QUANTUM AND ATOMISTIC EFFECTS IN NANOELECTRONIC TRANSPORT DEVICES

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Neophytos Neophytou

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2008

Purdue University

West Lafayette, Indiana

To my family ....

ACKNOWLEGEMENTS

I am thankful that during my PhD years I was given the opportunity to collaborate and learn from many people from various backgrounds. This for me was a remarkable experience and fun. I would like to express my deep gratitude and appreciation to my advisor, Prof. Mark Lundstrom, who supervised my work. Prof. Lundstrom's vision and approach towards research, directed towards "doing what is important and will provide help to the scientific community," helped me understand how to acquire and use my problem solving skills and attitude. His guidance, through strict professionalism as well as care towards his students, is really admirable, and constitutes true mentoring lessons, beyong the scientific aspect.

I would also like to express my great thankfulness to my co-advisor, Prof. Gerhard Klimeck for his guidance and all the knowledge I gained from him. The knowledge about atomistic simulations as well as about many other new approaches and areas of research that he provided, showed me how broad the science field is. His attitude and persistence towards "getting the project done, and as good as it can get," as well as his care for his students, is another big plus for me all these years.

I would also like to thank Prof. Ashraful Alam and Prof. Ahmed Sameh, for serving on my advisory committee.

Special thanks goes to Dr. Diego Kienle for his endless kindness and patience towards explaining various physics related issues, and providing guidance especially in the early period of my work. His attitude towards research and learning, and his persistence is a great lesson that I got from him.

I would like to thank Prof. Eric Polizzi for his help in the development of the three-dimensional atomistic simulator for nanotube transistors (CNTFET). I would also like to thank Prof. Shaikh Ahmed for his patience towards the work on the CNTFET simulator deployment. The discussions with him were so beneficial to me in academic and non-academic areas.

Thanks go to Dr. Titash Rakshit, Dr. Mark Stettler and Martin Giles for providing me the opportunity to intern at INTEL Corporation twice and initiate my work on III-V HEMTs.

I would like to thank Prof. Supriyo Datta for all the learning I got about the NEGF formalism through his class, Prof. M.P. Anantram for the collaboration in the CNT defect work, and Prof. Gengchiau Liang and Dr. Dmitri Nikonov for the collaboration on the graphene nano-ribbon (GNR) work.

I would like to thank my colleagues and labmates for their friendship and fruitful discussions all these years. Thanks goes to Prof. Jing Guo for the initial MATLAB carbon naotube scripts, Dr. Jing Wang, Dr. Anisur Rahman and Dr. Mathieu Luisier for the initial tight-binding scripts. I would also like to thank Dr. Sayed Hassan, Siyuranga Koswatta, Yang Liu, Kurtis Cantley, Himadri Pal, Raseong Kim, Yunfei Gao, Changwook Jeong, Tony Low, Nick Vargo, Abhijeet Paul, Lutfe Siddiqui, Sayeef Salahuddin, Ninad Pimparkar, Bhaskaran Muralidharan, Neerav Kharche, and all the others in EE350. I feel lucky and grateful for meeting these people from whom I learnt so many things. Thanks to Cheryl Haines for her help on travel/meeting arrangements.

Thanks goes to the NSF funded Network for Computational Nanotechnology (NCN) center for the computational resources.

Finally, I would like to thank my family who stood by me during the whole period of my studies. Special thanks to *Στέλλα* for her support, patience, and understanding through my entire graduate studies.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

ABSTRACT

Neophytos Neophytou Ph.D., Purdue University, December 2008. Quantum and Atomistic Effects in Nanoelectronic Transport Devices. Major Professors: Mark S. Lundstrom, Gerhard Klimeck.

As devices scale towards atomistic sizes, researches in silicon electronic device technology are investigating alternative structures and materials. As predicted by the International Roadmap for Semiconductors, (ITRS), structures will evolve from planar devices into devices that include 3D features, strong channel confinement, strain engineering, and gate all around placement for better electrostatic control on the channel. Alternative channel materials such as carbon nanotubes (CNT), nanowires (NW) and III-V based channel materials are considered to be possible candidates for future device technology nodes because of their potentially superior to silicon transport properties. For nanoscale dimensions, and under the operating conditions mentioned above, both atomistic and quantum effects become important in determining the electronic structure and transport properties of the devices. Detailed modeling and simulation that capture these new physics will be essential in providing understanding and guidance to the device operation and optimization. We have used the non-equilibrium Green's function (NEGF) formalism for quantum transport simulations and real space atomistic tight-binding techniques (p_z, sp3d5s*-SO) to investigate transport properties in CNT, NW and III-V HEMT field-effect transistors. Specifically, we have investigated the effect of atomistic defects such as atomic vacancies, and charged impurities in 1D CNT, and dangling bonds in NW channels. It was found that the presence of single defects, severely degrades the transport performance of 1D channels. We have further investigated the effect of physical quantization on the electronic structure of NW field-effect transistors and identified the main electronic structure factors that influence their performance. It was found that structural and quantization below 10nm can severely affect the electronic properties of

NW channels by changing the effective masses and altering degeneracies through valley splitting. Different wire orientations have different transport properties. The [110] and secondly [100] oriented nanowires are found to perform better than the [111] wires in terms of ON-current capabilities for n-type wires, whereas the [111] and [110] significantly outperform the [100] wires in the case of p-type nanowires. Explanations for this behavior can be extracted from the non-parabolicity and anisotropy of the Si 3D bulk *E(k)*. Finally, we present an analysis of recent experimental data for III-V HEMT devices using the NEGF formalism and address several issues related to the operation of HEMT devices. Interestingly, a 60nm HEMT device can be though to first order as a ballistic channel connected to two series resistances.

1. INTRODUCTION

1.1 Background

The need for higher device performance that is driving the silicon electronics industry has necessitated the scaling of device sizes to submicron dimensions over the last decade – a trend embodied in Moore's law. As devices scale further towards atomistic sizes, existing CMOS devices, suffer from severe short channel effects which lead to device performance degradation, variations and increased power dissipation. These are major issues that CMOS technology faces, and drawbacks to further scaling. To gain more control over these critical issues, while still retaining the path of advancing performance, CMOS devices as predicted by the International Roadmap for Semiconductors, (ITRS) [1], will evolve from planar devices into devices that include strong strain engineering for carrier velocity enhancement and 3D features and gate all around placement for better electrostatic control on the channel. High mobility materials, that can operate at lower power supply, such as III-V compounds and carbon nanotubes are also been considered as alternative channel materials.

Several device structures and geometries are being investigated for replacements for the planar Si transistor, both theoretically and experimentally. Planar structures such as ultra-thin-body devices and III-V HEMT devices [2-6], as well as multi-gate FinFET type geometries [7-9], gate all around nanowire devices [10-15], and carbon nanotube FETs [16-21] are all potential candidates (Fig. 1.1). To predict the potential benefit of these devices over conventional Si MOSFETs, and further optimize their operation, however, deep understanding is necessary, and modeling and simulation can offer tremendous insight. Short channel effects must be accurately captured through proper electrostatic treatment, many times in three-dimensions. Quantum mechanical effects will dominate transport, and need to be treated through a full quantum mechanical model.

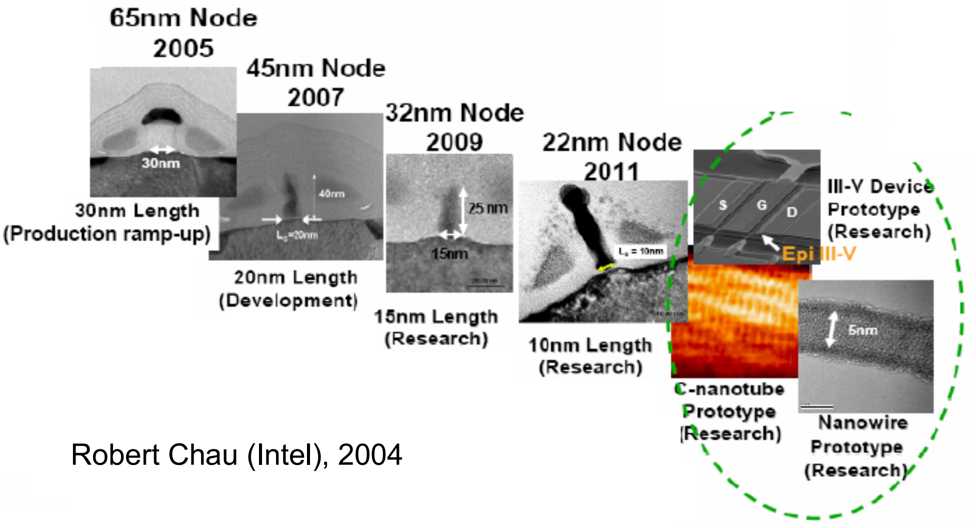


Fig. 1. 1. The development trend and future projections of CMOS devices and structures. Adopted from Robert Chau (INTEL Corp.[22]).

Suggestions have been made on that what will determine the ultimate scaling of devices will be the OFF-state source to drain tunneling, a purely quantum mechanical effects [23]. On top of that, the ability to predict the accurate electronic structure of the device, which might consist of material different from Si, oriented in arbitrary directions and quantized in various cross sectional shapes and directions and influenced from strong strain fileds will be another determining factor for evaluating the performance of a certain device.

1.1.1 3D device geometry

The 3D structure geometry is essential in the scaling of the device, since it provides the necessary gate control on the channel, which is weakened in the case of ultra scaled planar devices due to short channel effects. The electrostatics of the device needs to be optimized and carefully studied, since it ought to be the major factor that controls the device behavior. Potential variations in the cross section of the device as well as along the channel will also have a large effect on the device properties. A proper 3D structure description therefore, will be a main component needed in device modeling.

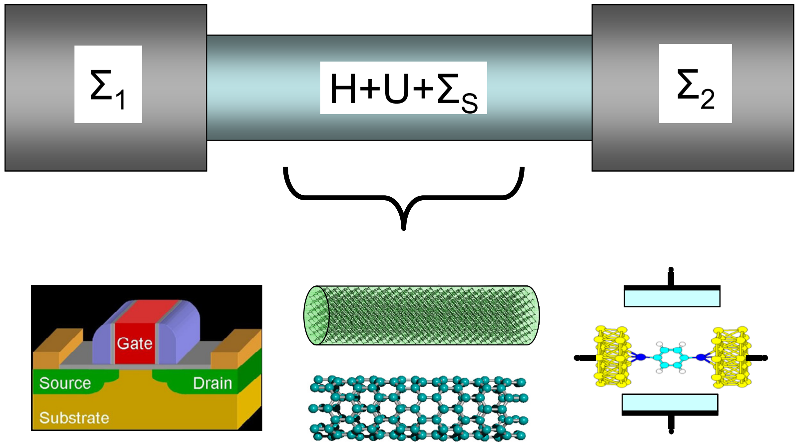


Fig. 1. 2. Applications of the NEGF formalism: The Non-Equilibrium-Greens' Function (NEGF) formalism has been extensively used in transport modeling in a variety of nanoscale devices, MOSFETs, nanowires, nanotubes and molecular electronics.

1.1.2 Quantum mechanical effects

Extensive simulation tools based on the non-Equilibrium Greens' Function (NEGF) formalism have been developed in order to study quantum effects in MOSFET devices [24]. NEGF has been demonstrated to correctly account and explain the quantum behavior of a variety of nanoscale devices such as 2D MOSFETs [25], 1D nanowires [26-28], carbon nanotubes [29-32], and 2D graphene based transistors [33] (Fig. 1.2). A variety of quantum mechanical effects can be captured by this treatment, such as quantization in density of states and transmission, tunneling, quantum mechanical reflections and resonance states, edge states (as in the case of graphene transistors) and metal induced gap states (MIGS), (as in the case Schottky Barrier contacts) [29, 33]. These capabilities have established NEGF as a powerful (and often necessary) tool in modeling quantum effects in nanostructures. Mode space and real space techniques are being used in device modeling within the NEGF framework. Mode space approaches provide certain parts of the necessary information, however, they fail in situations in which a lot of modes are mixing, as in situations with large potential variation in the devices, or near the surfaces where edge states and MIGS accumulate. In some of these cases, the coupled mode approaches might provide more accuracy, in others they do not. Compromises between accuracy and computation efficiency have to be made in some cases.

1.1.3 Atomistic modeling

Necessity of atomistic modeling: The effective mass approximation (EMA) has been extensively used in evaluating the performance of traditional MOSFET devices. New physics need to be included in treatment of the next stage of CMOS devices, however, which will make the EMA inadequate. The problem of identifying the correct bandstructure and effective masses of nanowires has been addressed by various authors in references [26, 34-36] with qualitative agreement on the main features of the electronic structures. Other sophisticated atomistic techniques for electronic structure calculation also mention mass variations in nanostructures from their bulk values, which results in different threshold voltages and ON-current densities [34, 37]. By adjusting the effective masses to map masses extracted from atomistic calculations, however, the EMA can still be used [34, 38] for the conduction band only, for specific wafer orientations only, and for specific structures and structure sizes only. In general, however, this method is not always valid, and atomistic simulations are more appropriate for nanoscale devices.

More accurate atomistic models that inherently include all relevant information about the electronic structure of materials and structures, therefore, should be utilized. Atomistic models can provide information about non-parabolicity, confinement level position beyond the effective mass approximation (for both structural and electrostatic confinement), the effects of strain in the electronic structure (uniform and non-uniform), as well as a more accurate distribution of charge in the device channel. In addition to that, atomistic description of the device in arbitrary orientations, has the advantage of being able to automatically capture the valley projections and extract the dispersions of the channels in the transport orientation. This model also automatically includes information about band coupling and mass variations as functions of quantization.

Motivation for an empirical tight-binding (TB) model: The choice of the atomistic model to be used depends on the problem of interest. At the nanometer scale the concept of a “new device” and a “new material” are blurred. Quantum mechanics of the electronic structure, crystal symmetry, atomic composition and spatial disorder are important. A certain electronic structure model needs to satisfy several requirements to accurately capture nanoscale device physics, in which the traditional effective mass approximation (EMA) will fail to capture. The finite extent of the devices rather than the infinite periodic nature speaks for the choice of a local basis set rather than a plane wave basis set. The stability of the bands in typical semiconductor devices speaks for a reduced model that takes the existence of bands for granted. The need to model complicated man-made heterostructures speaks for a nearest neighbor model to eliminate ambiguities of long-range coupling elements. The need to simulate large extensive structures containing tenths of millions of atoms [39], requires a reduced order model. The need to accurately model bandgaps (within a few meV) and masses (within a few %) speaks for an empirical bandstructure model rather than an ab-initio model. All these requirements have led to the choice of empirical TB in this work.

The $sp^3d^5s^*$ *TB model:* The basis set of the $sp^3d^5s^*$ nearest neighbor TB model used in this work, is composed of orthogonal localized orbitals. This type of basis makes it very attractive for accurate electronic structure of truncated nanostructures of finite sizes and composition variations on the nm-scale. It is a very convenient method to treat material and potential variations as well as strain fields at the nanoscale. The parameterization was performed using a genetic algorithm in [40, 41], and the parameters extracted can reproduce the band edges of the bulk silicon bandstructure over the entire Brillouin zone. The model is described in detail in references [39, 41-45]. The energy bands obtained for nanostructures, as well as in the bulk case in energy regions away from the bulk minima, are in good quantitative agreement with other theoretical calculations using pseudopotential and ab-initio GW methods [46].

Challenges and limitations: At the 22nm node (year 2011), the device channel length is predicted to be 10nm, which is only about 60 atomic layers in length. In such mesoscopic dimensions, the essence of electronic structure of devices, based mainly on effective mass approximation is definitely questionable. Atomistic effects, interface details and surface roughness, discrete dopant and impurity fluctuations will dramatically influence the device properties. The statistical nature of imperfections is lost, and the devices cannot be characterized by a collective behavior. Each device will be different from the rest, and variations due to imperfections will be a major drawback in the controlling their operation. An understanding of the impact of atomistic effects, therefore, will be of importance. In doing so, atomistic modeling provides this information inherently, since the devices can be modeled atom by atom and all imperfections can be are included. Direct modeling of quantum transport in 3D devices of realistic sizes using atomistic basis sets and quantum transport in the NEGF formalism is, however, prohibitive due to the extensive computational burden. This can be succeeded by using fast and parallel numerical techniques, with the involvement of computer scientists. A large part of the necessary information however, can be extracted relatively fast, from bandstructure simulations of nanostructures (cross section of infinite length devices). Relevant information can be extracted and then be used in the 3D NEGF simulation scheme in less computationally expensive approaches (i.e modifications in effective mass, or mode space approaches).

1.2 Summary and contributions

In this thesis, we address general issues that will impact the performance of the possible candidate structures and devices for silicon MOSFET alternatives. The three device families addressed are CNTs, NWs, and III-V material devices. These are all ultra scaled, low dimensional devices that will operate in the quantum capacitance regime, and possibly very close to the ballistic limit.

We investigate the sensitivity of these low dimensional device structures to defects and process variations and identify the important factors that will influence their performance. The key findings are that the performance of 1D channels is very sensitive to atomistic defects and can suffer from the presence of just a single defect or imperfection in the channel. In order to develop this general conclusion, simulations on a variety of channels with a variety of imperfections have been performed.

We also investigate the sensitivity of the electronic structure of scaled channels as a function of quantization, channel orientation and potential variations in the structure. Examination of these effects in several device families allowed us to identify which of the electronic structure features affect the performance of the devices and to what degree. We also identified at what extent the effect of variations in these features impact the device performance. What has emerged as a general conclusion out of this work is the importance of the semiconductor capacitance for devices that operate in the quantum capacitance regime, which degrades the total gate capacitance by a factor of about two. Variations in the capacitances of the devices, however, (oxide capacitance, quantum capacitance) do not appear as variations of similar order in the performance of the device. What can affect the performance of different channels is the transport effective mass. For short channel devices, light mass can increase the carrier velocity and improve performance, however, it can also enhance tunneling which slows carriers down. For this reason, the carrier velocity is lower than can be expected from bandstructure calculations. This dissertation, finally, other than the channel, addresses the importance of the contact resistance and the design of the source in low dimensional / low density of states devices.

From modeling point of view, the thesis provides descriptions of comprehensive models to analyze the effect of electronic structure in nanowires using an atomistic Hamiltonian self consistently coupled to a Poisson solver. It also provides detailed descriptions of how to build generic models for simulators utilizing atomistic treatment of nanodevices, 3D electrostatics and quantum transport. Most of these details are presented in the Appendices.

1.3 Outline of the dissertation

In chapter 2 a real-space atomistic 3D quantum (NEGF) simulator for experimentally realized CNT devices is used to investigate the impact of single atomistic defects in the 1D CNT channels. The chapter is based on references [31, 32, 47]. The effect of a variety of defects such as vacancies, and both positively and negatively charged impurities on the CNT channel performance is described. In addition to semiconducting CNTs, the effect of vacancy defects on metallic CNTs, as well as the effect of dangling bonds on the performance of nanowire devices is investigated. The results presented, therefore, provide a general understanding of the effect of defects in 1D systems, independent of the nature of the defects or the channel. The final conclusion of this study is that single atomistic defects have a large impact on the performance of 1D nanoscale devices.

Chapter 3 describes the self-consistent ballistic model that is used for examining the transport properties of nanowire transistors based on their electronic structure properties. The chapter is based on references [38, 48, 49]. In the model, the nanowire channel is described atomistically, in various transport and quantization directions. A semiclassical ballistic model is used for transport calculations. Atomisitc modeling is essential, since as it is explained in the chapter, the simple effective mass treatment can only work for some transport cases for the conduction bands of nanowires, only after the effective masses are adjusted to the TB masses. The new contribution to the semiclassical model is to include the potential variations in the cross section of the device through through a 2D Poisson solution for the solution of the electrostatics. The key finding is that consideration of potential variations in the cross section of the device, as well as the exact shape of the charge distribution is very important, for devices of sizes even down to 3nm, and can impact the electronic structure, the semiconductor capacitance and the device performance at large.

Chapter 4 describes explicitly the bandstructure effects on the transport properties of NMOS nanowires oriented in different transport orientations and it is based on reference [49]. It is shown that atomistic effects are dominant in the transport of ultra scaled nanowires. The effective masses and carrier velocities of nanowires, other than depending on the wire orientation, are also strong functions of quantization and change differently in different wire orientations. Explanation for this behavior is provided through features in the first Brillouin zone of Si. Band coupling effects such as valley splitting, lift the degeneracies of nanowires, and are also of different magnitude in different orientations. In addition, it is shown that the semiconductor capacitance in nanowires is important, and can degrade the total gate capacitance by ~30%. This is effect appears to be very similar to nanowires in all transport orientations. The final important conclusion of this chapter is that when comparing the relative performance of different nanowires, the carrier velocity is the dominant deciding factor.

Chapter 5 focuses on the ballistic transport of PMOS nanowires and is based on [50]. The valence band of semiconductors in general is much more complicated to describe compared to its conduction band counterpart because of the enhances spin-orbit coupling, the strong anisotropy and non-parabolicity of the heavy hole subband. It is shown that this problem becomes much larger in the case of nanowires and especially under potential variations in the lattice. The dispersions of nanowires in different orientations look very different, and qualitative explanations for their shapes are extracted from the Si bulk Brillouin zone. It is shown that the large change in the dispersion shapes under potential variations makes the effective mass description impossible. The effect of anisotropy is evident in the charge distribution in the cross section of the wires, in which the charge tends to accumulate on the surface with the largest quantization mass. The effect of the semiconductor capacitance on the total gate capacitance is somewhat less severe than in the case of the conduction band. The key contribution of this work is to provide a qualitative understanding of the dispersion shapes and quantization behavior of of p-type nanowires, and their implications in the device performance.

Chapter 6 presents a simulation analysis of recent experimental data for high-performance $In_{0.7}Ga_{0.3}As$ high electron mobility transistors (HEMTs). It is based on reference [51]. The bulk mobility in InGaAs is of the order of ~10,000 cm^2/V-s, so these devices are expected to reach the ballistic limit in much longer channel lengths than Si devices. It is shown that an $In_{0.7}Ga_{0.3}As$ HEMT with a gate length of L_G = 60nm can be understood as a ballistic channel FET with two series resistances attached at the source and drain contacts. Reduction of the series resistance, reduction of the insulator thickness, and optimization of the source design are singled out as the primary drivers to further improve III-V HEMT device performance (at a constant channel effective mass). The importance of this work lies in the theoretical demonstration that the specific devices analyzed can be understood as devices that operate ballistically, or very close to the ballistic limit.

Chapter 7 concludes the thesis work, resumes the main findings and points towards future directions.

2. QUANTUM SIMULATION OF 3D CNTFET DEVICES – THE EFFECT OF ATOMISTIC DEFECTS IN TRANSPORT PROPERTIES OF 1D CHANNELS

2.1 Introduction

As devices shrink to the nanoscale, 1D channel devices such as carbon nanotube field effect transistors (CNTFETs) and nanowire transistors are becoming potential candidates for electron devices. Carbon nanotubes (CNTs) have excellent device characteristics and are candidates for future digital switches and RF transistors [16-20, 52]. Simple circuits based on CNTFETs have already been demonstrated [21]. Nanowire transistors have also been demonstrated with cross sections even down to 3nm. It is expected, however, that the sensitivity of 1D channels to device variations and atomistic defects will be increased. At the mesoscopic regime, the nature of statistical averaging is lost, and each device is a different channel/material on its own. Slight variations to the size of the devices will correspond to large percentages of their sizes, and can affect their properties significantly. Understanding the effect of single defects on the performance of 1D channels will therefore be essential in addressing design issues on the nanoscale. In this chapter, the effect of a variety of defects (negatively/positively charged impurities, vacancies) on 1D CNTFET channels is examined. The effect of dangling bond defects on the transport properties of nanowires is also investigated. The conclusion is common for all devices and defects examined: In one-dimensional systems, the effect of individual defects can have severe effects on the device performance.

For the cases of defects in CNTFETs, a real space 3D ballistic quantum simulator is used. The simulator is based on the Non-Equilibrium Greens' Function method (NEGF) for quantum transport, and 3D electrostatics based on the finite element method

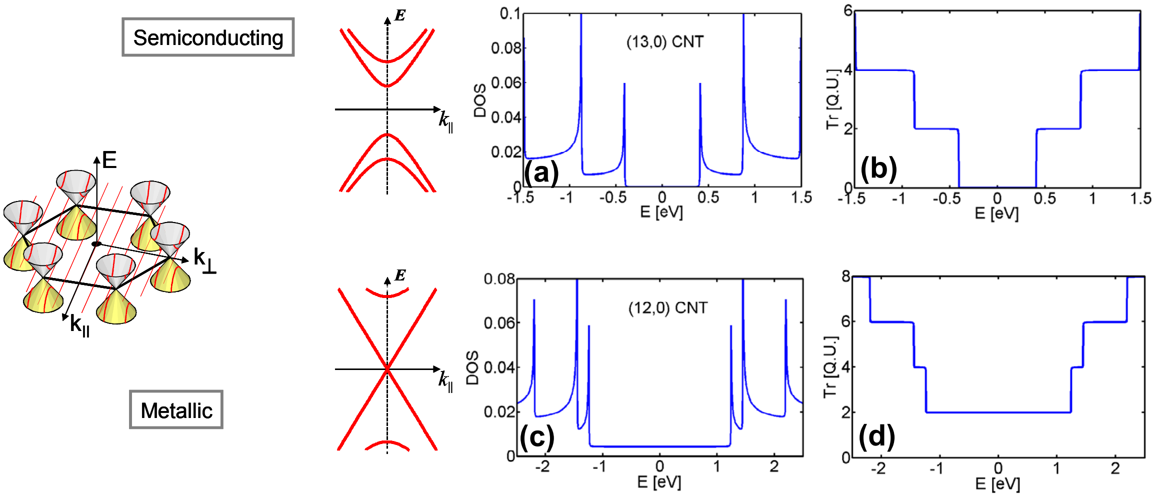


Fig. 2. 1. Electronic features of semiconducting and metallic zig-zag CNT devices: Rolling up the 2D graphene sheet, will quantize the *k*-space according to the direction that the material is rolled up, and according to number of atoms in the ring. (a-b) Semiconducting CNTs i.e. (13,0) have a parabolic dispersion at low energies, and are direct gap materials. (a) The (13,0) CNT *DOS(E)*. (b) The (13,0) CNT *T(E)*. (c-d) The metallic CNTs i.e. (12,0), have a linear dispersion that passes through the fermi level (E_f=0eV). (c) The (12,0) CNT *DOS(E)*. (d) The (12,0) CNT *T(E)*.

(FEM). The channel is described using an atomistic *pz*-orbital nearest-neighbor (NN) tight-binding (TB) description for the device Hamiltonian. Details of the simulator are given in Appendix A.

The chapter is based on references [31, 32, 47] and is organized as follows: The transmission, *T(E)*, and 1D density of states, *DOS(E)*, of CNTs are presented, for the control (un-defected) zig-zag CNT devices, as well as devices with atomistic defects in the channel. The effect of vacancy and charged impurity defects on the CNTFET transport properties is then discussed. Further on, the effect of vacancy defects in metallic zig-zag CNT devices is investigated. Finally, it is shown that dangling bond atomistic defects in nanowire devices show similar effects on the *T(E)* and *DOS(E)* as in the case of CNTs.

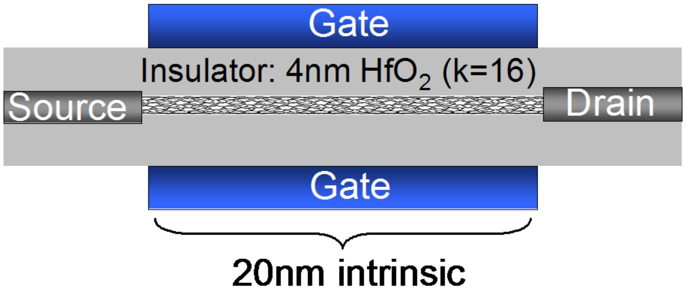


Fig. 2. 2. The 3D coaxial CNTFET with source/drain extensions: The channel length is 20nm. The channel is a zigzag CNT with ~1 nm diameter. The gate oxide is assumed to be 4 nm thick HfO_2 with dielectric constant k=16.

2.2 1D CNT electronic features: Density of states (*DOS(E)*) and transmission coefficient (*T(E)*).

The density of states (*DOS(E)*) and transmission coefficients (*T(E)*) for two different chirality zig-zag CNTs are shown in Fig. 2.1. The calculation was performed using the *pz*-orbital TB description of the CNT and the NEGF formalism. Rolling up the 2D graphene sheet, will quantize the *k*-space according to the direction that the material is rolled up, and according to number of carbon atoms in the ring that is formed. The *k*-space quantization will determine the properties of the CNT. The (13,0) CNT has semiconducting behavior, with zero density of states and transmission around the bandgap region as shown in Fig. 2.1 (a-b). The (12,0) CNT is metallic (zig-zag CNTs in which the chirality number *n* is a multiple of 3 are metallic), with a finite *DOS(E)* around the middle of the energy spectrum, and finite transmission of two quantum units per spin (Fig. 2.1 (c-d)). These results agree well with previously published results for zig-zag CNTs [29, 53].

2.3 The effect of atomistic defects on the transport properties of CNTFETs

In this section, we investigate how different vacancy defects affect the device

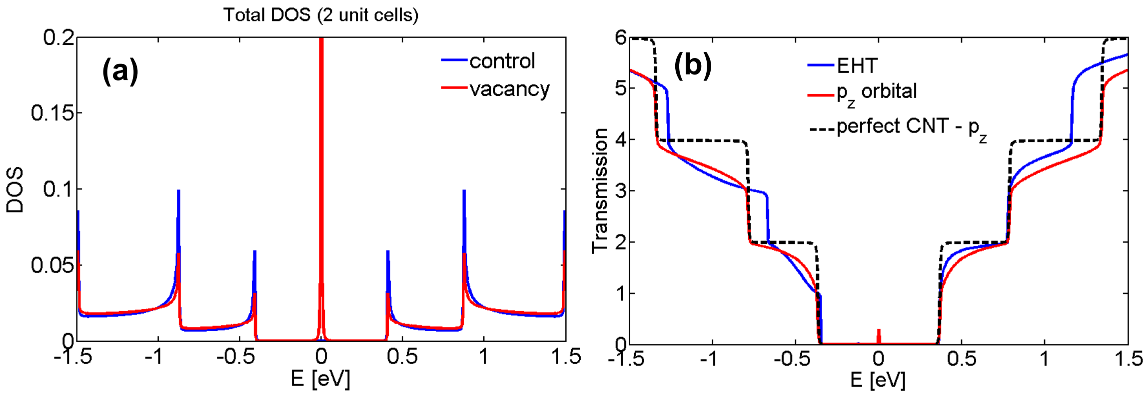


Fig. 2. 3. The effect of vacancies in the electronic properties of CNTs: (a) The density of states of the (13,0) CNT with and without a vacancy defect presence in the channel using a *pz* orbital approximation. The results are drawn on a two unit cell structure. (b) The transmission probability for two unit cells of (13,0) CNT. Results for a perfect CNT, a CNT with a vacancy in the channel modeled in the simple *pz* orbital approximation, and by using the four orbital non-othogonal Extended Huckel Theory (EHT) [54] model.

characteristics of semiconducting zig-zag CNTFETs. We consider the cases of vacancies and both positively and negatively charged impurities.

The model device considered consists of a (13,0) zigzag CNT, with 1nm diameter and 0.8eV bandgap (Fig. 2.2). The length of the undoped channel and source/drain region extensions are 25 and 22.5nm respectively. The source/drain regions are doped uniformly with N_D=10^9 dopants/m. The surrounding gate oxide is a 4nm thick HfO_2 high κ dielectric material (κ=16), whereas the interior of the CNT is vacuum (κ=1). We assume a gate workfunction that produces flat band conditions at $V_G = -E_g/4$ where E_g is the band gap of the CNT ($\Phi m_{Gate} = \Phi m_{CNT} + E_g/4$). The applied drain ($V_d$) and gate ($V_g$) biases vary from 0 to 0.45V, which provide an excellent I_{on}/I_{off} =10^5 and drive current (I_{on}) of 20μA, which are comparable to experimentally reported CNTFET values [17]. To model the vacancy, we set the onsite potential to 10^6eV at the carbon vacancy site, which ensures that a channel electron is effectively repelled from the location of the vacancy [55].

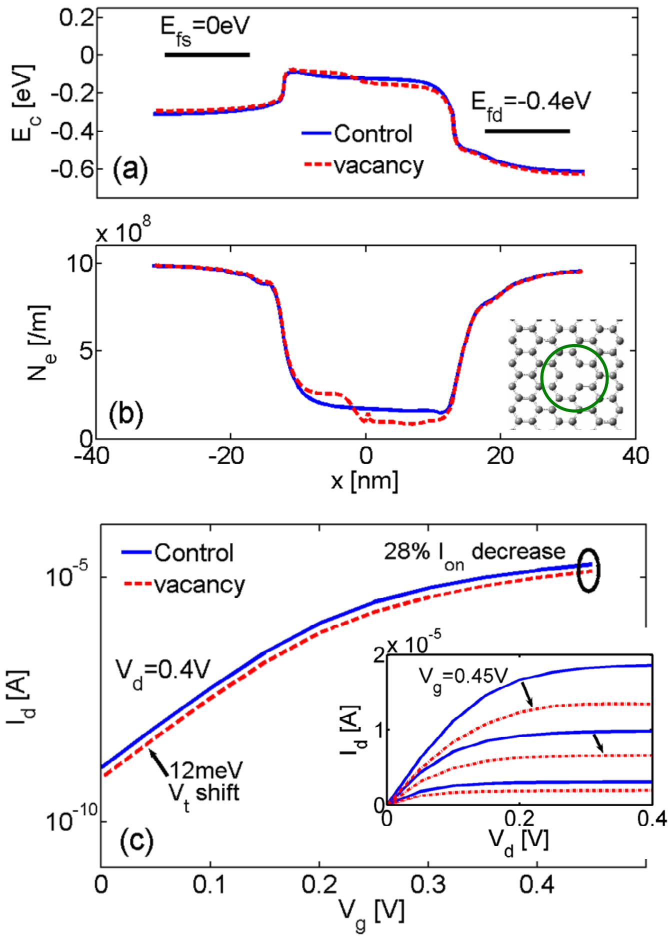


Fig. 2. 4. Effect of the vacancy on the transport properties of the CNTFET. (a) The conduction band of the control vs. vacancy defected CNT at V_g=0.4 V and V_d=0.4 V. (b) The carrier distribution along the channel of the two devices for the case (a). (c) The I_d-V_g characteristics at V_d= 0.4V. Inset: The I_d-V_d for V_g = 0.45, 0.35, 0.25V.

2.3.1 The effect of a vacancy defect

It is well known that vacancies arise in graphite at low concentrations during defective growth or as part of the thermal equilibrium concentration [56, 57]. They are much more prevalent in irradiated materials and are believed to be the predominant defects on irradiated graphite surfaces and CNTs [58-61].

We find that the vacancy creates a localized state in the bandgap and a reduction of the transmission probability in both the conduction and valence bands as shown in Fig.

2.3. The reduction in the transmission probability based on π-orbital model has been confirmed by non-orthogonal extended Hückel calculations using sp^3-orbitals for each carbon [54]. The localized state in the bandgap appears in the Hückel calculations too, however it is slightly shifted towards the valence band. This behavior is also verified experimentally [62].

In this study, we placed the vacancy in the middle of the CNT channel. When the transistor is in the OFF-state, the conduction band edge (E_c) in the channel is 0.2eV above the source Fermi level, and the channel is almost empty of charge. In this case, the vacancy changes the transport characteristics mainly through the reduction of the transmission probability. As the transistor is turned on (V_g=V_d=0.4V), the localized state in the band gap gets partially filled, and causes changes in the potential profile and the carrier occupancy of the channel. There is pile up and depletion of the carrier density (N_e) to the left and right of the vacancy respectively, due to carrier reflections from the localized state as shown in Fig. 2.4 (b). This causes the conduction band profile to float up/down at the left/right sides of the vacancy, as shown in Fig. 2.4 (a). The combination of transmission reduction and the slight increase in the source injection barrier, reduces the drive current from 18.5μA to 13.5μA, a reduction of 28% (Fig. 2.4 (c)). We note that a 12meV shift in the threshold voltage (V_t) is also observed. This shift, however, is responsible for approximately only 8% of the total 28% of Ion reduction. We also considered the case where the vacancy sits near the source and drain ends, and find that the drive current decreases by a similar amount, indicating that the decrease in drive current is independent of the vacancy location.

2.3.2 The effect of a negatively charged impurity defect

A second important source of scattering are charged impurities that reside either on the surface of the CNT, or in the gate oxide. These impurities can be charged ions, molecules, alkali metals or dopants that exchange charge with the CNT [63-66]. Here we investigate the effect of a negatively charged impurity having charge $-|q|$, where q is the charge of an electron. Three different locations of the impurity are considered: (A) the interior of the CNT, (B) the middle of the oxide, and (C) the top of the oxide, 0.5nm from

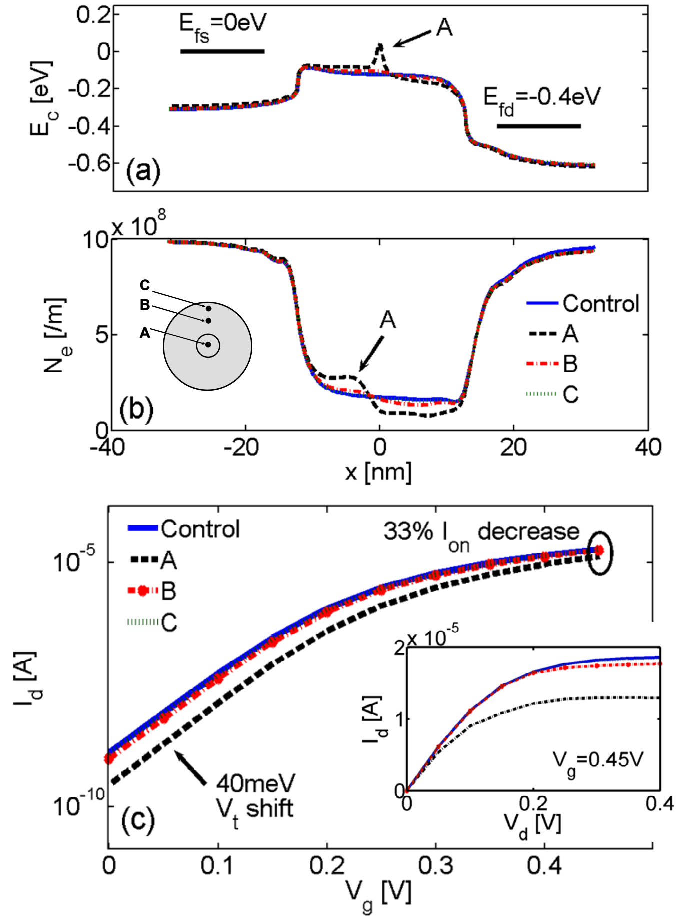


Fig. 2. 5. Effect of a negatively charged impurity in different locations of the device. The charged impurity is placed (A) in the middle of the CNT, (B) in the middle of the oxide and (C) near the gate electrode. (a) The conduction band of the control vs. the different charged impurity cases at V_g=V_d=0.4V. (b) The carrier distribution along the channel for all the impurity cases at the case corresponding to (a). (c) The I_d-V_g characteristics for V_d=0.4V. Inset: I_d-V_g for V_g=0.45V.

the gate electrode (inset of Fig. 2.5 (b)). We find that case (A) gives rise to a large scattering center in the conduction band as shown in Fig. 2.5 (a). Similar to the case of the vacancy localized state, electrons traveling from the source to the drain reflect from this barrier and pile-up/deplete to the left/right of the impurity (Fig. 2.5 (b)). The comparison to the case of vacancy is interesting. While the drive current is reduced by about 33%, which is similar to the case of the vacancy, there is now a large shift in the threshold voltage by about 40meV. For an operating bias of 0.4V, the threshold voltage

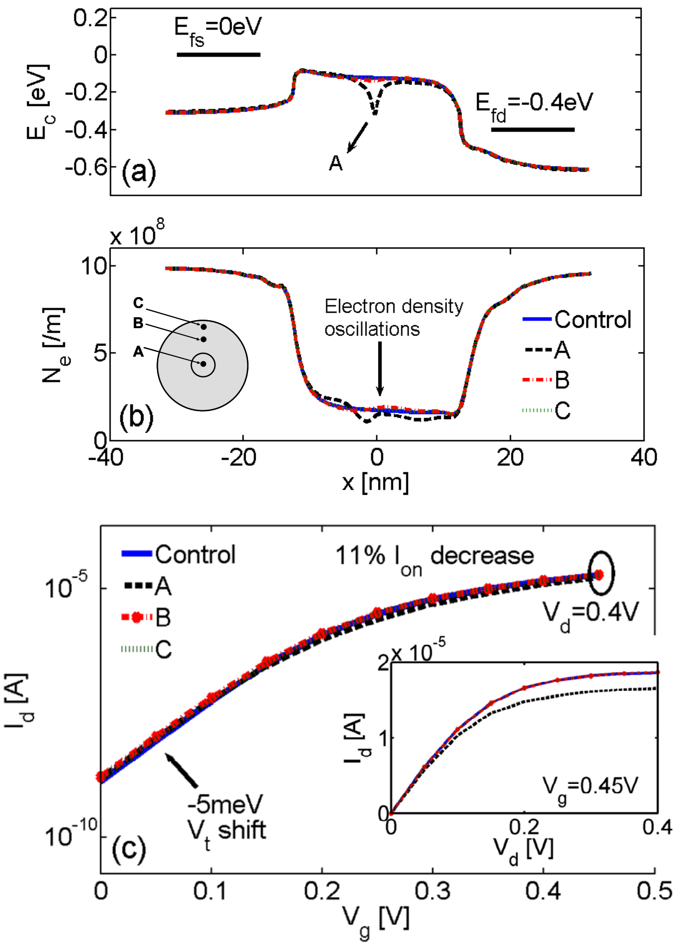


Fig. 2. 6. Effect of a positively charged impurity in different locations of the device. The charged impurity is placed (A) in the middle of the CNT, (B) in the middle of the oxide and (C) near the gate electrode. (a) The conduction band of the control vs. the different charged impurity cases at V_g=V_d=0.4V. (c) The charge distribution along the channel for all the impurity cases at the case corresponding to (a). (c) The I_d-V_g characteristics for V_d=0.4V. Inset: I_d-V_g for V_g=0.45V.

shift is about 10%, a magnitude that might not be tolerated in circuit design and can lead to large I_{off} variations. This can be a very serious problem since it degrades the low bias performance, for which CNTFETs are advantageous compared to conventional silicon MOSFETs, because they can provide large current densities at lower operating biases. In comparison to case (A), the drive current and threshold voltage are affected very little in cases (B) and (C) (5% and 0.5% reduction respectively). This behavior can be attributed to an effective screening of the impurity by the surrounding gate electrode, enhanced by the high k-dielectric of the gate insulator.

2.3.3 The effect of a positively charged impurity defect

Next, we investigate the effect of a positively charged impurity placed at the same locations as in Fig. 2.5 (b). While a negatively charged impurity in the center of the CNT creates a large barrier in the conduction band, a positive impurity creates a large well as shown in Fig. 2.6 (a). The carrier density in the channel (Fig. 2.6 (b)) slightly oscillates, and electrons are attracted around the positive impurity site in the middle of the channel. The top of the barrier is not affected significantly in this case, and as a result the drive current deceases by only 11% as shown in Fig. 2.6 (c). Quantum mechanical scattering from a potential well is weaker than scattering from a barrier. It is also interesting to see here that the shift in V_t is much smaller (5meV) and now negative. Finally, we find that cases (B) and (C), for which the impurity is placed in the oxide further away from the CNT shell, have an insignificant effect on the device performance. We would like to mention here, that for a p-type device, the relative role of the positive and negative impurities would be reversed.

2.3.4 Vacancy defects in metallic CNT channels.

Introduction of a vacancy defect in the channel of metallic CNTs, similarly to the case of semiconducting CNTs, alters both the density of states and the transmission near the Fermi level due to the change in bonding between the vacancy and its nearest neighbors, and introduces localized states. In Fig. 2.7 (a-b), the non-self consistent simulated density of states (*DOS(E)*) and transmission (*T(E)*) for a (12,0) CNT is shown respectively, with and without a single vacancy present in the channel. The model device in this case, is just a two unit cell channel, with open boundary conditions for the contacts.

The density of states around the Fermi level (Fig. 2.7 (a)) rises due to the broken π-network at the vacancy site. The transmission on the other hand (Fig. 2.7 (b)) decreases

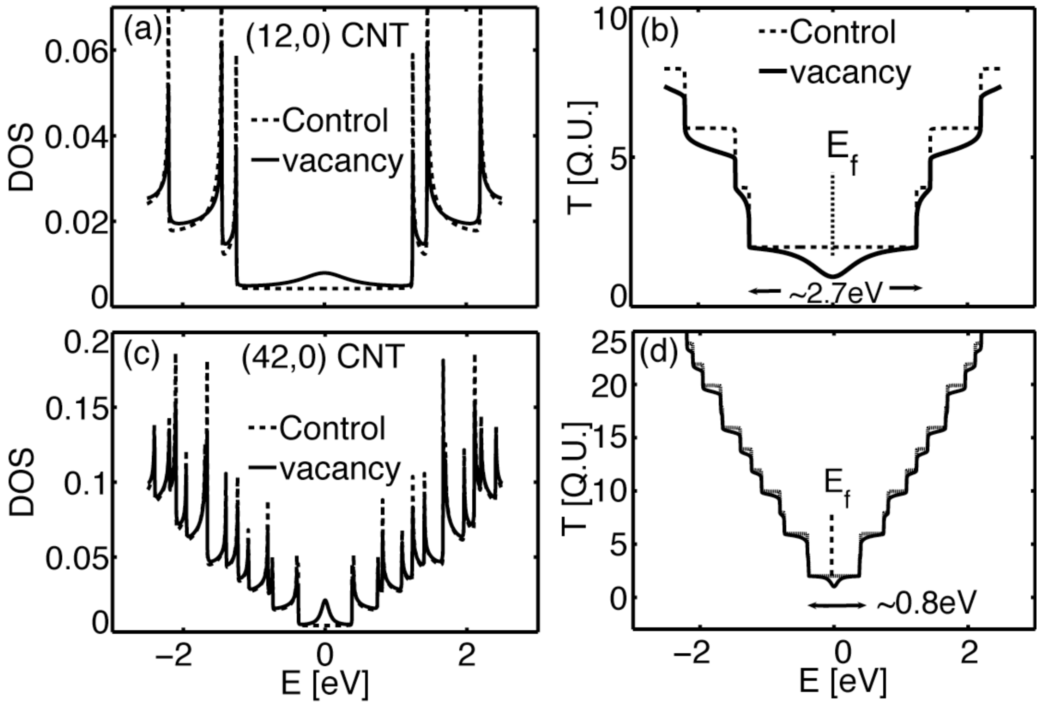


Fig. 2. 7. Density of states (*DOS(E)*) and transmission (*T(E)*) for metallic CNT channels consisting of two unit cells with open boundary conditions. (a) The *DOS(E)* for a (12,0) CNT with ~1 nm diameter for the control (defect-free – dash line), vs. vacancy included (solid line). The unit cell consists of 48 atoms. (b) The T(E) of the control (12,0) CNT (dash) vs. the defected one (solid). (c) The *DOS(E)* for a (42,0) CNT with ~4nm diameter for the control (defect-free - dash), vs. vacancy included (solid). The unit cell consists of 168 atoms. (d) The *T(E)* of the control (42,0) CNT (dash) unit cell vs. the defected one (solid).

close to one quantum unit (q^2/h per spin, where q is the charge of the electron and h is Plank's constant) in the vicinity of the Fermi level in agreement with Chico et. al. [53]. For CNTs with larger diameters, i.e. (42,0) with 4nm diameter, similar behavior is observed as in the case of the (12,0) CNT as shown in Fig. 2.7 (c-d). In this case, however, the reduction in transmission is more localized near the middle of the energy spectrum of the CNT. The same effect is observed for the increase in the density of states, which is mostly localized in the middle of the energy spectrum. These differences will significantly affect in the *I-V* characteristics of the two devices.

The self-consistent simulated charge and potential energy distribution along the channel of the coaxial (12,0) CNTFET of Fig. 2.2 are shown in Fig. 2.8 (a-b). The

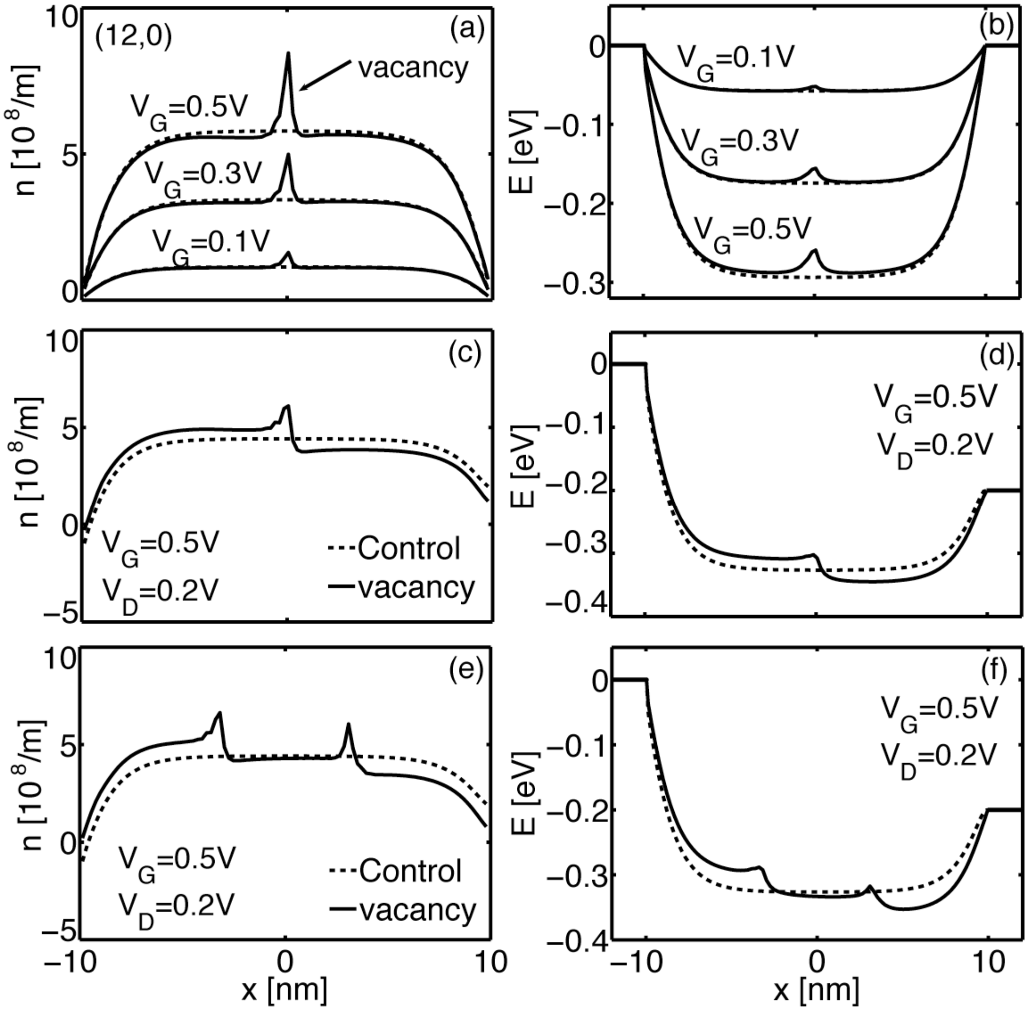


Fig. 2. 8. The effect of the vacancy on the electronic properties of a metallic (12,0) CNTFET. (a) The charge distribution along the channel of the control (dash) vs. the vacancy contained (solid) CNTFET for V_G=0.1, 0.3, and 0.5V, under equilibrium conditions (V_D=0 V). (b) The corresponding potential energy profile for the cases described in (a). (c) The charge distribution along the channel of the control and the defected CNTFET for V_G=0.5 V, under non-equilibrium conditions (V_D=0.2 V). Large scattering is observed at the vacancy site. Charge is accumulated in the left of the vacancy. (d) The corresponding potential energy profile for the case described in (c). (e) The charge distribution along the channel of the control and the defected with two vacancies CNTFET for V_G=0.5 V, under non-equilibrium conditions (V_D=0.2 V). (f) The corresponding potential energy profile for the case described in (e).

vacancy is placed in the middle of the channel. Fig. 2.8 (a) indicates an accumulation of charge at the vacancy location, which is associated with the increased local *DOS(E)*. The amount of charge increases as the bias on the gate electrode increases, because more of the vacancy induced states are now filled with electrons. This charge accumulation raises the electrostatic energy of the channel locally (Fig. 2.8 (b)). Under non-zero gate and drain biases (V_G=0.5V, V_D=0.2V), shown in Fig. 2.8 (c-d), the local accumulation of charge near the defect site creates a scattering center, and causes changes in the potential profile and the carrier occupancy of the channel in its entire length. There is accumulation and depletion of the carrier density (n) to the left and right of the vacancy, respectively, due to carrier reflections from the localized state as indicated in Fig. 2.8 (c). This causes the potential energy profile to float up (down) at the left (right) sides of the vacancy (Fig. 2.8 (d)). Introduction of two vacancies in the channel of the device as shown in Fig. 2.8 (e-f), has a similar local effect in the places where the vacancy is present. In this case the vacancies are located at $L/3$ (~6.7nm) and $2L/3$ (~13nm) from the source contact, where L is the length of the channel (20nm). The two peaks in the charge distribution indicate the location of the defects, which cause pile up of the charge and the potential energy to the left side of the device, and depletion to the right side.

The I_D-V_D characteristics are computed by self consistent NEGF simulations performed in the ballistic limit. The current is calculated in the Landauer formalism as $I = \int T(E)(E_{f_S} - E_{f_D})dE$, where ($Ef_S$ -Ef_D) is the difference between the Fermi levels of the source and the drain electrodes. As indicated in Fig. 2.9 (a), the self consistent transmission of the control (12,0) CNT device is constant at two quantum units over the entire energy range of interest (within the Ef_S - Ef_D window). For a device with a single vacancy the transmission is reduced to one quantum unit, and for the case of two vacancies in the channel, it reduces even further. Since the reduction happens over a large energy range, shifting the Ef_S-Ef_D window (for example by using a different gate bias or different workfunction for the gate), will not help in improving the defected devices' transmission probabilities. Figure 2.9 (b) shows the self consistently computed I_D-V_D characteristics for a (12,0) CNT with and without vacancy defects in the channel. Because of the constant transmission of the control CNT around the Fermi level, the metallic CNT I_D-V_D characteristics are linear with the slope being the conductance of the channel, as shown in Fig. 2.9 (b) (solid line). A single vacancy defect in the middle of the channel reduces the conductance by a factor of two (dashed line). Two and three vacancies are considered, placed in equal distances from one another in the channel. The

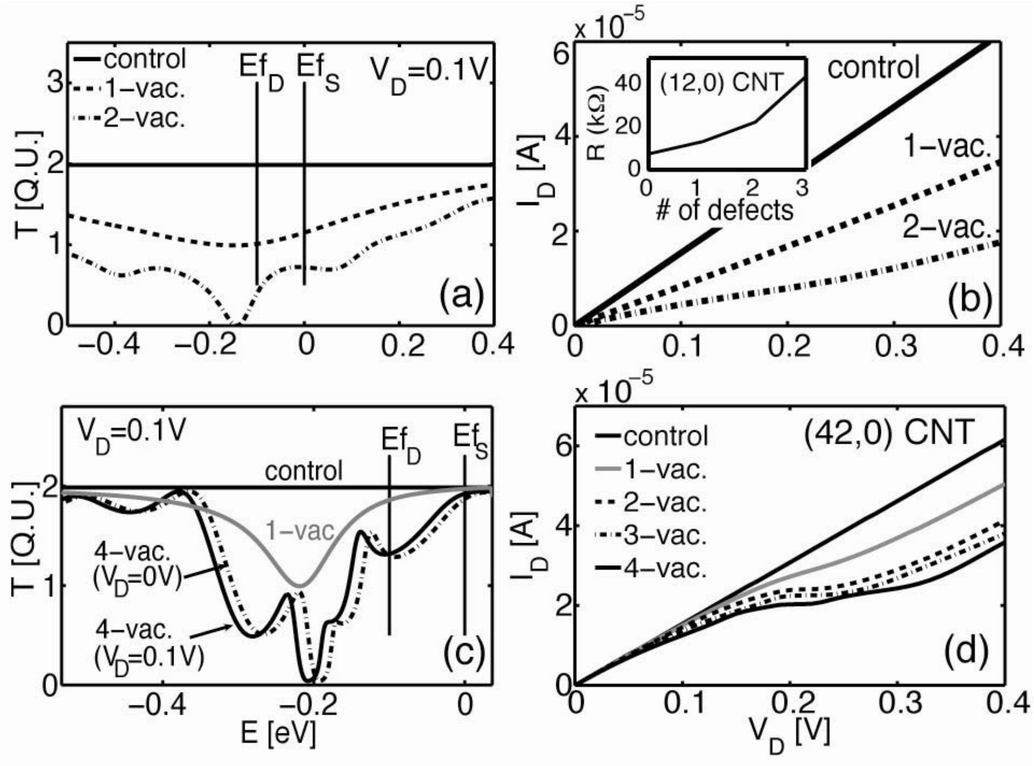


Fig. 2. 9. The effect of the vacancy defects on the drive current capabilities of the metallic CNTFET. (a) The self consistent transmission coefficient of the (12,0) CNTFET with none (solid), one (dash) and two (dash-dot) vacancies in the channel. The window *Ef_S-Ef_D* indicates the important electron transport energy region. (b) I_D vs. V_D for control vs. vacancy defected (12,0) CNTFETs. Control (solid), single vacancy (dash) and two vacancies (dash-dot) in the channel. INSET: The resistance of the channel as a function of the number of vacancies. (c) The self-consistent transmission coefficient of the (42,0) CNTFET for the control (solid), single vacancy (solid-grey), four-vacancy device (solid) at V_D=0.1V. Also, the four-vacancy device at V_D=0.1V (dash-dot). In all cases, the vacancies are placed in equal distances from each other in the channel, i.e. in the case of the two vacancies, they are placed at $L/3$ and $2L/3$ distances in the channel, where L is the channel length. (d) I_D vs. V_D for control vs. vacancy defected (42,0) CNTFETs.

drive current reduces by almost half every time another defect is introduced in the channel and the corresponding channel resistance increases (inset of Fig. 2.9 (b)). It is also found, that the exact position of the defect does not significantly alter the I_D-V_D characteristics. Variations in the defect position can cause small shifts and oscillations in the transmission spectrum; however they cannot affect the overall drive current characteristics significantly.

The control (42,0) CNT, with larger diameter (~4nm), has the same transmission as the (12,0) CNT (2 quantum units), and therefore delivers the same amount of current under small drain biases. (The (42,0) CNT will deliver more current at higher drain biases since other energy bands appear in much lower energies than the (12,0) CNT). The transmission plots of the control CNT, the CNT with one and four vacancies for the case of $V_D = 0.1V$, are shown in Fig. 2.9 (c). The transmission curve for the case of the four-vacancy CNT at $V_D = 0V$ (dotted line) is also shown for comparison to the four-vacancy device at $V_D = 0.1V$. This shows that the shape of the transmission of a certain channel does not significantly change with V_D, except from a small shift in its energy spectrum (smaller than the V_D window). The transmission curves for the cases of two and three vacancies fall in between these curves, however for simplicity they are not shown here.

The I_D-V_D characteristics for both control and defected (42,0) CNTs are shown in Fig. 2.9 (d). (Due to numerical convergence issues at high biases for the defected devices, the I_D-V_D characteristics were extracted for the entire bias range by using the self consistent transmission coefficient derived at $V_D = 0.1V$, considering that the shape of the transmission profile does not change significantly with V_D as shown in Fig. 2.9 (c)). At low drain biases, the drive current in all the devices is almost unaffected. The reason is that the reduction in transmission for all the defected devices is more localized within a smaller energy range (compared to the (12,0) CNT case) and initially falls out of the Ef_S-Ef_D window of width 0.1V. This, however, depends on the choice of the initial placement of this energy window (by changing the gate bias or the workfunctions of the electrodes). As the drain bias is increased and the reduced transmission region (around -0.2eV), falls within the Ef_S-Ef_D window, the drive current of the defected devices reduces. At even higher biases, the transmission, as well as the conductance of the channel (slope of the I_D-V_D) is rectified. Due to this dynamic dependence of the current on the drain bias, a single value for the channel resistance cannot be easily defined as in the case of the (12,0) CNT. Physically this difference between the performances of the two different diameter defected devices, can be explained by the fact that in the larger diameter CNT, the electrons have larger phase space and more paths to avoid scattering off the vacancy, making its effect less important.

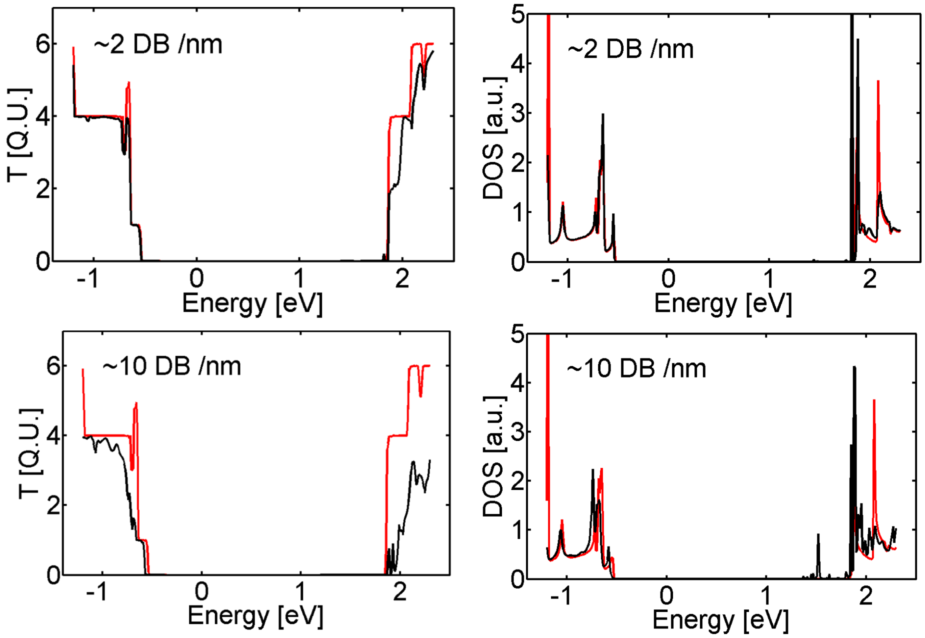


Fig. 2. 10. Density of states (*DOS(E)*) and transmission (*T(E)*) for 1.5nm cross section, 7nm length [110] oriented Si nanowire channels with open boundary conditions. (a) The *T(E)* for the control (defect-free – red line), vs. the same nanowire with 2 dangling bond defects/nm included (blue line). (b) The *DOS(E)* for the case described in (a). (c) The *T(E)* for the control (defect-free – red line), vs. the same nanowire with 10 dangling bond defects/nm included (blue line). (b) The *DOS(E)* for the case described in (c).

2.4 Dangling bond defects in ultra scaled cross section nanowires

The sensitivity of to single atomistic defects is observed in ultra scaled nanowires channels too. Here, the behavior of dangling bond defects in square nanowires with 1.5nm cross sections and 7nm in length is examined using an atomistic $sp^3d^5s^*$ TB model (see Appendix B) for the Hamiltonian description and the NEGF approach for the transport properties calculation. The transmission *T(E)* and density of states *DOS(E)* is presented in Fig. 2.10. The whole range all the way from the valence to the conduction band is presented in the energy axis. The control device with no dangling bond defects characteristics is shown in red, whereas cases for 2 dangling bond defects (Fig. 2.10 (a,b)) and 10 dangling bond defects per nanometer are presented (Fig. 2.10 (c,d)). Introduction of defects are in the case of CNTs, lowers the transmission significantly in both 2 defect and 10 defect cases. The *DOS(E)* loses the perfect 1D behavior, and oscillations are now observed in the plots. In the case of the 10 defects/nm, localized states start to appear in the bandgap, which are both the unpassivated bonds' states, and an effect of 1D localization due to imperfections in the wire.

2.5 Summary

In this chapter, a 3D real space quantum simulator for CNTFET electron transport was used to study the effect of charged impurity defects and vacancy defects in altering nanotube transistor device characteristics from the ballistic limit. A single vacancy can cause drive current reduction by approximately 28%, independent of the location of the vacancy in the channel.

While a single negatively charged impurity near the channel also decreases the drive current by a similar amount, it leads to a much larger threshold voltage shift, comparable to 40meV (10% of the power supply). The scattering strength of the charged impurity weakens, and finally becomes negligible when the scatterer is placed away from the CNT channel, close to the gate electrode. For a n-type device, a localized positively charged impurity causes a much less performance degradation and V_t shift (only 5meV) compared to the negative impurity.

Vacancy defects in metallic CNTs have similar effects. In small diameter metallic CNTs, a single vacancy defect can reduce the drive current capabilities of the CNT by half. Multiple defects can decrease the performance even more. The reduction, however, is much less in larger diameter CNTs. Atomistic defects affect the transport of nanowires similarly by reducing the transmission coefficient.

It is quite remarkable that a single defect can cause such large degradation in drive current and threshold voltage shift. Design of circuits using these quasi-one dimensional transistors should take this into consideration.

3. SIMULATIONS OF BALLISTIC TRANSPORT IN NANOWIRE TRANSISTORS: A SELF-CONSISTENT ATOMISTIC MODEL

3.1 Introduction

The physical dimension of 1D nanowires and the enhanced cross sectional quantization, call for models beyond effective mass approximation (EMA) in evaluating their electron transport performance. Tight-binding atomistic models, as explained in chapter 1, account for the new physics that need to be included, such as non-parabolicity, confinement level position beyond the effective mass approximation (for both structural and electrostatic confinement), the effects of strain in the electronic structure (uniform and non-uniform), the effect of surface roughness, as well as a more accurate distribution of charge in the device channel. Atomistic treatment provides automatically the dispersion of nanowires in arbitrary orientations and includes information about band coupling and mass variations as functions of quantization. In this chapter a model for evaluating the ballistic transport properties of nanowire transistors is described. The model is based on atomistic representation of the lattice to capture all essential physics described earlier. Cases where the EMA can still be applied are presented; however, EMA is not always valid. A crucial step in the model is accounting of the potential variations of in the cross section of the nanowire, which affect the device properties at large.

3.2 The empirical nearest-neighbor (NN) $sp^3d^5s^*$ tight-binding (TB) model

The atomistic model used is the $sp^3d^5s^*$ empirical nearest-neighbor (NN) tight-binding (TB) model. The basis set of the model is composed of orthogonal localized orbitals. This type of basis makes it very attractive for accurate electronic structure of truncated nanostructures of finite sizes and composition variations on the nm-scale. It is a very convenient method to treat material and potential variations as well as strain fields at the nanoscale. The parameterization was performed using a genetic algorithm in [40, 41], and the parameters extracted can reproduce the band edges of the bulk silicon bandstructure over the entire Brillouin zone. (Figure 3.1 (a) shows the constant energy surface ellipsoids of the first Brillouin zone of Si). The model is described in detail in references [39, 41-45]. The energy bands obtained for nanowires, as well as in the bulk case in energy regions away from the bulk minima, are in good quantitative agreement with other theoretical calculations using pseudopotential and ab-initio GW methods [46]. In this work, the electrostatic potential for charge self consistency is also included on the on-site energies of the Hamiltonian in an effective potential approach which shifts the bands with no further change in connectivity.

Validation of the model through experimental data: Since the accuracy of the results presented here strongly depends on the validity of the TB model used, and especially on its transferability to nanostructures, it is convincing to mention that the same model and calibration parameters were used to explain experimental data in a variety of applications with excellent qualitative agreement. Some examples include explaining resonant tunneling diode applications for transport under high bias with charge self consistency [67-69], explaining experimental data for the bandgap of ultra scaled nanowires [70-72] (shown in Fig. 3.1 (b)), valley splitting of tilted and disordered quantum wells [73], and the electronic structure of silicon systems with phosphorus impurities [74]. Further theoretical work presented in reference [75] examines the performance of core shell nanowires and validates against experimental data. Specifically, the theoretical calculation of experimental measurements of the bandgap of ultra scaled [112] oriented nanowires in [70-72] as shown in Fig. 3.1 (b), is a strong validation that the model captures the essential non-parabolicities in a large part of the Brillouin zone of Si. As it will be shown later on, non-parabolicites and anisotropies at high energies strongly influence the masses and band edges of nanowires. Since the

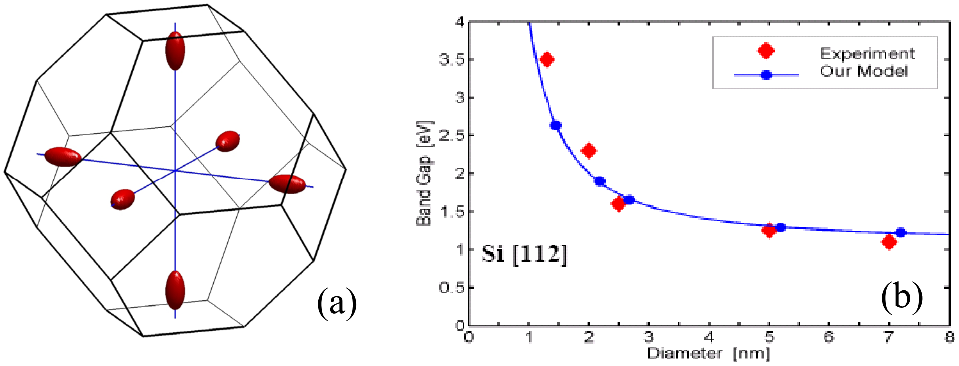


Fig. 3. 1. The $sp^3d^5s^*$ model and validity: (a) The constant energy surfaces of the first Brillouin zone of Si. (Adopted from Anisur Rahmans' dissertation [76]). (b) Comparison between the calculated and experimentally resolved bandgaps in [112] oriented silicon nanowires. Experimental results are extracted form [70]. TB calculation indicates good match (Adopted Jing Wang's dissertation [71]).

bandgap of quantized systems is a strong function of the quantization masses in the two transverse directions, a verification of the experimentally deduced nanowires' bandgap, supports the theoretical prediction for the behavior of the wires' masses under strong quantization, and in extent the validity of the model.

3.3 The simulation scheme

The simulation approach: The devices that can be treated are nanowires of arbitrary cross sectional shapes and transport orientations. The atomic arrangement of wires in different orientations is different as shown for the [100], [110] and [111] nanowires in Fig. 3.2 (b). Atomistic description of the device in arbitrary orientations, has the advantage of automatically capturing the valley projections and extract the dispersions of the channels in arbitrary transport orientations.

The simulation procedure consists of three steps as shown in Fig. 3.2 (a) and described below:

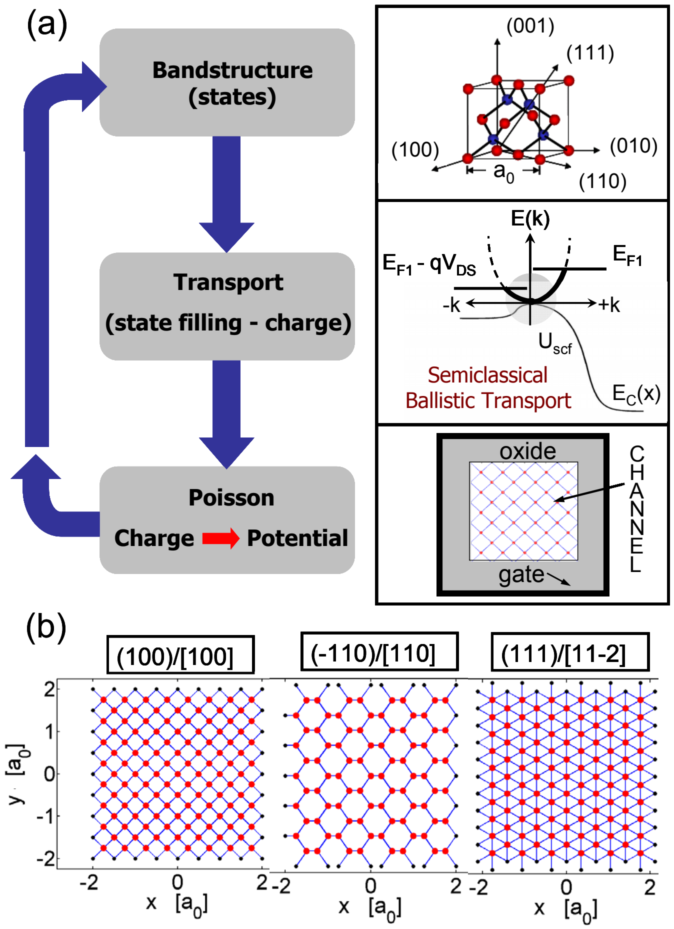


Fig. 3. 2. The 2D self-consistent model: (a) Simulation procedure schematic. Using an atomistic $sp^3d^5s^*$-SO tight-binding model, the bandstructure of the nanowire under consideration is calculated. A semiclassical ballistic model is then used to calculate the charge distribution in the wire from the source and drain Fermi levels. The charge is used in a 2D Poisson for the electrostatic solution of the potential in the cross section of the wire. The whole process is done self consistently. (b) The lattice in the wire transport orientations (surfaces) used – [100], [110] and [111].

1. First, the bandstructure of the wire is calculated using an atomistic tight-binding model. In this case, each atomic side in the zincblende lattice is represented by a $sp^3d^5s^*$ basis in the wire Hamiltonian. (In the case of conduction band calculations the spin-orbit coupling can be ignored. This approximation favors computational efficiency, without affecting the accuracy of the results [26]). The atoms that reside on the surface of the nanowire are passivated in the sp^3 hybridization scheme [45]. This technique successfully removes all dangling

bonds which otherwise will create surface states with eigen-energies in the bandgap of the device. Any effect of surface reconstruction or surface imperfections is not considered in this study. Only the channel atoms enter the atomistic calculation in the Hamiltonian construction. At this step, the energy of the dispersion states and their wavefunctions are computed. Bandstructure effects such as valley splitting and effective mass change under physical quantization are investigated at this step for the nanowire of interest, using the equilibrium dispersion (flat electrostatic potential in the Hamiltonian).

2. A semiclassical top-of-the-barrier ballistic model is used to fill the dispersion states and compute the transport characteristics [77, 78]. This model assumes that the positive going states are filled according to the source Fermi level, whereas the negative going states according to the drain Fermi level. Once the occupancy of the dispersion states is computed, using their wavefunction from step 1, the charge distribution in each of the orbital sites of the system (and therefore the spatial distribution of charge) is obtained (see Appendix C for details).

3. Using the charge distribution obtained in step 2, the 2D Poisson equation is solved in the cross section of the wire to obtain the electrostatic potential. The Poisson's equation is solved in 2D and all the atomic locations are collapsed on the 2D plane [79]. The Poisson domain is described by a finite difference mesh and contains the nanowire core on an atomistic mesh, the dielectric and the metal. The electrostatic potential is added to the diagonal elements of the atomistic Hamiltonian for recalculating the bandstructure until self consistency is achieved (numerical details are provided in Appendix C). An insulator surrounds the channel, which is assumed to be SiO_2 of 1.1nm thickness throughout the results presented in chapters 3, 4 and 5 of the dissertation. This dielectric is not included in the Hamiltonian, but only treated in the Poisson equation as a continuum medium. Any effects due to the potential variations along the transport direction are ignored. This falls under the assumption that at the ballistic limit the carrier injection at the top of the barrier is of most importance to the transport properties.

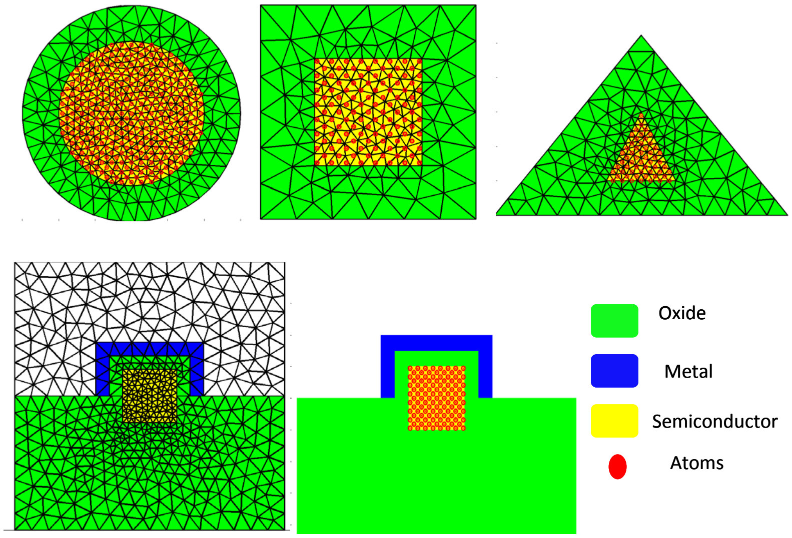
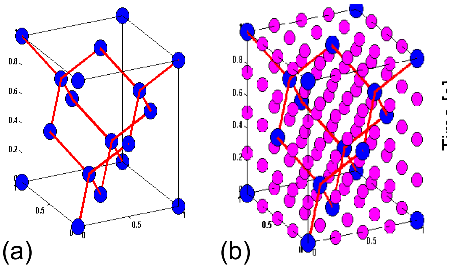


Fig. 3. 3. The 2D cross sections of the devices treated. The finite element mesh and the atomic positions (dots) are indicated. Top row: (a) Cylindrical, (b) Rectangular, (c) Triangular. Bottom row: (d) The tri-gate device structure, (e) The device domains for tri-gate structure. [80]

Further numerical details of the model are described in Appendix C.

Although the transport model used is simplistic, it allows for examining how the bandstructure of the nanowire alone will affect its ballistic transport characteristics, ignoring any short channel effects or quantum mechanical tunneling under the potential barrier. The same conclusion to this work can be obtained from full 3D quantum (NEGF) simulations [27, 81], but the simple model used here provides physical insight. Although the results presented in this thesis are for rectangular nanowires, the basic conclusions will hold for wires with different cross sectional shapes [34]. The model, however, has been extended to treat arbitrary cross sectional shapes as shown in Fig. 3.3 using the Finite Element Method for meshing the structure for solution of the Poisson equation [80]. It is noted here that in all the calculations performed the potential variations between the atomic locations in the wire are small compared to the tight-binding parameters used, so the tight-binding approximation is still valid.

The Poisson solution on a 3D zincblende lattice poses an interesting challenge for typical regular mesh solvers, zincblende is not a space-filling mesh. The lattice can be



(a) **3D atom positions** (b) **Poisson Mesh points**

Fig. 3. 4. The 3D zincblende lattice and Poisson mesh: (a) The position of the atoms in the zincblende lattice. (b) 3D mesh nodes for the equivalent 2D Poisson. (c) Time comparison between a full 3-D atomistic lattice and a quasi 2-D solution. (d) I_D-V_G characteristics computed for quasi 2-D and full 3-D Poisson solution. [80]

symmetrised and solved in standard finite difference/element methods in 3D (Fig. 3.4 (a,b)). The 2D solution (with all the zincblende atomistic charge collapsed on the 2D plane) corresponds in the 3D representation of Fig. 3.4 (a,b) indicating that the 2D approximation spreads out charge in real space significantly different than the original zincblende lattice. We have conducted extensive tests to validate the 2D Poisson solution compared to the actual 3D solution. A maximum deviation of 2% on band edges between the two approaches was found, and ignorable deviation in the I-V characteristics [80]. However, the 2D method reduces the Poisson computational time by almost 5X compared to the 3D solution, and it is the one used in the subsequent studies in this thesis.

3.4 Electronic structure of Si nanowires under bias

Potential and charge variation in the device cross section is a crucial step in device performance evaluation, even for nanowires of 3nm cross sectional sizes. This has a large impact on the semiconductor capacitance of the device (C_S). Potential variations cause changes in the equilibrium bandstructure of the wire. Similar to effective mass calculations, the levels shift according to the electrostatic confinement. The shifts in the

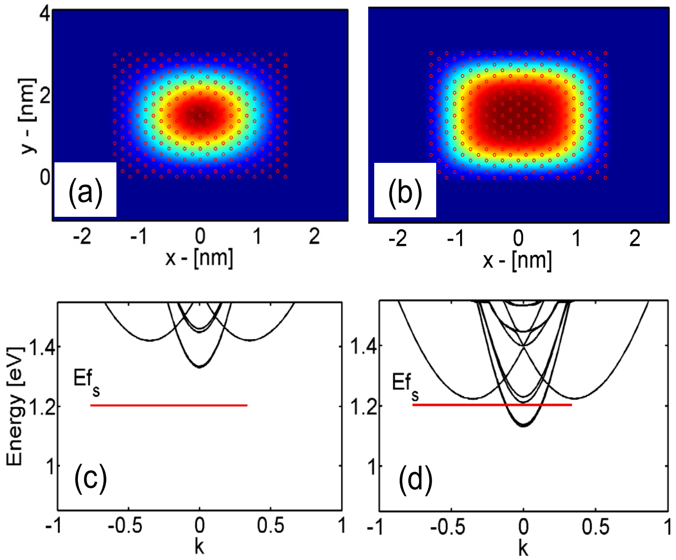


Fig. 3. 5. Device features for a 3nm rectangular wire. (a-b) The 2D cross section showing the charge distribution under low and high gate biases, respectively. The dots indicate the underlying atomic positions. (c-d) *E(k)* plots for the cases (a-b). Ef_s is the source Fermi level.

levels are, however, qualitatively and quantitatively different that the shifts obtained from an effective mass solution. Details on these issues will be addressed in chapters 4 and 5. In this section, however, it is indicated that considering these effects in transport through nanowires, one obtains quantitatively significantly different results than just considering transport through the empty lattice bandstructure as it was performed in previous models [34, 82], for both, small nanowires (3nm width) as well as in wires with larger cross sections.

Figure 3.5 shows the self consistent results for the charge distribution in a 3nm x 3nm square, [100] transport direction, Si nanowire, under low and high gate bias conditions, i.e. when the channel is partially and fully inverted. The underlying structure of the atoms is evident in the case where the channel is fully inverted (dots). A nanowire dispersion curve is usually considered to be a material and geometry dependent quantity, independent of the filling of the states. The difference between Fig. 3.5 (c-d), however, indicates that the dispersion in Fig. 3.5 (d) is not a solid shift in energy from Fig. 3.5 (c). The first set of excited states shifts below the band-edge minima at k=0.405. The filling

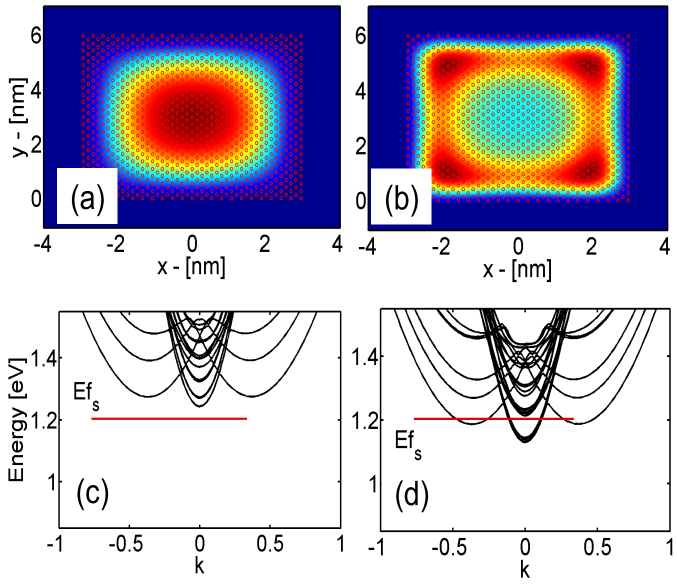


Fig. 3. 6. Device features for a 6nm rectangular wire. (a-b) The 2D cross section showing the charge distribution under low and high gate biases, respectively. The dots indicate the underlying atomic positions. (c-d) The corresponding *E(k)* plots for the cases (a-b). Ef_s is the source Fermi level.

of the states in the device changes the electrostatic potential, which in turn changes the lateral confinement. The change in the lateral confinement in turn changes the dispersion in the transport direction.

Larger nanowires show another interesting behavior under inversion conditions. With increasing gate biases the charge shifts from being confined in the center of the wire to be confined in the corners of the wire. The electrostatics of the device force these corner regions into stronger inversion (Fig. 3.6 (a,b)). Figures 3.6 (c-d) also shows significant changes in the bandstructure of the nanowire between the low and high bias conditions cases. This “corner effect” is observed in corners of strongly inverted multi gate devices.

Changes in the bandstructure and spatial distribution of charges reflect on the *I-V* device characteristics. Figure 3.7 shows a comparison between the I_D-V_G characteristics of the device for two simulation approaches: (i) the Poisson equation is solved in the cross section of the rectangular nanowire and the potential variation is considered in the

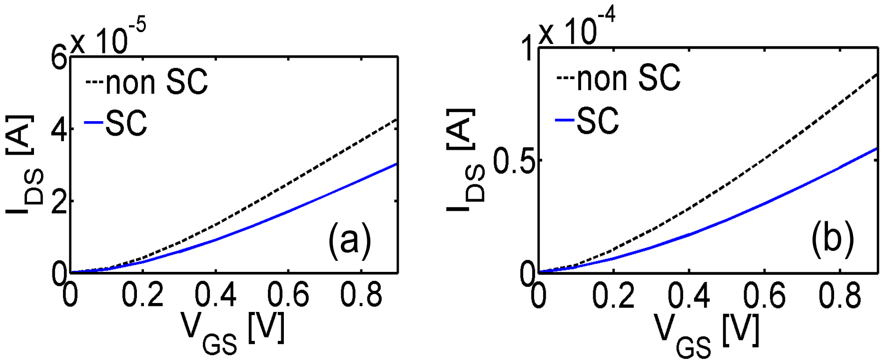


Fig. 3. 7. The I_D-V_G characteristics calculated using the empty (non-SC) E(k) diagram and using the charge filled (SC) *E(k)* diagram for rectangular nanowires. (a) The 3nm device. (b) The 6nm device.

bandstructure calculation, and (ii) a simple planar capacitance (oxide capacitance) is considered for the electrostatics of the device (no potential variation is considered in the cross section). The spatial variation of the charge, however, (i.e. volume inversion for small wires) makes the simple capacitance assumption inaccurate. This results in significant differences in the I_D-V_G characteristics obtained by the two methods, for both, small 3nm diameter (Fig. 3.7 (a)), and larger, 6nm diameter (Fig. 3.7 (b)) wires. Considering only a simple shift of the bands due to the gate bias will result in overestimating the drive current of the device. The small variation of the *E(k)* levels through self-consistency (especially for larger wires) by itself, does not have large effect in the transport characteristics in this case. It should be though, however, as connected to its corresponding wavefunction that causes the particular changes in the *E(k)*. The variations themselves can have a larger effect in the case of the valence band, where the bands deform drastically under potential variation in the lattice, as it will be shown in Chapter 5. In addition, in low drain bias and low temperature cases, where the transport window can be resolved within a few meV, these changes can show up in the transport characteristics.

3.5 $sp^3d^5s^*$ TB vs. effective mass approximation (EMA) for nanowires

Although in ultra scaled nanowire channels the atomistic approach will be more appropriate, in certain cases, the effective mass approximation (EMA) can give very similar results, once the masses are adjusted to masses extracted from $sp^3d^5s^*$. In this section, a few cases are presented, in which the EMA can match the atomistic results quite good. The examples presented are only for the conduction band. The strong interactions in the valence band, as will be shown in chapter 5, make the use of effective mass extremely difficult.

The EMA is formulated in the same model way as the $sp^3d^5s^*$ formalism. The only exception is that the Hamiltonian elements are a single 1x1 elements rather than a multi-orbital block. The bandstructure under such an approach, consists of parabolic bands centered at the Γ point, with the specified transport direction effective mass. An effective mass Hamiltonian can be built on a zincblende lattice, in which case the on side and the coupling parameters need to be changed in such a way in order to get the correct effective mass and quantization levels in the directions of interest. However, it is more convenient for the EMA Hamiltonian to be built on a cubic lattice, where the coupling parameters and the on-sites of the Hamiltonian are directly related to the effective masses in the specified directions of interest. The formulation of the 3D Schrodinger equation using a 3D Hamiltonian based on a 3D cubic lattice can be described as follows:

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

An ellipsoidal parabolic energy band with a diagonal effective mass tensor is assumed. This equation needs to be solved three times due to the three different types of ellipsoids in the Si conduction band bandstructure. The 3D Hamiltonian is expressed as:

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

where m_x^*, m_y^*, and m_z^* are the electron effective masses in the x, y, and z directions respectively, and $U(y,z)$ is the electrostatic potential energy in the 2D cross section. The

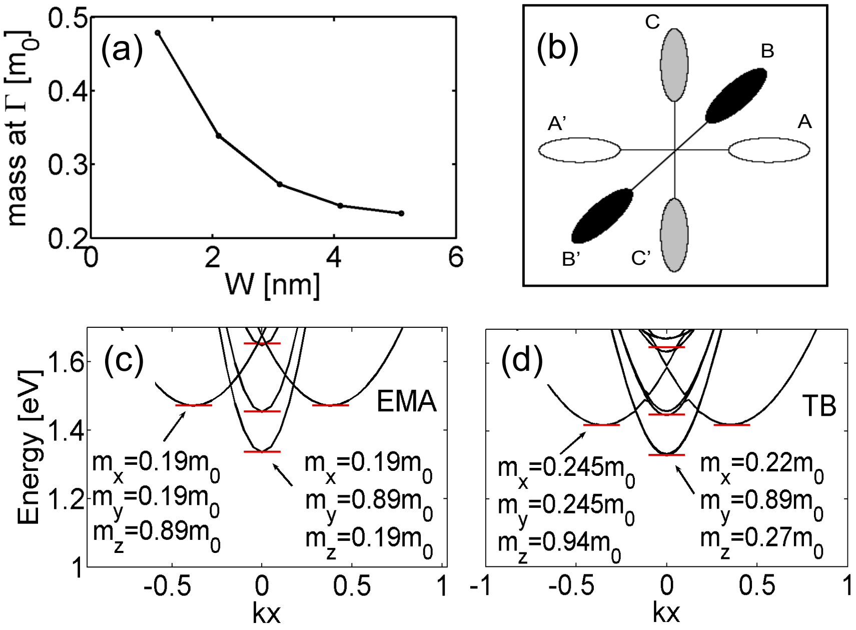


Fig. 3. 8. [100] transport orientation wire features. (a) The effective mass at the Γ point increases, as the dimensions of the wires shrink below 5nm. (b) The Si conduction band. (c) The *E(k)* for a 3nm wire in the EMA using the bulk Si masses. The 2D quantization levels are indicated. (d) The *E(k)* for a 3nm wire in the TB model. Different masses than the bulk ones are needed to calculate the 2D quantization levels. m_x and m_y denote the quantization masses, and m_z the transport mass.

reference of this potential energy is assumed to be the valence band edge of the bulk band profile. (This is arbitrary within a constant). The solutions of this equation will give the resulting bandstructure for the quantized structure according to the masses specified in the three directions.

Using the atomistic model, Fig. 3.8 (a) shows that the effective mass of NWs in the [100] transport orientation strongly depends on their diameter, which can be attributed to non-parabolicity in the Si bandstructure. Since both quantization and transport masses are affected, this will affect both the positioning of the quantized levels and the injection velocities, and will reflect on the *I-V* characteristics. To compare the TB model to the EMA, all types of ellipsoids (Γ and off-Γ) in the Si conduction band need to be included (Fig. 3.8 (b)). The transport and quantization masses used for each valley are

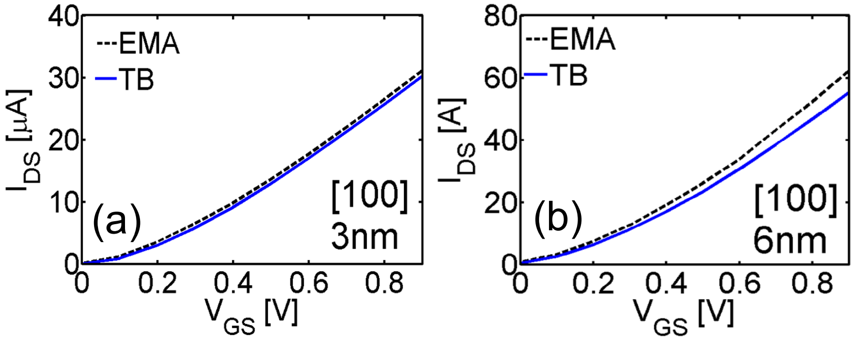


Fig. 3.9. I_D-V_G characteristics for the tuned EMA vs TB after the masses are calibrated. (a) The 3nm device. (b) The 6nm device. V_D=0.5V.

obtained from the TB model. Figure 3.8 (c) shows the *E(k)* of Si for a 3nm rectangular NW in the [100] direction. The dispersion is drawn using the bulk effective masses (m_l=0.89m$_0$ and m_t=0.19m$_0$). The in-plane pairs B and C, are shifted to k=0.41 for direct comparison to the sp^{3}d^{5}s* solution since in the EMA model all parabolas in the dispersion are centered at k=0. As shown in Fig. 3.8 (c), the subband levels agree well with the values obtained from a 2D quantization analytical calculation (horizontal lines) using the bulk quantization masses, noted m_x and m_y in the figure. In the atomistic model, however, the quantization masses are no longer the bulk masses. To map the subband levels, using the simple analytical 2D quantization formula:

$$E = \frac{\hbar^2\pi^2}{2}\left(\frac{{n_y}^2}{m_y{L_y}^2} + \frac{{n_z}^2}{m_z{L_z}^2}\right), \tag{3.3}$$

heavier quantization masses need to be used (Fig. 3.8 (d)). After the correct quantization (m_x, m_y) and transport masses (m_z) are extracted from TB, they are used in the EMA model.

After this adjustment in the masses, the I_D-V_G characteristics obtained by the two methods show very good agreement for both, small (3nm cross section) and larger (6nm cross section) [100] NWs (Fig. 3.9). In the case of the 6nm device, the masses are closer

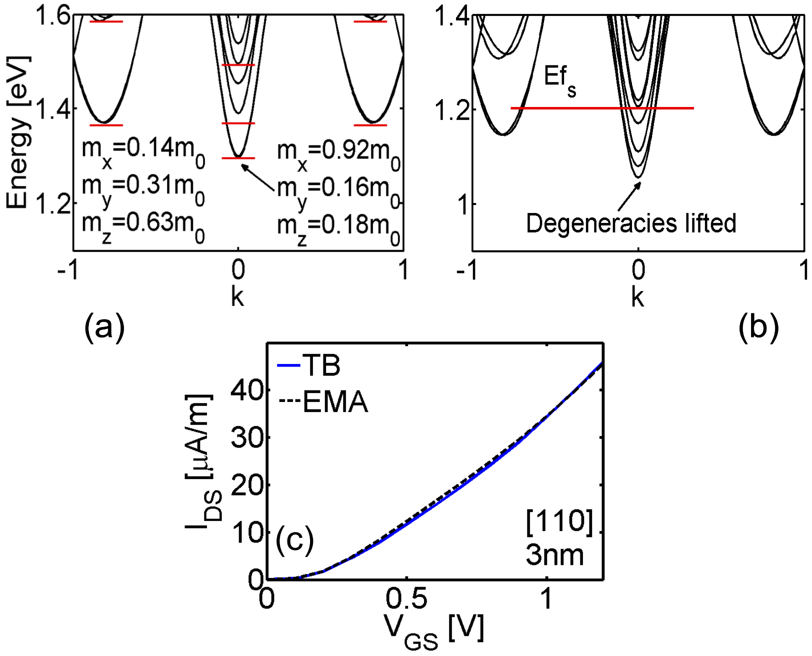


Fig. 3. 10. [110] transport orientation wire features: (a) The *E(k)* for a 3nm wire obtained from TB. The 2D quantization levels and the quantization masses are indicated (m_z is the transport mass). (b) The *E(k)* for the wire in (a) under high gate bias. V_D=0.5V. (c) The I_D-V_G for high V_D for the EMA vs. TB. m_x and m_y denote the quantization masses, and m_z the transport mass.

to the bulk values, as expected. More subbands are occupied as the device now starts to move from 1D towards a 3D device, and the interractions between them increase. The small divergence between the tuned EMA and the TB model in the 6nm case, is attributed to this different nature of band coupling between the two models, i.e. the valley splitting captured in the atomistic model, and enhanced under high biases, and the slightly different shifting of the subbands in the two models under potential variations in the lattice. Using a correctly calibrated EMA will result in large computational savings, while still including bandstructure effects to a large degree.

The [100] orientation is an example of how the EMA can successfully match with the TB model results. This is not true, however, in general. The 3nm [110] orientation dispersion shown in Fig. 3.10 (a), obtained from TB will look different than the [100] dispersion. The degeneracy in this case is 2 at Γ, and the mass is m^*=0.16m$_0$, reduced from the bulk value. The off- Γ valleys also have degeneracy of 2 each. A certain combination of quantization masses can be extracted to match the quantization levels, however, at least at Γ, once the first level is matched, the second cannot be matched accurately. To match this level, m_x=0.92m_0 and m_y=0.16m_0 are used. Other, very different combinations of masses can be used, however values similar to the bulk masses are more reasonable, especially for the heavy quantization mass which is less sensitive to structural quantization. Under self-consistent simulations, however, there is significant valley splitting in the [110] wire case, and all valleys became gradually single degenerate as more charge is introduced in the lattice (Fig. 3.10 (b)), an effect that cannot be captured in EMA. The effective mass model, under self consistency, results in different valley placement in energy, and none of the valley splitting is captured.

The overall current, however, using the TB and EMA still matches very nicely (Fig. 3.10 (c)). Energetic dispersion details might be pronounced at low temperatures and biases and the two models might not agree well. In the examples presented here, at room temperature and under high biases, the carriers are injected over a large energy range and bandstructure details are smeared out in transport calculations.

What had been described above were conduction band examples for which the EMA can be successfully implemented. However, this is not always the case. Next, the bandstructure of ultra scaled 1.5nm cross section wires is investigated and two examples in which the EMA will fail to reproduce the TB results are indicated. The NEGF [24] approach, is also used to calculate the transmission *T(E)* of the wires, still described in the $sp^3d^5s^*$ TB approximation. Figure 3.11 shows the *E(k)* and the corresponding *T(E)* for wires in the [100], [110] and [111] directions. The effect of valley splitting is particularly evident in the [110] wire case (Fig. 3.11 (b)) which makes all bands single degenerate and the *T(E)* for these wires to start at 1 quantum unit (Q.U.) per spin channel (Fig. 3.11 (e)) rather than 4Q.U. as in the case of the [100] wires. (Since we are interested in the conduction band properties of Si, we can safely ignore spin-orbit coupling and reduce the compute time without loss in accuracy). The splitting is expected to reflect on the *I-V* characteristics. Under such small cross sections, the EMA might need further adjustments to map correctly to the atomistic model and proper degeneracies need to be used. In the [111] wire case, things are different. For large dimension (>3nm) [111] wires, there are 6

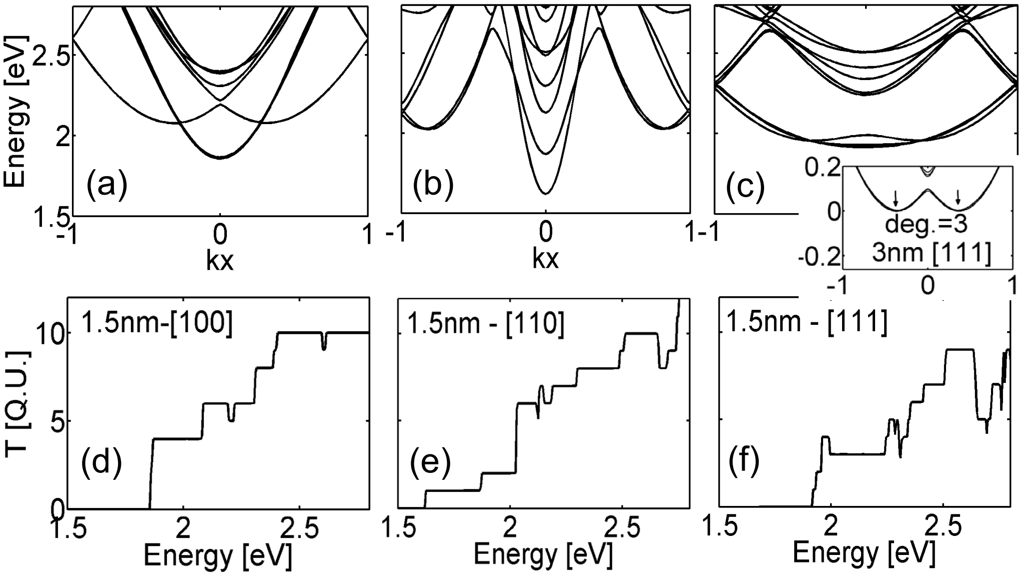


Fig. 3. 11. The effect of different transport orientations on the *E(k)* (upper row) and transmission coefficient, *T(E)*, (bottom row) of 1.5nm cross section wires. (a,d) [100]. (b,e) [110]. (c,f) [111] wire orientations. Inset of (c): The dispersion of a 3nm [111] wire.

degenerate valleys resulting from the 6 degenerate Si ellipsoids that are all quantized similarly. These are shown as (3+3) in the *E(k)* of a 3nm [111] wire in the inset of Fig. 3.11 (c). For [111] wires of 1.5nm diameter, however, at the band minima, due to interactions between the valleys, the dispersion is almost flat and there are only 3 much heavier bands now (Fig. 3.11 (c)). The *T(E)* in this case captures the three-fold degeneracy as a *T=3* Q.U. after *E=2eV*. Just before 2eV, the transmission is at *T=4 Q.U.* for a few *meV* because for that small energy region two of the subbands have not collapsed into a heavier flat band yet. Effects such as this type of band interactions are difficult to be treated in EMA.

3.6 Summary

An atomistic TB approach to calculate the electronic structure of nanowire devices self consistently with the 2D Poisson equation is described. The $sp^3d^5s^*$ TB model is used for the electronic structure calculations. Using a semiclassical model, the transport characteristics are computed. The lattice potential is calculated using a 2D Poisson solver and is self consistently included on the on-site energies of the Hamiltonian. Correctly accounting for the charge distribution in the wire affects the device characteristics at large, compared to models which only use the bandsructure features to assess the device performance. It is also shown that although atomistic description of the bandstructure is essential in nanoscale devices, the effective mass approximation can still be used in some cases, for the conduction band, after the masses are calibrated to the TB masses. This, however, is in general not possible.

In chapters 4 and 5, this model will be used to investigate the ballistic transport properties of n-type and p-type Si nanowire devices accordingly, in different transport orientations.

Finally, it is mentioned that the simulator used in this study will be released as an enhanced version of the Bandstructure Lab on nanoHUB.org [83]. This simulation engine will allow any user to duplicate the simulation results presented here and in chapters 4 and 5. Over 800 users have utilized the already existing Bandstructure Lab, which does not contain the charge self-consistent models, in the year 2007 alone.

4. BANDSTRUCTURE EFECTS IN SILICON NANOWIRE ELECTRON TRANSPORT

This chapter investigates the ballistic transport characteristics of square n-type nanowires of 3nm width, oriented in [100], [110] and [111] transport directions using the self-consistent model described in chapter 3. The results presented are based on reference [49].

The chapter is organized as follows: Section 4.1 examines the behavior of bandstructure under charge filling of the lattice for nanowires in different orientations for 3nm cross section square nanowires. Section 4.2 compares the performance of the nanowires in terms of total gate capacitance, quantum capacitance, injection velocity and drive current capabilities for nanowires in different orientations.

Section 4.3 examines how structural quantization will impact the dispersions of wires with different cross sectional areas. The valley splitting and the mass variation in wires of different cross sections are examined. In general, the degeneracies of and masses of nanowires are controlled by the orientation of the wire. Strong band coupling, however, can cause valley splitting of degenerate valleys, which can be possibly identified under low temperature and drain bias measurements [84]. Strong quantization can also result in effective mass variations. Section 4.4 provides an intuitive explanation of the dispersion mass variation using extracted subbands from the Si bulk bandstructure. Section 4.5 summarizes and concludes the chapter.

4.1 Effect of potential variations on the NW dispersion and charge distribution

Description of the dispersion in [100] oriented wires: The dispersion of a [100] oriented nanowire is shown in Fig. 4.1 (c). It has a four-fold degenerate valley at the Γ

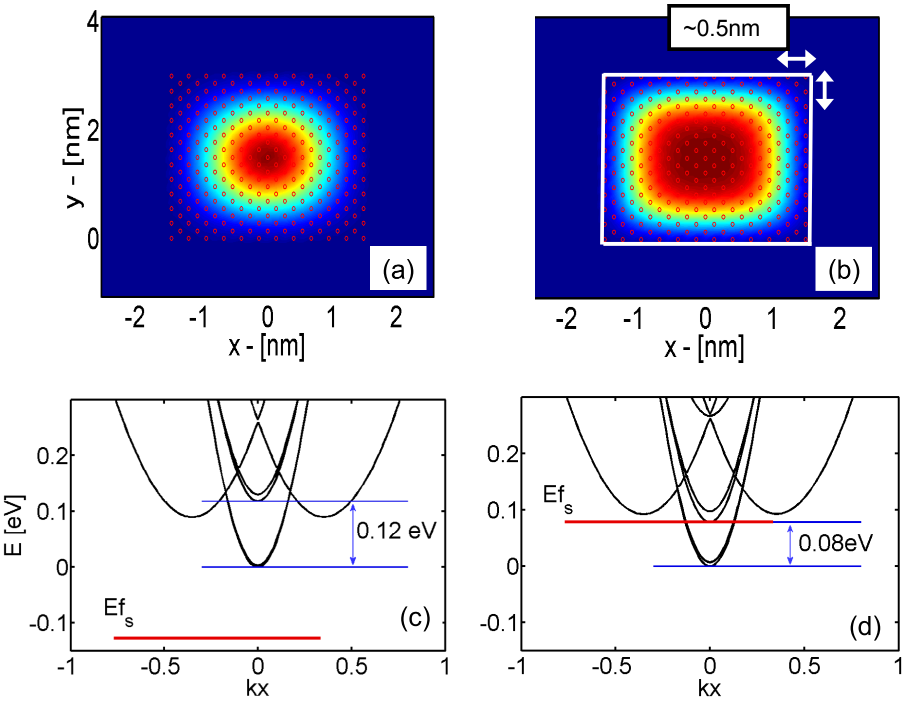


Fig. 4. 1. Device features for a 3nm [100] rectangular wire. (a-b) The 2D cross section showing the charge distribution under low and high gate biases, respectively. Even under high bias, the charge distribution is located almost half a nanometer away from the oxide. This causes degradation in the total capacitance of the wire. The dots indicate the underlying atomic positions. (c-d) *E(k)* plots for the cases (a-b). The bandstructure features change under self consistency. E_{fs} is the source Fermi level. (Zero energy indicates the conduction subband edge.)

point (k_x=0) resulting from the k-space projection of the four silicon ellipsoids that reside in the plane of quantization (here the *y-z* plane). There are two more valleys residing off-Γ (one in the positive and one in the negative k_x axis), that result from the two off-plane ellipsoids. The first four appear lower in energy because of their heavy quantization mass (m_y~m_l=0.89m$_0$ and m_z~m_t=0.19m$_0$) and have lighter transport mass (m_x~m_t=0.19m$_0$). The other two appear at higher energies because of the lighter quantization masses (m_y~m_t=0.19m$_0$ and m_z~m_t=0.19m$_0$) and have heavier transport mass (m_x~m_l=0.89m$_0$). (The wire masses m_x, m_y, m_z are close, but not exactly the bulk longitudinal and transverse masses for reasons that will be addressed later on).

Change of the [100] wire dispersion due to potential variations / charge filling: The first part of the results section investigates how potential variations in the cross section of a wire can change the dispersion and how the wavefunction shape changes as the lattice fills up with charge. Figure 4.1 shows device features for a 3nm square [100] oriented nanowire under low and high gate biases. (The drain bias used is V_D=0.5V in all cases throughout this work). Under low gate biases, the lattice is almost empty of charge (Fig. 4.1 (a)) and the dispersion relation (Fig. 4.1 (c)) is the equilibrium dispersion. Under high biases, there is significant charge filling of the lattice as shown in Fig. 4.1 (b). The charge distribution takes the shape of the underlying atomic positions. In these simulations, even under high inversion conditions, the wavefunction is pushed almost 0.5nm away from the Si/SiO_2 interface. The dispersion of this small size nanowires, on the other hand, is usually considered to be a material parameter, and under strong confinement a property of the geometry, but independent of charge filling of the lattice. It is shown, however, in Fig. 4.1 (d), that charge filling of the lattice causes changes in the dispersion of the nanowire even at the 3nm wire length scale. Here, the excited states at Γ shift down, and reside now below the off-Γ point valleys. In this case the change in the dispersion is small, but since it is associated with the wavefunction shape that gives rise to the charge distribution in the wire cross section, it can affect the devices capacitance and to some extent its transport characteristics.

Change of the [110] wire dispersion due to potential variations / charge filling: The change in the dispersion under potential variations is also observed in different wire orientations, which have different dispersion relations. The position of the bands shifts and degeneracies can also be lifted. Figure 4.2 (a-b) shows the *E(k)* of a [110] oriented nanowire under low and high biases. The dispersion looks different from the [100] dispersion, with a two-fold degenerate band at Γ, and pair of two-fold degenerate bands off-Γ. A larger variation in the dispersion under charge filling of the lattice is observed compared to the [100] wire case. The band degeneracies are lifted (from 2 to 1) by several meV. This is an effect that cannot be captured in a simple EMA treatment.

Change of the [111] wire dispersion due to potential variations / charge filling: Figures 4.2 (c,d) show the same features for a [111] oriented wire. The degeneracy of the bands of this wire is 3 (for each valley) because of the symmetry between the transport axis (or equivalently the quantization plane in the perpendicular direction) and the three pairs of ellipsoids in the Si bandstructure. High biases increase band coupling, which

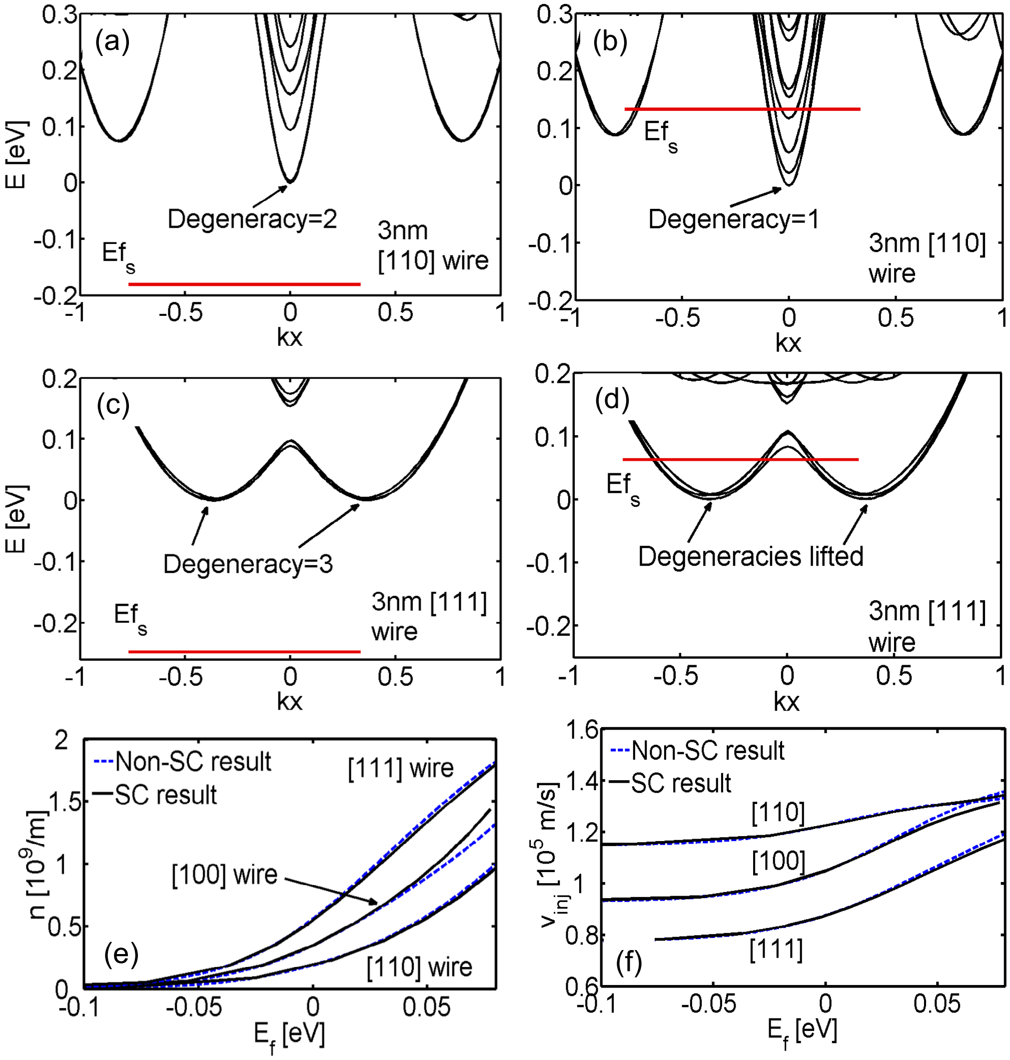


Fig. 4. 2. Self-consistent vs. non self-consistent bandstructure features: (a-b) The bandstructure of a 3nm [110] oriented nanowire under low bias (V_G=0V) (a) and high bias (V_G=0.8V) (b), and V_D=0.5V. Under high biases the degeneracies of the Γ valley are lifted from 2 to 1. (c-d) The bandstructure of a 3nm [111] oriented nanowire under low bias (V_G=0V) (c) and high bias (V_G=0.8V) (d) and V_D=0.5V. (e) The charge in the wire as a function of the difference of the conduction band edge from the Fermi level for two cases: (1) The Fermi level “scans” the equilibrium bandstructure and the charge is extracted, and (2) the charge is extracted from the self-consistent calculations with potential variations in the lattice taken into consideration. (f) The injection velocity for the same case as (e). The changes in the dispersion themselves do not reflect much on the charge distribution or the injection velocities. The differences between the two models result form the spatial information of the wavefunction that corresponds to the bandstructure changes.

slightly lifts the degeneracies. It is noted that in the case of [100] and [110] wires the conduction band minima is located at the Γ point since the quantized Δ valleys project there. In the [111] case, however, the conduction band minimum is located at 0.37 of the Brillouin zone (normalized to 1) as seen in Fig. 4.2 (c,d) for reasons explained in [85].

Charge / velocity are invariant to self-consistency: Just by looking at these variations in the dispersion, however, it is not clear that these will result in changes in the transport characteristics. Indeed, Fig. 4.2 (e,f) compares the density of states and velocities at the same E_f-E_c (difference of the Fermi level from the conduction band edge) between the equilibrium dispersion and the dispersion at various biases and little difference is observed. Quantities for two cases are calculated: (a) The Fermi level "scans" the equilibrium bandstructure and the charge and injection velocities are extracted, and (b) the results are extracted from the self-consistent calculations with potential variations in the lattice taken into consideration. The charge and injection velocity is plotted as a function of E_f. (E_c is shifted to zero for all wires). There is no significant difference in these extracted quantities due to the potential variations, and the self-consistent vs. non-self-consistent curves fall almost on top of each other. For this example a large drain bias (V_D=0.5V) is used. Under low drain biases (V_D=1meV) and low temperatures, however, where the transport energy window can be comparable or even smaller than the changes in the bandstructure, evidence of the bandstructure differences in these two quantities as well as other quantities such as the transconductance are more likely to appear.

Charge distribution is strongly dependent on self-consistency: Although the charge and velocity appear to be only weakly modified by the self-consistent calculation, the self consistently extracted bandstructure corresponds to a different wavefunction shape which reflects to a different charge distribution in space. This is the quantity that causes degradation of the total gate capacitance as will be shown later and affects the transport characteristics, and not the dispersion changes by themselves. One therefore, has to also consider the change in the wavefunction that is associated with the dispersion changes. (As shown in chapter 3 and in [48], the current-voltage characteristics can be significantly overestimated if the spatial variation of the charge is not considered).

Orientation differences in the charge: The fact that the charge in Fig. 4.2 (e) for any position of the Fermi level is always the highest in the [111] wire case, is due to the higher density of states and valley degeneracy. This particular wire orientation has the valleys with the heaviest mass ($0.47m_0$, where m_0 is the free electron mass) and the largest degeneracy (D=6). Therefore, at a certain energy level (E_f-E_c), there are more states occupied compared to the other wires. The [100] wire with mass $0.27m_0$ and D=4 of the lowest valleys, follows. The [110] wire has the lowest charge density at a certain energy level because of its lighter mass ($0.16m_0$) and lower degeneracy (D=2) at Γ.

Orientation differences in the velocity: The reverse trend is observed in Fig. 4.2 (f), where [110] wire has the highest velocity due to its lighter mass ($0.16m_0$). As higher k-states are occupied, the velocity increases since it is proportional to the slope of the bands. Noticeable here, is the fact that the carrier velocity in the [100] wire approaches the [110] velocity as the Fermi level is pushed into the conduction band. The lighter masses ($0.16m_0$) of the two-fold Γ valleys in the [110] wire give an initial advantage over the heavier ($0.27m_0$) [100] wire Γ valley masses. Once the heavier four-fold degenerate off-Γ valleys (with mass $0.61m_0$) of the [110], and the heavy two-fold degenerate off-Γ valleys (with mass $0.94m_0$) of the [100] start to populate, the carrier velocities become comparable in the two cases. The exact reasons why the masses have these values will be addressed later on in the paper, however this analysis can guide through the reasons why wires in different transport orientations have different properties.

4.2 Device performance comparison of NWs in different orientations

One of the points made in the previous paragraph, are comparisons of the different wire orientations at the same Fermi level position into the dispersion of the wires. Although this is a rough estimate of the wires' properties, the Fermi level is not at the same position for all devices, except under special cases. In this section, the full self consistent model is implemented to compare the performance of the nanowires. Figure 4.3 shows a performance comparison between the wires in the [100], [110] and [111] orientations. The various performance quantities shown further on, are all compared at the same OFF current (I_{OFF}) for all devices.

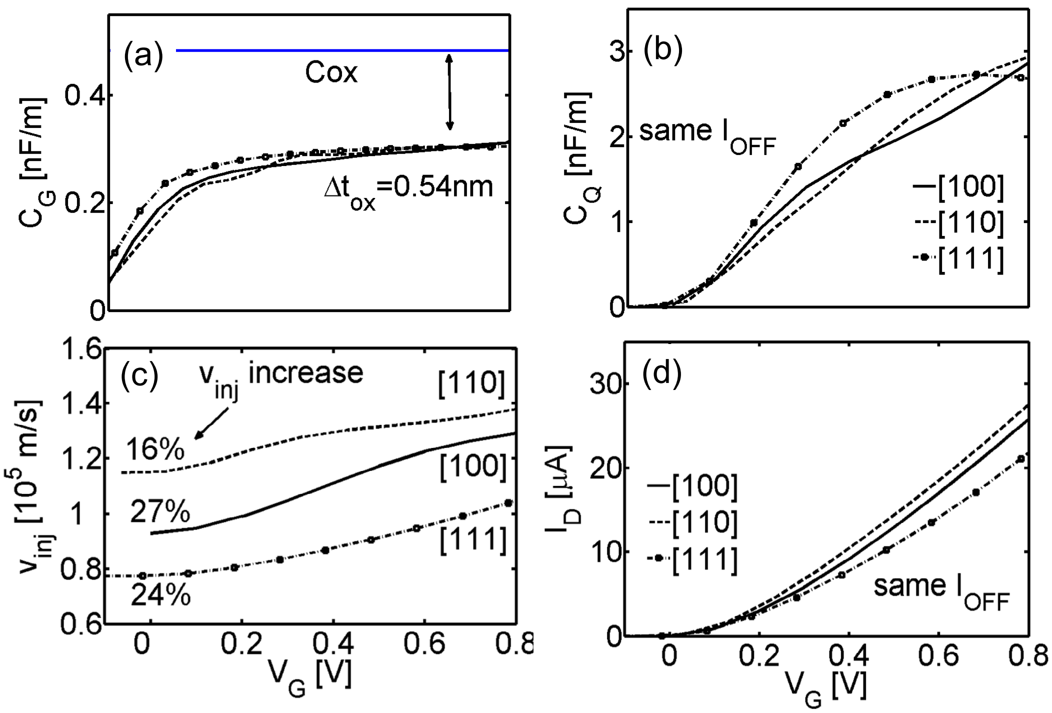


Fig. 4. 3. Performance comparison of the 3nm square wires in the [100], [110] and [111] directions at the same OFF-current (I_{OFF}). (a) The gate capacitance C_G vs. gate bias (V_G). The capacitance is similar for all wires, and degraded from the oxide capacitance ([86]) by an amount that corresponds to an increase in the effective oxide thickness of 0.54nm. (b) The quantum capacitance C_Q vs. V_G of the three devices, which is a measure of the density of states at the Fermi level. (c) Comparison between the injection velocities of the nanowires vs. V_G. In all cases, the velocity is not constant, but increases as the gate bias increases. The increase is calculated by the difference between the value at high V_G and the value at low V_G. (d) The I_{DS} vs. V_G for the three wires at the same I_{OFF}. The velocity difference directly reflects on the current differences.

Gating induces same capacitance / charge in all wire directions: Figure 4.3 (a) shows the total gate capacitance (C_G) ([86]) vs. gate bias (V_G) of the three wires at the same I_{OFF}. The total capacitance in the three wires is very similar for all gate biases for reasons we will explain later on. However, this is an indication that the same amount of inversion charge is accumulated in all wires irrespective of their orientations. Our calculation supports this argument too, showing that the charge difference between the wires at high inversion does not exceed 2%. In a relative performance comparison for wires in different orientations, therefore, the amount of charge will not affect the relative performance.

Low semiconductor capacitance (C_S) degrades the gate capacitance (C_G) by 30%: It is important to notice that for all three wire cases, the capacitance value is degraded from the oxide capacitance by almost 30%. This is an amount that corresponds to an effective increase in the oxide thickness of 0.54nm, which is 50% of the physical gate oxide thickness (t_{ox}=1.1nm). This large gate control reduction is evidence of low semiconductor capacitance (C_S) in low dimensional channels. The gate capacitance of a device is the series combination of the oxide capacitance (C_{OX}) and the semiconductor capacitance (C_S) given by the simple relation $C_G = \frac{C_S C_{OX}}{C_S + C_{OX}}$. For an electrostatically well behaved MOSFET device, C_S should be an order of magnitude larger than C_{OX} so that the C_G and therefore the charge in the device is totally controlled by the gate. In this example, the oxide capacitance of the rectangular structure is 0.483nF/m, numerically calculated using a 2D Poisson solver that takes the fringing at the edges into consideration. With C_G = 0.3nF/m (maximum value of Fig. 4.3 (a)), C_S can therefore be computed to be $C_S = 0.8nF/m$, which is only twice the value of the oxide capacitance (less than an order of magnitude difference).

Cs controlling factors: Charge distribution peak, small C_Q: C_S is defined as the differential of the charge in the device with respect to the surface potential (ψ_S). In 1D systems, under a single band effective mass approximation, the charge is the integral of the 1D density of states (g_{1D}) convoluted with the Fermi function (*$f(E_f$-$E)$*) over energy as:

$$C_S = \frac{\partial(qn_s)}{\partial \psi_s} = \frac{\partial}{\partial \psi_s}\left(\int qg_{1D} f[(E_f - E_c - \varepsilon_i)/k_B T]dE\right) \tag{4.1}$$

where q is the charge of the electron, ψ_s is the surface potential, E_f is the Fermi level, E_c is the conduction band edge and ε_i is the distance of the i^{th} quantized subband above E_c in energy. Carrying on the integration, the equation above results in:

$$C_S = \frac{q^2}{\pi}\left(\frac{2m}{\hbar^2}\right)^{1/2} \sqrt{k_B T}\sqrt{\pi}\,\frac{\partial}{\partial q\psi_s}\left(\Im_{-1/2}[(E_f - E_c - \varepsilon_i)/k_B T]\right) \tag{4.2}$$

$$= q^2\left(\frac{2m}{\pi\hbar^2}\right)^{1/2} \sqrt{k_B T}\ \Im_{-3/2}[(E_f - E_c - \varepsilon_i)/k_B T]\left(1 - \frac{\partial \varepsilon_i}{\partial \psi_s}\right) \tag{4.3}$$

$$= C_Q\left(1 - \frac{\partial \varepsilon_i}{\partial \psi_s}\right) \tag{4.4}$$

The first part of Eqn. 4.3, C_Q, is the quantum capacitance, which is a measure of the density of states at the Fermi level. C_S is degraded from C_Q by a factor that is proportional to how much ε_i (the difference of the i^{th} subband to E_c) changes. Ideally, at high inversion conditions ε_i should be constant, meaning that the quantized levels and E_c shift by same amount and the subband levels can easily get in the potential well that forms at the Si/SiO$_2$ interface. This directly translates on the wavefunction been able to come closer to the interface as the surface is inverted more and more. However, ε_i can float up as charge accumulates in the device, giving rise to the differential term in Eqn. 4.4, and the wavefunction stays away from the interface. As shown earlier on in Fig. 4.1 (b), this shift is almost 0.5nm. Other than the wavefunction shift, C_Q being small is the second degrading factor of C_S as indicated in Eqn. 4.4. Figure 4.3 (b) shows the C_Q of the three nanowires as a function of V_G, calculated as the density of states at the Fermi level. Clearly, for all wires the maximum value is below 3nF/nm, not even an order of magnitude above C_{OX} = 0.48nF/nm. The fact, that the position of the charge distribution degrades C_S from C_Q by almost four times, ($C_S = 0.8nF/m$), indicates its large significance on the device's capacitance. (Similar deviations of the semiconductor capacitance from the quantum capacitance have also been observed in thin body devices [87]).

Variations in C_Q between different wire orientations: As shown in Fig. 4.3 (b), in all wire cases, C_Q is not constant, but undergoes large transitions as the Fermi level is pushed inside the subbands at large gate biases. This is expected, since C_Q is a measure of the density of states at the Fermi level, and the differences in the dispersion cause differences in C_Q. Comparing C_Q for different wire orientations, the [111] wire has the largest C_Q for most of the bias range because of the higher mass ($m^*=0.47m_0$) and higher degeneracy of its valleys ($D=6$). The C_Q drop at high biases in the [111] case is associated with the decreasing 1D density of states away from the band edges, and due to the fact that its bands flatten out at Γ and do not extent as parabolic bands in *k*-space as shown in Fig. 4.2 (d). On the other hand, the [100] and [110] wires initially have lower C_Q, because of their lower density of states (lighter masses and lower degeneracies). At high biases, the upper valleys of the [100] and [110] wires start to get populated, which allows a continuous increase in C_Q for these wires. More specifically, since the charge in all cases is almost the same at a given bias, the same number of states in each wire need to be occupied. The Fermi level in the [110] wire with lower mass and smaller valley degeneracies reaches the upper valleys faster (at a lower gate bias) than the [100] wire in order to occupy the same number of states. Once this happens, the C_Q of the [110] wire surpasses the C_Q of the [100] wire (around V_G=0.4V).

Variations in C_Q do not cause variations in C_G: The differences in C_Q, between wires in different orientations, however, are not large enough to result in differences in the total capacitances. As seen earlier, C_Q is only partially responsible for the total capacitance degradation. The small differences in C_Q are smeared out in C_G by the oxide capacitance, and the charge shift from the interface, that is very similar for all the above wires. (This observation can of course be different in the case of high-k dielectric oxides, in which the importance of C_Q can be more pronounced).

Velocity controls the transport differences in different orientated wires: As explained above, the charge is almost the same in all three nanowires. Since in the ballistic limit the ON-current performance is given by the product of "charge *times* velocity", if the charge is the same, any performance differences will result from differences in the carrier velocities. Figure 4.3 (c) shows the injection velocities of the wires vs. gate bias (V_G). The [110] wire has the largest velocities whereas the [111] wire has the lowest velocities. In all cases, the injection velocities are not constant, but increase as the lattice is filled with charge because faster high energy carrier states are being populated. This increase in velocities, calculated form the initial value at low gate biases to the final value at high gate biases can reach up to 17% in the [110] wires and even up to 27% in the [100] and 24% in the [111] wire orientation cases. When comparing the velocities of the different wires, however, the masses of the valleys determine the velocities of the carriers. (In 1D, under the parabolic band approximation, the velocity is proportional to $v \sim 1/\sqrt{m^*}$). As a result, the [110] wire with m^*=0.16m$_0$ has the highest velocity, followed by the [100] wire with mass m^*=0.27m$_0$, and finally by the [111] wire of mass m^*=0.47m$_0$. The larger density of states of the [111] wire and its larger degeneracy do not allow the Fermi level to be pushed far into the conduction band. Therefore, only the lower energy and slower carries are used, and the velocity in this case is low. In the [110] wire case, the degeneracy is 2, and the subband density of states low, therefore the Fermi level will be pushed far into the conduction band, and faster carries will be utilized as shown in Fig. 4.3 (c).

Velocity differences affect the I-V differences: The velocity difference directly reflects on the I_{DS} as shown in Fig. 4.3 (d) in which the drive current capabilities of the wires are compared at the same I_{OFF}. The [110] and [100] wires perform better than the [111] wire in terms of ON-current capabilities. The current in the [110] wire stands ~5% higher than the [100] wire and ~20% higher than the [111] wire because of its lower mass. This result must be qualified since the bandstructure of the wires is a very sensitive function of their quantization. The results presented here are for these specific 3nm wire examples. In cases where important dispersion parameters such as the relative placement of the valleys in energy, masses and degeneracies, are altered, different conclusions might be drawn, especially for the relative performance of the [100] and [110] wires which is not that large. In the next section, an analysis is performed on how exactly these parameters (valleys splittings that lift degeneracies, and masses) are affected by quantization.

4.3 Quantization influence on valley splitting and mass variation

Quantization strongly affects both factors that control the performance, the degeneracies and masses. In this section of the paper, the effect of quantization on these parameters is examined. Degeneracies are controlled mainly by the orientation, but can be lifted due to valley splitting [88-90] under strong quantization (both electrostatic and structural).

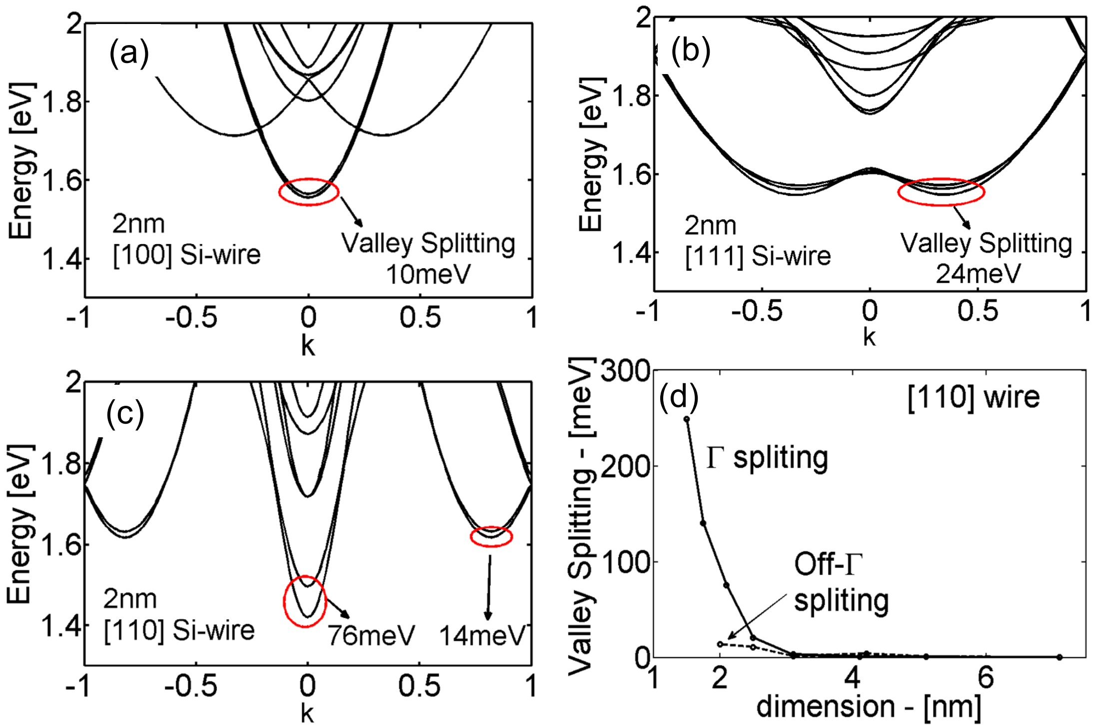


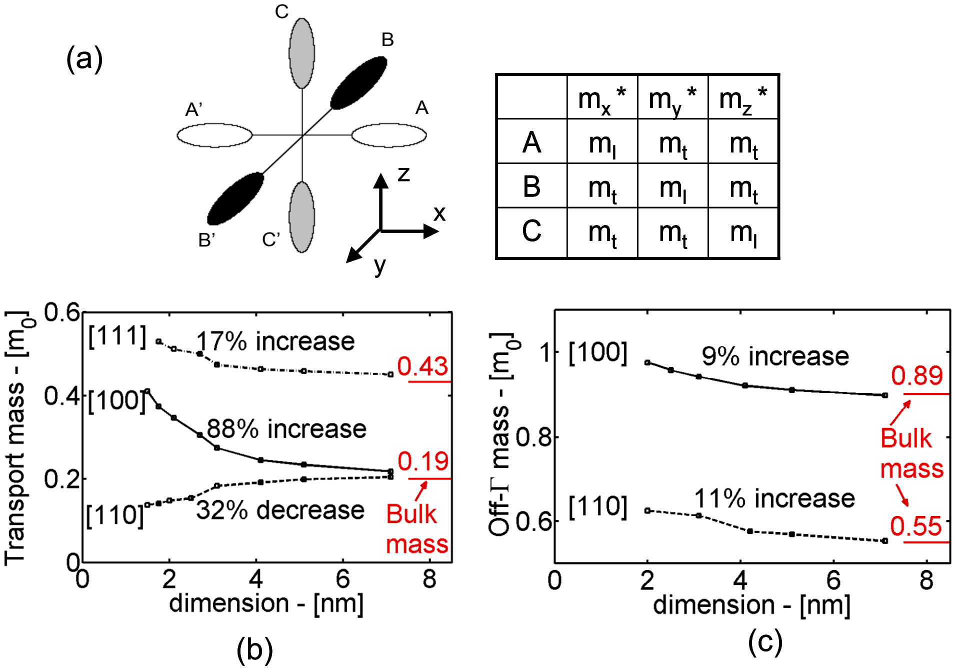
Fig. 4. 4. The effect of valley splitting in small nanowires. (a-c) The *E(k)* of a 2nm wire in the [100], [111] and [110] orientations respectively. (d) The effect of valley splitting in the [110] wire as the dimensions decrease. The Γ valleys are severely affected at cross sections below 3nm, whereas the off-Γ valleys are not affected as much.

Weak valley splitting in [100] and [111] quantized wires: Figure 4.4 (a,b) shows the *E(k)* of a 2nm wire in the [100] and [111] orientations. A slight valley splitting of the degenerate valleys under quantization is observed. In the case of the [100] the splitting is 10meV and in the case of the [111] wire, 24meV. These values are less than the room temperature k_BT = 26meV and are not expected to have a significant effect in the transport properties of the nanowires at room temperature [84].

Strong valley splitting in [110] quantized wires: In the case of [110] nanowires, valley splitting is significantly larger. As shown in Fig. 4.4 (c) in the *E(k)* of a 2nm [110] wire, Γ and off-Γ valleys experience valley splitting of their degeneracies by 76meV and

14meV respectively. Figure 4.4 (d) shows how this effect varies with the spatial confinement in the [110] wire. Although large nanowires (>5nm) are not affected, the valleys splitting can reach up to 200meV for the Γ valleys of narrow wires with sizes as narrow as 1.5nm. The valley splitting of the off-Γ valleys, on the other hand, is not affected as much. Only a few tenths of meVs of splitting are observed in this case. (It is noted here that the splitting in the other wire orientations is smaller than the [110] wires of similar quantization sizes even for wires below 2nm [90]).

Generally, masses increase with increase in quantization: The effective mass is the second important transport performance dispersion property that is affected by quantization of the nanowire cross section. The injection velocity and quantum capacitance strongly depend on the masses. Both the quantization and the transport masses of nanowires under arbitrary wire orientations are certain combinations of the longitudinal (m_l=0.89m$_0$) and the transverse effective masses (m_t=0.19m$_0$) of the Si ellipsoids. Figure 4.5 (a) shows the three pairs of ellipsoids that form the conduction band minima in Si, each characterized by the *x*, *y* and *z* directional masses. The masses of the valleys that appear in the nanowire dispersion are automatically included in tight-binding. What will be shown is that under quantization, the exact values of these masses are changed from their bulk values. In most cases, quantization results in an increase in the effective mass. Figure 4.5 (b) shows the variation in the lowest valley transport masses as the dimension of the wire cross section reduces. At large wire cross sections, the mass of the [100] valley that is located at Γ, approaches the bulk transverse mass m_t=0.19m$_0$. The bulk mass of the [111] wire is larger since it is a combination of m_t=0.19m$_0$ and m_l=0.89m$_0$ (the bulk value is 0.43m$_0$) [36, 91]. The mass in the [100] case almost doubles as the dimension of the wire's side decreases from 7.1nm to 1.5nm (88% increase). (The 3nm wire has m^*=0.27m$_0$ as mentioned earlier). The corresponding increase in the [111] wire's mass is 17%, with the 3nm wire having m^*=0.47m$_0$. The off-Γ valley masses (upper valleys) of both [100] and [110] wires also increase as the dimension reduces as shown in Fig. 4.5 (c). In the [100] off-Γ valley case, a slight mass increase of 9% between the 7.1nm and the 1.5nm is extracted from the bandstructure calculations. The off-Γ valley mass increase in the [110] case is 11%.



	m_x*	m_y*	m_z*
A	m_l	m_t	m_t
B	m_t	m_l	m_t
C	m_t	m_t	m_l

Fig. 4. 5. Mass variation in nanowires: (a) The three equivalent pairs of ellipsoids in the conduction band of Si are described by the longitudinal and transverse masses. A combination of these masses results in the quantization and transport masses of nanowires under arbitrary orientations. (b) The transport masses oriented in [100], [110] and [111] vs. wire dimension as calculated from TB. At large wire cross sections, the [100] and [110] that are located at Γ, approach the bulk m_t=0.19m$_0$. The mass of the [111] wire is larger since it is a combination of m_t and m_l=0.89m$_0$. As the wire dimensions shrink, the mass of the [110] wire reduces, whereas the masses of the other two wires increase. (c) The off-Γ valley masses for the cases of the [110] and [100] wires. Both increase as the dimensions decrease. (The expected bulk mass values for every orientation are denoted). The percentage change denoted is the change in the effective masses between the 1.5nm mass value (mostly scaled wire) and the 7.1nm wire.

[110] wire Γ valley masses decrease with increase in quantization: In contrast to the rest of the valleys, the Γ valley mass of the [110] oriented wires decreases with increase in quantization. As shown in Fig. 4.5 (b) the mass decreases by 32% as the side of the wire reduces from 7.1nm to 1.5nm. As mentioned earlier, the mass of a 3nm [110] wire is m^*=0.16m$_0$, which gives an enhanced injection velocities and transport characteristics of [110] wires over the rest of the wires. Anisotropy and non-parabolicity

in the Si conduction band Brillouin zone cause this unintuitive behavior as explained in the next section.

4.4 Understanding the nanowire mass variation as a function of quantization

Semi-analytical construction of the wire's dispersion: This distinctly different observation in the masses of wires is a result of the non-parabolicity and anisotropy of the Si bandstructure. Under any physical quantization, the subband levels will follow the "particle in a box" quantization, as shown in Fig. 4.6 (a). The smaller the physical domain, the larger the corresponding quantized *k*, and the higher the energy levels of the subbands. To estimate the quantization levels of the Si conduction band ellipsoid quantized along the longitudinal direction, the energy contour in the *x-y* plane near the band minima is plotted in Fig. 4.6 (b). ("Cut" through the ellipsoid along its longitudinal axis). Similarly to Fig. 4.6 (a), quantization of *Lx* of 2nm, 3nm, and 5nm will shift the energy levels to the vertical lines shown in the figure. The energy levels at these lines will be the relevant subbands in an ultra-thin-body (UTB) quantization – with one quantized dimension. Figure 4.6 (c) now, shows the energy contour taken at the 3nm line, perpendicular to the contour of Fig. 4.6 (b) in the *y-z* plane. An extra quantization in the *z*-direction (the second quantized dimension, as in the wire case) will leave only one allowed *k*-space variable, the transport direction one. This forms the 1D dispersion of the wire. The relevant 1D bands are the ones located at the horizontal lines of Fig. 4.6 (c). Lines for Lz=2nm, 3nm and 5nm are shown. The solid line indicates a relative subband for an UTB device with *Lx*=3nm and $Lz = \infty$ (k_z=0, only one quantization dimension).

Mass and band edge extraction from the semi-analytical construction: The 1D subbands of Fig. 4.6 (c) are plotted in Fig. 4.6 (d) for the cases of *Lz*=2nm, 3nm and 5nm. (The *x*-direction quantization is *Lx*=3nm in all cases). The mass of these bands is the transport mass (*y*-direction) that the wire has in the [010] orientation (equivalent to the [100] wire orientation described in the previous sections). Smaller cross sections raise the subband energy, and increase the masses. Through this process, both, the transport masses and the placement of the subband edges in energy can be deduced. From the subband edges the quantization masses can be extracted. The more non-parabolic the

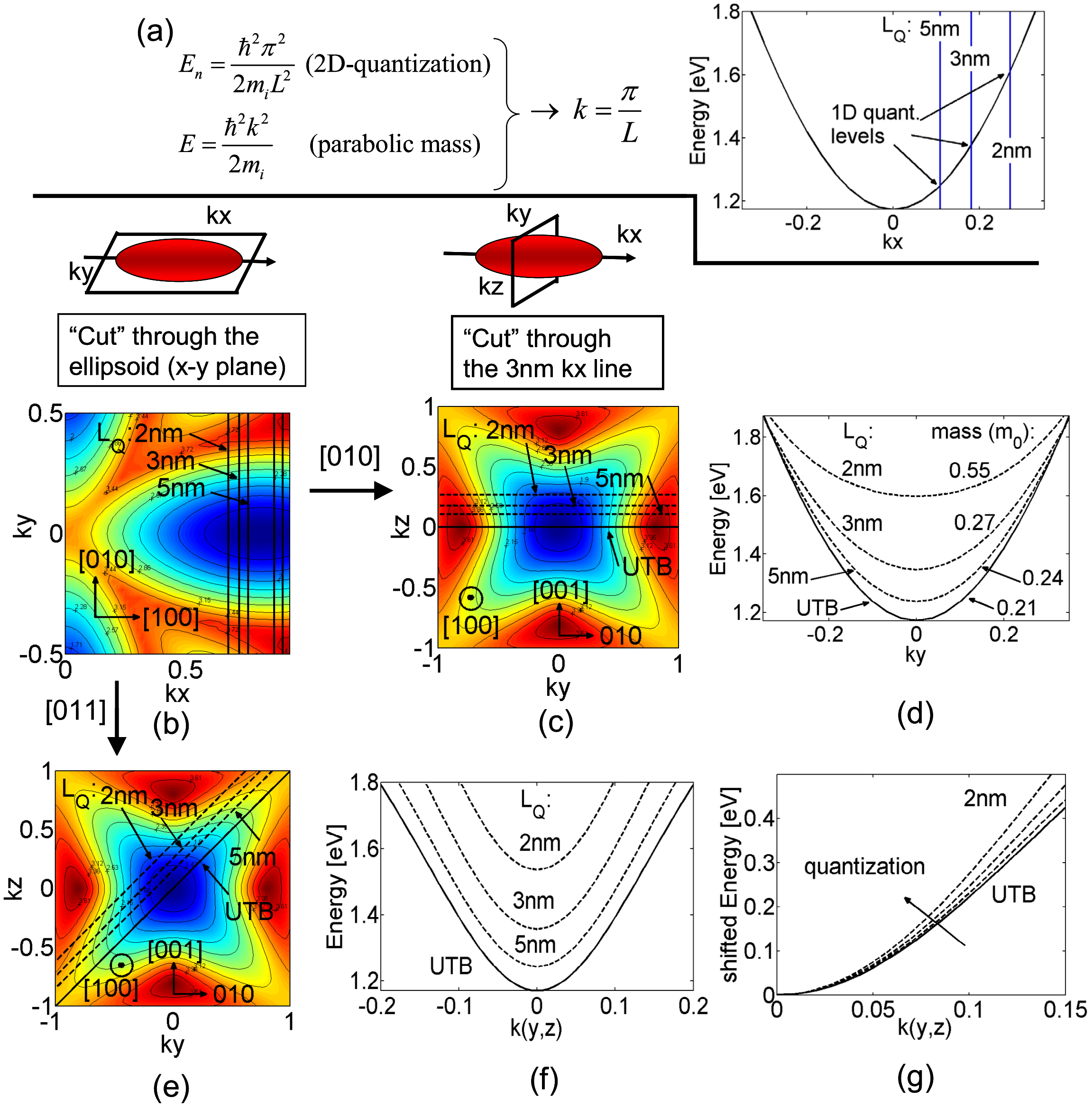


Fig. 4. 6. Semi-analytical understanding of the nanowire's dispersion: (a) The energy levels of a quantized structure using the "particle in a box" picture. Under quantization, the subband edges and masses can be deduced from the materials' bulk dispersion with a numerical *E(k)* diagram. (b) Energy contour at the middle of one of the Si Brillouin zone ellipsoids calculated using the full 3D k-space information of the Si Brillouin zone. A "cut" through the Si ellipsoid along its longitudinal axis is shown. Under quantization in *Lx*=2nm, 3nm and 5nm, the relevant subband energies are indicated by the vertical k_x=constant lines. (c) A "cut" through the Si ellipsoid perpendicular to its longitudinal axis at the k_x line corresponding to the 3nm quantization line of (b). The non-parabolicity and anisotropy is evident in this figure. The horizontal lines indicate the relevant energy regions under another quantization in the y-direction for *Ly*=2nm, 3nm and 5nm quantized structures. The solid line labeled UTB is the relevant band for an ultra-thin-body (UTB) device of *Lx*=3nm thickness in [001] with $Lz = \infty$. This is only quantized in

the *x*-direction. (d) The dispersions of the vertical lines in (c). The masses and the band edge of the dispersions will be the ones that appear in a quantized wire. (e) The 2D plot is the same as in (c). The 45° lines correspond to a quantization in the [0-11] for *Lyz*=2nm, 3nm and 5nm. The solid line labeled UTB is the relevant band for an ultra-thin-body (UTB) device of *Lx*=3nm thickness in [001] with $Lyz = \infty$. (f) The dispersions of the 45° lines in (e). The non-parabolicity is evident in this orientation. (g) Zoom of the right (positive momentum) branch of (f) with all dispersions shifted to the origin for comparison. As the structure is quantized in [0-11], the mass becomes lighter. The anisotropy in the Brillouin zone is directly reflected on the masses in the different wire orientations (as in Fig. 4.5 (b)).

bulk bandstructure is at higher energies in the direction of quantization, the slower the subbands rise in energy with quantization compared to the parabolic band case. This results in larger quantization masses. The more non-parabolic the bulk bandstructure is in the transport direction, the larger the transport masses will be. All these effects appear in thin body channel devices (UTB of Fig. 4.5 (d)), however, they are significantly more enhanced in the case of nanowires because of the extra quantization of one more physical dimension [37, 38].

Different orientations, different anisotropies: The transport masses of wires in other orientations can be explained similarly. Evident in the bandstructure is the anisotropy which results in different behavior in the quantization of the [100] to quantization of the [110] axes. In [110] oriented wires, the [100] and [0-11] directions are quantized. The [100] quantization is the same step as the one in Fig. 4.6 (b). Quantizing the [0-11] direction, will result in extracting 1D bands by lines that cross Fig. 4.6 (d) at 45°, (in *Lyz*, instead of horizontal). Figure 4.6 (e) shows the first subband of the dispersions of structures with *Lx* = 3nm and *Lyz* = 2nm, 3nm, 5nm, similar to Fig. 4.6 (c). Evident in this case is the non-parabolicity of the dispersion, as it is also evident in Fig. 4.4(c). For comparison purposes, Fig. 4.6 (g) shows the positive k_{yz} branch of the dispersion, with all the bands shifted to the origin. Clearly, as the structure is quantized in the [0-11] direction, the curvature of the dispersion increases, corresponding to a lowering of the transport mass of the wires. In contrast to the [001] quantization case of Fig. 4.6 (c), here the anisotropy in the bandstructure results in a reduction of the transport masses with increase in quantization, in agreement with the calculation for the actual nanowire mass shown in Fig. 4.5 (b). The magnitude of the mass variation is however smaller in the [0-11] quantization direction compared to the [001] direction. (Similar anisotropic results have been also obtained using empirical non-local pseudopotential and ab-initio GW calculations [46]).

Limitations of the semi-analytical construction: This construction method can provide a rough guidance as to what the dispersion of a nanowire will look like. The method, however, does not include any of the interactions between the bands/valleys (which are enhanced when the material is physically confined in a nanowire), and lacks any band coupling information. Effects such as valley splitting, that are a consequence of band coupling, cannot be captured. The extracted mass values, as well as their variation trends under quantization, are however quite accurate. In the case of nanowire electronic transport for nanowires larger than 3nm, where the mass is an important transport parameter, a first order estimation of the nanowires' performance can be drawn by using this analytical mass extraction.

4.5 Conclusions

Transport properties of nanowires in different transport orientations ([100], [110] and [111]) were examined using a 10 orbital $sp^3d^5s^*$ atomistic TB model self consistently coupled to a 2D Poisson solver. A semiclassical ballistic model was used to calculate the current-voltage characteristics of the nanowires. The dispersions of the nanowires undergo changes under gate bias, which at some cases can cause large lift of degeneracies and small subband shifts. Although these changes under self-consistency do not alter the velocity and density of states of the wires, they are associated with the spatial distribution of charge that together with the small 1D density of states can degrade the nanowire's capacitance by 30%. The quantum capacitance of the different oriented 3nm wires that were investigated is a strong function of gate bias, but of similar magnitude in all wires. Almost the same is also the total gate capacitance of all nanowire devices in different orientations investigated as well as the inversion charge. Due to their lighter mass, 3nm [110] oriented wires have the maximum injection velocities, whereas [111] oriented wires the lowest injection velocities due to their higher masses. The injection velocity reflects directly on the current capabilities of the wires, where the [110] and [100] oriented wires indicate the best performance in terms of ON-current capabilities compared to the [111] wires which are the worst.

The masses of the wires are a sensitive function of the wire dimensions (below 7nm), and strongly influence the output performance of nanowire devices. This is an effect that resides in the non-parabolicity and anisotropy of the Si Brillouin zone that is particularly important in strongly quantized devices. Valley splitting is another effect strongly dependent on quantization. [110] nanowires of dimensions below 3nm are extremely sensitive to this.

5. BANDSTRUCTURE EFFECTS IN SILICON NANOWIRE HOLE TRANSPORT

This chapter investigates the ballistic transport characteristics of square p-type nanowires of 3nm and 6nm width, oriented in [100], [110] and [111] transport directions using the self-consistent model described in chapter 3. This chapter is a continuation of chapter 4, for p-type nanowires and it based on [50]. The $sp^3d^5s^*$ atomistic tight-binding model self-consistently coupled to a 2D Poisson solver is used as described in chapter 3.

The problem of identifying the correct bandstructure for the valence band of Si in the inversion layers is much more complicated than the corresponding conduction band counterpart because of the strong non-parabolicity and anisotropy of the heavy-hole. *k.p* methods have been traditionally used [92, 93] for both un-strained and strained channels. De Michielis recently in [94] described the valence band through semi-analytical non-parabolic and anisotropic description calibrated to *k.p*. Our work, shows that the valence band problem is exceedingly more complicated in the case of nanowires. Here, enhanced coupling due to structural quantization, and large sensitivity of the band shape/curvature to potential variations in the lattice, especially in the [100] wire orientation, make the effective mass approximation (EMA) and the *k.p* method insufficient, and call for atomistic treatment [26, 34, 35]. (This is in contrast to the variations of the curvature of the conduction bands which show much less sensitivity to lattice potential variations).

The chapter is organized as follows: Section 5.1 examines the behavior of bandstructure under charge filling of the lattice for nanowires in different orientations for 3nm and 6nm cross section square nanowires. The anisotropy of the Si heavy-hole valence band strongly affects the preference of charge placement in the wires' cross section and differs for different oriented wires. The heavy quantization masses in the

[110] and [112] orientations cause preferable charge accumulation on the (110) and (112) surfaces rather than the (100) surfaces controlled by a lighter quantization mass.

Section 5.2 compares the performance of the nanowires in terms of total gate capacitance, quantum capacitance, injection velocity and drive current capabilities for nanowires in different orientations. The semiconductor capacitance C_S, as also in the case of NMOS nanowires [49], is important and can degrade the capacitance of the device by up to ~30% for wires in all transport orientations. In terms of ON-current capabilities, p-type nanowire transport will be preferable in the [111] oriented devices that have the largest carrier velocities closely followed by the [110] devices. [100] nanowires indicate much lower performance due to the enhanced band warping and curvature variations in their dispersions.

Section 5.3 provides an intuitive explanation of the dispersion shapes, curvatures and inversion charge distribution features under quantization in 2D and 1D for different orientation p-type devices by extracting energy surfaces from the 3D bulk bandstructrue.

5.1 Effect of potential variations on the NW dispersion and charge distribution

In this section, an analysis of the transport properties of 3nm and 6nm square nanowires in [100], [110] and [111] orientations is performed. Strong interactions between the nanowire valence bands, as well as potential variations introduce non-trivial features in the dispersions. Their dispersion properties (in terms of carrier velocities, and quantum capacitance) and the shape of the charge distribution in the cross section of the each wire are analyzed. The relative performances are then compared.

5.1.1 Hole transport in [100] oriented nanowires

The 3nm [100] wire: Figure 5.1 (a-d) shows device features for a 3nm square [100] oriented nanowire under low (V_G=0V) and high (V_G=1.2V) gate biases. (The drain bias used is V_D=0.5V in all cases throughout this work). Under low gate biases, the lattice is almost empty of charge (Fig. 5.1 (a)) and the dispersion relation (Fig. 5.1 (b)) is almost the empty lattice equilibrium dispersion. The dispersion is a mixture of heavy- and light-

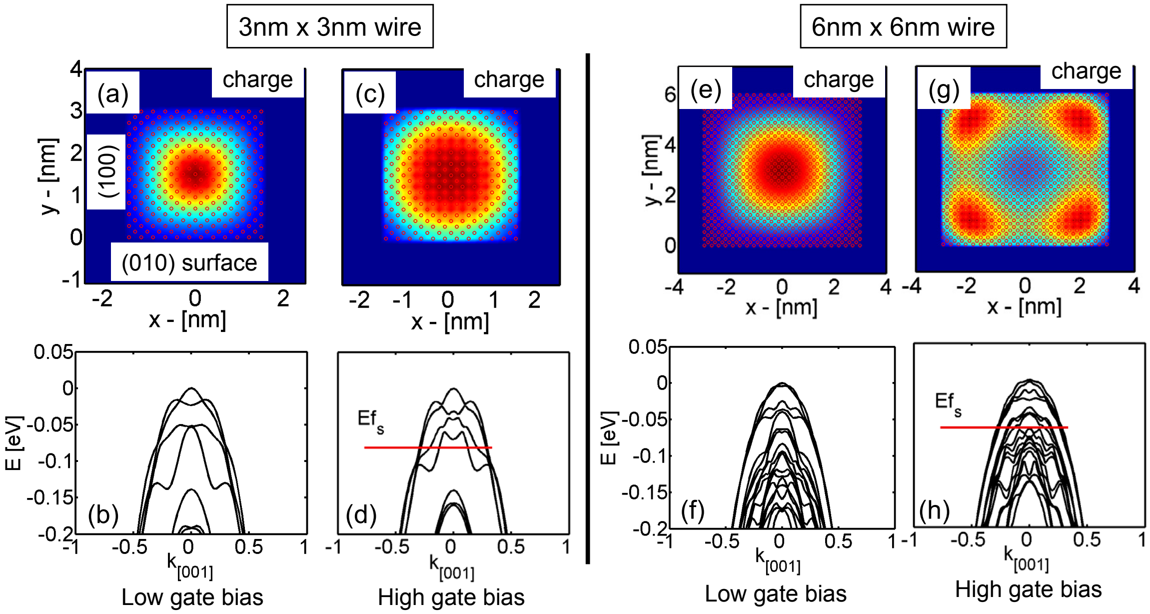


Fig. 5. 1. Device features for a [100] rectangular wire. (a-d) The 3nm square wire. (a-b) The 2D cross section showing the charge distribution and dispersion *E(k)* under V_G=0V gate bias conditions. (c-d) The 2D cross section showing the charge distribution and dispersion *E(k)* under V_G=1.2V, high gate bias conditions. The dots indicate the underlying atomic positions. E_{fs} is the source Fermi level. (Zero energy indicates the conduction subband edge). (e-h) Same features as in (a-d) for the 6nm square wire.

hole bands, indicating strong band coupling and warping. Under high biases, there is significant charge filling of the lattice as shown in Fig. 5.1 (c). Figure 5.1 (d) shows the dispersion of the nanowire under high gate bias. Similar to the case of electron conduction band transport [49], charge filling of the lattice causes changes in the dispersion of the nanowire even at the 3nm wire length scale (Fig. 5.1 (d)). The nature of the changes in the dispersions is not a simple shift in the bands' position, but rather strong warping, splittings and change of shape/curvature. Qualitative explanations on the shape of the dispersion will be given in section 3 of this chapter.

The 6nm [100] case: Figure 5.1 (e-h) shows for a 6nm square wire the same characteristics as shown for the 3nm wire of Fig. 5.1 (a-d). Under low bias conditions, the charge shown in Fig. 5.1 (e) is also placed in the middle of the channel. As the bias increases, the charge shifts towards the corners of the device as shown in Fig. 5.1 (g).

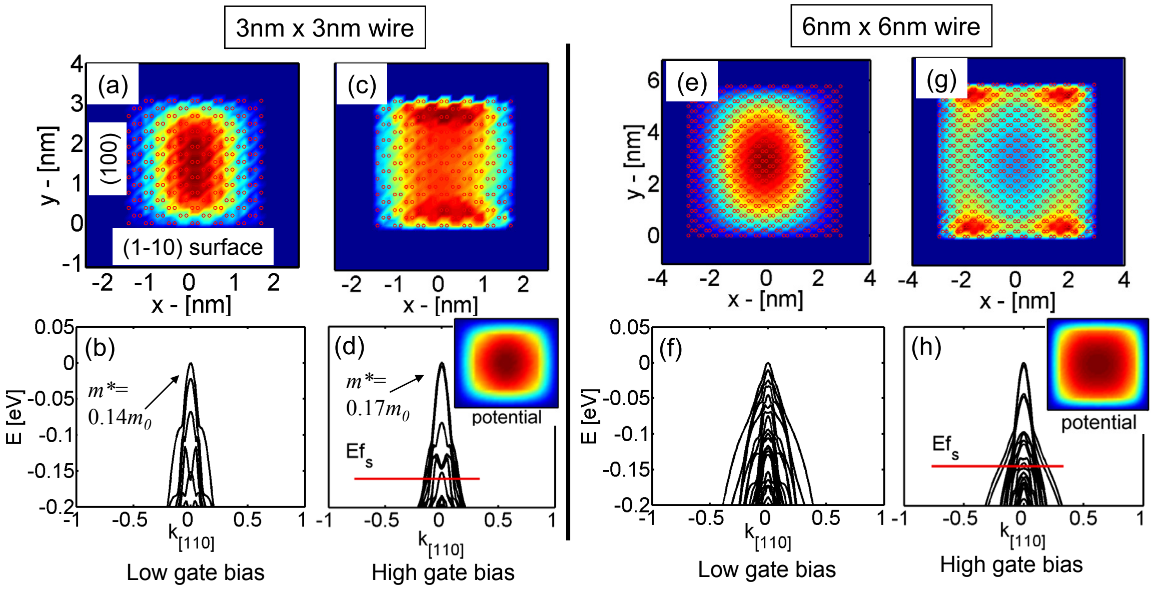
Similar corner effects in 6nm wires have also been calculated for the case of NMOS nanowire devices [48]. The corner effect is an electrostatic one, since the corners can reach inversion faster than the rest of the surfaces. Significant changes are also observed in the dispersion of the charge filled lattice (Fig. 5.1 (h)) compared to that of the empty one (Fig. 5.1 (f)).

5.1.2 Hole transport in [110] oriented nanowires

The 3nm [110] case: Fig. 5.2 (a-d) shows the same quantities for the 3nm [110] oriented wire, as shown in Fig. 5.1 (a-d) for the 3nm [100] oriented wire. The [110] wire's valence band dispersion shown in Fig. 5.2 (b) is significantly different than that of the [100] wire. Its' highest energy valley can be roughly described by effective mass, estimated to be ~0.14$m_{0.}$ (This is done using a parabolic dispersion constructed to fit the dispersion up to 0.1eV). The change in the dispersion under potential variations is also evident in this wire as shown in Fig. 5.2 (d). An estimate of the effective mass of the highest energy valley of this dispersion is calculated to be ~0.17m_0.

An obvious difference between the [110] and the [100] wire is observed in the charge distribution in the cross section of the wire. As clearly observed in Fig. 5.2 (a), under low bias the charge distribution is centered in the middle of the channel, preferably along its vertical *y*-axis along the [1-10] direction rather than the *x*-axis along [001]. This is an effect resulting from the anisotropy of the quantization mass in the [1-10] and [001] directions. Quantization in the $\langle 100 \rangle$ equivalent direction is subject to a lighter mass, and the wavefunction is shifted farther away from the Si/SiO_2 interface. The quantization mass in $\langle 110 \rangle$ is heavier, forcing the wavefunction more towards the (1-10) surface (see Fig. 5.2 (a)). This is even more evident under high biases as shown in Fig. 5.2 (c). Although the potential distribution in the cross section of the wire (inset of Fig. 5.2 (d)) is symmetric, the charge distribution in the cross section is preferably accumulated on the (1-10) surface (top/bottom) rather than the (001) surface (left/right).

This is an effect that should be taken into serious consideration when evaluating multi-surface type of devices. Kobayashi et al. in [15], showed experimentally that



Low gate bias High gate bias Low gate bias High gate bias

Fig. 5. 2. Device features for a [110] rectangular wire. (a-d) The 3nm square wire. (a-b) The 2D cross section showing the charge distribution and dispersion *E(k)* under V_G=0V gate bias conditions. (c-d) The 2D cross section showing the charge distribution and dispersion *E(k)* under V_G=1.2V, high gate bias conditions. Inset of (d): The electrostatic potential contour in the cross section of the nanowire. The dots indicate the underlying atomic positions. E_{fs} is the source Fermi level. (Zero energy indicates the conduction subband edge). The masses indicated are approximate effective masses for the highest energy level subband. (e-h) Same features as in (a-d) for the 6nm square wire.

different quantizations impact the performance of nanowire devices through V_T fluctuations and ON-current variations. In that work, PMOS nanowire devices of height 30nm were built on (100) surfaces. Both [110] and [100] transport orientated wires were investigated as a function of width fluctuations ([1-10] width direction for the [110] wire, and [010] width direction for the [100] wire. Fluctuations in the [1-10] width of the [110] wire had much smaller effect on the wire characteristics than fluctuations of the [010] width of the [100] oriented wires. This is evidence of the heavier mass quantization that does not allow large subband variations with size fluctuations.

The 6nm [110] case: Figure 5.2 (e-h) shows the same features for the 6nm x 6nm [110] wire as Fig. 5.2 (a-d) for a 3nm x 3nm device. Under low bias conditions (Fig. 5.2 (e)), the charge distribution still shows the preferential distribution along the *y*-[1-10] axis as in Fig. 5.2 (a). The low bias dispersion at energies close to the valence band edge indicates light masses that, however, get heavier for lower energies (Fig. 5.2(f)). At high biases, as shown in Fig. 5.2 (g), the charge distribution shifts towards the corners of the device. Unlike the 6nm [100] wire case in Fig. 5.1 (g), however, the charge is closer to the interface, and preferably accumulates on the top/bottom (1-10) surfaces rather than the left/right (001) surfaces. Comparing the charge distribution in Fig. 5.2 (g) with that of Fig. 5.2 (c), it appears that the inversion lobes that form along the (1-10) surface of Fig. 5.2 (c) are now on the four corners of Fig. 5.2 (g). It is reasonable to assume that the 6nm wire behaves as two 3nm wires, with two inversion lobes in the left/right each. (Looking at the dispersion of Fig. 5.2 (h), a second pair of light mass subbands above the Fermi level now appears as compared to the dispersion of the 3nm wire in Fig. 5.2 (d)).

5.1.3 Hole transport in [111] oriented nanowires

The 3nm [111] case: In Fig 5.3, the same analysis is performed for the [111] oriented wire. The [111] wire's valence band of Fig. 5.3 (b) is again different than that of the previous two described wires. Its highest valley can be roughly described by an approximate effective mass, estimated to be ~0.13m0. Under high bias, the dispersion changes (Fig. 5.3 (d)), and now the effective mass of the highest energy valley becomes ~0.11m0. Under low bias, the charge distribution in Fig. 5.3 (a) is centered in the middle of the channel. Under high bias (Fig. 5.3 (c)), the charge is still mostly symmetric in the cross section. This originates from the similarity in the dispersions along the quantization surfaces of this wire which are the (1-10) and the (11-2) surfaces, although a small degree of asymmetry can be observed. More details on this will be provided in section 3.

The 6nm [111] case: Fig 5.3 (e-h) again shows the same analysis for the 6nm square [111] wire. Under low bias conditions, the charge distribution is almost symmetric in the center of the wire as shown in Fig. 5.3 (e). The dispersions' curvature at energies close to the valence band edge is large, indicating light masses. As the energy increases, however, the curvature reduces, and the bands become heavier (Fig. 5.3 (f)). At high biases, the charge accumulates close and along the surfaces, slightly preferable along the

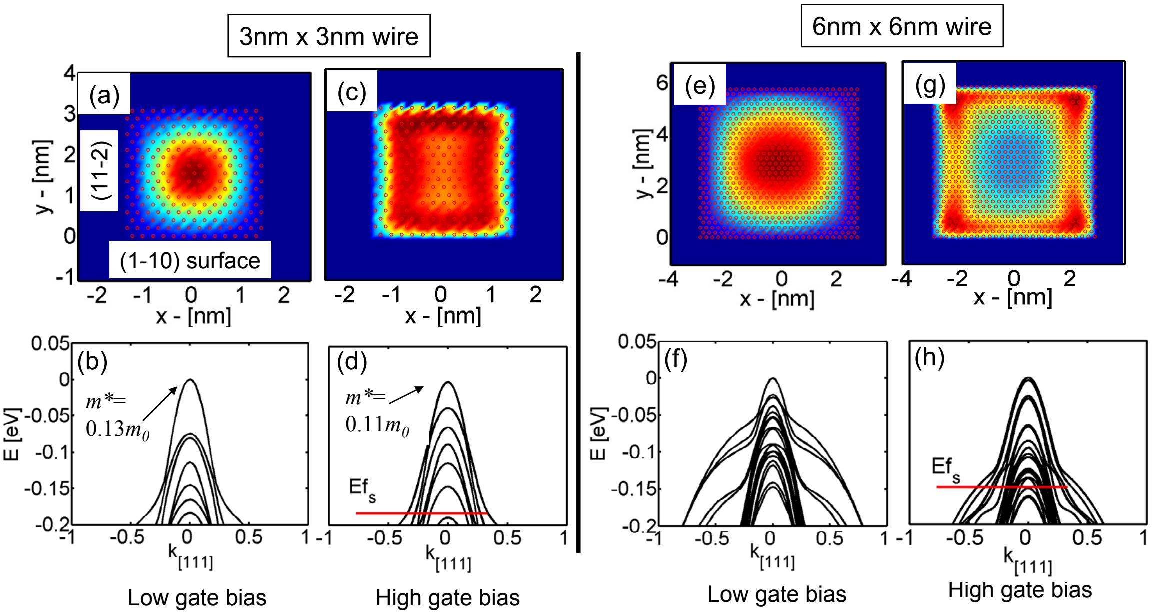


Fig. 5. 3. Device features for a [111] rectangular wire. (a-d) The 3nm square wire. (a-b) The 2D cross section showing the charge distribution and dispersion *E(k)* under V_G=0V gate bias conditions. (c-d) The 2D cross section showing the charge distribution and dispersion *E(k)* under V_G=1.2V, high gate bias conditions. The dots indicate the underlying atomic positions. E_{fs} is the source Fermi level. (Zero energy indicates the conduction subband edge). The masses indicated are approximate effective masses for the highest energy level subband. (e-h) Same features as in (a-d) for the 6nm square wire.

(1-10) top/bottom surfaces rather than the (11-2) left/right surfaces. The quantization masses of (1-10) and (11-2) surfaces are similar, which makes the charge distribution more symmetric. They are heavier than the (100) quantization masses, which allows the charge placement closer to the interfaces (comparing at the same gate overdrive). The high bias dispersion of this wire in Fig. 5.3 (h), as also observed in the case of the [110] wire, has light dispersion features for energies above the Fermi level, similar to the light bands in the low bias case for the 3nm wire of Fig. 5.3 (d). Just by looking at the dispersions, the 6nm wire under inversion behaves similarly to the 3nm wire, with a larger number of subbands.

5.2 Device performance comparison of NWs in different orientations

Figure 5.4 shows a performance comparison between the wires in the [100], [110] and [111] orientations for the 3nm and 6nm square nanowires. The various performance quantities shown further on, are all compared at the same OFF current (I_{OFF}) for all devices.

Gating induces same capacitance / charge in all wire directions: Figure 5.4(a) shows the total gate capacitance (C_G) vs. gate bias (V_G) of the three wires at the same I_{OFF}. The total capacitance in the three wires is very similar for all gate biases, as also observed in the case of the n-type transport in nanowires [49]. This is an indication that the same amount of inversion charge is accumulated in all wires irrespective of their orientations.

Low semiconductor capacitance (C_S) degrades the gate capacitance (C_G) from the oxide capacitance (C_{OX}) by ~30%: This amount corresponds to an effective increase in the oxide thickness of 0.35nm, (~30% of the physical gate oxide thickness of t_{ox} = 1.1nm). From $C_G = C_S C_{OX} /(C_S + C_{OX})$, with C_G= 0.35nF/m (maximum value of Fig. 5.4 (a)), and C_{OX}= 0.483 nF/m [86], C_S can is calculated to be $C_S = 1.27 nF / m$, which is only ~3 times larger than the value of the oxide capacitance. Comparing the C_S value to the quantum capacitance of the nanowires shown in Fig. 5.4 (b), (C_Q<3nF/nm for all wire orientations), the C_S is reduced from C_Q by a factor of ~2. This situation is the same in the case of the conduction band transport properties, and has also been observed in thin body devices and even bulk MOSFETs [87]. As discussed by Pal, this happens whenever the potential well formed in the channel is bias dependent. Equivalently, the charge centroid is placed further away from the interface, and C_S is reduced. C_Q being very similar for all nanowires, on the other hand, results in very similar final gate capacitances for all wires as well.

Velocity controls the transport differences in different orientated wires: Figure 5.4 (c) shows the injection velocities of the wires vs. gate bias (V_G). The [111] wire has the largest velocities, followed by the [110] wire, whereas the [100] wire has the lowest velocities, as would be expected from the subband masses estimated earlier for the [110] and [111] wires. (In 1D, under the parabolic band approximation, the velocity is

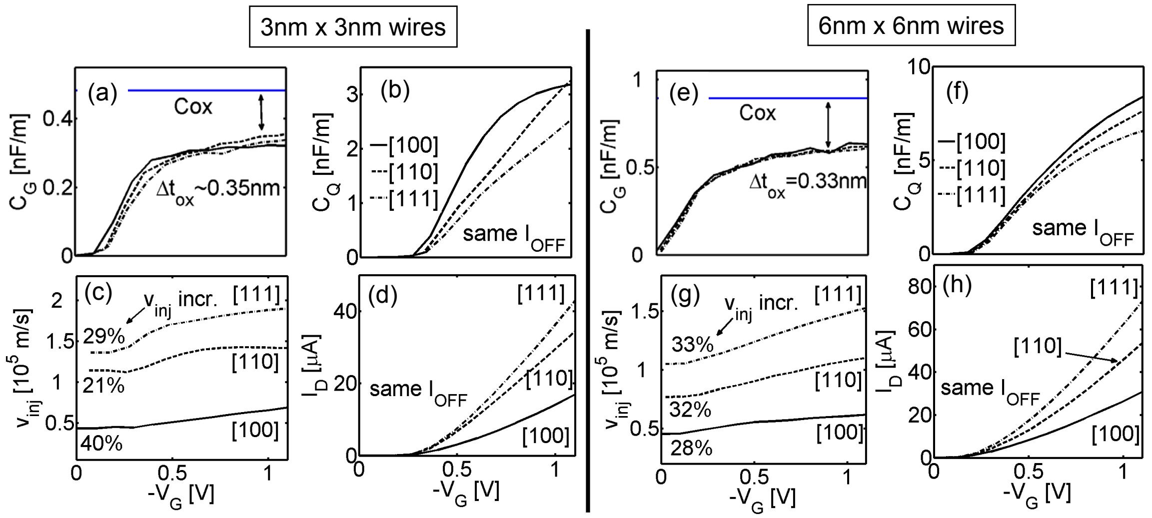


Fig. 5. 4. Performance comparison of the square wires in the [100], [110] and [111] directions at the same OFF-current (I_{OFF}). (a-d) The 3nm square wire: (a) The gate capacitance C_G vs. gate bias (V_G). The capacitance is similar for all wires, and degraded from the oxide capacitance by an amount that corresponds to an increase in the effective oxide thickness of 0.35nm. (b) The quantum capacitance C_Q vs. V_G of the three devices, which is a measure of the density of states at the Fermi level. (c) Comparison between the injection velocities of the nanowires vs. V_G. In all cases, the velocity is not constant, but increases as the gate bias increases. The increase is calculated by the difference between the value at high V_G and the value at low V_G. (d) The I_{DS} vs. V_G for the three wires at the same I_{OFF}. The velocity difference directly reflects on the current differences. (e-h) The 6nm square wire: Same features as in (a-d) respectively.

proportional to $v \sim 1/\sqrt{m^*}$). The lowest velocity in the [100] wire, is a result of the enhanced warping in its dispersion, which slows carriers down.

The injection velocities increase as high energy carrier states are being populated. This increase in velocity with gate bias can reach up to 40% in the [100] wires and even up to 29% in the [111] and 21% in the [110] wire orientation cases.

Velocity differences affect the I-V differences: Since the charge is the same in all devices, the velocity difference directly reflects on the ballistic I_{DS} as shown in Fig. 5.4 (d) in which the drive current capabilities of the wires are compared at the same I_{OFF}. The [111] and [110] wires perform better than the [100] wire in terms of ON-current capabilities. Comparing to the [100] wire, the [110] wire can transport almost 2 tims as much current, whereas the [111] wire almost 2.5 times higher current.

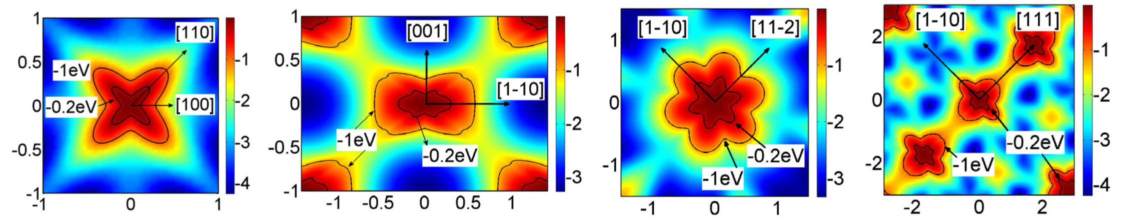
Transport characteristics for the 6nm wires: The same analysis has been performed in Fig. 5.4 (e-h) for the 6nm x 6nm wires. Similar observations as in the 3nm wire case are concluded for the total capacitance (Fig. 5.4 (e)), the quantum capacitance (Fig. 5.4 (f)), the velocities (Fig. 5.4 (g)), and the relative I_{DS} performance between the different orientation wires. The relative differences in the performance are slightly different than in the 3nm case, however, qualitatively the same conclusions can be drawn. Comparing to the 3nm wire case, the gate capacitance and the quantum capacitance doubles for the 6nm wire. The velocity is lower than that in the 3nm wire cases because slower carrier velocity subbands are now been occupied. Finally, the current is increased by ~60% (comparing at the same gate overdrive) when going from the 3nm x 3nm wire to the 6nm x 6nm wire (while the area is increased by 4 times).

5.3 Understanding the nanowire valence band dispersions and quantization through the 3D bulk *E(k)*

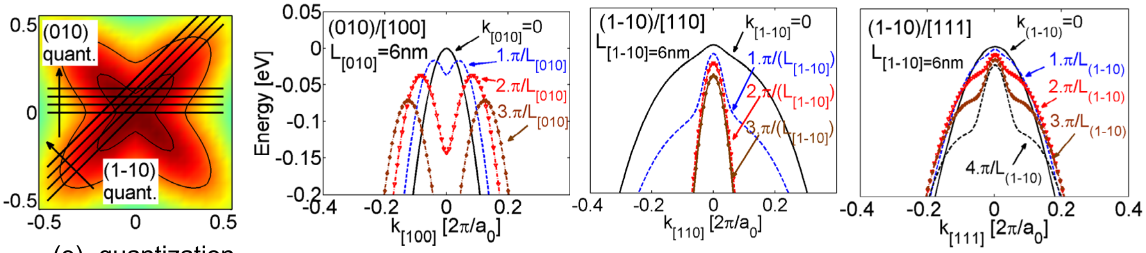
5.3.1 The *k*-space energy surfaces under different orientations-quantization behavior

In this section, intuitive explanations to the quantization behavior and the shape of the transport dispersions in different transport orientations are provided. The study of the curvatures and masses of the surface slices through the 3D bulk $E(k)$ in different directions can guide some understanding of the nanowire quantization behavior. The main reasoning comes from the anisotropy of the heavy-hole bandstructure, which defines the transport and quantization masses in different orientations. The heavier the quantization mass of a surface, the closer the charge will accumulate near the surface under inversion conditions. The heavy-hole (100), (110), (112) and (111) surfaces are investigated in Fig. 5.5. (The light-hole band is almost isotropic and does not force preferable charge distribution on any surface).

The (100) E(k) surface: Figure 5.5(a) shows the 2D $E(k_{[100]},k_{[010]})$ of the (100) surface. The contours at $E=0.2eV$ and $E=1eV$ are plotted. The bandstructure is



(a)-(100) surface (b)-(110) surface (c)-(111) surface (d)-(112) surface



(e)- quantization (f)-[100] subbands (g)-[110] subbands (h)-[111] subbands

Fig. 5. 5. Semi-analytical understanding of surface quantization: (a-d) Energy surface contours of the heavy hole calculated using the full 3D *k*-space information of the Si Brillouin zone. The energy contours for *E*=-0.2eV and *E*=-1eV are plotted. (a) The (100) surface. The anisotropy is evident in the [100] and [110] directions. (b) The (110) surface. (Or equivalently, 45° "cut" through the center of (a) into the surface. (c) The (111) surface. (d) The (112) surface. The [1-10] and [111] in plane directions are indicated. (e-h) Extraction of relevant quantization subbands. (e) The (100) energy surface with a few relevant quantization lines under quantization in the (010) (vertical $k_{[100]}$=constant lines) and (1-10) (45° lines) surfaces. Under quantization in (010), the shift in the *k*-value is given by $\Delta k=n.\pi/L$, where *n* is the subband index, and *L* is the quantization length (6nm in this case) (f) Relevant subbands in (010) surface quantization and [100] transport. (g) Relevant subbands in (1-10) surface quantization and [110] transport. (f) Relevant subbands in (1-10) surface quantization and [111] transport.

anisotropic, with heavier mass dispersion along the [110] direction ($m^*_{[110]}$ = -0.581m_0) and lighter along the [100] direction ($m^*_{[100]}$ = -0.276m_0). Quantization in [110] will utilize heavier quantization masses, than quantization in [100] directions. At the same inversion conditions in a nanowire, the charge distribution will preferably reside closer to the (110) surface than the (100) surface, which is exactly what observed in Fig. 5.2 (d) and 5.2 (g). On the other hand, the symmetry of the heavy hole in the [001] and [010] directions, results in the symmetric charge distribution in the [100] oriented wires of Fig. 5.1. Of course, the [100] wire can be quantized in the [110] and [1-10] directions. The charge distribution in this case will reside closer to the interface than what is shown in Fig. 5.1 (c,g).

The (110) E(k) surface: The differences in the quantization behavior in the [110] directed wire (with (100) and (110) equivalent quantization surfaces) is more evident in the $E(k_{[1\text{-}10]},k_{[001]})$ shown in Fig. 5.5 (b). This energy contour is a cross section along the 45° diagonal line of Fig. 5.5 (a). The [010] and [10-1] orientations indicated, are the quantization directions of the [110] oriented wire.

The (111) E(k) surface: The quantization of the [111] channels described earlier is determined from the (111) *E(k)* surface, with [1-10] and [11-2] quantized sides, shown in Fig. 5.5 (c). The energy surface does not look very different in the two directions, although minor details can be found. The charge distribution of Fig. 5.3 is due to this reason almost symmetric in the two quantization directions (with some minor differences).

5.3.2 Understanding the transport dispersion features

Relevant band extraction method: The bulk energy contour surfaces can provide indications for the shape of the dispersions in different orientations (either quantum wells or nanowires). Under any physical quantization, the relative energy/subband levels will follow the "particle in a box" quantization, and move away from the center of the band minima at a rate of $k_n=n.\pi/L$, where L is the quantization size. The smaller the physical domain, the larger the corresponding quantized k. For example, the relevant energy dispersion for a *quantum well* quantized in the [010]-direction, will be the horizontal energy surfaces in the [100]-[001] plane, passing through the lines drawn in Fig. 5.5 (e). Similarly, the relevant energy dispersions in [110] transport direction with quantization of the (1-10) surface are given by surfaces drawn through the 45° lines shown in Fig. 5.5 (e). (The relevant dispersions in the [111] transport direction with quantization of the (1-10) surface, will be surfaces perpendicular to the 45° lines of the (112) surface of Fig. 5.5(d) along [111]). Figures 5.5 (f-h) show the transport direction subbands for the [100], [110] and [111] orientations as the (010), (1-10) and (1-10) respectively are quantized, and for the *k*-vector in the remaining quantization direction set to $k=0$ (i.e. the quantization lines in Fig. 5.5 (e)). The quantization assumed is *L*=6nm (equivalently 12, 17 and 7 unit cells in the [110], [110] and [111] wires respectively).

Band shape in [100]: Figure 5.5 (f) shows energy subbands of the (010)/[100] structure. These will be relavant subbands in [100] orientation and partially explain the oscillating behavior observed in the subbands of the [100] wire dispersions of Fig. 5.1. (This is just an indication on where these subbands originate from. The method for an exact reconstruction of the nanowire subbands from the bulk bandstructure will be presented in the next section). Strong band coupling however, allows only qualitative understanding for the subband form, in contrast to the case of n-type wells, for which a good agreement can be achieved between this semi-analytical method and the full TB dispersions.

Band shape in [110]: The (110)/[110] direction subbands are drawn in Fig. 5.5 (g). (45° lines in Fig. 5.5 (e)). The subband passing through the center of the energy contour is of heavy-mass, however, as the structure is quantized in the [110] direction, the subbands shift away from the center and become lighter. This explains partially the shape of the subbands in the [110] wires of Fig. 5.2 (c), where the 3nm quantized wire has light mass subbands, whereas the 6nm quantized wire in Fig. 5.2 (g) has heavier subbands. Of course under electrostatic quantization in Fig. 5.2 (h), the highest energy subbands become lighter again (similar to the case of structural quantization).

Band shape in [111]: The [111] case, is similar to the [110] case as shown in the subbands of Fig. 5.5(h). Here, the subbands are drawn by taking lines at 45° in Fig. 5.5 (d) along [111] perpendicular to the [1-10] direction. Lines passing from the center (weak quantization), are heavy subbands, but as the structure is more quantized the subbands become lighter. The subband features in this case shine light into the [111] nanowire dispersion features of Fig. 5.3 (b) under strong quantization, and Fig. 5.3 (f) under weak quantization. Stongly quantized subbands will look more like the large n-value subbands (with lighter masses), while weakly quantizated subbands will look like the smaller n-value subbands (with heavier masses). Electrostatic quantization will have a similar effect as physical quantization in the curvature of the subbands, transforming the equilibrium dispersion of the 6nm wire (Fig. 5.3 (f)) into a dispersion with subbands similar to the 3nm wire subbands.

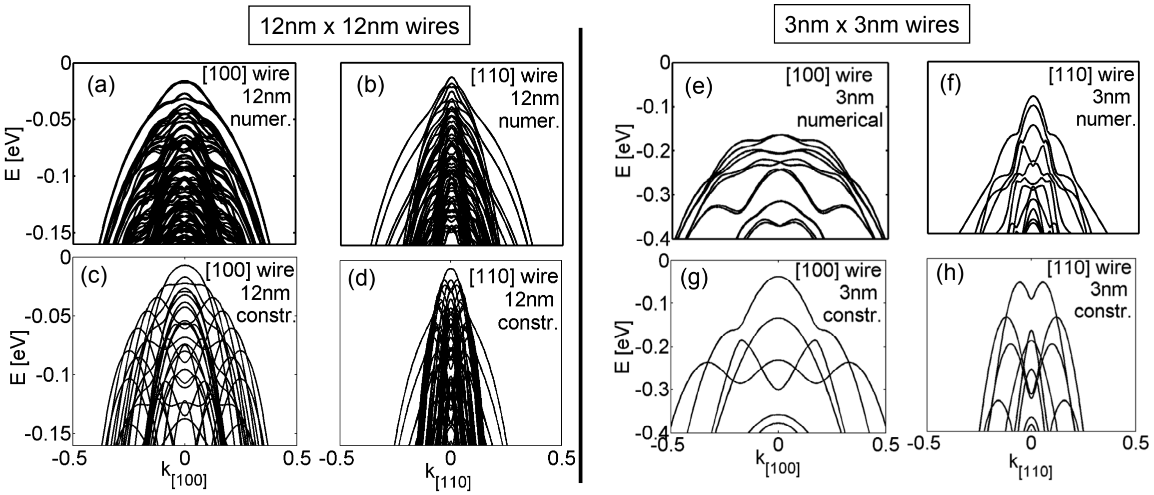


Fig. 5. 6. Semi-analytical understanding of nanowire dispersion: (a-d) The *E(k)* dispersions for square nanowires calculated using the full numerical TB model. (a) 12nm [100] wire. (b) 12nm [110] wire. (c) 3nm [100] wire. (d) 3nm [110] wire. (e-h) The *E(k)* dispersions for square nanowires calculated using the semi-analytical construction from the bulk bandstructure. (a) 12nm [100] wire. (b) 12nm [110] wire. (c) 3nm [100] wire. (d) 3nm [110] wire.

5.3.3 Semi-analytical construction of the NW dispersion from the bulk *E(k)*

The study of the surface slices through the bulk *E(k)* in different directions can guide some understanding of the nanowire dispersions. After picking a surface in a particular direction, the two remaining *k*-directions are quantized due to the lateral nanowire confinement. The quantized point on the plane corresponds to a single *k*-point in the transport direction dispersion. To obtain all *k*-points of the 1D subband in the transport direction, the surface slice needs to be translated further into the bulk dispersion and the lateral quantization must be redone. All relevant 1D subbands from the bulk 3D *E(k)* along the transport direction are obtained by shifting in *k*-space in both two relevant quantization directions $k_A=n.\pi/L_A$ and $k_B=m.\pi/L_B$, where k_A , k_B and L_A, L_B are the two quantization *k*-space directions and quantization lengths, for all *n, m*.

Figure 5.6 compares the full numerical dispersions to the semi-analytical dispersions constructed from the *bulk E(k)* for different wire cross sections (3nm, 12nm) and wire directions [100] and [110]. All heavy-hole, light-hole and split-off bands are included in this construction. In the cases of the 12nm wires, the envelope of the bands in the semi-analytical dispersion in both the [100] and [110] cases approximates the actual nanowire dispersion very closely. Enhanced band interactions in the case of the dispersions of the actual wires create qualitative differences in these multi-band dispersion figures. Potential variations are however not part of this analysis, which can have large impact on the dispersions in the actual wire. In the case of the smaller (3nm) wires, both in [100] and [110] orientations (Fig. 5.6 (c,d,g,h)), some of the trends and band shapes are captured, however due to the enhanced coupling, only small qualitative similarities can be observed between the actual and the semi-analytical constructed dispersions.

The semi-analytical construction method can only provide qualitative indications as of the subband form of the different oriented nanowires. Band interactions and the effect of potential variations is not included in this approach, however, reasonable understanding about the subbands of nanowires can be extracted from the simple bulk bandstructure. It is noted here, that in the case of the conduction band dispersion, or valence band of 2D quantum wells, where the band coupling is reduced, the semi-analytical construction method gives a much closer result to the actual dispersion of the structure [49, 95].

5.4 Conclusions

Transport properties of nanowires in different transport orientations ([100], [110] and [111]) are examined using a 20 orbital $sp^3d^5s^*$-SO atomistic TB model self-consistently coupled to a 2D Poisson solver. A semiclassical ballistic model is used to calculate the current-voltage characteristics of the nanowires. The dispersions of the nanowires cannot be described within the EMA because of the enhanced band coupling that induces large warping in the dispersions, especially for the [100] oriented wires. In addition, the dispersion shapes are strongly bias dependent.

The [111] wire has the largest carrier velocities and ON-current capabilities, followed by the [110] wire. The [100] wire is the worst in terms of both carrier velocities and drive current capabilities. The semiconductor capacitance (C_S) is important for nanowire devices and degrades the gate capacitance by ~30%. This effect is very similar for all wire orientations, for both 3nm and 6nm square wires.

The shape of the charge distribution in the cross section of the different oriented wires differs for each wire according to the quantization mass that each surface feels. The [100] wire examined with (010)/(001) quantization surfaces has a symmetric charge distribution in its cross section. In the case of [110] wires, with (1-10)/(100) quantization surfaces, the charge preferably accumulates on the (1-10) surface due to its largest quantization mass. The [111] wire, with similar quantization mass in the (1-10)/(11-2) surfaces has only slightly non-symmetric charge distribution profile on the two surfaces. These observations can give guidance towards the design of multi surface devices such as nanowires and FinFETs.

6. PERFORMANCE ANANLYSIS OF 60nm GATE LENGTH III-V InGaAs HEMTs: SIMULATIONS vs. EXPERIMENTS

6.1 Introduction

Field-effect transistors with III-V channel materials have recently received much attention because of their potential as switching devices for future digital technology nodes. Both heterostructure based high electron mobility transistors (HEMTs) [2-4] and MOSFETs [5, 6, 96] have been reported. Due to their higher mobility, the III-V channel materials should reach the ballistic limit at longer channel lengths than Si devices. The low effective mass of the III-Vs should also boost the ballistic carrier velocity and improve the I_D-V_D characteristics. Trade-offs are involved (e.g. the light effective mass leads to a density-of-states bottle neck [97, 98] and to source-drain tunneling [23]), but III-V FETs have the potential to outperform Si MOSFETs under low-voltage operation. In that regard, high-performance HEMTs based on III-V compounds with channel lengths below 90 nm have recently been demonstrated [3, 4, 99, 100]. Good control of the wide bandgap insulator thickness down to 3nm was achieved while still maintaining relatively low gate leakage currents – even under high biases. This paper is a simulation study of the results reported by Kim *et al.* [3]. Our objective is to examine the experimental data with a fully quantum mechanical, ballistic model in order to understand what can and cannot be explained.

In this chapter, a two-dimensional, ballistic quantum transport HEMT simulator based on the real space Non-Equilibrium Green's Function (NEGF) approach [24, 25] is employed. Simulation results show that for these $In_{0.7}Ga_{0.3}As$ HEMTs with a gate length of 60 nm and zero series resistance, a ballistic device of this kind would deliver about twice the on-current of the measured device. With external series resistors added, the simulated I-V characteristics are close to the measured results, except at the highest gate voltage. The simulations show good agreement with the subthreshold swing and drain-induced barrier lowering vs. channel length, but they do not show the drop in on-current with increasing gate length that is observed experimentally. The simulations shed light on the internal physics of these devices and identify issues for further study.

6.2 Approach

For simulation purposes, the device geometry is simplified as shown in Fig. 6.1. In the experimental device [3], the source and drain contacts are located on the top of the device, and the current flow is two-dimensional through a doped heterostructure stack. Rather than attempting to simulate the contacts (and the associated metal-semiconductor contact resistance), we placed ideal contacts at the two ends of the channel as shown in Fig. 6.1 and added extrinsic series resistors to the source and drain. The simulated HEMT consists of a 15nm $In_{0.7}Ga_{0.3}As$ layer between two $In_{0.52}Al_{0.48}As$ buffer layers. The gate electrode in the simulated device is placed on top of the $In_{0.52}Al_{0.48}As$ layer, which (in the simulated structure) has the same thickness throughout the entire length of the device. A silicon δ-doped layer [101] in the $In_{0.52}Al_{0.48}As$ buffer layer effectively dopes the source/drain regions of the device. The δ-doped layer is located 3nm away from the channel layer. Devices with insulator thickness of t_{ins} = 3nm, 7nm and 11nm were described in [3]. Later, better estimate of the 7nm and 11nm devices were given as 6.5nm and 10nm [101], and these were the values used in our simulations. In the simulating the t_{ins} = 3nm device, the δ-doped layer was placed on top of the insulator, and the gate electrode on top of a thin layer on top the δ-doped layer. In the simulation, the δ-doped layer is given a finite thickness of 0.40 nm and the thin layer on top of the d-doped layer was 0.40 nm. The result was an insulator thickness of t_{ins} = 3.8nm in the simulation. This is within the experimental uncertainty in the insulator thicknesses of +/- 1nm [101]. The uncertainty in insulator thickness is not substantial for the thinnest insulator device and this is discussed in Sec. 6.4.

The far left/right regions of the δ-doped layer are doped to $5x10^{13}/cm^2$ to mimic additional doping from the n+ cap layers ($In_{0.7}Ga_{0.3}As/In_{0.52}Al_{0.48}As$) used in the experimental device to facilitate ohmic contacts to the source and drain. There are, therefore, two different doping regions in the simulated device. The region that is directly adjacent to the channel to its left/right, has a carrier density of $2.1 \times 10^{12}/cm^2$, which is the value specified by the experimental group [101]. The far left/right region has a larger doping of $10^{13}/cm^2$. This level of doping is unrealistically high for this type of material, but it favors numerical stability of the simulation and does not affect the device. This is a

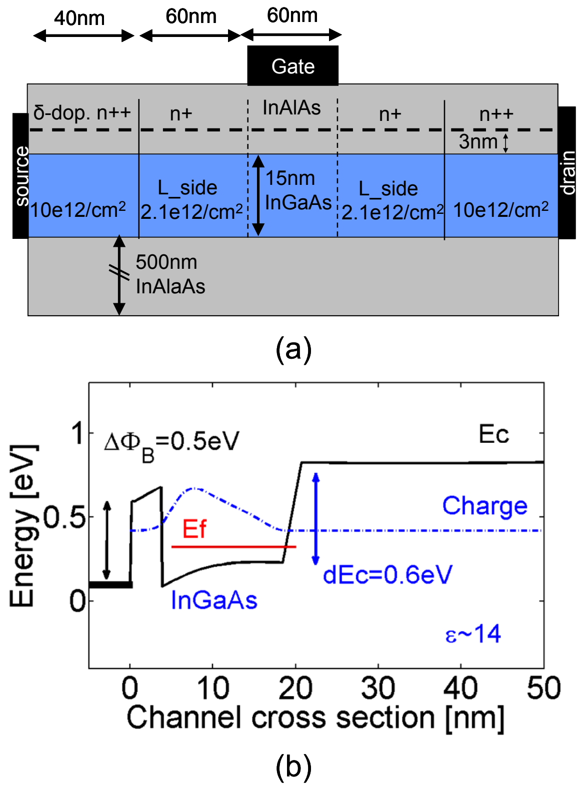


Fig. 6. 1. HEMT device description: (a) The simplified HEMT device structure. An $In_{0.7}Ga_{0.3}As$ between two $In_{0.52}Al_{0.48}As$ layers acts as the channel. A δ-doped layer 3nm away from the channel layer, effectively dopes the source/drain regions of the device to $2.2 \times 10^{12}/cm^2$. (b) The conduction band profile taken at a cross section of the HEMT device at the region of the source/channel boundary when the device is under large gate bias. The workfunction difference between the gate and the $In_{0.52}Al_{0.48}As$ buffer layer is adjusted to $\Delta\Phi_B = 0.5eV$. The conduction band discontinuity between the $In_{0.7}Ga_{0.3}As/In_{0.52}Al_{0.48}As$ layer is assumed to be $\Delta E_c = 0.6eV$. The dielectric constant of $In_{0.52}Al_{0.48}As$ is assumed to be $\varepsilon = 14$ and of the $In_{0.7}Ga_{0.3}As$ $\varepsilon = 14.5$.

way to mimic the extended source/drain regions of the actual device. In addition, although the lightly doped region in the experiment has L_{side} = 1μm [101], for computational efficiency L_{side} is set to L_{side} = 60nm in the simulations. As will be discussed later, under high gate bias, the source design becomes important. Because of the simplified source design used in the simulation, we will be able to draw only qualitative conclusions about the high gate bias performance. The channel region is the region directly under the gate electrode and has L_G = 60 nm as in the experimental device. Longer channel lengths were also examined experimentally, and these devices are briefly considered in Sec. 6.4.

It should be pointed out that the source and drain contacts in the simulated device should not be regarded as real contacts with an associated contact resistance. Rather, they are idealized contacts to the extended source/drain regions which are assumed to be maintained in thermodynamic equilibrium by strong scattering. Venugopal et al. examined this assumption for silicon transistors and found that scattering in typical contacts of heavily doped silicon is sufficient to maintain thermodynamic equilibrium [102]. Nevertheless, this assumption may need to be reconsidered as devices continue to shrink and for new channel materials such as the III-V's considered here. Indeed, Fischetti has discussed the phenomenon of "source starvation" which is a manifestation of non-equilibrium contacts in III-V FETs [103]. For this study, we assume extended contacts that are maintained in thermodynamic equilibrium.

Figure 6.1b shows the simulated conduction band profile normal to the channel taken at a location near the source end of the channel (about at the top of the potential energy barrier between the source and the channel) when the device is under large gate bias. The workfunction difference between the gate and the $In_{0.52}Al_{0.48}As$ buffer layer is adjusted to $\Delta\Phi_B = 0.5eV$ in order to match the threshold voltage of the simulated devices to the experimental measurements. The thickness of the $In_{0.52}Al_{0.48}As$ layer in this case is 3nm, and the δ-doped layer is adjacent to the gate/$In_{0.7}Ga_{0.3}As$ interface. The conduction band discontinuity between the $In_{0.7}Ga_{0.3}As/In_{0.52}Al_{0.48}As$ layer is assumed to be ΔE_c = 0.6eV [101, 104]. The dielectric constant of $In_{0.52}Al_{0.48}As$ is assumed to be ε = 14 and that of the $In_{0.7}Ga_{0.3}As$ is ε = 14.5 [105].

The effective mass of the $In_{0.7}Ga_{0.3}As$ channel is an input to the simulation. Because of conduction band non-parabolicity, quantum confinement will increase the effective mass as compared to its value in the bulk. In principle, the appropriate effective mass could be extracted by sophisticated atomistic calculations (e.g. tight-binding atomistic methods), however, this is a difficult task because the masses can be a function of the exact placement of the atoms in the structure and the distortions within the structure. In this work a simplified approach is followed; we extract the effective mass from atomistic, tight-binding [106] calculations for a 15 nm wide $In_{0.7}Ga_{0.3}As$ quantum well structure without assuming any lattice distortions. The dispersion of the quantum well is shown in Fig. 6.2. The wafer orientation is (100) and the transport orientation is [011]. The parabolic band drawn on top of the first valley is adjusted to match the density of states up to 0.2eV above the conduction band edge and results in an effective mass of $m^* = 0.048m_0$, which is the effective mass used in the simulations. (We chose to fit from the bottom of the conduction band to 0.2 eV above the bottom because the maximum position of the Fermi level above the conduction band edge under high gate bias is usually close to or below 0.2eV). Similar parabolic bands that match the *bulk* $E(k)$ bandstructures of the InAs and GaAs at Γ up to 0.2eV above the conduction band minima

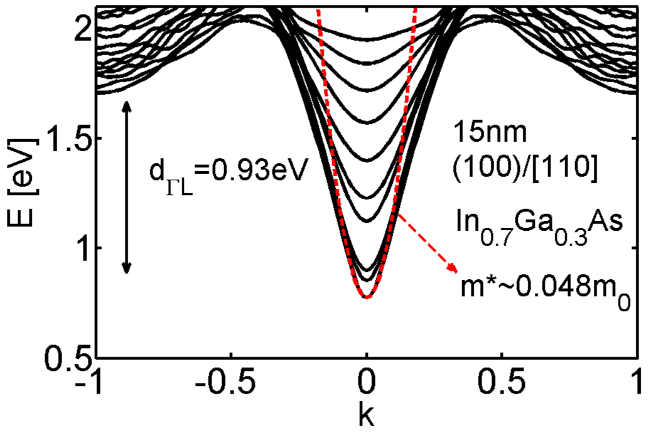


Fig. 6. 2. The dispersion of the composite 15nm thick $In_{0.7}Ga_{0.3}As$ structure calculated using atomistic tight-binding calculations with no distortions taken into account. The wafer orientation is (100) and the transport orientation is [011]. The parabolic band (red-dotted) of $m^* = 0.048\ m_0$ is adjusted to match the density of states up to 0.2eV above the conduction band edge.

were also extracted. A weighted average of these masses according to the 70% indium and 30% gallium composition results in a very similar value for the effective mass. The mass value is higher than the weighted average of the literature bulk masses, which is m^*=0.037m$_0$, (m^*_{InAs} = 0.027m$_0$, m^*_{GaAs} = 0.063m$_0$). Our use of a larger effective mass accounts for the effect of non-parabolicity in an approximate way. In addition, as shown in Fig. 6.2, the L valleys are very high in energy compared to the Γ valleys and are, therefore, ignored in our simulations. (This is expected since the composite channel in this case due to the 70% indium composition has stronger InAs properties rather than GaAs properties which will tend to place the Γ and L valleys closer in energy.)

The non-equilibrium Green's function approach [24, 25] for ballistic quantum transport is self consistently coupled to a 2D Poisson solver for treatment of the electrostatics. Since the channel is relatively thick (15nm), significant potential variations are expected in the cross section along the transport orientations. The NEGF Hamiltonian uses a real space technique in the parabolic, effective mass approximation (EMA) and accurately accounts for the mode coupling when large potential variations exist. The NEGF transport equation is solved in the channel area as well as in the upper $In_{0.52}Al_{0.48}As$ buffer layer in order to capture the wavefunction penetration in that layer. The 2D Poisson's equation is solved in the entire cross section of the device in order to accurately capture the 2D electrostatics of the device.

6.3 Results

In order to compare the measured to the simulated data, two fitting parameters were used, a value of external series resistance (R_{SD}) that is added to the device, and the workfunction difference ($\Delta\Phi_B$) between the gate and the $In_{0.52}Al_{0.48}As$ buffer layer. The effect of the series resistance will be explained in the following section. The workfunction difference, $\Delta\Phi_B$, is used to adjust the V_T of the simulated to that of the experimental data. The adjustment is done once for the L_G = 60 nm device with a 3nm thick insulator. The result, $\Delta\Phi_B = 0.5eV$, is a reasonable number for the workfunction

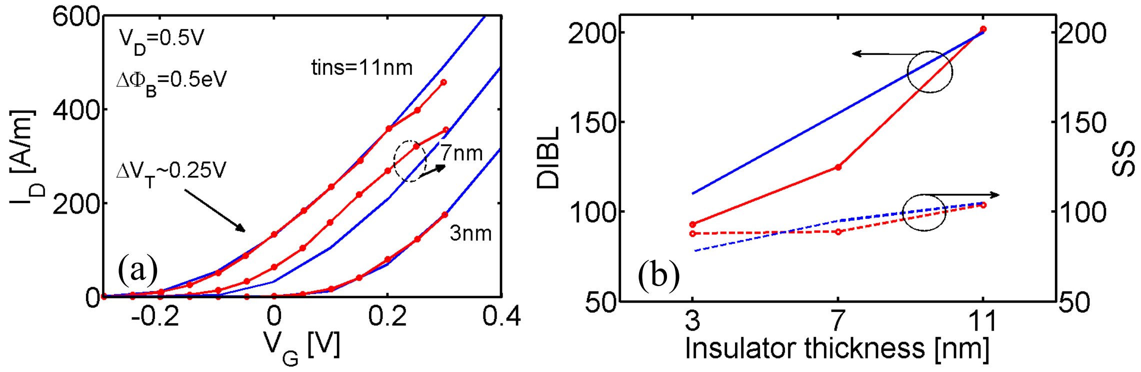


Fig. 6. 3. 2D electrostatics of the HEMT: (a) The experimental (red-circle) and simulated (blue-solid) I_D-V_G data for the L_G= 60nm, t_{ins}= 3nm, 7nm, 11nm devices. A workfunction difference between the gate and the $In_{0.52}Al_{0.48}As$ layer of $\Delta\Phi_B = 0.5eV$ is used in order to match the V_T for all devices. A negative shift in V_T by 0.25V is observed as the oxide thickness increases. (b) The *DIBL* and subthreshold swing *(SS)* of the experimental and simulated data.

difference between the two materials. Figure 6.3a shows the I_D-V_G characteristics for the three devices – each with a 60nm channel length but with three different insulator thicknesses. The measured and simulated curves agree fairly well. In the case of the 7nm insulator device, the simulated and measured V_T differs by ~0.04V. This small deviation might be due to various reasons such as interface traps, charged impurities, or uncertainties in the thickness of the layers in the experimental device. As the insulator thickness increases, there is a large negative shift in the V_T by almost 0.25V, which is attributed to the δ-doped layer and its increasing effect on the electrostatics of the channel as the gate electrode moves farther away. The threshold voltage shift is well-described by

$$\Delta V_T = \frac{qN_W}{C_{ins}} \frac{\hat{x}}{T_{ins}}, \tag{6.1}$$

where N_W is the δ-doping concentration per cm^2, and $\hat{x}$ is the centroid of the charge distribution in the insulator [107].

The Drain Induced Barrier Lowering (*DIBL)* and the subthreshold swing (SS) extracted from the simulated data are shown in Fig. 6.3b, and both are seen to increase as the insulator thickness increases, which is expected from 2D electrostatics. The simulated results agree with the experimental data both qualitatively and quantitatively.

The second adjustable parameter in the simulation is the series resistance. The series resistance originates from the complicated ohmic contact between the n^+ cap layer, the $In_{0.52}Al_{0.48}As$ layer and the barrier between the interface of the $In_{0.52}Al_{0.48}As$/ $In_{0.7}Ga_{0.3}As$ layers. Figure 6.4a shows the experimental I_D-V_D data and the simulated ballistic I_D-V_D characteristic at the same gate overdrive (V_G = 0.5V). The simulated, ballistic ON-current is almost double than the experimental value, and the channel resistance of the simulated ballistic device is R_B = 170 Ω-μm (inverse slope of the linear region). In order fit the simulated results to the experimental data, a series resistance (source plus drain) of R_{SD} = 400 Ω-μm was added to the ballistic data in order to match the total resistance measured in the experimental data (inverse slope of the high V_G experimental I_D-V_D).

Once the series resistance is fit to the linear region of the highest gate voltage data, the simulated data at low drain voltages shows very good agreement with the experimental observations for all three gate bias cases reported experimentally (V_G = 0.1V, 0.3V, 0.5V). The agreement at high drain voltage is also good, except for a ~15% discrepancy between ON-current of the measured and simulated data. For this L_G = 60 n nm device, the experimental results can, to a reasonable approximation, be explained by an intrinsic, ballistic FET with two series resistors attached to it, except for the overestimate of the on-current, which will be discussed in Sec. IV. Longer channel lengths appear to operate at a lower fraction of the ballistic limit, as will also be discussed in Sec. 6.4.

Similarly, the experimental I_D-V_D data for the t_{ins} = 7 nm and t_{ins} = 11nm devices can be explained by using slightly different values of R_{SD} (R_{SD} = 350 Ω-μm and R_{SD} = 310 Ω-μm, respectively). The value of the fitted series resistance increases as the insulator thickness decreases. This was also observed in the experiments and was attributed to the isotropic etching that was used before gate deposition to produce the three different insulator thicknesses [3]. As more of the insulator sidewall is etched, the series resistance tends to increase. Figures 6.4b,c show the experimental I_D-V_D for various V_G values compared to the simulated results after the series resistance has been fit.

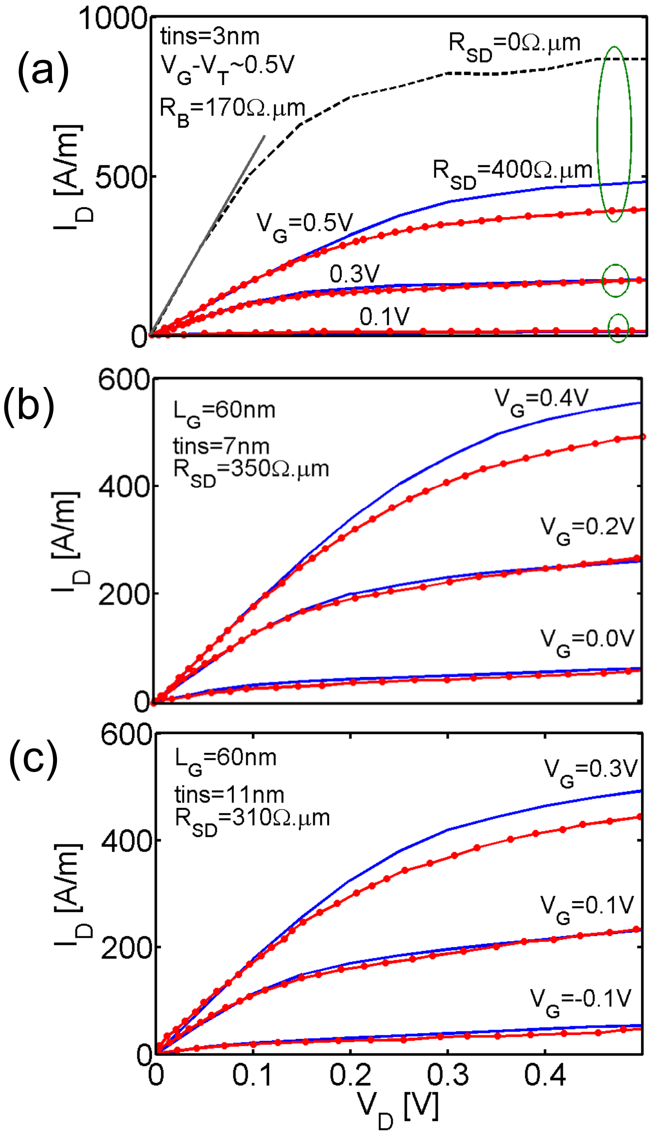


Fig. 6. 4. Comparison between the experimental (red-circle) and simulated I_D-V_D with series resistance added to them (blue-solid). (a) The t_{ins} = 3nm device. Data for V_G = 0.1V, 0.3V and 0.5V are shown. The black-dashed curve indicates the ballistic I_D-V_D at V_G = 0.5V with R_{SD} = 0 Ω-μm. A R_{SD} = 400 Ω- μm is added to the simulated data. (b) The t_{ins} = 7nm device. Data for V_G = 0V, 0.2V and 0.4V are shown. A R_{SD} = 350 Ω-μm is added to the simulated data. (c) The t_{ins} = 11nm device. Data for V_G = -0.1V, 0.1V and 0.3V are shown. A R_{SD} = 310 Ω-μm is added to the simulated data.

Good agreement between the experimental and simulated data is observed, but for each of the three cases, the ON-current of the simulated device is ~10-15% more than that of the measured device.

The mobility of a field-effect transistor is often extracted from the linear region current. Although mobility has no physical meaning in our ballistic simulations, the simulated ballistic drain current is linearly proportional to the drain voltage at low V_{DS}, so we can extract a "mobility" by equating the channel resistance to a conventional MOSFET expression,

$$R_{ch} = \frac{V_{DS}}{I_{DS}/W} \equiv \frac{L}{\mu_B C_{ins}(V_G - V_T)}, \tag{6.2}$$

where μ_B is the so-called ballistic mobility by [108-110]. From our simulations, R_{ch} at high gate bias (before adding the effect of R_{SD}), varies between R_{ch} = 170 Ω-μm – 240 Ω-μm as the insulator thickness varies from 3nm to 11nm. From these channel resistances, a value of the ballistic mobility is extracted to be μ_B~ 170-450 cm^2/V-s. Although the mobility of bulk $In_{0.7}Ga_{0.3}As$ is measured to be ~10,000 cm^2/V-s, the "apparent" mobility (in the sense of eqn. (2)) that a short channel HEMT can display is limited to a few hundred]. Alternatively, one could deduce a mobility for the device by plotting the total resistance between the source and drain as a function of channel length. The *y*-intercept of this curve would be the fixed, external series resistance and the inverse of the slope would be proportional to the channel mobility. In that case, a ballistic FET would show zero slope, corresponding to an infinite mobility.

6.4 Discussion

6.4.1 The effect of R_{SD}

Within the uncertainties of the simplified structure used in the simulations and in our knowledge of various device parameters, the results presented in the previous section show that the L_G = 60 nm HEMTs reported by Kim et al. [3] can be described as ballistic FETs with two external series resistors. The only significant discrepancy between the simulated and experimental results is the consistent 10-15% over-estimate of the

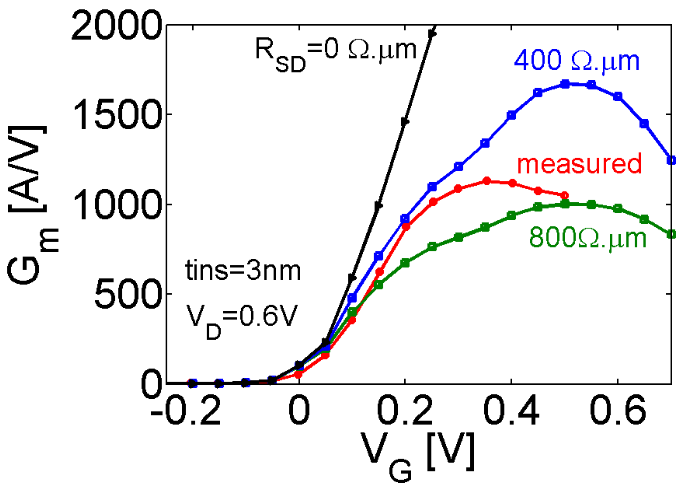


Fig. 6. 5. The g_m vs. V_G data for the t_{ins} = 3nm device. Measured data (red-circle), and simulated data with R_{SD} = 0 Ω-μm (black-solid), R_{SD} = 400 Ω-μm (blue-square), and R_{SD} = 800 Ω-μm (green-square) are shown.

ON-currents. The experimental transconductance, g_m, vs. gate voltage characteristic is shown in Fig. 6.5 for the t_{ins} = 3nm device. The observed degradation in g_m *at* high gate voltages might be attributed to various causes. Scattering at high gate biases could reduce mobility and degrade g_m. Another possibility is population of heavy effective mass upper valleys.

Figure 6.2 shows, however, that the L valleys are too high in energy to be populated. Parallel conduction in the upper layer, which has much heavier masses (~5 times heavier) than the channel layer, could also be a possibility. As shown in Fig. 6.1b, however, our simulations show no significant wavefunction penetration in the upper layer – even under high inversion conditions. Series resistance could be yet another possibility. Figure 6.5 shows the simulated g_m vs. V_G characteristics for three different values of series resistance (R_{SD} = 0, 400, 800 Ω-μm). For the R_{SD} = 0 and 400 Ω-μm cases, the g_m follows the experimental curve, but saturates at much higher V_G than the experimental curve. For the 800 Ω-μm characteristic, we obtain roughly the correct magnitude of g_m, but this value of R_{SD} is too large to be consistent with the experimental measurement. The fact that g_m degradation occurs even in the ballistic simulation tells us, however, that there might be other possibilities. Two other plausible causes, the design of the source,

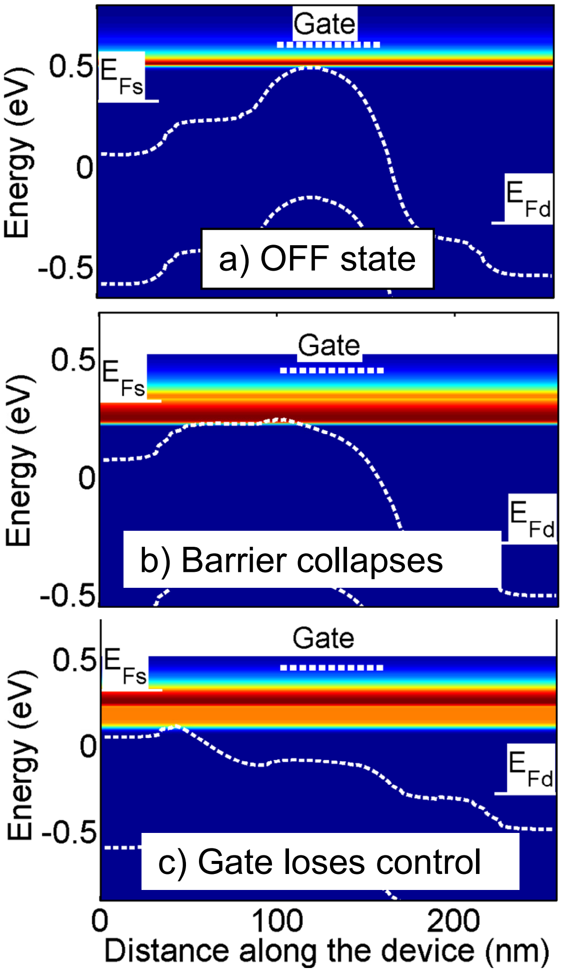


Fig. 6. 6. The source exhaustion mechanism. The energy resolved current spectrum is shown. (a) The device at OFF-state. (b) The barrier collapses as V_G is applied at ON-state. (c) Further increase in V_G causes the lightly doped region to collapse. The top-of-the-barrier that has now shifted to the highly doped region and the gate loses control over the device.

and the effects of non-parabolicity are discussed below.

6.4.2 Source design and source exhaustion

For III-V transistors, the design of the source can be an important factor [103, 111]. Transistors operate by modulating potential energy barriers [112, 113]. As the gate voltage increases, the potential energy barrier decreases, and the charge in the channel increases. When the gate voltage increases to the point where the barrier is removed and the channel charge is equal to the charge in the source, transistor action degrades significantly. Simply stated, there can't be more charge in the channel than in the source. For the HEMT under consideration here, the charge in the source ($2.1\text{x}10^{12}/\text{cm}^2$), is much lower than typical for Si MOSFETs, so these source exhaustion effects become apparent at relatively low gate voltages.

Source design limits are illustrated by the ballistic simulation shown in Fig. 6.6. Figures 6.6a, b, c show the energy-resolved current vs. position for the HEMT device under different gate voltages. The conduction and valence bands are indicated (white-dot lines), and the current flows above the top of the conduction band. The source/drain regions consist of two portions, an n^{++} region near the ideal contacts and an n^{+} region adjacent to the channel. Figure 6.6a shows the OFF-state of the device, where the source Fermi level (E_{fs}) is well below the top of the source to channel energy barrier. As V_G increases, the barrier in the channel decreases – eventually reaching the same level as the n+ source region (Fig. 6.6b). The top of the barrier has in this case shifted to the beginning of the n^{++} source region. When V_G increases even more (Fig. 6.6c), the gate can only modulate the energy barrier at the n^{++} to n^{+} junction through weak fringing fields. Transistor action is lost, and g_m drops as shown in Fig. 6.5 for both the simulated and measured characteristics. In our simulations, these effects are exaggerated by the assumption of ballistic transport in the n^{+} source, but the effect is primarily an electrostatic one and is also observed in drift-diffusion simulations [114].

The gate voltage at which the transconductance begins to degrade is strongly dependent on the barrier between the channel and the source, which depends on the doping of the source. Figure 6.7 shows the simulated g_m for structures with different δ-doping densities above the source/drain. As the doping in the source decreases, this effect shows up at smaller gate voltages. The low gate bias part of the g_m vs. V_G characteristic is not doping dependent because under low gate voltage, the source is able to supply the charge demanded by the gate voltage. In the experimental results, the n^{+} source region was L_{side} = 1μm in length, whereas in our simulation, L_{side} = 60nm was used. The differences in the source doping profiles may explain why the transconductance is experimentally observed to degrade ~0.2V before the simulated transconductance.

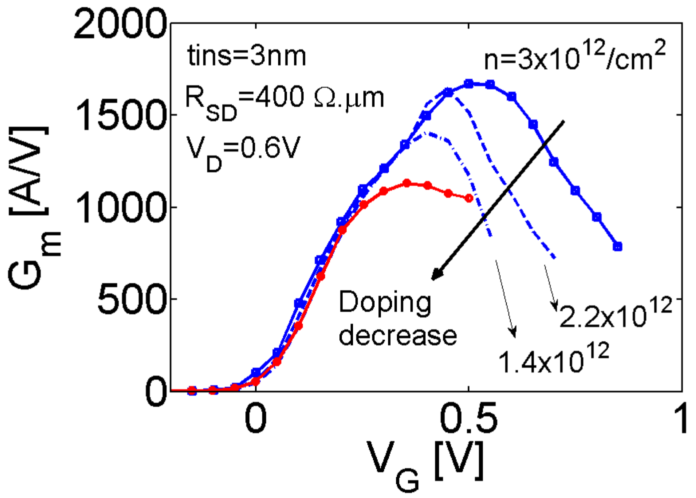


Fig. 6. 7. The effect of source/drain electron charge on g_m degradation. As the “doping” decreases the degradation starts in lower gate biases.

Although we cannot unambiguously conclude that the observed transconductance degradation is due to source exhaustion, our simulations do clearly demonstrate that source design is an important issue for III-V MOSFETs. Finally, note that the effects discussed here are purely electrostatic in nature and occur in both ballistic and drift-diffusion simulations. Fischetti has discussed “source starvation,” which results from a difficulty in injecting carriers into transverse momentum states in the channel [103]. Those effects were not included in our study and would only make source design an even more important issue.

6.4.3 Charge and velocity

Two important parameters for a FET are the charge and velocity at the beginning of the channel. Two questions arise. The first is: How close is the charge at the top of the potential barrier to the equilibrium MOS capacitor value of $Q = C_G(V_G - V_T)$? The second question is: How the velocity extracted from the numerical simulator compares to the ballistic injection velocity expected from the bandstructure of the channel. To answer both of these questions, the top of the potential barrier in the numerical results needs to be identified. Doing so is not as trivial, because of the large variation of the E_C across the depth of the 15nm channel width. We employ two different methods to locate the top of the barrier. The first is to take the weighted average of the charge distribution with the 2D $E_C(x,y)$ profile with the 2D charge density $n(x,y)$ according to

$$\langle E_C(x)\rangle = \frac{\int n(x,y)\,E_C(x,y)dy}{\int n(x,y)\,dy}. \tag{6.3}$$

Figure 6.8a shows the resulting $\langle E_C(x)\rangle$ (white-dotted line) superimposed on the electron density spectrum plot. Figure 6.8a is plotted at $V_G = 0.4$V, and $V_D = 0.35$V, which are the estimated intrinsic device voltages at the ON-state (after accounting for the effect of R_{SD}). From Fig. 6.8a, the top-of-the-barrier can be identified to reside at 105nm (5nm inside the channel from the point where the gate electrode begins).

A second way to identify the top-of-the-barrier is by identifying the point of maximum gate control by locating the position where $dN_S(x)/dV_G$ is maximized (where $N_S(x)$ is the charge in the channel per cm^2). This method places the top-of-the-barrier at 104 nm. Both approaches give very similar results, so we take the top-of-the-barrier to be at 104.5nm. The corresponding charge and the velocity (defined as $I_{ON}/N_S(x)$) at the-top-of-the-barrier are $N_S \approx 1.3\times10^{12}$ per cm^2 and $\upsilon_{ave} \approx 2.7\times10^{7}$ cm/s as shown in Fig. 6.8b,c respectively. The charge density and velocity are rather low due to the fact that the source Fermi level is less than 0.1eV above E_C under ON-state conditions. Figure 6.8 shows that these quantities are very sensitive to the precise location of the beginning of the channel. This information is available in our simulator, but it is not available when analyzing experimental data.

To answer the first question about how close the charge is to $Q = C_G(V_G - V_T)$, the simulated equilibrium carrier density vs. gate voltage is plotted in Fig. 6.8d (solid-blue). The quantity $Q = C_{ins}(V_G - V_T)$ with C_{ins} = 0.032 F/m^2 is shown as the solid-square-black line of Fig. 6.8d. Assuming that $C_G = C_{ins}$ clearly over-estimates the charge.

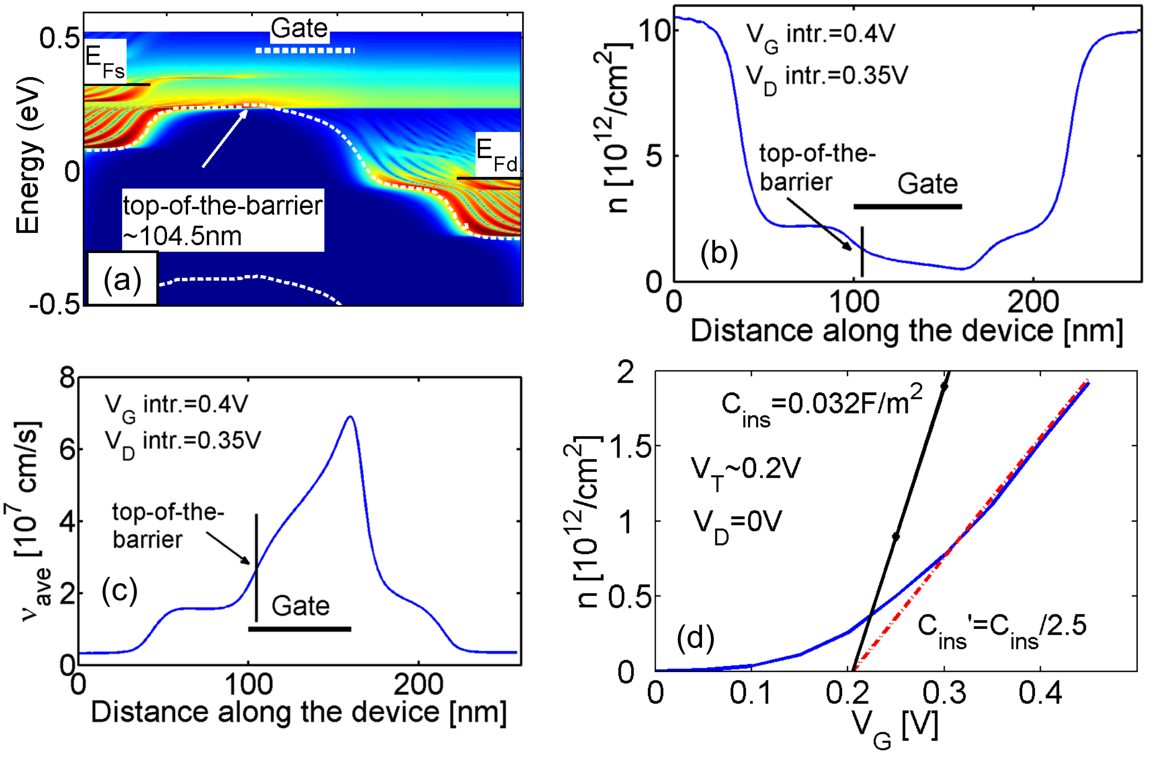


Fig. 6. 8. The intrinsic device parameters at ON-state. (a) The electron density spectrum. The density weighted E_C and E_V profiles are shown (dot-white lines). The top of the barrier is identified at 104.5nm. (b) The charge density along the length of the channel. (c) The average velocity along the length of the channel. (d) The equilibrium (V_D = 0V) carrier density vs. V_G (solid-blue). The charge as C_{ins}*(V_G-V_T) is shown in solid-square-black. The charge as (C_{ins}/2.5)*(V_G-V_T) is shown in dot-red.

From the slope of the C_G vs. V_G plot (dashed-red line), we observe that C_G is 2.5 times smaller than C_{ins}. From $C_G = C_{ins}C_S \big/ \left(C_{ins} + C_S\right)$, we obtain a semiconductor capacitance of $C_S = 0.67C_{ins}$. A simple calculation of the quantum capacitance, however, shows that $C_Q \sim 1.5C_{ins}$, which indicates that C_S is a factor of ~2 less than C_Q. As discussed by Pal [87], this occurs when the shape of the quantum well is bias-dependent.

According to Fig. 6.8d, at V_{GS} = 0.4V, the charge at the top of the barrier under equilibrium conditions is $N_S \approx 1.5\times10^{12}$ per cm^2. The value found from the simulation under V_{DS} = 0.35V is $N_S \approx 1.3\times10^{12}$ per cm^2, which is lower than the equilibrium value. It might be expected that *DIBL* would reduce V_T and therefore increase the charge. Part of the reason for the lower charge under drain bias could be that only the positive velocity states are occupied at high V_D. The quantum capacitance, therefore, decreases under large drain bias by a factor of two. The lower C_Q lowers the semiconductor capacitance C_S and offsets the *DIBL*. The result is that the charge at the top of the barrier is somewhat less under high V_{DS}.

The second question had to do with the value of the ballistic velocity from the numerical simulation as compared to the value expected from the bandstructure. For a given $E(k)$ and Fermi level, we can determine the corresponding N_S and $\langle \upsilon \rangle = \upsilon_{ave}$ under ON-state conditions where only $+k$ states are occupied. Figure 6.9 shows the result for the parabolic effective mass (EMA) dispersion used in the quantum simulations (square-blue). For comparison, the InAs and GaAs velocities are shown, calculated using dispersions extracted from an atomistic tight-binding model [39]. The weighted average of these two results is also shown in Fig. 6.9 (solid-brown). The weighted average tight-binding results resemble the effective mass results for the $In_{0.7}Ga_{0.3}As$ channel. The EMA velocity is in good agreement with the "weighted average" curve at low carrier densities, but at higher densities, the EMA velocity is higher, because non-parabolicity reduces the velocity in the tight-binding model. At an inversion charge density of $N_S = 1.3\times 10^{12}$ per cm^2, which corresponds to the charge at the top of the barrier in the numerical simulation, the velocity for the EMA is $\upsilon_{ave} = \upsilon_{inj} \approx 4\times 10^7$ cm/s, while for the weighted average tight-binding curve it is $\upsilon_{ave} = \upsilon_{inj} \approx 3.6\times 10^7$ cm/s. These values are both higher than the $\upsilon_{ave} \approx 2.7\times 10^7$ extracted from the NEGF simulation.

The difference in the velocities deduced from the bandstructure and that extracted from the NEGF simulation might have to do with tunneling currents and quantum mechanical reflections around the top-of-the-barrier, which tend to reduce the average velocity. (In support of this conjecture, we note that the *Fermi level* in the quantum model is almost a k_BT closer to E_C than in the semiclassical model at the *same* carrier density, which indicated a carrier population below the top-of-the-barrier, and/or "negative" going state population in the quantum model). It is also evident in Fig. 6.9 that nonparabolicity can be important at this bias regime and can cause about 10%

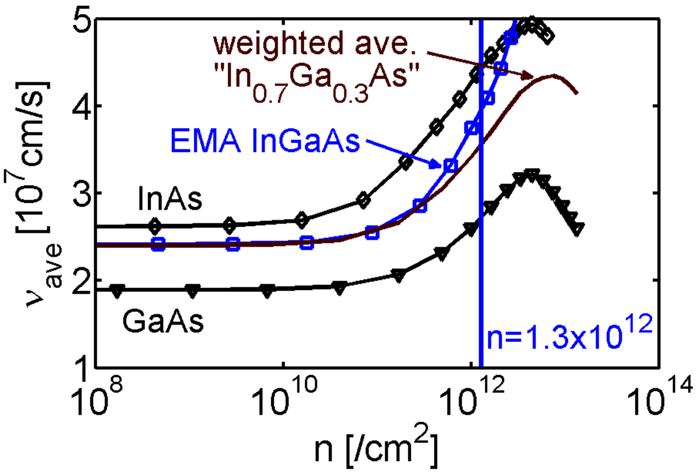


Fig. 6. 9. The “positive going” average bandstructure velocity vs. inversion carrier density of a 15nm thick quantum well, using a simple semiclassical ballistic model. The velocities of InAs and GaAs are shown in solid-square-black. Their bandstructures are calculated using an atomistic tight-binding model. The EMA bandstructure velocity for the dispersion used in the quantum simulation is shown in solid-circle-blue. The weighted average of the InAs and GaAs (In$_{0.7}$Ga$_{0.3}$As) velocity is shown in solid-brown.

degradation in the average carrier velocity. Nonparabolity is another possible contribution to the g_m degradation observed in the experimental data but not captured in the EMA treatment.

6.4.4 Scaling issues: Insulator thickness and gate length scaling

The main analysis of the discussion section up to now considered the t_{ins} = 3nm and L_G = 60nm device. The experimental data show variations in both changes in the InAlAs insulator thickness as well as gate length dependence. These two issues are briefly discussed here. Figure 6.10a shows how the insulator thickness affects the performance of the L_G = 60nm device. The equilibrium carrier density in the channel under V_{GS} = 0.4V is shown in solid-square-blue, extracted as in Fig. 6.8d for all devices at the same V_G - V_T. The carrier density in the channel doubles as the insulator thickness decreases from t_{ins} = 10 nm to t_{ins} = 3 nm – as expected. Under a high drain bias of V_D = 0.35V, however, the carrier density at the top of the barrier (dash/dot-diamond-black),

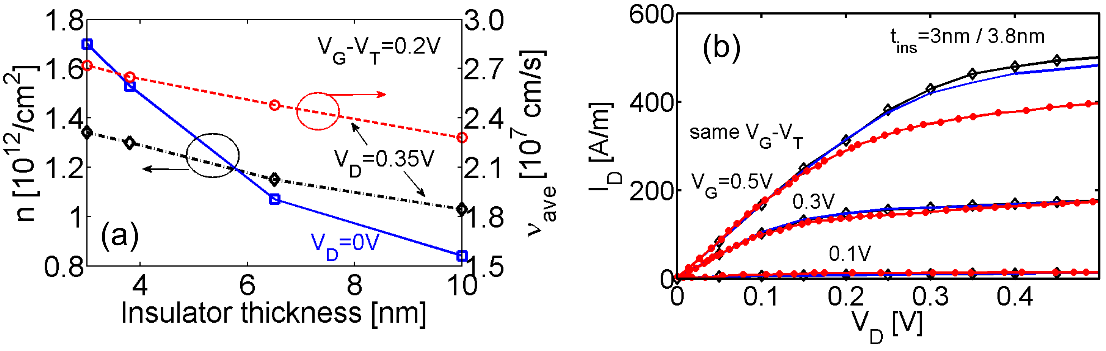


Fig. 6. 10. t_{ins} dependence of the HEMTs: (a) The simulated carrier density and average velocity at the same V_G-V_T = 0.2V as a function of insulator thickness for the L_G = 60nm device. Carrier densities for the V_D = 0V (solid-square-blue) and V_D = 0.35V (dash/dot-diamond-black) are presented. The gate bias is V_G = 0.4V. The V_G-V_T = 0.2V is the same for all insulator thickness devices at V_D = 0V. No further V_T adjustment was performed for the V_D = 0.35V case. The average velocity (dash-circle-red) is calculated at V_D = 0.35V. (b) The simulated and measured data are presented in a similar way to Fig. 6.4a for V_G = 0.1V 0.3V and 0.5V and for t_{ins} = 3nm (black-diamond), and 3.8nm (blue-solid). Variations in the insulator thickness do not introduce significant variations in the on-current.

does not vary as significantly with insulator thickness. This occurs because under high drain biases C_Q decreases by a factor of ~2, which drives the device toward the quantum capacitance limit in which variations in C_{OX} are not as significant. Larger DIBL in the low C_{OX} devices lowers the V_T and increases the charge in the channel. An interplay between these two effects reduces the charge variations as a function of $t_{ins.}$ The increase in charge as the t_{ins} is scaled from 10nm to 3nm is only ~30%. The velocity at the top of the barrier (dash-circle-red) shows an increase of ~20% with insulator thickness scaling. Scaling the insulator thickness down to 3nm can, therefore improve performance. Further scaling of the insulator, however, might not offer additional advantage at the on-state. Figure 6.10b shows the effect of scaling the insulator from t_{ins} = 3.8nm to t_{ins} = 3.0nm. This figure is the same as Fig. 6.4a, with the t_{ins} = 3.0nm result also shown in black-diamond, plotted at the same V_G-V_T. The difference at the on-state is less than 5%.

We next investigate the gate length dependence of the t_{ins} = 3 nm HEMTs. Experimentally in [3], L_G = 60nm, 85nm and 135nm devices were reported. Significant gate length dependence was observed experimentally, with the ON-current decreasing as

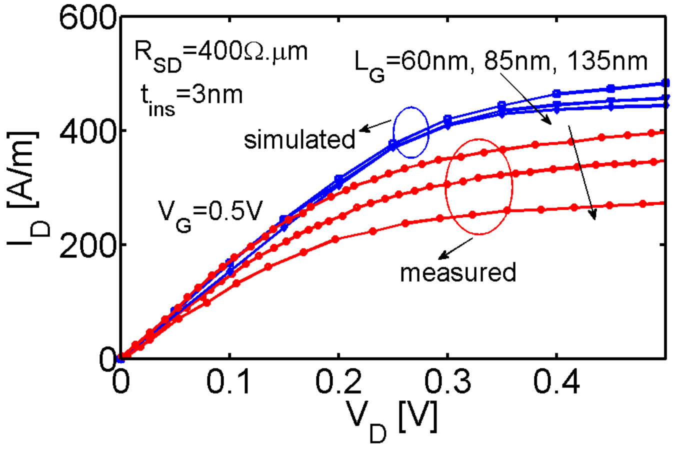


Fig. 6. 11. Gate length dependence of the 3nm oxide thickness device. The simulated and measured data are presented in a similar way to Fig. 6.4a for V_G= 0.1V, 0.3V and 0.5V and for L_G= 60nm, 85nm, 135nm. The simulated and measured data are in good agreement for the lower gate bias cases. Significant deviation is observed for the high bias cases, which is reduced as the gate length reduces. The simulated data do not show significant gate length dependence.

the gate length increases. This trend is shown in Fig. 6.11. This figure is the same as Fig. 6.4a, with all the three gate length data included (for clarity, we have shown only the highest gate voltage in each case). The solid-circle-red lines present the experimental data for the different gate lengths and for V_G = 0.5V, 0.3V, and 0.1V. The solid-blue lines present the simulated results for the same devices after the series resistance was included. Although it is not shown in the plots, a good match was observed between the simulated and measured data for lower gate biases. For the high gate bias case, the simulated results show little gate length dependence – as it is expected from a ballistic model. The small differences originate from the changes in the electrostatics. The measured high V_G data, however, show a significant gate length dependence. The longest device is about 40% below the ballistic simulation while the shortest device is only ~15% below. These results indicate increased scattering in the L_G = 85nm device and even stronger scattering in the L_G = 135nm device.

Finally, we should mention once again some of the uncertainties and simplifications that affect our analysis. The first is the ±1 nm uncertainty in the etched AlInAs layer thickness, which however does not introduce considerable uncertainty at the

ON-state. Second, the simplified device structure for the simulation had the source/drain regions that were only 60 nm long rather than 1 μm as in the experimental device. This simplification is likely to affect the high current region, where source design issues are expected to become important. Lattice distortions and the effect of strain in the channel were not considered and may have an impact on the effective mass of the channel. Except for two-dimensional electrostatics, a ballistic model should show no channel length dependence. The observed channel length dependence of the experimental devices suggests that the longer channel length devices are not as close to the ballistic limit as the 60 nm device that we examined.

6.5 Conclusion

The performance of recently demonstrated high-performance $In_{0.7}Ga_{0.3}As$ HEMTs was investigated using a quantum ballistic model self consistently coupled to a 2D Poisson solver for electrostatics. With the addition of external series resistors, reasonable agreement between the ballistic simulation and the experimental data was obtained. All of the *6*0 nm channel length devices with insulator thicknesses of 3nm, 7nm, and 11nm can be explained in this way using values of series resistance consistent with those measured in the experiments. Despite the simplifications in the model and the uncertainties in the exact values of the insulator thickness, series resistance and channel effective masses, these results suggest that 60nm channel length III-V HEMTs operate close to the ballistic limit. The on-current performance of longer channel lengths HEMTs appears to be degraded by scattering although they still operate at over one-half of the ballistic limit.

For operation near the ballistic limit, the ballistic injection velocity rather than bulk mobility becomes the parameter of interest. The ballistic injection velocity for this device was found to be relatively low for this light effective mass material, because of the relatively low inversion charge operating conditions, quantum tunneling and reflections, and conduction band non-parabolicity. The semiconductor capacitance also plays an important role by increasing the effective oxide thickness (EOT) of the thinnest insulator device by 2.5 times. Source exhaustion, an effect related to the sheet carrier density in the source, was identified as a possible explanation for the transconductance degradation that is experimentally observed in these devices. The results reported here suggest that source design is an important factor for III-V FETs, as has been recently pointed out by Fischetti [103]. They also demonstrate that the key factors in improving III-V HEMT performance are reduction of the parasitic series resistance, optimization of the source design, and reduction of the insulator thickness.

7. CONCLUSIONS AND FUTURE WORK

7.1 New physics in nanoelectronic devices

The focus of this thesis was to investigate issues that will influence the performance of a variety of possible future generation nanoelectronic devices. Quantum and atomistic effects were investigated for carbon nanotube, nanowire and III-V FETs. Quantum treatment that includes full atomistic effects or atomistic information is essential when the device dimensions scale down to the nanometer size. The effect of structure imperfections and atomistic defects on CNTFETs was investigated using a full quantum mechanical (NEGF) model and atomistic description of the channel. The results presented should be understood as a more general trend of the sensitivity of 1D transport channels to single atomistic deformations that hold for CNTs, NWs, and even as recently observed, graphene based channels [33, 115].

In the context of atomistic effects and their influence on nanoscale devices, the effect of bandstructure on the properties of the electronic structure of nanowires, both n- and p-type, was investigated using atomistic TB models. Important considerations on the influence of bandstructure on the dispersions of wires are addressed that clearly indicated that new physics appear at the nanoscale that cannot be captured by the traditional effective mass approximation, especially for the p-type devices. Understanding the scale at which these effects influence the device operation, can point towards directions of improving their operation. For example, understanding the shape of nanowire dispersion, can provide insight into the most efficient operating regime for nanowires, which includes the transport and quantization directions, as well as the quantization dimension and shapes. In the case of III-V devices, using the least amount of fitting parameters (a series resistance and the workfucntion difference between the gate electrode and the insulator), and the most amount of physics, we were able to determine the correct parameters that have the largest impact on the performance of the device and provide useful guidance to the experimentalists.

7.2 Carbon based electronics

Carbon electronics consisting of CNT and more recently graphene and nanoribbon based channels (GNR) [33, 115] have been investigated by various groups as a potential replacement of silicon, which served as the main electronic platform for decades. Much of the interest originated from the extremely high mobilities observed in CNT and GNR materials, and the indications of close to ballistic electron transport operation. Fabrication challenges, however, concerning issues such as the precise control over the CNT chirality, metallic vs. semiconducting tube separation, alignment and growth, still need to be overpassed for carbon devices to become technologically useful as computation elements. In additions to that, this work has rigorously shown that any atomistic variations and charged impurities in the CNT will have a large effect on the channel. This is up to the circuit designers finally, to implement variation insensitive designs to host such channels.

Other applications regarding the use of CNTs for electronic applications have recently drawn a lot of attention. One of them is the use of CNTs as band-to-band-tunneling (BTBT) FET devices. The idea is to bias the CNT in a different way in order to utilize the tunneling current for the device operation. This approach has been both experimentally [116] and theoretically demonstrated [117], and seems to provide some advantage over the conventional thermionic FET operation in the OFF-state by reducing the subthreshold swing below 60mV/dec. This however, is still subject to the same fabrication issues that make the entire carbon technology inattractive at the moment. The second application is the use of CNT bundle network, of CNTs randomly distributed on a substrate for microelectronic applications such as sensors, solar cells, or their possibility as thermoelectric materials. This type of applications might not require the sensitivity in the CNT type as for logic applications.

7.3 Nanowire based electronics

Nanowire and multi-gate devices are promising devices for future technology nodes that can provide electrostatic integrity for controlling the short channel effects. Their properties are sensitive to channel variations, but not in a degree that cannot be controlled. FinFET type of devices scaled down to 10nm x 10nm cross section sides can be thought as nanowires, and have been already demonstrated [7-9]. The important about these is that at some degree, a designer can alter the electronic structure of these devices using strain engineering and the proper surface quantizations and transport orientations. This can offer some electrostatic advantage over the planar devices and since it is still based on silicon, potentially much easier to be implemented rather than alternative device architectures and materials.

7.4 III-V based electronics

III-V electronics is a promising candidate for low power, high speed devices, and can offer some advantage over Si devices in gate delay, f_t and power dissipation. InGaAs devices demonstrated down to 40nm gate lengths indicate that this is possible once the device is properly optimized [4]. III-V based devices, however, have their own problems that need to be addressed. In our work, we indicated that the source design is exceedingly more important than in the case of Si devices, and needs to be one o the main issues to be addressed. Large part of the problem comes from the low density of states of III-Vs, which does not allow a large amount of carriers in the source of the device. A properly optimized III-V device, can offer some advantage over Si in certain properties, however, the light mass advantage might be proven to be a burden at extreme scaling of the channel lengths down to the 5nm gate lengths, since tunneling might then be enhanced and dramatically increase the power dissipation of the transistor. Nevertheless, if not for logic devices at the end of the roadmap, III-V devices might be utilized as high speed, low power applications.

7.5 Future outlook

A list of possible future work is presented below:

Chapters 4 and 5 in this thesis have investigated bandstructure effects of n-type and p-type nanowire devices and compared the performance of devices in different orientations. A proper investigation of the performance of these devices needs to include treatment of the experimental device cross section, with rounded edges, surface imperfections, strain fields and distortion. FinFET type geometries are strong candidates for electronic devices, and similar studies on those geometries need to be performed in order to improve the understanding of nanowire and FinFET type devices. NEGF simulations on the full 3D device instead of the 2D cross section might need to be performed for this task. Limited size cross section devices can be simulated with NEGF with enough compute power [27, 81], however for realistic cross sections, approximations such as mode space techniques need to be implemented. The zone unfolding method [118, 119] has been recently developed within the tight-binding context in order to provide approximate bandstructures in systems with distortions. This technique can be used in order to extract approximate effective masses to be used in effective mass simulations, in order to avoid the computationally expensive NEGF formalism on atomistic TB Hamiltonians.

Recently a lot of attention has been drawn in III-V type of devices and an increase in the corresponding publications in IEDM, TED and EDL. This large interest, calls for the collaboration between the experimental and the simulation groups to identify crucial issues in III-V device operation. The work in chapter 6 points to several III-V HEMT design directions. The effect of variations and uncertainties in various parameters such as the insulator thickness, the confinement and transport masses, lattice distortions and strain, channel width need to be investigated. Devices with higher mobility InAs channels have also recently been reported. It is of technological interest to identify if these devices are much closer to the ballistic limit and can utilize in larger degree the high channel mobility.

Another direction will be the ballistic quantum treatment of p-type quantum well devices. For this, the full atomistic Hamiltonian with spin-orbit coupling needs to be used within the NEGF formalism. This will include the non-parabolicity and anisotropy effects that are enhanced in p-type devices as shown in chapter 5. NEGF treatment on the atomistic Hamiltonian including all transverse *k*-space modes is computationally challenging, but can be simulated for thin (a few nm) quantum wells. For thicker quantum wells, advanced numerical algorithms and parallelization might need to be implemented.

Finally, the effect of strain in nanowire and III-V devices in various transport and confinement orientations is an additional direction of technological interest. Strain engineering is fully utilized in current technology nodes and it will continue towards future device generations, whether nanowire, FinFET, III-V quantum wells are utilized.

LIST OF REFERENSES

LIST OF REFERENCES

[1] ITRS, "ITRS Public Home Page. http://www.itrs.net/reports.html."

[2] R. Chau, S. Datta, M. Doczy, B. Doyle, J. Jin, J. Kavalieros, A. Majumdar, M. Metz, and M. Radosavljevic, "Benchmarking nanotechnology for high-performance and low-power logic transistor applications," *IEEE Transactions on Nanotechnology*, vol. 4, pp. 153-158, 2005.

[3] D. H. Kim and J. A. del Alamo, "Scaling behavior of In0.7Ga0.3AsHEMTs for logic," *2006 International Electron Devices Meeting, Vols 1 and 2*, pp. 587-590, 2006.

[4] K. Dae-Hyun and J. A. del Alamo, "Logic performance of 40 nm InAs HEMTs," Washington, DC, USA, 2007.

[5] Y. Xuan, Y. Q. Wu, T. Shen, T. Yang, and P. D. Ye, "High performance submicron inversion-type enhancement-mode InGaAs MOSFETs with ALD Al_2O_3, HfO_2 and HfAlO as gate dielectrics," Washington, DC, USA, 2007.

[6] M. Passlack, P. Zurcher, K. Rajagopalan, R. Droopad, J. Abrokwah, M. Tutt, Y. B. Park, E. Johnson, O. Hartin, A. Zlotnicka, P. Fejes, R. J. W. Hill, D. A. J. Moran, X. Li, H. Zhou, D. Macintyre, S. Thorns, A. Asenov, K. Kalna, and I. G. Thayne, "High mobility III-V MOSFETs for RF and digital applications," Washington, DC, USA, 2007.

[7] B. S. Doyle, S. Datta, M. Doczy, S. Hareland, B. Jin, J. Kavalieros, T. Linton, A. Murthy, R. Rios, and R. Chau, "High performance fully-depleted tri-gate CMOS transistors," *IEEE Electron Device Letters*, vol. 24, pp. 263-265, 2003.

[8] Y. Bin, C. Leland, S. Ahmed, W. Haihong, S. Bell, Y. Chih-Yuh, C. Tabery, H. Chau, X. Qi, K. Tsu-Jae, J. Bokor, H. Chenming, L. Ming-Ren, and D. Kyser, "FinFET scaling to 10 nm gate length," San Francisco, CA, USA, 2002.

[9] R. Chau, "Benchmarking nanotechnology for high-performance and low-power logic transistor applications," *2004 4th IEEE Conference on Nanotechnology*, pp. 3-6, 2004.

[10] J. Xiang, W. Lu, Y. J. Hu, Y. Wu, H. Yan, and C. M. Lieber, "Ge/Si nanowire heterostructures as high-performance field-effect transistors," *Nature*, vol. 441, pp. 489-493, 2006.

[11] N. Singh, F. Y. Lim, W. W. Fang, S. C. Rustagi, L. K. Bera, A. Agarwal, C. H. Tung, K. M. Hoe, S. R. Omampuliyur, D. Tripathi, A. O. Adeyeye, G. Q. Lo, N. Balasubramanian, and D. L. Kwong, "Ultra-narrow silicon nanowire Gate-All-Around CMOS devices: Impact of diameter, channel-orientation and low temperature on device performance," *2006 International Electron Devices Meeting, Vols 1 and 2*, pp. 294-297, 2006.

[12] N. Singh, A. Agarwal, L. K. Bera, T. Y. Liow, R. Yang, S. C. Rustagi, C. H. Tung, R. Kumar, G. Q. Lo, N. Balasubramanian, and D. L. Kwong, "High-

performance fully depleted silicon-nanowire (diameter <= 5 nm) gate-all-around CMOS devices," *IEEE Electron Device Letters*, vol. 27, pp. 383-386, 2006.

[13] K. H. Cho, S. D. Suk, Y. Y. Yeoh, M. Li, K. H. Yeo, D. W. Kim, S. W. Hwang, D. Park, and B. I. Ryu, "Observation of single electron tunneling and ballistic transport in twin silicon nanowire MOSFETs (TSNWFETs) fabricated by top-down CMOS process," *2006 International Electron Devices Meeting, Vols 1 and 2*, pp. 290-293, 2006.

[14] K. H. Cho, K. H. Yeo, Y. Y. Yeoh, S. D. Suk, M. Li, J. M. Lee, M. S. Kim, D. W. Kim, D. Park, B. H. Hong, Y. C. Jung, and S. W. Hwang, "Experimental evidence of ballistic transport in cylindrical gate-all-around twin silicon nanowire metal-oxide-semiconductor field-effect transistors," *Applied Physics Letters*, vol. 92, 2008.

[15] M. Kobayashi and T. Hiramoto, "Experimental study on quantum confinement effects in silicon nanowire metal-oxide-semiconductor field-effect transistors and single-electron transistors," *Journal of Applied Physics*, vol. 103, 2008.

[16] P. L. McEuen, M. S. Fuhrer, and H. K. Park, "Single-walled carbon nanotube electronics," *IEEE Transactions on Nanotechnology*, vol. 1, pp. 78-85, 2002.

[17] A. Javey, R. Tu, D. B. Farmer, J. Guo, R. G. Gordon, and H. J. Dai, "High performance n-type carbon nanotube field-effect transistors with chemically doped contacts," *Nano Letters*, vol. 5, pp. 345-348, 2005.

[18] J. Chen, C. Klinke, A. Afzali, and P. Avouris, "Self-aligned carbon nanotube transistors with charge transfer doping," *Applied Physics Letters*, vol. 86, 2005.

[19] Z. Yao, C. L. Kane, and C. Dekker, "High-field electrical transport in single-wall carbon nanotubes," *Physical Review Letters*, vol. 84, pp. 2941-2944, 2000.

[20] P. J. Burke, "An RF circuit model for carbon nanotubes," *Ieee Transactions on Nanotechnology*, vol. 2, pp. 55-58, 2003.

[21] V. Derycke, R. Martel, J. Appenzeller, and P. Avouris, "Carbon nanotube inter- and intramolecular logic gates," *Nano Letters*, vol. 1, pp. 453-456, 2001.

[22] www.intel.com.

[23] J. Wang and M. Lundstrom, "Does source-to-drain tunneling limit the ultimate scaling of MOSFETs?," *International Electron Devices 2002 Meeting, Technical Digest*, pp. 707-710, 2002.

[24] S. Datta, *Electronic Transport in Mesoscopic Systems*: Cambridge MA: Cambridge Univ. Press, 1997.

[25] Z. B. Ren, R. Venugopal, S. Goasguen, S. Datta, and M. S. Lundstrom, "nanoMOS 2.5: A two-dimensional simulator for quantum transport in double-gate MOSFETs," *IEEE Transactions on Electron Devices*, vol. 50, pp. 1914-1925, 2003.

[26] M. Luisier, A. Schenk, W. Fichtner, and G. Klimeck, "Atomistic simulation of nanowires in the $sp^3d^5s^*$ tight-binding formalism: From boundary conditions to strain calculations," *Physical Review B*, vol. 74, 2006.

[27] M. Luisier, A. Schenk, and W. Fichtner, "Full-band atomistic study of source-to-drain tunneling in Si nanowire transistors," *Sispad 2007: Simulation of Semiconductor Processes and Devices 2007*, pp. 221-224, 2007.

[28] M. Luisier, A. Schenk, and W. Fichtner, "Atomistic treatment of interface roughness in Si nanowire transistors with different channel orientations," *Applied Physics Letters*, vol. 90, 2007.

[29] J. Guo, S. Datta, M. Lundstrom, and M. P. Anantam, "Towards Multi-Scale Modeling of Carbon Nanotube Transistors," 2005.

[30] S. O. Koswatta, S. Hasan, M. S. Lundstrom, M. P. Anantram, and D. E. Nikonov, "Nonequilibrium green's function treatment of phonon scattering in carbon-nanotube transistors," *IEEE Transactions on Electron Devices*, vol. 54, pp. 2339-2351, 2007.

[31] N. Neophytou, D. Kienle, E. Polizzi, and M. P. Anantram, "Influence of defects on nanotube transistor performance," *Applied Physics Letters*, vol. 88, 2006.

[32] N. Neophytou, S. Ahmed, and G. Klimeck, "Influence of vacancies on metallic nanotube transport properties," *Applied Physics Letters*, vol. 90, 2007.

[33] G. C. Liang, N. Neophytou, M. S. Lundstrom, and D. E. Nikonov, "Ballistic graphene nanoribbon metal-oxide-semiconductor field-effect transistors: A full real-space quantum transport simulation," *Journal of Applied Physics*, vol. 102, 2007.

[34] J. Wang, A. Rahman, A. Ghosh, G. Klimeck, and M. Lundstrom, "On the validity of the parabolic effective-mass approximation for the I-V calculation of silicon nanowire transistors," *IEEE Transactions on Electron Devices*, vol. 52, pp. 1589-1595, 2005.

[35] E. Gnani, S. Reggiani, A. Gnudi, P. Parruccini, R. Colle, M. Rudan, and G. Baccarani, "Band-structure effects in ultrascaled silicon nanowires," *IEEE Transactions on Electron Devices*, vol. 54, pp. 2243-2254, 2007.

[36] M. Bescond, N. Cavassilas, and M. Lannoo, "Effective-mass approach for n-type semiconductor nanowire MOSFETs arbitrarily oriented," *Nanotechnology*, vol. 18, 2007.

[37] Y. Liu, N. Neophytou, T. Low, G. Klimeek, and M. S. Lundstrom, "A tight-binding study of the ballistic injection velocity for ultrathin-body SOI MOSFETs," *IEEE Transactions on Electron Devices*, vol. 55, pp. 866-871, 2008.

[38] N. Neophytou, A. Paul, M. Lundstrom, and G. Klimeck, "Simulations of nanowire transistors: Atomistic vs. effective mass models," *Journal of Computational Electronics*, vol. DOI10.1007/s10825-008-0188-4, 2008.

[39] G. Klimeck, S. S. Ahmed, H. Bae, N. Kharche, R. Rahman, S. Clark, B. Haley, S. H. Lee, M. Naumov, H. Ryu, F. Saied, M. Prada, M. Korkusinski, and T. B. Boykin, "Atomistic simulation of realistically sized nanodevices using NEMO 3-D - Part I: Models and benchmarks," *IEEE Transactions on Electron Devices*, vol. 54, pp. 2079-2089, 2007.

[40] G. Klimeck, R. C. Bowen, T. B. Boykin, C. Salazar-Lazaro, T. A. Cwik, and A. Stoica, "Si tight-binding parameters from genetic algorithm fitting," *Superlattices and Microstructures*, vol. 27, pp. 77-88, 2000.

[41] T. B. Boykin, G. Klimeck, and F. Oyafuso, "Valence band effective-mass expressions in the $sp^3d^5s^*$ empirical tight-binding model applied to a Si and Ge parametrization," *Physical Review B*, vol. 69, 2004.

[42] G. Klimeck, F. Oyafuso, T. B. Boykin, R. C. Bowen, and P. Allmen von, *Computer Modeling in Engineering and Science (CMES)*, vol. 3, pp. 601-642, 2002.

[43] J. C. Slater and G. F. Koster, "Simplified LCAO Method for the Periodic Potential Problem," *Physical Review*, vol. 94, pp. 1498-1524, 1954.

[44] P. O. Lowdin, "On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals," *Journal of Chemical Physics*, vol. 18, pp. 365-375, 1950.

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

[46] Personal communication with Prof. Mark Schilfgaarde (Arizona State University) for the GW calculations and with Dr. Tony Low (Purdue University) for the pseudopotential calculations.

[47] N. Neophytou, A. Shaikh, and G. Klimeck, "Non-Equilibrium Green's Function (NEGF) Simulation of Metallic Carbon Nanotubes: The Effect of the Vacancy Defect," *Journal of Computational Electronics*, 2007.

[48] N. Neophytou, A. Paul, M. Lundstrom, and G. Klimeck, "Self-consistent simulations of nanowire transistors using atomistic basis sets," *Sispad 2007: Simulation of Semiconductor Processes and Devices*, pp. 217-220, 2007.

[49] N. Neophytou, A. Paul, M. Lundstrom, and G. Klimeck, "Bandstructure effects in silicon nanowire electron transport," *IEEE Transactions on Electron Devices,*2008. Also in: *arXiv:0801.0125v2*, 2008.

[50] N. Neophytou, A. Paul, and G. Klimeck, "Bandstructure effects in silicon nanowire hole transport," *IEEE Transactions on Electron Devices, submitted*, 2008.

[51] N. Neophytou, T. Rakshit, and M. Lundstrom, "Performance Analysis of 60nm gate length III-V InGaAs HEMTs: Simulations vs. experiments," *IEEE Transactions on Electron Devices, submitted*, 2008.

[52] P. L. McEuen, M. Bockrath, D. H. Cobden, and J. G. Lu, "One dimensional transport in carbon nanotubes," *Microelectronic Engineering*, vol. 47, pp. 417-420, 1999.

[53] L. Chico, L. X. Benedict, S. G. Louie, and M. L. Cohen, "Quantum conductance of carbon nanotubes with defects," *Physical Review B*, vol. 54, pp. 2600-2606, 1996.

[54] D. Kienle, J. I. Cerda, and A. W. Ghosh, "Extended Huckel theory for band structure, chemistry, and transport. I. Carbon nanotubes," *Journal of Applied Physics*, vol. 100, 2006.

[55] M. P. Anantram and T. R. Govindan, "Conductance of carbon nanotubes with disorder: A numerical study," *Physical Review B*, vol. 58, pp. 4882-4887, 1998.

[56] A. A. El-Barbary, R. H. Telling, C. P. Ewels, M. I. Heggie, and P. R. Briddon, "Structure and energetics of the vacancy in graphite," *Physical Review B*, vol. 68, 2003.

[57] A. Hashimoto, K. Suenaga, A. Gloter, K. Urita, and S. Iijima, "Direct evidence for atomic defects in graphene layers," *Nature*, vol. 430, pp. 870-873, 2004.

[58] A. V. Krasheninnikov, K. Nordlund, and J. Keinonen, "Production of defects in supported carbon nanotubes under ion irradiation," *Physical Review B*, vol. 65, 2002.

[59] Y. W. Fan, B. R. Goldsmith, and P. G. Collins, "Identifying and counting point defects in carbon nanotubes," *Nature Materials*, vol. 4, pp. 906-911, 2005.

[60] A. V. Krasheninnikov, K. Nordlund, M. Sirvio, E. Salonen, and J. Keinonen, "Formation of ion-irradiation-induced atomic-scale defects on walls of carbon nanotubes," *Physical Review B*, vol. 6324, 2001.

[61] C. Gomez-Navarro, P. J. De Pablo, J. Gomez-Herrero, B. Biel, F. J. Garcia-Vidal, A. Rubio, and F. Flores, "Tuning the conductance of single-walled carbon nanotubes by ion irradiation in the Anderson localization regime," *Nature Materials*, vol. 4, pp. 534-539, 2005.

[62] S. Lee, G. Kim, H. Kim, B. Y. Choi, J. Lee, B. W. Jeong, J. Ihm, Y. Kuk, and S. J. Kahng, "Paired gap states in a semiconducting carbon nanotube: Deep and shallow levels," *Physical Review Letters*, vol. 95, 2005.

[63] J. Lu, S. Nagase, S. Zhang, and L. M. Peng, "Energetic, geometric, and electronic evolutions of K-doped single-wall carbon nanotube ropes with K intercalation concentration," *Physical Review B*, vol. 69, 2004.

[64] M. Radosavljevic, J. Appenzeller, P. Avouris, and J. Knoch, "High performance of potassium n-doped carbon nanotube field-effect transistors," *Applied Physics Letters*, vol. 84, pp. 3693-3695, 2004.

[65] T. Takenobu, T. Takano, M. Shiraishi, Y. Murakami, M. Ata, H. Kataura, Y. Achiba, and Y. Iwasa, "Stable and controlled amphoteric doping by encapsulation of organic molecules inside carbon nanotubes," *Nature Materials*, vol. 2, pp. 683-688, 2003.

[66] C. Jo, C. Kim, and Y. H. Lee, "Electronic properties of K-doped single-wall carbon nanotube bundles," *Physical Review B*, vol. 65, 2002.

[67] R. C. Bowen, G. Klimeck, R. K. Lake, W. R. Frensley, and T. Moise, "Quantitative simulation of a resonant tunneling diode," *Journal of Applied Physics*, vol. 81, pp. 3207-3213, 1997.

[68] C. Bowen, C. L. Fernando, G. Klimeck, A. Chatterjee, D. Blanks, R. Lake, J. Hu, J. Davis, M. Kulkarni, S. Hattangady, and I. C. Chen, "Physical oxide thickness extraction and verification using quantum mechanical simulation," Washington, DC, USA, 1997.

[69] G. Klimeck, T. B. Boykin, R. C. Bowen, R. Lake, D. Blanks, T. Moise, K. Yung-Chung, and W. R. Frensley, "Quantitative simulation of strained and unstrained InP-based resonant tunneling diodes," Fort Collins, CO, USA, 1997.

[70] 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, 2003.

[71] J. Wang, "Device physics and simulation of silicon nanowire transistors," in *Electrical and Computer Enginnering*. West Lafayette, IN: Purdue University, 2006.

[72] X. Y. Zhao, C. M. Wei, L. Yang, and M. Y. Chou, "Quantum confinement and electronic properties of silicon nanowires," *Physical Review Letters*, vol. 92, 2004.

[73] N. Kharche, M. Prada, T. B. Boykin, and G. Klimek, "Valley splitting in strained silicon quantum wells modeled with 2 degrees miscuts, step disorder, and alloy disorder," *Applied Physics Letters*, vol. 90, 2007.

[74] R. Rahman, C. J. Wellard, F. R. Bradbury, M. Prada, J. H. Cole, G. Klimeck, and L. C. L. Hollenberg, "High precision quantum control of single donor spins in silicon," *Physical Review Letters*, vol. 99, 2007.

[75] G. C. Liang, J. Xiang, N. Kharche, G. Klimeck, C. M. Lieber, and M. Lundstrom, "Performance analysis of a Ge/Si core/shell nanowire field-effect transistor," *Nano Letters*, vol. 7, pp. 642-646, 2007.

[76] A. Rahman, "Exploring new channels materials for nanoscale CMOS devices: A simulation approach," in *Electrical and Computer Engineering*. West Lafayette, IN: Purdue University, 2006.

[77] M. Lundstrom and J. Guo, *Nanoscale transistors: Device physics, modeling and simulation*: Springer, 2006.

[78] A. Rahman, J. Guo, S. Datta, and M. S. Lundstrom, "Theory of ballistic nanotransistors," *IEEE Transactions on Electron Devices*, vol. 50, pp. 1853-1864, 2003.

[79] We have conducted extensive tests to validate the 2D Poisson solution compared to the actual 3D solution. A maximum deviation of 2% on band edges between the two approaches was found, and ignorable deviation in the I-V characteristics. However, the 2D method reduces the Poisson computational time by almost 5X compared to the 3D solution.

[80] In collaboration with Abhijeet Paul.

[81] M. Luisier and A. Schenk, "Atomistic simulation of nanowire transistors," *Journal of Computational and Theoretical Nanoscience*, vol. 5, pp. 1-15, 2008.

[82] A. Rahman, G. Klimeck, and M. Lundstrom, "Novel channel materials for ballistic nanoscale MOSFETs-bandstructure effects," Washington, DC, MD, United States, 2005.

[83] Bandstructure lab on nanoHUB.org (https://www.nanohub.org/tools/bandstrlab/).

[84] R. Kim, N. Neophytou, A. Paul, G. Klimeck, and M. Lundstrom, "Nanowire Electronics: What's special about 1D?," *Journal of Vacuum Science and Technology*, vol. B, 2008.

[85] The length of the unit cell in [111] is $L_{111} = \sqrt{3}a_0$. Therefore, the Brillouin zone of a 1D nanowire in [111] extends from $-\frac{\pi}{\sqrt{3}a_0}$ to $\frac{\pi}{\sqrt{3}a_0}$. The Δ valleys in bulk Si are located at $k_x = 0.815 \cdot 2\pi / a_0$. Under quantization in (111) surface they project on the [111] axis at $k_{[111]} = 0.815 \cdot 2\pi / \left(\sqrt{3}a_0\right) = 1.63\pi / \left(\sqrt{3}a_0\right)$. The valley projection point falls in the 2nd Brillouin zone. It is then folded in the first Brillouin zone as $k_{[111]} = 1.63\pi / \left(\sqrt{3}a_0\right) - 2\pi / \left(\sqrt{3}a_0\right) = 0.37\pi / \left(\sqrt{3}a_0\right)$. After normalization to the length of the Brillouin zone, the valleys appear at $k_{]111]} = 0.37$.

[86] The oxide capacitance of the rectangular structure is 0.483 nF/m, numerically calculated using a 2D Poisson solver that takes the fringing at the edges into consideration.

[87] H. S. Pal, K. D. Cantley, S. S. Ahmed, and M. S. Lundstrom, "Influence of bandstructure and channel structure on the inversion layer capacitance of silicon and GaAs MOSFETs," *IEEE Transactions on Electron Devices*, vol. 55, pp. 904-908, 2008.

[88] T. B. Boykin, G. Klimeck, M. A. Eriksson, M. Friesen, S. N. Coppersmith, P. von Allmen, F. Oyafuso, and S. Lee, "Valley splitting in strained silicon quantum wells," *Applied Physics Letters*, vol. 84, pp. 115-117, 2004.

[89] 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," *Physical Review B*, vol. 70, 2004.

[90] A. Rahman, G. Klimeck, M. Lundstrom, T. B. Boykin, and N. Vagidov, "Atomistic approach for nanoscale devices at the scaling limit and beyond - Valley splitting in Si," *Japanese Journal of Applied Physics Part 1-Regular Papers Brief Communications & Review Papers*, vol. 44, pp. 2187-2190, 2005.

[91] A. Rahman, M. S. Lundstrom, and A. W. Ghosh, "Generalized effective-mass approach for n-type metal-oxide-semiconductor field-effect transistors on arbitrarily oriented wafers," *Journal of Applied Physics*, vol. 97, 2005.

[92] M. V. Fischetti, Z. Ren, P. M. Solomon, M. Yang, and K. Rim, "Six-band k center dot p calculation of the hole mobility in silicon inversion layers: Dependence on surface orientation, strain, and silicon thickness," *Journal of Applied Physics*, vol. 94, pp. 1079-1095, 2003.

[93] E. X. Wang, P. Matagne, L. Shifren, B. Obradovic, R. Kotlyar, S. Cea, M. Stettler, and M. D. Giles, "Physics of hole transport in strained silicon MOSFET inversion layers," *IEEE Transactions on Electron Devices*, vol. 53, pp. 1840-1851, 2006.

[94] M. De Michielis, D. Esseni, Y. L. Tsang, P. Palestri, L. Selmi, A. G. O'Neill, and S. Chattopadhyay, "A semianalytical description of the hole band structure in inversion layers for the physically based modeling of pMOS transistors," *IEEE Transactions on Electron Devices*, vol. 54, pp. 2164-2173, 2007.

[95] G. Klimeck, R. C. Bowen, and T. B. Boykin, "Strong wavevector dependence of hole transport in heterostructures," *Superlattices and Microstructures*, vol. 29, pp. 187-216, 2001.

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

[97] M. V. Fischetti and S. E. Laux, "Monte-Carlo Simulation of Transport in Technologically Significant Semiconductors of the Diamond and Zincblende Structures .2. Submicrometer Mosfets," *IEEE Transactions on Electron Devices*, vol. 38, pp. 650-660, 1991.

[98] P. M. Solomon and S. E. Laux, "The ballistic FET: design, capacitance and speed limit," Washington, DC, USA, 2001.

[99] N. Waldron, K. Dae-Hyun, and J. A. del Alamo, "90 nm self-aligned enhancement-mode InGaAs HEMT for logic applications," Washington, DC, USA, 2007.

[100] C. I. Kuo, H. T. Hsu, E. Y. Chang, C. Y. Chang, Y. Miyamoto, S. Datta, M. Radosavljevic, G. W. Huang, and C. T. Lee, "RF and logic performance improvement of In0.7Ga0.3As/InAs/In0.7Ga0.3As composite-channel HEMT using gate-sinking technology," *IEEE Electron Device Letters*, vol. 29, pp. 290-293, 2008.

[101] Personal communication with Dr. Dae-Hyun Kim and Prof. Jesus del Alamo.

[102] R. Venugopal, S. Goasguen, S. Datta, and M. S. Lundstrom, "Quantum mechanical analysis of channel access geometry and series resistance in nanoscale transistors," *Journal of Applied Physics*, vol. 95, pp. 292-305, 2004.

[103] M. V. Fischetti, T. P. O'Regan, S. Narayanan, C. Sachs, S. Jin, J. Kim, and Y. Zhang, "Theoretical study of some physical aspects of electronic transport in nMOSFETs at the 10-nm gate-length," *IEEE Transactions on Electron Devices*, vol. 54, pp. 2116-2136, 2007.

[104] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, "Band parameters for III-V compound semiconductors and their alloys," *Journal of Applied Physics*, vol. 89, pp. 5815-5875, 2001.

[105] O. Madelung, *Semiconductors: Basic Data*, vol. 2nd rev. edition. New York: Springer 1996.

[106] G. Klimeck, S. S. Ahmed, N. Kharche, M. Korkusinski, M. Usman, M. Prada, and T. B. Boykin, "Atomistic simulation of realistically sized nanodevices using NEMO 3-D - Part II: Applications," *IEEE Transactions on Electron Devices*, vol. 54, pp. 2090-2099, 2007.

[107] R. Pierret, *Semiconductor Device Fundamentals*: Addison-Wesley Publishing Company, 1996.

[108] M. S. Shur, "Low ballistic mobility in submicron HEMTs," *IEEE Electron Device Letters*, vol. 23, pp. 511-513, 2002.

[109] J. Wang and M. Lundstrom, "Ballistic transport in high electron mobility transistors," *IEEE Transactions on Electron Devices*, vol. 50, pp. 1604-1609, 2003.

[110] M. Zilli, D. Esseni, P. Palestri, and L. Selmi, "On the apparent mobility in nanometric n-MOSFETs," *IEEE Electron Device Letters*, vol. 28, pp. 1036-1039, 2007.

[111] S. E. Laux, "A simulation study of the switching times of 22-and 17-nm gate-length SOI nFETs on high mobility substrates and Si," *IEEE Transactions on Electron Devices*, vol. 54, pp. 2304-2320, 2007.

[112] E. O. Johnson, "Insulated-Gate Field-Effect Transistor - Bipolar Transistor in Disguise," *Rca Review*, vol. 34, pp. 80-94, 1973.

[113] M. Lundstrom and Z. B. Ren, "Essential physics of carrier transport in nanoscale MOSFETs," *IEEE Transactions on Electron Devices*, vol. 49, pp. 133-141, 2002.

[114] Personal communication with Himadri Pal.

[115] G. Fiori and G. Iannaccone, "Simulation of graphene nanoribbon field-effect transistors," *IEEE Electron Device Letters*, vol. 28, pp. 760-762, 2007.

[116] J. Appenzeller, Y. M. Lin, J. Knoch, and P. Avouris, "Band-to-band tunneling in carbon nanotube field-effect transistors," *Physical Review Letters*, vol. 93, 2004.

[117] S. O. Koswatta, M. S. Lundstrom, M. P. Anantram, and D. E. Nikonov, "Simulation of phonon-assisted band-to-band tunneling in carbon nanotube field-effect transistors," *Applied Physics Letters*, vol. 87, 2005.

[118] T. B. Boykin and G. Klimeck, "Practical application of zone-folding concepts in tight-binding calculations," *Physical Review B*, vol. 71, 2005.

[119] T. B. Boykin, N. Kharche, and G. Klimeck, "Brillouin-zone unfolding of perfect supercells having nonequivalent primitive cells illustrated with a Si/Ge tight-binding parameterization," *Physical Review B*, vol. 76, 2007.

[120] http://www.photon.t.u-tokyo.ac.jp/~maruyama/nanotube.html
http://www.pa.msu.edu/cmp/csc/nanotube.html

[121] http://www-dinma.univ.trieste.it/nirftc/research/easymesh/

[122] M. P. L. Sancho, J. M. L. Sancho, and J. Rubio, "Highly Convergent Schemes for the Calculation of Bulk and Surface Green-Functions," *Journal of Physics F-Metal Physics*, vol. 15, pp. 851-858, 1985.

[123] E. Polizzi and N. Ben Abdallah, "Subband decomposition approach for the simulation of quantum electron transport in nanostructures," *Journal of Computational Physics*, vol. 202, pp. 150-180, 2005.

[124] A. Svizhenko, M. P. Anantram, T. R. Govindan, B. Biegel, and R. Venugopal, "Two-dimensional quantum mechanical modeling of nanotransistors," *Journal of Applied Physics*, vol. 91, pp. 2343-2354, 2002.

APPENDICES

APPENDIX A

A THREE-DIMENSIONAL (3D) ATOMISTIC QUANTUM SIMULATIOR FOR REALISTIC PLANAR CARBON NANOTUBE FET STRUCTURES

A.1 Introduction

In this section a real space 3D quantum simulator based on atomistic nearest neighbor (NN) tight binding (TB) approach for treatment of electron transport in zigzag carbon nanotubes (CNT) is described. The code is based on the Non-Equilibrium Greens' Function method (NEGF) for quantum transport, and 3D electrostatics based on a finite element mesh (FEM) technique. The simulator is general enough to account for generic device geometries, coaxial as well as the experimentally realised planar device structures. This captures all the electrostatic effects that arise from the geometric features of the device. Furthermore, the real space atomistic representation of the CNT channel, allows for treatment of charge and potential variations around the CNT perimeter, as well as the study of atomistic defects in the device.

The appendix is organized as follows: The basic geometrical and electronic structure features of the CNT are presented. The construction of the 3D quantum simulator and its capabilities are then discussed. Finally, the parallelization scheme and the deployment of the simulator on nanoHUB.org is discussed.

A.2 Structure of the SWNT - Geometry construction

A carbon nanotube (CNT) can be visualized as a rolled-up graphene sheet in the

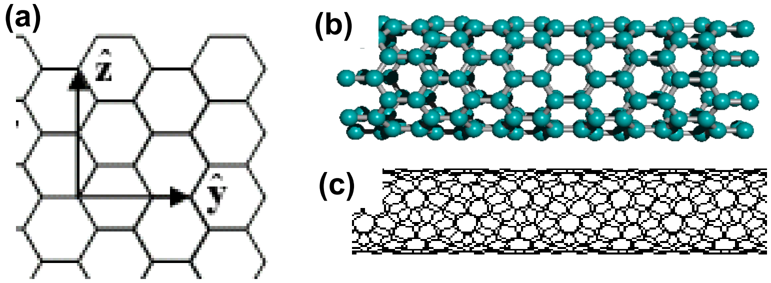


Fig. A. 1. The CNT formed by rolling up the graphene sheet of carbon atoms. Depending on the chirality, *zig-zag*, *armchair* or *chiral* tubes can exist.

shape of a cylinder of certain diameter and chirality. The chirality of a CNT can be characterized by the *(n,m)* numbers that control the direction that the sheet will "roll-up" and the size of the diameter. Depending on the diameter and the chirality, the properties of the CNT can vary from metallic to semiconducting. CNTs in which $n=m$ are called *armchair* CNTs and are metallic. CNTs with $m=0$ are called *zig-zag*, and are metallic when n is a multiple of 3, otherwise they are semiconducting. Semiconducting CNTs are direct gap materials, with the gap depending on the diameter D, with roughly $E_G \sim 1/D$. Figure A.1 (a) shows the graphene sheet structure, in which the carbon atoms are placed in a hexagonal scheme, with the carbon-carbon (C-C) bond of length 1.44Å, and the form of a zig-zag (b) and a chiral (c) CNT.

Once the chiral vector *(n,0)* is indicated, then the coordinates of the unit cell of the CNT can be constructed in 3D space. This is the first step in the atomistic simulation process. The unit cell for the specific zigzag CNT under consideration consist of four rows of carbon atoms, which form the CNT rings when the graphene is rolled into a cylinder. The coordinates of the unit cell atoms and the bonding information, are constructed in the simulator from geometrical considerations by using commercially available codes [120]. The unit cell is then repeated in the CNT transport axis to create all the atomic coordinates of the full length of the nanotube. The CNTFET device (including the gate/source/drain electrodes and the dielectric constants) is then constructed around the nanotube utilizing the geometric information of the carbon atoms.

Figure A.2 shows the device structures the simulator can handle, the simple coaxial and the planar devices. The cylindrical device (Fig. A.2 (e)) is theoretically the one that offers the best performance because of excellent electrostatics, however, what is experimentally achieved is the planar device, both bottom and top gated ones (BG and TG). Figure A.2 (a) shows the TG structure. The BG structure has the same geometry, but the top gate electrode is missing. In the case of the coaxial device, the CNT is surrounded by the gate oxide and a coaxial gate. In the cases of the planar devices, the CNT lies on a rectangular substrate which is treated as an insulator with a certain dielectric constant. For the BG device, the bottom gate electrode is used as the gate electrode of the device. In the case of the TG device, a top gate electrode is placed on the top region of the device, and is separated from the CNT channel through another dielectric material. The interior of the CNT is assumed to be a third dielectric (vacuum in general). The CNT channel is connected to the source/drain reservoirs which are assumed to be metallic electrodes. In order to capture the correct parasitics and fringing effects in the device, the metallic electrodes used for the gate, source and drain are treated as rectangular boxes with a certain width and thickness, rather than just metallic surfaces, which gives a more accurate treatment of electrostatics for the 3D treated devices. The source/drain contacts experimentally are metallic electrodes that connect to the CNT channel through Schottky Barriers. Doped source/drain contacts have however been realized, by deposition of potassium on the left/right sides of the CNT [63-66]. This, results in partial charge exchange between the potassium atoms and the CNT, such that the conduction band lowers enough, making that part of the channel acting as MOSFET type doped source/drain contacts. In the simulator, the option to choose the doping in the entire CNT channel is available, in which case, doped source/drain contacts can be treated, rather than Schottky barrier ones. In this case, the metallic electrodes may or may not used in the simulator domain.

Having all the information of the atomic position and the positions of the electrodes and the dielectric regions, the mesh of the device is constructed using the commercially available mesh 2D finite element mesh (FEM) creator Easymesh 1.4, [121]. The choice of FEM instead of other methods such as finite difference method

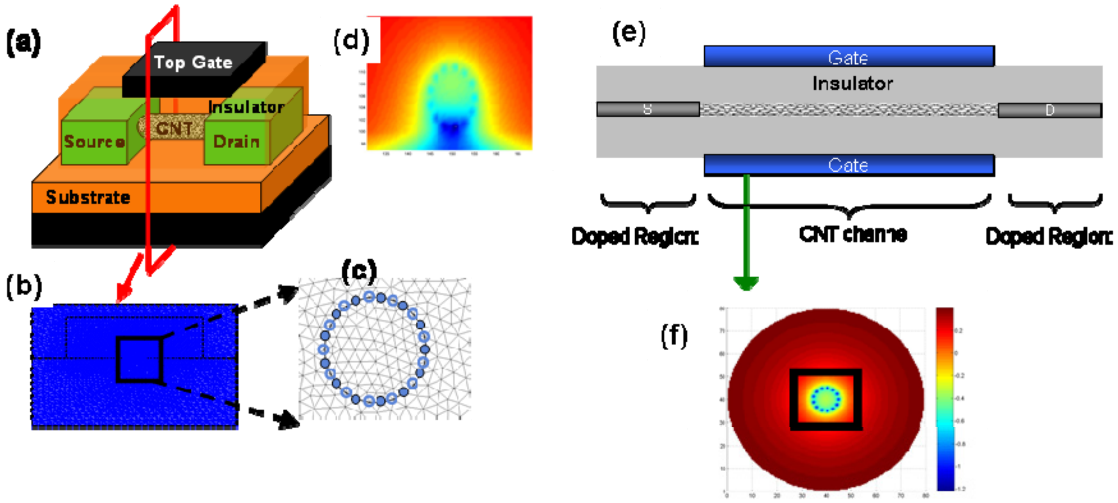


Fig. A. 2. The device structures, mesh, and the atomistic capabilities of the 3D CNT simulator. (a) The planar TG structure. BG device is realized without the top gate electrode. (b) The mesh on the cross section of the planar device (created using EASYMESH). (c) The mesh near the carbon atomic positions. The mesh is built on the atoms. (d) A solution for the potential variation around the CNT under strong asymmetric bottom and top gate biases. (e) The coaxial structure. (f) The potential distribution in a cross section of the device. The atomic positions are evident.

(FDM) has to do with the irregular geometry of the device. In the case of the cylindrical geometry, the mesh can be easily constructed with FEM rather than FDM which is more convenient for rectangular geometries. Even in the case of the planar structures, however, the merging of the cylindrical geometry of the CNT to the rectangular geometry of the surroundings of the device, is better accomplished using FEM. A 2D mesh is created, which includes information about all the regions of the 3D device (i.e. all the interfaces of the various regions are projected on a 2D plane). A cross section of the mesh and a zoom-in is shown in Fig. A.2 (b-c). The 2D mesh is then repeated in the transport direction, and each predefined region on the mesh is assigned to the properties of the specific cross section regions. For example, a 2D mesh cross section that is cut through the metallic electrodes, will be assigned to the properties of the metal contacts (properties include dielectric constant and boundary conditions, which are different for different device regions). When the same slice is cut through the middle of the channel, then it is assigned to the properties of the CNT channel and the insulators.

Since the CNT atoms are placed in regular positions along the channel of the devices in rings of carbon atoms, the 2D mesh is built by constructing mesh nodes on the positions of the CNT atoms, as shown in Fig. A.2 (c) (here we assume an unrelaxed CNT geometry). In this way a proper merging of the TB and the Poisson mesh for correct charge placement during the self consistent simulations is allowed. A zigzag *(n,0)* CNT unit cell consists of two types of carbon atom rings (type *A* an *B*), both consisting of n carbon atoms, with the position of the *B* type ring shifted a certain angle with respect to the position of the atoms of the *A* type ring as shown in Fig. A.2 and A.1. The mesh at every slice is built on the atomic positions of all 2*n* atoms of both types of rings. In Fig. A.2 (c) the full and empty dots indicate the atomic positions of the two types of rings. The 2D mesh created, is then very fine in order to accommodate for atomistic resolution in that region. In the regions near the contacts and the gate electrodes, the mesh is sparser since less accuracy is needed there. In the transport direction, the 2D grid is repeated on every place that carbon atom sites are placed. For example, a zigzag CNT consists of rings of carbon atoms placed 1.44Å (ring *A* to the next ring *B*) or 0.72Å (ring *B* to the next ring *A*). The 2D mesh is placed directly on each ring. In this way, all atomic sites of the CNT reside on the full 3D mesh that is created by considering all 2D meshes.

A.3 The real space non-equilibrium-Greens' function (NEGF) approach

The Hamiltonian of the device is described in the simple *pz,* tight binding (TB) orbital, nearest neighbor (NN) approximation by:

$$H = \sum_i \left(|t\rangle\langle t| \right), \tag{A.1}$$

with zeros for the on-site elements, *t* on the elements connecting the nearest neighbor sites and zeros elsewhere. Although there are four orbitals in the outer electron shell of carbon atoms (s, p_x, p_y, p_z), only the p_z orbital is sufficient in electron transport because the bands involving the p_z orbital are largely uncoupled from the bands involving the other orbitals. In addition to that, the bands with major contributions from the other orbitals are far away from the Fermi level, and they do not play a role in transport. The coupling parameter is given by $t = 3eV$, (default value), to give the correct bandstructure. The Green's function for the device is given by:

$$G(E) = \left[\left(E + i0^{+}\right)I - H - \Sigma_{1} - \Sigma_{2}\right], \tag{A.2}$$

where I is the identity matrix, and Σ_1 and Σ_2 are the self energies for the left and right reservoirs. In the real space p_z orbital approximation, the Hamiltonian is of total size $N \text{ x } N$, where N is the number of atoms in the channel. Zig-zag CNTs, however, as shown in Fig. A.1 are composed of rings of carbon atoms (rolled up from the graphene). In real space Hamiltonian, every ring in the CNT structure can be identified as an n x n, block in the Hamiltonian, where n is the number of atoms in the ring (same as the chirality index). The Hamiltonian can be written therefore in a block tridiagonal matrix form as:

$$H = \begin{bmatrix} \alpha_1 & \beta_2^{+} & & & & \\ \beta_2 & \alpha_2 & \beta_1 & & & \\ & \beta_1 & \alpha_3 & \beta_2 & & \\ & & \beta_2^{+} & \alpha_4 & \beta_1 & \\ & & & \beta_1 & \alpha_5 & \cdots \\ & & & & \cdots & \cdots \end{bmatrix}, \quad \text{with} \tag{A.3}$$

$$\alpha_i = \begin{bmatrix} E_{m_1} & & & \\ & E_{m_2} & & \\ & & \cdots & \\ & & & E_{m_n} \end{bmatrix}, \quad \beta_2 = t\begin{bmatrix} 1 & 1 & & & 1 \\ 1 & 1 & 1 & & \\ & 1 & 1 & \cdots & \\ & & \cdots & \cdots & 1 \\ & & & 1 & 1 \end{bmatrix}, \quad \beta_1 = t\begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \cdots & \\ & & & & 1 \end{bmatrix},$$

(A.4a-c)

where Em_i is the midgap potential energy of the i^{th} atom on the shell of the CNT.

The local density of states is given by $D_{S(D)}(E) = \left(1/2\pi\right) G \Gamma_{S(D)} G^{+}$, where $\Gamma_{S(D)} = i\left(\Sigma_{S}(D) - \Sigma_{S}(D)^{+}\right)$, is the energy level broadening due to the source/drain contacts. The charge density per unit length, $n_L(z)$, in the CNT is calculated by integrating the density of states in the channel over all energies,

$$n_L(z) = (-e) \int_{-\infty}^{\infty} dE.sgn(E).D(E).f\left[sgn(E).(E - \widetilde{E_F})\right], \tag{A.5}$$

where e is the electron charge, sgn is the sign function and $\widetilde{E_F}(z) = E_F - Em(z)$ is the Fermi level minus the mid-gap energy of the nanotube. The grounded source electrode Fermi level is set to zero and acts as the reference of the entire simulation, (E_F = 0). The nanotube middle gap energy is potential computed from the electrostatic potential at the CNT by $Em(z) = -eV(z) - Eg/2$, where Eg is the band gap of the nanotube. The current is calculated after convergence in the Landauer formalism by using

$$I = \frac{4e}{h} \int T(E)\left[f_s(E) - f_d(E)\right], \tag{A.6}$$

where $T(E) = Trace\left(\Gamma_S G \Gamma_D G^+\right)$. In the case that the source/drain are assumed to be infinite reservoirs, (MOSFET like devices), the self energies in the Green's function are computed as $\Sigma_S = \beta_1 g_1 \beta_1^+$ where g_1 is the surface Green's function, calculated recursively using:

$$g_m = \left[(E + i\eta)I - \alpha_m - \tau g_{m+1} \tau^+\right]^{-1}. \tag{A.7}$$

Here g_m is the surface Greens' function of the m^{th} ring of carbon atoms and α_m is the Hamiltonian block matrix of the m^{th} ring. The infinitesimal $\eta \to 0^+$ added to the energy ensures that the solution is the retarded Greens' function, rather than the advanced one. Applying this relation to the first two rings of the CNT, results to:

$$g_2 = \left[(E + i\eta)I - \alpha_1 - \beta_1 g_1 \beta_1^+\right]^{-1}, \; and \;\; g_3 = \left[(E + i\eta)I - \alpha_2 - \beta_2 g_2 \beta_2^+\right]^{-1} \tag{A.8}$$

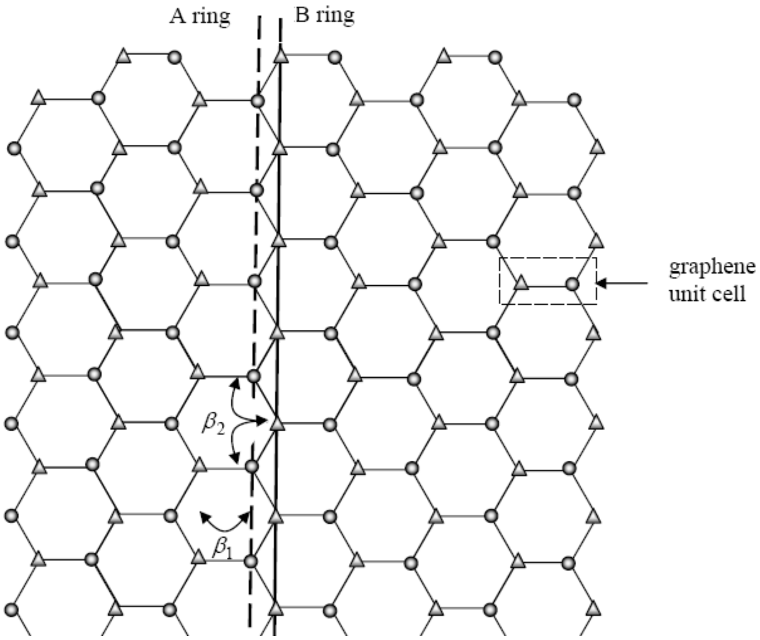


Fig. A. 3. The graphene lattice structure. When rolled up it forms the CNT device. When the "A/B ring" lines roll up, the atoms along these lines form the *zig-zag* CNT rings. The "circle" noted atoms form the 'A' ring type, whereas the "triangle" noted atoms form the 'B' ring type. The positions of the atoms of the two rings are shifted with respect to each other as shown. (Adopted from [29]).

(Note here that the self energies are n x n, complex symmetric matrices and added on the first/last block of the N x N, Hamiltonian matrix). Since the first atom ring of the source end, and the last atom ring in the drain end belong to the infinite reservoirs, the potential around the carbon atoms is the same (due to azimouthial symmetry). Due to this simple form of α_1, and by observing that all matrices that are involved in the calculation of g_1 are diagonal except β_2, (that is in the *zig-zag* CNT case only), the surface Green's function can be calculated simply by diagonalizing the β_2 by $\beta_2^D = V\beta_2 V^+$. The eigenvalues of β_2 form the β_2^D matrix, which replaces β_2 in the recursive scheme for the surface Greens' function solution. This makes the calculation of the inverse a trivial inverse of a 1 x 1 element. Assuming that α_2 is also equal to α_1, and can be written as a number times the identity, then again the g_3 is just the inverse of 1 x 1 blocks. Using the formula (A.7), the exact surface Green's function for the CNT can be computed in two iteration steps as also described in [29]:

$$g^D = \frac{(E-\alpha)^2 + (\beta_1)^2 - (\beta_2^D)^2 \pm \sqrt{(E-\alpha)^2 + (\beta_1)^2 - (\beta_2^D)^2 - 4(E-\alpha)^2(\beta_1)^2}}{2(\beta_1)^2(E-\alpha)} \qquad \text{(A.9)}$$

A back transformation will give the surface Greens function as a full complex symmetric matrix as: $g = V^+ g^D V$. (This process is equivalent in transformation from real space to mode space representations; connect all the modes, and then back-transform into the real space. It can be exact only when the azimouthial symmetry of the potential around the ring is retained (uncoupled modes), and the transformation that diagonalizes the β_1 is the same as the one that diagonlizes β_2. Otherwise, a recursive scheme will need to be implemented. The computational efficiency for reducing the calculation of the self energy to an analytical function of energy can be very large compared to algorithms that will compute the self energies in an iterative scheme (i.e. the Sancho-Rubio [122], or the direct inversion scheme of Eqn. A.7). In the case of armchair or chiral CNTs the calculation of the self energy cannot be reduced to a single analytical function of energy, therefore the computation expense of such simulations is much larger (since depending on the case, a large portion of the computation time is spent in self energy calculations). It is noted here that the unit cell of a zig-zag CNT consists of four rings, however, the later two are equivalent to the first two and are redundant in the calculation of the self energy.

A.4 Electrostatics – Poisson equation solution

The 3D Poisson equation is solved by the FEM method. And a modified version of the 3D FEM simulator NESSIE is used. The mesh is created using the commercially available mesh generator software EASYMESH [121]. Since the coupled Schrödinger-Poisson system is highly non-linear, in order to achieve convergence, the Poisson equation is transformed into a non-linear equation. The implicit scheme used is that for a given potential V^n at the step *n*, the new potential V^{n+1} is given by [123]:

$$-\nabla\left(\varepsilon_r(\vec{r})\nabla V^{n+1}(\vec{r})\right)=\frac{q}{\varepsilon_0}\left(n_D(\vec{r})-n(\vec{r})\frac{T\left[V^{n+1}\right]}{T\left[V^n\right]}\right), \tag{A.10}$$

where *T* is a function of *V*, and $n_D(r)$ is the doping charge and $n(r)$ is the mobile charge in the device entered as point charges on the mesh node that connects to the corresponding atomic position. Because of the exponential behavior of the electron density as a function of the potential *V*, a suitable choice for $T\left[V^n\right]=\exp\left(q\beta V\right)$, with $\beta=1/k_BT$. Linearization of this coupled system leads to the Gummel iterative scheme, where for a given potential V^n at the step *n*, the new potential V^{n+1} is given by:

$$-\nabla\left[\varepsilon_r(\vec{r})\nabla V^{n+1}(\vec{r})\right]+\frac{q}{\varepsilon_0}n(\vec{r})\frac{V^{n+1}}{V_{ref}}=\frac{q}{\varepsilon_0}\left[n_D(\vec{r})-n(\vec{r})\left(1-\frac{V^n}{V_{ref}}\right)\right], \tag{A.11}$$

with $V_{ref}=k_BT/q$.

This method makes the coupled system of equations to converge in most of the cases. In studies of atomistic defects in the structures, when a large number of defects is introduced, at high biases the convergence behavior might suffer, however these are cases when severe oscillations in the potential and charge are introduced in the device.

A.5 The simulation scheme

The overall simulation chart is as follows:

(a) Input definition:

1. a) Define the type *(n,0)* of CNT and generate the coordinates of the unit cell.
   b) Repeat the unit cell to create the coordinates of the entire CNT structure (*x,y,z,*

azimouthial angle theta)

c) Define the type and geometry of the structure.

d) Create the 2D mesh according to the coordinates of the CNT and the geometry of the structure. Repeat the 2D mesh to create the full 3D mesh.

2. a) Define the simulation parameters (V_G/V_D biases, electrode workfunction differences between materials used, convergence criteria, number of iterations)

(b) Simulation:

1. a) Initial guess simulation. A semiclassical ballistic model is used for calculation of the charge density self consistently with the Poisson equation (Eqn A.11).

b) Quantum simulation: Calculate charge density using Eqn. A.5 self consistently with the poisson equation Eqn. A.11 using the semiclassical solution calculated in the previous section as an initial guess. Iterate between the poisson and transport solutions till convergence as shown in the scheme of Fig. A.4

c) Upon convergence, calculate the current using Eqn. A.6.

The most time consuming part of the simulation comes from the energy integration in the calculation of the Green's function. We employ two techniques in obtaining the Green's function, the Recursive Greens' Function (RGF) [124] algorithm, and solution of N-value problem using the ZGBEV banded matrix N-value problem solver routine from LAPACK, with similar performance. The RGF scheme recursively marches block by block from the upper left (source) to the lower right (drain) of the Hamiltonian, and calculating the left connected surface greens' function by inverting the individual blocks (rings). It then recursively marches backwards towards the source, formulating the total Greens' function elements of certain blocks that are needed in the computation of the charge density and the transmission coefficient. The advantage of this scheme is that each block can be built individually and the entire Hamiltonian does not need to be stored. In terms of memory, this method is only restricted by the size of the block of the device representing the individual rings. Also, the diagonal elements of the Greens' function are calculated trivially on the backward march. This can be particularly

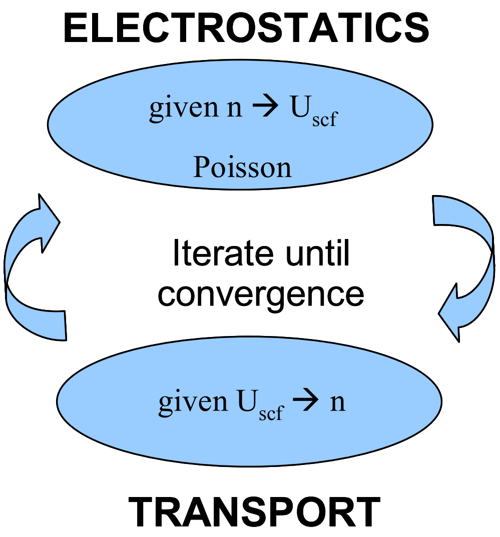


Fig. A. 4. The simulation scheme. Given an initial potential distribution (U_{scf}), the charge (n) is calculated using the NEGF equations. The charge is fed back into the Poisson solver for an updated version of the potential distribution. The process continuous till self consistent convergence is achieved.

helpful in the treatment of scattering, in which the diagonal elements are needed. When using LAPACK routines, all the Hamiltonian elements need to be stored. This which can be a disadvantage when simulating structures of larger number of atoms or a larger number of atomic basis sets. In addition to that, this method gives the columns of the Greens' function that are requested, i.e. the first/last "*n*" (number of atoms in the ring) which are adequate in the calculation of the charge density and the transmission coefficient in the ballistic case. In the case where incoherent scattering is implemented, the entire full Green's function matrix will need to be calculated by LAPACK, although only the diagonal elements are useful. This can be costly both for memory and computational time, and prohibitive for large scale simulations.

Finally, message passing interface (MPI) parallelization is implemented (through the use of the mpif90 module) in the integration procedure to calculate the charge density over the energy spectrum while the Green's function at each energy point is calculated by a serial algorithm. Various MPI inter-processor communication calls are employed in this integration phase making the scheme a truly parallel one. The resulting speed-up factor

(Fig. A.5) shows a satisfactory scaling behavior for up to 20 processors. This is also depicted in the inset of Figure A.5 where the simulation time is plotted w.r.t. the number of processors used. Please note the growing deviation from an ideal linear behavior. The reason of this declining behavior beyond 20 processors is attributed to the increased amount of time used in various MPI communication calls with large amount and size of data among the processors.

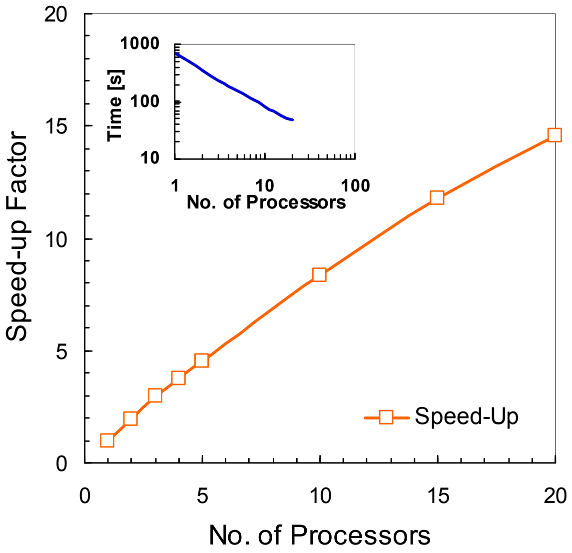
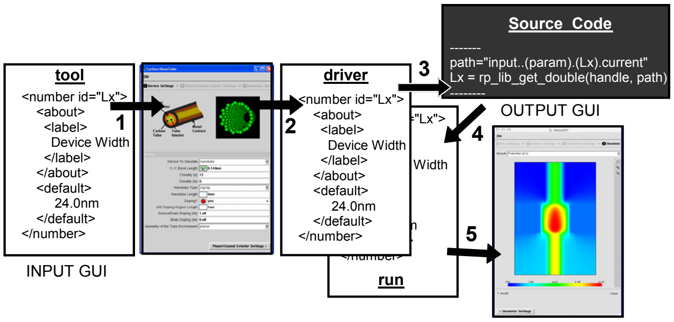


Fig. A. 5. The speed-up factor for the MPI parallelization scheme of the CNT simulator.

A.6 Deployment as a community CNTFET software

The Carbon nanotube simulation project is now a part of a wider initiative, the NSF Network for Computational Nanotechnology (NCN). The Network for Computational Nanotechnology (http://www.ncn.purdue.edu) is a multi-university, NSF-funded initiative with a mission to lead in nanotechnology research and education as well as outreach to students and professionals by offering a set of cyber services (accessible through the nanoHUB portal www.nanoHUB.org) including interactive online simulation, tutorials, seminars, and online courses packaged using e-learning standards.



```
<number id="Lx">
  <about>
    <label>
      Device Width
    </label>
  </about>
  <default>
    24.0nm
  </default>
</number>
```

```
-------
path="input..(param).(Lx).current"
Lx = rp_lib_get_double(handle, path)
--------
```

Fig. A. 6. Rappture: Revolutionizing tool development.

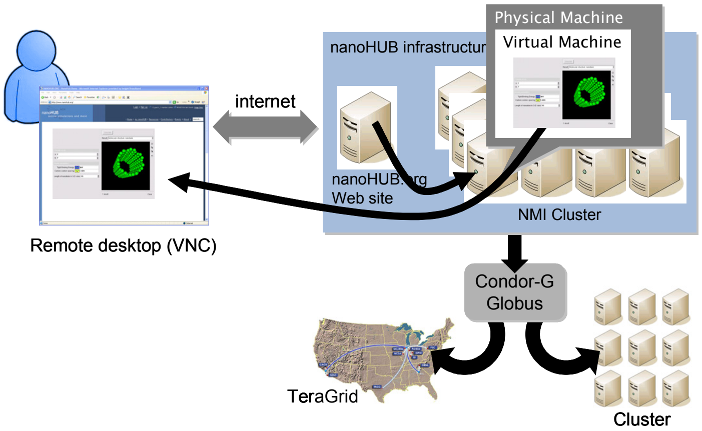


Fig. A. 7. Remote access to simulators and compute power.

In the year 2006, the educational and outreach services were accessed by over 16,200 users. More than 3,500 users performed over 94,000 online simulations. Over 30 applications are available online ranging from toy models to sophisticated simulation engines not yet available commercially. All the NCN services are freely open to the public.

NanoHUB is dedicated to offer services to real users such as experimentalists and educators, not to computational scientists alone. Therefore, a user friendly GUI is required for these tools to be operated by non-experts. The tool should be available for anybody without any installation requirements. From a deployment perspective the tool development and GUI development must be streamlined, the codes must be benchmarked, and quality tested, and adequate computational resources must be available. Rappture is the new nanoHUB.org created toolkit that enables the rapid development of GUIs for applications. Two approaches can be followed: (1) The legacy application is not modified at all and wrapper script translates Rappture I/O to the legacy code. (2) Rappture is integrated into the source code to handle all I/O (see Fig. A.6). The first step is to declare the parameters associated with one's tool by describing Rappture objects in the Extensible Markup Language (XML). Rappture reads the XML description for a tool and generates the GUI automatically. The second step is that the user interacts with the GUI, entering values, and eventually presses the *Simulate* button. At that point, Rappture substitutes the current value for each input parameter into the XML description, and launches the simulator with this XML description as the *driver* file. The third step shows that, using parser calls within the source code, the simulator gets access to these input values. Rappture has parser bindings for a variety of programming languages, including C/C++, Fortran, Python, Perl, Tcl, and in the near future, MATLAB. And finally, the simulator reads the inputs, computes the outputs, and sends the results through *run* file back to the GUI for the user to explore.

A preliminary version of the simulator within a framework called CNTFET has been recently deployed on the nanoHUB.org for community use. The process of *web-based deployment* of this and other tools is depicted in Fig. A.7. A user visits the *www.nanohub.org* site and finds a link to a tool. Clicking on that link will cause our middleware to create a virtual machine running on some available CPU. This virtual machine gives the user his/her own private file system. The middleware starts an application and exports its image over the Web to the user's browser. The application looks like an Applet running in the browser. The user can click and interact with the application in real time taking advantage of high-performance distributed computing power available on local clusters at Purdue University. Large scale calculations will soon be launched on the NSF TeraGrid or the open science grid.

APPENDIX B

ATOMISTIC MODELING OF NANOWIRE BASED STRUCTURES: GENERIC HAMILTONIAN CONSTRUCTION

B.1 Introduction

The $sp^3d^5s^*$–SO TB model formulation for calculating the electronic structure of nanostructures (3D bulk, 2D quantum wells, 1D nanowires) is described. The description below is a generic formalism for the formulation of a generic TB Hamiltonian of a semiconductor under arbitrary crystal orientations and types, bonding, lattice deformations, TB model and arbitrary basis set number, and cross sectional structural shape. It is valid for EMA as well as more complex sp^3s^* and $sp^3d^5s^*$ TB models, for cubic, zincblende and wurtzide lattices (as long as the appropriate parameterization is provided). It is restricted only by the computational burden, which can be eliminated by using fast and parallel numerical techniques. Once the Hamiltonian is constructed, obtaining the dispersion of the structure becomes a simple eigenvalue problem. The non-equilibrium Greens' function NEGF [24] formalism can be also used on the atomistic Hamiltonains, however, this is computationally expensive in terms of both memory and computation time and maybe prohibitive for realistic device sizes. The Hamiltonian construction here is for the case of nanowires, however, a generic formulation for quantum wells can be formed in a similar way.

B.2 The nanowire atomistic description for arbitrary crystal orientations

B.2.1 The unit cell of a semiconductor lattice

The starting point for the calculation is the creation of the atomistic description of the lattice and extraction of the connectivity information.

(a) Zincblende lattice

Semiconductors such as silicon, germanium and the III-V materials as InAS, GaAS, InSb are built on a zincblende lattice. The unit cell of zincblende structures consists of two atoms, the anion and the cation. The coordinates of the two atoms in the orthogonal Cartesian coordinate system are:

Anion: (0, 0, 0),

Cation: (a_0/4, a_0/4, a_0/4),

where a_0 is the length of the zinblende atomic lattice. Each atom is connected to four neighboring atoms in the perfect structure. The directions of the four bonds are:

Anion: $\vec{a}_1 = [1\ 1\ 1],\ \vec{a}_2 = [1\ \bar{1}\ \bar{1}],\ \vec{a}_3 = [\bar{1}\ 1\ \bar{1}],\ \vec{a}_4 = [\bar{1}\ \bar{1}\ 1]$,

Cation: $\vec{b}_1 = [\bar{1}\ \bar{1}\ \bar{1}],\ \vec{b}_2 = [\bar{1}\ 1\ 1],\ \vec{b}_3 = [1\ \bar{1}\ 1],\ \vec{b}_4 = [1\ 1\ \bar{1}]$.

The cation to anion bonds are in the opposite directions since the anion is connected to a cation and backwards. The entire bulk atomic structure can be constructed by repetition of the primitive unit cell in the 3D space in the direction of the basis vector set of the lattice. In the zincblende case, this basis vector set is:

$\vec{v}_1 = (1/2)[1\ 1\ 0],\ \vec{v}_2 = (1/2)[1\ 0\ 1],\ \vec{v}_3 = (1/2)[0\ 1\ 1]$

(b) Cubic lattice

Effective mass approaches in Tight Binding are mostly built on a cubic lattice (although by using the appropriate parameters the effective masses in certain directions can be

achieved). The unit cell of the cubic structure consists of two atoms (one atom in the case where the material is not a compound of two materials), the anion and the cation. The coordinates of the two atoms in the orthogonal Cartesian coordinate system are:
Anion: (0, 0, 0),
Cation: (a_0, 0, 0),
where a_0 is again the length of the cubic atomic lattice. Each atom is connected to six neighboring atoms in the perfect structure. The directions of the six bonds are:

Anion: $\vec{a}_1 = [1\ 0\ 0],\ \vec{a}_2 = [\bar{1}\ 0\ 0],\ \vec{a}_3 = [0\ 1\ 0],\ \vec{a}_4 = [0\ \bar{1}\ 0],\ \vec{a}_5 = [0\ 0\ 1],\ \vec{a}_6 = [0\ 0\ \bar{1}]$

Cation: $\vec{b}_1 = [\bar{1}\ 0\ 0],\ \vec{b}_2 = [1\ 0\ 0],\ \vec{b}_3 = [0\ \bar{1}\ 0],\ \vec{b}_4 = [0\ 1\ 0],\ \vec{b}_5 = [0\ 0\ \bar{1}],\ \vec{b}_6 = [0\ 0\ 1]$.

Again, the cation to anion bonds are in the opposite directions since the anion is connected to a cation and backwards. The entire bulk atomic structure can be constructed by repetition of the primitive unit cell in the 3D space in the direction of the basis vector set of the lattice. In the cubic case, these basis vector set are:

$\vec{v}_1 = [1\ 1\ 0],\ \vec{v}_2 = [1\ 0\ 1],\ \vec{v}_3 = [0\ 1\ 1]$.

B.2.2 Construction of the nanowire channel under arbitrary orientations

The geometry of the device can be constructed using the unit cell information which can be repeated in the whole 3D space in the directions of the basis vector set of the lattice. In order to create a certain wire structure, the approach that is used is as follows:

Step 1 – Identify the lattice/unit cell:

The lattice in which the device is build from, is indicated and the unit cell information identified (cubic or zincblende).

Step 2 – Rotate the unit cell:

The unit cell is then rotated from its original $[x_0, y_0, z_0]$ axis to the $[x_r, y_r, z_r]$ axis, for which the x_r indicates the transport direction and y_r the growth direction of the wire. For example, the wire can be oriented for growth/transport in the (100)/[110]

directions. This is achieved by a rotation transformation as of: $[x_r, y_r, z_r] = R\,[x_0, y_0, z_0]$, where $[x_r, y_r, z_r]$ denotes the new coordinates of the rotated unit cell atoms with initial coordinates in the lattice 3D space $[x_0, y_0, z_0]$. R is the rotation transformation matrix given by $R = [r_x, r_y, r_z]$, formed of the vectors of the new axis basis. The same rotation transformation is applied on the bond directions and the lattice basis vector set in order to be rotated in the new coordinate system.

Step 3 – Strain the unit cell:

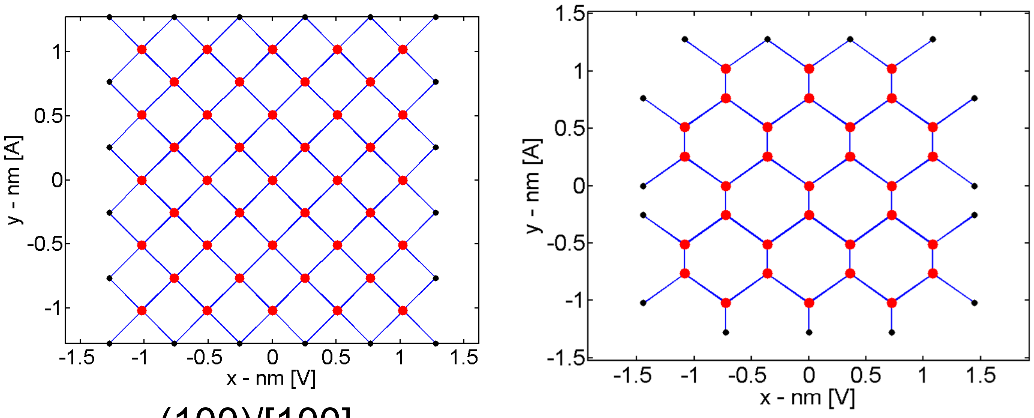
The unit cell is strained according to a specified 3D strain tensor. The strain can be uniaxial, biaxial or hydrostatic. A new unit cell is then generated with displaced atomic coordinates and bond lengths and directions. In the case of uniaxial and biaxial strain, the atomic displacement in the directions which are not strained is calculated using piezoelectric coefficients according to Poisson's ratio (volume conservation). In the case of hydrostatic strain, all *x*, *y* and *z* directions are strained similarly. The strain transformation is done as: $[x_s, y_s, z_s] = S\,[x_r, y_r, z_r]$, where $[x_s, y_s, z_s]$ denotes the new coordinates of the strained unit cell atoms resulting from the already rotated coordinates of the 3D lattice, $[x_r, y_r, z_r]$. The matrix *S,* as used in this scheme, is assumed to be diagonal with the diagonal elements been $S = \left(1 + e_{ii}\right)\delta_{ii}$, where e_{ii} is the fraction of the lattice displacement in the *ii* direction under strain (*i* can take *x,y* or *z*). In the same way as the coordinates of the atoms, the bond vectors and the lattice basis vector set are also transformed.

Step 4 – Create a large set of atoms:

Using the new information for the rotated and strained unit cell, a large "bulk" size material is created by repeating the rotated and strained unit cell in the three directions of the basis set of the new lattice structure.

Step 5 – "Cut" the device cross section:

The next step is to identify the region of the unit cell of the wire that will be used



(100)/[100]

(100)/[110]

Fig. B. 1. The cross sections of nanowires oriented in different directions. (a) The [100] oriented wire. (b) The [110] oriented wire. The red dots (larger) represent the atomic positions. Black dots (smaller) are the hydrogen atoms for *H*-passivation. The bond connectivity between the atoms is indicated.

as the channel of the device. A cross section of the device channel to be constructed is identified (i.e. a rectangular or a circular cross section). The “bulk” size structure is then “cut” in the shape of the cross section specified. The atoms that are located insight the cross section will be used as the device, whereas the atoms that are outside of the cross section are thrown away since they are not needed any more. In the transport direction, at this point only a unit cell of the wire is kept. In order to identify a unit cell, one has to choose a specific atom, and move in the transport direction until the same type of atom is found in the lattice. Once this happens, a unit cell is identified. The technique implemented here is as follows: Starting from an initial atom of certain type (anion or cation), the vector along the transport direction passing from that atom is calculated. All the perpendicular distances of all atoms on that vector are then calculated, and the ones that are zero (the atom is on the same axis as the original atom) are identified. The nearest same type atom that lies on the initial vector is then found. The distance of this atom to the initial atom indicates the length of the unit cell. All atoms that are located within the initial and final atoms, excluding the atoms that lie on the plane of the last atom, form the unit cell of the wire. Figure B.1 shows the cross-sections of two nanowires oriented in the [100] and [110] directions created using the above description.

Step 6 – Connectivity:

Once the unit cell of the wire is created, the entire wire which will form the channel of the device is formed by repetition of the unit cell in the transport direction until a wire of a specified length is formed. The information about all the atoms and their bonds can then be extracted. Particularly, all locations of the atoms and all directions of the bonds are known. The connectivity matrix can then be constructed by locating for each bond of each atom its connection. Starting from the location of each atom, and the direction and length of each bond, the "target" location for a potential connection is identified. The coordinates of the rest of the atoms are then checked to identify if there is an atom in the nearby region (for numerical issues) of that specific location. The information of this atom is then identified. Similarly, this is done for all bonds in the wire. There are two more types of bonds. If a bond of an atom reaches outside of the cross sectional area specified previously for the structure, then that particular bond is not connected anywhere, and is identified as bond type subject to passivation (in the $sp^3d^5s^*$ TB model) as shown in Fig. B.1 (black/smaller dots). If a bond is located in the cross sectional area, but in the left and right outside the transport length of the wire, then this bond will be subject for applying periodic boundary conditions on it (in the case of bandstructure calculation).

For speeding up this process of creating the connectivity matrix information, especially in the case of large structures with large number of atoms, the information is extracted on the unit cell, and then is translated to the rest of the unit cells of the device, rather than recalculated for all the unit cells. This however, is not possible in the case of non-uniform variations in the wire length. In this case the connectivity information for the entire wire needs to be computed. Figure B.2 shows an example of a 3D wire built on the atomistic lattice, indicating the bonds that connect all atoms and the bonds that are to be passivated in the simulations.

In this way the geometry of the underlying atomistic structure is created, and information about the bonding is identified. This will function as the basis for both the creation of the Laplacian for the solution of Poisson's equation and the creation of the Hamiltonian matrix for the quantum part of the problem.

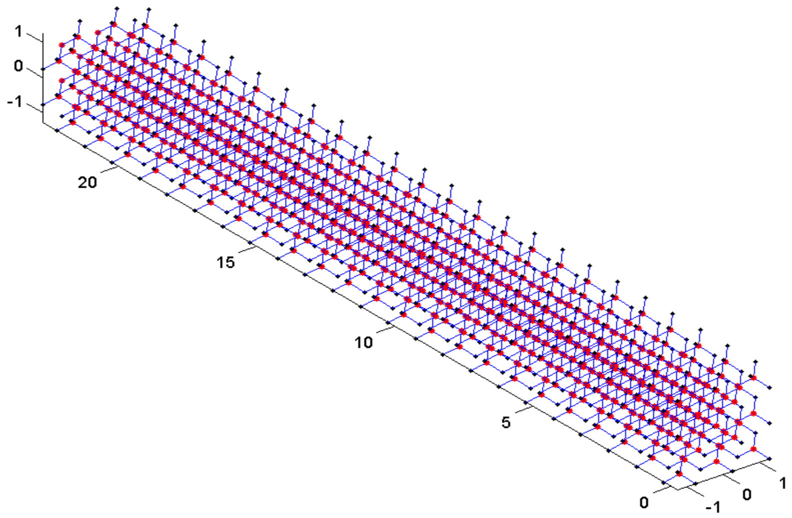
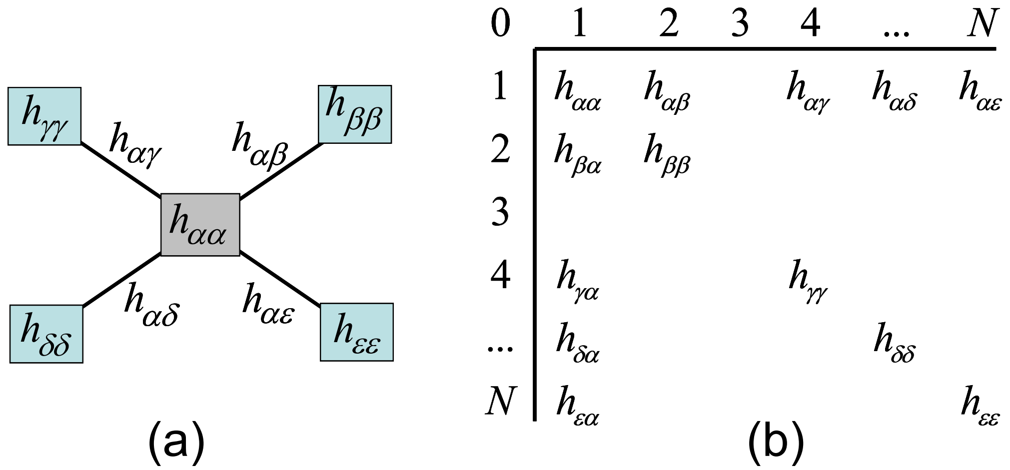


Fig. B. 2. The 3D wire built on an atomistic lattice representation. The red dots (larger) are the atomic positions. The black dots (smaller) are the hydrogen positions (for *H*-passivation). The bond connectivity is indicated.

B.3 The multi-orbital Hamiltonian structure

B.3.1 Construction of the Hamiltonian from the atomistic structure representation

The Hamiltonian of the device is constructed according to the connectivity information. The atoms are given an index according to their appearance in the device structure. The indexing of the atoms scans the rows (*x*) and columns (*y*) of each cross section of the wire, always starting from the beginning of each row. This is the way they also appear in the Hamiltonian. Each atom will occupy a diagonal block in the device Hamiltonian of size equal to the number of basis sets (number of orbitals) that are used for its' description. Once the diagonal elements have all been placed, the off-diagonal blocks of the Hamiltonian are then populated, according to the connectivity information as shown in Fig. B.3 (a,b). If a certain bond is not connected to another atom, then there is no diagonal block associated with that bond, and special treatment (either periodic BCs or passivation is later on applied). The Hamiltonian structure in 3D is block penta-



diagonal, where each cross section of the device constitutes a block-tridiagonal part of the Hamiltonian, which is then connected to the next cross section through another block, finally creating a penta-diagonal structure (as shown in Fig. B.3 (c)). The size of the entire matrix depends on the size of the underlying atomic structure. Since each atom is represented by a block of n_b basis sets, the Haniltonian is of size ($N^* n_b$ x $N^* n_b$), where N

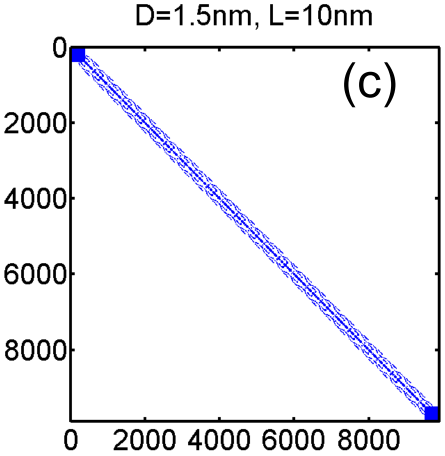


Fig. B. 3. The construction of the wire Hamiltonian from the lattice discritization. Once the connectivity is identified, the on-site and the coupling elements are built (a), the Hamiltonian can be built by placing the relevant elements in a matrix form. (c)The Hamiltonian structure for the entire wire device. The periodicity in the atomic layers is evident in the periodicity of the Hamiltonian down the main diagonal. The highly sparse matrix makes it possible for its solution through highly optimized parallel linear algorithms.

is the number of atoms in the wire. A wire of cross section (3nm x 3nm) and 10 nm length will consist of approximately 3000 atoms. In the $sp^3d^5s^*$ TB representation each atom is represented with n_b =10 (without spin orbit - SO) or n_b =20 (with SO). Therefore, the total size of the Hamiltonian matrix can easily reach 30,000 x 30,000 (w/o SO), or 60000 x 60000 (with SO). The system however, is sparse enough to allow sparse matrix solution techniques and recursive algorithms to be applied for extraction of relevant quantities for electron transport applications. In effective mass approaches, where the basis set is only n_b=1, the matrix size is much smaller and can be relatively easily handled.

B.3.2 The nearest-neighbor (NN) tight-binding (TB) $sp^3d^5s^*$-SO atomistic description

Description of the Model:

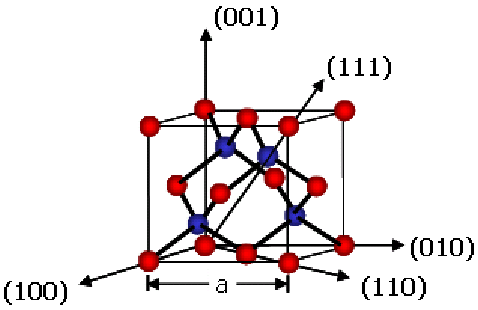


Fig. B. 4. The zincblende lattice: Consists of two FCC lattices (one for the anions, and one for the cations), misplaced by a quarter of the main diagonal in the direction of the main diagonal.

TB is an empirical model for bandstructure calculation based on local combination of atomic orbitals (LCAO).The pioneer work for application of TB in semiconductor devices was done by Slater/Koster (1954) [43]. In this model, the Hamiltonian of the device is based on local combination of atomic orbitals. Each atom in the lattice is represented by 10 orbitals ($sp^3d^5s^*$) if spin orbit coupling is not included, and 20 orbitals ($sp^3d^5s^*$-S0) if spin orbit is included. The on-site terms of the Hamiltonian and the coupling between the various orbitals between nearest-neighbors (NN) are fitting parameters. The TB parameters are calibrated using a genetic algorithm to match the bulk bandstructure over the entire BZ [40, 41].

The zincblende lattice consists of two FCC lattices misplaced by a quarter of the main diagonal along the direction of the main diagonal as shown in Fig B.4. The atoms that form the two lattices are identified as the anions and the cations. In materials such as silicon and germanium, the anions and cations both consist of the same atoms and are indistinguishable. In compound materials however, such as GaAs, InAS, InSB, etc, the two lattices are different, and the parameterization upon which the TB Hamiltonian is constructed is different. Therefore the unit cell of the TB model for the bulk consists of these two atoms (anion/cation). Each atom has four bonds (sp^3 hybridization), connecting each anion to four cations and reversely, each cation to four anions.

On-site and coupling terms in the Hamiltonian:

The first step towards the construction of the Hamiltonian blocks is the connection of an anion to the cation. The bond that connects the two atoms is described by the coupling between them. The two atom Hamiltonian can be described as shown in the equation B.1 below. (Note that the coupling described here is for one of the four bonds only):

$$H = \begin{bmatrix} H_{aa} & H_{ac} \\ H_{ca} & H_{cc} \end{bmatrix}, \text{ where } H_{aa} = \left[E_{i,j} \right] \delta_{i,j}, \text{ and } H_{ac} = \left[g^B_{i,j} \; V^{ac}_{i,j} \right], \quad \text{(B.1a-c)}$$

with *i,j* running through all the orbital indexes i.e.:

$$s, p_x, p_y, p_z, s^*, d_{xy}, d_{yz}, d_{zx}, x^2 - y^2, z^2 - r^2,$$

$E_{i,j}$ are the on-site elements describing the anion orbitals. H_{cc} is composed of the corresponding on-site elements for the cation atomic orbitals, which in the case of elemental semiconductors like Si is equal to H_{aa} (which is an n_b x n_b diagonal matrix). H_{ca} is defined as $H_{ca} = H_{ac}^{\dagger}$ such that the total Hamiltonian matrix remains hermitian. H_{ac} is the coupling between the anion and cation matrices, and is a full n_b x n_b matrix, since all anion orbitals couple to all cation ones and the reverse. The elements that form the coupling elements are formed as $V_{i,j}^{ac} = f_{i,j}(l,m,n)V_{u,v}$. (l,m,n) are the directional cosines of the lattice, defined as $(l,m,n) = \frac{(a_x, a_y, a_z)}{\sqrt{a_x^2 + a_y^2 + a_z^2}}$. In the case of the undistorted structure, $(l,m,n) = 1/\sqrt{3}$. In the case in which strain is applied, the (l,m,n) are changed accordingly. $f_{i,j}(l,m,n)$ are the two center Slater-Koster energy integrals [43, 76] that transform the orthogonal Lowdin [44] based described orbitals into the σ, π, and δ bonding. $V_{u,v}$ are the TB fitting parameters that are generated using genetic algorithm in order to correctly describe the BZ of the material of interest [40]. The *u,v* indexes, still run over the individual orbitals, however, due to their symmetries, a lot of parameters are repeatedly used, and much less than n_b x n_b elements are needed.

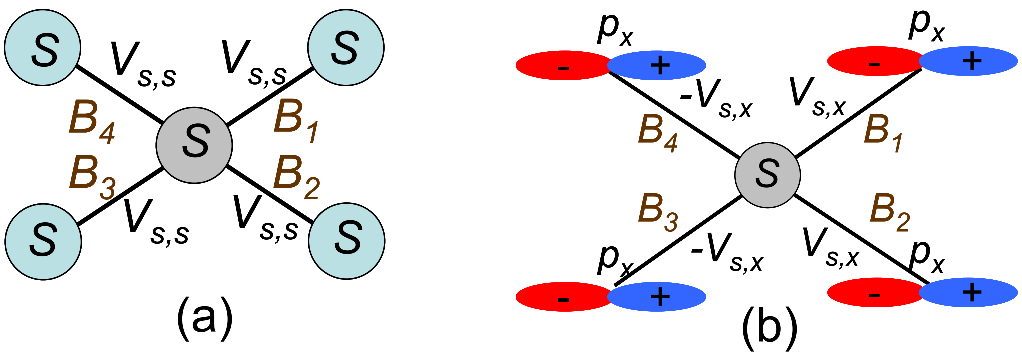


Fig. B. 5. An example of the symmetry and parity of the orbitals. (a) The coupling between *s,s* orbitals is the same for all four bonds. (b) The coupling between *s,p_x* orbitals is different for each bond.

The $g_{i,j}^{B}$, are the elements of an n_b x n_b matrix, which is a *sign* matrix, different for every one of the four bonds (*B*) between the anion and the cations. The reason is the different parity of the matrix coupling elements of orbitals residing on different atoms. This is explained in the Fig. B.5. The coupling between $V_{s,s}$ orbitals is the same for all bonds between anion and cations. In the case of the V_{s,p_x} coupling, although the magnitude is the same, the sign of the matrix element changes according to the bond considered. In order to build the *sign* matrix, the following four components are defined:

$$\begin{aligned} g_0 &= [1 \;\; 1 \;\; 1 \;\; 1] \\ g_1 &= [1 \;\; 1 \; \text{-1} \; \text{-1}] \\ g_2 &= [1 \; \text{-1} \;\; 1 \; \text{-1}] \\ g_3 &= [1 \; \text{-1} \; \text{-1} \;\; 1] \end{aligned} \tag{B.2}$$

These four components can actually describe all the bonding parity information of all orbitals. For example, g_0 represents orbitals with the same coupling parity for all four bonds as the $V_{s,s}$ shown in Fig. B.5. g_1 represents orbitals with the same coupling parity for bonds *B₁*, *B₂* as the V_{s,p_x} shown in Fig. B.5. Using these four components, a matrix G^B that includes all the symmetry information about all the bonds can be constructed as:

$$G^B = \begin{matrix} g_0 & g_1 & g_2 & g_3 & g_0 & g_3 & g_1 & g_2 & g_0 & g_0 \\ g_1 & g_0 & g_3 & g_2 & g_1 & g_2 & g_0 & g_3 & g_1 & g_1 \\ g_2 & g_3 & g_0 & g_1 & g_2 & g_1 & g_3 & g_0 & g_2 & g_2 \\ g_3 & g_1 & g_1 & g_0 & g_3 & g_0 & g_2 & g_1 & g_3 & g_3 \\ g_0 & g_2 & g_2 & g_3 & g_0 & g_3 & g_1 & g_2 & g_0 & g_0 \\ g_3 & g_1 & g_1 & g_0 & g_3 & g_0 & g_2 & g_1 & g_3 & g_3 \\ g_1 & g_0 & g_3 & g_2 & g_1 & g_2 & g_0 & g_3 & g_1 & g_1 \\ g_2 & g_3 & g_0 & g_1 & g_2 & g_1 & g_3 & g_0 & g_2 & g_2 \\ g_0 & g_1 & g_2 & g_3 & g_0 & g_3 & g_1 & g_2 & g_0 & g_0 \\ g_0 & g_1 & g_2 & g_3 & g_0 & g_3 & g_1 & g_2 & g_0 & g_0 \end{matrix} . \tag{B.3}$$

The matrix elements $g_{i,j}^{B}$ can be extracted from the matrix G^{B} by taking all the n_b rows with all first elements of the g vectors when considering bond B_1, the second elements of the g vectors for B_2, etc.

Finally, these on-site and coupling matrix elements for each bond, will serve as the basic block in the Hamiltonian construction.

The Bulk Hamiltonian:

The bulk Hamiltonian can be obtained by applying periodic BCs in the equation of the two atom Hamiltonian (eqn. B.1). The BCs are applied only on the coupling elements as phase factors as:

$$\begin{aligned} H_{ac} = \left[g_{i,j}^{B_1} \; V_{i,j}^{ac} \right] e^{\left(i\vec{k}.\vec{x}_0\right)} + \left[g_{i,j}^{B_2} \; V_{i,j}^{ac} \right] e^{\left(i\vec{k}.\vec{x}_1\right)} + \\ \left[g_{i,j}^{B_3} \; V_{i,j}^{ac} \right] e^{\left(i\vec{k}.\vec{x}_2\right)} + \left[g_{i,j}^{B_4} \; V_{i,j}^{ac} \right] e^{\left(i\vec{k}.\vec{x}_3\right)} \end{aligned} \tag{B.4}$$

Each term of the above equation represents the contribution due to each of the four bonds of the anion, connected to four cations. The dispersion can be calculated then by solving an eigenvalue problem of the Hamiltonian for the k-values of interest as:

$$E(k_x, k_y, k_z) = eig\left[H(k_x, k_y, k_z) \right] \tag{B.6}$$

Hydrogen – Passivation (sp^3 - Hybridization) [45]:

The passivation of the bonds that reside outside the domain of the device, is done using a *sp^3* hybridized scheme. The construction of the Hamiltonian, assumes the individual orbitals as the basis set. This means that each on-site element (orbital) has contributions from four bonds (couplings). In order to passivate a specific bond, a transformation to the hybridized space is performed. This means that the transformed matrix will have the bonds as the basis, with contribution on each bond from all orbitals. The on-side element of the bond that is to be passivated is then raised to a large value, in order to be placed away from the energies of interest and does not affect the bandstructure calculation. The bonds from an anion to the four cations and vice versa, are formed primarily by *sp*^{3}-hybridization as a linear combination of the only the *s* and *p* orbitals. The bonds from an anion to the cations are:

$$\begin{aligned}
\left|sp^3\right\rangle_{[111]}^{a\to c} &= \frac{1}{2}\left(\left|s\right\rangle+\left|p_x\right\rangle+\left|p_y\right\rangle+\left|p_z\right\rangle\right)\\
\left|sp^3\right\rangle_{[\bar{1}\bar{1}1]}^{a\to c} &= \frac{1}{2}\left(\left|s\right\rangle-\left|p_x\right\rangle-\left|p_y\right\rangle+\left|p_z\right\rangle\right)\\
\left|sp^3\right\rangle_{[1\bar{1}\bar{1}]}^{a\to c} &= \frac{1}{2}\left(\left|s\right\rangle+\left|p_x\right\rangle-\left|p_y\right\rangle-\left|p_z\right\rangle\right)\\
\left|sp^3\right\rangle_{[\bar{1}1\bar{1}]}^{a\to c} &= \frac{1}{2}\left(\left|s\right\rangle-\left|p_x\right\rangle+\left|p_y\right\rangle-\left|p_z\right\rangle\right)
\end{aligned}
\Rightarrow
\begin{bmatrix}
\left|sp^3\right\rangle_{[111]}^{a\to c}\\
\left|sp^3\right\rangle_{[\bar{1}\bar{1}1]}^{a\to c}\\
\left|sp^3\right\rangle_{[1\bar{1}\bar{1}]}^{a\to c}\\
\left|sp^3\right\rangle_{[\bar{1}1\bar{1}]}^{a\to c}
\end{bmatrix}
=\frac{1}{2}
\begin{bmatrix}
1 & 1 & 1 & 1\\
1 & -1 & -1 & 1\\
1 & 1 & -1 & -1\\
1 & -1 & 1 & -1
\end{bmatrix}
\begin{bmatrix}
\left|s\right\rangle\\
\left|p_x\right\rangle\\
\left|p_y\right\rangle\\
\left|p_z\right\rangle
\end{bmatrix}
=V_{sp^3}^{a\to c}$$

(B.7a)

whereas the bonds from a cation to the four anions are:

$$\begin{aligned}
\left|sp^3\right\rangle_{[\bar{1}\bar{1}\bar{1}]}^{c\to a} &= \frac{1}{2}\left(\left|s\right\rangle-\left|p_x\right\rangle-\left|p_y\right\rangle-\left|p_z\right\rangle\right)\\
\left|sp^3\right\rangle_{[11\bar{1}]}^{c\to a} &= \frac{1}{2}\left(\left|s\right\rangle+\left|p_x\right\rangle+\left|p_y\right\rangle-\left|p_z\right\rangle\right)\\
\left|sp^3\right\rangle_{[\bar{1}11]}^{c\to a} &= \frac{1}{2}\left(\left|s\right\rangle-\left|p_x\right\rangle+\left|p_y\right\rangle+\left|p_z\right\rangle\right)\\
\left|sp^3\right\rangle_{[1\bar{1}1]}^{c\to a} &= \frac{1}{2}\left(\left|s\right\rangle+\left|p_x\right\rangle-\left|p_y\right\rangle+\left|p_z\right\rangle\right)
\end{aligned}
\Rightarrow
\begin{bmatrix}
\left|sp^3\right\rangle_{[\bar{1}\bar{1}\bar{1}]}^{c\to a}\\
\left|sp^3\right\rangle_{[11\bar{1}]}^{c\to a}\\
\left|sp^3\right\rangle_{[\bar{1}11]}^{c\to a}\\
\left|sp^3\right\rangle_{[1\bar{1}1]}^{c\to a}
\end{bmatrix}
=\frac{1}{2}
\begin{bmatrix}
1 & -1 & -1 & -1\\
1 & 1 & 1 & -1\\
1 & -1 & 1 & 1\\
1 & 1 & -1 & 1
\end{bmatrix}
\begin{bmatrix}
\left|s\right\rangle\\
\left|p_x\right\rangle\\
\left|p_y\right\rangle\\
\left|p_z\right\rangle
\end{bmatrix}
=V_{sp^3}^{c\to a}$$

(B.7b)

The passivation is then achieved by a transformation as follows:

$$\left[H\right]_{Hybrid}=V_{sp^3}\left[H_{E(sp)}\right]V_{sp^3}^{\dagger}, \tag{B.8a}$$

where:

$$H_{E(sp)} = \begin{bmatrix} E_s & & & \\ & E_{px} & & \\ & & E_{py} & \\ & & & E_{pz} \end{bmatrix}$$

, is the on-site matrix consisting only of the *s* and *p* orbitals. Once the transformation takes place, the on-site elements of the hybridized space matrix of the bonds to be passivated are raised by $(h_{sp3})_{i,i}=30eV$. Finally, a back transformation into the orbital space will give the passivated matrix elements:

$$\left[H\right]_{Passiv.} = V_{sp^3}^{\dagger}\left[H_{Hybrid} + h_{sp^3}\right]V_{sp^3} \text{ , where} \qquad \text{(B.8b)}$$

$$h_{sp^3} = \begin{bmatrix} a_1 & & & \\ & a_2 & & \\ & & a_3 & \\ & & & a_4 \end{bmatrix}$$

, with a_i been 30eV or zero, depending on whether the bond *i* is passivated or not.

Spin-orbit (SO) interactions in the Hamiltonian (more details in [76]):

In the previous section, the TB Hamiltonian is built without spin orbit (SO) coupling interactions included. Here, we will describe the formalism that includes SO interactions. In this formalism, spin orbit interactions follow the following properties:

- They affect only orbitals associated with the *s*- and *p*-type orbitals. The spin orbit interaction of the *d*-type orbitals is ignored for semiconductors, because it does not have any effect on energies near the bandgap.
- They affect only orbitals with different spins sitting on the *same* atom and not in different atoms. Therefore the off-diagonal blocks, coupling the different atoms are not affected by SO. Only the diagonal blocks H_{aa}, and H_{cc} of the Hamiltonian are modified.

Without SO, the bulk unit cell Hamiltonian is given by:

$$H = \begin{bmatrix} H_{aa} & H_{ac} \\ H_{ca} & H_{cc} \end{bmatrix} \tag{B.9}$$

With SO coupling, each block H_{aa} (and H_{cc}) will be modified to:

$$H = \begin{bmatrix} H_{a\uparrow a\uparrow} & \\ & H_{a\downarrow a\downarrow} \end{bmatrix} \tag{B.10}$$

for the spin-up and spin-down orbitals. The SO interaction, affecting only the spin orbitals on the same atom, has the following form:

$$H_{SO} = \begin{bmatrix} H_{SO\uparrow\uparrow} & H_{SO\uparrow\downarrow} \\ H_{SO\downarrow\uparrow} & H_{SO\downarrow\downarrow} \end{bmatrix}. \tag{B.11}$$

This will be added to the Hamiltonian without SO coupling as:

$$H_{TOT} = H_{no-SO} + H_{SO}. \tag{B.12}$$

The various blocks H_{SO} need to be calculated. The spin-orbit Hamiltonian is given by:

$$H_{SO} = \frac{q\hbar}{4m^2c^2}\vec{\sigma}.\left(\vec{E}\ \mathrm{x}\,\vec{p}\right), \tag{B.13}$$

where E is the nuclear electric field and p is the momentum operator. σ are the Pauli matrices. Now,

$$
\begin{aligned}
H_{SO} &= \frac{q\hbar}{4m^2c^2}\vec{\sigma}.\left(\vec{E}\ \mathrm{x}\ \vec{p}\right) = \frac{q\hbar}{4m^2c^2}\left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\vec{x} + \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}\vec{y} + \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\vec{z}\right).\begin{vmatrix} \vec{x} & \vec{y} & \vec{z} \\ E_x & E_y & E_z \\ p_x & p_y & p_z \end{vmatrix} \\
&= \frac{q\hbar}{4m^2c^2}\begin{bmatrix} E_x p_y - E_y p_x & \left(E_y p_z - E_z p_y\right) - i\left(E_z p_x - E_x p_z\right) \\ \left(E_y p_z - E_z p_y\right) + i\left(E_z p_x - E_x p_z\right) & -\left(E_x p_y - E_y p_x\right) \end{bmatrix} \\
&= \begin{bmatrix} H_{SO\uparrow\uparrow} & H_{SO\uparrow\downarrow} \\ H_{SO\downarrow\uparrow} & H_{SO\downarrow\downarrow} \end{bmatrix}
\end{aligned}
$$

where $p_x = -i\frac{\partial}{\partial x}$, $p_y = -i\frac{\partial}{\partial y}$, $p_z = -i\frac{\partial}{\partial z}$. (B.14)

The interaction between the different atomic orbitals of the *spin-up* to *spin-up* orbitals can then be calculated from:

$$\left\langle p_i \middle| H_{SO} \middle| p_j \right\rangle = \left\langle \left[p_i\ 0\right] \middle| H_{SO} \middle| \left[p_j\ 0\right] \right\rangle \tag{B.15}$$

The interaction between the different atomic orbitals of the *spin-down* to *spin-down* orbitals can be calculated from:

$$\left\langle p_i \middle| H_{SO} \middle| p_j \right\rangle = \left\langle \left[0\ p_i\right] \middle| H_{SO} \middle| \left[0\ p_j\right] \right\rangle \tag{B.16}$$

The interaction between the different atomic orbitals of the *spin-down* to the *spin-up* (and vise versa) orbitals can then be calculated from:

$$\left\langle p_i \middle| H_{SO} \middle| p_j \right\rangle = \left\langle \left[0\ p_i\right] \middle| H_{SO} \middle| \left[p_j\ 0\right] \right\rangle \quad \text{(for down/up)} \tag{B.17}$$

$$\left\langle p_i \middle| H_{SO} \middle| p_j \right\rangle = \left\langle \left[p_i\ 0\right] \middle| H_{SO} \middle| \left[0\ p_j\right] \right\rangle \quad \text{(for up/down)} \tag{B.18}$$

In order to evaluate the different integrals for the SO elements, we need to take into account the properties below:

- The *E*-field and the *p* orbitals are odd spatial functions along their respective axis.

- Products of even functions are even, products of even number of odd functions are even, and products of odd number of odd functions are odd functions.
- The derivative of an even function is an odd function, and the derivative of an odd function is an even function.

For example:

$$\begin{aligned}\langle p_x \uparrow | H_{SO} | p_y \uparrow \rangle &= \left\langle \begin{bmatrix} p_x & 0 \end{bmatrix} \right| H_{SO} \left| \begin{bmatrix} p_y \\ 0 \end{bmatrix} \right\rangle \\ &= \frac{q\hbar}{4m^2c^2}\left(-i\langle p_x | \left(E_x \frac{\partial}{\partial y} - E_y \frac{\partial}{\partial x} \right) | p_y \rangle \right) \\ &= -i\delta \end{aligned} \tag{B.19}$$

After evaluating all the matrix elements, we get the following results: the H_{SO} blocks that are added to the Hamiltonian are as follows:

$$\begin{aligned} &\langle p_x \downarrow | H_{SO} | p_y \downarrow \rangle = i\delta \\ &\langle p_y \downarrow | H_{SO} | p_z \downarrow \rangle = -i\delta \\ &\langle p_x \uparrow | H_{SO} | p_z \downarrow \rangle = \delta \\ &\langle p_z \downarrow | H_{SO} | p_x \uparrow \rangle = \delta \\ &\langle p_y \uparrow | H_{SO} | p_z \downarrow \rangle = -i\delta \quad , \\ &\langle p_z \downarrow | H_{SO} | p_y \uparrow \rangle = i\delta \\ &\langle p_z \uparrow | H_{SO} | p_z \downarrow \rangle = -\delta \\ &\langle p_x \downarrow | H_{SO} | p_z \uparrow \rangle = -\delta \\ &\langle p_z \uparrow | H_{SO} | p_y \downarrow \rangle = i\delta \\ &\langle p_y \downarrow | H_{SO} | p_z \uparrow \rangle = -i\delta \end{aligned} \tag{B.20}$$

with all other matrix elements been zero.

The $H_{SO} = \begin{bmatrix} H_{SO\uparrow\uparrow} & H_{SO\uparrow\downarrow} \\ H_{SO\downarrow\uparrow} & H_{SO\downarrow\downarrow} \end{bmatrix}$, with each of the four blocks in the H_{SO} matrix consisting of the interactions between all of the $sp^3s^*d^5$ orbitals:

$$H_{SO\uparrow\uparrow}/H_{SO\downarrow\downarrow}/H_{SO\uparrow\downarrow}/H_{SO\downarrow\uparrow}=\begin{bmatrix}[h]_{sp^3s^*-sp^3s^*} & [h]_{sp^3s^*-d^5}\\ [h]_{d^5-sp^3s^*} & [h]_{d^5-d^5}\end{bmatrix}, \qquad \text{(B.21)}$$

where $[h]$ takes the values $h_{SO\uparrow\uparrow}/h_{SO\downarrow\downarrow}/h_{SO\uparrow\downarrow}/h_{SO\downarrow\uparrow}$ as follows:

$$\left[h_{SO\uparrow\uparrow}\right]_{sp^3s^*-sp^3s^*}=\begin{bmatrix}0&0&0&0&0\\0&0&-i\delta&0&0\\0&i\delta&0&0&0\\0&0&0&0&0\\0&0&0&0&0\end{bmatrix} \qquad \text{(B.22(a))}$$

$$\left[h_{SO\downarrow\downarrow}\right]_{sp^3s^*-sp^3s^*}=\begin{bmatrix}0&0&0&0&0\\0&0&i\delta&0&0\\0&-i\delta&0&0&0\\0&0&0&0&0\\0&0&0&0&0\end{bmatrix} \qquad \text{(B.22(b))}$$

$$\left[h_{SO\uparrow\downarrow}\right]_{sp^3s^*-sp^3s^*}=\begin{bmatrix}0&0&0&0&0\\0&0&0&\delta&0\\0&0&0&-i\delta&0\\0&-\delta&i\delta&0&0\\0&0&0&0&0\end{bmatrix} \qquad \text{(B.23(c))}$$

$$\left[h_{SO\downarrow\uparrow}\right]_{sp^3s^*-sp^3s^*}=\begin{bmatrix}0&0&0&0&0\\0&0&0&-\delta&0\\0&0&0&-i\delta&0\\0&\delta&i\delta&0&0\\0&0&0&0&0\end{bmatrix} \qquad \text{(B.23(d))}$$

Finally, the total Hamiltonian is modified to include the SO part as in Eqn. B.12.

The wire Hamiltonian and bandstructure calculation:

To construct the TB Hamiltonian for a nanowire, the unit cell is identified as explained earlier in the chapter. Each atom enters the Hamiltonian as an (n_b x n_b) block and connected to the other atoms through the connectivity information that directly comes from the underlying structure as shown in Fig. B.1. If a bond points in the direction out of the wire, where there supposed to be an infinite wall (SiO_2 or H) then that bond is passivated as explained above.

Once the Hamiltonian of the device is identified, the $E(k)$ relationship of the nanowire can be extracted by imposing periodic BCs on the unit cell of the wire (i.e. only in 4 atomic layers in the [100] direction). For this, the coupling of the unit cell to the two neighboring unit cells, H_{01}, is identified, and used to impose periodic BCs in the Hamiltonian. Assuming that the quantization happens in the y-z plane, translational symmetry is broken in the y, z directions, but retained in the transport x direction (k_y, k_z are therefore not longer good quantum numbers). The Hamiltonian is translation invariant only in the x direction and the $E(k_x)$ calculation is therefore as follows:

$$E(k_x) = eig\left(H_{00} + \sum_n H_{01} e^{ik_x a_0} \right) = eig\left(H_{00} + H_{01} e^{ik_x a_0} + H_{01}' e^{-ik_x a_0} \right), \qquad \text{(B.24)}$$

where only the k_x component is retained.

B.4 Summary

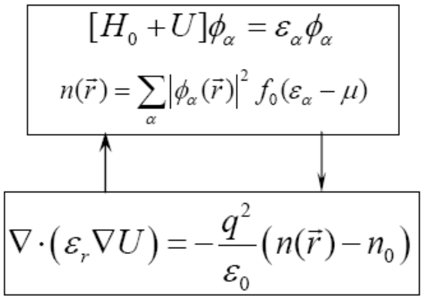
In this chapter the creation of the atomic underlying lattice and the construction of the bulk and the nanowire Hamiltonians based on this underlying lattice are presented. The construction of the sp^{3}d^{5}s* Hamiltonian, the coupling, the boundary conditions and the calculation of the bandstructure are also presented. Although the chapter elaborates more on the formation of the sp^{3}d^{5}s* TB-NN Hamiltonian, the method presented is a generic scheme that can be used for any number of orbitals (single orbital results to the effective mass approximation). The Hamiltonian for a finite wire length can be constructed by using the connectivity and BC information of all atoms. For a uniform wire, in the interest of speed and memory, this information can be calculated on the unit cell and repeated through the wire length. The BCs in this case at the most left/right sides are the infinite BCs. In the case where the *E(k)* dispersion is computed, the BCs describe the periodicity of the lattice through the *k*-value described phase factor. During the NEGF formalism they are imposed as self energies.

APPENDIX C

THE SELF-CONSISTENT BANDSTRUCTURE MODEL

C.1 Extracting the charge distribution from the wire bandstructure

Calculation of the bandstructure of the nanowire is performed by calculating the eigenvalues of the Hamiltonian with periodic boundary conditions (BCs) imposed as:



$$[H_0 + U]\phi_\alpha = \varepsilon_\alpha \phi_\alpha$$

$$n(\vec{r}) = \sum_\alpha |\phi_\alpha(\vec{r})|^2 f_0(\varepsilon_\alpha - \mu)$$

$$\nabla \cdot (\varepsilon_r \nabla U) = -\frac{q^2}{\varepsilon_0}(n(\vec{r}) - n_0)$$

(C.1)

Fig. C. 1. The self consistent (SC) loop between the transport and the Poisson equations. The charge is calculated by summing the contribution of each energy level over all bands of the k-space multiplied by its occupation probability (Fermi Function). The charge serves as an input to the Poisson equation to obtain the potential till convergence.

In the above model, n_0 is the equilibrium charge density when $U = 0V$. The subscript α stands for each eigenvalue degree of freedom, the subband index and the $k-value$, i.e. $\alpha \equiv \{m, k_x\}$ (assuming the $\vec{x}$ to be the infinite direction). The total wavefunction (assuming a planewave basis in the infinite direction can be written as:

$$\phi_\alpha = \frac{e^{ik_x \vec{\alpha}_x}}{\sqrt{L_x}} \phi_m(y,z). \tag{C.2}$$

The eigenvalues at each k point will for the bands of the nanowire, whereas the eigenfunctions are the expansion coefficients of expansion of the basis set which will give the total wavefunction. In TB a basis set is not defined. One has only access to the coefficients of expansion. Any complete basis can be used in the expansion for the actual wavefunction, and in most cases Slater type functions are utilized. The charge distribution on the atomic locations can however be obtained by using only the expansion coefficients as follows:

The charge density is therefore calculated as:

$$\begin{aligned} n(z) = \sum_\alpha |\phi_\alpha|^2 f_0\left(E_\alpha - E_f\right) &= \sum_\alpha \left| \frac{e^{ik_x \vec{\alpha}_x}}{\sqrt{L_x}} \phi_\alpha(y,z) \right|^2 f_0\left(E_\alpha - E_f\right) \\ &= \frac{1}{L_x} \sum_\alpha |\phi_\alpha(y,z)|^2 f_0\left(E_\alpha - E_f\right) \end{aligned} \tag{C.3}$$

Converting the sum over the $k-space$ into an integral, equation (C.3) becomes

$$n(z) = \left(\frac{L_x}{2\pi}\right) \frac{1}{L_x} \int_\alpha |\phi_\alpha(y,z)|^2 f_0\left(E_\alpha - E_f\right) dk_x. \tag{C.4}$$

Using the normalization properties of the wavefuntion:

$$\begin{aligned} &\int_\Omega \left| \frac{e^{ik_x \vec{\alpha}_x}}{\sqrt{L_x}} \right|^2 |\phi_\alpha(y,z)|^2 d^3\vec{r} = 1, \\ &\Rightarrow \frac{1}{L_x} \int_S |\phi_\alpha(y,z)|^2 d^2\vec{r} = 1 \end{aligned} \tag{C.5}$$

In the TB model, one has no explicit knowledge on the actual wavefunctions, but rather on the normalized coefficients of expansion of a basis which can be assumed to be any complete basis set. In the sp^{3}d^{5}s* model, the confined wavefunction $\phi_\alpha(y,z)$ is the expansion of the orbitals that represent the Hamiltonian's basis set as:

$$\phi_\alpha(y,z) = c_s\left|s\right\rangle + c_{p_x}\left|p_x\right\rangle + c_{p_y}\left|p_y\right\rangle + \ldots + c_{z^2-r^2/3}\left|d_{z^2-r^2/3}\right\rangle. \tag{C.6}$$

Assuming normalized and orthogonal orbitals (although their explicit shape is not relevant in the model),

$$\left|\phi_\alpha(y,z)\right|^2 = \left|c_s\right|^2 + \left|c_{p_x}\right|^2 + \left|c_{p_y}\right|^2 + \ldots + \left|c_{d_{z^2-r^2/3}}\right|^2 = \sum_i \left|c_i\right|^2. \tag{C.7}$$

The basis set is not defined, in this model. The charge density therefore, is calculated as point charges on the atomic locations which have contributions from all orbitals. Inserting (C.7) into (C.4):

$$n(z) = \frac{1}{2\pi}\int_\alpha \sum_i \left|c_{\alpha,i}\right|^2 f_0\left(E_\alpha - E_f\right)dk_x, \tag{C.8}$$

where the integral is over all subbands m and all $k-values$ of the first brillouin zone (BZ) ($k_x = -\pi/a_x : \pi/a_x$) and the coefficients depend on both of these quantities as well as the individual orbitals.

The above described information about the coefficients $c_{\alpha,i}$ and the subband energies $E_\alpha = E(k)$, where the $E(k)$ is the energies of the subbands at the specific $k-value$. Calculating the charge distribution in a 10 or 20 orbital model such as the sp^{3}d^{5}s*-SO model, generates a large number of data. Each eigenvalue of the bandstructure gives contributions to N sides (where N is the size of the total Hamiltonian). In a typical simulation of a small (3nm x 3nm) wire, using 10 orbitals per atom, the Hamiltonian size can reach up to *N*~5000. By using 100 *k*-points for the bandstructure calculation, and 30 bands in the simulation, the number of charge carrying elements is: 5000 x 30 x 100 = 15x10^6. In this the resolution in the individual orbital is not needed, so by summing it up, the number of elements reduce to 1.5x10^6, still a large number. Therefore, storing all these will require large memory space and is unnecessary. However, in order to calculate the self consistent bandstructure with the electrostatic potential solution, the spatial resolution of charge on each atom is needed, and in order to apply the top-of-the-barrier ballistic model, the charge distribution in k-space and bands is needed. Therefore, one stores the charge in two different ways. In the first case, the charge is saved by summing over the *k*-space and keeping the spatial variation, and in the second case by summing over the spatial variables and keeping the k-space variation. The first will be used in the Poisson loop, the second in the top-of-the-barrier model. In this way the sizes of the charge elements can be easily stored.

C.2 The 2D Poisson solver and the non-linear convergence scheme

In 2D simulations the electrostatic potential in the cross section domain of the wire is obtained using a 2D poisson solver. Solvers based on both, Finite Difference (FDM) and Finite Element (FEM) methods are implemented, with unnoticeable differences in the solutions between the two. Although FDM solvers are much easier to code, the FEM method can easily treat circular structures as well as structures with non-regular cross-sections.

In this section, the connection of the poisson solution for the electrostatic potential to the TB atomistic mesh, and the implementation of the non-linear scheme that is used for convergence are described. A brief description of the 2D poisson equation formulation in FDM is first described.

The 2D Poisson in FDM:

The 2D poisson is given by:

$$\nabla.\left(\varepsilon_r \nabla U\right) = -\frac{\rho}{\varepsilon_0} = \frac{q}{\varepsilon_0}\left(n(r) - n_0\right), \tag{C.9}$$

where ε_r is the dielectric constant of the material at position *r*, *n(r)* and n_0 are the mobile and fixed charge densities respectively. The 2D domain is discritized as shown in Fig. C.1 using a regular square mesh (in the FDM case). There are two regions treated, the nanowire channel and the oxide. The interfaces of these regions are built on the nodes of the mesh themselves. The dots in the figure represent the position of the mesh points and the position of the atoms. The region between the nodes is the elements of the domain, for which a dielectric constant is assigned to $\varepsilon_{SiO_2} = 3.9$, $\varepsilon_{Si} = 12$ in the case of Si/SiO_2 devices.

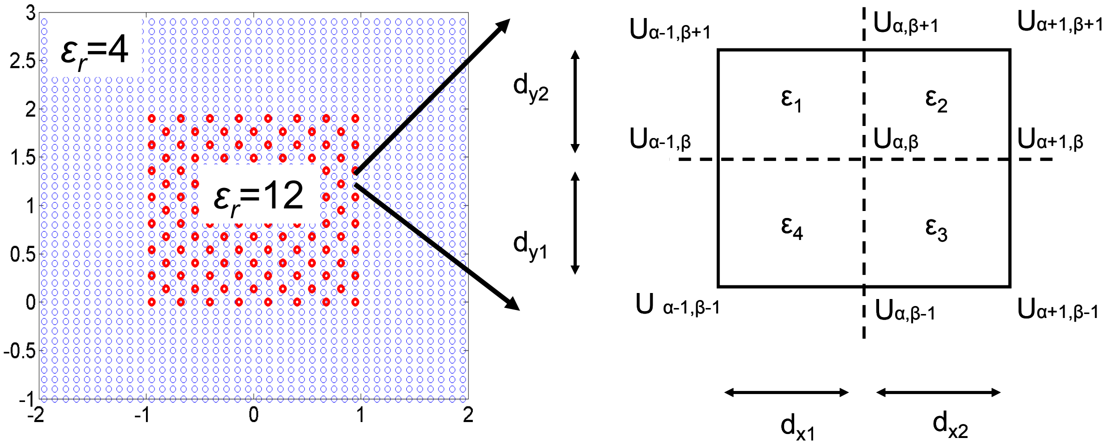


Fig. C. 2. Creating the Laplacian: (a) The 2D cross sections of the Poisson domain for a rectangular geometry. The positions of the mesh nodes and the atoms are indicated in dots. (b) The assignment of the potentials and the dielectrics in the elements of the domain.

A second order differential equation is then written for each internal node on which a solution is to be computed as:

$$\begin{aligned}
&\left(\frac{\varepsilon_1 dy_1+\varepsilon_4 dy_2}{dy_1+dy_2}\right)\left(\frac{U_{\alpha,\beta}-U_{\alpha-1,\beta}}{dx_1}\right)\left(\frac{2}{dx_1+dx_2}\right)+\\
&\left(\frac{\varepsilon_1 dx_1+\varepsilon_2 dx_2}{dx_1+dx_2}\right)\left(\frac{U_{\alpha,\beta}-U_{\alpha,\beta+1}}{dy_2}\right)\left(\frac{2}{dy_1+dy_2}\right)+\\
&\left(\frac{\varepsilon_2 dy_1+\varepsilon_3 dy_2}{dy_1+dy_2}\right)\left(\frac{U_{\alpha,\beta}-U_{\alpha+1,\beta}}{dx_2}\right)\left(\frac{2}{dx_1+dx_2}\right)+\\
&\left(\frac{\varepsilon_3 dx_1+\varepsilon_4 dx_2}{dx_1+dx_2}\right)\left(\frac{U_{\alpha,\beta}-U_{\alpha,\beta-1}}{dy_1}\right)\left(\frac{2}{dy_1+dy_2}\right)=\\
&-\frac{q}{\varepsilon_0}\left(n(r)-n_0\right).
\end{aligned} \tag{C.10}$$

Upon forming a differential equation for each node, the Laplacian can be constructed for the solution of the system. In 2D, this is a penta-diagonal matrix since each diagonal element of the Laplacian has 4 neighbors. The pre-factors of every potential component in the equation can be seen as an average dielectric in the direction of the first derivative of the potential. The second derivative is obtained by dividing by the average of the local discritization of the domain in the direction of the first derivative. (In this scheme, the mesh can be non-linear). The Laplacian, finally, enters the Poisson equation as:

$$\left[L\right]\left\{U\right\}=\frac{q^2}{\varepsilon_0}\left\{n(r)-n_0\right\}+\left\{U_B\right\}, \tag{C.11}$$

where U_B is the boundary conditions on the gate of the device. (The Laplacian involves only an equation for each unknown potential. The known part of the equation, (fixed boundary conditions), appears at the right hand side. Boundary conditions are placed on the outer nodes of the domain, and represent the gate bias.

The solution of this will give the potential distribution in the domain. The potential on the atomic positions will be used in the construction of the Hamiltonian in order to account for any variation in the electronic structure of the material due to potential variation in the cross section.

Convergence scheme:

The Poisson equation can easily be solved using Matlab “solve-\” command, which solves using the LU decomposition method. A direct solution is therefore achievable. However, due to the fact that the charge is exponentially dependent on the potential through Fermi-Dirac statistics, the resultant potential will have a small probability of driving the charge calculation into convergence. Especially in high bias cases the solution will diverge. A non-linear scheme based on the Newton-Ralphson method is therefore implemented for the convergence of the self consistent simulation between the Poisson and the transport parts. This method is based on a variable transformation, in which the charge in the device is expressed as a function of the potential, through some dummy function. It essentially builds into the Poisson the information that the charge calculated in transport, is exponential with respect to the potential. The dummy function however, needs to have some physical meaning for better behavior. In this case, Fermi-Dirac integrals are the most suitable since they include information about the exponential behavior of the density to the potential.

The charge can be expressed as:

$$n(r) = Nc\mathfrak{I}_{\alpha}(\eta_f), \tag{C.12}$$

where $\mathfrak{I}_{\alpha}(\eta_f)$ is the Fermi-Dirac integral of order α. and *Nc* is the density of the conduction band states. The Fermi Dirac integral is a function of the potential, and can be expressed as follows:

$$\mathfrak{I}_{\alpha}(\eta_F) = \frac{1}{\Gamma(\alpha+1)}\int_0^{\infty}\frac{\xi^{\alpha}d\xi}{1+e^{(\xi-\eta_F)}} \tag{C.13}$$

The dummy variable η_f can be expressed therefore as:

$$\eta_{Fn0}(r) = \mathfrak{I}_{\alpha}^{-1}\left(\frac{n(r)}{Nc}\right). \tag{C.14}$$

The quasi Fermi level can then be defined as:

$$Fn(r) = -U_0(r) + E_G + (kT/q)\Im_\alpha^{-1}\left(\frac{n(r)}{Nc}\right), \tag{C.15}$$

where $U_0(r)$ is some value of the electrostatic potential (usually the initial guess), all in units of electron volts (eV). Under this variable transformation, the charge is transformed to a quasi Fermi level value that will be serving as the input to the Poisson equation. (Notice that Fn is a quantity that logarithmically depends of the charge, and therefore it can be numerically more stable). The dummy variable η_f can be expressed as a function of Fn and the potential $U_0(r)$ as:

$$\eta_{Fn0}(r) = \frac{Fn - (-U_0(r) + E_G)}{kT/q}. \tag{C.16}$$

The total electronic charge that enters the Poisson equation can then be calculated by an inverse transformation as:

$$\rho(r) = N_{Fixed}(r) - n(r) + p(r) = N_{Fixed} - Nc\Im_\alpha(\eta_{Fn}) + Nv\Im_\alpha(\eta_{Fp}), \tag{C.17}$$

where $N_{Fixed}(r)$ is the fixed charge in the device (doping), $n(r)$ is the electron mobile charge, $p(r)$ is the hole mobile charge that enters in the charge equation through the hole parameter η_{Fp} and the Nv (density of states of the valence band), derived in the same way as for the electron case.

In this way, the Poisson equation is transformed into a non-linear equation, since the charge is a function of the potential through η_{Fn0}. $U_0(r)$, however, is an initial guess potential. Equations (C.16), (C.17), and the Poisson equation (C.11) are then solved self consistently ($U_0(r)$ is replaced by the unknown potential $U(r)$ and η_{Fn0} by η_{Fn}, while the quasi Fermi level Fn remains constant) until convergence in the potential $U(r)$ is achieved as shown in the diagram below:

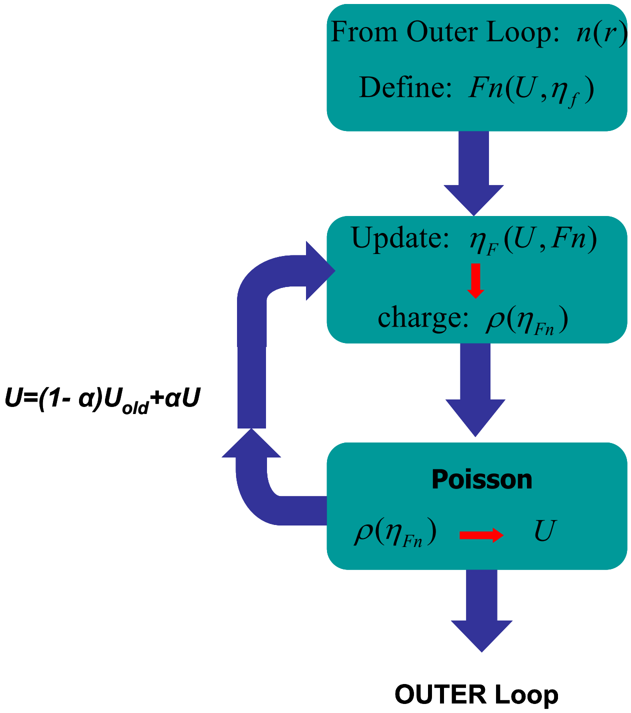


Fig. C. 3. The non-linear Poisson loop (inner loop). Using the charge density from the Schrödinger solution the Quasi Fermi Level (Fn) is determined. Based on that variable transformation, the charge density is defined as a function of the potential. The potential is determined iteratively between the Quasi Fermi Level and the Poisson equation.

In this case, the information of the exponential dependence of the charge on the potential is built in the solver. The inner loop does not solve for overall neutrality of the entire system. This can only be achieved through the outer loop iterations. The potential distribution solution that the non-linear scheme (inner loop) converges to, is in general different than the direct solution that the linear poisson equation will give. In this scheme, two different equations need to be solved, one for the potential and one for the charge. Both need to converge as it will be shown below. Upon convergence of both, then the direct solution of the linear poisson equation and the non-linear will agree. The potential that the non-linear Poisson converges to, although initially does not satisfy the transport

TB equation, it prevents it from diverging, until final convergence is achieved in the outer loop.

The choice of the dummy function (Fermi-Dirac in this case) can help in faster convergence. The closer to the physical dependence between the charge and the potential is described in the dummy function, the faster the overall simulation (outer loop) converges. Since the charge is exponentially dependent on the potential, Fermi-Dirac statistics functions are suitable dummy functions with no significant difference between Fermi Dirac functions of different orders.

The convergent criterion used in the inner loop is *dU*=1e-9. Strict criterion is essential in order to avoid any large unphysical oscillations in the charge because of its' exponential behavior on the potential.

The convergence behavior of the potential (black) and the charge (red) in the outer loop of the simulator is shown in Fig. C.3. The error and relative error in the potential decreases by an order of magnitude every two iterations. The same happens to the charge, where the relative error (*dN/N*) also shows a similar trend. The error in the charge however, stands almost two orders of magnitude higher than the error in the potential. As shown in the figure, the ratio of the relative errors between the charge and the potential is one to two orders of magnitude (red dotted). Since the charge is $N \sim e^{U/k_B T}$, then $\frac{\partial N}{N} \sim \frac{\partial U}{k_B T} \sim 40\ \partial U$. If the charge is needed to converge to accuracy of 0.1% (red-sold-circle), since this is what determines the current in the device, the potential needs to converge to an error of dU=10^{-5} (black-solid-circle). As a result the convergence criterion on the potential should be low enough in order to achieve good charge convergence.

The outer loop is much more time consuming than the inner loop in general because it involves calculation of the bandstructure of the wire, which results to an eigenvalue problem of size (*N_atoms*n_orbitals x N_atoms*n_orbitals*), which can be extremely time consuming. Therefore, the reduction of the outer loop iterations is of significant importance. Although Fermi-Dirac integrals can be used as the dummy functions in the inner loop, an analytical expression using the actual equilibrium

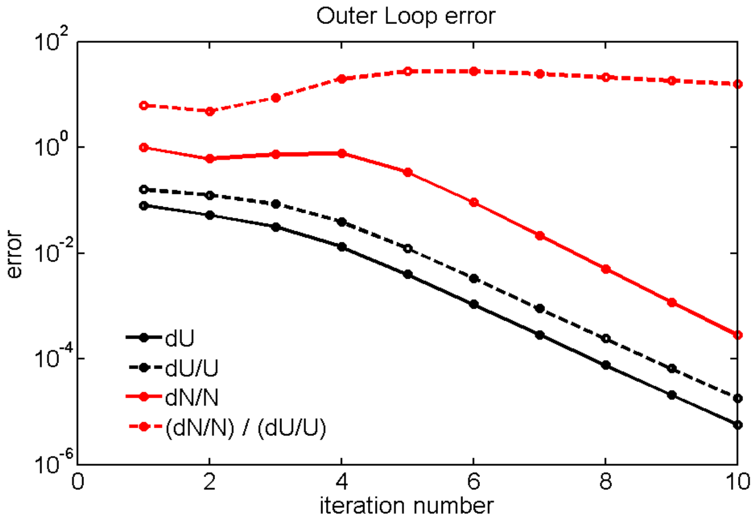


Fig. C. 4. The convergence behavior of the outer loop. Black-Solid: The error in the potential. Black-Dotted: The relative error in the potential. Red-Solid: The error in the charge. Red-Dotted: The ratio of the relative errors of the charge and the potential.

bandstructure information for the specific device, rather than just parabolic approximations for the bands that result to the Fermi-Dirac integrals can be potentially proven more useful.

VITA

VITA

Neophytos Neophytou was born in Cyprus in 1978. He received his BSc in Electrical Engineering from Purdue University, West Lafayette, IN, in 2001. He received his MSc in Electrical Engineering from Purdue University in 2003. From 2004-2008 he has been perusing his PhD degree at Purdue University in the area of Microelectronics and Nanotechnology. His research interests include theoretical modeling and simulation of nanoelectronic switching devices for future logic technology nodes.