of our simulator and makes it a practical model for extensive device simulation and design [4]. (The simulation of a ballistic SNWT with a 10nm gate length and a 3nm Si body thickness normally takes <15 minutes per bias point on one 1.2GHz ATHLON processor).

3.3 Results

In this section, we first verify the validity of the FUMS approach by comparing its results with those obtained by the UMS and CMS approaches. Then we adopt the FUMS as a simulation tool to explore device physics (i.e., both internal characteristics and terminal currents) of ballistic SNWTs with various cross-sections (e.g., triangular, rectangular and cylindrical).

3.3.1 Benchmarking of the FUMS approach

As mentioned in Sec. 3.2, for both CMS and UMS approaches, we need to solve a 2D Schrödinger equation at each slice of the SNWT to obtain the electron subbands and the corresponding eigenfunctions (modes). Fig. 3.2 shows the electron wavefunctions at

a slice of the SNWTs with a triangular, rectangular or cylindrical cross-section, respectively. After solving all the N_X 2D Schrödinger equations, the electron subband levels are obtained (see Fig. 3.3, circles). For the FUMS approach, however, only one 2D Schrödinger equation needs to be solved, and the subband profile can then be calculated by Eq. (3.42). Fig. 3.3 clearly illustrates that this approximation (solid lines) provides excellent agreement with the rigorous calculation (circles), which shows that the FUMS approach correctly computes the electron subbands in SNWTs.

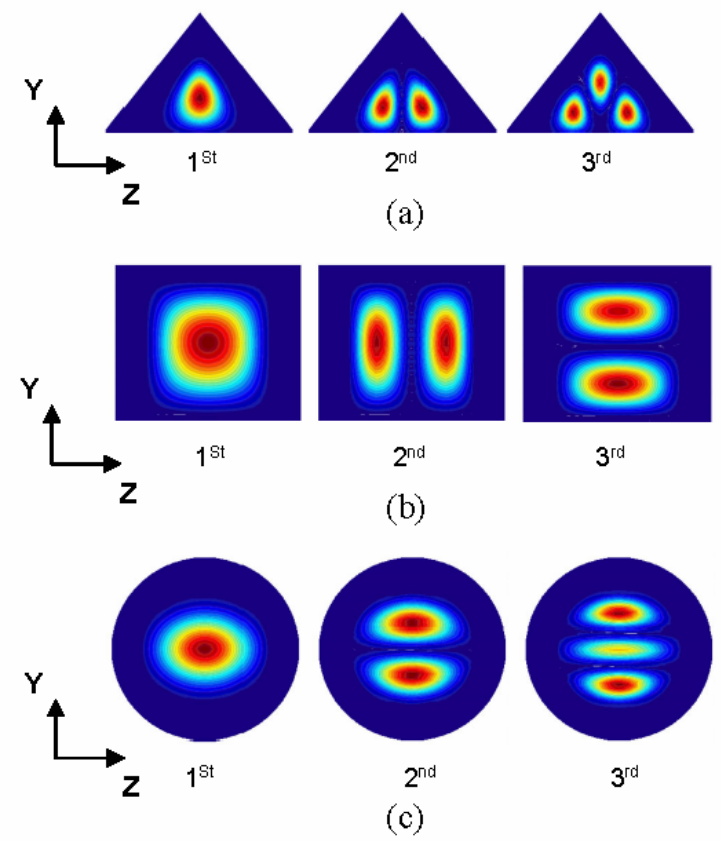


Fig. 3.2 The 2D modes (the square of the modulus of the electron wavefunctions in the (010) valleys) in a slice of (a) triangular wire (TW), (b) rectangular wire (RW) and (c) cylindrical wire (CW) transistors. For clarity, the SiO_2 substrates for TW and RW FETs are not shown here.

Figure 3.4 compares the computed I_{DS} vs. V_{GS} characteristics for the simulated cylindrical SNWT by the FUMS (dashed lines), UMS (circles) and CMS (crosses) approaches, respectively. It is clear that all the three approaches are in excellent agreement (<0.5% error), thus indicating that the FUMS approach, which has much higher computational efficiency than CMS and UMS, is an attractive simulation tool for modeling Si nanowire transistors. Although the sample device structure we use in Fig. 3.3 and Fig. 3.4 is a cylindrical SNWT, our conclusion is also applicable for SNWTs with arbitrary cross-sections (assuming the shape of the Si body is uniform along the x direction). In the following parts of this chapter, we will use the FUMS approach to investigate the device physics in various SNWTs.

3.3.2 Device physics and characteristics

The NEGF transport model we use in this chapter provides an opportunity to illustrate the LDOS of the simulated SNWTs. Fig. 3.5 shows the LDOS together with the electron subbands for a ballistic cylindrical SNWT with 10nm gate length and 3nm Si body thickness. Strong oscillations in the LDOS plot are clearly observed, which is due to the quantum mechanical reflection. [41] To be specific, the states injected from the drain are reflected off the drain-to-source barrier at the high drain bias and these reflected states strongly interfere with the injected ones. At the source end, the states injected at energies around the source barrier are also reflected and interfere. It should be noted that the occurrence of quantum inference in ballistic SNWTs relies on the quantum coherence (complete preservation of electron phase information) inside the devices. If scattering (dephasing mechanism) is included, as we will see in Appendix, the quantum interference and the oscillations in the LDOS are smeared out. In addition, the presence of states below the first electron subband is also visible in the LDOS plot, which is caused by source-to-drain tunneling [17].

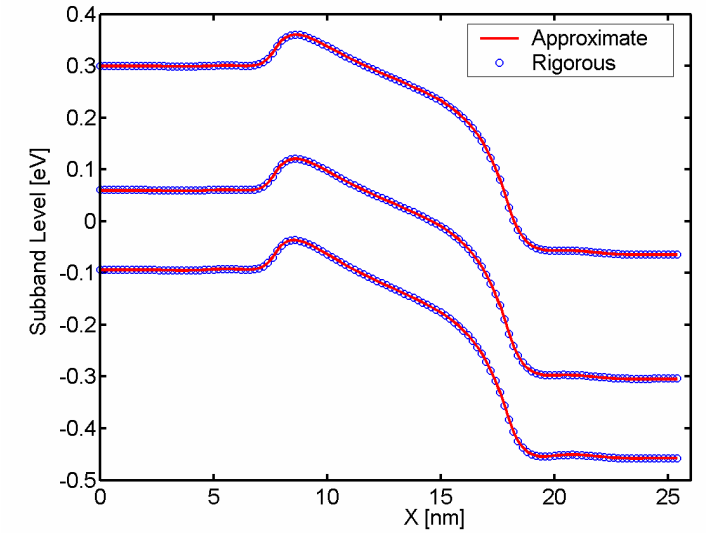


Fig. 3.3 The electron subband profile in a cylindrical SNWT with 10nm gate length (V_{GS}=0.4V and V_{DS}=0.4V). The silicon body thickness, T_{Si} (as shown in Fig. 3.1c), is 3nm, and the oxide thickness is 1nm. The S/D doping concentration is $2\cdot10^{20}$cm^{-3} and the channel is undoped. The solid lines are for the approximation method (solving a 2D Schrödinger equation only once) used in the FUMS approach, while the circles are for the rigorous calculation (solving 2D Schrödinger equations N_X times) adopted in the UMS and CMS approaches.

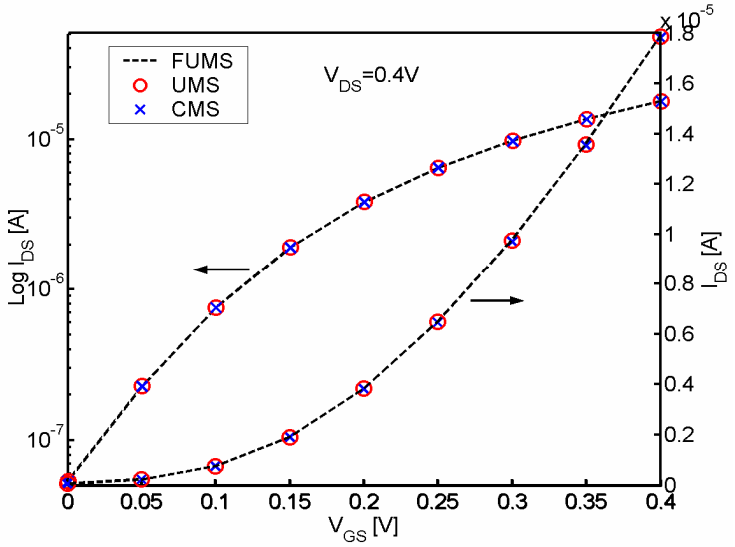


Fig. 3.4 The I_{DS} vs. V_{GS} curves for a cylindrical SNWT in the logarithm (left) and linear (right) scales (V_{DS}=0.4V). The device structure is the same as that in Fig. 3.3. The crosses are for the CMS approach, the circles are for the UMS approach, and the dashed lines are for the FUMS approach.

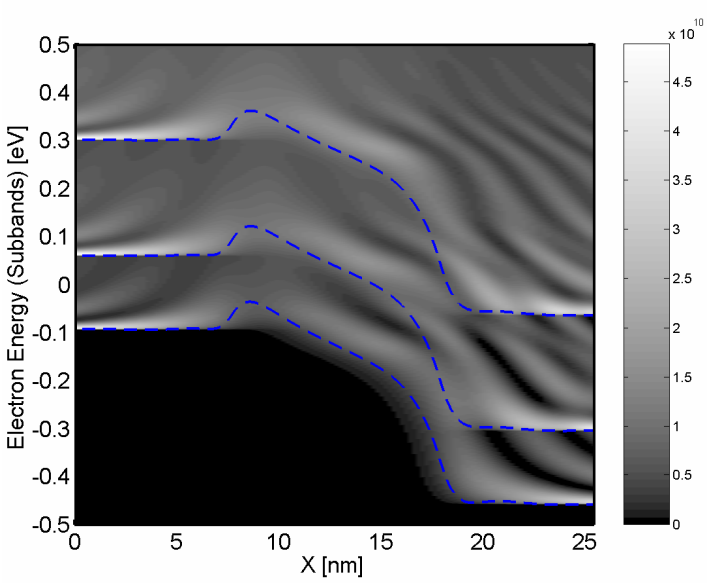


Fig. 3.5 The computed LDOS (in 1/(eV·m)) and electron subbands (dashed lines) of a ballistic cylindrical SNWT with 10nm gate length and 3nm Si body thickness (the details of the device geometry are described in the Fig. 3.3 caption). (V_{GS}=0.4V and V_{DS}=0.4V).

Figure 3.6 plots the 1D electron density (in m^{-1}) profile along the channel of the simulated cylindrical SNWT. It is clearly observed that the oscillations in the LDOS of the device result in an oscillation in the 1D electron density, even at the room temperature and more apparent at low temperature (77K). In general, such an oscillation in the electron density profile occurs in all kinds of transistors with 1D channels (e.g., the carbon nanotube transistor [70]). It is interesting to mention that there is no evident oscillation in the electron density profile in a planar MOSFET (see Fig. 8 on p. 3736 in [41]) although its LDOS also bears strong oscillations (see Fig. 4 on p. 3735 in [41]). The reason is that in a planar MOSFET there is a transverse direction (normal to both the Si/SiO_2 interfaces and the channel direction), in which the electron wavefunction is assumed to be a plane wave, thus resulting in numerous transverse modes in the device. These transverse modes wash out the oscillations in the LDOS and cause a smooth electron density profile. So the oscillation in the electron density profile is a special property of SNWTs as compared with planar MOSFETs.

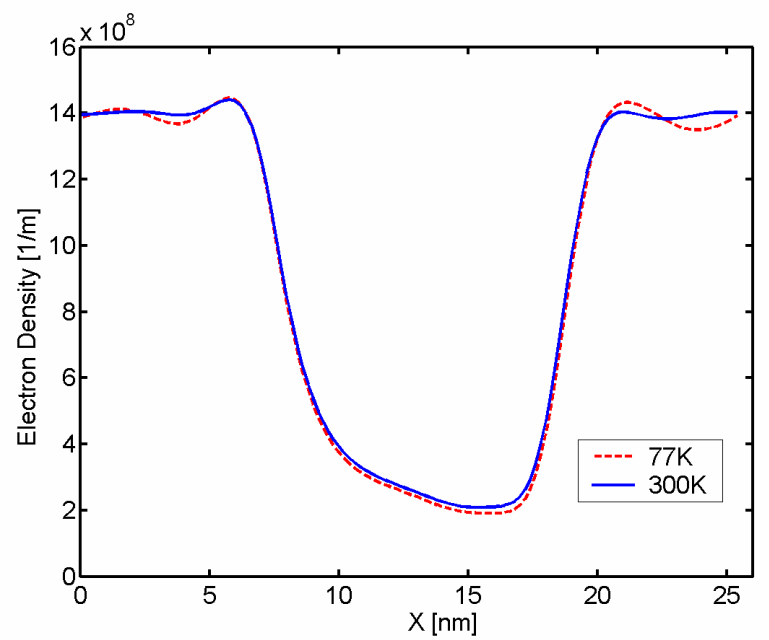


Fig. 3.6 The 1D electron density profile along the channel of the simulated cylindrical SNWT (the details of the device geometry are described in the Fig. 3.3 caption). The solid line is for T=300K while the dashed line is for T=77K. (V_{GS}=0.4V and V_{DS}=0.4V).

Figure 3.7 illustrates the transmission coefficient, calculated from Eqs. (3.38) and (3.39), for the simulated cylindrical SNWT. When the total electron energy increases above the source end of the first subband, the electrons start to be injected into the channel, so the transmission coefficient begins to increase from zero. As the electron energy continues to go up, the second and third subbands (modes) become conductive successively, which results the step-like shape of the transmission coefficient curve. We also observe that the transmission coefficient is above zero even when the total electron energy is below the top of barrier of the first subband, which is the evidence of source-to-drain tunneling [17].

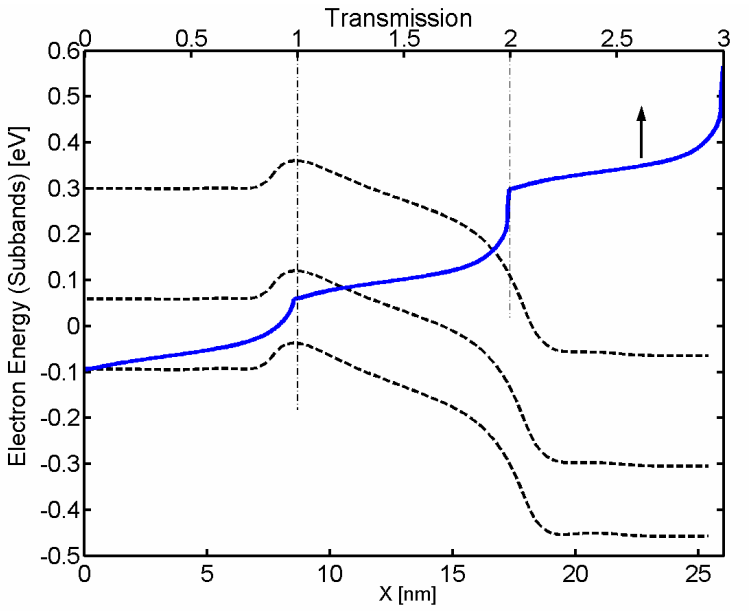


Fig. 3.7 The transmission coefficient and electron subbands in the simulated cylindrical SNWT (the details of the device geometry are described in the Fig. 3.3 caption). (V_{GS}=0.4V and V_{DS}=0.4V).

In Fig. 3.8, we compare the I_{DS} vs. V_{GS} characteristics for SNWTs with triangular, rectangular and cylindrical cross-sections. Two interesting phenomena are evidently visible:

1) The cylindrical wire (CW) and triangular wire (TW) transistors have higher threshold voltages, V_T (that is defined as $I_{DS}(V_{GS}=V_{TH})=10^{-8}$A when V_{DS}=0.4V), than the rectangular wire (RW). The reason is that the cross-sectional areas of the CW and the TW are smaller that that of the RW, which leads to stronger quantum confinement in the CW and the TW as compared with the RW.
2) The CW, a gate-all-around structure, offers the best subthreshold swing and the highest ON-OFF current ratio under the same gate overdrive, V_{GS}-V_T, due to its best gate control among all the three simulated structures.

These results clearly show that our simulator correctly treats the 3D electrostatics, quantum confinement and transport in ballistic SNWTs with arbitrary cross-sections.

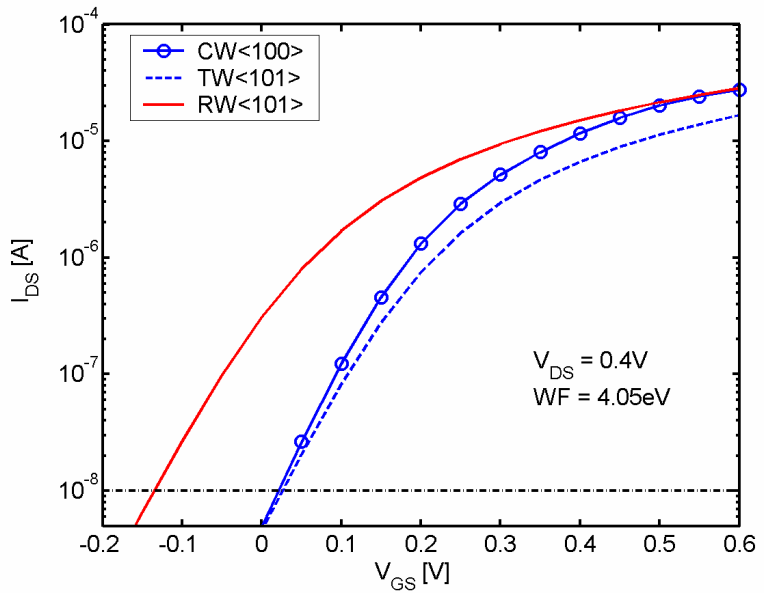


Fig. 3.8 The I_{DS} vs. V_{GS} curves for the triangular wire (TW) FET with <101> oriented channels, rectangular wire (RW) FET with <101> oriented channels and cylindrical wire (CW) FET with <100> oriented channels. (V_{DS}=0.4V). All the SNWTs have the same silicon body thickness (T_{Si}=3nm), oxide thickness (T_{ox}=1nm), gate length (L=10nm) and gate work function (WF=4.05eV). The Si body width, W_{Si}, of the RW is 4nm. In the calculation of the TW and RW FETs, whose channels are <101> oriented, the effective-masses of electrons in the (100) and (001) valleys are obtained from [71] as $m_x^* = 0.585m_e$, $m_y^* = 0.19m_e$ and $m_z^* = 0.318m_e$.

3.4 Summary

In this chapter, we presented a computationally efficient, three-dimensional quantum simulation of various ballistic silicon nanowire transistors based on the effective-mass approximation. [43] [44] The coupled/uncoupled mode space approaches [41] [45] [47] were adopted to decompose the 3D device Hamiltonian, which greatly reduces the simulation time while keeping excellent computational accuracy. The use of a fast uncoupled mode space approach further scales down the computational complexity and makes our simulator executable on a single processor. This enables our approach to be used as a practical 3D quantum model for extensive device simulation and design. [4]

Although we focused on ballistic simulations in this chapter, a simple treatment of scattering with the Büttiker probes [42] [57], previously applied to MOSFET simulations, can also be implemented into our SNWT simulator. The basic equations for the Büttiker probes and the relevant simulation results will be shown in Appendix.

4. BALLISTIC PERFORMANCE LIMITS AND SCALING POTENTIAL OF SILICON NANOWIRE TRANSISTORS

In this chapter, we explore the ballistic performance limits and scaling potential of silicon nanowire transistors (SNWTs) by using the simulation capability developed in Chapter 2 and Chapter 3 (based on the effective-mass approximation [43] [44]). Three different topics will be discussed. Sec. 4.1 shows a comparison between the upper performance limit of SNWTs with that of the planar double-gate (DG) MOSFET. [4] In Sec. 4.2, we propose a general approach to compare planar vs. non-planar (nanowire) FETs with the consideration of both Electrostatic integrity (gate control) and Quantum confinement (so called the 'EQ approach'). [58] Sec. 4.3 introduces a conceptual study of the channel material optimization for both planar MOSFETs and nanowire FETs based on the effective-mass approximation. [59]

4.1 The Ballistic Performance Limits of Silicon Nanowire Transistors

In this section, we adopt the ballistic simulator that has been described in Chapter 3 to investigate the upper performance limits of SNWTs with various cross-sections (i.e., triangular, rectangular and cylindrical). The simulation results are compared with those for a ballistic planar DG MOSFET [39] [48] [72] [73], which is simulated by a two-dimensional (2D) quantum simulator, NanoMOS-2.5 [40].

4.1.1 Device structures

Silicon nanowire transistors with various types of cross-sections are being extensively explored by a number of experimental groups. In [13], the authors reported

triangular parallel wire channel transistors; the device was built on a Si (001) wafer and the Si body layer was etched along the (111) surfaces, so the cross-section of the Si body becomes an isosceles triangle and the channel of the device is <110> oriented. At the same time, different types of tri-gate/gate-all-around FETs were fabricated by using wires with *rectangular* cross-sections. [11] [12] [14] [15] The diameters of these wire transistors are around 30-100nm and the gate lengths are >50nm. In this section, we extract the geometry configurations from those experimental structures and project them to the device structures at the scaling limit, where the transistor gate length is <10nm [17] and the Si body of these wire FETs becomes small nanowires with triangular/rectangular cross-sections. In addition, we also simulate the cylindrical SNWT [16], a gate-all-around structure, which offers the optimum gate control and scaling potential among all kinds of SNWTs. In Fig. 4.1, the cross-sections of all the simulated SNWTs and planar DG MOSFETs are schematically illustrated.

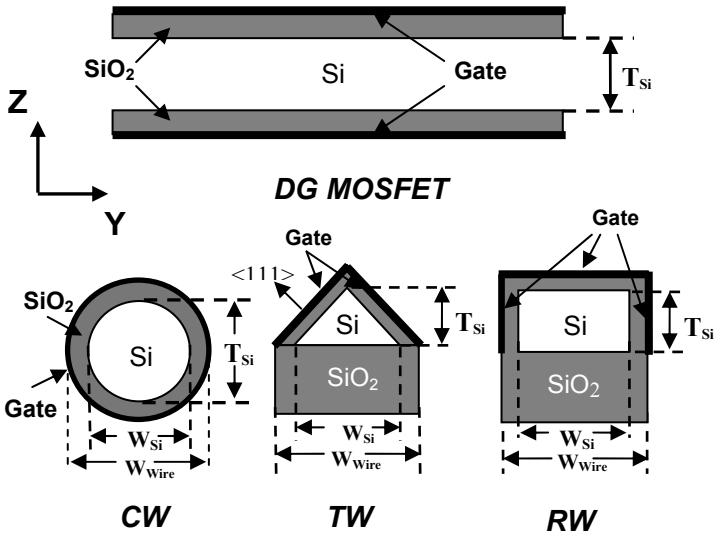


Fig. 4.1 The cross-sections of the symmetric planar double-gate (DG) MOSFET and various nanowire FET structures – the cylindrical wire (CW), triangular wire (TW) and rectangular wire (RW) FETs. (The channel (x) direction is perpendicular to the paper surface.) Here T_{Si} is the silicon body thickness, W_{Si} is the silicon body width and W_{Wire} is the wire width. For the CW FET, T_{Si}= W_{Si} is equal to the diameter of the circular Si body.

4.1.2 Methodology

The simulation tool for SNWTs we use in this section is the ballistic simulator we developed in Chapter 3. For the simulated SNWT structures, the fast uncoupled mode space (FUMS) approach offers excellent accuracy (<0.5%) as compared with the rigorous coupled mode space (CMS) approach. Therefore, we adopt the FUMS model for all the SNWT simulations in this work, which greatly reduces the computation time. For the planar DG MOSFETs, we adopt a 2D quantum simulator, NanoMOS-2.5, which was developed by Z. Ren and coworkers [40].

In order to compare SNWTs with planar MOSFETs, we need to convert the current for a single nanowire into current per unit width. In this work, based on the device structures in the experiments [13] [15], we assume that in a SNWT a number of nanowires are placed in parallel to form the channel of the FET. So the current for a single wire can be converted into current per unit width by dividing it by the wire spacing parameter, ρ, which is the distance between the geometrical centers of adjacent wires. Fig. 4.2 shows the definitions of the wire spacing parameters for different wires. It should be noted that the wire spacing is assumed to be W_{Wire} for a triangular SNWT and $2W_{Si}$ for a rectangular SNWT, which are both extracted from the relevant experiments [13] [15]. For a cylindrical SNWT, we select a minimum wire spacing, $\rho = W_{Wire}$, to achieve the optimum device metrics (to be discussed in Subsection 4.1.3). Hence, the cylindrical SNWT with a minimum wire spacing provides an upper performance limit for all types of FETs (i.e., both SNWTs and planar MOSFETs).

By dividing the current for a single nanowire by the wire spacing parameter, we actually guarantee that all the devices (i.e., both SNWTs and planar MOSFETs) have the same integration density. The explanation is as follows. In our comparison, the SNWTs and the DG MOSFETs always possess the same gate (channel) length. For this reason, to achieve an equal integration density, a SNWT and a DG MOSFET must have an identical channel width, W_0. Consequently, the current for the SNWT is obtained as

$$I_{SNWT}^{W_0} = I_W \cdot \frac{W_0}{\rho}, \tag{4.1}$$

where I_W is the current for a single nanowire, which is an output of our SNWT simulator, while the current for the DG MOSFET is evaluated as

$$I_{MOS}^{W_0} = I_M \cdot W_0, \tag{4.2}$$

where I_M (in μA/μm), the current per unit width for the simulated DG MOSFET, is directly computed by the MOSFET simulator, NanoMOS-2.5 [40]. When comparing the currents for the two devices, i.e., $I_{SNWT}^{W_0}$ vs. $I_{MOS}^{W_0}$, we can simply eliminate W_0 from both Eq. (4.1) and Eq. (4.2). Thus, the quantities we are actually comparing are I_W / ρ vs. I_M, which implies that we should convert the current for a single wire (I_W) into current per unit width by dividing it by the wire spacing parameter (ρ), if we require the SNWT and the DG MOSFET have an equal integration density.

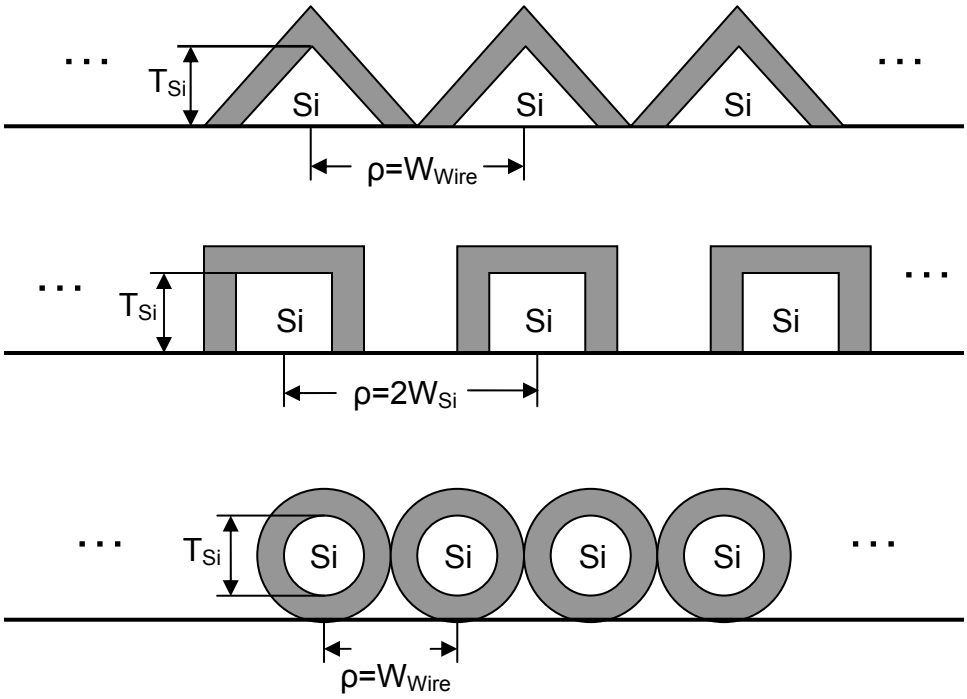


Fig. 4.2 The definitions of wire spacing parameter, ρ, for SNWTs with various cross-sections (from the top to the bottom – triangular, rectangular and cylindrical).

The basic concepts of our comparison methodology are briefly summarized as below. As we know, the performance of an integrated circuit chip is evaluated by its integration density, power consumption and switching speed [2]. In our comparison between SNWTs and DG MOSFETs, we fix the integration density of both devices (as described earlier) as well as the static power density per unit chip area (we adopt a different gate work function for each device to achieve a specified OFF-current, 10μA/μm [74]). Then we compare the switching speed of each device; there are two parameters that determine the switching speed of the circuit: 1) the ON-current, which determines the interconnect delay, and 2) the intrinsic device (transistor) delay, τ. [2] For a planar MOSFET, an analytical expression of τ is obtained as [2]

$$\tau = \frac{C_G\left(V_{DD} - V_T\right)}{I_{ON}}, \tag{4.3}$$

where V_T is the threshold voltage and C_G is the gate capacitance, which, at most cases (i.e., if quantum confinement is not so significant), can be simply expressed as [2]

$$C_G = \frac{\varepsilon_{ox}}{T_{ox}} W, \tag{4.4}$$

where T_{ox} is the oxide thickness, ε_{ox} is the SiO_2 dielectric constant and W is the MOSFET channel width. Eqs. (4.3) and (4.4) are widely used to compute the intrinsic device delay of conventional MOSFETs. For SNWTs, however, which are multi-gate structures, Eq. (4.4) is not valid anymore, especially when the Si body thickness/width is comparable to the oxide thickness. Consequently, a new equation for the intrinsic device delay, τ, is adopted in this work,

$$\tau = \frac{Q_{ON} - Q_{OFF}}{I_{ON}}, \tag{4.5}$$

where Q_{ON} (Q_{OFF}) is the total electron charge inside the device at the ON (OFF) state. (For conventional planar MOSFETs, Eq. (4.5) gives the same results as the previous equations, Eqs. (4.3) and (4.4)). Since this is a simulation study, Q_{ON} and Q_{OFF} can be directly obtained from the numerical simulators, so Eq. (4.5) can be easily used to calculate the intrinsic device delays of any types of FETs (e.g., planar MOSFETs,

SNWTs with arbitrary cross-sections, and other types of FETs like carbon nanotube FETs [5] [6] or molecular transistors [9]).

4.1.3 Results

In Table 4.1, we list the device characteristics of the simulated DG MOSFETs and SNWTs with the same gate length (10nm), oxide thickness (1nm) and Si body thicknesses (3nm and 6nm). The results show that when the Si body is thin (3nm, see Table 4.1a), gate control for the DG MOSFET is good, which endows the DG MOSFET with good performance at the 10nm-scale. The triangular wire (TW) and rectangular wire (RW) FETs (with no bottom gate) provide comparable subthreshold swings, better DIBL and intrinsic device delays, and slightly lower ON-currents. If we choose a thicker body (6nm, see Table 4.1b), which is easier to realize in practice and offers less device-device variation, the DG MOSFET loses its good gate control (at the 10nm-scale) because the two gates are relatively far apart. For thick Si bodies, the TW and RW FETs offer better device metrics, both in the subthreshold region and at the ON-state. So we conclude that the SNWT offers more benefits when the silicon body is relatively thick.

Another interesting phenomenon is that as the channel length decreases from 10nm to 4nm, the advantages of the SNWTs over the DG MOSFET (with a 3nm-thick body) become more and more significant (Fig. 4.3). This occurs because the silicon body looks thicker (compared with the channel length) as the channel length scales down. As a result, the SNWT provides better scaling capability than the planar DG MOSFET. We also observed that both the TW/RW FETs and DG MOSFETs operate well below the cylindrical SNWTs with a minimum wire spacing, which provide the optimum gate control (although they could be difficult to fabricate in practice). These results show that there is still room for optimizing gate geometry configuration. For rectangular SNWTs, we compare the device design in <100> oriented channel with that in <110> one and observe that the <100> oriented channel does offer 10-15% higher ballistic ON-current (see Table. 4.1). (A similar effect was illustrated in [12] based on a calculation of carrier mobility.)

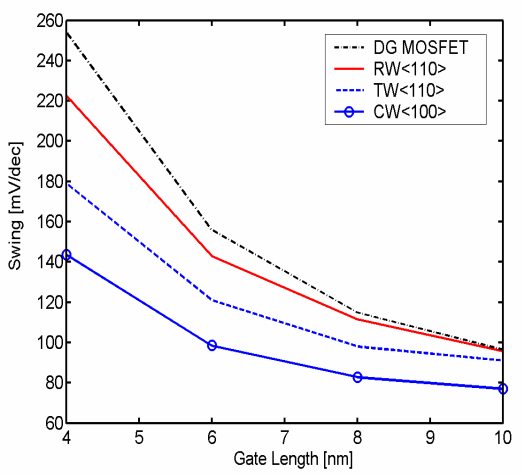
Table 4.1 Geometrical parameters and device characteristics of the DG MOSFETs and various SNWTs. (V_{DS}=0.4V). The silicon body thicknesses, T_{Si} (as shown in Fig. 4.1), are (a) 3nm and (b) 6nm. For all the simulated structures, the source/drain (S/D) doping concentration is $2 \cdot 10^{20} cm^{-3}$ and the channel is undoped. There is no S/D overlap with channel and the gate length is always equal to the channel length (L=10nm). The currents for SNWTs are divided by the wire spacing parameter, ρ. For each device, a proper gate work function is selected to obtain a specified *OFF-current*, 10μA/μm [74]. No parasitic capacitance is included in the calculation of intrinsic device delays.

(a) T_{Si}=3nm, T_{ox}=1nm and L=10nm

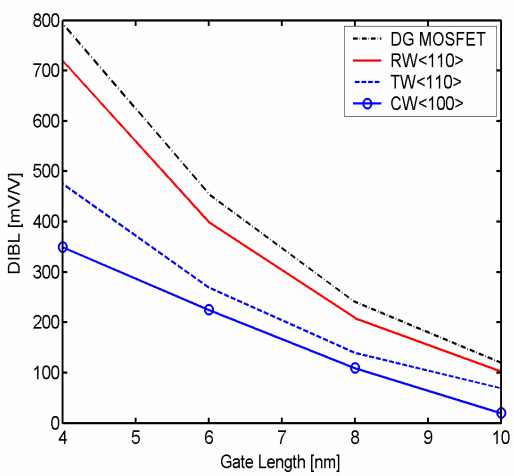
	DG	TW <110>	RW <110>	RW <100>	CW <100>
W_{Si} (nm)	n/a	4.2	4.0	4.0	3.0
W_{Wire} (nm)	n/a	6.7	6.0	6.0	5.0
ρ (nm)	n/a	6.7	8.0	8.0	5.0
I_{ON} (μA/μm)	2000	1630	1460	1650	3550
S (mV/dec)	97	91	96	97	77
DIBL (mV/V)	120	69	104	109	21
Delay (ps)	0.075	0.056	0.069	0.062	0.046

(b) T_{Si}=6nm, T_{ox}=1nm and L=10nm

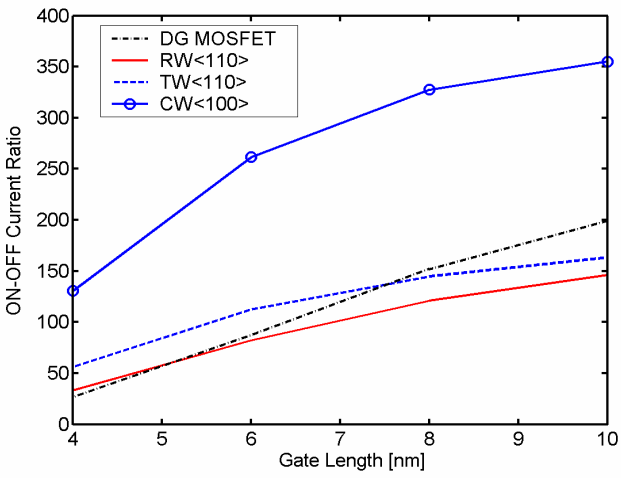
	DG	TW <110>	RW <110>	RW <100>	CW <100>
W_{Si} (nm)	n/a	8.5	4.0	4.0	6.0
W_{Wire} (nm)	n/a	10.9	6.0	6.0	8.0
ρ (nm)	n/a	10.9	8.0	8.0	8.0
I_{ON} (μA/μm)	930	940	1680	1860	2290
S (mV/dec)	144	122	101	102	96
DIBL (mV/V)	396	195	157	163	129
Delay (ps)	0.167	0.114	0.092	0.082	0.079



(a) Swing vs. gate length



(b) DIBL vs. gate length



(c) ON-OFF current ratio vs. gate length

Fig. 4.3 The device characteristics (i.e., (a) swing, (b) DIBL, and (c) ON-OFF current ratio) vs. gate length curves for SNWTs and DG MOSFETs (T_{Si}=3nm and T_{ox}=1nm). For each device structure, a proper gate work function is selected to achieve a specified OFF-current of 10μA/μm [74]. The wire spacing parameters for the SNWTs are listed in Table 4.1a.

Figure 4.4 illustrates the variations of OFF (ON) currents for the DG MOSFET and SNWTs due to the fluctuation of T_{Si}. It is shown that the TW and CW FETs suffer larger performance variations than the DG MOSFET because their Si body widths (W_{Si}) and body thickness (T_{Si}) always fluctuate in the same direction (i.e., both increase or decrease). For the RW FET, W_{Si} and T_{Si} are uncorrelated. If we only consider the

fluctuation in T_{Si}, the results show that the RW FET displays comparable – even smaller performance variations than the DG MOSFETs, which can be an advantage of the RW structure.

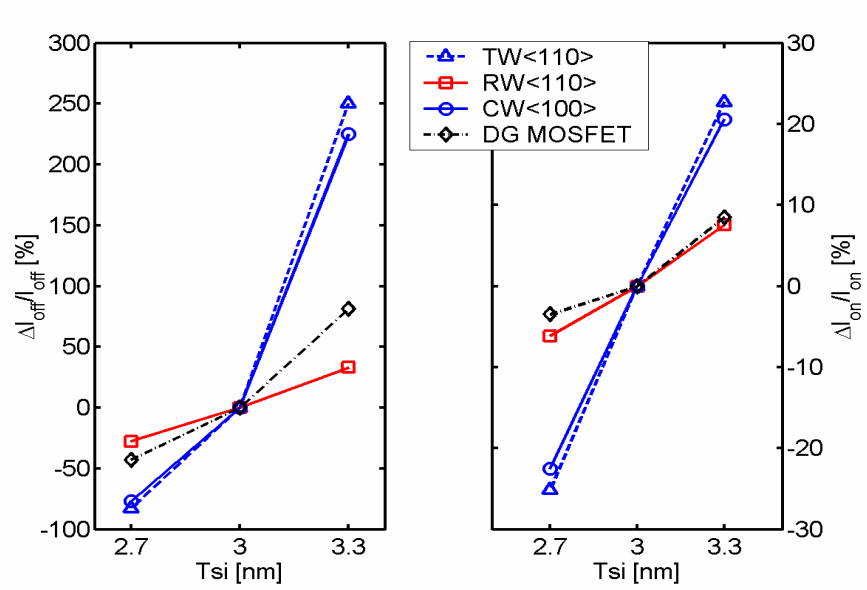


Fig. 4.4 The variations of OFF-current (left) and ON-current (right) vs. T_{Si} curves for the simulated DG MOSFETs, TW<110>, RW<110> and CW<100> FETs in Table. 4.1a (L=10nm and T_{ox}=1nm). For each type of device structure, a proper gate work function is selected to achieve a specified OFF-current, 10μA/μm [74], for the *nominal* (T_{Si}=3nm) device.

Finally, we examine the role of wire spacing (wire density) on SNWT performance. Fig. 4.5 plots the ON-OFF current ratio vs. ρ (left) and the subthreshold swing vs. ρ (right) curves. Since we design each device for a *fixed* OFF-current (10μA/μm [74]), a different gate work function is needed for each wire spacing parameter, ρ. So the subthreshold swing and ON-OFF current ratio become dependent on ρ. The results show that the ON-OFF current ratio degrades and the subthreshold swing increases as the wire spacing is raised. As a result, maintaining a high wire density is important for a SNWT to achieve a high drive current and a small subthreshold swing.

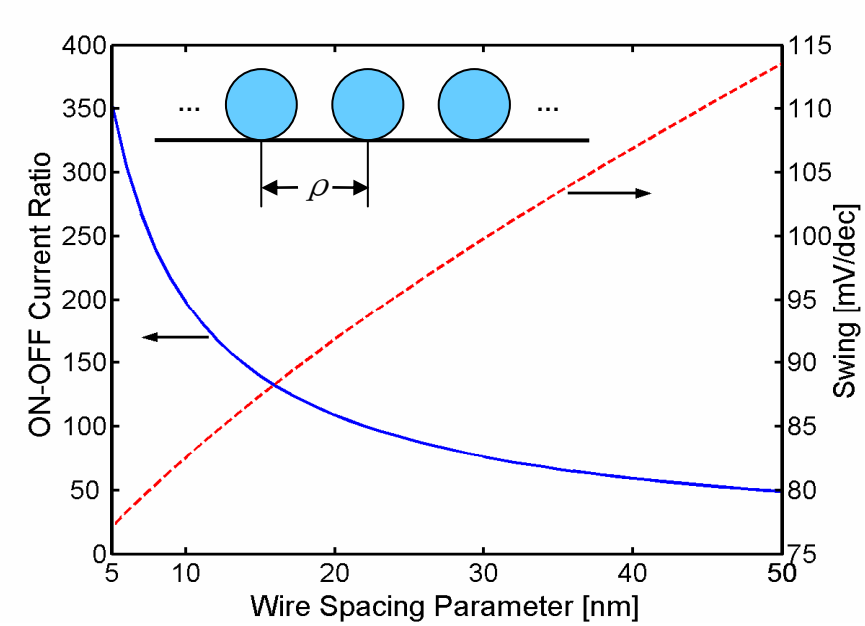


Fig. 4.5 The dependence of device performance (i.e., ON-OFF current ratio (left) and subthreshold swing (right)) on the wire spacing parameter. The simulated device is a CW<100> FET with T_{Si}=3nm, T_{ox}=1nm and L=10nm.

4.1.4 Conclusions

In this section, we explored the device characteristics and scaling potential of ballistic silicon nanowire transistors by using the three-dimensional (3D) numerical simulator developed in Chapter 3. The results were compared with those of the planar double-gate MOSFET and it shows that

1) Triangular and rectangular nanowire FETs, without a bottom gate, offer better electrostatics and scaling potential than planar DG MOSFETs, especially when the silicon body is thick.
2) The cylindrical SNWT structures with a minimum wire spacing offer substantially better device metrics than any other structures (i.e., the DG MOSFET or the triangular/rectangular SNWTs), which implies that there is still room for the optimization of gate geometry configurations.

3) Maintaining a high wire density is important for nanowire FETs (or any FETs with one-dimensional (1D) channels) to achieve good device performance.
4) While providing better electrostatic integrity (gate control), the SNWT structure may also suffer larger device performance variation than the planar DG MOSFET due to the stronger quantum confinement in nanowires. This effect will be seriously considered in Sec. 4.2.

4.2 Assessment of the Scaling Potentials of SNWTs vs. Planar MOSFETs Based on a General Approach

4.2.1 Introduction

In Sec. 4.1, we performed a 3D numerical simulation to show that SNWTs do offer better gate control (e.g., a better subthreshold swing and a higher ON-OFF current ratio) than planar DG MOSFETs. The results also showed, however, that the SNWT structures suffer larger threshold voltage variation than the planar DG MOSFET due to stronger quantum confinement in nanowires. In other words, the advantage of multi-gate structures (SNWTs) in terms of electrostatic integrity may be compromised by their sensitivity to quantum confinement (which aggravates threshold voltage variation and surface roughness scattering). This effect should be seriously considered when selecting gate geometries at the scaling limit.

In this section, we propose a general approach, which considers both Electrostatic integrity and Quantum confinement (so called the "EQ approach"), to compare the device performance of nanoscale Si FETs with various gate geometry configurations (i.e., planar MOSFETs vs. SNWTs). This approach is based on the use of a L_Q vs. L_E plot, where L_Q is the quantum confinement length to be defined below, and L_E denotes the electrostatic scale length. As defined in [75] [76] [77], the electrostatic scale length, L_E, describes the significance of the 2D (in planar FETs) or 3D (in non-planar FETs) effects such as DIBL and V_T rolloff in the device [2]. To be specific, the larger the ratio $L_E/L_{Channel}$, the more serious the 2D (3D) effects. (It was shown in [75] that devices with acceptable

characteristics could be designed as long as the gate length is greater than ~1.5× the electrostatic scale length.) The quantum confinement length, L_Q, is defined as

$$L_Q = \left| k_B T / (dE_0 / dT_{Si}) \right|, \tag{4.6}$$

where E_0 is the lowest electron subband level in the quantum mechanically confined Si body (to be discussed later), T_{Si} is the Si body thickness, k_B is Boltzmann constant and T is the ambient temperature. (When computing the L_Q of a SNWT, we assume the Si body width, W_{Si}, as shown in Fig. 4.1, to be proportional to T_{Si}, so changing T_{Si} does not alter the shape of the cross-section but only its area). The physical meaning of L_Q is explained as follows. If the silicon body thickness variation in a FET is ΔT_{Si} (caused by process variations), then the threshold voltage variation, ΔV_T, can be obtained approximately as

$$\Delta V_T \cong \left| \Delta E_0 / q \right| = \left| \frac{dE_0}{dT_{Si}} \frac{\Delta T_{Si}}{q} \right| = \frac{k_B T \cdot \Delta T_{Si}}{q \left| k_B T / (dE_0 / dT_{Si}) \right|} = \frac{\Delta T_{Si}}{L_Q} \frac{k_B T}{q}, \tag{4.7}$$

where q is the charge of a single electron. As we can see from Eq. (4.7), the larger the L_Q is, the smaller the threshold voltage variation will be. So L_Q explicitly represents the sensitivity of threshold voltage variation to Si body thickness fluctuation. Based on the definitions of L_Q and L_E, it is easily concluded that a well-tempered device structure should have a large L_Q (e.g., $L_Q > \Delta T_{Si}$) as well as a small L_E (e.g., $L_E < L_{Channel}$). As a result, when plotting the L_Q vs. L_E curves for different device structures on the same figure, the one that has the largest L_Q at the same L_E offers the best device characteristics (considering both electrostatics and quantum confinement).

In the following subsections, we first provide the equations we use to compute the L_Q and L_E for different nanoscale planar/non-planar structures, and then we compare their scaling potentials based on our general approach.

4.2.2 Device structures

There are three types of device structures to be investigated in this work: 1) the symmetric planar DG MOSFET, which is the optimum planar structure in terms of gate control, [39] [48] [72] [73] 2) the gate-all-around cylindrical wire (CW) FET that

provides the best electrostatic integrity among SNWTs [4] [16] [78], and 3) the rectangular wire (RW) FET, a tri-gate structure, which is being extensively explored in recent experimental work [11] [12]. (The device structures have been illustrated in Fig. 4.1.) Two wafer orientations, (001) and (011), are considered in the simulations. For the (001) wafer, the z direction (shown in Fig. 4.1) is in $[001]$, while for the (011) wafer the z direction is in $[011]$. In CW FETs, the W_{Si}/T_{Si} ratio is always equal to 1. In contrast, the W_{Si}/T_{Si} ratio in the RW structure can take arbitrary values. In practice, to make all three gates effective in modulating the channel conductance, the W_{Si} in the RW FET is normally selected to be comparable to T_{Si}. [11] In this work, for simplicity, $W_{Si}/T_{Si}=1$ is always assumed for the RW structure.

4.2.3 Basic equations for the calculation of L_E and L_Q

A) *Equations for the Electrostatic Scale Length L_E:*

The electrostatic scale length, L_E, is a widely used parameter to describe the electrostatic coupling between the drain bias and the potential inside the channel (i.e., the larger the L_E, the stronger the coupling). [16] [75] [76] [77] For a symmetric DG MOSFET, as indicated in [75], the L_E satisfies the equation

$$1=\frac{\varepsilon_{Si}}{\varepsilon_{ox}}\tan(\pi T_{ox}/L_E)\tan(\pi T_{Si}/2L_E), \tag{4.8}$$

where T_{ox} is the oxide layer thickness, ε_{Si} and ε_{ox} are the dielectric constants for Si and SiO_2, respectively. By numerically solving Eq. (4.8), the L_E is computed for given T_{Si} and T_{ox}.

For a cylindrical wire (CW) FET, following the same procedure as in [75], we find that the L_E satisfies the following equation

$$\varepsilon_{Si}\frac{J_0\left(-\pi T_{ox}/L_E+\chi_{01}\right)}{J_1\left(-\pi T_{ox}/L_E+\chi_{01}\right)}-\varepsilon_{ox}\frac{J_0\left(\pi T_{Si}/2L_E\right)}{J_1\left(\pi T_{Si}/2L_E\right)}=0, \tag{4.9}$$

where $J_0(x)$ and $J_1(x)$ are both Bessel functions [79] and $\chi_{01}=2.4048$.

In general, to obtain L_E for a SNWT with arbitrary cross-sections, a numerical solution of the 3D Laplace's equation,

$$\nabla \cdot \left(\varepsilon \left(y,z \right) \nabla V \left(x,y,z \right) \right) = 0 , \tag{4.10}$$

is necessary (here $V\left(x,y,z\right)$ is the electrostatic potential and $\varepsilon\left(y,z\right)$ denotes the dielectric coefficient in the structure). To do this, the solutions to Eq. (4.10) are decomposed into two domains: one in the channel (x) direction and one at the cross-section (the y-z plane, see Fig. 4.1) of the wire (i.e., $V\left(x,y,z\right) = X\left(x\right) \cdot u\left(y,z\right)$). Thus, two separate equations are obtained as

$$\frac{\partial^2 X\left(x\right)}{\partial x^2} = k_n^2 X\left(x\right), \tag{4.11}$$

and,

$$-\nabla_{yz} \cdot \left(\varepsilon \left(y,z \right) \nabla_{yz} u\left(y,z\right) \right) = k_n^2 \varepsilon \left(y,z \right) \cdot u\left(y,z\right), \tag{4.12}$$

where k_n are coefficients to be determined by satisfying the boundary conditions (to be discussed below). As described in [75], the electrostatic scale length, L_E, is defined as $k_1 = \pi / L_E$, where k_1^2 can be viewed as the lowest eigenvalue of Eq. (4.12). To evaluate k_1, we discretize Eq. (4.12) by the finite element method (FEM) [45] subject to the boundary conditions, $u\left(y,z\right) = 0$ at the gate contacts and $\vec{n} \cdot \left(\vec{\nabla} u\left(y,z\right) \right) = 0$ elsewhere. (For the simulated RW structure, which has a floating SiO_2 substrate, we follow previous work [76] [80] [81] and assume that the buried oxide substrate is thick enough so the electric field at the interface between the Si body and the buried oxide substrate is negligible. The validity of this approximation will be discussed in Subsection 4.2.4.) Thus, Eq. (4.12) becomes the following linear system

$$\left[A\right]_{N \times N} \left[U\right]_{N \times 1} = k_n^2 \left[S\right]_{N \times N} \left[U\right]_{N \times 1}, \tag{4.13}$$

where N is the number of the nodes in the simulated cross-section (the 2D mesh in the y-z plane) excluding those at the gate contacts (where $u\left(y,z\right) = 0$). By evaluating the eigenvalues of Eq. (4.13), k_1^2 (the lowest eigenvalue) and then the L_E are computed. (To

verify the validity of this approach, we used it to compute the L_E for the CW structure and obtained the same result as that from Eq. (4.9).)

B) *Equations for the Quantum Confinement Length L_Q:*

According to Eq. (4.6), to evaluate the quantum confinement length, L_Q, of a device, we first need to obtain an expression for the lowest electron subband level, E_0. To do this, an ellipsoidal parabolic energy band for electrons is assumed and the effective-mass approximation [43] [44] is adopted in this work. For simplicity, we neglect the penetration of electron wavefunction into the oxide layers (i.e., the electron wavefunction is zero at the Si/SiO_2 interfaces). We also assume that the potential profile variations in the cross-section (see Fig. 4.1) are negligible, which is appropriate for ultra-thin-body structures as confirmed by self-consistent Schrödinger-Poisson simulations. With these assumptions, an analytical expression can be obtained for a planar SOI MOSFET (e.g., the DG MOSFET) as

$$E_0 = \frac{\hbar^2 \pi^2}{2 m_z^* T_{Si}^{\ 2}}, \tag{4.14}$$

where m_z^* is the effective-mass in the z direction ($m_z^* = m_l = 0.98 m_e$ for the (001) wafer and $m_z^* = 2 m_l m_t / (m_l + m_t) = 0.318 m_e$ [71] for the (011) wafer). Inserting Eq. (4.14) into Eq. (4.6), the L_Q for a planar MOSFET is evaluated as

$$L_Q = \left| k_B T / (dE_0 / dT_{Si}) \right| = \frac{k_B T m_z^* T_{Si}^{\ 3}}{\hbar^2 \pi^2}. \tag{4.15}$$

For a SNWT, in general, a 2D Schrödinger equation (in the y-z plane) needs to be numerically solved to obtain E_0. As mentioned earlier, we assume that the Si body width is always proportional to T_{Si} when calculating L_Q, so the E_0 in a SNWT is also proportional to $1/T_{Si}^{\ 2}$ (as in a planar FET), that is,

$$E_0 = \frac{C}{T_{Si}^{\ 2}}, \tag{4.16}$$

where C is independent of T_{Si} and it can be evaluated after numerically computing E_0. (To be concise, the mathematical proof of Eq. (4.16) is not shown here.) The value of C depends on the electron effective-masses in the y and z directions and the geometry of the

wire cross-sections (e.g., the shape and the W_{Si}/T_{Si} ratio). (The C values for the simulated SNWT structures are listed in Table 4.2.) Inserting Eq. (4.16) into Eq. (4.6), the quantum confinement length, L_Q, for a SNWT is obtained as

$$L_Q = \frac{k_B T}{2C} T_{Si}^{\ 3}. \tag{4.17}$$

It is clearly shown in Eqs. (4.15) and (4.17) that in either a planar or a non-planar FET, the quantum confinement length, L_Q, is proportional to T_{Si}^3, thus indicating that L_Q is quite sensitive to Si body thickness. So a trade-off between L_E and L_Q arises; in order to achieve better electrostatic integrity (smaller L_E), a smaller body thickness (T_{Si}) is preferred, which will, however, cause larger threshold voltage variation (smaller L_Q, see Eq. (4.7)). As we will see, the approach proposed in this section provides an opportunity to properly evaluate this trade-off when comparing different gate geometry configurations at the scaling limit.

Table 4.2 The C values (defined by Eq. (4.16)) for the simulated CW and RW ($W_{Si}/T_{Si}=1$) structures. Two wafer orientations, (001) and (011), are considered in this calculation.

	CW		RW	
	(001)	(011)	(001)	(011)
C [eV-nm^2]	2.70	2.70	2.35	2.11

4.2.4 Results

Figure 4.6 (left) plots the electrostatic scale length L_E vs. Si body thickness T_{Si} curves for the cylindrical wire (CW) FET (dashed) and the symmetric planar double-gate (DG) MOSFET (solid) on the (001) wafer. It is clear that the CW FET, a gate-all-around structure, has significantly smaller electrostatic scale length than the planar DG MOSFET

at the same Si body thickness, thus indicating that the CW FET offers better gate control [4] [16]. At the same time, Fig. 4.6 (right) illustrates that the CW FET (dashed) has much smaller quantum confinement length than the DG MOSFET (solid), which implies stronger quantum confinement and consequently larger threshold voltage variation for the CW FET [4]. It is important to mention that these qualitative conclusions inferred from the L_E vs. T_{Si} and L_Q vs. T_{Si} curves are consistent with those obtained from the detailed numerical simulations in Sec. 4.1. It shows that L_E and L_Q are well-defined parameters that can be used to represent the key metrics (i.e., gate control and quantum confinement) of the device structure.

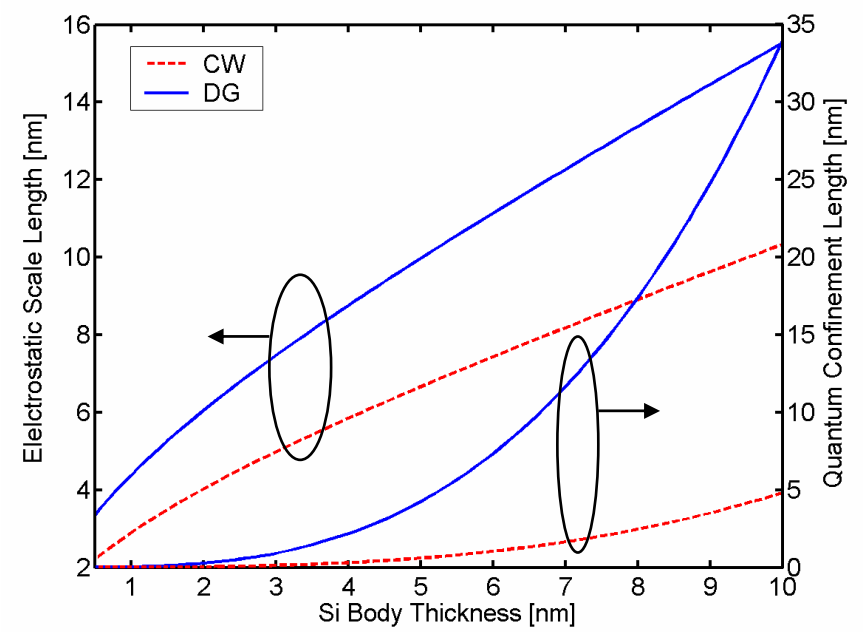


Fig. 4.6 The electrostatic scale length L_E vs. T_{Si} curves (left) and the quantum confinement length L_Q vs. T_{Si} curves (right) for the cylindrical wire (CW) FET (dashed) and the symmetric planar DG MOSFET (solid) on the (001) wafer. The geometry configurations of the two structures have been illustrated in Fig. 4.1 (the oxide thickness is fixed to be 1nm) and the effective-masses we use in this calculation are $m_y^* = 0.19m_e$ and $m_z^* = 0.98m_e$.

In Fig. 4.7, we plot the L_Q vs. L_E curves for the CW FET (dashed) and the planar DG MOSFET (solid) on the (001) wafer. Interestingly, the two curves are quite close to each other, which indicates that the CW FET and the DG MOSFET perform equally well.

Based on the results shown in Figs. 4.6 and 4.7, we conclude that the cylindrical wire geometry has much better electrostatic integrity than the planar DG structure, but this is offset by the fact that the CW geometry also results in much stronger quantum confinement. On balance, the two structures are nearly equal in their scaling potential (according to our EQ approach) for the (001) wafer.

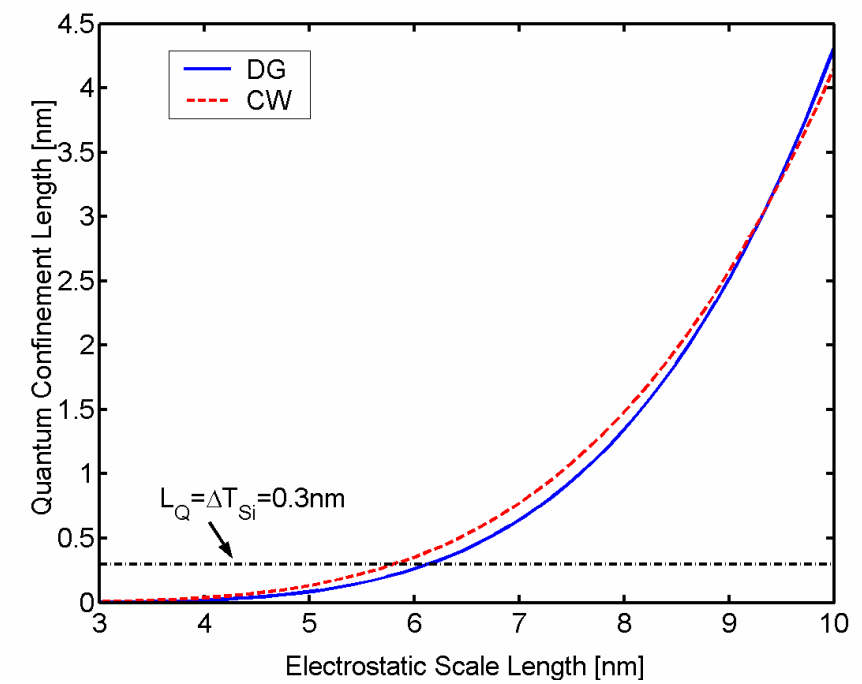


Fig. 4.7 The L_Q vs. L_E plot for the cylindrical wire (CW) FET (dashed) and the symmetric planar DG MOSFET (solid) on the (001) wafer. The geometry configurations of the two structures have been illustrated in Fig. 4.1 (the oxide thickness is fixed to be 1nm) and the effective-masses we use in this calculation are $m_y^* = 0.19m_e$ and $m_z^* = 0.98m_e$.

The L_Q vs. L_E plot (e.g., Fig. 4.7) can also be used to identify the minimum electrostatic scale length (L_{min}) for an assumed Si body thickness variation ΔT_{Si} (e.g., $\Delta T_{Si} = 0.3\text{nm}$ for one monolayer variation). To be specific, if we require $\Delta V_T < k_B T / q$, then L_Q must be larger than ΔT_{Si}, as implied by Eq. (4.7). So we can obtain the minimum electrostatic scale length, L_{min}, for a given device structure by interpolating the

corresponding L_Q vs. L_E curve (see Fig. 4.7). From Fig. 4.7, we find that corresponding to $\Delta T_{Si} = 0.3\text{nm}$, $L_{min} = 5.8\text{nm}$ for the CW FET and $L_{min} = 6.2\text{nm}$ for the DG MOSFET on a (001) wafer. Using the same approach, we also compute the L_{min} for the RW structure. Fig. 4.8 compares the L_{min} for all the simulated structures, and the results show that for the (001) wafer the DG MOSFET has a little larger L_{min} than the CW FET and both of them provide smaller L_{min} than the RW (tri-gate) structure. For the (011) wafer, however, the performance of the planar DG structure degrades so that the advantage of the non-planar (e.g., CW and RW) FETs becomes more apparent. The reason for this is that for the (001) wafer the effective-mass in the confinement (z) direction ($m_z^* = 0.98m_e$) of a planar DG FET is much larger than that in the transverse (y) direction ($m_y^* = 0.19m_e$), which is one of the two confinement directions in a non-planar structure. So the planar DG FET displays much weaker quantum confinement. For the (011) wafer where the m_z^* is relatively small (i.e., $m_z^* = 0.318m_e$ [71]), the quantum confinement advantage of the DG structure is greatly diminished and consequently the non-planar FETs have significantly better overall performance (according to our general approach) due to their better electrostatic integrity. In addition, Fig. 4.8 also illustrates that the cylindrical wire FET (gate-all-around) performs substantially better than the rectangular wire structure (tri-gate), which agrees with the detailed numerical simulations in Sec. 4.1.

As indicated in part A of Subsection 4.2.3, when calculating L_E for the RW structure that has a floating SiO_2 substrate, we assume that the electric field at the interface between the Si body and the SiO_2 substrate is zero (so called the 'zero-field' assumption). By doing detailed benchmarking simulations with our 3D numerical simulator developed in Chapter 3, we find that for the RW structure simulated in this work, the 'zero-field' assumption underestimates the subthreshold swing by <2% and the DIBL by ~10% . So this assumption *slightly* overestimates the electrostatic integrity of the RW FETs with an acceptable error, which will not affect any qualitative conclusions of the device comparison.

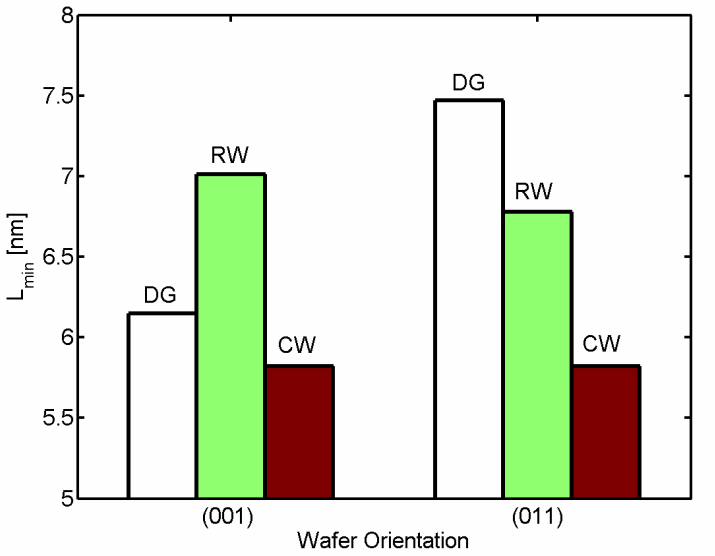


Fig. 4.8 The minimum electrostatic scale length, L_{min}, for the DG MOSFET, CW FET and RW FET. The geometry configurations of the simulated structures have been illustrated in Fig.4.1 (the oxide thickness is fixed to be to 1nm). Two wafer orientations, (001) and (011), are examined. In this calculation, the Si body thickness variation, ΔT_{Si}, is assumed to be equal to 0.3nm, which roughly accounts for one monolayer variation.

4.2.5 Conclusions

In this section, we proposed a general approach, the EQ approach, to assess the scaling potential of SNWTs vs. planar MOSFETs based on both gate control (electrostatics) and quantum confinement. We illustrated this general approach by evaluating three different device geometries. For (001) Si wafers, the results show that the non-planar nanowire structures (e.g., the cylindrical wire FET) provides better gate control while displaying stronger quantum confinement than planar devices (e.g., double-gate MOSFETs). On balance, the CW FET and the planar DG MOSFET perform nearly equally well for (001) Si wafers. For (011) Si wafers, however, where the effective-mass in the confinement (z) direction of the planar FET is relatively small, the non-planar wire structures are significantly advantageous.

4.3 Channel Material Optimization of Planar and Nanowire MOSFETs: A Conceptual Study

4.3.1 Introduction

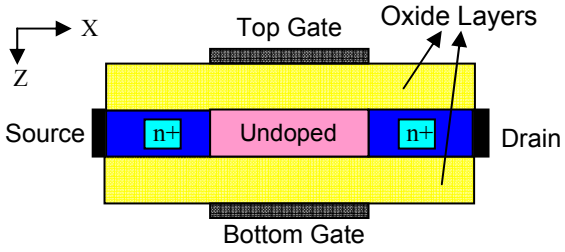
As conventional Si CMOS transistors are approaching their scaling limit, many researchers are exploring new materials and device structures to push MOS technology towards fundamental limits. MOSFETs with strained silicon, SiGe, or even III-V channels are possibilities [7] [8] [82], as are 1D channels made from nanowires [3] [4] or nanotubes [5] [6]. In this section, we theoretically examine the impact of the channel material property (i.e., the energy dispersion relation) and device structure (i.e., 2D planar vs. 1D nanowire) on the ultimate performance of ballistic MOSFETs. The results will show that when the transport effective-mass is small, it degrades device performance, and that planar and nanowire MOSFETs behave differently.

Different channel materials display different energy dispersion relations and different effective-mass at the band-edge. To achieve high device performance, one might expect that a light transport effective-mass would be best since it offers a high carrier injection velocity. [36] [67] On the other hand, a light effective-mass also leads to a lower quantum (or semiconductor) capacitance [55] [56], which degrades the ON-current (I_{ON}) of the device. When the channel length is sufficiently small, strong source-to-drain (S/D) tunneling occurs at a small transport effective-mass. [17] Tunneling degrades the subthreshold characteristics of the FET and consequently lowers the ON-current for the same OFF-current (I_{OFF}). For these reasons, an optimum transport effective-mass may exist for a given device structure.

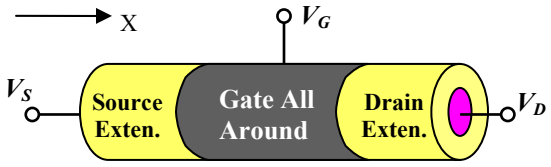
4.3.2 Device structure and simulation results

The simulated structures of the planar and nanowire FETs are sketched in Fig. 4.9. We first employ a ballistic, semiclassical, FET model, 'FETToy' [52] [53], introduced in Chapter 2. It is assumed that states at the top of the barrier are filled according to the

source and drain Fermi levels, and the electron density and current are computed by integrating over filled states (using Landauer's formula [37]). The model treats the quantum capacitance effect [55] [56] but not S/D tunneling [17]. Fig. 4.10a plots I_{ON} vs. transport effective-mass, m_{eff}, for the nanowire FET. (Different gate work functions are used for each m_{eff} to achieve the *same* I_{OFF}.) The valley degeneracy, *Nval*, is equal to 1 and the electrical effective oxide thickness (EOT) is 0.6nm. When m_{eff} is large enough, the device works at the Charge Control Limit (CCL) – the mobile charge at the top of the barrier is determined by the gate insulator capacitance and the gate overdrive (not m_{eff}), so I_{ON} is roughly proportional to the electron thermal velocity, which is proportional to $m_{eff}^{-1/2}$ [36] [67]. When m_{eff} is sufficiently small, the device reaches the Quantum Capacitance Limit (QCL) – the barrier height is solely determined by the applied bias (the mobile charge is sensitive to m_{eff}), and I_{ON} is independently of m_{eff}. For a 2D planar FET, as shown in Fig. 4.10b, I_{ON} is roughly proportional to $m_{eff}^{-1/2}$ at CCL, as in the 1D case, but $I_{ON} \sim m_{eff}^{1/2}$ at the QCL. [56] For this reason, a 2D FET displays an optimum effective-mass, m_{op}, which provides the highest I_{ON} for the same I_{OFF}. Fig. 4.10c shows that m_{op} is inversely proportional to EOT and *Nval*.

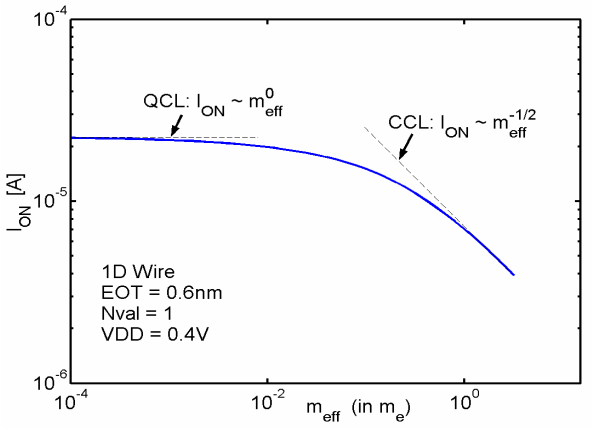


(a)

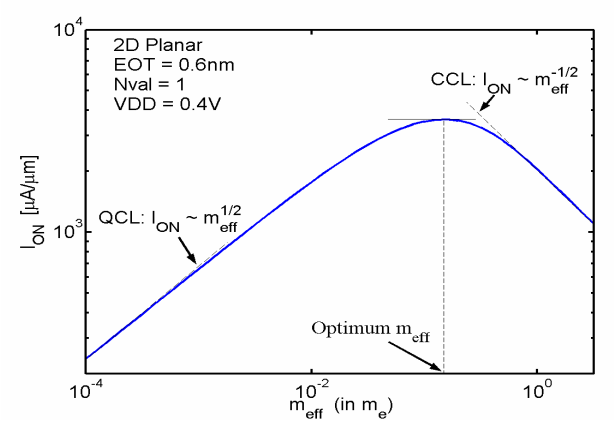


(b)

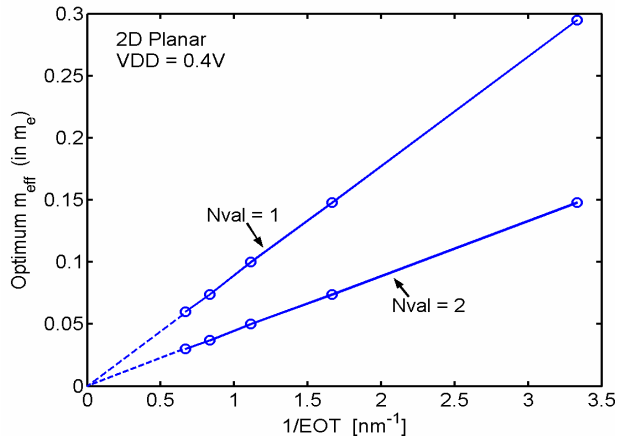
Fig. 4.9 Device structures of a planar double-gate MOSFET (a) and a gate-all-around cylindrical nanowire FET (b).



(a)



(b)



(c)

Fig. 4.10 (a) I_{ON} vs. m_{eff} (log-log) for the nanowire FET. (b) I_{ON} vs. m_{eff} (log-log) for the planar FET. (c) Optimum m_{eff} vs. 1/EOT. The supplied voltage is VDD=0.4V. All the simulations for this figure are done by using the FETToy model [52] [53].

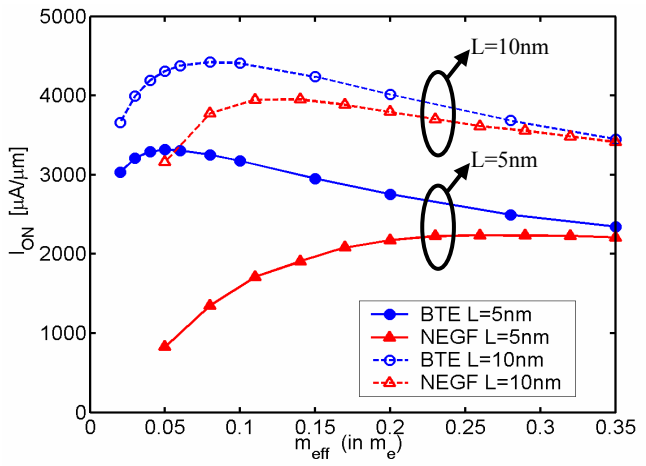
To treat S/D tunneling, we employ detailed numerical simulators for planar (NanoMOS-2.5 [40]) and nanowire (developed in Chapter 3) FETs. Ballistic quantum transport is treated with the nonequilibrium Green's function approach (NEGF) [37] [38]. For comparison, a ballistic Boltzmann Transport Equation (BTE) model [36] is also used, in which S/D tunneling is not included. Two channel lengths, L=5nm and L=10nm, are used for both the planar and nanowire FETs. Fig. 4.11a clearly shows that both NEGF and BTE predict an m_{op} for the planar FET, and m_{op}(NEGF)>m_{op}(BTE) due to the S/D tunneling. When L=5nm, S/D tunneling is more serious and the difference between m_{op}(NEGF) and m_{op}(BTE) is larger than for L=10nm. For the nanowire FET (Fig. 4.11b), the semiclassical BTE model does not give an m_{op} (as was also observed in Fig. 4.10a). When S/D tunneling is included, however, there does exist an m_{op} for the nanowire FET, and its value increases with a decreasing channel length.

4.3.3 Conclusions

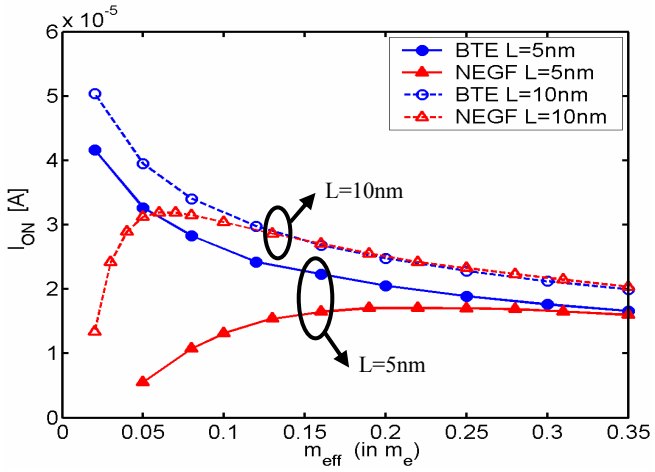
In conclusion, we performed a theoretical exploration of how the channel material (i.e., the energy dispersion relation) affects the ultimate performance of planar and nanowire MOSFETs. We found that:

1) *Without* the consideration of S/D tunneling, there exists an optimum effective-mass, which offers the highest I_{ON} for the same I_{OFF}, for a planar FET. For a nanowire FET, I_{ON} increases with a decreasing m_{eff} and saturates when m_{eff} is sufficiently small.
2) When S/D tunneling is considered, an optimum effective-mass can be defined for both planar and nanowire FETs, and its value increases when the channel length scales down.

These results suggest that III-V MOSFETs [82], which display a light transport effective-mass, might prove advantageous, but only if the channel length is not reduced well below 10nm.



(a)



(b)

Fig. 4.11 I_{ON} vs. m_{eff} for a planar FET (a) and a nanowire FET (b) calculated by using the detailed numerical simulators (i.e., NanoMOS-2.5 [40] for planar DG MOSFETs and the ballistic SNWT simulator developed in Chapter 3).

5. A THEORETICAL INVESTIGATION OF SURFACE ROUGHNESS SCATTERING IN SILICON NANOWIRE TRANSISTORS

This chapter describes a theoretical investigation of the effects of surface roughness scattering (SRS) [83] [84] [85] on the device characteristics of silicon nanowire transistors (SNWTs). To do this, we adopt the full three-dimensional (3D), quantum mechanical simulator, developed in Chapter 3. The microscopic structure of the Si/SiO_2 interface roughness [61] [62] is directly treated by using a 3D finite element technique. The results show that 1) SRS reduces the electron density of states in the channel, which increases the SNWT threshold voltage, and 2) the SRS in SNWTs becomes less effective when fewer propagating modes are occupied, which implies that SRS is less important in small-diameter SNWTs with few modes conducting than in planar MOSFETs with many transverse modes occupied.

5.1 Introduction

To explore the realistic performance limits of SNWTs, understanding carrier transport in Si nanowires becomes increasingly important. Careful studies are needed to experimentally characterize transport in SNWTs, but it is also clear that a theoretical understanding is similarly important. In this chapter, we present a theoretical exploration of the Si/SiO_2 interface roughness scattering, or surface roughness scattering (SRS) [83] [84] [85], in SNWTs.

It is well-known that scattering due to Si/SiO_2 interface roughness is important in planar silicon MOSFETs, and it is expected to be even more important in ultra-thin body silicon-on-insulator (UTBSOI) MOSFETs [83]. For bulk MOSFETs, electrons are

confined at the Si/SiO_2 interface by an electrostatic potential well. Under high gate bias, the potential well is thin, electrons are confined very near the interface, SRS increases, and the effective mobility decreases. For UTBSOI MOSFETs, the confining potential is determined by the film thickness, and SRS can be enhanced by the roughness at the two interfaces. [83] In a SNWT, the channel is surrounded by the Si/SiO_2 interfaces, so one might expect SRS to dominate transport. We will show, however, that SRS may be less important in SNWTs than in planar devices because of the one-dimensional (1D) nature of the SNWT channel.

5.2 Methodology

In Chapter 3, we developed a self-consistent, full 3D, quantum mechanical simulator of SNWTs based on the effective-mass approximation. [43] [44] In this chapter, to investigate the effects of SRS, we apply this simulator for small-diameter (~3nm) SNWTs with physically rough Si/SiO_2 interfaces. The simulated structure is a gate-all-around SNWT with a rectangular cross-section and a $[100]$ oriented channel (see Fig. 5.1). Following previous work on SRS [83] [84] [85], we assume an abrupt, randomly varying interface between the Si and SiO_2, parametrized by a root mean square (*rms*) amplitude and an autocovariance function [61] [62]. The statistical nature of the roughness will depend on the nanowire fabrication methods and may differ considerably from that arising during the high temperature oxidation of a planar Si surface. Nevertheless, since our objective is to discuss general insights into the physics of SRS in SNWTs, we will employ the roughness parameters for a planar (100) Si/SiO_2 interface obtained from [61]. Our use of a continuum level description may be questioned, but we believe that it is a useful first step that gives insight into how the magnitude and spatial coherence of potential fluctuations influence carrier transport. In contrast to previous work [83] [84] [85], which made use of perturbation theory to compute the surface roughness scattering rate, we treat the physically rough structure directly.

The microscopic structure of the Si/SiO_2 interface roughness is implemented into the 3D simulator according to the following procedure. We first discretize the simulation

domain with a 3D finite element mesh [44] [45]; each element is a triangular prism with a 2Å height and edge length, comparable to the size of roughness at the (100) Si/SiO_2 interface. [61] Next, we generate a two-dimensional (2D) random distribution across the *whole* Si/SiO_2 interface (unfolding the four interfacial planes into a sheet) according to an exponential autocovariance function [61],

$$C(x) = \Delta_m^2 e^{-\sqrt{2}x/L_m}, \tag{5.1}$$

where L_m is the correlation length, Δ_m is the *rms* fluctuation of the roughness and x is the distance between two sampling points at the interface. Based on the 2D random distribution, the types of the elements at the Si/SiO_2 interfaces may be changed from Si to SiO_2, or reversely, to mimic the rough interfaces (see Fig. 5.1b).

After the roughness is implemented, electron transport through the rough SNWT is simulated by using the non-equilibrium Green's function approach [37] [38]. With a coupled mode space (CMS) representation [41] [44] [45] [47], the wavefunction deformation due to the Si/SiO_2 interface roughness is treated. (The simulation methodology has been discussed in detail in Chapter 3.) To emphasize the role of SRS on electron transport, we do not include any other scattering mechanisms, so coherent transport is assumed inside the device. (Oscillations in the current due to quantum interference might be expected, but the averaging over a thermal distribution of wavelengths that occurs is sufficient to suppress them.) The length of the channel ($L = 10\text{nm}$) is long enough to ensure that sufficient averaging takes place so that sample specific effects are not observed. The simulated results for the rough SNWT are then compared with those for a device with the same geometrical parameters (e.g., nominal oxide thickness and Si body thickness) but smooth Si/SiO_2 interfaces. By doing this, the effects of SRS on SNWT device characteristics can be clearly identified.

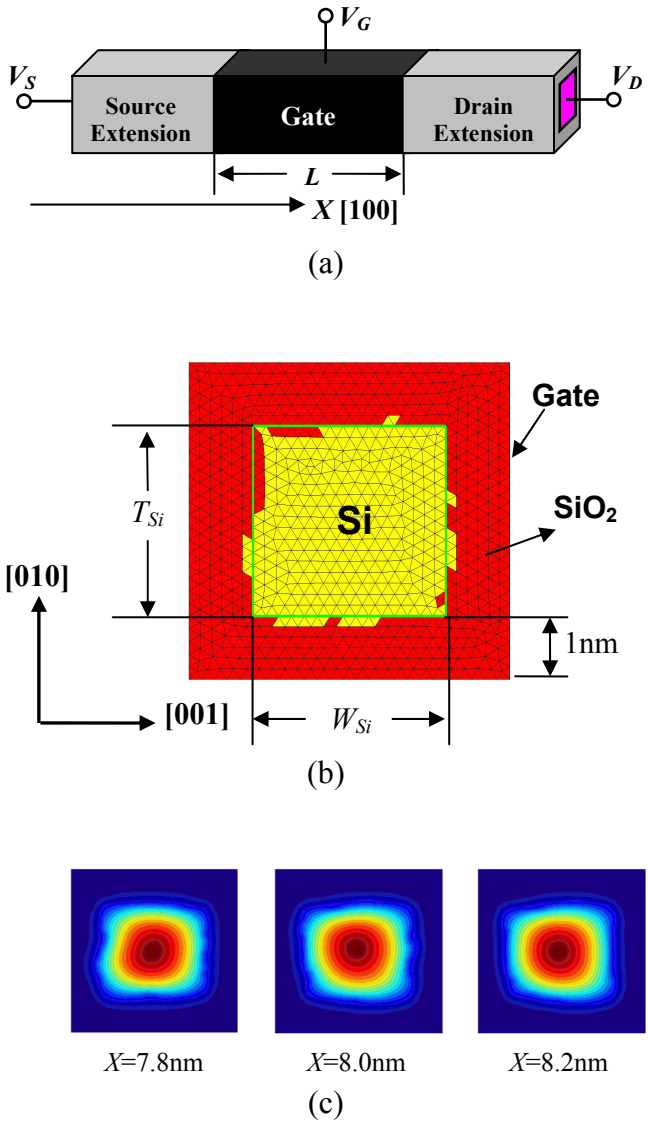


Fig. 5.1 (a) The schematic diagram of the simulated gate-all-round SNWT. The source/drain doping concentration is $2 \cdot 10^{20}$ cm^{-3} and the channel is undoped. There is no source/drain overlap with the channel and the gate length is L=10nm. V_S, V_D, V_G are the applied voltage biases on the source, the drain and the gate, respectively. (b) The cross-section of the SNWT with a specific interface roughness pattern for the slice at X=9.0nm. For the device with smooth Si/SiO_2 interfaces, the Si body thickness is T_{Si}=3nm, the wire width is W_{Si}=3nm, and the oxide thickness is 1nm. (c) The confined wavefunctions for the slices at X=7.8nm, X=8.0nm, and X=8.2nm, respectively. The shape of the wavefunctions changes from slice to slice due to the Si/SiO_2 interface roughness.

5.3 Results

Figure 5.2 plots the electron subband profile (left column) at the ON-state (V_{GS}=V_{DS}= 0.4V) in the simulated SNWT with rough and smooth Si/SiO_2 interfaces. The corresponding transmission coefficients (right column) for both the rough and smooth SNWTs are also shown. Note that the modes are coupled in the simulation; we show them separately for illustrative purposes only. It is clearly seen in the Energy vs. X plot that the presence of the roughness introduces significant fluctuations in the electron subbands, which lead to fluctuating elements in the diagonal terms of the device Hamiltonian (for details, see Eq. (3.7) in Chapter 3) and act as a scattering potential. At the same time, the shape of the confined wavefunction also alters from slice to slice in the rough SNWT (see Fig. 5.1c for an example), which produces deformation and coupling elements in both diagonal and off-diagonal terms of the device Hamiltonian (for details, see Eqs. (3.7), (3.8b) and (3.8c) in Chapter 3), and consequently lowers the transmission. (This effect has been named 'wavefunction deformation scattering' [86] [87] [88].) To examine the significance of wavefunction deformation scattering, we plot an Energy vs. Transmission curve (dot-dashed) for the rough SNWT calculated by the uncoupled mode space (UMS) approach (see Chapter 3 for details), in which only the variations in the electron subbands are included while the deformation and coupling terms are completely discarded. The fact that the UMS approach significantly overestimates the transmission for the rough device infers that wavefunction deformation scattering dominates the transport. This is an important finding because common perturbation theory treatments [83] [84] [85] of SRS scattering typically treat the subband energy fluctuations but not the wavefunction deformation scattering.

From the Energy vs. Transmission plot, we find that the difference between the transmission curve for the rough SNWT and that for the smooth device becomes more and more noticeable as energy increases. This occurs because as energy increases, more subbands (modes) become conductive and the coupling between different modes efficiently reduces the transmission in the rough SNWT. In other words, SRS becomes

more significant as more modes conduct. As we will show later, this effect has an important impact on the role of SRS on SNWT device characteristics.

Figure 5.3a plots the I_{DS} vs. V_{GS} curves in a semi-logarithmic scale for both the rough and smooth SNWTs. The results show that there is a distinct threshold voltage (V_T) increase caused by the SRS. In the low gate bias region, the lateral displacement of the smooth and rough characteristics implies a V_T increment of ~30mV for the roughness parameters we used (L_m=0.7nm and *rms*=0.14nm) and varies little from sample to sample. The increase in V_T due to SRS was unexpected and the reason for it is as follows. Due to SRS, injections at low energies are blocked in the rough SNWT, which reduces the density-of-states (DOS) near the band-edge (see Fig. 5.3b). The lowered DOS near the band-edge reduces the charge density in the subthreshold regime, and consequently increases V_T in the rough SNWT. This effect would be modest in a conventional MOSFET with an energy-independent DOS above the band-edge, but it becomes pronounced in a 1D wire with a singularity in the DOS at the band-edge.

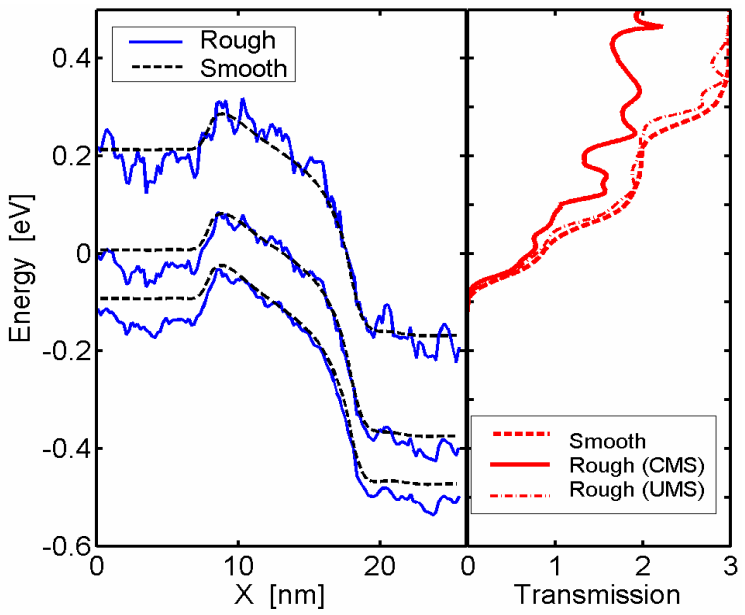
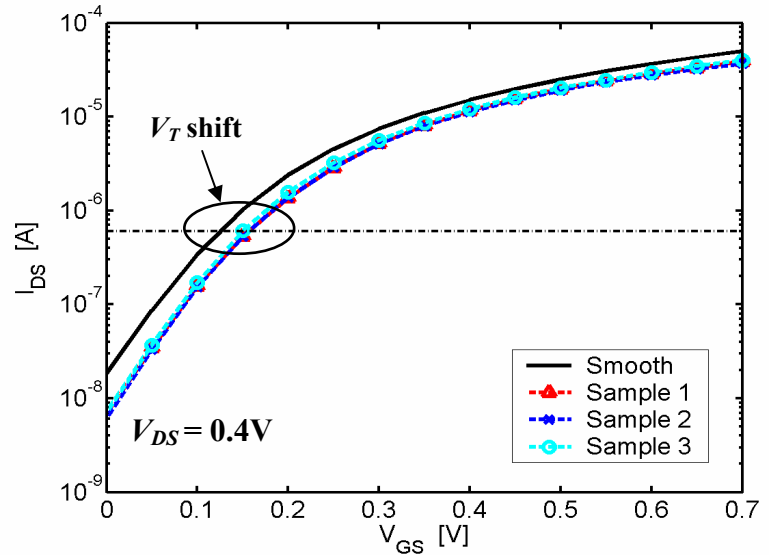
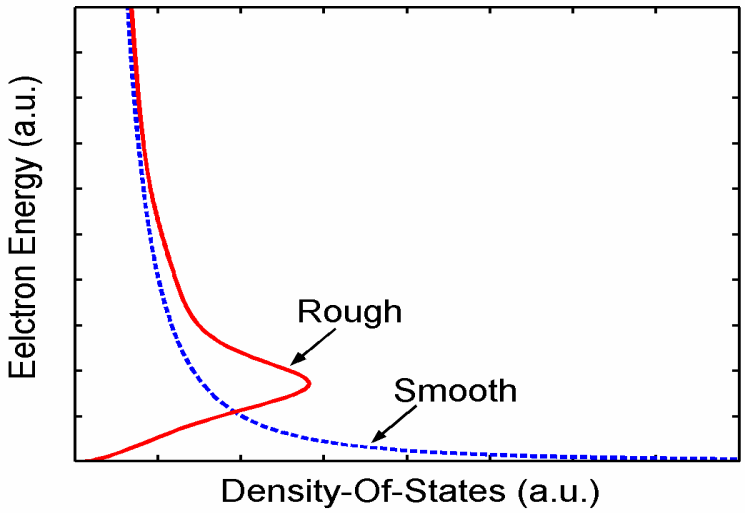


Fig. 5.2 The electron subband profile (for the (010) valleys) and the corresponding transmission coefficients for the simulated SNWT (T_{Si}=W_{Si}=3nm) with smooth and rough Si/SiO_2 interfaces. The roughness parameters used are L_m=0.7nm and *rms*=0.14nm [61]. The device is at the ON-state (V_{GS}=V_{DS}=0.4V), so the source and drain Fermi levels are equal to 0eV and -0.4eV, respectively.



(a)



(b)

Fig. 5.3 (a) I_{DS} vs. V_{GS} curves for the simulated SNWT (T_{Si}=W_{Si}=3nm) with smooth (solid) and rough (dashed with symbols) Si/SiO_2 interfaces. (V_{DS}=0.4V). Three samples (triangles, crosses and circles) of the rough SNWT are generated based on the same roughness parameters (L_m=0.7nm and *rms*=0.14nm) but different random number seeds. The SNWT threshold voltage (V_T) is defined as $I_{DS}(V_{DS}=V_T, V_{DS}=0.4V)=2\cdot10^{-7}\cdot W_{Si}$ (A), where W_{Si} is in nm. (b) The reduction of electron DOS at low injection energies caused by SRS.

Finally, we explore the effects of SRS on the SNWT drain current above threshold. To do this, we compute a current ratio $\beta = I_{DS}^{Rough} / I_{DS}^{Smooth}$ at the *same* gate overdrive, V_{GS}-V_T, for both rough and smooth SNWTs. By comparing currents (rough vs. smooth) at the same gate overdrive, the effect of the V_T increasing induced by SRS is removed. This allows us to examine whether the roughness can cause a significant reduction of SNWT ON-current by back-scattering. Fig. 5.4 shows the β vs. gate overdrive curves for the SNWTs with different wire widths and roughness parameters. Several interesting phenomena are observed. First, all the simulated structures display a decreasing β with an increasing gate overdrive. This occurs because more modes become conductive under higher gate bias, which, as described earlier, enhances SRS in the SNWTs.

Second, based on the roughness parameters, L_m=0.7nm and *rms*=0.14nm, which are typical of an oxidized Si/SiO_2 interface [61], the SNWT with W_{Si}=3nm (solid) achieves a surprisingly high $\beta \approx 0.9$ at a typical ON-state condition (gate overdrive = 0.3V for a 0.4V supplied voltage). The same amount of surface roughness scattering severely degrades the mobility of a planar MOSFET under a high gate bias. [86] To explore the effects of the correlation length L_m, two additional values (1.4nm for circles and 3.0nm for triangles) were examined. The results show that β is insensitive to L_m, as expected from the averaging over a thermal distribution of electron wavelengths that occurs at room temperature and high drain bias (V_{DS}=0.4V). In contrast, doubling the *rms* (diamonds) clearly degrades β at the same gate overdrive, indicating the importance of maintaining relatively smooth Si/SiO_2 interfaces for the high performance applications of SNWTs.

Third, increasing the wire width reduces the strength of quantum confinement and thus increases the number of conducting modes in the SNWT. Our results (solid vs. dashed) clearly show that with a larger number of conducting modes in the wider (W_{Si}=9nm) SNWT, SRS is much stronger than in the narrower (W_{Si}=3nm) device. This observation also suggests that SRS is more serious in a planar MOSFET, which can be viewed as a SNWT with a very large wire width.

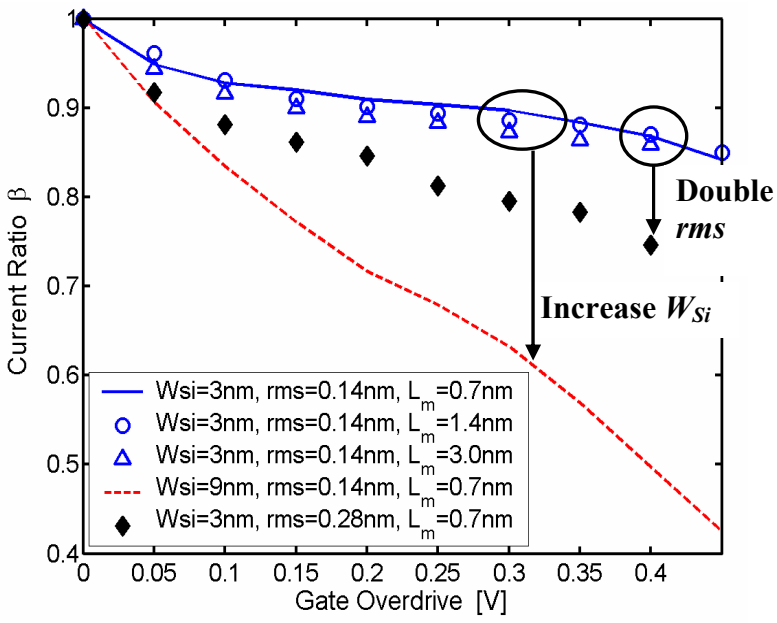


Fig. 5.4 Current ratio β vs. gate overdrive curves for the simulated SNWTs with different wire widths (W_{Si}) and roughness parameters (L_m and *rms*). At all the cases, the Si body thickness is fixed to be T_{Si}=3nm and the drain bias is V_{DS}=0.4V.

5.4 Summary

In summary, we theoretically investigated the surface roughness scattering in silicon nanowire transistors by using a full 3D, self-consistent, quantum mechanical simulator, developed in Chapter 3. The microscopic structure of the Si/SiO_2 interface roughness [61] [62] was implemented into the simulator using the 3D finite element method [44] [45]. We found that

1) SRS significantly deforms electron wavefunction and subbands, and the deformation and coupling terms, instead of the variations in electronic subbands, dominate the transport,
2) SRS reduces the electron density of states in the channel, which increases the SNWT threshold voltage,

3) SRS is insensitive to the correlation length of the Si/SiO_2 interface roughness but very sensitive to the *rms*, so a smoother Si/SiO_2 interface is highly preferred for the high performance applications of SNWTs,
4) SRS in SNWTs becomes less serious when fewer propagating modes conduct, implying that SRS will be less important in small-diameter SNWTs than in planar MOSFETs with many transverse modes occupied.

This work provides important insights into the nature of SRS in SNWTs and suggests that SRS may not be as important in nanowires as it is in conventional, planar MOSFETs.

6. ATOMISTIC SIMULATIONS OF SILICON AND GERMANIUM NANOWIRE TRANSISTORS

6.1 Introduction

In nanowires, due to the two-dimensional (2D) quantum confinement, the bulk crystal symmetry is not preserved any more. For this reason, the atomistic bandstructure effects may play an important role on the device characteristics of semiconductor nanowire transistors, especially when the diameters are small (e.g., <3nm). In this chapter, we perform an atomistic simulation of silicon and germanium nanowire transistors to explore their ballistic performance limits with the consideration of bandstructure effects.

Instead of doing a full three-dimensional (3D), atomistic simulation within the non-equilibrium Green's function (NEGF) formalism [37] [38], which is discouraged by its huge computational burden, we accomplish our atomistic simulations according to the following 'two-step' procedure.

1) Step 1 is the calculation of the energy dispersion (*E-k*) relations of silicon and germanium nanowires by using a nearest-neighbor $sp^3d^5s^*$ tight binding (TB) approach [49] [50] [51]. Within this tight binding approach, 20 orbitals, consisting of an $sp^3d^5s^*$ basis with spin-orbital coupling, are used to represent each atom in the nanowire Hamiltonian. The orbital-coupling parameters we use are from [49], which have been optimized by Boykin et al. to accurately reproduce the band gap and effective-masses of bulk Si and Ge (within a <5% deviation from the target values [49]). (It should be mentioned that bulk bond lengths are assumed in our calculations. In real nanowires, the crystal structures will relax to obtain a minimum energy [89]. We expect that the

general results of this study will also apply to relaxed structures while some quantitative differences may appear.) At the Si/Ge surfaces, a hard wall boundary condition for the wavefunction is applied and the dangling bonds at these surfaces are passivated using a hydrogen-like termination model of the sp^3 hybridized interface atoms. [90] As demonstrated in [90], this technique successfully removes all the surface states from the semiconductor band gap. Based on this $sp^3d^5s^*$ tight binding approach, we have developed a simulator that can treat Si and Ge nanowires with *arbitrary* wire orientations and cross-sectional shapes.

2) Step 2 is to evaluate the *I-V* characteristics of the Si/Ge nanowire FETs based on the *E-k* relations obtained in Step 1. Here we exploit a semi-numerical ballistic FET model, 'FETToy' [52] [53], introduced in Chapter 2. This model captures 3D electrostatics, quantum capacitance [55] [56] and bias-charge self-consistency in ballistic FETs. (Since the model assumes a semiclassical transport, source-to-drain tunneling [17] and band-to-band tunneling [2] [36] are not considered.) In the past, this model was used to evaluate the *I-V* characteristics of Si MOSFETs [52] and high electron mobility transistors [54] with parabolic energy bands and Ge MOSFETs with numerical *E-k* relations [91]. More details about this model can be found in the cited references [52] [53] [54] [91] and Chapter 2 of this thesis. With this method, ballistic *I-V* characteristics of both n-type and p-type Si/Ge nanowire FETs can be evaluated.

This chapter is divided into the following sections. Sec. 6.2 shows the calculated energy dispersion relations of Si and Ge nanowires with various wire orientations, Sec. 6.3 presents a performance evaluation and optimization of Si and Ge nanowire FETs, Sec. 6.4 explores the validity of the widely used parabolic effective-mass approximation [43] [44] [92] for the *I-V* calculation of Si nanowire transistors, and Sec. 6.5 summarizes the chapter.

6.2 Calculated Energy Dispersion Relations

Figure 6.1 shows an example of one of the simulated nanowire structures in this work. The transport orientation of the wire is along the $[100]$ direction, the shape of the cross-section is circular (or more strictly, octangular) and the wire diameter is 3nm. A unit cell of the nanowire crystal consists of four atomic layers along the x (transport) direction and has a length of a_0=5.43Å (for Si) and a_0=5.65Å (for Ge). [49] It should be noted that although Fig. 6.1 is only for a specific nanowire structure with a particular wire orientation and cross-sectional shape, nanowires with various wire orientations (e.g., $[100]$, $[110]$, $[111]$ and $[112]$) and cross-sectional shapes (e.g., circular and rectangular) are explored in this work.

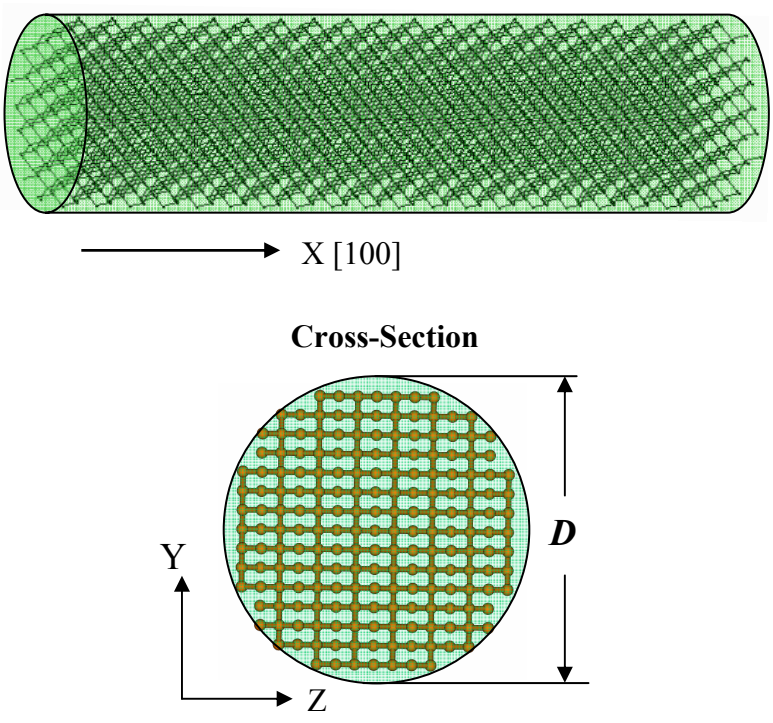
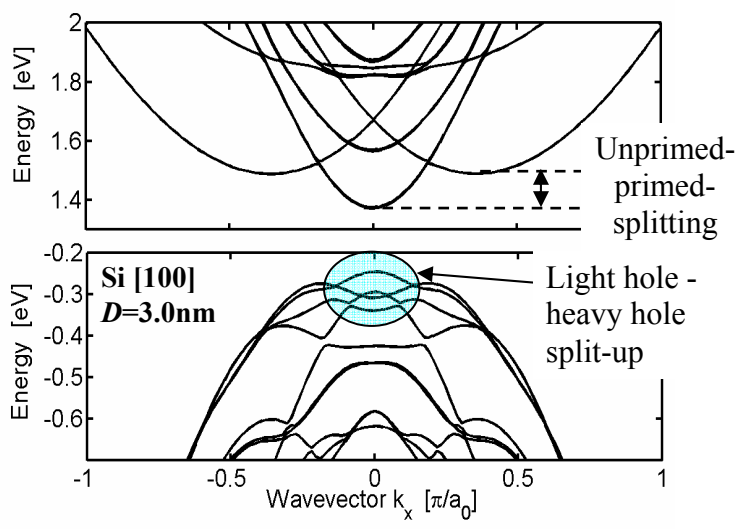


Fig. 6.1 The atomistic structure of a $[100]$ orientation nanowire with a circular cross-section and a 3nm wire diameter (D=3nm).

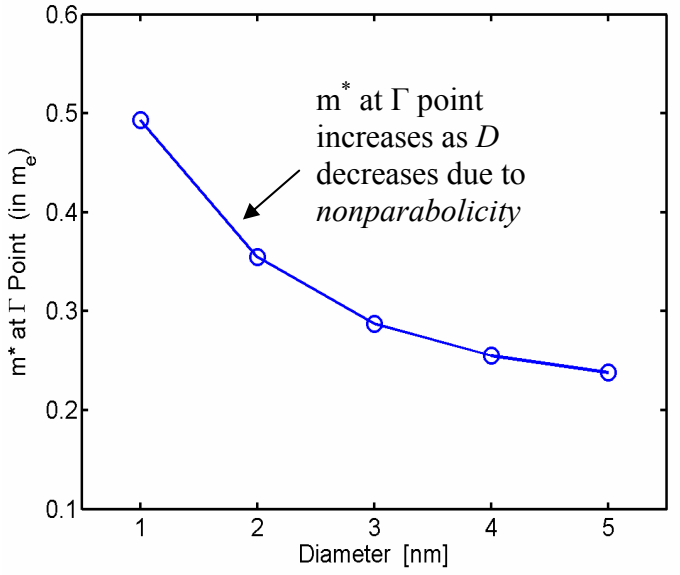
Figure 6.2a illustrates the *E-k* relation for the Si nanowire structures shown in Fig. 6.1. For the wire conduction band, it is clear that the six equivalent Δ valleys in the bulk Si conduction band split up into two groups due to quantum confinement [93] [94], which we call 'unprimed-primed-splitting' here. Four unprimed valleys, $[010]$, $[0\bar{1}0]$, $[001]$, and $[00\bar{1}]$, are projected to the Γ point (k_x=0) in the one-dimensional (1D) wire Brillouin zone ($-\pi/a_0 \le k_x \le \pi/a_0$) to form the conduction band-edge. Two primed valleys (i.e., $[100]$ and $[\bar{1}00]$), located at $k_x = \pm 0.815 \cdot 2\pi/a_0 = \pm 1.63\pi/a_0$ in the bulk Brillouin zone, are zone-folded to $k_x = \pm 0.37\pi/a_0$ in the wire Brillouin zone to form the off-Γ states. (A similar observation has been reported in [93] and [94] for square Si nanowires with a $[100]$ transport direction and four confinement directions along the equivalent <110> axes.) For the wire valence band, the degeneracy between the light hole and heavy hole in bulk Si is lifted up by quantum confinement. Fig. 6.2b plots the effective-mass m^* at the Γ point in the wire conduction band vs. wire diameter *D* for a $[100]$ oriented Si nanowire with a circular cross-section. It is clearly shown that m^* increases with a decreasing *D*, due to the nonparabolicity of the Δ valleys in the bulk Si conduction band. [36] [95] (The effects of this nonparabolicity on the device performance of Si nanowire FETs will be discussed in detail in Sec. 6.4.)

In Fig. 6.3a, we plot the *E-k* relation for a $[111]$ oriented Ge nanowire with a circular cross-section and a 4nm diameter. The range of the 1D Brillouin zone for this wire is $-\pi/\sqrt{3}a_0 \le k_x \le \pi/\sqrt{3}a_0$, since the length of the unit cell is $\sqrt{3}a_0$ for this structure. Three L valleys in bulk Ge, $[\bar{1}1\,1]$, $[1\bar{1}1]$ and $[1\,1\bar{1}]$, are projected to the point $k_x = \pi/\sqrt{3}a_0$, while another three L valleys in bulk Ge, $[\bar{1}\bar{1}1]$, $[1\bar{1}\bar{1}]$ and $[\bar{1}1\bar{1}]$ are projected to $k_x = -\pi/\sqrt{3}a_0$. These states form the lowest valleys in the conduction band of this Ge nanowire. Fig. 6.3b shows the *E-k* relation for a $[112]$ oriented Si nanowire with a circular cross-section and a 2nm diameter. Fig. 6.3c compares the calculated band

gaps (solid with circles) for [112] oriented Si nanowires with a circular cross-section vs. the experimental results (diamonds) obtained from [20]. The results clearly show that our tight binding calculation provides good agreement with the measured data [20].



(a)



(b)

Fig. 6.2 (a) The *E-k* relations for a [100] oriented Si nanowire with a circular cross-section and a 3nm wire diameter (the atomistic structure of this nanowire is shown in Fig. 6.1). (b) The effective-mass m^* at the Γ point in the wire conduction band vs. wire diameter D for a [100] oriented Si nanowire with a circular cross-section.

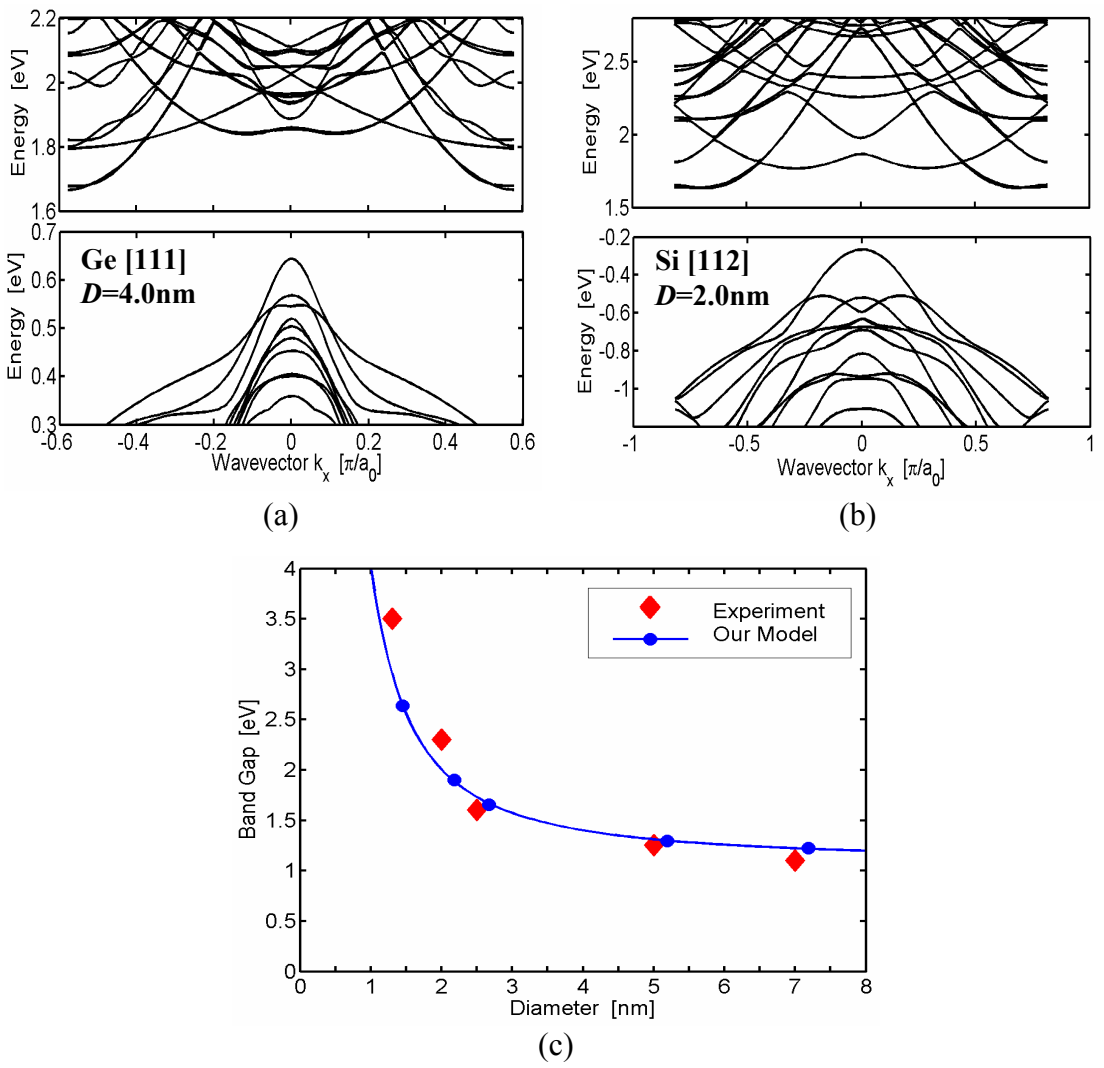
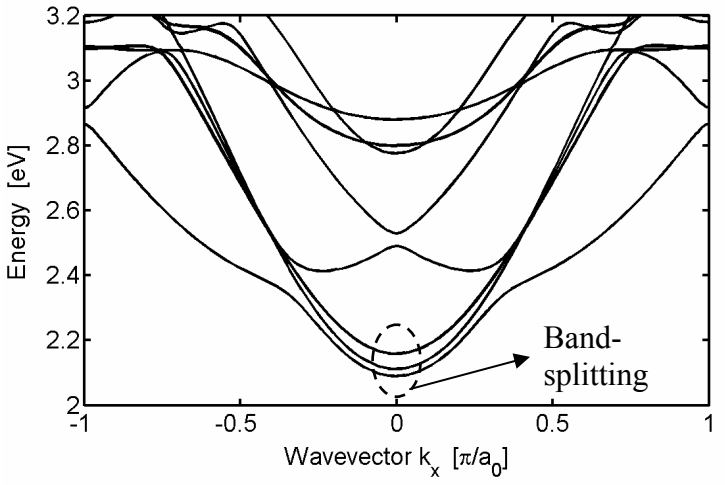


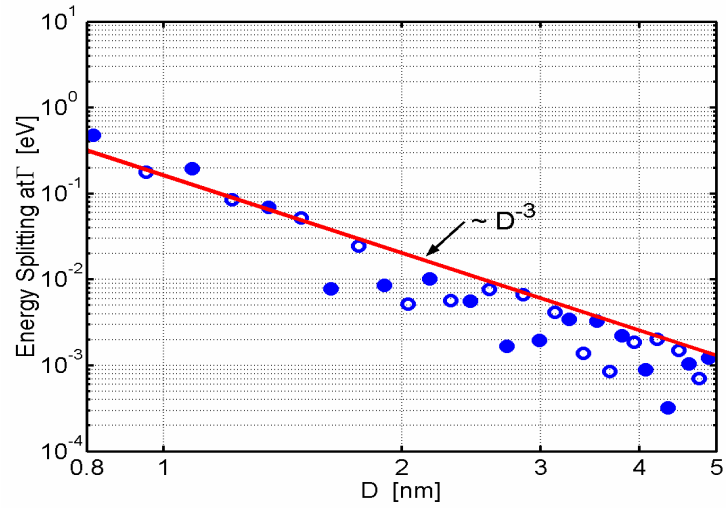
Fig. 6.3 (a) The *E-k* relations for a [111] oriented Ge nanowire with a circular cross-section and a 4nm wire diameter. (b) The *E-k* relations for a [112] oriented Si nanowire with a circular cross-section and a 2nm wire diameter. (c) Comparison between the calculated band gaps (solid with circles) for [112] oriented Si nanowires with a circular cross-section and experimental data (diamonds) [20].

In the final part of this section, let us discuss an interesting bandstructure effect, 'band-splitting' [50] [51] [96], in silicon nanowires with small diameters. Fig. 6.4a shows the calculated conduction band for a [100] oriented Si nanowire with a square cross-section (four confinement directions along the equivalent <100> axes) and a 1.36nm wire width. The results clearly show that the degeneracy of the 4-fold Γ valleys

in this [100] oriented square wire is lifted by the interaction between the four equivalent valleys. Fig. 6.4b plots the wire width (D) dependence of the splitting energy, defined as the difference between the highest and the lowest energy (at the Γ point) of the four split conduction bands. The splitting energy is seen to fluctuate as a function of the number of atomic layers and the envelope decreases with the wire width according to D^{-3}, analogous with the band-splitting observed in Si quantum wells [50] [51] [96].



(a)



(b)

Fig. 6.4 (a) The calculated conduction band for a [100] oriented Si nanowire with a square cross-section and a 1.36nm wire width. (b) The splitting energy (at the Γ point) vs. wire width (D) for the simulated Si nanowires. The closed circles are for the wires with an odd number of atomic layers while the open circles are for the ones with an even number of atomic layers.

6.3 Performance Evaluation and Optimization of Silicon and Germanium Nanowire Transistors

In this section, we adopt the semi-numerical ballistic FET model, 'FETToy' [52] [53], introduced in Chapter 2, to calculate the *I-V* characteristics of Si and Ge nanowire FETs based on the numerical *E-k* relations calculated by our tight binding model. The device performance (e.g., ON-current and intrinsic device delay) of ballistic n-/p-type Si and Ge nanowire FETs will be compared for different wire orientations. By doing this, an optimum wire orientation in terms of device performance can be defined for both Si and Ge nanowire FETs. Finally, the dependence of the device performance on wire diameter will also be explored.

As mentioned earlier in Chapter 2, there are two parameters, α_G and α_D, used in the FETToy model to describe the electrostatic couplings between the top of the barrier and the gate and the drain, respectively. To treat the short channel effects (SCEs) [2] in FETToy simulations, the values of α_G and α_D are tuned for a given structure in such a way that the FETToy model provides the same sub-threshold swing and drain induced barrier lowering (DIBL) as our detailed numerical simulator developed in Chapter 3. All the values of α_G and α_D used in this section are listed in Table 6.1. (When extracting α_G and α_D, a simple parabolic energy band is assumed for both the FETToy model and the detailed numerical simulator. And it is found that the values of α_G and α_D are not sensitive to the bandstructures of the channel materials.)

Figure 6.5 plots the I_{DS} vs. V_{GS} curves in a semi-logarithmic scale for the simulated n-type and p-type Si nanowire FETs with a circular cross-section and a 3nm wire diameter. The channel orientation is $[100]$ and the gate length is assumed to be L=8nm. In our simulations, we assume a different gate work function for each structure to achieve a specified OFF-current, $10\mu\text{A}/\mu\text{m}\bullet(2D)$, where D is the wire diameter. Then we can compare the device performance of Si and Ge nanowire FETs with various wire orientations. Fig. 6.6 shows the ON-currents (a) and intrinsic device delays (b) for the simulated n-/p-type Si and Ge nanowire FETs with four different channel orientations – $[100]$, $[110]$ and $[111]$, which are the three major crystal orientations, as well as $[112]$,

which is one of the favored growth orientations for bottom-up nanowires [19] [20] [97]. All the devices have the same wire diameter, 3nm, which is selected with the consideration of the tradeoff between the SCEs control and the threshold voltage fluctuation (see Sec. 4.2 for the details). The results clearly show that $[110]$ is the *optimum* channel orientation, which offers the highest ON-current and the fastest intrinsic device delay for the same OFF-current, for both n-type and p-type Si/Ge nanowire FETs. With this optimum channel orientation, the n-type (p-type) Ge nanowire FET displays a ~40% (~30%) higher ON-current than the Si counterpart.

In Fig. 6.7, we explore the dependence of nanowire FET performance on the wire diameter. As we know, when the wire diameter varies, 1) it affects the strength of the quantum confinement in the wire and consequently alters its bandstructure (e.g., band-edges and effective-masses), and 2) it also changes the electrostatic scale length [58] of the nanowire FET and, therefore, affects the SCEs of the device. In our work, to separate these two effects, we first simulate the devices *without* the consideration of SCEs (assuming a perfect gate control, α_G=1 and α_D=0). Fig. 6.7a shows the intrinsic device delays of the simulated Si and Ge nanowire FETs at three different wire diameters, 2nm, 3nm and 4nm. Four channel orientations, $[100]$, $[110]$, $[111]$ and $[112]$, are explored. It is clear that for pFETs, the device performance is improved with a decreasing wire diameter at all the cases. The reason is as follows. As shown in Fig. 6.2a, quantum confinement in a nanowire lifts up the degeneracy between the light hole and heavy hole in the valence band, which reduces the average hole effective-mass and consequently raises the injection velocity of the holes. When the wire diameter scales down, quantum confinement becomes stronger, which leads to a higher hole injection velocity. This is why the p-type nanowire FETs with smaller wire diameters display higher device performance. For nFETs, however, the dependence of device performance on wire diameter is sensitive to the material type (e.g., Si or Ge) and the wire orientation. To explain this, we have to understand how quantum confinement affects the conduction band properties of Si and Ge nanowires.

1) *Nonparabolicity*. Due to the nonparabolicity of the bulk Si/Ge conduction bands [36] [95], the effective-masses of the lowest conduction-band valleys in

a nanowire increase as the strength of quantum confinement increases (or the wire diameter decreases). (An example for a $[100]$ oriented Si nanowire is illustrated in Fig. 6.2b.) Since the electron injection velocity is inversely proportional to the square root of the effective-mass, *nonparabolicity* leads to a decreasing device performance as the wire diameter scales down.

2) *Unprimed-primed-splitting*. As shown in Fig. 6.2a, quantum confinement in a $[100]$ oriented Si nanowire lifts up the degeneracy between the unprimed and primed valleys in the bulk Si conduction band and reduces the *average* effective-mass in the transport direction of the wire. In fact, this unprimed-primed-splitting effect also occurs in Si/Ge nanowires with other orientations. In general, *unprimed-primed-splitting* increases the average injection velocity of electrons and consequently causes an increasing device speed with a decreasing wire diameter.
3) *Band-splitting*. For most of the wire orientations of a Si/Ge nanowire, band-splitting is modest in the diameter range of 2nm – 4nm. However, for some particular wire orientations (e.g., Si $[110]$), band-splitting becomes significant when the wire diameter is ~3nm. It increases the curvature of the lowest conduction band in the wire and, therefore, reduces the effective-mass at the conduction band-edge of the wire. For this reason, *band-splitting* in nanowires with some particular wire orientations may lead to an increasing device performance with a decreasing wire diameter.

In a real Si/Ge nanowire, all these three effects mentioned above co-exist, and the competition between them finally determines the dependence of the device performance on wire diameter. This explains why for n-type Si/Ge nanowire FETs, there is no general trend to describe how the wire diameter affects device performance. Fig. 6.7b shows the same results as in Fig. 6.7a except that the SCEs are considered. (The values of α_G and α_D used are listed in Table 6.1.) It is shown that the SCEs degrade the performance of nanowire FETs and this effect becomes more serious when the wire diameter is relatively large (e.g., 4nm).

Table 6.1 The values of α_G and α_D for different structures simulated in this section. The gate length is L=8nm and the oxide thickness is 1nm.

Wire Diameter (nm)	2.0	3.0	4.0
α_G	0.91	0.84	0.77
α_D	0.014	0.032	0.059

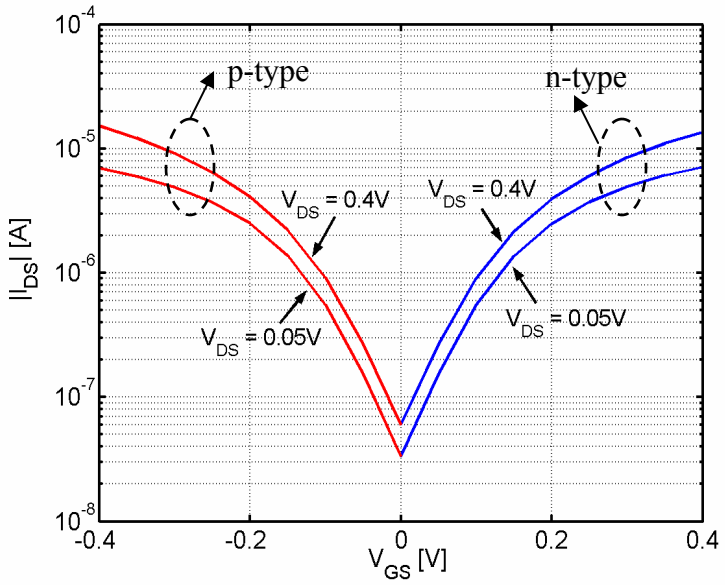
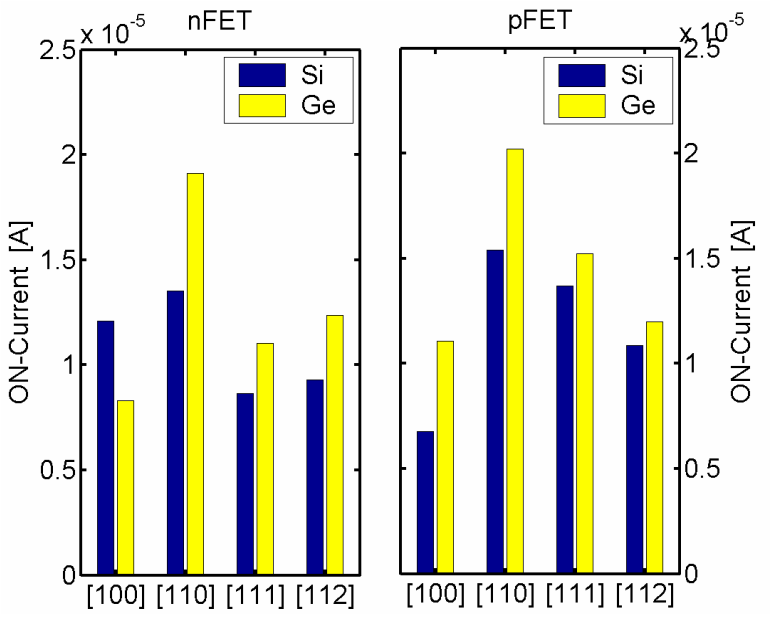
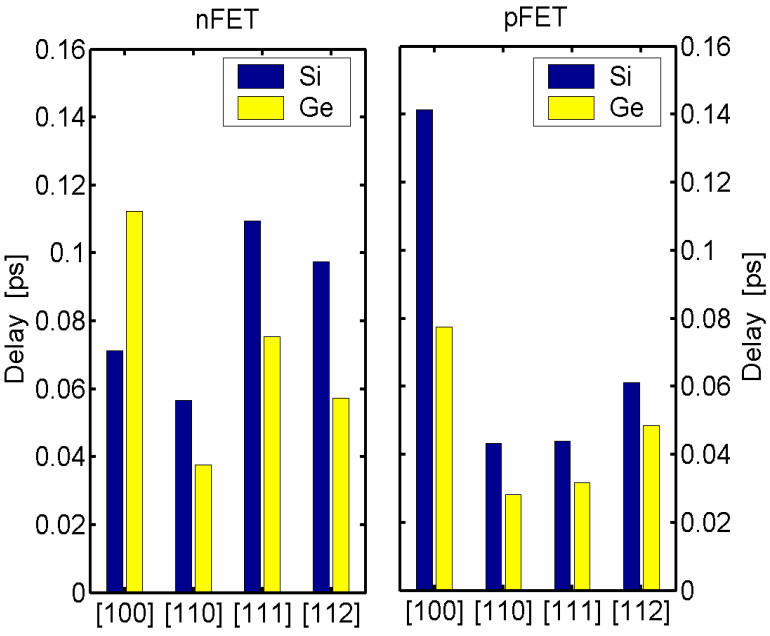


Fig. 6.5 I_{DS} vs. V_{GS} curves in a semi-logarithmic scale for the simulated n-type (right) and p-type (left) Si nanowire FETs with a circular cross-section and a 3nm wire diameter. The channel orientation is $[100]$ and the gate length is L=8nm. α_G=0.84 and α_D=0.032 are adopted to account for the SCEs. Two drain biases, V_{DS}=0.4V and V_{DS}=0.05V, are used in the simulation.

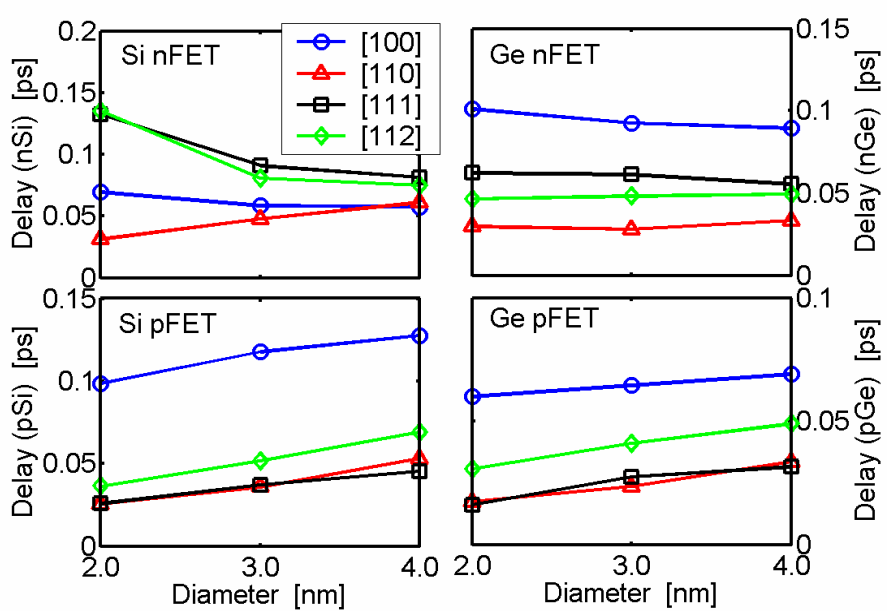


(a)

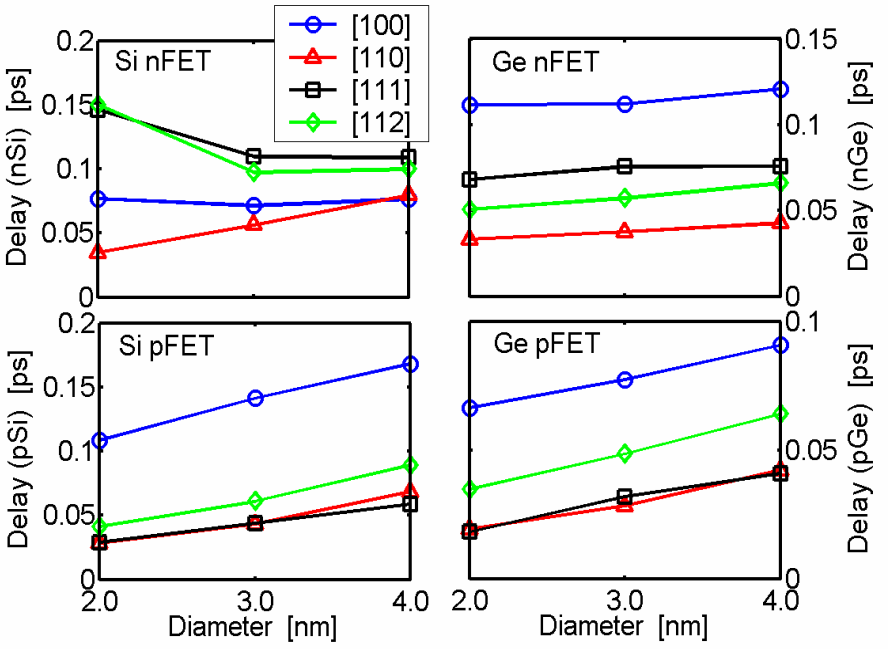


(b)

Fig. 6.6 The ON-currents (a) and intrinsic device delays (b) for the simulated n-/p-type Si and Ge nanowire FETs with four different channel orientations, $[100]$, $[110]$, $[111]$ and $[112]$. For all the devices, the wire diameter is 3nm, the gate length is 8nm and the oxide thickness is 1nm. The supplied voltage (VDD) is 0.4V.



(a)



(b)

Fig. 6.7 The dependence of the intrinsic device delay on the wire diameter for the simulated Si and Ge nanowire FETs (a) without and (b) with the consideration of the SCEs. Four channel orientations, $[100]$, $[110]$, $[111]$ and $[112]$, are explored. For all the devices, the gate length is 8nm and the oxide thickness is 1nm. The supplied voltage (VDD) is 0.4V.

6.4 On the Validity of the Parabolic Effective-Mass Approximation for the Current-Voltage Calculation of Silicon Nanowire Transistors

In this section, we explore the validity of the parabolic effective-mass (pEM) approximation [43] [44] [92] for the current-voltage (*I-V*) calculation of silicon nanowire transistors. To do this, we first compute the energy dispersion relations of Si nanowires by our tight binding (TB) approach. The *I-V* characteristics of n-type SNWTs are then evaluated by the FETToy model [52] [53] using both the tight binding *E-k* relations and parabolic energy bands. By comparing the results for the two types of *E-k* relations, the validity of the parabolic effective-mass approximation is examined.

Figure 6.8 shows an example of the simulated nanowire structures in this section. The transport orientation of the wire is along the $[100]$ direction (see Fig. 6.8a), the shape of the cross-section is square, and the faces of the square are all along the equivalent <100> axes (see Fig. 6.8c). Fig. 6.8b illustrates a unit cell of the nanowire crystal, which consists of four atomic layers along the x (transport) direction and has a length of a_0=5.43Å. It should be noted that although Fig. 6.8 is only for a nanowire with a wire width D=1.36nm, nanowires with various wire widths (from 1.36nm to 6.79nm) are explored in this section.

Figure 6.9 plots the I_{DS} vs. V_{GS} curves for a square SNWT with D=1.36nm in both (a) a semi-logarithmic scale and (b) a linear scale. The dashed lines are for the results based on the tight binding *E-k* relations while the solid lines are for the parabolic effective-mass (pEM) results. In the parabolic effective-mass approach, all six conduction-band valleys in bulk Si are considered, and the effective-masses used in the calculation (m_l=0.891m_e and m_t=0.201m_e) are extracted from the bulk *E-k* relation evaluated by our tight binding approach with the parameters obtained from [49]. (By doing this, the <~5% deviation in bulk Si effective-masses caused by the TB parameters [49] are prevented from affecting our comparison between TB and pEM.) If we define a threshold voltage, V_T, as

$$I_{DS}(V_{GS}=V_T, V_{DS}=0.4\text{V})=300\text{nA}, \tag{6.1}$$

and an ON-current of SNWTs as

$$I_{ON}=I_{DS}(V_{GS}-V_T=0.3\text{V},\ V_{DS}=0.4\text{V}), \tag{6.2}$$

we find that pEM significantly overestimates the threshold voltage by $V_T^{pEM}-V_T^{TB}=0.28\text{V}$ and the ON-current by $\left(I_{ON}^{pEM}-I_{ON}^{TB}\right)/I_{ON}^{TB}=42\%$ as compared with the tight binding results. Fig. 6.10 compares pEM (solid) vs. tight binding (circles) for the *I-V* calculation of a thicker SNWT with D=6.79nm. It is clear that pEM provides nearly identical *I-V* characteristics as tight binding except for a small overestimation of ON-current by ~5%. The solid lines with circles in Fig. 6.11 show the wire width (D) dependence of the errors, $V_T^{EM}-V_T^{TB}$ in (a) and $\left(I_{ON}^{EM}-I_{ON}^{TB}\right)/I_{ON}^{TB}$ in (b), associated with pEM. It is clear that pEM starts to overestimate threshold voltage by >0.03V when D scales below 3nm and ON-current by ≥10% when D is ≤5nm.

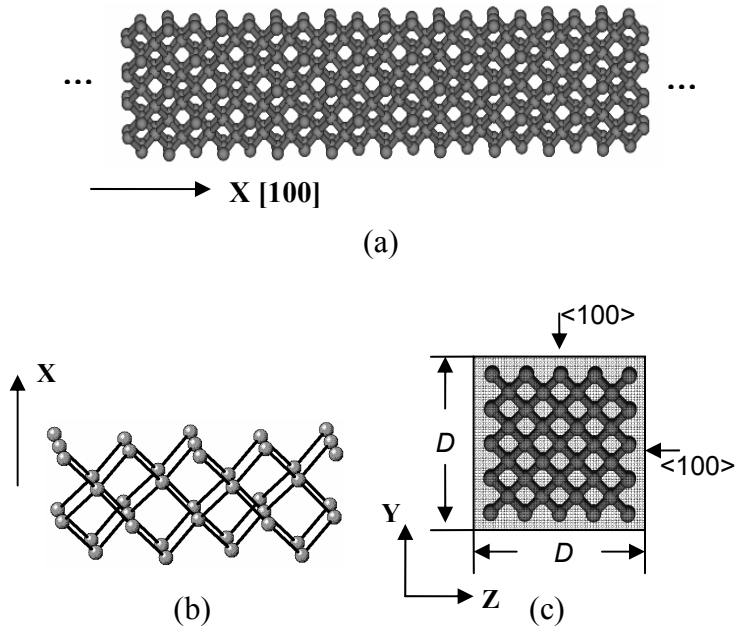
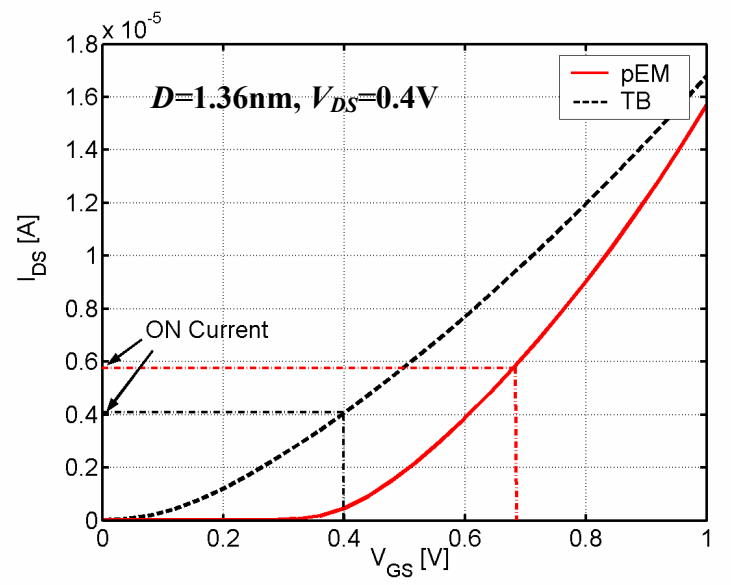
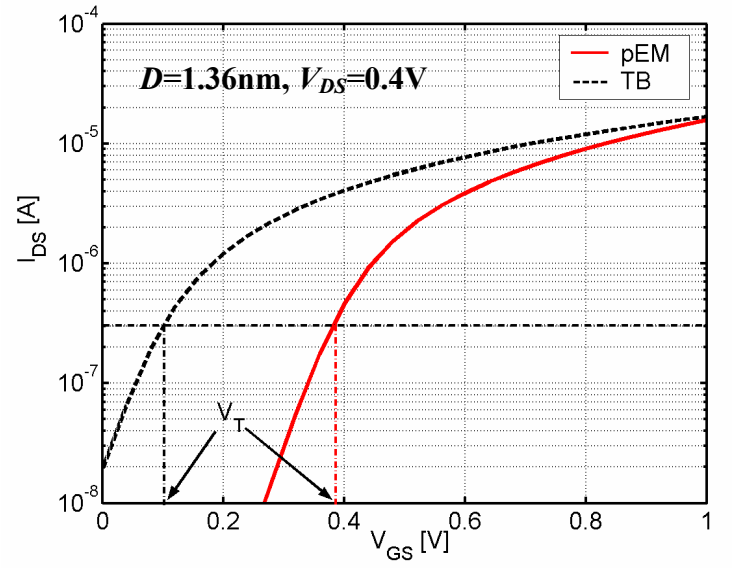


Fig. 6.8 (a) The atomic structure of a square nanowire (D=1.36nm) with a $[100]$ transport direction. (b) A unit cell of the square nanowire illustrated in (a). (c) The schematic diagram of the cross-section of the square nanowire. D demotes the edge length of the square cross-section and the four faces of the square are all along the equivalent <100> axes.



(a)



(b)

Fig. 6.9 The I_{DS} vs. V_{GS} curves for a square SNWT with D=1.36nm in both (a) a semi-logarithmic scale and (b) a linear scale. The oxide thickness is 1nm, the temperature is 300K, and the drain bias is 0.4V. The dashed lines are for the results based on the tight binding (TB) E-k relations while the solid lines for the parabolic effective-mass (pEM) results.

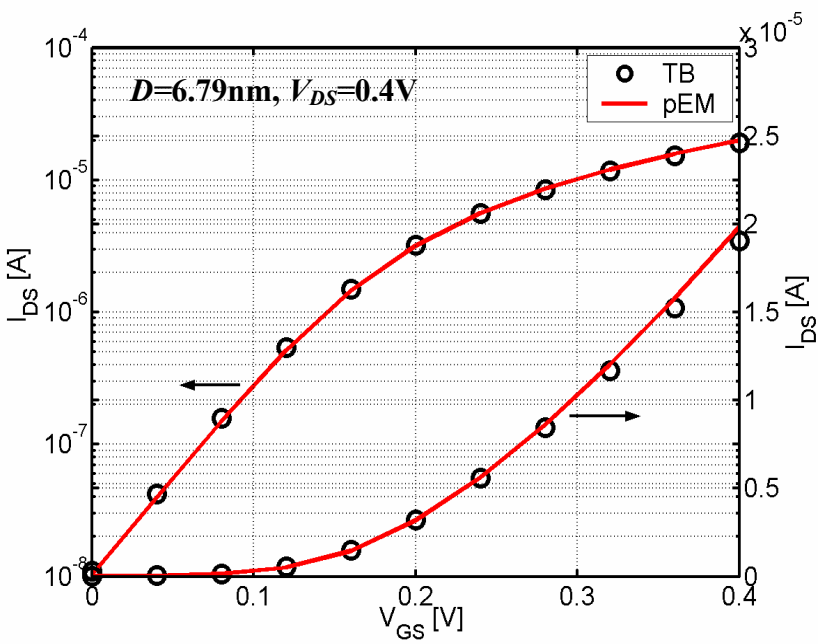


Fig. 6.10 The I_{DS} vs. V_{GS} curves for a square SNWT with D=6.79nm in both a semi-logarithmic scale (left) and a linear scale (right). The oxide thickness is 1nm, the temperature is 300K, and the drain bias is 0.4V. The circles are for the results based on the tight binding (TB) E-k relations while the solid lines for the parabolic effective-mass (pEM) results.

To understand the above observations, we plot the D dependence of the wire conduction band-edges, E_C, and the transport effective-mass, m_x^*, at the Γ point in the wire conduction band (Fig. 6.12). (Note that the square wires are the nominal structures we focus on in this section. For comparison, we also show the results for circular wires with the same ([100]) wire orientation. The results illustrate that the energy dispersion relations are nearly invariant when the cross-sectional shape changes from square to circular, indicating that the conclusions in this work also apply to wires with a circular cross-section.) As we can see in Fig. 6.12a, when D>4nm (Area>16nm^2), the E_C obtained from the tight binding calculations (solid with squares) is well reproduced by pEM (dashed). (In pEM, the wire conduction band-edge is determined by the lowest subband level of the four unprimed valleys). At smaller wire widths, however, pEM overestimates E_C due to the nonparabolicity [36] [95] of the bulk Si bands. This overestimation of E_C by pEM directly leads to the overvalued threshold voltages of the

simulated SNWTs. The solid line with squares in Fig. 6.12b shows an increasing m_x^* (extracted from the tight binding *E-k* relations) with a decreasing *D*, which is also a result of the nonparabolicity of the bulk Si *E-k* relations. When *D*<3nm (Area<9nm^2), m_x^* extracted from tight binding is >40% larger than the corresponding bulk value used in pEM. Since the electron thermal velocity is inversely proportional to the square root of the transport effective-mass, the pEM calculations, which adopt a smaller m_x^* than the tight binding approach, overestimate the carrier injection velocity and consequently the SNWT ON-currents. In short, the nonparabolicity of the bulk Si bands plays an important role when quantum confinement is strong (small *D*). The use of parabolic energy bands overestimates the wire conduction band-edge and underestimates the transport effective-mass, and consequently provides a higher SNWT threshold voltage and ON-current as compared with the tight binding approach.

Although we have shown that the *parabolic* effective-mass approach does not perform well at small wire widths, it is still interesting to know whether it is possible to modify the effective-mass approach to obtain a better agreement with the tight binding calculation, since the effective-mass approximation significantly reduces computation time as compared to atomistic treatments. To do this, we first define a quantum confinement energy as the difference between the wire conduction band-edge, E_C and that for bulk Si ($E_C^{bulk} = 1.13\text{eV}$ [49]). Fig. 6.13a shows the quantum confinement energy computed by pEM (E_{QC}^{pEM}) vs. that obtained from the tight binding calculation (E_{QC}^{TB}). It is evident that for small wire widths, the data points (circles) stand above the *y*=*x* curve, indicating that pEM overestimates the quantum confinement energy when it is large. Inspired by the expressions for the nonparabolicity of the bulk Si bands [36] [95], we propose the following quadratic equation to analytically describe the E_{QC}^{pEM} vs. E_{QC}^{TB} relation,

$$E_{QC}^{TB} \cdot \left(1 + \alpha \cdot E_{QC}^{TB}\right) = E_{QC}^{pEM}, \tag{6.3}$$

where α is treated as a fitting parameter and $\alpha = 0.27\text{eV}^{-1}$ is used for the solid line in Fig. 6.13a for the best agreement with the extracted data. Similarly, the E_{QC}^{TB} dependence of

the transport effective-mass m_x^* at the Γ point (Fig. 6.13b) can also be described by the following equation,

$$m_x^* = m_{bulk}^* \left(1 + \beta \cdot E_{QC}^{TB}\right), \tag{6.4}$$

where $m_{bulk}^* = m_t = 0.201 m_e$ is the transport effective-mass in the unprimed valleys in bulk Si and $\beta = 1.5\text{eV}^{-1}$ is chosen to achieve the best match between the extracted data points (circles) and the analytical expression (solid) up to $E_{QC}^{TB} = 1\text{eV}$, which is sufficient for the *I-V* calculation of the simulated Si nanowire transistors.

After knowing Eqs. (6.3) and (6.4), the effective-mass approximation can be tuned for a better fit with tight binding in the following steps.

Step 1) Calculate the quantum confinement energy, E_{QC}^{pEM} by the parabolic effective-mass approach with the bulk effective-masses (i.e., m_y^* and m_z^*).

Step 2) Solve Eq. (6.3) for the updated quantum confinement energy, E_{QC}^{new}, as

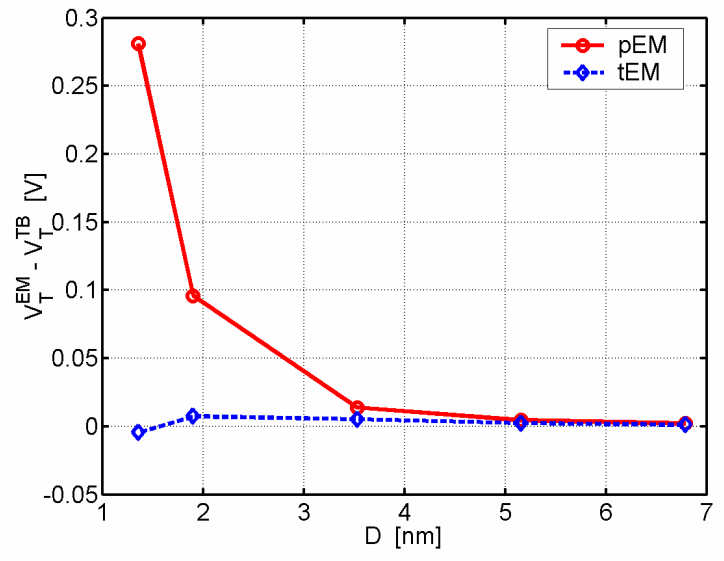
$$E_{QC}^{new} = \frac{-1 + \sqrt{1 + 4\alpha \cdot E_{QC}^{pEM}}}{2\alpha}. \tag{6.5}$$

Step 3) Evaluate the tuned transport effective-mass at the Γ point by Eq. (6.4),

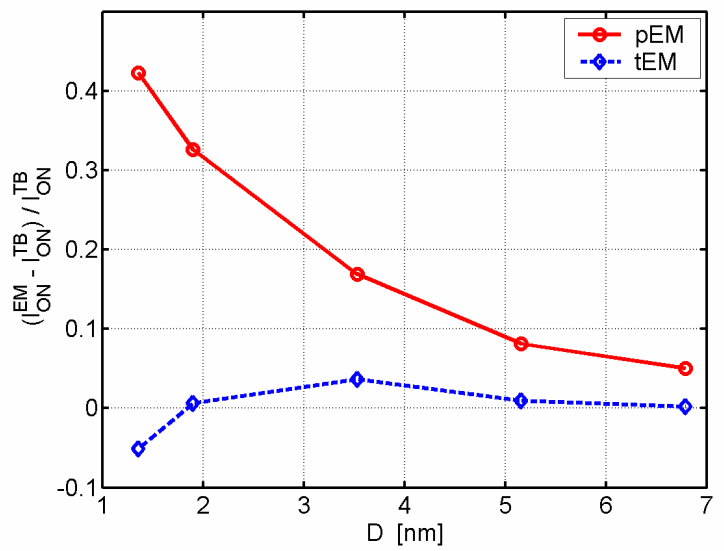
$$m_x^* = m_{bulk}^* \left(1 + \beta \cdot E_{QC}^{new}\right). \tag{6.6}$$

Step 4) Use the computed E_{QC}^{new} and m_x^* for the *I-V* calculation of SNWTs.

It should be noted that the above tuning process is only necessary for the four unprimed valleys because 1) at large wire widths, the quantum confinement energy is small and nonparabolicity is insignificant in both unprimed and primed valleys, so the parabolic effective-mass approach performs well, and 2) at small wire widths, the two primed valleys are well separated from the unprimed ones due to stronger quantum confinement (smaller effective-masses in the *y* and *z* directions) in these primed valleys, so the electron density and current contributed by the primed valleys are negligible (e.g., when $E_{QC}^{TB} > 0.15\text{eV}$, over 97% electrons are distributed in the unprimed valleys).

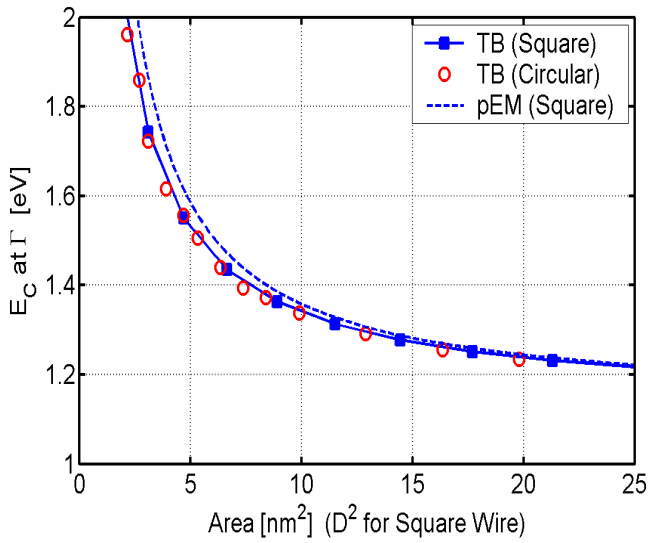


(a)

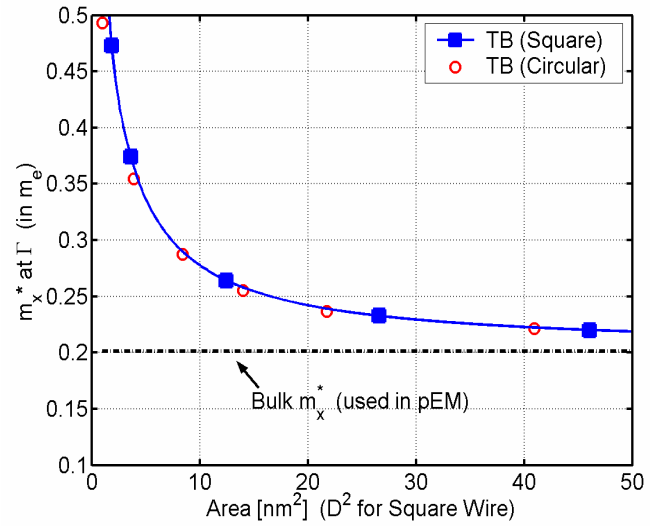


(b)

Fig. 6.11 The wire width (D) dependence of the errors, $V_T^{EM} - V_T^{TB}$ in (a) and $\left(I_{ON}^{EM} - I_{ON}^{TB}\right) / I_{ON}^{TB}$ in (b), associated with the effective-mass approximations. The solid lines with circles are for the parabolic effective-mass (pEM) approximation while the dashed lines with diamonds are for the tuned effective-mass (tEM) approach.

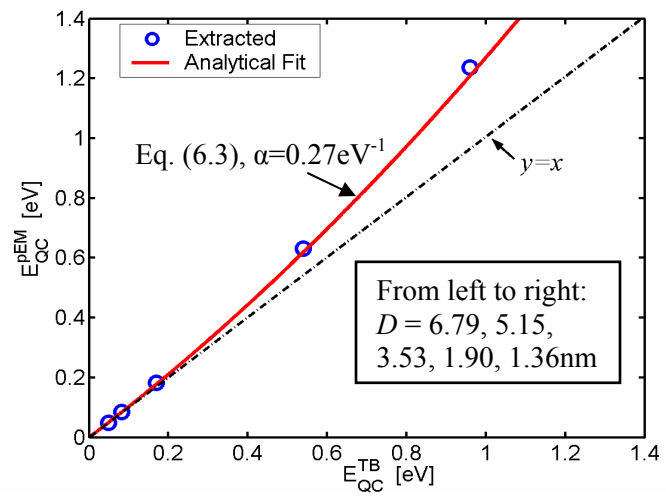


(a)

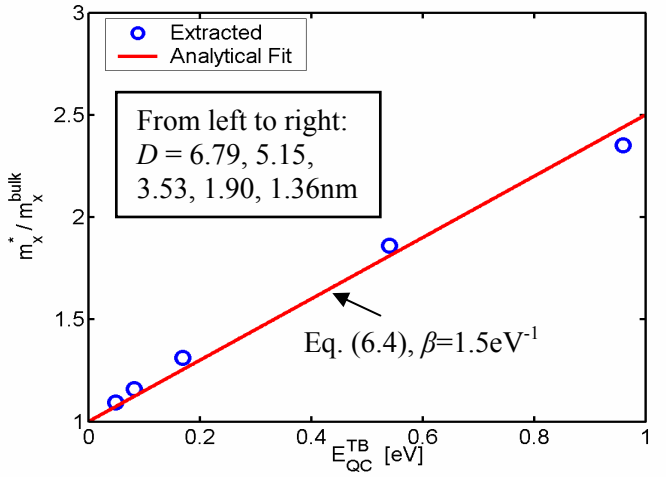


(b)

Fig. 6.12 (a) The conduction band-edges, E_C, for the simulated wires with different wire widths. The solid line with squares is for the values for the square wires obtained from the tight binding (TB) *E-k* relations and the dashed line is for the corresponding parabolic effective-mass (pEM) results. For comparison, the TB values for the circular wires with a $[100]$ wire orientation are also shown (circles). (b) The wire width (D) dependence of the transport effective-mass, m_x^*, at the Γ point in the wire conduction band (extracted from the tight binding energy bands by $m_x^* = \hbar^2 / \left(\partial^2 E / \partial^2 k_x\right)$, where $\hbar$ is the Plank constant). The solid line with squares is for the square wires while circles are for the circular wires. For comparison, the bulk value of m_x^* for the unprimed valleys (used in pEM) is shown by the dash-dot line.



(a)



(b)

Fig. 6.13 (a) The quantum confinement energy computed by parabolic effective-mass (E_{QC}^{pEM}) vs. that obtained from the tight binding calculation (E_{QC}^{TB}). (b) The ratio of the transport effective-mass, m_x^* to the bulk value, m_x^{bulk} vs. E_{QC}^{TB}. In both plots, the circles are for the data points extracted from the tight binding and parabolic *E-k* relations (from left to right: *D*=6.79nm, 5.15nm, 3.53nm, 1.90nm, and 1.36nm), while the solid lines are for the analytical fit based on Eqs. (6.3) and (6.4).

The dashed lines with diamonds in Fig. 6.11 show the wire width (*D*) dependence of the errors, $V_T^{EM} - V_T^{TB}$ in (a) and $\left(I_{ON}^{EM} - I_{ON}^{TB}\right)/I_{ON}^{TB}$ in (b), associated with the *tuned* effective-mass approximation. For wire widths ranging from 1.36nm to 6.79nm, the

tuned effective-mass approach provides an excellent match with the tight binding calculation – less than 10mV error for V_T and less than 5% error for I_{ON}. So far, we have shown that the effective-mass approximation can be modified by introducing two *D*-independent parameters, α and β, to accurately reproduce the *I-V* results computed by tight binding. It must be mentioned that the *values* of α and β used in this work were obtained for SNWTs with one particular channel orientation (i.e., $[100]$) and one specific cross-sectional shape (i.e., square with all faces along the equivalent <100> axes). The important point is that for *I-V* calculation it is possible to simply tune the effective-mass approach to fit the tight binding model. We expect that this conclusion may apply to other SNWTs with different transport directions and cross-sections while the *values* of the tuning parameters (α and β) are subject to change.

In this section, by using our tight binding approach as a benchmark, we examined the validity of the parabolic effective-mass approximation [43] [44] [92] for the *I-V* calculation of n-type silicon nanowire transistors. It was found that the simple parabolic effective-mass approach with bulk effective-masses significantly overestimates SNWT threshold voltages when the wire width (*D*) is <3nm, and SNWT ON-currents when *D*<5nm. However, by introducing two analytical equations with two tuning parameters, the effective-mass approximation can well reproduce the tight binding *I-V* results over a wide range of wire widths – even at *D*=1.36nm. In conclusion, bandstructure effects begin to manifest themselves in silicon nanowire transistors with small diameters, but with a simple tuning procedure, the parabolic effective-mass approximation may still be used to assess the performance limits of silicon nanowire transistors.

6.5 Summary

In this chapter, based on a nearest-neighbor $sp^3d^5s^*$ tight binding approach [49] [50] [51], we developed a simulator that can calculate the energy dispersion relations for Si and Ge nanowires with arbitrary wire orientations and cross-sectional shapes. The *I-V* characteristics of various Si and Ge nanowire FETs were computed by using the FETToy

model [52] [53] based on the tight binding *E-k* relations. The results show that 1) $[110]$ is the optimum channel orientation for both n-type and p-type Si/Ge nanowire FETs, which offers the highest ON-current and the fastest intrinsic device delay for the same OFF-current, and 2) the device performance of p-type Si/Ge nanowire FETs is improved as the wire diameter scales down, while for n-type Si/Ge nanowire FETs, the dependence of the device performance on wire diameter is sensitive to the material type and the wire orientation. In addition, we also examined the validity of the widely used parabolic effective-mass approximation [43] [44] [92] for the *I-V* calculation of SNWTs. It is shown that the parabolic effective-mass approach significantly overestimates SNWT threshold voltages and ON-currents at small wire widths. By introducing a simple tuning procedure, however, the effective-mass approximation can well produce accurate *I-V* results over a wide range of wire widths (even when D=1.36nm), as calibrated with the tight binding approach.

7. SUMMARY AND OUTLOOK

This work addressed device physics, modeling and design issues of silicon nanowire transistors (SNWTs). The main accomplishments of this thesis are:

1) Chapter 2 discussed an analytical theory for ballistic nanowire FETs, which is derived by modifying a semi-numerical, ballistic FET model, 'FETToy', proposed by A. Rahman and coworkers [52] [53] for ballistic planar MOSFETs. Based on this simple approach, the essential physics and peculiarities of 1D wire FETs were investigated. This FETToy model also played an important role when evaluating the ballistic performance limits of SNWTs with the atomistic bandstructures (Chapter 6).
2) In Chapter 3, we described a self-consistent, full three-dimensional (3D) quantum simulator of ballistic SNWTs based on the non-equilibrium Green's function (NEGF) formalism [37] [38] and the effective mass (EM) approximation. [43] [44] The coupled/uncoupled mode space approach [41] [45] [47] was introduced, which significantly reduced the computational expense while maintaining great accuracy as compared with the full 3D real space representation. This makes our simulator a practical tool for the simulation and design of ballistic SNWTs with various cross-sections (e.g., triangular, rectangular and cylindrical). [4] Within the NEGF framework shown in this chapter, scattering in SNWTs can be phenomenologically treated by a simple model, so called the 'Büttiker probes' [42] [57]. The details of this method are discussed in Appendix.
3) Chapter 4 discussed the performance limits and scaling potential of ballistic SNWTs. It addressed three different topics. Sec. 4.1 showed a comparison between the upper performance limit of SNWTs with that of the planar

double-gate MOSFET. [4] The results showed that SNWTs scale better than planar devices. In Sec. 4.2, we proposed a general approach to compare planar vs. non-planar (nanowire) FETs with the consideration of both Electrostatic integrity (gate control) and Quantum confinement (so called the 'EQ approach'). [58] The scaling capabilities of different structures (e.g., cylindrical wire FETs, double-gate planar FETs and tri-gate FETs) were compared using this approach. Sec. 4.3 introduced a conceptual study of the channel material optimization for both planar MOSFETs and nanowire FETs based on the effective-mass approximation. [59] The results implied that an optimum effective-mass can be defined for a given structure, which provides the highest ON-current for the same OFF-current.

4) In Chapter 5, we presented a microscopic simulation of surface roughness scattering (SRS) in SNWTs by using the 3D quantum simulator developed in Chapter 3. [60] The microscopic structure of the Si/SiO_2 interface roughness was directly implemented based on the relevant auto-covariance function [61]. The results showed that SRS in SNWTs becomes less serious when fewer propagating modes conduct, implying that SRS will be less important in small-diameter SNWTs than in planar MOSFETs with many transverse modes occupied.
5) In Chapter 6, we performed atomistic simulations of Si and Ge nanowire transistors. [63] [64] Based on a nearest-neighbor $sp^3d^5s^*$ tight binding approach [49] [50] [51], we developed a simulator that can calculate *E-k* relations for unrelaxed Si and Ge nanowires with arbitrary wire orientations and cross-sectional shapes. The *I-V* characteristics of various Si and Ge nanowire FETs were computed by using the FETToy model [52] [53], introduced in Chapter 2, based on the tight binding *E-k* relations. The impact of bandstructure effects on SNWT performance was investigated and the channel orientation optimization was done for both Si and Ge nanowire FETs. Finally, the validity of the parabolic effective-mass approximation [43] [44] [92] for *I-V* calculation of n-type SNWTs was examined.

This work can be extended as follows in the future.

1) Electron-phonon interactions. Understanding carrier transport in Si nanowires is of great importance for the assessment of the performance limits of SNWTs. In this thesis, we completed a microscopic treatment of surface roughness scattering in SNWTs with small diameters. Furthermore, the electron-phonon interaction [98] is another important scattering mechanism that may limit the ultimate performance of SNWTs. Chapter 6 in this thesis describes a simulator that can calculate the *E-k* relations for Si and Ge nanowires with arbitrary wire orientations and cross-sectional shapes. This work offers a useful basis for the calculation of phonon scattering rate and phonon-limited carrier mobility using Fermi's Golden Rule [36], assuming the phonon modes in Si nanowires are correctly obtained. More work can also be done to solve a coupled phonon-electron/hole Boltzmann transport equation [36] to explore the effects of phonon scattering on SNWT device characteristics.
2) Nanowire heterostructures. Recently, different experimental groups reported coaxial or longitudinal nanowire heterostructures fabricated in a bottom-up approach [22] [23] [25]. These structures are expected to have potential applications in nanoelectronics and nano-optics. Following the work in Chapter 6, more work can be done to extend our current tight binding simulator to calculate the energy dispersion relations for nanowire heterostructures, which will be important for exploring the essential physics and design issues of nanowire heterostructure devices. In general, although we focused on the field-effect transistor application of Si (Ge) nanowires in this thesis, the simulation capabilities developed here may also be used or extended to explore novel applications of semiconductor nanowires.

LIST OF REFERENCES

LIST OF REFERENCES

[1] G. Moore, "Progress in digital integrated electronics," *IEEE International Electron. Dev. Meeting (IEDM) Tech. Digest*, pp. 11-13, 1975.

[2] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge, UK: Cambridge University Press, 1998.

[3] Y. Cui, Z. Zhong, D. Wang, W. Wang and C. M. Lieber, "High performance silicon nanowire field-effect transistors," *Nano Lett.*, vol. 3, no. 2, pp. 149-152, Feb. 2003.

[4] J. Wang, E. Polizzi and M. Lundstrom, "A computational study of ballistic silicon nanowire transistors," *IEEE International Electron Dev. Meeting (IEDM), Tech. Digest*, pp. 695-698, Dec. 8-10 2003.

[5] A. Javey, J. Guo, Q. Wang, M. Lundstrom and H. Dai, "Schottky barrier free nanotube transistors near the ballistic transport," *Nature,* vol. 424, pp. 654-657, Aug. 2003.

[6] J. Guo, M. Lundstrom and S. Datta, "Performance projections for ballistic carbon nanotube field-effect transistors," *Appl. Phys. Lett.*, vol. 80, no. 17, pp. 3192-3194, Apr. 2002.

[7] C. W. Leitz, M. T. Currie, M. L. Lee, Z.-Y. Cheng, D. A. Antoniadis, and E. A. Fitzgerald, "Hole mobility enhancements in strained $Si/Si_{1-y}Ge_y$ p-type metal oxide-semiconductor field-effect transistors grown on relaxed $Si_{1-x}Ge_x$ ($x<y$) virtual substrates," *Appl. Phys. Lett.*, vol. 79, no. 25, pp. 4246-4248, Dec. 2001.

[8] A. Rahman, A. Ghosh and M. Lundstrom, "Assessment of Ge n-MOSFETs by quantum simulation," *IEEE International Electron Dev. Meeting (IEDM), Tech. Digest*, pp. 471-474, Dec. 8-10 2003.

[9] M. Reed, "Molecular-scale electronics," *Proc. of the IEEE*, vol. 87, pp. 652-658, 1999.

[10] F. –L. Yang, D. –H. Lee, H. -Y Chen, C. –Y. Chang, S.-D. Liu, C. –C. Huang, et al., "5nm gate nanowire FET," *IEEE Symposium on VLSI Tech., Digest of Technical Papers*, pp. 196-197, Jun. 2004.

[11] B. S. Doyle, S. Datta, M. Doczy, S. Hareland, B. Jin, J. Kavalieros et al, "High performance fully-depleted tri-gate CMOS transistors," *IEEE Electron Dev. Lett.*, vol. 24, no. 4, pp. 263-265, Apr. 2003.

[12] H. Majima, Y. Saito and T. Hiramoto, "Impact of quantum mechanical effects on design of nano-scale narrow channel n- and p-type MOSFETs," *IEEE International Electron Dev. Meeting (IEDM) Tech. Digest*, pp. 951-954, Dec. 3-5 2001.

[13] T. Saito, T. Saraya, T. Inukai, H. Majima, T. Nagumo and T. Hiramoto, "Suppression of short channel effects in triangular parallel wire channel MOSFETs," *IEICE Trans. Electron.*, vol. E85-C, no. 5, pp. 1073-1078, May 2002.

[14] M. Je, S. Han, I. Kim and H. Shin, "A silicon quantum wire transistor with one-dimensional subband effects," *Solid-State Electronics*, vol. 28, pp. 2207-2212. Dec. 2000.

[15] S. H. Zaidi, A. K. Sharma, R. Marquardt, S. L. Lucero and P. M. Varangis, "Multiple nanowire field-effect transistors," *Proc of the 2001 1st IEEE Conference on Nanotechnology*, pp. 189-194, Oct. 2001.

[16] C. P. Auth and J. D. Plummer, "Scaling theory for cylindrical, fully-depleted, surrounding-gate MOSFETs," *IEEE Electron Dev. Lett.*, vol. 18, no. 2, pp. 74-76, Feb. 1997.

[17] J. Wang and M. Lundstrom, "Does source-to-drain tunneling limit the ultimate scaling of MOSFETs?" *IEEE International Electron Dev. Meeting (IEDM) Tech. Digest*, pp. 707-710, Dec. 7-9 2002.

[18] D. Wang, Q. Wang, A. Javey, R. Tu, H. Dai, H. Kim, P. McIntyre, T. Krishnamohan, and K. Saraswat, "Germanium nanowire field-effect transistors with SiO_2 and high-kappa HfO_2 gate dielectrics," *Appl. Phys. Lett.*, vol. 83, no. 12, pp. 2432-2434, Sep. 2003.

[19] Y. Wu. Y. Cui, L. Huynh, C. J. Barrelet, D. C. Bell, C. M. Lieber, "Controlled growth and structures of molecular-scale silicon nanowires," *Nano Lett.,* vol. 4, no. 3, pp. 433-436, Feb. 2004.

[20] D. D. D. Ma, C. S. Lee, F. C. K. Au, S. Y. Tong and S. T. Lee, "Small-diameter silicon nanowire surfaces," *Science*, vol. 299, pp. 1874-1877, Mar. 2003.

[21] N. A. Melosh, A. Boukai, F. Diana, B. Gerardot, A. Badolato, P. M. Petroff and J. R. Heath, "Ultrahigh-density nanowire lattices and circuits," *Science*, vol. 300, pp. 112-115, Apr. 2003.

[22] L. Samuelsona, M. T. Bjorka, K. Depperta, M. Larssonb, B. J. Ohlssonc, N. Paneva, A. I. Perssona, N. Skolda, C. Thelandera and L. R. Wallenbergb, "Semiconductor nanowires for novel one-dimensional devices," *PHYSICA E*, vol. 21, pp. 560-567, Mar. 2004.

[23] Y. Wu, R. Fan and P. Yang, "Block-by-block growth of single-crystalline Si/SiGe superlattice nanowires," *Nano Lett.* vol. 2, no. 2, pp. 83-86, Feb. 2002.

[24] S. Chung, J. Yu and J. R. Heath, "Silicon nanowire devices," *App. Phys. Lett.,* vol. 76, no. 15, pp. 2068-2070, Apr. 2000.

[25] M. S. Gudiksen, L. J. Lauhon, J. Wang, D. C. Smith and C. M. Lieber, "Growth of nanowire superlattice structures for nanoscale photonics and electronics," *Nature*, vol. 415, pp. 617-620, Feb. 2002.

[26] Y. Huang, X. Duan, Y. Cui, L. J. Lauhon, K. H. Kim and C. M. Lieber, "Logic gates and computation from assembled nanowire building blocks," *Science*, vol. 294, pp. 1313-1317, Nov. 2001.

[27] X. Duan, Y. Huang and C. M. Lieber, "Nonvolatile memory and programmable logic from molecule-gated nanowires," *Nano Lett.* vol. 2, no. 5. pp. 487-490, May 2002.

[28] Z. Zhong, D. Wang, Y. Cui, M. W. Bockrath and C. M. Lieber, "Nanowire crossbar arrays as address decoders for integrated nanosystems," *Science*, vol. 302, pp. 1377-1379, Nov. 2003.

[29] Y. Huang, D. Duan, Q. Wei and C. M. Lieber, "Directed assembly of one-dimensional nanostructures into functional networks," *Science*, vol. 291, pp. 630-633, Jan. 2001.

[30] X. Duan, C. Niu, V. Sahi, J. Chen, J. W. Parce, S. Empedocles and J. L. Goldman, **"**High-performance thin-film transistors using semiconductor nanowires and nanoribbons," *Nature*, vol. 425, pp. 274-278, Sep. 2003.

[31] X. Duan, Y. Huang, Y. Cui, J. Wang and C. M. Lieber, "Indium phosphide nanowires as building blocks for nanoscale electronic and optoelectronic devices," *Nature*, vol. 409, pp. 66-69, Jan. 2001.

[32] X. Duan, Y. Huang, R. Agarwal and C. M. Lieber, "Single-nanowire electrically driven lasers," *Nature*, vol. 421, pp. 241-245, Jan. 2003.

[33] M. H. Huang, S. Mao, H. Feick, H. Yan, Y. Wu, H. Kind, E. Weber, R. Russo and P. Yang, "Room-temperature ultraviolet nanowire nanolasers," *Science*, vol. 292, pp. 1897-1899, Jun. 2001.

[34] J. Wang, M. S. Gudiksen, X. Duan, Y. Cui and C. M. Lieber, "Highly polarized photoluminescence and photodetection from single indium phosphide nanowires," *Science*, vol. 293, pp. 1455-1457, Aug. 2001.

[35] Y. Cui, Q. Wei, H. K. Park and C. M. Lieber, "Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species," *Science*, vol. 293, pp. 1289-1292, Aug. 2001.

[36] M. Lundstrom, *Fundamentals of Carrier Transport*, 2nd edition, Cambridge, UK: Cambridge University Press, 2000.

[37] S. Datta, *Electronic Transport in Mesoscopic Systems*, Cambridge, UK: Cambridge University Press, 1997.

[38] S. Datta, "Nanoscale device modeling: the Green's function method," *Superlattices and Microstructures*, vol. 28, pp. 253-278, Oct. 2000.

[39] Z. Ren, R. Venugopal, S. Datta, M. Lundstrom, D. Jovanovic and J. G. Fossum, "The ballistic nanotransistor: a simulation study," *IEEE International Electron Dev. Meeting (IEDM), Tech. Digest*, pp. 715-718, Dec. 2000.

[40] 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. Electron. Dev.*, vol. 50, no. 9, pp. 1914-1925, Sep. 2003.

[41] R. Venugopal, Z. Ren, S. Datta and M. Lundstrom, "Simulating quantum transport in nanoscale transistors: real versus mode-space approaches," *J. Appl. Phys.*, vol. 92, no. 7, pp. 3730-3739, Oct. 2002.

[42] 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, no. 9, pp. 5613-5625, May 2003.

[43] S. Horiguchi, "Validity of effective mass theory for energy levels in Si quantum wires," *Physica B*, vol. 227, pp. 336-338, Sep. 1996.

[44] J. Wang, E. Polizzi and M. Lundstrom, "A three-dimensional quantum simulation of silicon nanowire transistors with the effective-mass approximation," *J. Appl. Phys.*, vol. 96, no. 4, pp. 2192 – 2203, Aug. 2004.

[45] E. Polizzi and N. Ben Abdallah, "Self-consistent three-dimensional models for quantum ballistic transport in open systems," *Phys. Rev. B*, vol. 66, no. 24, pp. 245301.1 – 245301.9, Dec. 2002.

[46] E. Polizzi and S. Datta, "Multidimensional nanoscale device modeling: the finite element method applied to the non-equilibrium Green's function formalism," *Proc. of the 2003 3rd IEEE Conference on Nanotechnology*, pp. 40-43, Aug. 2003.

[47] E. Polizzi E and N. Ben Abdallah, "Subband decomposition approach for the simulation of quantum electron transport in nanostructures," *J. Computational Phys.*, vol. 202, no. 1, pp. 150-180, Jan. 2005.

[48] Z. Ren, *Nanoscale MOSFETs: Physics, Simulation, and Design*, Ph. D. Thesis, Purdue University, West Lafayette, IN, 2001.

[49] T. B. Boykin, G. Klimeck and F. Oyafuso, "Valence band effective-mass expressions in the sp3d5s* empirical tight-binding model applied to a Si and Ge parametrization," *Phys. Rev. B*, vol. 69, pp. 115201.1-115201.10, Mar. 2004.

[50] T. B. Boykin, G. Klimeck, M. A. Eriksson, M. Friesen, S. N. Coppersmith, P. Von Allmen, F. Oyafuso and S. Lee, "Valley spitting in strained silicon quantum wells," *App. Phys. Lett.*, vol. 84, no. 1, pp. 115-117, Jan. 2004.

[51] T. B. Boykin, G. Klimeck, M. Friesen, S. N. Coppersmith, P. von Allmen, F. Oyafuso and S. Lee, "Valley splitting in low-density quantum-confined heterostructures studied using tight-binding models," *Phys. Rev. B*, vol. 70, pp. 165325.1-165325.12, Oct. 2004.

[52] A. Rahman, J. Guo, S. Datta and M. Lundstrom, "Theory of ballistic nanotransistors," *IEEE Trans. Electron Dev.*, vol. 50, no. 9, pp. 1853-1864, Sep. 2003.

[53] www.nanohub.org, 2005.

[54] J. Wang and M. Lundstrom, "Ballistic transport in high electron mobility transistors," *IEEE Trans. Electron Dev.*, vol. 50, no. 7, pp. 1604-1609, Jul. 2003.

[55] S. Luryi, "Quantum capacitance devices," *Appl. Phys. Lett.*, vol. 52, pp. 501-503, Feb. 1988.

[56] P. M. Solomon and S. E. Laux, "The ballistic FET: design, capacitance and speed limit," *IEEE International Electron Dev. Meeting (IEDM), Tech. Digest*, pp. 95-98, Dec. 3-5 2001.

[57] M. Büttiker, "4-terminal phase-coherent conductance," *Phys. Rev. Lett.*, vol. 57, pp. 1761-1764, Oct. 1986.

[58] J. Wang, P. Solomon and M. Lundstrom, "A general approach for the performance assessment of nanoscale silicon field effect transistors," *IEEE Trans. Electron Dev.*, vol. 51, no. 9, pp. 1366-1370, Sep. 2004.

[59] J. Wang and M. Lundstrom, "Channel material optimization for the ultimate planar and nanowire MOSFETs: a theoretical exploration," presented at *Device Research Conference (DRC)*, Santa Barbara, CA, June 20-22 2005.

[60] J. Wang, E. Polizzi, A. Ghosh, S. Datta and M. Lundstrom, "A microscopic quantum simulation of Si/SiO_2 interface roughness scattering in silicon nanowire transistors," presented at the 10^{th} International Workshop on Computational Electronics (IWCE-10), West Lafayette, IN, Oct. 25-27 2004.

[61] S. M. Goodnick, D. K. Ferry, C. W. Wilmsen, Z. Liliental, D. Fathy and O. L. Krivanck, "Surface roughness at the $Si(100)$-SiO_2 interface," *Phys. Rev. B*, vol. 32, no. 12, pp. 8171-8186, Dec. 1985.

[62] T. Yoshinobu, A. Iwamoto and H. Iwasaki, "Scaling analysis of SiO_2/Si interface roughness by atomic-force microscopy," *Japanese J. Appl. Phys.*, Part 1, vol. 33, no. 1B, pp. 383-387, Jan. 1994.

[63] J. Wang, A. Rahman, A. Ghosh, G. Klimeck and M. Lundstrom, "Performance evaluation of ballistic silicon nanowire transistors with atomic-basis dispersion relations," *Applied Physics Letters*, vol. 86, pp. 093113.1-093113.3, Feb. 2005.

[64] J. Wang, A. Rahman, A. Ghosh, G. Klimeck and M. Lundstrom, "On the validity of the parabolic effective-mass approximation for the current-voltage calculation of silicon nanowire transistors," to appear in *IEEE Trans. Electron Dev.*, 2005.

[65] K. Natori, "Ballistic metal-oxide-semiconductor field effect transistor," *J. Appl. Phys.*, vol. 76, no. 8, pp. 4879-4890, Oct. 1994.

[66] S. Ramo, J. R. Whinnery and T. Van Duzer, *Fields and Waves in Communication Electronics*, 3rd edition, New York, NY: John Wiley & Sons Inc., 1994.

[67] M. Lundstrom and Z. Ren, "Essential physics of carrier transport in nanoscale MOSFETs," *IEEE Trans. Electron Dev.*, vol. 49, no. 1, pp. 133-141, Jan. 2002.

[68] www-dinma.univ.trieste.it/nirftc/research/easymesh, 2003.

[69] A. Rahman, M. Lundstrom and A. W. Ghosh, "Generalized effective-mass approach for n-type metal-oxide-semiconductor field-effect transistors on arbitrary oriented wafers," *J. Appl. Phys.*, vol. 97, pp. 053702.1-053702.12, Mar. 2005.

[70] J. Guo, *Carbon Nanotube Electronics: Modeling, Physics and Applications*, Ph. D. Thesis, Purdue University, West Lafayette, IN, 2004.

[71] F. Stern and W. E. Howard, "Properties of semiconductor surface inversion layers in the electric quantum limit," *Phys. Rev.*, vol. 163, no. 3-15, pp. 816-835, Nov. 1967.

[72] H. Wong, D. Frank and P. Solomon, "Device design considerations for double-gate, ground-plane, and single-gated ultra-thin SOI MOSFETs at the 25 nm channel length generation," *IEEE International Electron. Dev. Meeting (IEDM) Tech. Digest*, pp. 407-410, Dec. 1998.

[73] L. Chang, S. Tang, T. King, J. Bokor and C. Hu, "Gate length scaling and threshold voltage control of double-gate MOSFETs," *IEEE International Electron. Dev. Meeting (IEDM), Tech. Digest*, pp. 719-722, Dec. 2000.

[74] Semiconductor Industry Association, *International Technology Roadmap on Semiconductors*, 2001 edition, www.itrs.net, 2001.

[75] D. J. Frank, Y. Taur and H. -S. P. Wong, "Generalized scale length for two-dimensional effects in MOSFETs," *IEEE Electron Dev. Lett.*, vol. 19, no. 10, pp. 385-387, Oct. 1998.

[76] R. –H. Yan, A. Ourmazd and K. F. Lee, "Scaling the Si MOSFET: from bulk to SOI to bulk," *IEEE Trans. Electron. Dev.*, vol. 39, no. 7, pp. 1704-1710, Jul. 1992.

[77] K. Suzuki, T. Tanaka, Y. Tosaka, H. Horie and Y. Arimoto, "Scaling theory for double-gate SOI MOSFETs," *IEEE Trans. Electron. Dev.*, vol. 40, no. 12, pp. 2326-2329, Dec. 1993.

[78] J. –T. Park, J. –P. Colinge and C. H. Diaz, "Pi-gate SOI MOSFET," *IEEE Electron Dev. Lett.*, vol. 22, no. 8, pp. 405-406, Aug. 2001.

[79] G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists*, 5th edition, Burlington, MA: Harcourt/Academic Press, 2001.

[80] K. K. Young, "Short-channel effect in fully-depleted SOI MOSFETs," *IEEE Trans. Electron. Dev.*, vol. 36, no. 2, pp. 399-402, Feb. 1989.

[81] J. Knoch, B. Lengeler and J. Appenzeller, "Quantum simulations of an ultrashort channel single-gated n-MOSFET on SOI," *IEEE Trans. Electron. Dev.*, vol. 49, no. 7, pp. 1212-1218, Jul. 2002.

[82] R. Chau, S. Datta, M. Doczy, B. Doyle, B. Jin, J. Kavalieros, A. Majumdar, M. Metz and M. Radosavljevic, "Benchmarking nanotechnology for high-performance and low-power logic transistor applications," *IEEE Trans. Electron Dev.*, vol. 4, no. 2, pp. 153-158, Mar. 2005.

[83] F. Gamiz, J. B. Roldan, J. A. Lopez-Villanueva, P. Cartujo-Cassinello and J. E. Carceller, "Surface roughness at the Si-SiO2 interfaces in fully depleted silicon-on-insulator inversion layers," *J. Appl. Phys.*, vol. 86, no. 12, pp. 6854-6863, Dec. 1999.

[84] M. V. Fischetti, Z. Ren, P. M. Solomon, M. Yang and K. Rim, "Six-band k·p calculation of the hole mobility in silicon inversion layers: dependence on surface orientation, strain, and silicon thickness," *J. Appl. Phys.*, vol. 94, no. 2, pp. 1079-1095, Jul. 2003.

[85] C. –Y. Mou and T. –M. Hong, "Transport in quantum wells in the presence of interface roughness," *Phys. Rev. B*, vol. 61 no. 19, pp. 12612-12615, May 2000.

[86] K. Uchida, H. Watanabe, A. Kinoshita, J. Koga, T. Numata and S. Takagi, "Experimental study on carrier transport mechanism in ultrathin-body SOI n- and p-MOSFETs with SOI thickness less than 5nm," *IEEE International Electron. Dev. Meeting (IEDM) Tech. Digest*, pp. 47-50, Dec. 7-9 2002.

[87] H. Sakaki, T. Noda, K. Hirakawa, M. Tanaka and T. Matsusue, "Interface roughness scattering in GaAs/AlAs quantum-wells," *Appl. Phys. Lett.*, vol. 51, no. 23, pp. 1934-1936, Dec. 1987.

[88] A. Gold, "Electronic transport-properties of a two-dimensional electro-gas in a silicon quantum-well structure at low-temperature," *Phys. Rev. B*, vol. 35, no. 2, pp. 723-733, Jan. 1987.

[89] X. Zhao, C. Wei, L. Yang and M. Y. Chou, "Quantum confinement and electronic properties of silicon nanowires," *Phys. Rev. Lett.*, vol. 92, no. 23, pp. 236805.1-236805.4, Jun. 2004.

[90] S. Lee, F. Oyafuso, P. von Allmen and G. Klimeck, "Boundary conditions for the electronic structure of finite-extent embedded semiconductor nanostructures," *Phys. Rev. B*, vol. 69, pp. 045316.1-045316.8, Jan. 2004.

[91] A. Rahman, G. Klimeck and M. Lundstrom, "Bandstructure effects in ballistic nanoscale MOSFETs," *IEEE International Electron Dev. Meeting (IEDM), Tech. Digest*, pp. 139-142, Dec. 13-15 2004.

[92] H. Takeda and N. Mori, "Three-dimensional quantum transport simulation of ultra-small FinFETs," presented at the 10th International Workshop on Computational Electronics (IWCE-10), West Lafayette, IN, Oct. 25-27 2004.

[93] Y. J. Ko, M. Shin, S. Lee and K. W. Park, "Effects of atomic defects on coherent electron transmission in Si nanowires: full band calculations," *J. App. Phys.*, vol. 89, no. 1, pp. 374-379, Jan. 2001.

[94] Y. Zheng, C. Rivas, R. Lake, K. Alam, T. B. Boykin and G. Klimeck, "Electronic properties of silicon nanowires," *IEEE Trans. on Elec. Dev.*, vol. 52, no. 6, pp. 1097-1103, Jun. 2005.

[95] L. Reggiani, *Hot-Electron Transport in Semiconductors*, New York, NY: Springer-Verlag, 1985.

[96] A. Rahman, G. Klimeck, M. Lundstrom, N. Vagidov and T. B. Boykin, "Atomistic approach for nano-scale devices at the scaling limit and beyond - valley splitting in Si," *Japanese J. Appl. Phys.*, Part 1, vol. 44, no. 4B, pp. 2187-2190, Apr. 2005.

[97] D. Wang, Y. –L Chang, Q. Wang, J. Cao, D. B. Farmer, R. G. Gordon and H. Dai, "Surface chemistry and electrical properties of germanium nanowires," *J. Am. Chem. Soc.*, vol. 126, pp. 11602-11611, Aug. 2004.

[98] R. Kotlyar, B. Obradovic, P. Matagne, M. Stettler and M. D. Giles, "Assessment of room-temperature phonon-limited mobility in gated nanowires," *Appl. Phys. Lett.*, vol. 84, no. 25, pp. 5270-5272, Jun. 2004.

[99] A. Rahman and M. Lundstrom, "A compact scattering model for the nanoscale double-gate MOSFET," *IEEE Trans. Electron Dev.*, vol. 49, no. 3, pp. 481-489, Mar. 2002.

APPENDIX

APPENDIX
A SIMPLE MODEL TO TREAT SCATTERING

In Chapter 3, we developed a three-dimensional (3D), quantum mechanical simulator for ballistic silicon nanowire transistors (SNWTs) based on the non-equilibrium Green's function (NEGF) approach [37] [38]. In this appendix, we add a dissipative transport model, so called the 'Büttiker probes' [42] [57], into this simulator. The Büttiker probes model is a simple, phenomenological approach to treat scattering within the NEGF framework and it was previously adopted in MOSFET simulations [42]. The simulation results in this appendix will show that this simple model captures the essential effects of scattering on both internal device parameters (e.g., charge distribution and electrostatic potential) and current-voltage characteristics. It should also be noted that the Büttiker probes model is only a phenomenological method, equivalent to the drift-diffusion model used in the semi-classical context. [36] So it is incapable of treating detailed scattering mechanisms in SNWTs.

1) Theory:

The simple treatment of scattering with the Büttiker probes has been adopted by Venugopal and coworkers [42] for the simulation of nanoscale MOSFETs. Due to the similarity between the transport calculations of a MOSFET and a SNWT, here we will follow the basic concepts and formalism of the method described in [42] while making necessary modifications and corrections for the case of SNWT simulation.

In the ballistic regime, as we know, electrons move through the device coherently, with their energies and phase information conserved. When scattering is present, however, electrons' momenta and energies could be altered and their phase information

may be lost. Based on this observation, virtual probes (Büttiker probes) are attached to the device lattice (in the channel direction), which serve as reservoirs that absorb electrons from the active device, modulate their momenta and/or energies, and then re-inject them back to the device. The difference between the probes and the S/D contacts is that the probes can only change the electron momentum/energy and not the number of electrons within the active device [42].

Figure A.1 shows the one-dimensional (1D) device lattice (in the channel direction) for a SNWT with the Büttiker probes attached. Each probe is treated as a virtual 1D lattice (in the x' direction) that is coupled to a node in the device lattice. The coupling energy, Δ_m^i, between this virtual lattice and the node it is attached to is called the Büttiker probe strength [42], which is determined by the ballisticity of the device. For instance, when Δ_m^i is zero, there is no coupling between the device and the probes, so the electrons can travel through the device ballistically. If this coupling energy is large, it means that the electrons in the active device region can easily scatter into the probes, which implies that the scatting in the device is strong. As we will show later, the Büttiker probe strength can be analytically related to the electron mean-free-path [36], which allows us to calibrate the parameters in our simulation to mimic a low-field mobility that can be measured experimentally [42]. It should also be noted that since we treat each probe as a reservoir, a Fermi level (μ_i, $i = 2,...,N_X - 1$) needs to be assigned to the probe, and the values of these probe Fermi levels have to be adjusted to achieve current continuity (i.e., the net current at each probe is zero). The mathematical formalism used to treatment this physical structure is described in the following paragraphs. (Since this is an appendix to Chapter 3, all the variables used here have the same definitions as those in Chapter 3.)

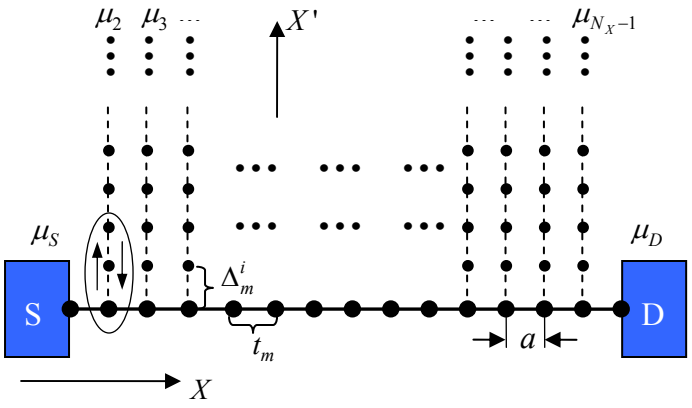


Fig. A.1 A generic plot of the 1D device lattice (solid line with dots, along the X direction) with the Büttiker probes attached. Each probe is treated as a virtual 1D lattice (dashed line with dots, along the X' direction) coupled to a node in the device lattice. The coupling energy between this virtual lattice and the node it is attached to is Δ_m^i, and that between two adjacent device lattice nodes is t_m. The probe Fermi levels are labeled as μ_i ($i = 2,3,...,N_X - 1$).

As we show in Sec. 3.2, the retarded Green's function for mode m is obtained as

$$G^m(E) = \left[ES^m - h_{mm} - \Sigma_S^m(E) - \Sigma_1^m(E) - \Sigma_2^m(E) \right]^{-1}. \tag{3.31}$$

If we discretize the matrices by the finite difference method (FDM) method, S^m is a $N_X \times N_X$ identity matrix and the device Hamiltonian h_{mm} is expressed as

$$h_{mm} = \begin{bmatrix} 2t_m + E_{sub}^m(0) & -t_m & 0 & \cdots & \cdots & 0 \\ -t_m & 2t_m + E_{sub}^m(a) & -t_m & \ddots & \ddots & \vdots \\ 0 & -t_m & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -t_m & 0 \\ \vdots & \ddots & \ddots & -t_m & 2t_m + E_{sub}^m\left[(N_X-2)a\right] & -t_m \\ 0 & \cdots & \cdots & 0 & -t_m & 2t_m + E_{sub}^m\left[(N_X-1)a\right] \end{bmatrix}$$

(FDM), (A.1)

where the coupling energy between adjacent lattice nodes (in the x direction) is $t_m = \left(\hbar^2/2a^2\right)\overline{a_{mm}}$ and $\overline{a_{mm}}$ is defined in Eq. (3.28a). In the *ballistic* limit, the scattering self-energy $\Sigma_S^m = 0$ so the total self-energy matrix is

$$\Sigma^m = \Sigma_S^m + \Sigma_1^m + \Sigma_2^m = \begin{bmatrix} -t_m e^{ik_{m,1}a} & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & \cdots & 0 & -t_m e^{ik_{m,N_X}a} \end{bmatrix} \text{ (FDM)}, \qquad \text{(A.2)}$$

where $k_{m,1}$ (k_{m,N_X}) is determined by $E = E_{sub}^m(0) + 2t_m\left(1-\cos k_{m,1}a\right)$ ($E = E_{sub}^m\left[\left(N_X-1\right)a\right] + 2t_m\left(1-\cos k_{m,N_X}a\right)$). After we attach the Büttiker probes to the device lattice (Fig. A.1), the device Hamiltonian h_{mm} becomes

$$h_{mm} = \begin{bmatrix} 2t_m + E_{sub}^m(0) & -t_m & 0 & \cdots & \cdots & 0 \\ -t_m & 2t_m + \Delta_m^2 + E_{sub}^m(a) & -t_m & \ddots & \ddots & \vdots \\ 0 & -t_m & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -t_m & 0 \\ \vdots & \ddots & \ddots & -t_m & 2t_m + \Delta_m^{N_X-1} + E_{sub}^m\left[\left(N_X-2\right)a\right] & -t_m \\ 0 & \cdots & \cdots & 0 & -t_m & 2t_m + E_{sub}^m\left[\left(N_X-1\right)a\right] \end{bmatrix}, \qquad \text{(A.3)}$$

and the total self-energy matrix turns to

$$\Sigma^m = \Sigma_S^m + \Sigma_1^m + \Sigma_2^m = \begin{bmatrix} -t_m e^{ik_{m,1}a} & 0 & \cdots & \cdots & 0 \\ 0 & -\Delta_m^2 e^{ik_{m,2}a} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & -\Delta_m^{N_X-1} e^{ik_{m,N_X-1}a} & 0 \\ 0 & \cdots & \cdots & 0 & -t_m e^{ik_{m,N_X}a} \end{bmatrix}, \qquad \text{(A.4)}$$

where $k_{m,i}$ (i=1,2,…,N_X) is determined by $E = E_{sub}^m\left[\left(i-1\right)a\right] + 2t_m\left(1-\cos k_{m,i}a\right)$, and Δ_m^i (i=2,3,…,N_X-*1*) is the Büttiker probe strength. For convenience, we prefer to keep the device Hamiltonian h_{mm} in its original form, as in Eq. (A.1), so we move the terms containing Δ_m^i in the diagonal elements of h_{mm} to the total self-energy matrix Σ^m. Thus,

$$\Sigma^m = \Sigma_S^m + \Sigma_1^m + \Sigma_2^m = \begin{bmatrix} -t_m e^{ik_{m,1}a} & 0 & \cdots & \cdots & 0 \\ 0 & -\Delta_m^2\left(e^{ik_{m,2}a}-1\right) & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & -\Delta_m^{N_X-1}\left(e^{ik_{m,N_X-1}a}-1\right) & 0 \\ 0 & \cdots & \cdots & 0 & -t_m e^{ik_{m,N_X}a} \end{bmatrix}. \tag{A.5}$$

Inserting Eqs. (A.1) and (A.5) into Eq. (3.31), the retarded Green's function G^m can be evaluated.

Knowing G^m, the state spectral function due to injection from the S/D and all probes for mode *m* is obtained as [42],

$$A_i^m(E) = G^m(E)\Gamma_i^m(E)G^{m\dagger}(E), \tag{A.6}$$

where *i* runs over all the reservoirs (including the S/D) and Γ_i^m is an $N_X \times N_X$ matrix defined as

$$\Gamma_i^m[p,q] = j\left[\Sigma^m[p,q] - \Sigma^{m\dagger}[p,q]\right]\delta_{p,i}\delta_{q,i}, \ (p,q=1,2,\ldots,N_X). \tag{A.7}$$

The local density of states (LDOS) due to injection from reservoir *i* is then obtained as

$$D_i^m[p] = \frac{1}{\pi a} A_i^m[p,p], \ (i=1,2,\ldots,N_X,\ p=1,2,\ldots,N_X), \tag{A.8}$$

and the 1D electron density (in m^{-1}) for mode *m* can be calculated by

$$n_{1D}^m = \sum_i \int_{-\infty}^{+\infty} D_i^m f(\mu_i, E)\,dE, \tag{A.9}$$

where *i* is the reservoir index that runs over all the probes and the S/D, and μ_i is the Fermi level for reservoir *i* (note that $\mu_1 = \mu_S$ and $\mu_{N_X} = \mu_D$).

The transmission coefficient between any two reservoirs *i* and *r* can be evaluated as

$$T_{i\leftrightarrow r}^m(E) = trace\left[\Gamma_i^m(E)G^m(E)\Gamma_r^m(E)G^{m\dagger}(E)\right]. \tag{A.10}$$

The net current density (at energy *E*) at reservoir *i* including contributions from all reservoirs (labeled by *r*), modes (labeled by *m*) and valleys is

$$\eta_i(E)=\frac{q}{\pi\hbar}\sum_m\sum_r T^m_{i\leftrightarrow r}(E)\left[f(\mu_i,E)-f(\mu_r,E)\right], \tag{A.11}$$

and the net current at reservoir *i* is

$$I_i=\int_{-\infty}^{+\infty}\eta_i(E)dE. \tag{A.12}$$

As mentioned in [42], while the S/D Fermi levels are determined by the applied voltages, the Fermi levels of the probes have to be adjusted to ensure current continuity, which implies that the net current at each probe must be zero,

$$I_i=\int_{-\infty}^{+\infty}\eta_i(E)dE=0,\ (i=2,3,...,N_X-1). \tag{A.13}$$

Inserting Eq. (A.11) into (A.13), we obtain

$$\frac{q}{\pi\hbar}\sum_m\sum_r\int_{-\infty}^{+\infty}T^m_{i\leftrightarrow r}(E)\left[f(\mu_i,E)-f(\mu_r,E)\right]dE=0,\ (i=2,3,...,N_X-1). \tag{A.14}$$

Solving this nonlinear equation group (A.14) by Newton's method [42], the Fermi levels ($\mu_i, i=2,3,...,N_X-1$) of all the probes are evaluated. It should be mentioned that if we implement the elastic Büttiker probes, which can only change the electron momentum and not the energy, to capture elastic scattering mechanisms in SNWTs (e.g., surface roughness scattering and ionized impurity scattering [36]), the net current for each probes has to be zero at any energy, so

$$\eta_i(E)=\frac{q}{\pi\hbar}\sum_m\sum_r T^m_{i\leftrightarrow r}(E)\left[f(\mu_i,E)-f(\mu_r,E)\right]=0,\ (i=2,3,...,N_X-1). \tag{A.15}$$

It implies that the probe Fermi levels are both position and energy dependent. In this case, the Fermi levels of probes at each energy can be computed by solving the linear equation group (A.15). Knowing the probe Fermi levels (by solving either Eq. (A.14) or Eq. (A.15)), the electron density and terminal current can be calculated from Eqs. (A.9) and (A.12).

Finally, we list the equations that relate the Büttiker probe strength, Δ^i_m, to the classical low-field electron mobility μ_0. Following the procedures in [42], for a single-mode 1D conductor with a uniform potential, we can obtain

$$\frac{\Delta^i_m}{t_m}=\frac{2a}{\lambda}, \tag{A.16}$$

where λ is the electron mean-free-path, which relates to the low-field electron mobility by the following equation for a 1D conductor (the $\lambda \sim \mu_0$ relation for a 2D conductor is described in [99]),

$$\lambda = \left(\frac{2\mu_0}{\upsilon_T} \frac{k_B T}{q} \right) \cdot \frac{[\Im_{-1/2}(\eta_F^i)]^2}{\Im_{-3/2}(\eta_F^i)\Im_0(\eta_F^i)}, \tag{A.17}$$

where $\upsilon_T = \sqrt{2k_B T / \pi m_x^*}$ is the uni-directional thermal velocity of non-degenerate electrons [36] [67]. The function, $\Im_n(x)$, is the Fermi-Dirac integral [36], and η_F^i is defined as $\eta_F^i = [\mu_i - E_{sub}^m(x_i)]/k_B T$, where x_i is the position of the *ith* reservoir (probe) of the device. It should be noted that the mean-free-path λ defined in Eq. (A.17) is position-dependent and consequently the Büttiker probe strength, Δ_m^i, is also position-dependent. As mentioned earlier, single-mode occupancy is assumed in our analysis. If more than one mode is occupied, the mean-free-path should be treated as an average mean-free-path over all the modes and valleys. (Please refer to Appendix B in [42] for details.)

2) Results:

Figure A.2 plots the LDOS together with the electron subbands for a dissipative cylindrical SNWT with a 10nm gate length and a 3nm Si body thickness. We assume that both elastic (e.g., surface roughness scattering and ionized impurity scattering) and inelastic (e.g., electron-phonon interactions) scattering mechanisms are present in the device (i.e., Eq. (A.14) is used for current continuity), and the equivalent mobility is $55\text{cm}^2/(\text{V}\cdot\text{s})$ at the S/D extension regions and $200\text{cm}^2/(\text{V}\cdot\text{s})$ in the channel. Compared with the ballistic case (Fig. 3.5), strong oscillations in the LDOS, which is due to quantum interference, are washed out. It is because scattering inside the SNWT randomizes the phase of the electrons and consequently destroys the quantum coherence in the device [37] [42]. Moreover, the slope of the electron subbands in the S/D extension regions manifests the S/D series resistances at the ON-state, which is caused by

the strong scattering (i.e., the S/D mobility is only $55\text{cm}^2/(\text{V}\cdot\text{s})$) at the heavily doped S/D regions. In Fig. A.3, we compare the I_{DS} vs. V_{GS} characteristics for this dissipative cylindrical SNWT (solid) with its ballistic limit (dashed). It is evidently shown that scattering lowers both OFF and ON currents. For the mobility values we use, the ON-current of the dissipative SNWT approaches ~70% of the ballistic limit.

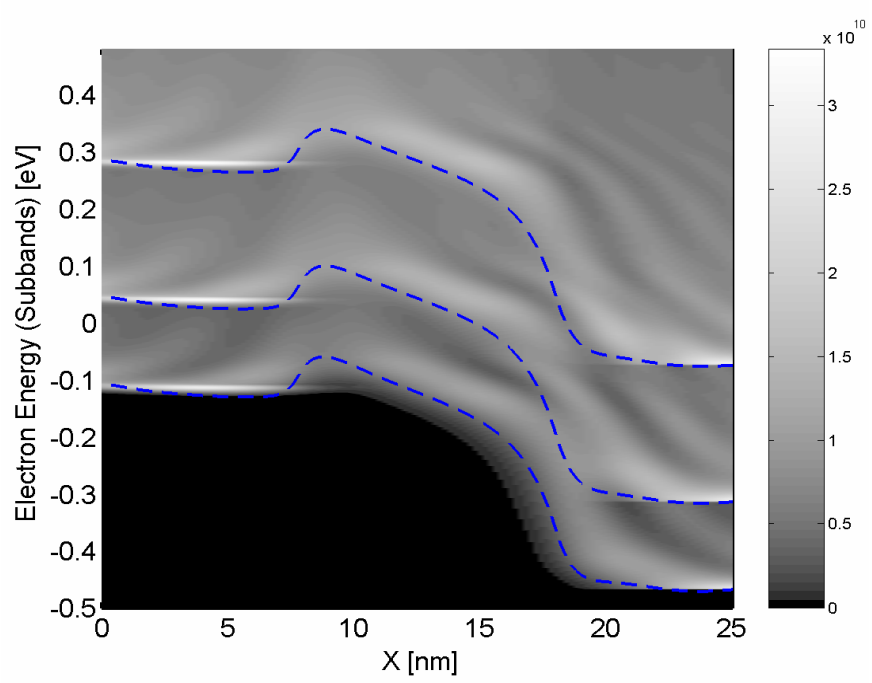


Fig. A.2 The computed LDOS, in $1/(\text{eV}\cdot\text{m})$, and electron subbands (dashed lines) of a dissipative cylindrical SNWT with a 10nm gate length and a 3nm Si body thickness (the details of the device geometry are described in the Fig. 3.3 caption). (V_{GS}=0.4V and V_{DS}=0.4V). The S/D mobility is $55\text{cm}^2/(\text{V}\cdot\text{s})$ and the channel mobility $200\text{cm}^2/(\text{V}\cdot\text{s})$.

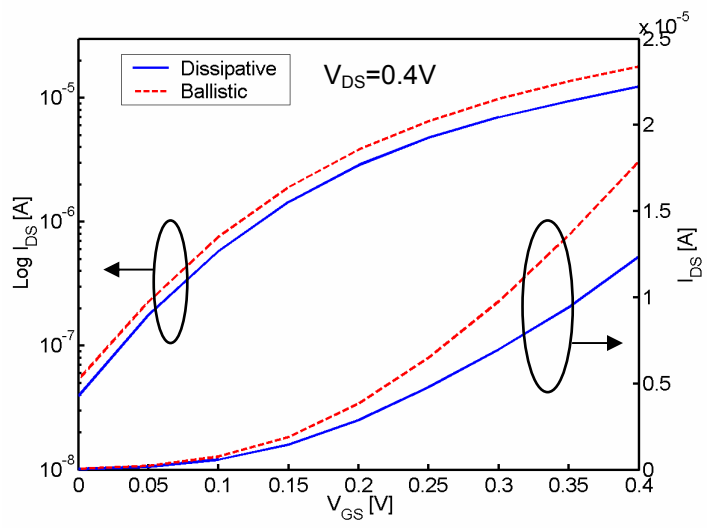


Fig. A.3 The I_{DS} vs. V_{GS} curves for a cylindrical SNWT with a 10nm gate length and a 3nm Si body thickness (the details of the device geometry are described in the Fig. 3.3 caption) in logarithm (left) and linear (right) scales (V_{DS}=0.4V). The dashed lines are for the ballistic limit while the solid lines are for the case with scattering (i.e., the S/D mobility is $55\text{cm}^2/(\text{V}\cdot\text{s})$ and the channel mobility is $200\text{cm}^2/(\text{V}\cdot\text{s})$).

The above results clearly indicate that the simple quantum treatment of scattering with the Büttiker probes captures the effects of scattering on both internal characteristics and terminal currents for SNWTs. The relation between the Büttiker probe strength, the only input parameter in this model, with the experimentally measurable low-field mobility enables this simple model to be used in engineering simulation and design. It should also be noted, however, that this phenomenological model is only a macroscopic description of scattering, which is similar as the drift-diffusion model that is used in the semiclassical context. [36] To quantum mechanically treat various scattering mechanisms in detail, a rigorous treatment of scattering within the NEGF formalism [37] is still needed.

VITA

VITA

Jing Wang was born in Shangqiu, Henan, China, in February, 1979. He received the Bachelor of Engineering degree in electronic engineering from Tsinghua University, Beijing, China, in 2001. In August, 2001, he became a direct Ph.D. student in the School of Electrical and Computer Engineering at Purdue University, West Lafayette, Indiana. After graduating from Purdue University, Jing Wang will be working at the IBM Semiconductor Research and Development Center (SRDC), Hopewell Junction, New York, as a Device Modeling Engineer.