

MODELING AND SIMULATION OF GRAPHENE NANORIBBON ELECTRONICS

By

YIJIAN OUYANG

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2011

© 2011 Yijian Ouyang

To my family

ACKNOWLEDGMENTS

First I would like to thank my advisor Dr. Jing Guo for his enormous contribution to my PhD research. Dr. Guo's great vision to the field of device physics in nanoscale regime has contributed and inspired many of my research works. He presented the way of reliably and efficiently solving research problems. Dr. Guo has been very accessible for technical help. He can almost answer all my questions or direct me where I can find answers. Besides great help in technical aspects, Dr. Guo also taught me the importance of communicating ideas with people and provided me many chances of presenting works in conferences and interacting with the people from academy and industry. The experience as a PhD student with Dr. Guo will benefit my whole life.

I am very grateful for the insightful discussions and fruitful collaborations with Dr. Xinran Wang and Dr. Hongjie Dai in Stanford University. The time with my former and current colleagues, Youngki Yoon, James Fodor, Seokmin Hong, Pei Zhao, Jyotsna Chauhan, Yang Lu, Qun Gao, Wenchao Chen and Gyungseon Seol, is pleasant. Our discussions and collaborations are very helpful to advance research progress and understand device physics. It has been very sweet to have great support from my wife Zongrui Ding, my parents, cousins, and grandparents. I would like to extend my sincere thanks to them.

Also I would like to thank Dr. Ant Ural, Dr. Huikai Xie, and Dr. Kirk Ziegler for serving on my committee. The computing resources and technical supports from University of Florida High Performance Computing are highly appreciated. Lastly, I thank our funding agencies: National Science Foundation, Semiconductor Research Corporation, and Office of Naval Research.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS.....	4
LIST OF TABLES.....	7
LIST OF FIGURES.....	8
LIST OF ABBREVIATIONS.....	11
ABSTRACT.....	13
CHAPTER	
1 INTRODUCTION	15
Carbon Nanotube and Carbon Nanotube Electronics.....	15
Graphene and Graphene Nanoribbon Electronics	18
Modeling and Simulation of Carbon-Based Materials and Devices	21
2 MULTILAYER GRAPHENE NANORIBBON AS TRANSISTOR CHANNEL MATERIAL.....	26
Monolayer and Multi-layer Graphene Nanoribbons	26
Modeling Method	27
Simulation Results	29
Discussions.....	33
3 EFFECTS OF EDGE CHEMISTRY DOPING ON GRAPHENE NANORIBBON MOBILITY	39
Edge Doping of Graphene Nanoribbons.....	39
Ab Initio Simulation and Tight Binding Parameterization	40
Simulation Results	45
Conclusions	51
4 SIMULATION OF GRAPHENE NANORIBBON TRANSISTORS WITH CARRIER SCATTERING.....	56
Scattering Mechanisms in Graphene Nanoribbons.....	56
Transport Simulations Based on NEGF	60
Extraction of Bandgaps.....	69
Fabry-Perot Simulation	72
Transport in Coulomb Blockade Regime	74

5	Scaling Behaviors of Graphene Nanoribbon FETs: A Three Dimensional Quantum Simulation Study	94
	Graphene Nanoribbons for Transistor Applications	94
	Simulation Approach.....	96
	Simulation Results	98
	Comparison to Si MOSFET	106
	APPENDIX NOTES ON FABRY-PEROT SIMULATIONS.....	115
	LIST OF REFERENCES	119
	BIOGRAPHICAL SKETCH.....	125

LIST OF TABLES

<u>Table</u>		<u>page</u>
2-1	Comparison of the (20, 0) CNTFET to 1-, 5-, and 10-layer $n=22$ GNR-FETs at the ballistic performance limits.....	35
2-2	The same comparison as Tab. 2-1 in the presence of OP scattering	35
4-1	The energy spacings from experiments and computed by the $E-k$ relation in Eq. 4-14.....	81

LIST OF FIGURES

<u>Figure</u>	<u>page</u>
2-1 The schematic structures of the modeled multilayer GNFET and the CNTFET.	35
2-2 The atomistic structure of an AB-stacking multi-layer GNR.....	36
2-3 Schematics of the top of the barrier transistor model	36
2-4 Electronic structures of multilayer GNRs	37
2-5 Effect of the gate insulator thickness on on-current for GNFETs of zero interlayer coupling and CNTFETs.....	37
2-6 Effect of interlayer coupling on on-current	38
3-1 Atomic structures of $n=15$ AGNR super-cells with a single dopant.....	52
3-2 Transmission as a function of energy	52
3-3 Configurations of multiple nitrogen dopants.	53
3-4 Resistance as a function of the channel length for $n=66$ nitrogen doped GNR with a doping concentration of 2%.....	53
3-5 Edge-doping-limited mobility as a function of the GNR width	54
3-6 Edge-doping-limited mobility as a function of the doping concentration	54
3-7 Edge-doping-limited mobility as a function of the carrier density	55
3-8 Electronic transmissions of (a) $n=15$ and (b) $n=66$ GNRs for oxygen edge doping	55
4-1 Schematics of the experiment device structures	81
4-2 Side-view of the source contact with the section plane normal to the width direction.....	82
4-3 Schematic for 2D Poisson problem	82
4-4 Energy difference between the Fermi level and the Dirac point inside the lead .	83
4-5 G_D -vs- V_g at $T=290K$ for device 1 ($L=86$ nm)	83
4-6 G_D -vs- V_g at $T=290K$ for $L=140$ nm device (device 2).....	84

4-7	G_d -vs- V_g simulated with variable SB heights (solid curves) and fixed SB (dashed curves).....	84
4-8	G_d -vs- V_g at $T=50K$ for device 1 by scaling transmission.....	85
4-9	G_d -vs- V_g at $T=50 K$ for device 1 with charge impurity and transmission scaling	85
4-10	Charge impurity configuration used to generate the curve in Fig. 4-9	86
4-11	G_d -vs- V_g at $T=50K$ for device 1 including various scattering mechanisms	86
4-12	Schematics showing transport gap measurement at cryogenic temperature	87
4-13	Schematic shows the bands bending causes uncertainty in bandgap measurement in Fig. 4-12.....	87
4-14	Experiment $\ln(G_{off}(T))$ vs $-1/T$	88
4-15	Relation between the SB height and thermal barrier height at off-state.....	88
4-16	Log scale conductance as a function of temperature in the range of 70K to 290K	89
4-17	Conductance as a function of gate voltage at zero source-drain bias for device 1 at $T=2K$	89
4-18	2D differential conductance plot for device 1 at $T=2K$	90
4-19	Experiment 2D differential conductance at $T=2K$ and small gate voltages	90
4-20	Energy level diagrams for single electron charging process.....	91
4-21	Experiment data: high resolution scan of conductance plot for a small range of biases	92
4-22	Simulated 2D differential conductance plot for Fig. 4-21.	92
4-23	Simulated 2D differential conductance plot for Fig. 4-19	93
5-1	Simulated device structures with different gate geometries.....	108
5-2	The schematic sketch of an armchair edge $N=12$ GNR channel with the source and drain contacts	108
5-3	The gate transconductance vs. channel length	109
5-4	Channel length dependence of (a) S and (b) DIBL.....	109

5-5	<i>I</i> - <i>V</i> characteristics of the single gate GNR-FET.	110
5-6	<i>I</i> - <i>V</i> characteristics of GNR-FETs with different channel lengths	110
5-7	<i>I</i> - <i>V</i> characteristics of GNR-FETs with different channel widths.....	111
5-8	Gate oxide dielectric constant dependence of (a) <i>S</i> and (b) <i>g_m</i>	111
5-9	Gate oxide thickness dependence of (a) <i>S</i> and (b) <i>g_m</i>	112
5-10	The <i>I_D</i> vs. <i>V_G</i> characteristics for single gate device with different Schottky barrier heights.....	112
5-11	Effects of contact width and height.	113
5-12	The intrinsic delay vs. channel length for a GNR-FET (the solid line) and a Si MOSFET (the dashed line).....	114
A-1	The schematics of the multi-mode conduction in GNR channel.	117
A-2	Schematics of the <i>E</i> - <i>k</i> with positive group velocity for phase calculation.	118
A-3	Conductance as a function of Fermi energy for different set of mode parameters	118

LIST OF ABBREVIATIONS

1D	1 dimensional
2D	2 dimensional
3D	3 dimensional
AP	Acoustic phonon
BSIM	Berkeley short channel IGFET model
BTE	Boltzmann transport equation
CAD	Computer aided design
CMOS	Complimentary metal-oxide-semiconductor
CNT	Carbon nanotube
DC	Direct current
DEP	Dielectrophoretic
DFT	Density functional theory
DG	Double gate
DIBL	Drain induced barrier lowering
DOS	Density of states
FEM	Finite element
FET	Field effect transistor
GNR	Graphene nanoribbon
ITRS	International technology roadmap for semiconductors
<i>I-V</i>	Current-voltage
mCNT	Metallic carbon nanotube
MEMS	Micro-electromechanical system
mfp	Mean free path
MoM	Method of moments

MOSFET	Metal semiconductor field effect transistor
NEGF	Non-equilibrium Green's function
OP	Optical phonon
RF	Radio frequency
SBFET	Schottky barrier field effect transistor
SG	Single gate
SWNT	Single walled carbon nanotube
TB	Tight binding
TEM	Transmission electron microscopy
TLE	Thermal light emission
VFET	Vertical field effect transistor
WG	Wrap-around gate
ZBP	Zone boundary phonon

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

MODELING AND SIMULATION OF GRAPHENE NANORIBBON ELECTRONICS

By

Yijian Ouyang

May 2011

Chair: Jing Guo

Major: Electrical and Computer Engineering

The dissertation presents author's simulation work on graphene nanoribbon (GNR) and GNR field effect transistors (GNRFET). The dissertation divides into 5 chapters. Chapter 1 reviews the experimental and theoretic progresses on carbon based materials and nanoscale electronics. Chapter 2 explores the performance limits of multi-layer GNRFETs. Motivated by the experiments of deriving multi-layer GNR through unzipping multi-walled carbon nanotubes (CNT), the performance limits of FETs with multi-layer GNR channels are evaluated by a semi-classic ballistic transistor model. The effects of the number of layers, interlayer coupling strength and optical phonon (OP) scattering are investigated. The simulations show that a large number of layers, zero interlayer coupling and large gate capacitance are desired to optimize device performance. Chapter 3 studies the effects of edge doping in GNRs on the carrier mobility. Specific perturbations due to nitrogen and boron dopants are calculated by ab-initio density functional theory (DFT) simulations and parameterized for tight-binding Hamiltonian. The resonant scattering plays an important role in the dependence of doping-limited mobility on the channel width, edge doping concentration, and 2D carrier density. The doping-limited mobility approximately scales linearly with width when it is

larger than 4nm. The mobility is roughly proportional to inverse of doping concentration. For small channel widths, the mobility increases as carrier density increases whereas the mobility shows non-monotonic behavior for large channel widths. Chapter 4 investigates the transport properties of GNR-FETs based on a recent experiment work. The simulations shows the experiment devices operate at 50% of ballistic limit. The important role of metal contacts on the number of transport modes, asymmetrical output characteristics, and on-state performance are studied. The performance degradation of the experiment devices is found due to charge impurities and structural defects. Effect of phonon scattering is negligible. Transport phenomena at cryogenic temperature such as Coulomb Blockade and Fabry-Perot oscillations are also explored. The simulation results agree well with the experiment measurements, indicating the high quality of GNR channels in the experiment devices. Chapter 5 explores the scaling behavior of GNR-FETs with different gate structures. The low dimensionality of GNR channels, which leads to small density-of-states, plays an important role in device scaling. Ballistic transport simulations predict the intrinsic performance of GNR can be better than Si MOSFET. However the scaling limit of GNR-FET is shorter than the one predicted for Si MOSFET.

CHAPTER 1 INTRODUCTION

The carbon based materials, carbon nanotubes, graphene and graphene nanoribbons, have received extensive attentions for their importance in fundamental physics and potential applications in many areas such as electronics, mechanics and medicine [1-4]. A monolayer graphene is composed of a single layer of carbon atoms in honeycomb structure. A single layer GNR can be obtained by pattern graphene into strips. And a single walled CNT (SWNT) can be conceptually viewed by rolling up a graphene sheet seamlessly. A particular way of patterning GNRs or rolling up CNT is called chirality. Both CNTs and GNRs are carbon based quasi-1D materials. They share some similarities. For a complete survey of carbon based electronics and a better understanding of the technologies evolve, the introduction on carbon nanotube and carbon nanotube electronics is presented below.

Carbon Nanotube and Carbon Nanotube Electronics

CNT was first experimentally observed by using transmission electron microscopy (TEM) in 1991 although people have long speculated its existence in the products of carbon fiber synthesis [1]. Because of its diameter can be made in the order of nanometer, it is a nearly ideal example of quasi-1D system. Depending on the chirality, it can be metallic or semiconducting [5]. Predicted by electronic bandstructure calculation in p_z orbital tight binding basis set, One third of SWNTs are metallic and the other two thirds are semiconducting [6]. The prediction is observed experimentally [7]. CNT can be multi-walled as well [1]. They are almost metallic which can be useful for interconnect applications. Semiconducting CNT can be used for transistor channel due to its excellent electrical properties. First it has high mobility. The bandstructure limited

velocity (group velocity of Bloch electrons) is high even at low energy range. The mean free path of acoustic phonon scattering (AP) is in micrometer range [8], [9]. Near ballistic transport can be achieved in low bias range. The energy of optical phonon is large ~ 0.18 eV, which makes OP scattering effective at relatively high bias regime [10].

Second the conduction bandstructures and valence bandstructures are almost symmetrical (In simplest TB, they are perfectly symmetrical), which makes them good for CMOS applications [6]. In 1998 the first CNT field effect transistor (CNTFET) was demonstrated [11], [12]. Since then significant experimental progress has been achieved in refining CNTFET [13], [14]. The channel length (tube length) is shortened from tens of micrometers to tens of nanometers. To reduce contact resistance, people studied various types of metal-CNT contacts and fabrication process. The earliest and easy-to-fabricate devices are the Schottky barrier FET (SBFET). Then MOSFET like devices which have heavily doped source and drain were developed, in which contact resistance is significantly lowered compared to SBFET and ambipolar I - V characteristic is greatly suppressed. The gating technology has also been improved. While simple experimental devices typically have a bottom gate with an oxide thickness of several hundred nanometers, the best solid state gating technology can make a top gate of several nanometer high- κ materials [15]. Although CNT has no dangling bonds and thus supposedly to work well in contact with high- κ material, in reality adverse effects are indeed observed and people developed DNA shield to protect carbon channel and wrap-around gating technology which further increases the gate electrostatic control [16]. The current delivered per tube in short channel devices, which can be as large as 60 μ A, well exceeds the saturation current ~ 20 μ A limited by OP (scattering) [13]. The on-off

current ratio and subthreshold slope are decent if not better than the state of the art Si MOSFET. In fact the on-current density per active device width (tube diameter) is much larger than any commercial transistors in use. The metallic CNTs can be used for interconnect applications due to its high current delivery capacity per tube [17].

However the major drawbacks of CNT are also obvious. First it has a low dimensionality. Although the current per tube is high, the CNTFETs and CNT interconnect can only compete with Si MOSFET and copper lines if they can be packed dense enough. Second the possible formation of metallic SWNT can totally destroy the functionality of a CNTFET. Third the fabrication process is not compatible with Si MOSFET because the synthesis of CNT requires a temperature well in excess of 500 C which will damage the Si transistors and interconnects. Experimentalists have tried hard to tackle those drawbacks. Significant progress has been achieved in a relatively short period although the match to Si MOSFETs, which has evolved for about half century, is not there yet. Almost perfectly aligned CNT array has been fabricated [18]. People can eliminate the metallic tubes by pumping large amount of current through the tubes or using plasma etching which selectively eliminates metallic tubes in a certain diameter range [18], [19]. To overcome the difference in temperature between the CNT and CMOS fabrication process, the SWNTs are fabricated separately, introduced to the CMOS fabrication process and assembled by electric field manipulation method such as dielectrophoretic (DEP) [20]. Alternatively people fabricate micro-thermal plates at specific locations of a chip [21]. Those locations are heated over 500 C to grow SWNT while other parts of the chip are kept cool and damage to Si devices is avoided.

Transistor applications require good control of CNT position and orientation. It is much less demanding in process to fabricate a CNT thin film, which is essentially a random CNT network. Such device has several applications. First due to its large ratio of surface area to volume, it is advantageous for absorbing molecules and potentially can be used for sensor application. Second the CNT thin film, which has many openings, is much more transparent to light and electrical field than a metal thin film while a decent electrical conductivity is preserved. This property makes it very suitable for electrodes in vertical FET (VFET) in low cost large area display application, in which the gate electrode plane is below the source electrode plane made of CNT thin film and the electrical field has to penetrate the source plane to modulate the channel. CNT also has potential applications in microelectrical-mechanic systems (MEMS) due to its excellent mechanic properties.

Graphene and Graphene Nanoribbon Electronics

Monolayer graphene was successfully isolated in 2005 and experimentally characterized to have large mobility $\sim 1e5 \text{ cm}^2/\text{V}\cdot\text{s}$. The monolayer or few-layer graphene of the best transport quality is obtained by mechanically exfoliation of graphite. Much progress has been achieved to exfoliate a large area of graphene and transfer it to wafers for further processing. On the other hand, synthesis of graphene on silicon-carbide substrate is also developed. But the mobility of the graphene grown on substrate is inferior to exfoliated graphene. Graphene is important in fundamental physics due to its linear $E-k$ relation around charge neutral level which can be used to study Dirac fermions and Berry phase [3]. Its high conductivity and mobility are also attractive for electronics applications. Similar to CNT thin film, it can be used for electrodes in light emitting device due to its high optical transparency. Its high mobility

makes it promising for RF and transistor applications. However monolayer graphene is a semimetal which lacks of bandgap. In order to create a bandgap in this 2D material, potential difference between A and B sublattices is needed. This can be achieved by placing graphene layer on the surface of a crystal that has approximately the same lattice constant and structures but with intrinsic potential difference between two sublattices. Experimentally a bandgap of as large as 200 meV is observed. But such intimate contacting between the graphene film and the substrate may degrade the mobility. Alternatively an electrical gate can be used to produce potential difference between sublattices but this cannot work with monolayer graphene. Instead a bilayer graphene with electrical field normal to the plane can have a bandgap. There is self-consistency relationship between the charge density on the two layers and the induced bandgap. Therefore this bandgap is not constant as the device works in different regimes. Also a sufficient bandgap for practical applications is very demanding on the gating technology to induce large enough charge density

Due to the difficulty in generating a bandgap in 2D graphene, people turn to the quasi-1 dimension material, GNR. As the graphene is planar material, CMOS fabrication process can be used to pattern graphene into GNR. As limited by the thickness of photoresist in lithography, the ribbon width cannot be smaller than 15 nm otherwise the ribbon will be broken into pieces. The theoretic calculation shows a sub-10 nm channel width is required to open a large enough bandgap in room temperature. The early fabricated GNR only shows a factor of about 10 in gate modulated channel conductance at room temperature and about several hundred in cryogenic experiments [22]. But even at extremely low temperature, a large source-drain bias will destroy the on-off

current or conductance ratio. Later people use a wet process, which involves expansion and sonication of graphite, to obtain sub-10 nm wide GNR channels with ultra-smooth edges. The Raman spectroscopy confirmed the ribbon geometry. The FET built on the nanometer wide GNR shows an on-off ratio up to $1e5$ under a large source-drain bias. And the on-current is as large as 7 μA per ribbon under a 10 nm SiO_2 bottom gate. The on-current is expected to be even larger when a high- κ gate is used. The phonon scattering in perfect GNR (all hydrogen terminated edges and regular shape) is similar to CNT. The AP scattering mfp is expected to be in micrometer range and OP scattering happened at relatively high gate voltage. But unlike CNT, a GNR has edges. The atomistic structure irregularity at edges can scatter electrons and degrade current. While ohmic contact can be consistently made for metal and CNT, the experimentally characterized contact resistance for GNR is much larger which indicates the existence of strong elastic scattering. Although GNR suffers from edge scattering, the chemically reactive edges can be utilized to functionalize the channel. An important application is channel doping which is crucial for complementary logic applications. Experimentally n-type GNR-FET has been demonstrated by using ammonia plasma etching of the channel [23]. Another advantage of GNR compared to CNT is the all semiconducting channels for sub-10 nm width. However the wet process for fabricating narrow GNRs is even less compatible with CMOS fabrication compared to the fabrication of CNT. Recently a new method of obtaining sub-10 nm wide GNR has been developed [24]. This method uses plasma to etch CNT and “unrolls” the tube into GNR. In this process multi-layer GNR can be obtained by etching multi-walled CNT and the number of layers is larger than the number of layers obtained from wet process, which is usually 1 or 2

layers. This natural stacking of GNR channels can help to remedy the low on-current per single-layer GNR.

Modeling and Simulation of Carbon-Based Materials and Devices

The significant progress in experiments is accompanied with substantial achievements in theoretic work, which help people to understand the device physics in the carbon based materials, evaluate and optimize device performance and predict the characteristics of the devices that are not experimentally available. Most of those works should be categorized into applied physics or science since the fundamentals of solid state physics, quantum mechanics and semiconductor physics have been there long time ago but now reappear under some new scenarios such as exotic materials and reduced dimensionality. Nonetheless the nature of applying fundamental physics does not degrade the importance of these theoretic works at all. On the contrary they bring the science to the experimental works otherwise the experiments are no difference than secrete cooking recipes. There are indeed a few breakthroughs in fundamental physics that are of great value to condense matter community and device physicists. For example the new type of hybrid functional HSE used in density functional theory simulations, which solves the bandgap underestimation problem in DFT simulations in many cases if not to all [25]. The simulation works on carrier transport falls into three large categories, material property simulations, device simulations and circuit simulations.

Among material property simulations the lowest level is based on ab-initio quantum chemistry methods. In this simulation all material properties such as relaxed atomistic structure, electronic structure, phonon structures, mass density, elastic modulus and so forth are all computed from the basic physical constants. But the

computational cost is extremely high. Among the ab initio methods, DFT is probably the computationally cheapest but very inaccurate method. However in many cases DFT simulations can give sufficient qualitative and quantitative descriptions. The matrix elements and full electronic bandstructures can be used to calculate carrier scattering rate thus the mobility through Fermi's golden rule or Kubo formalism. For example people use this strategy to study the mobility degradation in graphene when some carbon atoms are substituted by dopants. The bandstructures can be parameterized by simpler basis set such as empirical tight binding and effective mass for efficient computations. The p_z tight binding prescription of CNT, GNR and graphene is obtained from ab initio simulation. The scattering rate/mobility calculation can be calculated in simple basis as well. Another source of scattering is lattice vibration, of which the harmonic component is called phonon. By deriving the Hessian matrix (second derivative of the total energy with regard to displacement) one can obtain the force constant matrix. Then the phonon modes and normal modes amplitude can be calculated using Newton's law. With deformation potential (first derivative of total energy to displacement) one can obtain the perturbation potential or matrix element. The optical spectrum of a material is a scattering rate calculation in essence in which the quasi-particles emitted and absorbed during the electron transitions are the quantities of interest.

As a device, the geometrical size along transport direction is finite despite that on other directions the size may be treated as infinitely large. This is major difference from material property simulation although some materials like small molecules have a finite size. But isolated molecules do not have electronic transport properties. The device

boundaries are treated in simulation explicitly or implicitly. Equations of motions for device simulation basically have three types, classic, semi-classic and quantum mechanic. The classic equations of motion are Newton's laws without considering quantum states, which is rarely used in device simulation. The semiclassical equation of motion is the Boltzmann transport equation (BTE), which take the quantum states into consideration. Therefore the bandstructure effect is included and bandstructure is usually unchanged in simulation. A distribution function of particles at a given position, momentum and time, a point in phase space, is solved. The phase space is limited by all available quantum/microscopic states. For fermions, the Pauli exclusion principle is already included, for example, the use of Fermi distribution function. A particularly simple but useful model is the top-of-barrier nano-transistor model which is suitable for evaluation of performance limits [26]. What BTE is not included is the wave characteristics of particles and thus is unable to solve tunneling and interference problems. In conventional devices, the phase relaxation length is usually much smaller than device dimensions and thus electrons are treated as classic particles. The drift-diffusion theory in semiconductor physics can be viewed as derived from BTE with a lot assumptions and simplifications. When the device feature size is comparable to the phase relaxation length or tunneling is important, quantum equation of motion is required. WKB approximation is widely used to solve tunneling problem through potential barriers of irregular shapes. Care must be taken for the phase evolution of wave functions near tunneling points. One important drawback is that it assumes a well defined E-k relation in device just as BTE does. However this is not true when the device lacks of translational symmetry in transport direction, for example, at non-

equilibrium condition. The non-equilibrium Green's function formalism is able to solve the barrier tunneling problem and transport through device with translational symmetry breaking. But NEGF is not necessary when the transport is coherent since there is well defined scattering state across the channel. A numerical solution of Schrodinger equation can give the same results. The powerfulness of NEGF formalism lies in the ability to treat incoherent transport in which there is phase relaxation and sometimes energy relaxation as well. Phenomenological scattering process such as Buttiker probes and microscopic scattering process can both treated in NEGF. The elegance of NEGF is the ability of focusing on device of interest and incorporating the effects of outer sources such as contacts and phonon bath through a quantity called self-energy. The size of problem is the same as the device rather than the whole system which is practically unsolvable. The major shortage of NEGF compared to drift diffusion model is the computational cost. In NEGF the Green's function (operator) is the quantity to solve which is a matrix. Usually only the tri-diagonal elements of this matrix are needed. Despite of the efficient recursive algorithm to solve the matrix elements, which is equivalent to Gaussian elimination, the computation cost can still be very expensive depending on the size of the principal layers (PL) and size of Hamiltonian basis in use. Therefore one needs to carefully examine the problem and then decides whether a full quantum transport simulation is necessary or a semiclassical transport simulation will suffice. In the device simulations by technology CAD, quantum transport is not largely used as additional corrections for inversion charges at surface channel, tunneling through SB barriers at contacts, band-to-band tunneling and tunneling into oxide or floating gate as in memory applications.

The circuit simulation is the simulation of a system or network composed of many interconnected devices. To make the computational cost affordable, the model for each device must be further simplified from the device simulation, which is called compact modeling. The soundness of results and simplicity in arithmetic are more important than being rigorous in physics. Therefore a lot of phenomenological parameters are introduced such as in BSIM. The soundness of results is checked by comparisons with device simulation and experiments. The compact modeling of CNT circuitry has been developed despite CNTFET is still in the experimental phase [27].

CHAPTER 2
MULTILAYER GRAPHENE NANORIBBON AS TRANSISTOR CHANNEL MATERIAL

Monolayer and Multi-layer Graphene Nanoribbons

Graphene [2], [3], [28] is a monolayer of carbon atoms packed into a honeycomb lattice, which possesses atomically thin body and an area scale orders of magnitude greater, making it an ideal two-dimensional (2D) system. A 2D graphene is a semimetal without a bandgap, but a bandgap opens if a field-effect transistor (FET) channel is built on a nanometer-wide graphene nanoribbon (GNR) due to the width direction confinement [29], which leads to subband formation as well [30]. The high mobility (up to 200,000 cm²/Vs for suspended graphene at low temperature and 10,000 cm²/Vs for substrate-supported graphene at room temperature) and carrier velocity ($\sim 10^8$ cm/s) demonstrated in 2D graphene have stimulated strong interests on graphene electronics [2]. The transport properties of GNRs in experiments have been so far hindered by imperfect edges [31], [32], but excellent transport properties have been theoretically predicted for structurally perfect GNRs [33]. The GNERFET, however, suffers from the problem of a low on-current due to its nanometer-wide channel, and it has been shown previously that the ballistic performance limits of a monolayer GNERFET are not better than a CNTFET in terms of the on-current and on-off current ratio [34].

In this study we assess the performance limits of GNERFETs and CNTFETs using a well established ballistic transistor model, which has been applied to various two-dimensional and one-dimensional channel transistors before [26]. The performance limits of a transistor are achieved when the contacts are ideal and the channel is ballistic (no scattering). Schottky barriers can play an important role in CNTFETs [35] and GNERFETs. The Schottky barrier is known to lower the on-current and increase the

off-current due to ambipolar I - V characteristics. These detrimental effects in the Schottky barrier CNTFETs, however, can be eliminated by using a metal-oxide-semiconductor (MOS) FET device structure [36], which has heavily doped source and drain extensions as shown in Fig. 2-1a. The performance limits are therefore assessed for a ballistic MOSFET structure with semi-infinite source and drain extensions. In this condition, the simple semiclassical model agrees with detailed quantum mechanical transistor simulations for a channel length down to about 10 nm [26], as long as the direct source-drain tunneling current is relatively small compared to the total source-drain current.

Multilayer GNR were demonstrated in recent experiments [24], [37]. We show that the modeled ballistic performance limits of a multilayer GNFET can be significantly better than its corresponding CNTFET in terms of the on-current and on-off current ratio. The advantage becomes even bigger in the presence of optical phonon scattering, which is known to be strong in graphene-derived nanostructures. The important role of gating technology and interlayer coupling are investigated for achieving the performance advantage of multilayer GNFETs over CNTFETs. We found that weak interlayer coupling, which is accessible through incommensurate non-AB stacking, is desired for high performance multilayer GNFET.

Modeling Method

The modeled graphene nanoribbon and carbon nanotube have similar bandgaps for a fair comparison. The (20, 0) zigzag CNT is semiconducting, which results in a diameter of $d_{\text{CNT}} \approx 1.6$ nm and a bandgap of $E_g \approx 0.50$ eV. The $n=22$ armchair edge GNR (where n denotes the number of carbon dimer lines [38]) is also semiconducting,

which results in a width of $W_{\text{GNR}} \approx 2.7$ nm and a bandgap of $E_g \approx 0.52$ eV for a monolayer GNR. Both transistors have doped source and drain extensions, as schematically shown in Fig. 2-1a. For the GNR-FET, a wrapped-around gate is used as shown in Fig. 2-1b. For the CNT-FET, a coaxial gate is used as shown in Fig. 2-1c. The nominal gate insulator is a 3-nm-thick ZrO_2 , which has a relative dielectric constant of $\kappa \approx 25$. A power supply voltage of $V_{DD} = 0.5$ V is used.

The bandstructures of the CNT and GNR channels are required as an input to the transistor model. A nearest-neighbor p_z tight binding approach was used. The hopping integral is taken as $t_{cc} = -2.7$ eV. The edge effect plays an important role in the GNR, and a factor of 1.12 is used for the hopping parameter between two edge carbon atoms to count the edge relaxation effect [39]. For the AB-stacking structure of the multilayer graphene as shown in Fig. 2-2, an interlayer coupling of 0.3 eV is set only for vertically aligned two atoms in two neighboring layers [40]. The nominal device has an interlayer coupling of zero, which is the lower limit of interlayer coupling. Weak interlayer coupling could be achieved by incommensurate non-AB-stacking structures, and it is most preferable for better device performance as discussed below. The nominal value is also varied to examine the effect of interlayer coupling strength on the device performance.

A “top-of-barrier” transistor model is used to assess the performance limits of the multilayer GNR-FETs. The model fills the $+k$ states at the top of the barrier with the source Fermi level E_{FS} and the $-k$ states with the drain Fermi level E_{FD} as shown in Fig. 2-3a. Self-consistent electrostatics is treated by a simple capacitance model, in which the gate insulator capacitance value is computed by a numerical Poisson solver and source- (drain-) channel capacitance is ignored for simplicity. The optical phonon

scattering in carbon nanotubes [10] or graphene nanoribbons has a short mean free path of ~ 10 nm, which could play an important role even for transistors with a sub-100 nm channel length. We also consider the transistor performance in the presence of the OP scattering by setting an effective Fermi level at $E'_F = E_{FS} - \hbar\omega_{OP}$, which determines the population of the $-k$ states in the presence of OP scattering, as shown in Fig. 2-3b.

$\hbar\omega_{OP} \approx 0.18$ eV is the optical phonon energy in CNTs and GNRs

Simulation Results

Because the transistor performance strongly depends on the bandstructure of the channel material, we first examine the bandstructures of the multilayer GNRs. The upper left panel of Fig. 2-4a plots the bandstructure of the monolayer $n=22$ AGNR, which has a width of $W_{GNR} \approx 2.7$ nm and a band gap of $E_g \approx 516$ meV. Compared to a CNT in which a periodic boundary condition applies in the circumferential direction of the CNT and a valley degeneracy factor of 2 exists for each band, the monolayer AGNR does not have valley degeneracy due to a different quantization boundary condition in the width direction of the GNR. At the limit of zero interlayer coupling, an m -layer GNR has the same bandstructure as the monolayer, but each bands becomes a factor of m degenerated due to m uncoupled layers. The interlayer coupling lifts the subband degeneracy. Figure 2-4a also shows the bandstructure s of 2-, 5- and 10- layer multilayer GNRs with an AB stacking structure. As the number of the layer increases, the subband spacing reduces, and the upper subbands should become more accessible for carrier transport. Furthermore, as the layer number increases from 1 to 10, the bandgap decreases from 516 meV to 197 meV. We also examined the dependence of the bandgap on the interlayer coupling strength for the multilayer GNRs as shown in

Fig. 2-4b. As the interlayer coupling strength increases from the zero limit to the value for the AB stacking structure, the bandgap monotonically decreases regardless of the number of the GNR layers. The decrease is more significant for a multilayer GNR with a larger number of layers.

Next, we examine the effect of various gating on transistor performance by varying the SiO₂ gate oxide thickness t_{ox} from 0.23 nm to 20 nm, as shown in Fig. 2-5. A common off-current of 10 nA is specified for all transistors. The lower limit of the simulated SiO₂ thickness is an equivalent case of a 3-nm-thick ZrO₂ gate for the modeled gating structure as shown in Fig. 2-1, which results in the same gate insulator capacitance. The on-current increases only slightly by 12% as the number of the layers increases from 1 to 10 at $t_{ox}=20$ nm as shown in Fig. 2-5a. The transistor on-current is determined by the product of the charge density and the average carrier velocity at the top of the channel potential barrier. The relatively small improvement is largely due to a small increase in carrier density as the number of layers increases. The gate to channel capacitance C_g is a serial combination of insulator capacitance C_{ins} and the quantum capacitance C_q , and it is limited by C_{ins} since it is much smaller than C_q . A larger number of layers results in a slightly larger C_{ins} and thereby slightly larger C_g due to a thicker channel. For thick oxide, the advantage of using a multilayer GNR channel in terms of the on-current is insignificant.

In contrast, the advantage of the multilayer GNR channel in terms of the on-current becomes more significant as t_{ox} decreases, as shown in Fig. 2-5a. Table 2-1 compares the performance of various channels with a high- κ 3 nm ZrO₂. For a common off-current of 10 nA, the monolayer GNR channel has a slightly larger gate insulator

capacitance, however a 13% smaller on-current than the CNT channel, because of the lack of valley degeneracy in the monolayer GNR. As the number of the layer increases, the on-current surpasses that of the CNT. For a 5-layer GNR, the on-current is 97% larger than that of the CNT, and for a 10-layer GNR, the on-current is 180% larger than that of the CNT. For a similar gating technology, the ballistic on-current of the multilayer GNR can be significantly larger than the CNT with a similar bandgap.

In order to understand the factors that contribute to the increase of the on-current, we also computed the gate insulator capacitance and the average carrier velocity as shown in Tab. 2-1. The average carrier velocity decreases from 4.28×10^7 cm/s to 3.50×10^7 cm/s as the number of GNR layers m increases from 1 to 10 because more carriers populate closer to the bottom of the m -fold degenerate lowest subbands where the band-structure-limited velocity is low. The on-current, however, increases by a factor of 3.2 due to the increase of the gate capacitance C_g by a factor of 3.9. It is also interesting to notice that the increase of the gate capacitance significantly outpaces the increase of the gate insulator capacitance stemming from a thicker GNR body as the number of the GNR layer increases, because of the proportional increase of the quantum capacitance as a function of the number of the layers. Furthermore, for the modeled thin high- κ gate insulator, the quantum capacitance and the gate insulator capacitance becomes comparable, and both of them play a role in determining the gate capacitance.

The improvement of on-current is even more significant in the presence of OP scattering. As shown in Tab. 2-2, the on-current increases even more significantly as the number of layers increases in the presence of OP scattering. The on-current of the

10-layer GNR channel is 260% larger than that of the CNT. In addition, the average carrier velocity is also larger, which is different from the case of the ballistic channel as shown in Tab. 2-1. For a CNTFET, the OP scattering results in a saturation current close to $(4e^2/h) \times (\hbar\omega_{OP}) \approx 28\mu A$, where the OP energy $\hbar\omega_{OP} \approx 180$ meV when only the 1st subband conducts the current, which is the case for the CNTFET because $E_{FS} - E_1 \approx 216$ meV is smaller than the spacing between the 1st and the 2nd subbands $E_2 - E_1 \approx 226$ meV. The OP scattering results in considerably increased population of the $-k$ states, which also lowers the average carrier velocity at the top of the barrier from 4.61×10^7 cm/s to 2.72×10^7 cm/s. In contrast, the effect of OP scattering on the 10-layer GNRFET is small, because the value of $E_{FS} - E_1 \approx 120$ meV is smaller than the OP energy. In a multilayer GNR, more subbands are responsible for delivering the on-current, and therefore, the $E_{FS} - E_1$ value is smaller. As a result, the OP scattering only has a small effect on the on-current and average carrier velocity of the 10-layer GNRFET. Different from the case of the ballistic channel, the average carrier velocity of the 10-layer GNRFET is also about 20% larger than that of the CNTFET in the presence of OP scattering, which promises faster intrinsic transistor speed.

Finally, we examine the dependence of the on-current on the interlayer coupling. As shown in Fig. 2-6, the largest on-current is achieved at the zero interlayer coupling and the on-current decreases as the interlayer coupling increases. The ballistic on-current monotonically decreases from 51 μA to 36 μA for 2 layers and from 77 μA to 58 μA for 5 layers as the interlayer coupling increases to the value of AB stacking. The reason is that increase of the interlayer coupling increases the spacing between the subbands, which makes the higher subbands more difficult to be accessed for delivering

the on-current. In addition, the increase of interlayer coupling also decreases the bandgap, as shown in Fig. 2-4. As the bandgap decreases, the band-to-band tunneling can be turned on, which can significantly increase the off-current. (Modeling the band-to-band tunneling current is beyond the capability of the semiclassical model used here.) Non-AB-stacking multilayer structures have been recently observed in experiments for both GNRs unzipped from CNTs [24] and CVD-grown multilayer graphene structures. Weakening the interlayer coupling by non-AB-stacking structures, such as the randomly stacking structures, should be pursued for boosting the performance of the multilayer GNR FETs.

Discussions

The performance of the multilayer GNR MOSFETs can also be compared to that of the silicon MOSFETs. Such comparison is clouded by the different dimensionality of the channel material, and we simply discuss how dense a GNR array channel needs to be to meet the performance goal at the end of the ITRS roadmap [41]. The ITRS roadmap calls for an on-current of about $2700 \mu\text{A}/\mu\text{m}$ and an off-current of $0.60 \mu\text{A}/\mu\text{m}$ at a power supply voltage of 0.65 V for the technology nodes near year 2020. To reach the on-current of $2700 \mu\text{A}/\mu\text{m}$, a 10-layer array GNR FET with 26 GNRs per μm is needed for the on-current as shown in Tab. 2-2. The array has an off-current of $0.26 \mu\text{A}/\mu\text{m}$, which is less than a half of the ITRS goal, although the power supply used in the GNR FET simulation is only 0.5 V . A recent experimental study characterized the mobility degradation of monolayer graphene on a SiO_2 substrate [42]. Another experimental study showed that the multilayer graphene transistor is more immune to the noise compared to the monolayer graphene transistor [43], whose one-atom-thick body is susceptible to the charge impurities and oxide traps near the transistor channel.

Therefore the multilayer GNR-FET is expected to have a better immunity to the adverse effects from substrates than the monolayer GNR-FET.

The bandstructure of a multilayer GNR is assumed to be gate-voltage-independent in this study. Large electric field and a significant potential drop between the layers, however, can alter the bandstructure of a multilayer GNR. For the simulated gate-all-around structure and the low applied voltage below 0.5V, the electric field between the layers and its resulting potential drop between the layers is small (which is less than 30 meV as computed by a separate self-consistent atomistic simulation as described in Ref. [44], [45]). Its effect on the transistor I - V characteristics is negligible for the modeled gating structure and power supply voltage.

In summary, we have shown the important role of developing good gating technology and weakening the interlayer coupling for improving the performance of multilayer GNR-FETs. The thin high- κ gate insulator is already in production for Intel 45nm transistor, and its application to GNR-FETs remains to be developed. Weakening the interlayer coupling could be experimentally achieved in non-AB-stacking multilayer graphene. The performance limit of a well-designed multilayer GNR-FET can significantly outperform its CNT counterpart with a similar gate and bandgap in terms of the on-current, at either the ballistic limit or in the presence of OP scattering.

Table 2-1. Comparison of the (20, 0) CNTFET to 1-, 5-, and 10-layer $n=22$ GNRFETs at the ballistic performance limits.

	I_{OFF} (nA)	C_{ins} (F/m)	I_{ON} (uA)	$E_{FS}-E_1$ (eV)	I_{ON}/I_{OFF}	$\langle v(0) \rangle$ cm/s
CNT	10	8.42e-10	39.9	0.233	3900	4.28e7
1-layer GNR	10	8.98e-10	34.1	0.279	3400	4.61e7
5-layer GNR	10	1.27e-9	77.3	0.155	7700	3.69e7
10-layer GNR	10	1.73e-9	109.6	0.120	11000	3.50e7

The transistor structures are shown in Fig. 2-1. The gate insulator thickness is 3 nm and the dielectric constant is 25 (for ZrO_2). The off-current (I_{OFF}), gate insulator capacitance (C_{ins}), on-current (I_{ON}), the spacing between the source Fermi level and the top of the 1st subband barrier ($E_{FS}-E_1$), the on-off current ratio (I_{ON}/I_{OFF}), and the average carrier velocity at the top of the barrier ($\langle v(0) \rangle$) are compared. For the modeled 1-layer GNR, the spacing between the 1st and 2nd subband is $E_2 - E_1 = 0.121$ eV. For the modeled CNT, $E_2 - E_1 = 0.226$ eV.

Table 2-2. The same comparison as Tab.2-1 in the presence of OP scattering

With OP	I_{off} (nA)	C_{ins} (F/m)	I_{ON} (uA)	$E_{FS}-E_1$ (eV)	I_{ON}/I_{OFF}	$\langle v(0) \rangle$ cm/s
CNT	10	8.42e-10	29.3	0.216	2900	2.72e7
1-layer GNR	10	8.98e-10	24.4	0.260	2400	2.69e7
5-layer GNR	10	1.27e-9	71.0	0.151	7100	3.30e7
10-layer GNR	10	1.73e-9	105.4	0.118	10500	3.33e7

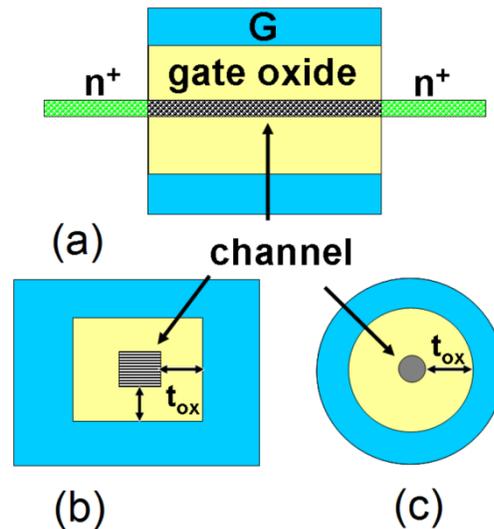


Figure 2-1. The schematic structures of the modeled multilayer GNRFET and the CNTFET. (a) The MOSFET has an intrinsic channel and heavily doped source and drain extensions. (b) The cross section of the wrapped-around gate GNRFET. (c) The cross section of the CNTFET with the coaxial gate.

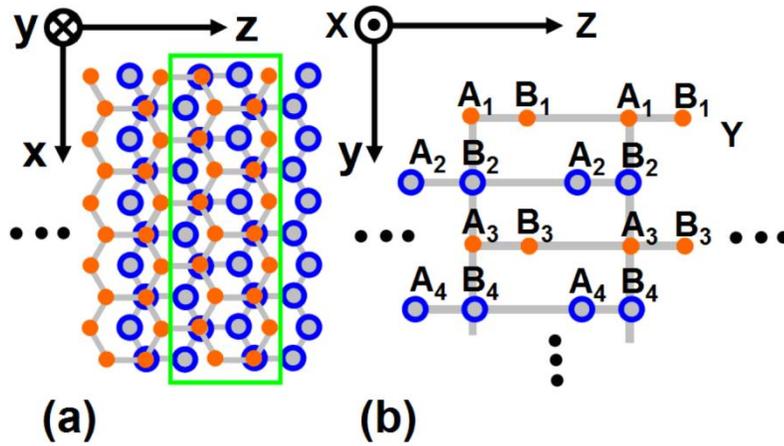


Figure 2-2. The atomistic structure of an AB-stacking multi-layer GNR. (a) Top view of GNR. The z direction is defined along the GNR transport direction, x direction along the width direction, and y direction along the thickness direction. (b) Side view of GNR. ‘A’ or ‘B’ denotes the atoms in A or B sublattice, respectively. And the subscripts denote the layer indices in a multilayer GNR. The modeled multilayer GNR has armchair edges.

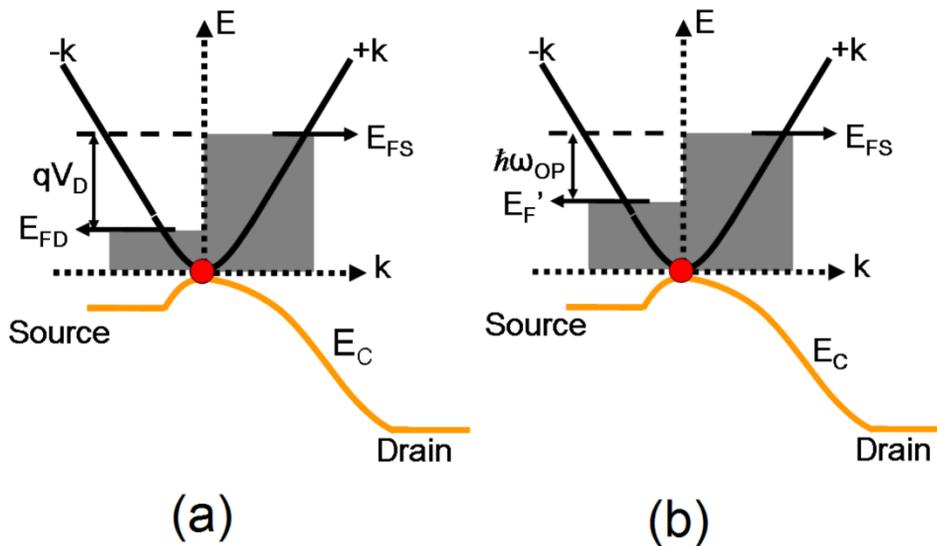


Figure 2-3. Schematics of the top of the barrier transistor model. (a) Ballistic transport. The $+k$ states are filled according to the source Fermi level E_{FS} whereas the $-k$ states are filled according to the drain Fermi level E_{FD} . (b) Transport in the presence of OP scattering. The $-k$ states are filled according to an effective Fermi level $E'_F = E_{FS} - \hbar\omega_{OP}$.

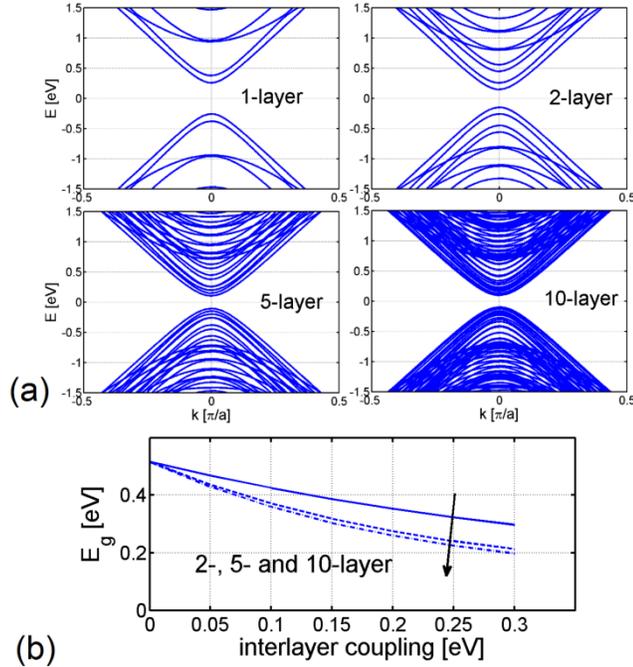


Figure 2-4. Electronic structures of multilayer GNRs. (a) The bandstructures of 1-, 2-, 5- and 10-layer $n=22$ multi-layer AGNRs with an interlayer coupling of 0.30 eV. The bandgap is 0.52 eV for 1-layer, 0.30 eV for 2-layer, 0.21 eV for 5-layer and 0.20 eV for 10-layer. (b) The bandgap as a function of interlayer coupling for 2-layer and 5-layer and 10-layer $n=22$ AGNRs.

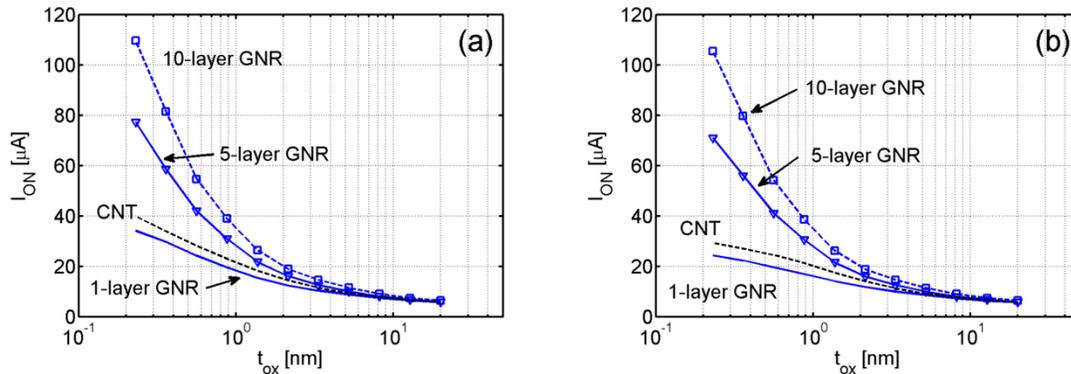


Figure 2-5. Effect of the gate insulator thickness on on-current for GNRFETs of zero interlayer coupling and CNTFETs. (a) The on-current as a function of the gate oxide thickness at the ballistic limit. (b) In presence of OP scattering. The multilayer $n=22$ AGNRs of zero interlayer coupling have a similar bandgap as the simulated (20, 0) for a fair comparison. The insulator dielectric constant is $\kappa_{SiO_2} \approx 3.9$. The smallest value of the oxide thickness, $t_{ox} \approx 0.23$ nm is equivalent to a 3-nm thick ZrO_2 with a dielectric constant of $\kappa \approx 25$, which results in the same gate insulator capacitance.

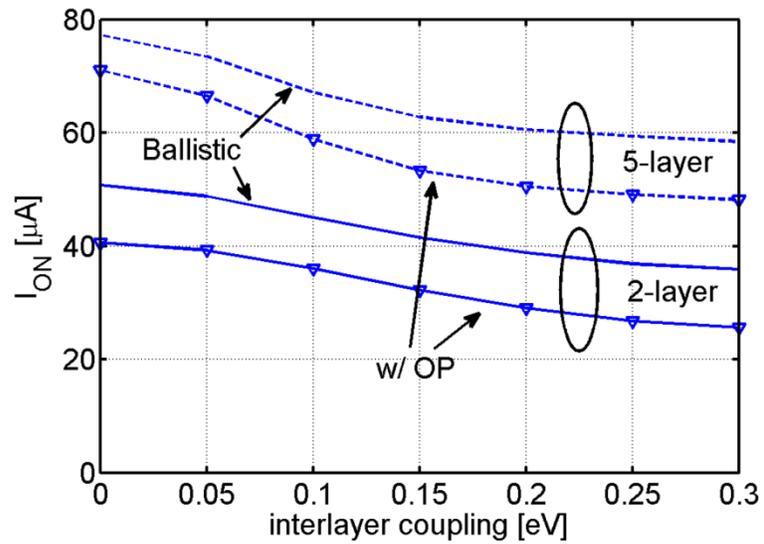


Figure 2-6. Effect of interlayer coupling on on-current. The on-current as a function of the interlayer coupling is computed for the 5-layer GNR-FET (dashed lines) and the 2-layer GNR-FET (solid lines). A 3-nm-thick ZrO_2 gate insulator is used. The lines without symbols are computed for ballistic transport and the lines with symbols are for transport with OP scattering.

CHAPTER 3
EFFECTS OF EDGE CHEMISTRY DOPING ON GRAPHENE NANORIBBON
MOBILITY

Edge Doping of Graphene Nanoribbons

The strong interest in graphene electronics was driven by its high mobility for potential electronics applications [2], [3], [28]. Graphene nanoribbons (GNR) are derived by patterning graphene into narrow strips. A significant bandgap can be opened due to the quantum confinement along channel width [29], [31]. The chemically active edges can be engineered for various functionalities. A recent work has shown that nitrogen and oxygen atoms can be used to dope GNR for n-type and p-type transistor channels through substitution of the edge atoms [23]. The impact of edge dopants on the electronic transport such as carrier mobility, however, is not yet clear. In this work we exclusively studied on the edge-doping-limited mobility which is not covered by previous studies on GNR mobility [46], [47].

The chemistry of different dopants is captured by using ab initio density functional theory simulations. For n-type doping, the substitution of nitrogen atoms for the edge carbon atoms is considered. For p-type doping, the passivation of edge carbon atoms by oxygen atoms is considered. The quantum transport equation is solved in non-equilibrium Green's function (NEGF) formalism with Hamiltonians in tight binding (TB) bases parameterized from ab-initio Hamiltonian and overlap matrices [48]. The dependences of the edge-doping-limited mobility on the dopant species, doping concentration, GNR channel width, and carrier density are investigated. For narrow GNRs (1-10 nm), the doping-limited mobility is found to be approximately proportional to the channel width, and inversely proportional to the edge doping concentration. The mobility is a non-monotonic function of 2D carrier density. The study shows that the

quasi-1D channel and resonant scattering are important in determining the doping-limited GNR mobility.

Ab Initio Simulation and Tight Binding Parameterization

The specific perturbation caused by one dopant (one nitrogen or one oxygen atom) at an edge to the electronic structure and transport properties of a perfect GNR is simulated by using the ab-initio transport program SMEAGOL [49], which combines density functional theory (DFT) method and NEGF method. The doped channel structures which are essentially super-cells have 5 unit cells with one dopant at the edge of the middle unit cell as shown in Fig. 3-1. The channel super-cells and a GNR unit cell are relaxed separately by usual SIESTA DFT simulations [50]. Due to the short range of the dopant perturbation (one can conceptually think of a Gaussian profile), the Hamiltonians of the two unit cells at the two ends of a super cell approximately recover to that of a perfect GNR unit cell. A double zeta polarized basis set is used. The energy cut-off is 200 Ryd. Local density approximation is used. For armchair GNR a spin restricted simulation is sufficient. For zigzag GNR whose edge states can cause spin polarization [51], a spin unrestricted simulation is needed. The convergence criterion for structure relaxation is 0.04 eV/Ångs. Based on the relaxed channel and GNR unit cell atomic structures, SMEAGOL self-consistently calculates the transmission through the channel (with a dopant) connected to two semi-infinite leads using NEGF method. At the end of a SMEAGOL simulation, one can obtain the electronic transmission which is used for resistance calculation later and the Hamiltonian/overlap matrices of the channel and leads in the representation of the pseudo atomic orbital (PAO) basis, i.e. DZP basis. In DZP basis set for each hydrogen atom, there are 5 orbitals and for each carbon, nitrogen, or oxygen atom, there are 13 orbitals.

To calculate the doping-limited mobility, the ensemble averaged resistance of a GNR with a given channel length, channel width and doping concentration needs to be calculated (more details on the calculation are given later). The large number of simulation runs for ensemble averaged quantities and the large GNR channel sizes to simulate make the *ab initio* simulations impractical. Therefore a tight binding parameterization is needed to reduce the size of the original Hamiltonian and overlap matrices (a large size non-orthogonal TB basis) used in ab-initio simulations. An orthogonal p_z orbital TB prescription (one orbital per carbon atom, and we call it the orthogonal p_z TB in this section) has been successfully applied in simulating CNT and GNR electronic structures at low energy [5], [39]. For GNR a correction to Hamiltonian elements because of edge bond relaxation is necessary to obtain a good agreement between the band structures calculated by the orthogonal p_z TB and ab-initio simulations [39]. After including the dopant, the bandstructure is neither a useful nor a meaningful criterion to test whether a TB parameterization is good or not due to the translational symmetry breaking in a doped channel (not a periodic super-cell structure). Therefore the transmission through a doped channel connected to two semi-infinite GNR leads can be used as the criterion of judging TB parameterization in presence of the dopants [52]. It has been shown that an on-site perturbation (the perturbation to the diagonal elements of a Hamiltonian) can be extracted for the orthogonal p_z TB basis in the previous study of doping of CNT [52]. We found that this method works well for nitrogen edge doping in GNR as shown in Fig. 3-2a. However for oxygen edge doping, the orthogonal p_z TB with on-site perturbations does not produce transmissions close enough to those calculated by SMEAGOL, possibly due to the fact that an oxygen atom

passivates an edge carbon atom (substitutes for a hydrogen atom) rather than substitutes for an edge carbon atom. Instead we found that the 1st zeta p_z orbital of the DZP basis set used in DFT simulations is able to produce transmissions reasonably close to those by SMEAGOL simulations as shown in Fig. 3-2b. The Hamiltonian and overlap matrices in the 1st zeta p_z orbital basis are essentially a subset of the original ab-initio basis set. Thus this p_z basis is non-orthogonal and has coupling between atoms beyond first nearest neighbors. We keep the non-orthogonality and limit the interaction up to the third nearest neighbors. Perturbations to both on-site elements and hopping elements are included. Despite the complexity in the TB parameterization for oxygen doping, the size of the non-orthogonal basis, one orbital per atom (hydrogen atoms are not in the Hamiltonian anymore), is the same as the orthogonal p_z TB for nitrogen doping. One thing to note is that oxygen atoms are included in the parameterized TB basis.

With the TB bases of a drastically reduced size, we can compute the transmissions of the channel structures of a size far beyond the capacity of ab-initio simulations. However cautions must be taken when the configurations of multiple dopants are randomly generated. The TB bases are derived from ab-initio simulations of a single dopant. When there are multiple dopants, it is assumed that each dopant causes a perturbation as the single dopant in ab-initio simulations and the total perturbation is a superimposition of individual perturbations. Therefore the channel Hamiltonian can only be reasonably approximated by performing superimposition of individual perturbations if the dopants are not too close to each other and a linear superimposition of perturbations can hold. In this study any two of the dopants are not

allowed at the same edge if they are in one unit cell or in two neighboring unit cells. The examples of allowed and prohibited dopant configurations are shown in Fig. 3-3a and 3-3b.

After constructing the channel Hamiltonian, the retarded Green's function of the channel at energy E is calculated as,

$$G(E) = [(E + i0^+)I - H - \Sigma_1 - \Sigma_2]^{-1} \quad (3-1)$$

where H is the Hamiltonian matrix in orthogonal (non-orthogonal) p_z TB basis for nitrogen (oxygen) doped channel, and Σ_1 (Σ_2) is the self-energy due to the semi-infinite dopant-free source (drain) lead. The electronic transmission per spin through the channel is calculated as

$$T_e(E) = \text{Tr}[\Gamma_1(E)G(E)\Gamma_2(E)G^+(E)] \quad (3-2)$$

where $\Gamma_{1,2} = -2\text{Im}(\Sigma_{1,2})$ is the broadening function of source/drain lead and G^+ is the advanced Green's function. NEGF formalism can rigorously treat the transport through the channel with translational symmetry breaking (e.g. by random dopants). Other non-idealities for transport such as phonon scattering and structural defects are neglected since this study focuses on the doping-limited mobility. With the transmission, the resistance can be computed as,

$$R = 1/G, \quad G = \frac{2q^2}{h} \int dE T_e(E) \left(-\frac{\partial f(E, \mu, T)}{\partial E} \right) \quad (3-3)$$

where G is the conductance and f is the Fermi distribution which is a function of energy, Fermi energy μ and temperature T . The factor of 2 counts the spin degeneracy. A proper Fermi energy μ is chosen to give a desired 2D carrier density. For a given

channel length, width, edge doping concentration, and 2D carrier density, 300 trial simulations with randomly generated dopant configurations are performed to obtain the ensemble averaged resistance, $\langle R \rangle$. Apparently not all of the resistance comes from the doped channel. Part of the resistance is the contact resistance, R_0 , or called ballistic resistance because it will exist even if the transport in the channel is ballistic. The resistance due to doping, defined as channel resistance, is $\langle R_{ch} \rangle = \langle R \rangle - R_0$. Then the channel length is changed to another value while other settings, such as the channel width, edge doping concentration, and 2D carrier density are fixed. Another 300 trial simulations are carried out and another $\langle R_{ch} \rangle$ can be obtained. Finally with enough points collected, a $\langle R_{ch} \rangle$ vs. L_{ch} curve can be plotted. A linear fitting can be done and the slope of the fitting line is the channel resistivity. The (electron) mobility is calculated as

$$\mu_n = \frac{dL_{ch}}{qn_{2D}d \langle R_{ch} \rangle} \quad (3-4)$$

where n_{2D} is the 2D electron density. The hole mobility is obtained by changing n_{2D} to p_{2D} , the 2D hole density.

The linear scaling of channel resistance with regard to the channel length is not guaranteed. The reason is that the transport simulation is phase-coherent since the scatters (dopants) in the channel are rigid scatters. If the channel length is much larger than the localization length, $L_c = ML_0$, where M is the number of modes in conduction and L_0 is in the order of the momentum relaxation length, then the resistance scales exponentially with the channel length. The transport is thus in the strong localization regime. Therefore the extraction of channel resistivity should be performed in the linear regime of the R vs. L_{ch} curve to avoid the onset of strong localization because at room

temperature operation, the phase relaxation processes (such as phonon scattering) ensure that the transport in the channel is not in the strong localization regime and only the linearly-scaled resistance matters to the carrier mobility [53]. The linear extrapolation of R -vs.- L_{ch} at zero channel length should give the ballistic resistance. As shown in Fig. 3-4, the extrapolated resistance is approximately equal to the ballistic resistance of the channel.

Simulation Results

First we show the transmission spectra of the structures in Fig. 3-1 calculated by ab-initio simulations and parameterized TB simulations. An energy range from -0.7eV to 0.7eV with regard to the charge neutral point (the middle of the bandgap) is chosen, which is relevant to electronics applications. The transmissions for nitrogen edge doping are displayed in Fig. 3-2a and 2b. The almost metallic GNRs in Fig. 3-2a have a small bandgap and a width index $n=3q+2$. The semiconducting GNRs in Fig. 3-2b have a relatively large bandgap and a width index $n=3q$. The $n=3q+1$ group is semiconducting and is similar to the $n=3q+2$ group. Generally the dopants cause the transmissions degradation at almost all energy levels (the perfect transmissions are step-wise and have integer number values per spin). However there is qualitative difference between the transmissions of valence subbands and conduction subbands. In both Fig. 3-2a and 3-2b, oscillations (peaks and dips) appear in the transmissions of the conduction subbands whereas the transmissions of the valence subbands monotonically increase as $|E|$ increases. As the channel width becomes smaller, the transmission peaks and dips occur at larger energies due to the larger energy spacing between subbands. For example the first dip of $n=9$ GNR does not show in the energy range due to its large bandgap. Qualitatively the transmissions of the $n=3q$ group and $n=3q+2$ are similar.

Quantitatively the valence-subband transmissions of the $n=3q+2$ group is less degraded than those of the $n=3q$ group.

For oxygen edge doping, oscillations of transmission similar to nitrogen edge doping are observed as well except that the conduction and valence subbands are switched. There are oscillations in the transmissions of valence subbands (the transmission dip of $n=9$ does not show in the energy range again) whereas the transmissions of conduction subbands monotonically increase as energy increases. The transmission oscillations can be qualitatively explained by a simple model. With a single parabolic band, a free electron ($E>0$) will be scattered by a square potential well ($U<0$) and the transmission as a function of E shows oscillations due to resonant scattering. For nitrogen edge doping, the ionized dopants (positively charged) induce a potential well for electrons and thus oscillations are observed in the transmissions of the conduction subbands. For oxygen edge doping, the ionized dopants (negatively charged) induce a potential well for holes and oscillations are observed in the transmissions of the valence subbands. Despite of the complications from the coupled subbands, non-square potential wells, and non-parabolic dispersions the fundamental physics is similar to the simple model. The shape of potential wells (profiles of on-site perturbations) can be conceptually thought of as a Gaussian distribution centered at a dopant but this is a very rough approximation. The magnitude of the potential well center is about 4 eV for nitrogen doping and 1eV for oxygen doping. This is possibly because an oxygen atom passivates a carbon atom while a nitrogen atom substitutes for a carbon atom. The magnitude difference in perturbations is the direct cause of that , as we will see later, the hole mobility in oxygen doped channel is consistently larger

than the electron mobility in nitrogen doped channel with the same channel width, doping concentration and carrier density.

The electronic structure of GNR is very sensitive to the channel width due to its extremely low dimensionality of quasi-1D channel. Therefore the mobility is expected to have a strong dependence on channel width. The channel width dependence of doping limited mobility is studied with a 2D carrier density of $1 \times 10^{13} \text{ m}^{-2}$ and an edge doping concentration of 2%. The edge doping concentration is defined as the ratio of the number of dopants to the number of edge carbon atoms. In Fig. 3-5a and 3-5b the mobility-vs.-channel-width curves are plotted for nitrogen doping and oxygen doping respectively. For each dopant specie, 2 curves for the $n=3q$ and $n=3q+2$ groups are presented. The general trend is very understandable that as the channel width increases, the doping limited mobility increases. This is because as the channel width increases, the overlap between the carrier wave function and dopants decreases and the propagation of carriers is less perturbed.

There is a roughly linear dependence of mobility on the channel width in the range of 4-10 nm. Below 4 nm, non-monotonic behaviors are observed which can be explained by the position of the Fermi energy with regard to the transmission dips and peaks caused by resonant scattering. With a fixed 2D carrier density of $1 \times 10^{13} \text{ cm}^{-2}$, the Fermi energy is about 0.3 eV above (below) the charge neutral point for electron (hole) conduction regardless of the channel width. Let us first check the curve for the $n=3q+2$ group for nitrogen doping. The Fermi energy $\mu \sim 0.3 \text{ eV}$ is near the first transmission peak of the small channel width (e.g., $n=8$). As the channel width increases, the position of first transmission dip is shifted toward μ . The transmission at μ decreases and

reaches the minimum value when μ is right aligned with the dip. As the channel width further increases, the first dip moves away from μ while the second transmission peak moves toward μ which causes the transmission at μ to increase. Thus a non-monotonic behavior of the mobility is observed due to the non-monotonic behavior of the conductance related to the transmission through Eq. 3-4. Although we are using the transmission by a single dopant (Fig. 3-2) to explain the non-monotonic behavior, the transmissions by multiple randomly generated dopants are qualitatively similar to that of a single dopant in terms of the positions of resonant peaks and dips. One may expect at large channel width, non-monotonic behavior in the mobility-vs.-channel width curves can still exist since there are a series of transmission peaks and dips in the conduction subbands and μ will keep sweep through them as the channel width increases. However except that the dip in the lowest subband can reduce the transmission to zero, the dips in higher subbands are much shallower and narrower which can be understood from the solution to the simple potential well problem. From Eq. 4-4 the thermal broadening ($T=300$ K) prevents the conductance and thus mobility to oscillate with the transmission peaks and dips of higher subbands. Using the same arguments, the non-monotonic behavior of the curve of the $n=3q$ group with oxygen doping can be explained. The $n=3q+2$ group of oxygen edge doping does not have the non-monotonic behavior because the transmission peak in the lowest valence subband is very low and narrow. When the transmission peak moves toward and then away from μ as the channel width decreases, the thermal broadening again makes the conductance insensitive to the small peak. Thus the mobility is a smooth and monotonic function of

the channel width rather than oscillating as the transmissions. At low temperature, the non-monotonic behavior is expected to be stronger.

Next we investigate the dependence of the doping-limited mobility on the doping concentration (the ratio of the number of dopants to the number of edge carbon atoms) at a fixed 2D carrier density of $1 \times 10^{13} \text{ cm}^{-2}$. As the doping concentration increases, the density of dopants increases and the carriers get more frequently backscattered so the mobility decreases as shown by the solid lines in Fig. 3-6a and 3-6b for nitrogen and oxygen doping respectively. The dashed lines are fitting lines proportional to $1/N_{A,D}$ where N_A/N_D is the acceptor/donor concentration. The $1/N_{A,D}$ is the dependence of the doping limited mobility on the doping concentration in silicon [17]. The doping-limited mobility in GNRs basically follows a similar dependence for both the narrow ($n=15$, $W=1.7 \text{ nm}$) and relatively wide ($n=66$, 8.1 nm) GNRs. The narrow ribbon ($n=15$) with nitrogen doping has extremely low mobility ($\sim 10 \text{ cm}^2/\text{Vs}$ at 1% doping concentration) as the doping concentration is larger than 1% whereas the mobility of oxygen doping is larger by one order of magnitude ($\sim 500 \text{ cm}^2/\text{Vs}$ at 1% doping concentration) for the same channel width and doping concentration, which is the result of the larger perturbation induced by nitrogen atoms than the oxygen atoms. In Fig. 3-6b for oxygen doping, the curve of $n=15$ GNR is very close to that of $n=30$ GNR in log scales. This is because the 2D hole density in use results a Fermi energy aligned to the first transmission dip of $n=30$ whereas the transmission dip of $n=15$ is away from the Fermi energy. It is the same reason that the mobility of $n=30$ ($W=3.6 \text{ nm}$) GNR is the dip of the curve for the $n=3q$ group in Fig. 3-5b. Typically with the same doping specie, the

mobility-vs.-concentration curve for a larger channel width is on top of the curve for a smaller channel width as shown in Fig. 3-6a.

In the study of the channel width dependence of doping limited mobility, it has been shown that the position of the Fermi energy with regard to the transmission dips and peaks are very important. The Fermi energy can be shifted by modulating the carrier density in the channel. The doping limited mobility as a function of 2D carrier density is plotted for several channel widths for nitrogen and oxygen doping in Fig. 3-7a and b respectively. The 2D carrier density is varied from 1×10^{12} to $2 \times 10^{13} \text{ cm}^{-2}$ and the upper limit of the magnitude of the Fermi energy is $|\mu| \sim 0.5 \text{ eV}$ regardless of channel widths. For a small channel width, $n=15$ of oxygen doping for example, the mobility increases as the carrier density increases. This is because the transmission dip in the lowest subband occurs far below $E = -0.5 \text{ eV}$ shown in Fig. 3-3b. As the Fermi energy moves downward (more negative) from a little above the valence subband edge into first subband (shown in Fig. 3-8a by the vertical lines), the transmission and thus conductance increases. More importantly, the increase of the conductance is faster than the increase of 2D carrier density and according to Eq. 4-5 the mobility increases. As the channel width increases, the transmission dips move to energies larger than -0.5 eV as shown in Fig. 3-8b. As the hole density increases, the Fermi energy passes the transmission dip which can cause the mobility to decrease since the carrier density increases whereas the conductance decreases. However a transmission dip does not necessarily reduce the mobility if it is very shallow and sharp. Another mechanism that can cause the doping limited mobility to decrease as the Fermi energy (carrier density) increases is the flat transmission. Because with a relatively flat transmission (the

transmission around E_{F2} in Fig. 3-8b), the conductance does not increase or increases too slow to keep mobility increasing as the Fermi energy moves down and carrier density increases. Therefore the mobility-vs.-carrier density curves show non-monotonic behavior for large channel widths.

Conclusions

We have presented a simulation study of the edge-doping-limited mobility. The edge chemistry of different dopant species is captured by ab-initio DFT simulations. A parameterized TB basis is obtained by calibrating transmissions calculated with the TB basis set to those calculated with large-size PAO basis set used in DFT simulations. The small size of parameterized TB basis (one orbital per atom) enables the construction of GNR channels of sizes beyond the ab-initio simulation capacity.

The nitrogen doping induces oscillations in the conduction subband transmissions whereas the oxygen doping induces oscillations in the valence subband transmissions. The oscillations are due to resonant scattering and can be explained by a simple potential well model. The resonant scattering plays an important role in the dependence of doping-limited mobility on the channel width, edge doping concentration, and 2D carrier density. The doping-limited mobility approximately scales as $W-W_C$ for $W>W_C$, where $W_C=4\text{nm}$. For $W<4\text{ nm}$, the mobility is small and is not a monotonic function of the channel width. The doping concentration dependence is similar to silicon, i.e., roughly proportional to $1/N_{A,D}$. For small channel widths, the mobility increases as 2D carrier density increases whereas the mobility shows non-monotonic behavior for large channel widths.

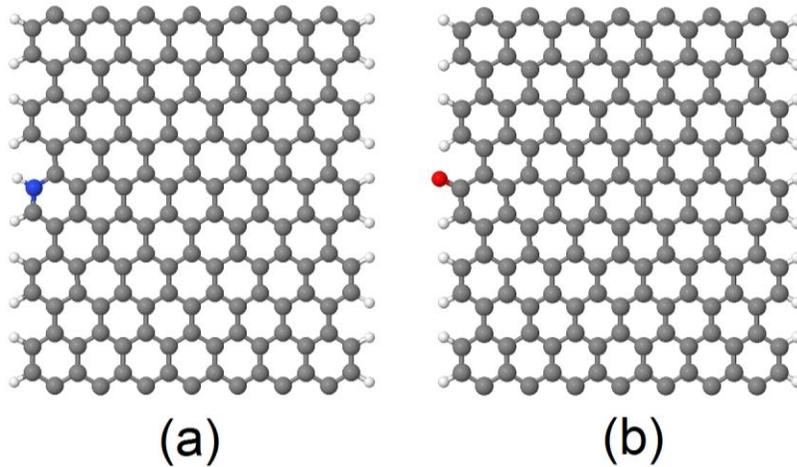


Figure 3-1. Atomic structures of $n=15$ AGNR super-cells with a single dopant. (a) Nitrogen edge doping by substitution for an edge carbon atom. (b) Oxygen edge doping by substitution for a hydrogen atom. There are five unit cells in a super-cell.

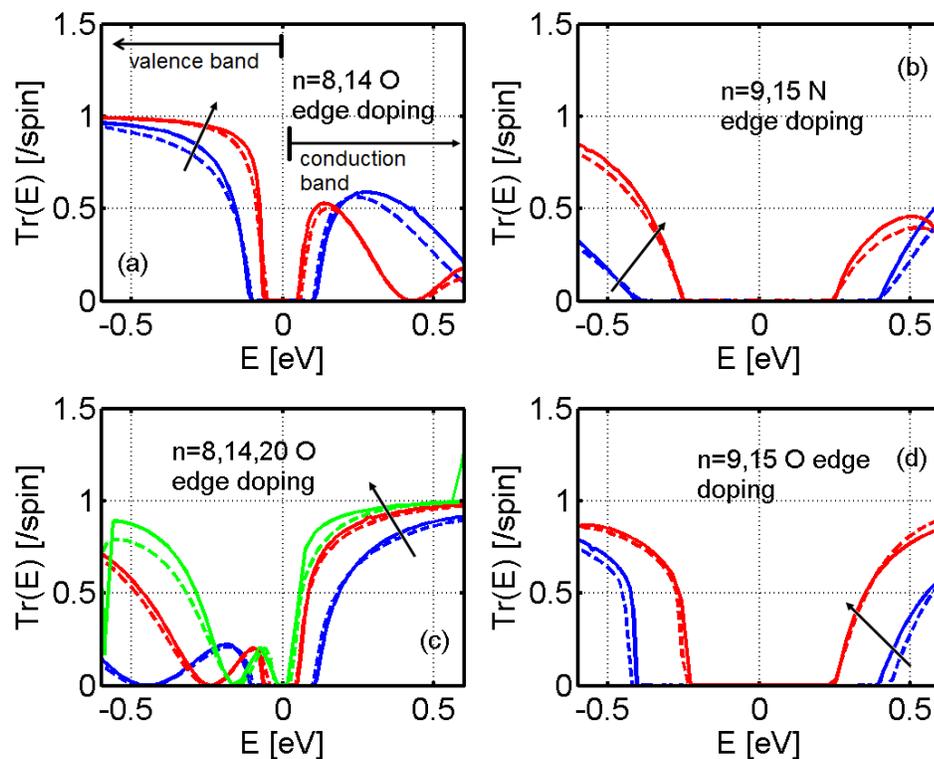


Figure 3-2. Transmission as a function of energy. Results are computed by ab initio transport simulations (solid lines) and the parameterized TB bases (dashed lines) for doped GNR segments shown in Fig. 3-1 of different channel widths. Nitrogen doping for (a) $n=3q+2$ group and (b) $n=3q$ group. Oxygen doping for (c) $n=3q+2$ group and (d) $n=3q$ group.

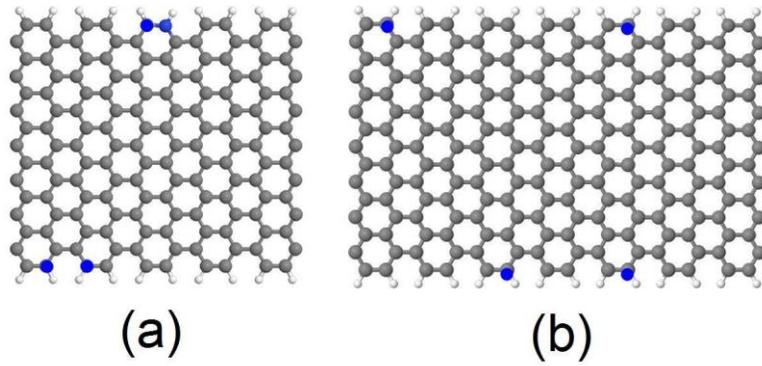


Figure 3-3. Configurations of multiple nitrogen dopants. (a) Prohibited configurations because there are dopants too close to each other. (b) Allowed configurations. The GNR axis is rotated by 90 degree from Fig. 3-1.

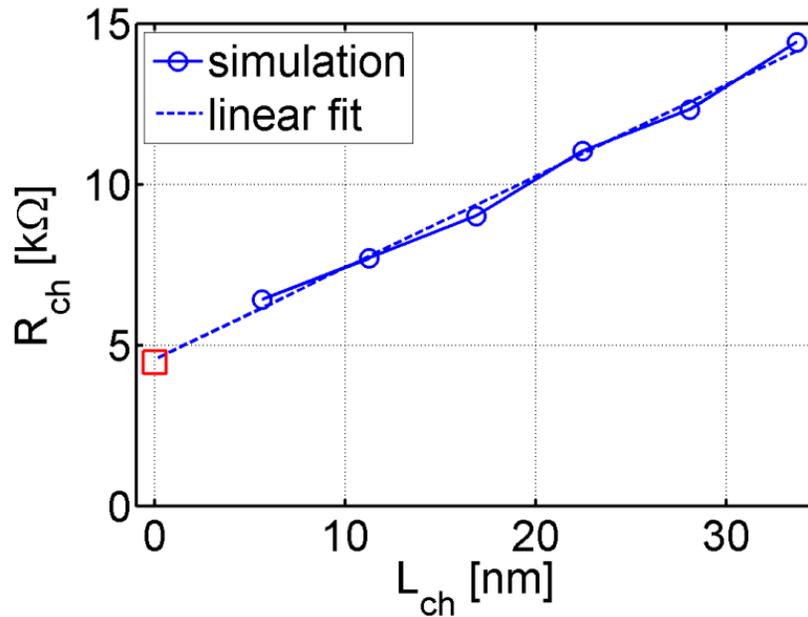


Figure 3-4. Resistance as a function of the channel length for $n=66$ nitrogen doped GNR with a doping concentration of 2%. The dashed curve is the linear fitting of the data points marked by blue circles. The red square shows the ballistic resistance and agrees well with the extrapolated resistance from the linear fitting at zero channel length.

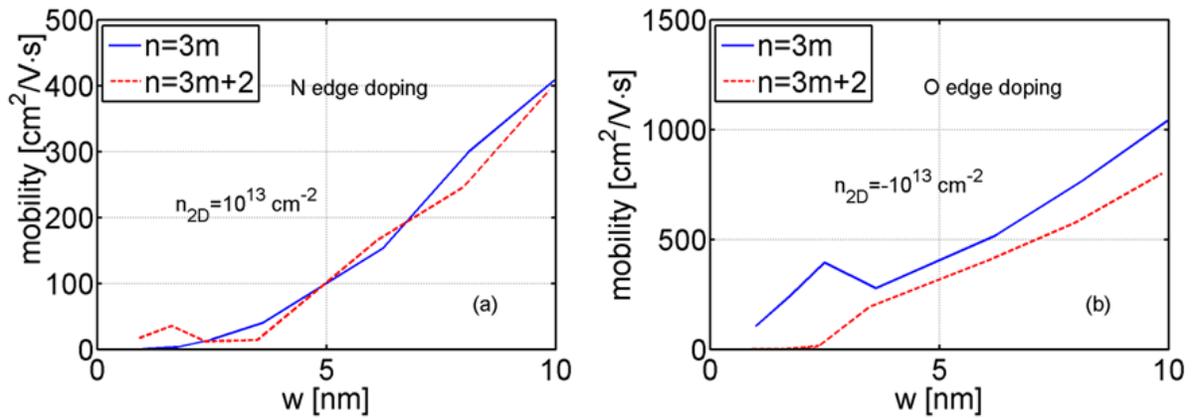


Figure 3-5. Edge-doping-limited mobility as a function of the GNR width. (a) n-type edge doping and for (b) p-type edge doping at a 2D carrier density of $1 \times 10^{13} \text{ cm}^{-2}$ and a doping concentration of 2%.

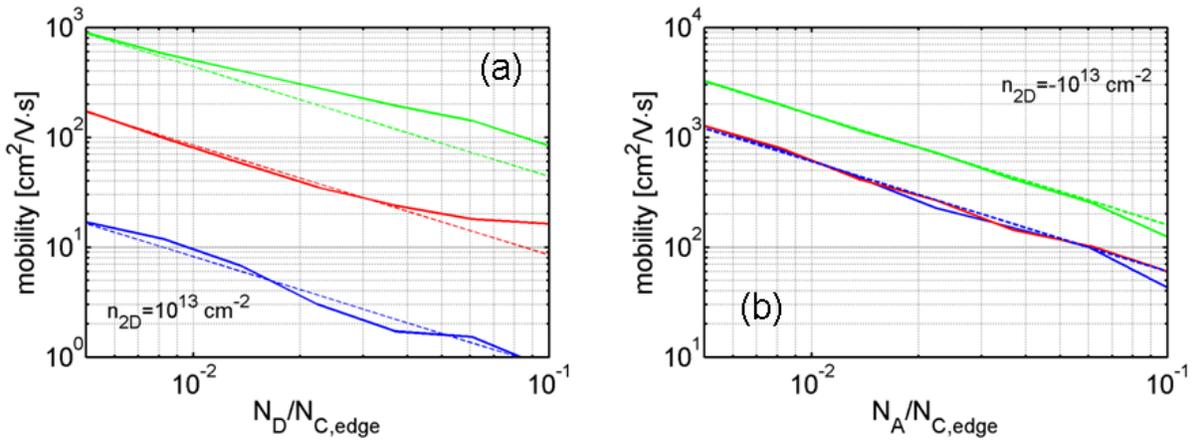


Figure 3-6. Edge-doping-limited mobility as a function of the doping concentration (a) nitrogen edge doping and (b) oxygen edge doping with a 2D carrier density of $1 \times 10^{13} \text{ cm}^{-2}$. The doping concentration is defined as the ratio of the number of dopants to the number of edge carbon atoms.

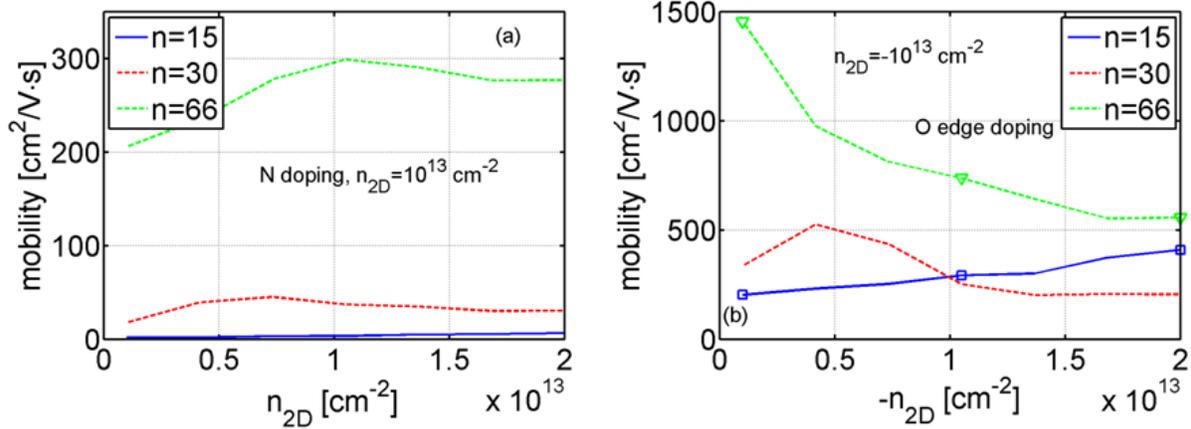


Figure 3-7. Edge-doping-limited mobility as a function of the carrier density. (a) Nitrogen edge doping and (b) oxygen edge doping at a doping concentration of 2%. The triangular and square symbols are marked at the carrier density of 1×10^{12} , 1×10^{13} , and 2×10^{13} cm⁻² which correspond to the Fermi energies labeled as E_{F1} , E_{F2} , and E_{F3} in Fig. 3-8 respectively.

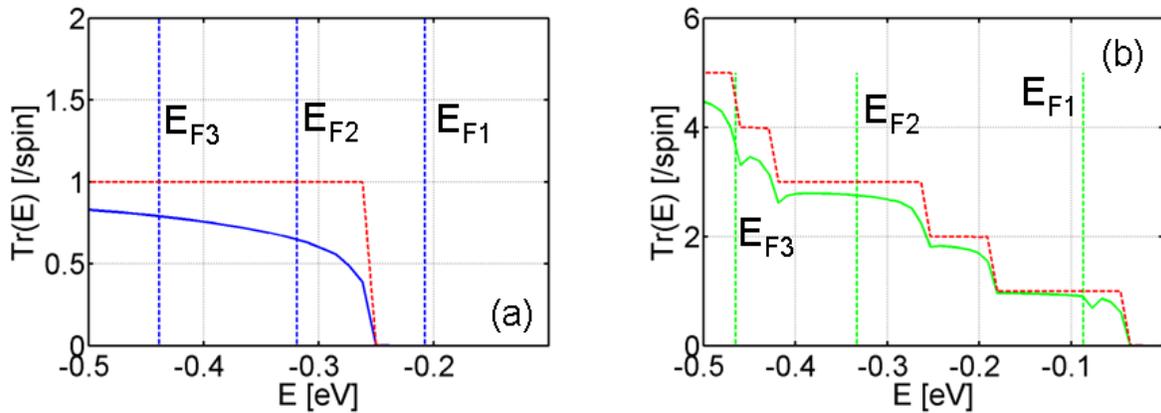


Figure 3-8. Electronic transmissions of (a) $n=15$ and (b) $n=66$ GNRs for oxygen edge doping. The red dashed curves are the step-wise perfect transmissions. The vertical lines labeled as E_{F1} , E_{F2} , and E_{F3} indicate the Fermi energies for the carrier density 1×10^{12} , 1×10^{13} , and 2×10^{13} cm⁻² respectively.

CHAPTER 4 SIMULATION OF GRAPHENE NANORIBBON TRANSISTORS WITH CARRIER SCATTERING

Scattering Mechanisms in Graphene Nanoribbons

The high mobility of graphene, which is a monolayer of carbon atoms packed into a two-dimensional (2D) honeycomb lattice, has stimulated strong interest not only in fundamental physics but also in potential technological applications due to its high carrier mobility [1-3]. A monolayer graphene sheet is a semimetal which does not have a bandgap. The graphene field effect transistors (FET) suffer a small on-off current (or conductance) ratio at room temperature. By patterning graphene into narrow strips, i.e., graphene nanoribbons (GNR), a bandgap, however, can be opened by quantum confinement in the width direction of a GNR [4-7]. With a channel width less than 5nm, GNR-FETs are able to provide an on-off current ratio of $\sim 10^4$ [31]. However with such a small channel width, the device performance is very sensitive to structural defects and impurities. The on-current and carrier mobility are thus degraded. The characterized field effect mobility is less than $300 \text{ cm}^2/\text{Vs}$. A recent work has demonstrated high quality GNRs derived from unzipping carbon nanotubes (CNT) [54]. The GNR width is around 14 to 17 nm which is about 4 times wider than previously reported ultra-narrow GNRs [31]. The bandgap for channel widths of this order is less than 100 meV. But even the moderately improved on-off current ratio by the small bandgap can significantly reduce the off-state power consumption because graphene FETs are very leaky at off-state. More importantly with such a small channel width of $\sim 14 \text{ nm}$, the GNR FET maintains a strikingly high field effect mobility of $1200 \text{ cm}^2/\text{Vs}$ with a channel length less than 90 nm at room temperature. This GNR mobility is much higher than those of previously reported GNRs with similar widths and the deep sub-micron Si MOSFETs of

which the typical mobility value is around $100 \text{ cm}^2/\text{Vs}$ [22], [29]. Therefore the high carrier mobility and improved on-off current ratio make the demonstrated high quality GNR-FETs with channels derived from unzipping CNTs very promising for high speed applications. Understanding the transport characteristics including the role of metal contacts and the scattering mechanism is vital to optimize device design and performance.

In Chapter 4 we model the experiment devices by solving the quantum transport equation. Unlike in Chapter 2 where ballistic performance is evaluated and scattering is treated in the semi-classic transport equation, scattering is included in quantum transport simulations to explain the observed conductance (G_σ vs. V_g) characteristics. The major findings from simulating experiment devices are briefed as follows. More details on the underlying physics and simulation method will be given in the following subsections.

We found the metal contacts play an important role in the transport characteristics. They shift the Dirac point in the GNR leads (i.e., GNR sections right underneath metal source/drain contacts) and cause the non-symmetrical conductance vs. gate voltage characteristics (with regard to Dirac voltage) [55]. The shift of the Dirac point is accompanied with a charge transfer from the metal contacts to GNR leads, which results doping of GNR leads. The doped GNR leads limit the number of subbands contributing to transport at on-state. By solving a 2D Poisson equation (details are given later), we found there are 4 to 5 subbands (spin degenerate) contributing to conduction at on-state which agrees well to the number of conductance steps in low temperature ($\sim 50\text{K}$) G_σ - V_g curves. The ballisticity (the ratio of measured on-state conductance to the

ballistic limit $\sim M \times 2 e^2/h$, where M is the number of conducting subbands) of the GNR-FETs is about 0.5. Due to the low density of states (DOS) around Dirac point, the Schottky barrier (SB) height is variable as the gate voltage changes. Including the change of SB height in simulations is necessary to obtain quantitative and qualitative agreement between simulations and experiments on the non-symmetrical $G_{\sigma} V_g$ characteristics.

Carrier scattering due to structural defects and charge impurities is treated, which causes ripples to the $G_{\sigma} V_g$ curves as observed experimentally. The magnitude of current degradation (i.e. rate of backscattering) is not the same for the two scattering mechanisms. The charge impurity scattering only slightly decreases the current with a typical impurity density of $\sim 10^{12} \text{ cm}^{-2}$ in graphene. The scattering due to structural defects such as edge roughness and vacancy can reduce the current to 50% of the ballistic limit when the defect sites are about 4% of total atoms in the channel. Phonon scattering is found not important in these short-channel GNR devices because theory-predicted phonon scattering mfp is comparable or larger than the channel length and the measured on-state conductance does not decrease as temperature lowers.

The off-state conductance strongly depends on temperature. However the dependence is not similar to the case of a Si Schottky diode which has a much larger bandgap. Therefore the bandgap of GNR-FETs cannot be extracted by the way of extracting SB heights of Schottky diodes. The simulations capture the qualitative features of the measured off-state conductance as a function of temperature. By best fitting to experiments, we are able to extract the bandgaps from simulations. The

extracted bandgaps are very close to the values given by a widely used p_z orbital tight binding schemes.

At low cryogenic temperature ($T < 5$ K), the GNR-FET comes into the regime of Coulomb blockade or single electron charging near the off-state, where NEGF (a mean field theory) is no longer applicable [56]. To simulate GNR-FETs in this regime, a master equation based on the multi-electron picture is solved. The simulated diamond patterns in 2D conductance (conductance as a function of gate and drain biases) agrees well with experiment. By mapping measured conductance patterns to simulated results, one can identify the origin of the excited states, i.e., which two single particle levels contribute to an excited state line in 2D conductance plot. The excited states can be fitted by the single particle levels derived from the electronic structure based on the tight binding calculation with quantized transport wave vectors (a particle-in-box problem). At high gate biases where the devices are not in Coulomb blockade regime, Fabry-Perot oscillations of conductance is observed in experiment. The transport is simulated by using the transmission matrix method [57]. The simulation shows the irregular patterns in 2D conductance plot emanating from multi-subband transport and scattering sites. By fitting to experiment, we identify that there are 2 scatters respectively located at $1/3L$ and $2/3L$ along the channel where L is the channel length.

The next 4 sections are organized as follows. Section 2 presents the self-consistent NEGF simulation method and results, which cover the conductance over entire gate voltage range and temperature above 100K. Section 3 shows the Fabry-Perot simulation results based on transmission matrix method at high gate biases and cryogenic temperature ($T < 5$ K). Section 4 gives the transport simulations at low gate

biases and cryogenic temperature where the devices are operated in Coulomb blockade.

Transport Simulations Based on NEGF

The experimental device structure is schematically shown in Fig. 4-1. The channel material is primarily identified as monolayer graphene by experiment means but possibility of bilayer is not totally excluded. We did simulations for bilayer channels and the fitting to experiment is much worse than the fitting by simulations based on monolayer channels. Theory has predicted that the two layers of bilayer graphene can be quite insulating to each other if there is a rotation between two layers [58]. Hence it's possible that even the experimental channel is bilayer graphene, the lower layer graphene is acting like an insulator and the channel is essentially monolayer. Therefore throughout Chapter 4 all simulations are performed for monolayer channels. There are two representative experiment devices. The channel sizes (length by width) are 86×14 nm² and 140×17 nm². The back gate has a SiO₂ thickness of 300 nm. The height of source/drain metal contacts is 20 nm. The simulation domain which is decided by the Poisson solver is 400 nm wide and $L+100$ nm long (that is a contact length of 50 nm on each side, source and drain) is included.

The quantum transport equation is solved in the non-equilibrium Green's function formalism. The retarded Green's function of the contact structure at energy E is calculated as [12],

$$G(E) = [(E + i0^+)I - H - U - \Sigma_1 - \Sigma_2]^{-1} \quad (4-1)$$

where H is the total Hamiltonian matrix of the contact structure, Σ_1 (Σ_2) is the self-energy of the semi-infinite GNR lead at source (drain) with a broadening of 10 meV, I is

the identity matrix and U is the self-consistent potential solved from a Poisson solver.

The electronic transmission per spin through the channel is calculated as

$$T_e(E) = \text{Tr}[\Gamma_1(E)G(E)\Gamma_2(E)G^+(E)] \quad (4-2)$$

where $\Gamma_{1,2}(E) = i(\Sigma_{1,2} - \Sigma_{1,2}^+)$ is the broadening function of the leads and G^+ is the advanced Green's function. The conductance G is computed through the Landau formula,

$$G = \frac{2q^2}{h} \int dE T_e(E) \left(-\frac{\partial f(E, E_F, T)}{\partial E} \right), \quad (4-3)$$

where f is the Fermi distribution which is a function of energy E , Fermi energy E_F and temperature T . The factor of 2 counts the spin degeneracy. The transmission and conductance of the end-contacts are calculated in a similar way. Above equations are quite generic and work for any basis sets.

The channel Hamiltonian can be modeled by two methods. One is to use the nearest p_z orbital tight binding basis set as Chapter 2 and 3. For the channel size of experiment devices, the number of atoms can be larger than 50,000. By partitioning the channel into many principal layers (PL) the Hamiltonian is organized into a tri-diagonal block form, which means one PL is coupled to only one neighboring PL on its right and one on its left. Inverting of a matrix in this form as in Eq. 4-1 can be efficiently done by applying the recursive Green's algorithm [49]. In real space each PL of minimum size P is about 60 by 60 (60 orbitals) for a width of 14 nm. In order to reduce the computational load, a mode space approach can be used. Then Equation 4-2 is modified to

$$T_e(E) = \sum_k \text{Tr}[\Gamma_{1,k}(E)G_k(E)\Gamma_{2,k}(E)G_k^+(E)] \quad (4-4)$$

where k is the subband index and the size of PL for each mode is just 1. The number of subbands, M , allowed in the GNR parts underneath contacts is 4~5. Hence only the lowest M 1D-subbands are needed in transport simulation if the mode space approach is applicable. However the condition to apply mode space approach is that the potential in width direction does not change or only varies a little [44]. This is basically satisfied for perfect AGNRs free of charge impurities. If charge impurities are present, there will be potential ripples corresponding to induced hole or electron puddles. At a given position along channel, potential variation can as large as 1 eV for impurities very close (~1 nm) to channel surfaces (computation details will be given later). The potential variation in width direction mixes transverse modes. To overcome this problem, we can solve the transport equation with coupled modes instead of solving 5 uncoupled modes [59]. The transport equation is modified to,

$$G_s(E) = [(E + i0^+)I - H_s - U_s - \Sigma_{1,s} - \Sigma_{2,s}]^{-1} \quad (4-5)$$

Where the subscript s means the quantity is a subspace of matrices in mode space representation. The potential matrix is now in the form of block-diagonal. A block is of size of $M \times M$ and is computed as

$$U_{s,ii} = \Phi_{s,i}^+ U_{ii} \Phi_{s,i} \quad (4-6)$$

Where U_{ij} is the potential matrix block of size $P \times P$ at i th PL in real space, and $\Phi_{s,i}$ is the subset of transformation matrix of size $P \times M$ at i th PL. If there is no potential variation in width direction, U_{ij} is a diagonal matrix with constant elements and after transformation the result is still diagonal. The computational load is much smaller than solving in real space but it is larger than uncoupled mode space. Strictly speaking there are more than M modes that are coupled to each other. For relatively slow potential

variation as induced by charge impurity, the coupling between the lowest M modes and other ($P-M$) modes are very small. This is analogous to the Fourier transformation of slow varying signals.

There are two ways to model the Hamiltonian with structural defects. One way is to remove the matrix elements corresponding to the defect sites. The other way is to add a very large potential (several hundred eV) at the defect sites. Apparently the coupled mode space approach is not applicable to the first one since there are no consistently defined transverse modes at PLs of different sizes due to removal of atoms. The latter one can work with coupled modes approach in principle because the size of Hamiltonian is the same as the perfect channel. However the number of modes needed is so large ($M \sim P$) that the coupled mode space does not provide much computational savings. Thus simulations with structural defects are performed in the real space presentation.

The charge density is calculated semi-classically by using the equilibrium DOS of GNR as,

$$n(r) = -q \int_{E_N(r)}^{+\infty} D_0(E) f(E, E_F) + q \int_{-\infty}^{E_N(r)} D_0(E) (1 - f(E, E_F)) \quad (4-7)$$

Where $n(r)$ is the charge density at position r , $D_0(E)$ is the equilibrium 2D DOS of GNR channel, $E_N(r)$ is the charge neutral energy, i.e. the Dirac point, at position r , and f is the Fermi distribution with a position-independent Fermi level because we are interested in the equilibrium or near equilibrium conductance.

The quantum transport equation is self-consistently solved with a three-dimensional (3D) Poisson equation for treating self-consistent potential on the channel.

The technique of method of moments (MoM) is used to solve the 3D Poisson equation [12]. The integral equation of MoM writes as,

$$\Phi(\vec{r}) = \int d\vec{r}' G(\vec{r}, \vec{r}') n_{3D}(\vec{r}') \quad (4-8)$$

where $\Phi(\vec{r})$ is the potential at r , $n_{3D}(r')$ is the 3D charge density at r' , and $G(r, r')$ is the Green's function for the Laplace operator. The physical meaning of the Green's function is the potential at point r produced by a point charge $\delta(\vec{r} - r')$ at r' . Following Ref. [12], the channel and surfaces of the electrodes, where charge can reside, are divided into 2D grids (the GNR channel is treated as a rectangular sheet of zero thickness). The Green's function for grid nodes is obtained analytically by applying the Coulomb law and method of image charge. The integral equation is therefore discretized into a set of linear equations suitable for numerical implementation,

$$\begin{bmatrix} \Phi_D \\ \Phi_B \end{bmatrix} = G(r; r') \begin{bmatrix} n_D \\ n_B \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} n_D \\ n_B \end{bmatrix} \quad (4-9)$$

where r and r' are geometrical centers of grids, Φ_D and Φ_B are the potentials on the grids of channel and boundaries, i.e. the surfaces of the metal source, drain, and gate respectively, n_D (n_B) is the charge (not charge density) at the grids of the device (boundaries), and A , B , C , and D are sub matrices of the Green's function in matrix version. The potentials on the source, drain, and gate, Φ_B , are the known quantities which are decided by applied source-drain and source-gate biases. Thus n_B can be eliminated by replacing it with Φ_B . After some algebra, the channel potential can be related to the charge density on the channel and the potential boundary conditions through

$$\Phi_D = (A - BD^{-1}C)n_D + BD^{-1}\Phi_B \quad (4-10)$$

The charge impurities can be included straightforwardly into the MoM solver by treating them similar to the grids of the channel (i.e., charge on the grids is known but not the potential). The difference is that the charge on these grids does not depend on the potential of the grids.

The above is good for solving the potential on the channel but not accurate for the potential on the GNR leads which are under the metal contacts as shown in Fig. 4-2, despite that the GNR leads only enter the transport equation through the self-energies. The potential at the two ends of the channel from MoM is not the potential on the leads. The grids at two ends are not right under the contacts. In addition the potential of a grid on the channel is the average potential over an area of a grid ($2 \times 2 \text{ nm}^2$) whereas the distance between a GNR lead and the metal contact above is only 0.34 Angs. Thus the potential at two ends is not a good approximation to the potential on the leads. The potential on the leads determines the non-symmetrical conductance characteristics and maximum conductance. An accurate evaluation is needed. In principle one can discretize the contact surfaces and GNR into grids as small as $\sim 0.3 \times 0.3 \text{ Angs}^2$ to obtain accurate result but this will impose enormous computational load to the 3D MoM Poisson solver. Instead we consider a 2D Poisson problem as shown in Fig. 4-3. It is essentially a MOS structure with the metal contact served as a top gate with $t_{ox}=0.34 \text{ Angs}$ and $\epsilon_r=1$. The applied voltage between the GNR lead and metal contact is the work function difference. Again equilibrium DOS of GNR is used to determine the self-consistent potential on the leads. Ideally the change of potential on the GNR lead should be negligible (a fixed SB height) as the gate voltage varies since the back gate is far away ($t_{ox}=300 \text{ nm}$). However the numerical result in Fig. 4-4 indicates that the

potential can change about 100 meV as the gate voltage is swept from -40V to 40V. The reason is that the bottom dioxide has a dielectric constant of 3.9 and the fringe field causes the capacitance several times larger than the simple estimation $\epsilon_r \epsilon_0 / t_{ox}$. Considering the maximum potential is about 280 meV, 100 meV is not very small. It affects the shape of the non-symmetrical conductance curve which will be shown later.

This 2D treatment assumes the third dimensionality (GNR axis) is homogeneous. This is not true for the part of leads near the contact edges. But at the position about $3t_{ox}$ (~ 1 nm) from the edge this 2D treatment becomes a reasonable approximation. The transition region (~ 1 nm) is much shorter than GNR leads. Therefore the self-energies are computed using the potential solved from the 2D Poisson equation.

The work function difference is set to 460 meV as derived from the work functions of graphite and palladium. First we check experiment data at $T=290$ K and compare them to simulations for the device of a channel length of 86 nm and a width of 14 nm (called as device 1 hereafter) in Fig. 4-5. The gate-channel capacitance computed by MoM is 0.41aF. The peak field effect mobility $\mu = L^2 / C_g \times dG / dV_g$ is 1600 cm^2/Vs . This value includes the quantum resistance due to the finite number of modes. To exclude the effects of ballistic resistance, G is changed to $G' = MG_0 / (MG_0 - G)$, where $G_0 = 2q^2/h$ is the conductance quanta and M is number of transport modes. This expression is physical as $G' = +\infty$ when $G = MG_0$ implying the channel is a perfect conductor. In Fig. 4-5 the ballistic conductance is about 2 times larger than the experiment value. So the channel mobility is about 3000 cm^2/Vs . To fit the experiment curve, the ballistic conductance is multiplied by a gate voltage-independent factor α ($\alpha < 1$, ~ 0.5 for device 1) to count scattering inside the channel. The fitting agrees well to experiment despite

that the approach is simply phenomenological. The non-symmetrical curve is caused by the polarity of carriers at different gate biases. At positive biases, electrons are induced in the channel whereas the carriers in the GNR leads are holes. Thus the band-to-band (BTB) tunneling causes extra resistance and the n-branch conductance is smaller than that of p-branch where no BTB occurs [55]. Conductance characteristics of similar qualitative features are also obtained for the device with $L=140\text{nm}$ and $W=17\text{nm}$ (called as device 2 hereafter) as shown in Fig. 4-6. Good fitting is obtained by scaling the ballistic conductance with $\alpha=0.5$ and a parasitic resistance of $3\text{ k}\Omega$. Compared to the conductance curves with fixed lead potential, the variable potential on the GNR leads has little effect on the conductance curve of device 1 whereas it enhances the curve's non-symmetry for device 2 as shown in Fig. 4-7. The reason is that at very negative gate voltages, the potential on the leads is increased. Hence the number of allowed transport modes inside the GNR leads increases. If the channel is the transport bottleneck, more conducting modes in the leads will not help boost the conductance as for device 1 because the potential in the channel is hard to increase due to the small C_g per length (small channel width). On the other hand the GNR leads are the transport bottleneck for device 2 which has a larger C_g per length and thus is more conducting modes at on-state. The slight increase (decrease) of lead potential at very positive (negative) biases increases (reduces) conductance.

Next we show the comparison of the experiment data and simulations for device one at $T=50\text{K}$ in Fig. 4-8. The scaled ballistic conductance by a factor of 0.5 basically agrees with the experiment curve, especially in terms of the number of conductance steps in p-branch but it misses some features. The scaled ballistic conductance is

smoother than the experiment curve, does not have small dips in p-branch (i.e. conductance never decrease as $|V_g|$ increases). This means the method to treat scattering by phenomenological scaling is no longer a good approximation at low temperature. Instead we need to treat the scattering realistically to better fit experiment. The maximum value of on-state conductance does not increase as temperature drops, which is indicative of negligible phonon scattering. Then a possible cause is charge impurities which are also important in graphene. The conductance curve with a particular configuration of charge impurities has dips in p-branch, which better fits the experiment data than the simple scaling as shown by Fig. 4-9. In this particular impurity configuration (Fig. 4-10), there are 5 impurities with a charge of $1q$ each, located in a plane 1nm above the channel. The corresponding 2D impurity density is about $4 \times 10^{11}/\text{cm}^2$, which is similar to the impurity density found in graphene. However we find that the conductance degradation due to charge impurity scattering is very small (about 10%). A scaling factor is still needed to fit experiment for the curves in Fig. 4-9.

Next we investigate the effects of structural defects. This is done by adding a very large potential at defect sites. The dominant structural defects are edge roughness for narrow GNRs. In Fig. 4-11, we show the simulated conductance curves with different percentages of edge roughness (the ratio of defect sites to the number of edge carbon atoms in the perfect structure). Edge roughness scattering introduces strong oscillations (dips and peaks) in conductance. With edge roughness of 4% the ballistic conductance is reduced by about 50%, close to the experiment. Thus we conclude edge roughness is the dominant cause for the conductance degradation. Just as charge impurity scattering, a particular defect configuration for edge roughness is need to fit the

experiment data more closely. The effects of charge impurities and edge roughness cannot be deconvoluted by simulations. At high temperature the detail of defect configurations is not important because large energy range for transport and strong phase relaxation will smooth conductance curves and only the ensemble average values are meaningful. This is possibly the reason that the simple scaling of ballistic conductance works so well at high temperature.

In conclusion of this section, the electrostatics simulation of GNR under metal contacts indicates the SB height is variable as gate voltage changes. Transport simulations including this mechanism can better fit the experiment data. The simulation also tells that the finite doping of GNR leads under metal contacts can be the transport bottleneck. Thus increasing the doping of GNR lead is as important as improving the gate efficiency. Transport simulations with carrier scattering show that improving edge quality can significantly boost the device performance.

Extraction of Bandgaps

It has been suggested that the energy gap appeared in narrow GNRs is actually not a pure bandgap. It is a transport [60]. There are still localized gap states in the gap region which does not deliver current. Transport gaps with localized gap states are not desired because they decrease carrier mobility. For these high quality GNRs which operate at the 50% of the ballistic limit, the transport gap is expected to be caused largely by the bandgap from width confinement rather than by scattering due to gap states. The bandgap can be characterized by conductance measurement as a function of gate and drain biases at cryogenic temperature as illustrated in Fig. 4-12a [61]. The idea is briefed as follows. With a proper gate voltage, the source and drain Fermi levels can respectively align to the conductance and valence band edges simultaneously with

a source-drain bias equal to the bandgap. At this particular gate voltage, conductance is very small when the drain bias is less than the bandgap and becomes much larger when the drain bias is larger than the bandgap. Therefore the largest $|V_d|$ value of an area (usually of a diamond shape) of very small conductance in the 2D differential conductance plot (Fig. 4-12b) gives the transport gap. However this method only works exactly for flat bands. The value yielded by this method may not be accurate due to band bending as explained in Fig. 4-13. With downward bending (Fig. 4-13) the change of conductance is gradual as the drain Fermi level moves around the valence band edge at the middle of channel due to the continuous change of tunneling distance dt . In contrast, with flat bands the tunneling distance for carrier decreases from the whole channel length to zero as the drain Fermi level moves across the valence band edge, which results in a abrupt change of conductance. Similarly a gradual change of conductance occurs as the source Fermi level moves around the conduction band edge due to BTB. The gradual change of conductance introduces large uncertainty in identifying maximum $|V_d|$ and hence the bandgap. For the reasons stated above, the relation $I_{off}(T) \propto \exp(-E_g / 2k_B T)$ where I_{off} is the off-state current and k_B is the Boltzmann constant, cannot be used to extract the transport gap either because it only works for flat bands and mid-gap SB.

First we inspect the experiment curve of $\ln(G_{d,off})$ vs. $1/T$ for device 1 in Fig. 4-14. Before using NEGF transport simulations which rigorously treat both thermionic and tunneling currents, to extract bandgaps. Apparently the curve is not linear, indicating that the $G_{d,off}$ is not a simple exponential function of $E_g/k_B T$. The slope of the curve increases as temperature increases. The slope is related to certain energy barrier

height, which is similar to the relation between the reverse saturation current and SB height in diodes. The relevant energy barrier here is the thermal barrier as shown in Fig. 4-15a, which is defined as the spacing between the Fermi level and the lowest valence band edge for p-type doped leads. In Fig. 4-15b the thermal barrier height is plotted as a function of SB height for holes at off-state (minimum conductance) for different temperatures. The thermal barrier height is equal to SB height only for mid-gap devices. At a given temperature, the thermal barrier height increases as SB height decreases. This is because as hole SB height decreases from $E_g/2$, i.e. the middle gap device, the transmission at energies above Fermi level decreases due to the blocking by the band bending near source/drain ends. The middle part of the bands can thus be further lowered to block more transmission at energies below the Fermi level, which leads to the increase of thermal barrier. For a given SB height ($<E_g/2$), the thermionic conduction dominates at energies below the Fermi level at off-state. As temperature increases, this part of conduction increases faster than the conduction above the Fermi level, where tunneling transport dominates. In order to minimize the conductance, the middle part of the bands is lowered to block more thermionic conduction. Hence the thermal barrier height increase as temperature increases and so is the slope of the curve. However there is no definite proportionality between the slope and thermal barrier height.

The procedure of extracting bandgaps is as follows. The minimum conductance for a given bandgap is found for each temperature point (in the range of 70 K to 290 K) by the self-consistent NEGF simulation. The scattering is included by scaling the ballistic conductance as discussed above. The scaling factor is both gate-voltage and

temperature independent. A root mean square (RMS) error is defined as $\sigma = (\langle [G(T) - G_{exp}(T)/G_{exp}(T)]^2 \rangle)^{1/2}$, where $G(T)$ and $G_{exp}(T)$ are the simulated and experimental conductance at T , respectively, and the average is taken over temperature. We found that the slope of the $\log(G)$ vs. $-1/T$ curve is dominantly determined by the bandgap for negative SB heights, and it is independent of the scaling factor. After a group of curves of different bandgaps are simulated, the best fitting bandgap, i.e., the extracted bandgap, is given by the curve that has the smallest RMS error, σ_{min} . For device 1 and 2, the extracted bandgaps are $E_g = 72 \text{ meV}$, and $E_g = 60 \text{ meV}$ respectively as shown in Fig. 4-16. By allowing a fitting error of $2\sigma_{min}$, the extracted E_g of the devices can be varied by $\pm 18 \text{ meV}$, $\pm 17 \text{ meV}$, from their best fitting values.

Fabry-Perot Simulation

At cryogenic temperature ($T < 5 \text{ K}$) the carrier phase relaxation length is large and electron wave function interference can be manifested via transport properties. At very negative gate biases the potential profile in the channel is relatively flat as the GNR leads are p-type doped. With contact resistance at two ends the GNR channel is thus like a Fabry-Perot interferometer for holes. In metallic CNT it has been observed that the differential conductance oscillates periodically with gate biases [62]. However the differential conductance pattern of the high quality GNFETs at high gate biases is not periodic. To understand this phenomenon we simulate the GNR using scattering matrix method by assuming uniform potential profile in the channel. In principle NEGF simulations can give all the results from scattering matrix method. However the 2D differential conductance plot requires a lot of simulation runs at different biases. The scattering matrix method assuming a uniform potential profile is much cheaper than NEGF since it works with plane waves rather than atomic orbitals.

The transmission spectrum is computed by coherently combining a number of scattering matrices [57]. The computational details are given in the Appendix. Here we give a brief description. The size of the scattering matrices is determined by the number of (spin-degenerate) conducting modes at a particular carrier energy level. The two contacts and impurities are described by the scattering matrices that can cause intra- and inter-mode scattering, i.e., having non-zero off-diagonal elements. The perfect channel segments sandwiched between the contacts or impurities are represented by the diagonalized scattering matrices, $[\exp(ik_{x,n}D)]_{nn}$, where the wave vector k_n (along GNR axis) at carrier energy E is found from a Dirac $E-k$ with the quantized transverse wave vector $k_{y,n}$ and D is the length of a segment. After obtaining the transmission spectrum we can calculate the differential conductance through Landau formula. The Fermi energy ε_F is modulated by the gate. The gate voltage is related to the Fermi energy through serial combination of the gate insulator capacitance C_g and the quantum capacitance of the GNR device which is proportional to the density-of-states (DOS) at Fermi energy at low temperature. For hole conduction at cryogenic temperature, the gate voltage $V_g \approx -e \int_{\varepsilon_F}^{\varepsilon_0} g(\varepsilon) d\varepsilon / C_g + \varepsilon_F / e$, where ε_0 is the valence band edge, and $g(\varepsilon)$ is DOS. The bias is applied to the source and drain as $\varepsilon_F - V/2$ and $\varepsilon_F + V/2$ respectively.

In Fig. 4-17 simulated G_d vs. V_g at $V_d=0$ for device 1 is compared to experiment data. The simulation involves 4 conducting modes in the displayed gate voltage range and 2 scatters located at $1/3L$ and $2/3L$ (the impurity configuration in Fig. 4-10 can be viewed as 3 scatters), where L is the channel length. The 2-scatter configuration results in conductance ripples that are similar to the experiment (small peaks come in pair to form a large peak). The F-P pattern is more complicated than that of metallic single

walled carbon nanotubes where the fluctuation is almost periodic. The difference is primarily caused by the multi-subband transport in the GNR. One can show that for one subband transport (spin degenerate or valley degenerate or both), the gate voltage period for the conductance peaks is constant no matter the channel is metallic or semiconducting because the group velocity $v(E)$ times DOS is constant at all energies. For multi-subband transport, the group velocity of each subband times the *total* DOS is no longer a constant and the pattern will not show constant gate voltage periods for the conductance peaks. In Fig. 4-18 the 2D differential conductance is shown. Patterns of diamond shapes are observed. The size of a pattern along the drain bias axis is defined as the distance between a conductance minimum and a nearest conductance maximum at same gate voltage, which is directly related to the carrier energy at certain transverse mode. For example the carriers (holes) in the lowest subband have a almost constant carrier velocity $\sim v_F^0$ around $V_g \sim -28V$. Therefore the pattern size is $V_C = hv_F^0/2Le \sim 21mV$ as one can observe there is a conductance peak at the point of $V_g = -28 V$ and $V_d = 21 mV$. However this peak is not clear cut since other subbands also results in patterns of different sizes (The higher subband has a smaller pattern). When those patterns are added up, the corners of resultant diamond pattern are blurred. In conclusion the multi-subband transport in the GNRs results in an irregular F-P pattern of conductance.

Transport in Coulomb Blockade Regime

Coulomb blockade happens when the single electron charging energy U_0 is much larger than both $k_B T$ and γ , where γ is the broadening due to source/drain leads to the channel. U_0 is defined as $U_0 = q/C_E$, where C_E is the total electrostatic capacitance from all electrodes [56]. For device 2 the channel-gate capacitance calculated by MoM is

1.1aF and channel-source (drain) capacitance is 2.3 aF. The computed U_0 is 28.1meV, much larger than $k_B T$ for $T < 5K$. Around off-state, the small channel conductance indicates a small γ , thus the transport could be in the Coulomb blockade regime. The measured 2D differential conductance is shown in Fig. 4-19. The diamond patterns in the figure are caused by single electron charging effect rather than the F-P oscillations discussed in last section. There several reasons. First the magnitude of the peak conductance is very small, in the order of $0.01 G_0$, where $G_0 = 2q^2/h$ is the conductance quanta. If it were F-P oscillation the peak conductance should be around $0.5 G_0$. At zero drain voltage, the conductance as a function of the gate voltage does not have a sinusoidal shape as the F-P pattern. Rather the conductance is close to zero everywhere except at some gate biases there are peaks of very narrow width. Lastly the size of F-P pattern (half height of a diamond) corresponds to a wave vector increment that equals to $2\pi/L$ for constructive interference. On the other hand the size of Coulomb blockade pattern is equal to the single electron charging energy plus difference between single particle levels. The size shown in Fig. 4-19 is obviously too large for F-P oscillation of the lowest subband. Therefore the diamond patters in Fig. 4-19 are a manifest of transport in Coulomb blockade regime.

First we address the factors that determine the shape and size of the diamond patterns in the 2D differential conductance plot. A two-single-particle-level system ($\epsilon_{\uparrow} = \epsilon_{\downarrow} = 0$, spin degenerate) is used to illustrate the points and one of the two states, say $\epsilon_{1,\uparrow}$, is populated at $V_d = 0$ and $V_g = 0$. The conductance peaks are due to changing of number of electrons in the system as shown in Fig. 4-20. To fill in another electron at $V_d = 0$, the gate voltage has to change by an amount equals to q/C_g . Another

interpretation is like this. Due to the presence of one electron, the one-particle energy of the down spin state is raised by an amount of single electron charging energy. To fill one electron in ϵ_{\downarrow} at $V_D=0$, the gate electrode has to lower the energy of the down spin to zero. The change of energy is given by

$$d\epsilon = -q dV_g \frac{C_g}{C_E} - q dV_s \frac{C_s}{C_E} - q dV_d \frac{C_d}{C_E} \quad (4-11)$$

With $d\epsilon=U_0$ and $U_0=q/C_E$, the change of gate voltage is simply q/C_g . At $V_g=V_1$ ($0 < V_1$ but $V_1 \ll q/C_g$) and $V_D=0$ the conductance is ~ 0 as depicted in Fig. 4-19a because the source and drain Fermi levels are positioned in the middle of the two levels shown by Fig. 4-20a. By increasing the drain voltage with V_g fixed, a conductance peak occurs at V_2 as shown in Fig. 4-20a. This conductance peak corresponds to removing one electron from the device. By lowering the drain Fermi energy (increasing V_D) to the energy of occupied state, the electron in that state exits to the drain contact and results in a conductance peak as illustrated in Fig. 4-20b. Hence we have the equation,

$$-qV_2 = -qV_1 \frac{C_g}{C_E} - qV_2 \frac{C_d}{C_E} \quad (4-12)$$

This leads to the slope $V_2/V_1 = C_g/(C_E - C_d)$. If V_D is decreased then the energy of occupied state will be increased. As the energy reaches zero at $V_D = -V_2$, equal to the source Fermi level, the electron exits to the source and a conductance peak occurs as shown in Fig. 4-20c. By this argument, a similar equation to Eq. 4-9 writes

$$0 = -qV_1 \frac{C_g}{C_E} + qV_2 \frac{C_d}{C_E} \quad (4-13)$$

And the slope is $-V_2/V_1 = -C_g/C_d$. For a device of symmetrical source and drain, $C_d = C_s$. Therefore the magnitudes of the slopes are not the same since C_E

$C_d = C_s + C_g > C_d$. This explains why the diamonds formed by conductance peaks are “tilted” in the conductance plot. From the diamond due to filling two electrons to a spin-degenerate state (ϵ_{\uparrow} and ϵ_{\downarrow}) one can extract the capacitances between the electrodes and channel. If the diamond emanates from filling two non-spin-degenerate states ($\epsilon_{1,\uparrow} \neq \epsilon_{0,\downarrow}$), then the diamond size is larger than that in the spin-degenerate case by $\epsilon_1 - \epsilon_0$ along drain bias axis and $(\epsilon_1 - \epsilon_0)C_E/C_g$ in gate bias axis. Above analysis is for electron conduction and is straightforward to adapt for hole conduction.

With the coupling capacitances between the electrodes and channel, single electron levels, and source/drain broadening, we can model the GNR-FET in Coulomb blockade regime by solving the many-particle Fock space master equation. For a system of N single particle levels (spin non-degenerate), the number of many-body states is 2^N . The weak coupling to the source and drain contacts causes the transition from one many-body state to another, which generates the current flow. The initial and final many-body states can only differ by one single level (one less or more). If the final state β has one more or one less electron in the single level ϵ_i than the initial state α , then the transition rate is,

$$R_{\beta\alpha} = \frac{\gamma_1}{\hbar} f_1(E_{F1}, \epsilon_i) + \frac{\gamma_2}{\hbar} f_2(E_{F2}, \epsilon_i) \quad (4-14)$$

Or,

$$R_{\beta\alpha} = \frac{\gamma_1}{\hbar} (1 - f_1(E_{F1}, \epsilon_i)) + \frac{\gamma_2}{\hbar} (1 - f_2(E_{F2}, \epsilon_i)) \quad (4-15)$$

where γ_1 (γ_2) is the coupling strength between the source (drain) contact and channel, f_1 (f_2) is the source (drain) Fermi function at the single level ϵ_i , and E_{F1} (E_{F2}) is the source (drain) Fermi level. The single particle level ϵ_i depends on other single particle levels in

the many-body state that has ε_i , and the external electric field exerted by the electrodes.

The equation for ε_i is,

$$\varepsilon_i = \varepsilon_i^0 + (N-1)U_0 - qV_g \frac{C_g}{C_E} - qV_d \frac{C_d}{C_E} \quad (4-16)$$

where ε_i^0 is the single level energy when voltages on electrodes are all zero, N is the number of electron in the many-body state (the initial or final many-body state). At steady state there is no net in-flow or out-flow for any many-body state. Hence we have following equation for a particular state α ,

$$\sum_{\beta} R_{\alpha\beta} P_{\beta} = \sum_{\beta} R_{\beta\alpha} P_{\alpha} \quad (4-17)$$

where P_{α} is the steady state probability that the system is in many-body state α . It states that the rate at which other states are scattered into α is equal to the scattering rate of α into other states, i.e. no net flow. Together with the normalization condition for probability, the probability at each state can be solved. A terminal current is the flow caused by a coupled electrode (source/drain). The output of the simulation is the conductance as a function of the drain and gate voltages. Due to the small energy scales, a high resolution scan of conductance (Fig. 4-21) is used to check the validity of the simulation shown in Fig. 4-22. The simulation captures Coulomb diamond shapes and sizes, as well as the excited state lines. By comparing the simulation with experiments, one can clearly identify where each excited state line comes from as labeled in the simulated conductance figure (Fig. 4-23). These excited state lines can come from adjacent (in energy scale) single particle levels (e.g. $\varepsilon_1 - \varepsilon_0$) as well as non-adjacent energy levels (e.g. $\varepsilon_3 - \varepsilon_0$). One can also extract capacitances and single particle levels by best fitting of the simulated results to experimental measurements.

For device 2, the gate capacitance is estimated as $C_g \sim 0.87$ aF from the gate voltage periods of conductance peaks at $V=0$. The drain capacitance $C_d \sim 2.13$ aF is found by the slope of the diamond boundary (the one that has negative slope). From single electron charging energy total capacitance can be computed, and source capacitance $C_s \sim 2.70$ aF is obtained. From the experiment data (Fig. 4-20), we can establish following mappings between single particle levels and energy values: $\epsilon_3 - \epsilon_0 = \Delta\epsilon_{30} \sim 19.44$ meV, $\epsilon_2 - \epsilon_0 = \Delta\epsilon_{20} \sim 10.08$ meV and $\epsilon_1 - \epsilon_0 = \Delta\epsilon_{10} \sim 2.52$ or 3.6 meV (ϵ_i is a spin-degenerate level). We use a simple hyperbolic $E-k$ model, which is a good approximation to the $E-k$ by tight binding at low energy, to compute the single particle levels due to quantization in longitudinal direction and to fit the level differences. The hyperbolic $E-k$ for the lowest subband is of the form [63],

$$E(k) = \sqrt{\left(E_g / 2\right)^2 + \hbar^2 v_F^2 k^2} \quad (4-18)$$

where E_g is the bandgap, k is the transport wave vector, v_F is the Fermi velocity, and \hbar is the reduced Plank constant. The k values that gives rise to the single particle levels are $k_n = n\pi/L$ ($n=1,2,\dots$). With $E_g=60$ meV and $L=135$ nm, the corresponding differences between single particle levels are $\Delta\epsilon_{30} \sim 20.02$ meV, $\Delta\epsilon_{20} \sim 10.15$ meV and $\Delta\epsilon_{10} \sim 2.83$ meV (Fig. 4-22), which are in good agreement with the experiment observation in Fig. 4-21. In the low resolution scan shown in Fig. 4-19, we are also able to assign all the observed excited states to the expected level spacings (Table 4-1). We note that in the scan shown in Fig. 4-19, not all the expected levels (such as $\Delta\epsilon_{30}$) showed up possibly due to the low resolution of the experiment data.

In conclusion at ultra low cryogenic temperature the GNR-FETs operate in the Coulomb blockade regime when gate voltages are small. The simulation based on the

many-electron picture is able to explain the experiment data qualitatively and quantitatively. The simulation tells the origins of the excited state lines in the experiment and hence makes the fitting to the single particle levels possible. We find that the GNR $E-k$ with longitudinal confinement can produce the single particle levels of which the level spacings agrees well with the experiment. It suggests that the GNRs of experiment devices are likely of high quality and the transport properties are dominated by the intrinsic electronic structure of GNR and contacts rather than the impurities and defects.

Table 4-1. The energy spacings from experiments and computed by the $E-k$ relation in Eq. 4-14

Level spacings (meV)	$\Delta\epsilon_{10}$	$\Delta\epsilon_{20}$	$\Delta\epsilon_{30}$	$\Delta\epsilon_{21}$	$\Delta\epsilon_{31}$	$\Delta\epsilon_{32}$	$\Delta\epsilon_{42}$
Experiment (from excited states)	2.56, 3.6	9.12, 10.08	20.16, 19.44, 19.8	5.76,	16.32, 16.8	9.6, 10.56, 11.52, 12.96	20.62, 19.68
Experiment (from E_{add} of even-odd diamonds)	1.68	6.24	17.28	4.56	15.6	11.04	N/A
Calculation using the E-k model ($E_g=60\text{meV}$, $L=140\text{nm}$)	2.64	9.51	18.86	6.87	16.22	9.35	20.02
Simulation to fit the experiment ($E_g=60\text{meV}$, $L=135\text{nm}$, $C_g=0.87\text{aF}$)	2.83	10.15	20.02	7.32	17.19	9.87	19.89

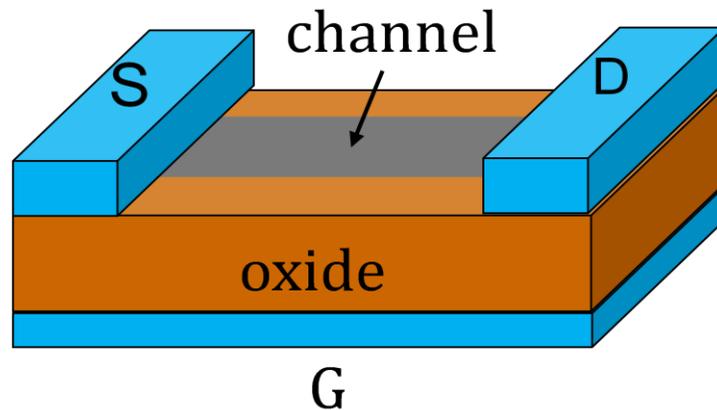


Figure 4-1. Schematics of the experiment device structures. The channel size (length times width) is $86 \times 14 \text{ nm}^2$ for device 1 and $140 \times 17 \text{ nm}^2$ for device 2. The metal source/drain 20 nm high and 1 μm wide (in simulation we use a width of 400 nm). The oxide (SiO_2) thickness is 300 nm.

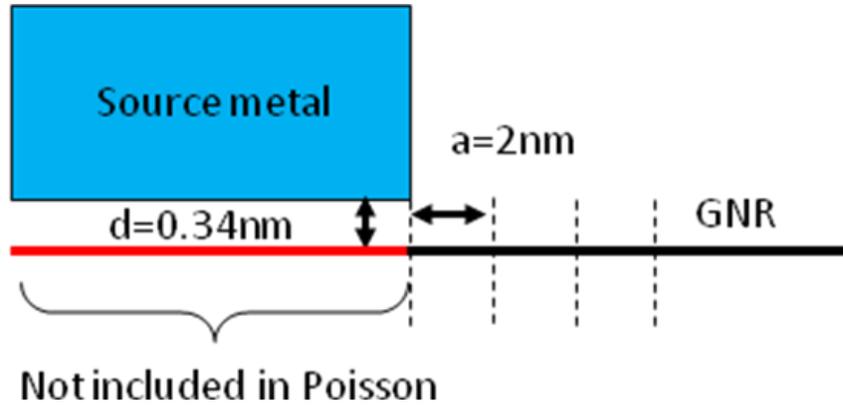


Figure 4-2. Side-view of the source contact with the section plane normal to the width direction. The GNR under the metal (lead) is not included in the Poisson solver because the small distance between the GNR and metal if properly treated will impose too large computational burden. Due to the large grid size (2 nm) the potential at end node is not a good approximation to the potential in the lead.

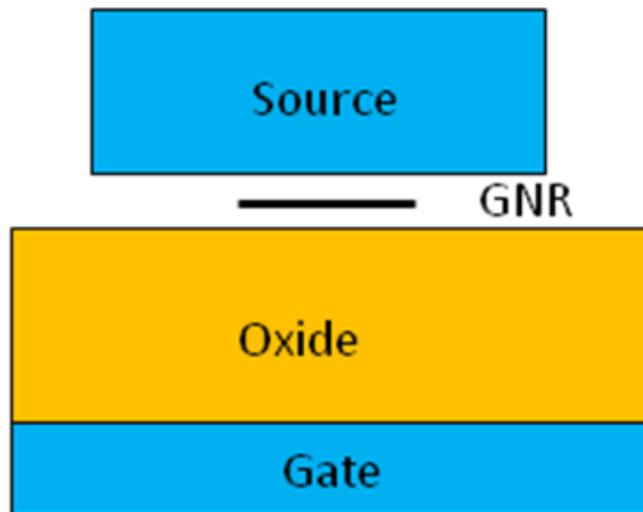


Figure 4-3. Schematic for 2D Poisson problem. The plane is normal to the GNR axis.

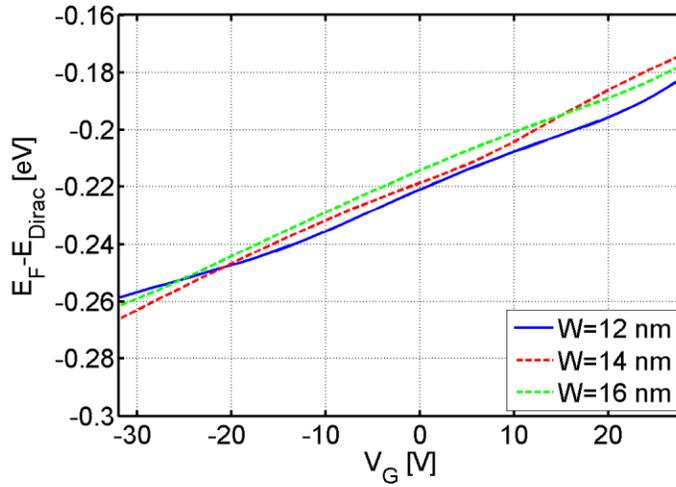


Figure 4-4. Energy difference between the Fermi level and the Dirac point inside the lead. The SB height for holes is equal to the value plus half band gap. It is calculated by a self-consistent 2D Poisson solver. The work function difference between the channel and metal is 0.46 eV.

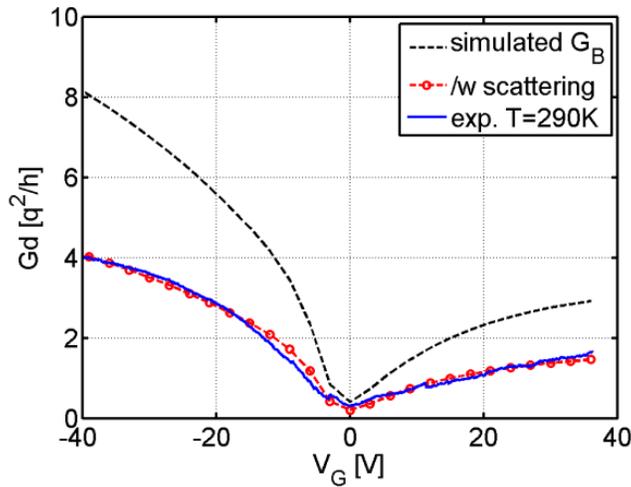


Figure 4-5. G_d -vs- V_g at $T=290K$ for device 1 ($L=86$ nm). To fit experiment data (solid curve) the simulated ballistic conductance (dashed curve) is scaled by a factor of $\frac{1}{2}$ to produce the best fitting (circles). The SB height inside GNR leads is varied self-consistently as gate voltage changes.

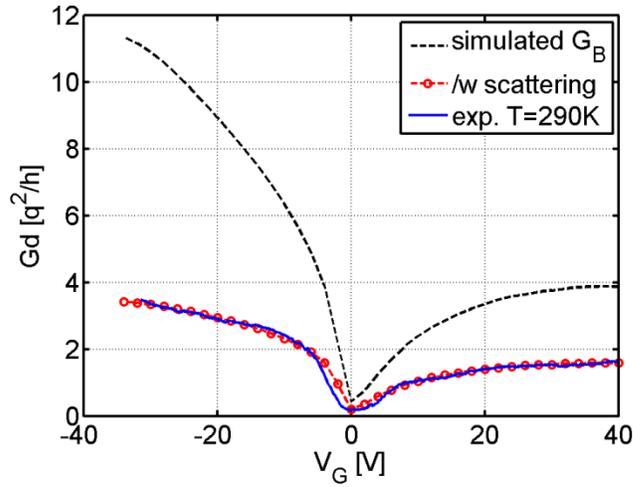


Figure 4-6. G_d -vs- V_g at $T=290\text{K}$ for $L=140\text{ nm}$ device (device 2). To fit experiment data (solid curve) the simulated ballistic conductance (dashed curve) is first scaled by a factor of $\frac{1}{2}$. Then a parasitic resistance of $3\text{ k}\Omega$ is included to produce the best fitting (circles). The SB height inside GNR leads is varied self-consistently as gate voltage changes.

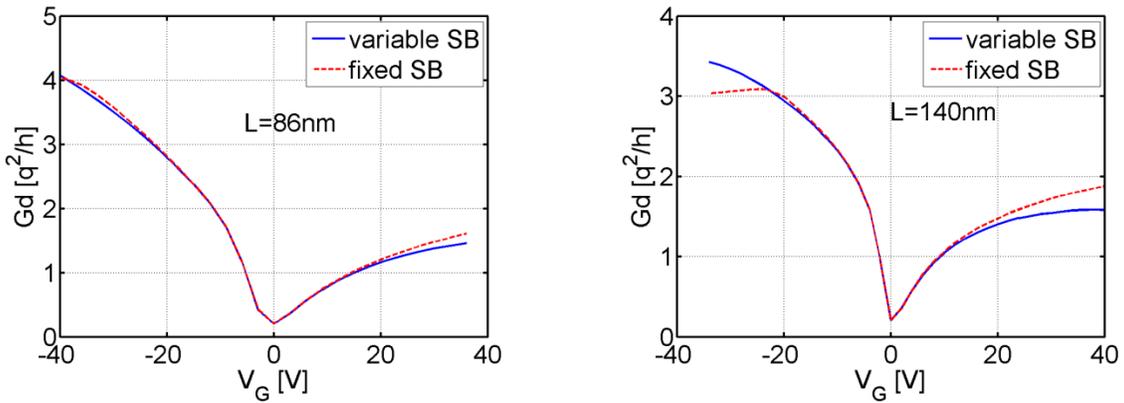


Figure 4-7. G_d -vs- V_g simulated with variable SB heights (solid curves) and fixed SB (dashed curves). A) Device 1 and B) Device 2. For device 1 the difference between two curves is small because the channel is the bottleneck of transport. For device 2 the difference is large at high gate biases as the leads under metal contacts become the bottleneck.

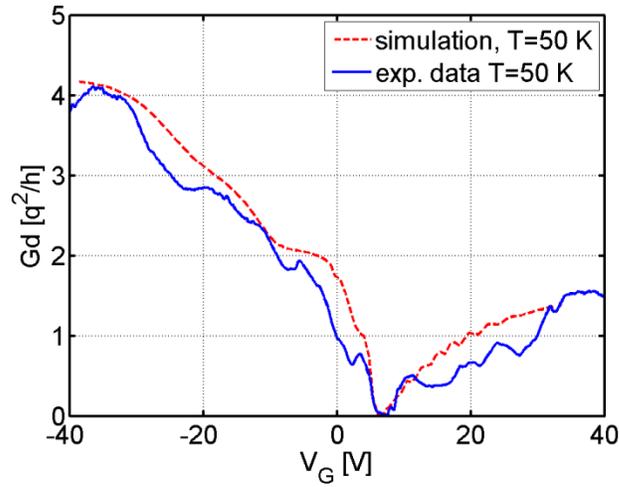


Figure 4-8. G_d -vs- V_g at $T=50\text{K}$ for device 1 by scaling transmission. Simulated curve is obtained by scaling the ballistic conductance with a factor of 0.5 as used in Fig. 4-5 and Fig. 4-6. The simulated curve is smoother than the experiment data.

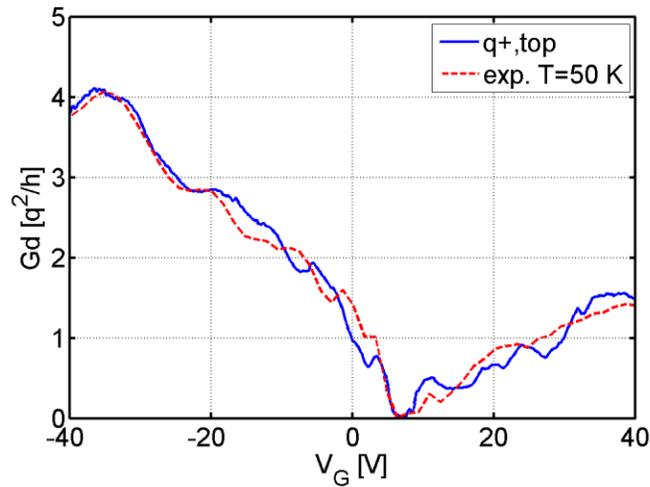


Figure 4-9. G_d -vs- V_g at $T=50\text{K}$ for device 1 with charge impurity and transmission scaling. Simulated curve is obtained by including charge impurity scattering and a conductance scaling factor of 0.55. It shows that the conductance degradation due to charge impurity scattering alone is only about 10%. The configuration of the charge impurities used in the simulation is shown in Fig. 4-10.

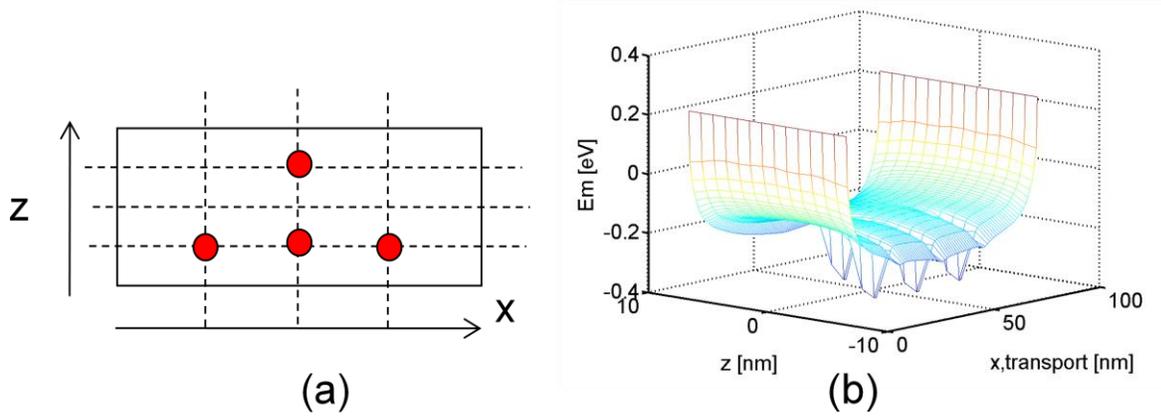


Figure 4-10. Charge impurity configuration used to generate the curve in Fig. 4-9. (a) The impurity positions projected to the channel plane and (b) 2D potential profile on the channel due to impurities. The dashed lines partition the channel length and width into grids of even sizes. Each impurity has a charge of $+1q$ and is 1 nm above the channel plane.

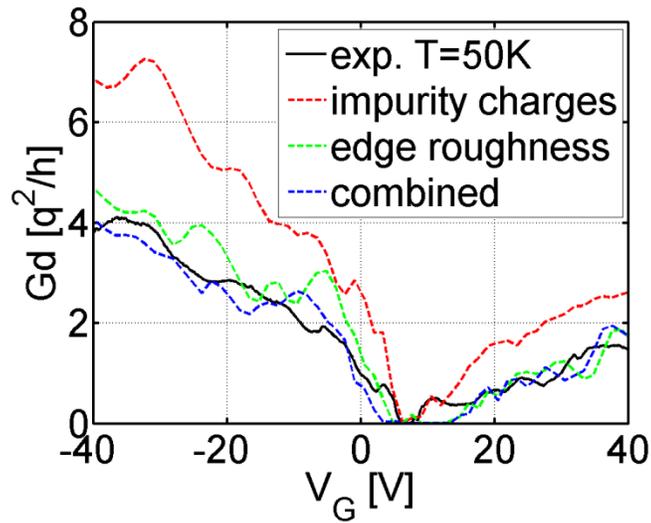


Figure 4-11. G_d -vs- V_g at $T=50K$ for device 1 including various scattering mechanisms. The percentage of edge roughness is 3%. Now scaling factor is needed after including structural defects, i.e. edge roughness.

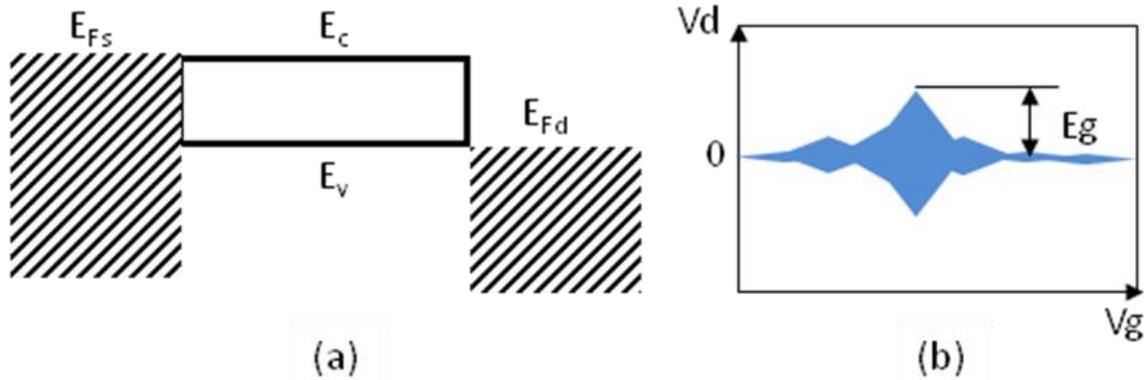


Figure 4-12. Schematics showing transport gap measurement at cryogenic temperature (a) A schematic of flat bands with the source/drain Fermi levels aligned to band edges and (b) corresponding 2D differential conductance plot. The show area indicates low conductance. The size of the largest pattern gives the bandgap

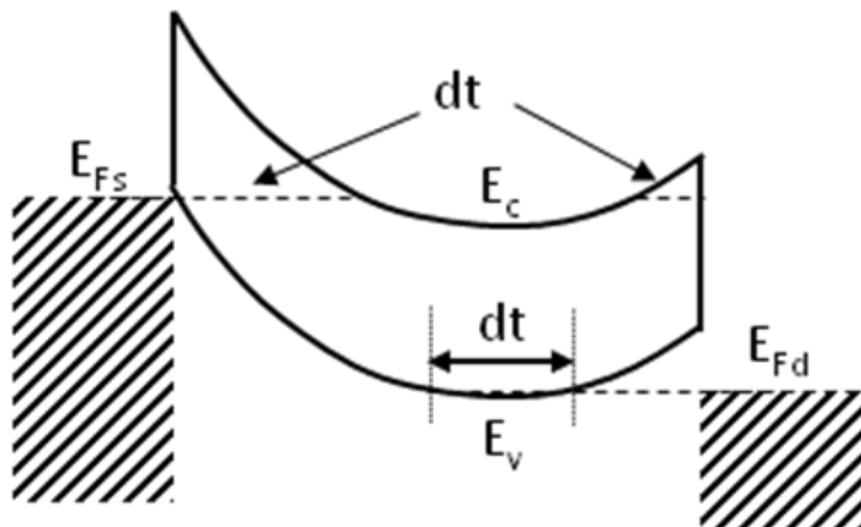


Figure 4-13. Schematic shows the bands bending causes uncertainty in bandgap measurement in Fig. 4-12. The tunneling distance dt continuously changes across the band edges. The transition from low to high conductance is thus blurred which results in large uncertainty in band gap estimation.

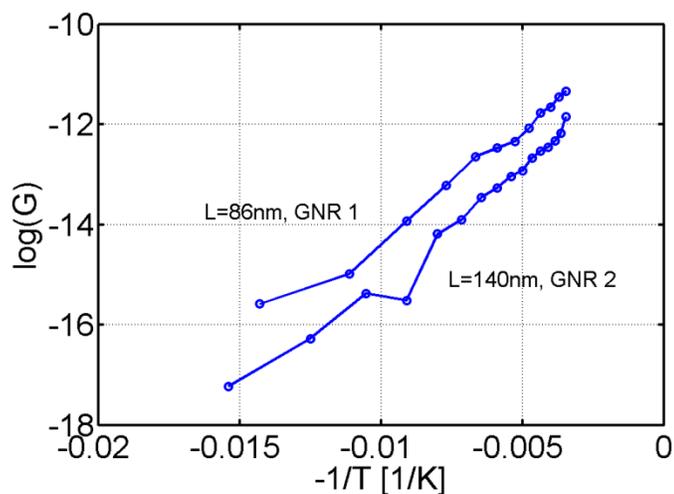


Figure 4-14. Experiment $\ln(G_{\text{off}}(T))$ vs $-1/T$. Clearly the experiment data do not form straight lines.

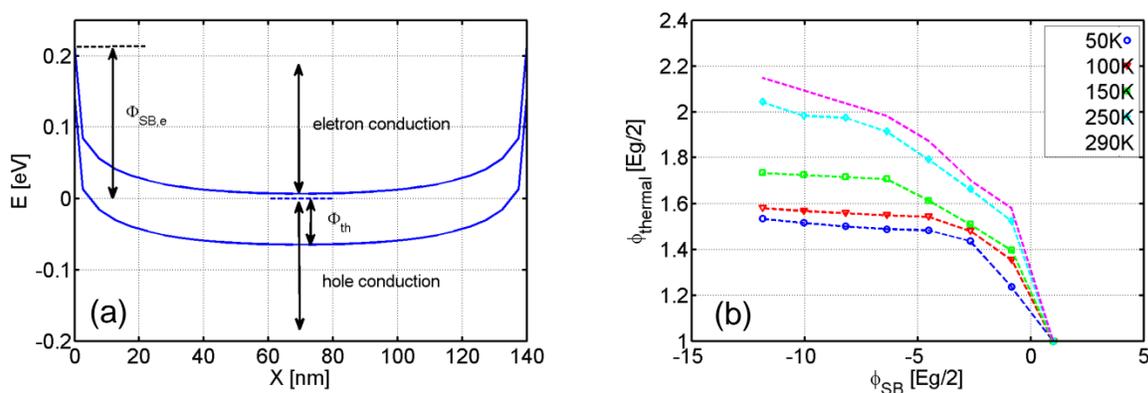


Figure 4-15. Relation between the SB height and thermal barrier height at off-state. (a) A self-consistent band diagram at showing the definitions of electron SB height and thermal barrier height. (b) Thermal barrier height as a function of SB height at different temperatures at off-state. It shows only for mid-gap devices ($\Phi_{\text{SB}} = E_g/2$) the thermal barrier height is equal to half of the bandgap and $I_{\text{off}}(T) \propto \exp(-E_g/2k_B T)$ can hold.

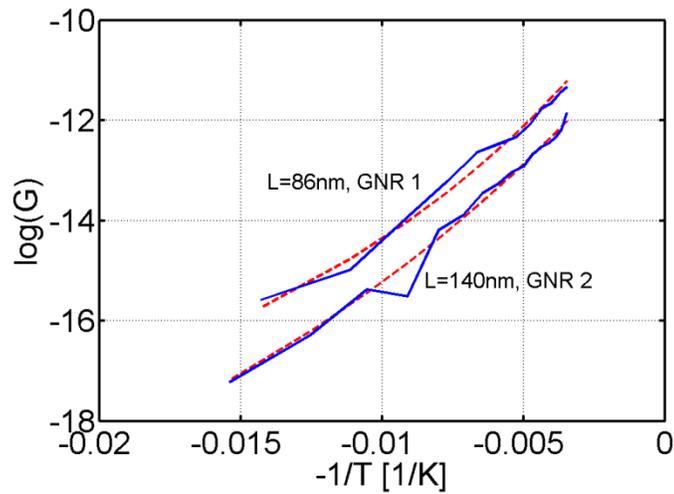


Figure 4-16. Log scale conductance as a function of temperature in the range of 70K to 290K. The dashed curves are simulated by NEGF. The solid curves are experiment data.

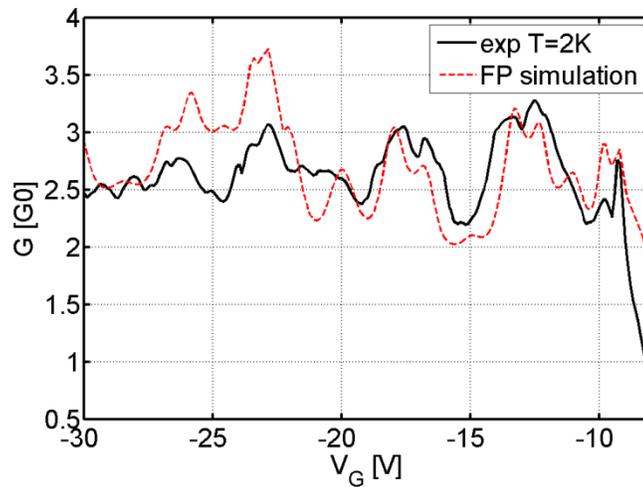


Figure 4-17. Conductance as a function of gate voltage at zero source-drain bias for device 1 at $T=2K$. Dashed curve is simulated using scattering matrix method. Solid curve is experiment data. Subband-dependent phases, and transmissions and reflection coefficients are used in the simulation. The gate electrostatics is treated by a capacitor model with the equilibrium GNR DOS.

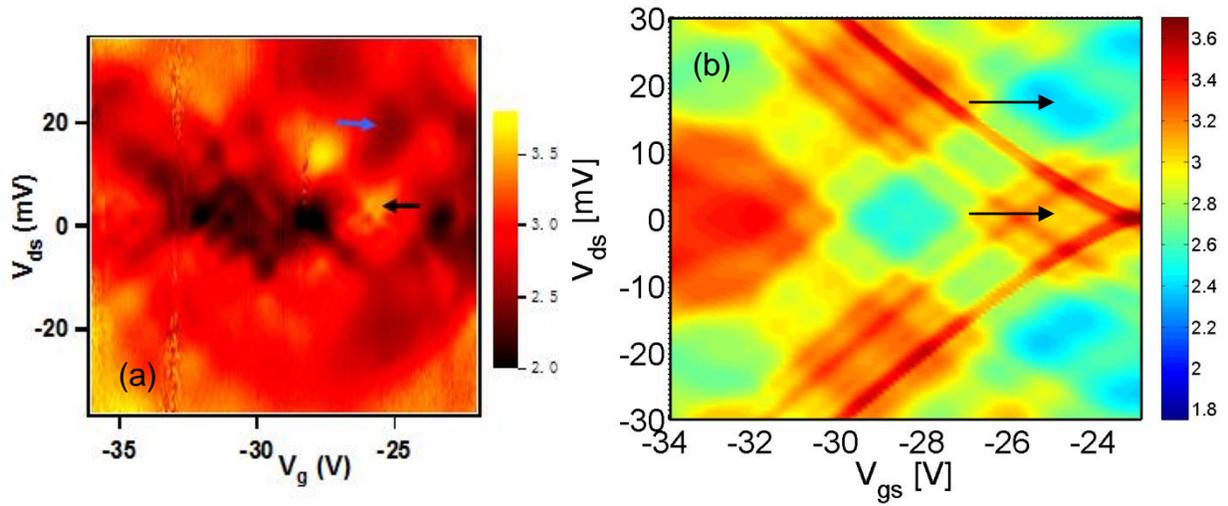


Figure 4-18. 2D differential conductance plot for device 1 at $T=2\text{K}$. (a) Experiment data (b) Simulated conductance by scattering matrix method.

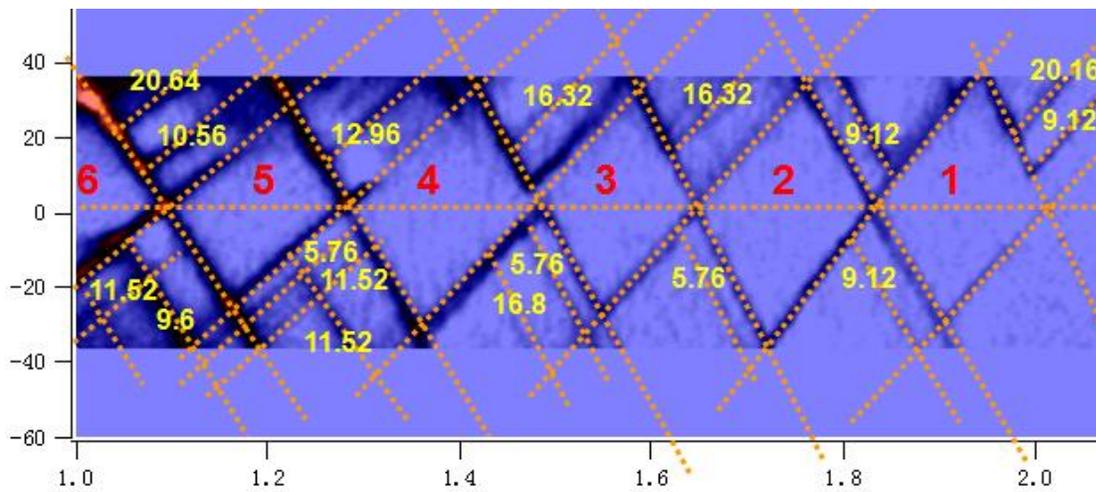


Figure 4-19. Experiment 2D differential conductance at $T=2\text{K}$ and small gate voltages. The diamond patterns and excited state lines are signatures of Coulomb blockade.

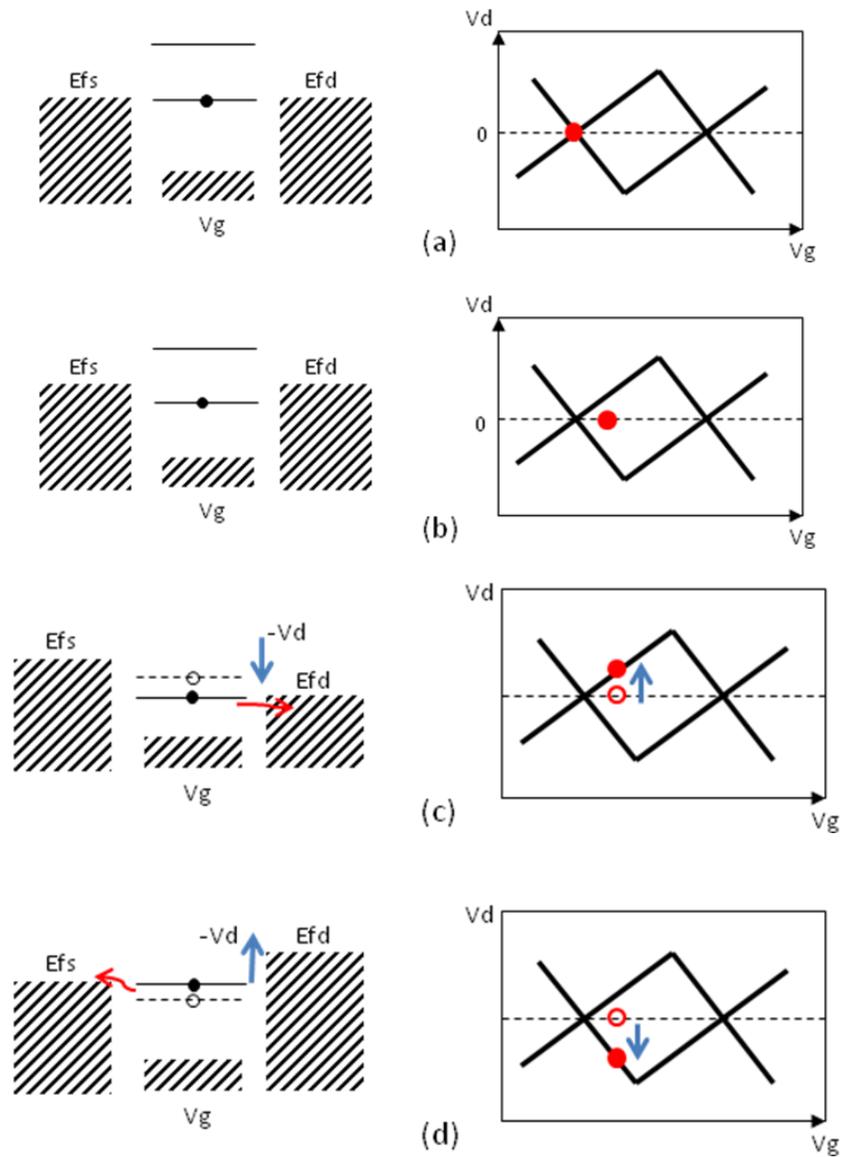


Figure 4-20. Energy level diagrams for single electron charging process (a) Energy level diagram corresponding to the circled conductance peak at $V_d=0$ axis. The dark lines present high conductance in 2D plot. The lowest single particle level is aligned to both Fermi levels, which leads to a conductance peak. (b) At $V_d=0$ as V_g increases the conductance decreases to zero (c) Increasing V_d lowers the single particle level to the drain Fermi level, leading to a conductance peak. (d) Decreasing V_d increases the single particle level to the source Fermi level Energy level and a conductance peak occurs.

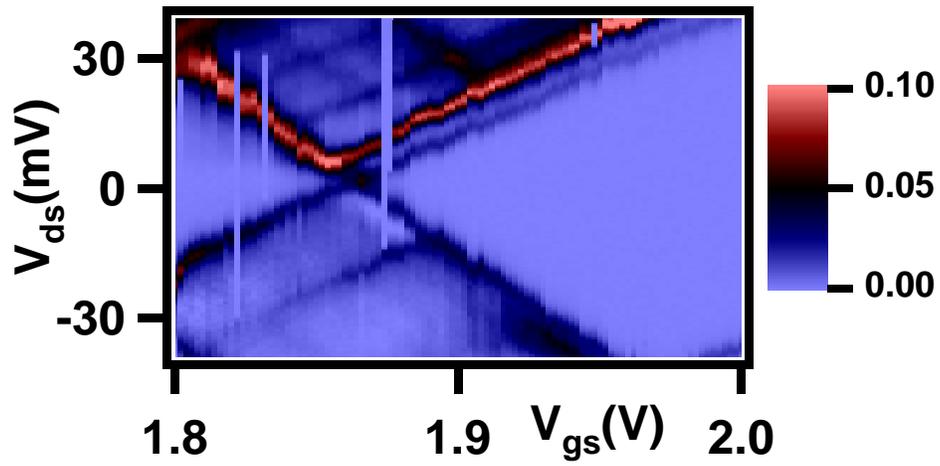


Figure 4-21. Experiment data: high resolution scan of conductance plot for a small range of biases

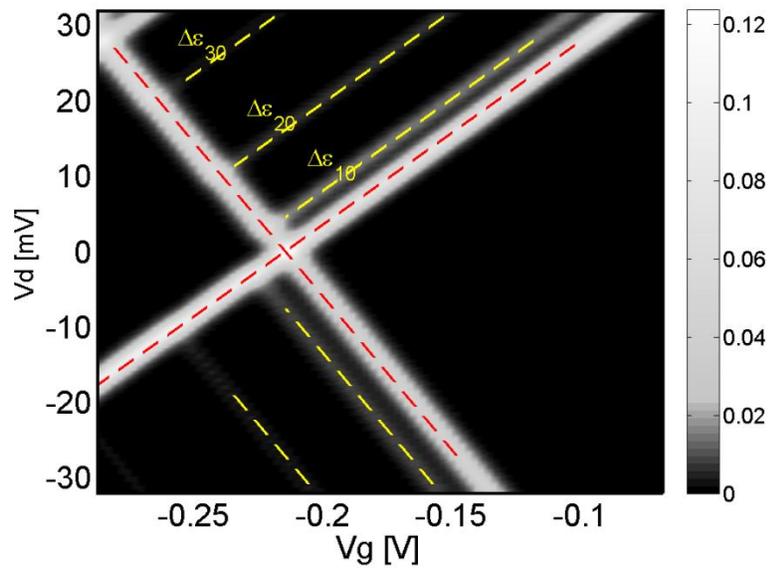


Figure 4-22. Simulated 2D differential conductance plot for Fig. 4-21.

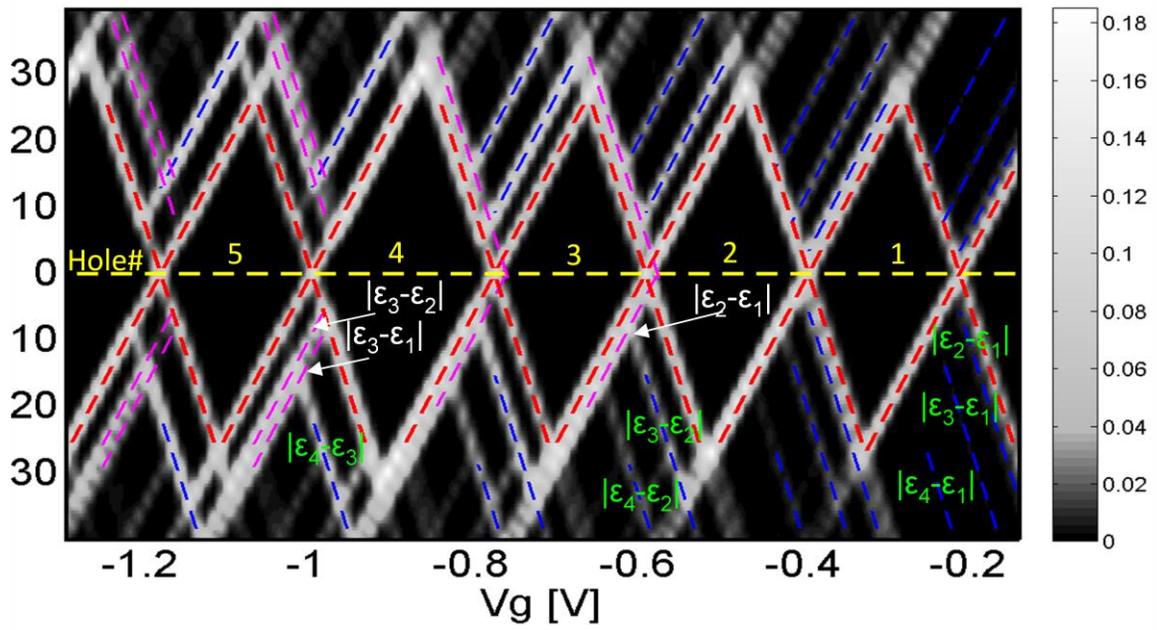


Figure 4-23. Simulated 2D differential conductance plot for Fig. 4-19

CHAPTER 5
SCALING BEHAVIORS OF GRAPHENE NANORIBBON FETS: A THREE
DIMENSIONAL QUANTUM SIMULATION STUDY

Graphene Nanoribbons for Transistor Applications

The continuous scaling of silicon transistors has been the driving engine for the exponential growth of digital information processing systems over the last decades. The Si transistor in production today is below the 100 nm scale and has entered the nanoelectronics regime. With the scaling limit of Si FETs in sight, a large group of emerging research devices are being extensively studied [41]. Among them, carbon-based nanostructure FETs are the forerunners due to their exceptionally excellent carrier transport properties [64]. Carbon nanotube (CNT) FETs with excellent device performance have been demonstrated [65]. With promising progress on fabricating and patterning a graphene layer recently, graphene electronics has been a topic of strong research interests [2-4]. A narrow strip of graphene, graphene nanoribbon (GNR), can be either metallic or semiconducting, depending on its structure [38]. The semiconducting GNR is being intensively explored for potential transistor applications. The channel geometry of a GNFET can be defined by lithography, which offers potentially better control over a CNTFET. The concept of all graphene circuits, in which GNFETs are connected by metallic GNR interconnects, has been proposed [4], [28]. An exceptionally high mobility ($\sim 10,000 \text{ cm}^2/\text{Vs}$) of graphene and GNRs has been demonstrated experimentally [2] and theoretically [66], which leads to the promise of near ballistic transport in a nanoscale GNFET. A recent theoretical study assessed the performance limits of GNFETs, but it is based on a semiclassical transport model coupled to a simple treatment of self-consistent electrostatics [67]. Quantum effects and

electrostatic short channel effects are not treated, which makes it difficult to explore scaling behaviors and ultimate scaling limits of GNRFETs.

In this work, a comprehensive study on the scaling behaviors of GNRFETs is performed by solving an atomistic quantum transport equation based on the non-equilibrium Green's function (NEGF) formalism self-consistently with a three-dimensional (3D) Poisson equation. The dependence of the I - V characteristics, transconductance, subthreshold swing, DIBL on the channel length are studied and compared for the single gate (SG), double gate (DG), and wrapped around gate (WG) geometries. The scaling characteristics of the gate insulator thickness and dielectric constant are explored. The roles of the contact size and Schottky barrier height are examined. The intrinsic delay of the GNRFET is simulated and compared to that projected for the Si FETs at the end of the roadmap.

Transistors with different device structures can operate in different ways. For a conventional MOSFET with heavily doped source and drain extensions, the gate modulates the channel conductance. If the heavily doped semiconductor source and drain are replaced by metal source and drain, Schottky barriers (SBs) form between the contacts and the channel and a SBFET is obtained. The transistor behavior above the threshold is achieved by modulating the tunneling current through the SBs at the two ends of the channel. For CNTs and GNRs, the fabrication process of the SBFET structure is simpler because it does not require doping of a monolayer of carbon atoms. Many CNTFETs demonstrated to date operate as a SBFET [68]. In this study, we focus our attention on GNR SBFETs with metal source and drain contacts. GNR MOSFETs with doped source and drain extensions will be the topic of a future study.

Simulation Approach

We simulated GNRFETs with three different gate geometries at room temperature ($T=300$ K) to explore the effect of gate geometry on the performances of GNRFETs. Figure 5-1a shows SG GNRFET, which has the advantage of easy fabrication but is not optimized for good gate control and suppression of short channel effects. Figure 5-1b is DG GNRFET, which sandwiches a graphene ribbon between two gates. Figure 5-1c shows the cross section of a WG GNRFET (in the plane normal to the channel direction), and the GNR is surrounded by the gate. The WG GNRFET is most challenging for fabrication, but it offers ideal gate control. The nominal device parameters are as follows. The SiO_2 gate oxide thickness is $t_{\text{ox}}=2$ nm and the relative dielectric constant is $\epsilon_r=4$. The GNR channel has armchair edges, as shown in Fig. 5-2. The ribbon index N denotes the number of carbon atom dimer lines, following the definition in Ref. [38]. The $N=12$ armchair edge GNR channel has a width of $\sim 13.5\text{\AA}$, which results in a bandgap of $E_g \approx 0.83$ eV. The channel length is $L_{\text{ch}}=20$ nm. The metal source/drain is directly attached to the GNR channel and the Schottky barrier height between the source/drain and the channel is a half of the GNR band gap, $\Phi_{Bn} = \Phi_{Bp} = E_g / 2$. The flat-band voltage is zero. A power supply voltage of $V_{\text{DD}}=0.5$ V is used. The nominal device parameters are varied to explore different scaling issues.

The DC characteristics of ballistic GNRFETs are simulated by solving the Schrödinger equation using the non-equilibrium Green's function (NEGF) formalism self-consistently with a 3D Poisson equation [56]. Ballistic transport is assumed. A tight binding Hamiltonian with a p_z orbital basis set is used to describe an atomistic physical observation of the GNR channel. One p_z orbital is enough for the atomistic physical

description since s , p_x , and p_y levels are far from the Fermi level and do not play important roles for carrier transport. A p_z orbital coupling parameter of 3 eV is used and only the nearest neighbor coupling is considered. The retarded Green's function of the device is computed as,

$$G(E) = [(E + i0^+)i - H - U - \Sigma_1 - \Sigma_2]^{-1}, \quad (5-1)$$

where H is the tight binding Hamiltonian matrix of the GNR channel, U is the self-consistent potential matrix determined by the solution of the Poisson equation, and Σ_1 and Σ_2 are the self-energies of the metal source and drain contacts, as shown in Fig. 5-2. The charge density can be computed as,

$$Q_i(x) = (-q) \int_{-\infty}^{+\infty} dE \cdot \text{sgn}[E - E_N(x)] \{ D_1(E, x) f(\text{sgn}[E - E_N(x)](E - E_{F1})) + D_2(E, x) f(\text{sgn}[E - E_N(x)](E - E_{F2})) \} \quad (5-2)$$

where q is the electron charge, $\text{sgn}(E)$ is the sign function, $E_{F1, F2}$ is the source (drain) Fermi level, and $D_{1,2}(E, x)$ is the local density of states due to the source (drain) contact, which is computed by the NEGF method. The charge neutrality level [69], $E_N(x)$, is at the middle of band gap because the conduction band and the valence band of the GNR are symmetric.

The self-consistent potential is computed from the charge density and the electrode potentials using the Poisson equation,

$$\nabla[\varepsilon \nabla U(\vec{r})] = qQ(\vec{r}) \quad (5-3)$$

where $U(\vec{r})$ is the electron potential energy which determines the diagonal entry of the potential energy matrix in Eq. 5-1, ε is the dielectric constant, and $Q(\vec{r})$ is the charge density. Because the electric field varies in all dimensions for the simulated device

structure, a 3D Poisson equation is numerically solved using the finite element method (FEM). The FEM has the advantage to treat complex device geometries and boundaries between different dielectric materials.

The source-drain current is computed using the NEGF method once the self-consistency between the NEGF transport equation and the Poisson equation is achieved. The on-current, off-current, and minimal leakage current can be extracted from the simulated I - V characteristics. We compute the intrinsic delay as

$\tau = (Q_{on} - Q_{off}) / I_{on}$, where Q_{on} and Q_{off} are the total charge in the channel at on state and off state respectively, and I_{on} is the on-current [70]. The total charge in the channel is

calculated by $Q_{ch} = \int_0^{L_{ch}} Q_i(x) dx$, where $Q_i(x)$ is the charge density as a function of

channel position. The off state is chosen as the state at which a GNR-FET delivers the minimum source-drain current, $V_G = V_D/2$, as discussed later. It can be shifted to $V_G = 0$ if a proper gate work function (and therefore, the threshold voltage) could be achieved.

The calculation of the intrinsic delay assumes zero parasitic capacitance. The parasitic capacitance between electrodes impairs the performance of nanoscale transistors in practice. The intrinsic delay, therefore, indicates the upper limit of the switching speed.

Simulation Results

We first compare the switching on and switching off characteristics of GNR-FETs with different gate geometries. The switching on characteristics are described by simulating the transconductance, and the switching off performance and the immunity of the short channel effects are studied by simulating the subthreshold swing and DIBL in Fig. 5-4. Figure 5-3 plots the transconductance, $g_m = \partial I_D / \partial V_G$, which is the ratio of current variation to the gate voltage variation at on state, as a function of the channel

length for three types of gate geometries with nominal device parameters. The following observations were made. First, the transconductance remains approximately constant for $L_{ch} > 10$ nm. The reason is that the current does not depend on the channel length in the ballistic transport regime, which is different from the diffusive transport regime. Second, the transconductances of all GNR-FETs decrease for $L_{ch} < 10$ nm because the gate has a worse control of the channel due to electrostatic short channel effects. Third, the transconductances of DG and WG GNR-FET are 35% and 55% larger than that of SG GNR-FET at $L_{ch} = 20$ nm. Using a DG structure increases the transconductance but does not double transconductance. In contrast, DG Si FETs have been actively studied for the promise of the on-current advantage over their SG counterpart. The reason is that for a transistor operating at the conventional MOSFET limit (in which the semiconductor capacitance is much larger than the gate insulator capacitance), a double gate structure leads to a two times larger gate insulator capacitance, and therefore, two times larger total gate capacitance and transconductance. For a transistor operating at the quantum capacitance limit (in which the gate insulator capacitance is much larger than the semiconductor capacitance of the channel) [26], [71], the total gate capacitance is limited by the semiconductor capacitance and is independent of the gate insulator capacitance. A double gate geometry does not lead to an improvement of the total gate capacitance and the transconductance. The GNR-FET has a one-dimensional channel with a monolayer of carbon atoms. Its small semiconductor capacitance (due to a low density of states) makes it operate closer to the quantum capacitance limit than Si FETs. The advantage of using multiple gate structures, therefore, is smaller.

Figure 5-4a shows the variation of subthreshold swing with different channel length. The subthreshold swing, S , is calculated by $\Delta V_G / \log(I_D(V_G + \Delta V_G) / I_D(V_G))$ at subthreshold region. With 10 nm GNR channel, the gate electrostatic control is not sufficiently good to keep the subthreshold swing and DIBL approximately the same as the longer channel devices. The subthreshold swing, however, remains approximately constant beyond $L_{ch}=15$ nm for all gate geometries. The gate geometry dependence of the subthreshold swing for the nominal device with $L_{ch}=20$ nm is not strong, and the variation of S is less than 15%. The advantage of multiple gate geometry for the immunity to the short channel effects, however, is more obvious at a shorter channel length. If we choose $S=100$ mV/decade as the criterion (the dotted line in Fig. 5-4a), the subthreshold swing of SG GNRFET at $L_{ch}=10$ nm does not meet it. The criterion, however, can be met by using a DG geometry ($S \sim 94$ mV/decade) or a WG geometry (~ 87 mV/decade).

Fig. 5-4b shows DIBL vs. channel length. DIBL is a feature of short channel effects and can quantitatively be expressed as $\Delta V_t / \Delta V_D$, where V_t is the threshold voltage. At a channel length of 10 nm, the short channel effect is severe and drain voltage affects the barrier at the beginning of the channel a lot. When channel length increases, DIBL, however, decreases drastically and it remains approximately constant beyond $L_{ch}=20$ nm. For the nominal device with $L_{ch}=20$ nm, DIBL has nearly no difference among all three gate geometries. If we specify 100 mV/V as the criterion (the dotted line in Fig. 5-4b), at $L_{ch}=10$ nm, DIBL of SG GNRFET is much larger than the criterion, whereas that of DG and WG GNRFET is less than the criterion, which means GNRFETs with multiple

gate geometries can extend the channel length scaling down more than SG GNR-FET. The observation is consistent with that of Fig. 5-4a.

We next study the effect of power supply voltages. The I_D vs. V_G characteristics for the SG nominal device is plotted in Fig. 5-5a. As the Schottky barrier height is half of the bandgap, the minimum currents occur approximately at a gate voltage of $V_{G,min}=1/2V_D$, at which the conduction band bending at the source end of the channel is symmetric to the valence band bending at the drain end of the channel, and the electron current is equal to the hole current. The condition of achieving the minimal leakage current is the same as CNT SBFET with middle gap SBs [72]. Increasing the drain voltage leads to an exponential increase of the minimal leakage current, which indicates the importance of properly designing the power supply voltage for achieving sufficiently small leakage current.

The minimal leakage bias point can be shifted to $V_G \approx 0$ in order to achieve a small off current (at $V_G=0$) by properly designing the transistor threshold voltage (e.g., by gate work function design). We, therefore, obtain the off state characteristics at $V_{G,min}=V_{DD}/2$ and the on state characteristics at $V_G=V_{G,min}+V_{DD}$ from the I - V simulated for GNR-FETs with a zero flat band voltage. The output characteristics in Fig. 5-5b show typical linear and saturation regimes. When V_G is increased, saturation current is increased due to a larger voltage drop between the gate and the source contact and a larger energy range for carrier injection from the source into the channel.

The effect of channel length scaling on SG GNR-FET I - V characteristic is explored in Fig. 5-6. At $V_D=0.5$ V, both the off-current and on-current are increased when channel length decreases to a small value. For a channel length of 5 nm, direct tunneling from

the source to drain leads to a large leakage current, and the gate voltage can hardly modulate the current. The transistor is too leaky to have appreciable difference between on and off states. For channel length of 10 nm, the on-off current ratio is greatly improved to near 600. Increasing the channel length to 15 nm, the on-off current ratio is further increased to about 3500 due to decreased tunneling current at off state. However, after channel length exceeds 15 nm, the increase in on-off current ratio is not significant. This is because the device is simulated at ballistic limit. (The state-of-the-art short channel CNTFETs already operate close to ballistic limit [65]) Further increasing the channel length hardly changes the on-current or off-current, nor does the on-off current ratio while for conventional MOSFET, increasing the channel length will cause the channel resistance to increase proportionally.

The I_D - V_D characteristics in Fig. 5-6b confirm the results in Fig. 5-6a. For 5 nm and 10 nm long channels, direct tunneling from the source to drain and electrostatic short channel effects are severe, so that no decent saturation of I_D - V_D characteristics is observed. The devices with extremely short channel operate more like a conductor rather than a transistor at $V_G=0.75$ V. For channel length of 15 to 25 nm, typical transistor I_D - V_D curves can be obtained and the difference in length results in only a little change in I_D - V_D curves. For DG and WG devices, the qualitative results are similar. Compared to single gate device, the channel length can be scaled to a little bit shorter before significant on-off current ratio is lost.

The scaling of channel width is achieved by changing the index N of the armchair edge GNR channel. Similar to the CNTs, there are semiconducting and metallic GNRs and only semiconducting ribbons are relevant here. In Fig. 5-7a, the I_D - V_G

characteristics for different channel widths are plotted. As the channel width is increased, both the off-current and on-current increase. A GNR with a larger channel width has a smaller bandgap [38], which results in a smaller Schottky barrier height in our simulation and consequently the current is increased. There is a trade-off between on-off current ratio and on-current. For an $N=9$ channel, the on-off current ratio is over one hundred thousand while the on-current is less than $1 \mu\text{A}$. For an $N=24$ armchair edge ribbon channel, although the on-current is $22 \mu\text{A}$, the on-off current ratio is only 11.8. But unlike the channel length scaling, where extremely short channel device with small on-off current ratio behaves like a conductor, the wide channel device with small on-off current ratio has I_D - V_D characteristics with clear linear regime and saturation regime as shown in Fig. 5-7b. This is because the channel length here is 20 nm, the gate, therefore, has good electrostatic control over the channel potential profile, and direct tunneling from source to drain in the band gap energy range is negligible.

The effect of gate oxide dielectric constant is explored. The subthreshold swing and transconductance versus gate oxide dielectric constant are shown in Fig. 5-8. The increase of gate oxide dielectric constant does not affect much the subthreshold swing and the transconductance. In this study, electrostatic potential in the channel region is carefully calculated by solving the Poisson equation with the Laplace solution combined with particular solution. The Laplace solution for the simulated devices is totally determined by the device geometry and boundary condition, whereas the particular solution is affected by the gate insulator capacitance with its thickness and dielectric constant. The effect of gate dielectric constant on total electrostatic potential is applicable only in the particular solution, but it does not affect much on total electrostatic

potential since the magnitude of the particular solution is very small compared to Laplace solution at the quantum capacitance limit [73]. The subthreshold swing and the transconductance, therefore, do not change a lot in the simulated channel length region.

The scaling of the gate oxide thickness is studied next. Fig. 5-9 shows the subthreshold swing and transconductance versus the gate oxide thickness. The increase of the gate oxide thickness results in worse gate electrostatic control, and hence larger subthreshold swing. A thinner oxide, therefore, is desirable for a larger on-current (due to a larger transconductance) and a larger maximum on-off current ratio (due to a smaller subthreshold swing). As the oxide thickness increases, the role of gate geometry is more important. For example, g_m of the WG GNR FET is about 25% larger than that of a SG GNR FET for $t_{ox}=1$ nm, but the improvement is 60% at $t_{ox}=2.5$ nm.

Figure 5-10 shows the I_G - V_G characteristics of GNR transistor with different Schottky barrier heights. As discussed above, at $V_D=0.5$ V, the minimal leakage current occurs at $V_G=0.25$ V for the Schottky barrier height equal to one half of the bandgap, and the I - V characteristic is ambipolar. Because the ambipolar characteristic increases the leakage current and is not preferred for CMOS applications, it is interesting to ask whether the ambipolar characteristic can be suppressed by designing the SB heights. For example, it could be expected that electron transport would be preferred when the SB height for electrons decreases, and the hole transport would be preferred if the SB height for electrons increases. The simulated I - V characteristics in Fig. 5-10, however, indicate that engineering the SB height does not change the qualitative ambipolar feature of the I - V characteristics when the gate oxide is thin. The reason is that the gate electrode can screen the field from the source and drain effectively for a thin gate oxide.

The Schottky barrier, whose thickness is approximately the same as the gate insulator thickness [74], is nearly transparent. Engineering the SB height, therefore, only has a small effect on the qualitative feature of the I_D - V_G characteristics. A similar phenomenon has been previously noticed and verified in CNT SBFETs.

The effect of the size of source and drain contacts on transistor I-V is studied in Fig. 5-11. In Fig. 5-11a, it shows both off-current and on-current decrease as the width of source/drain contact is increased from 1.4 nm to 4.8 nm while the on-off current ratio keeps approximately constant. To understand the I_D - V_G curves, the conduction band profiles at on state ($V_D=0.5$ V and $V_G=0.75$ V) in transport direction are plotted in Fig. 5-11b. Because the simulation shows little potential variation in the transverse direction of the channel, the potential is obtained by averaging the potentials of the atom sites in the channel width direction. It clearly shows the thickness of Schottky barrier at source end is decreased when the width of contacts is decreased. This is due to the fact that scaling down the width of the contacts can decrease the fringe field from gate to source contact and the penetration length of source electrical field is decreased. The Fig. 5-11c and Fig. 5-11d show how the height of contacts affects the transistor's I_D - V_G characteristics. It shows the same trend when the contact height is decreased. The DG and WG transistors basically follow the same qualitative conclusions. Compared to the SG FET, the contact size has smaller effect on the DG and WG GNFETs because the GNR channel is sandwiched or wrapped around by the gate, the field lines from the source and drain contacts are better screened by the gate electrodes, and the source and drain contact geometry has less impact. Generally, small size and low dimensional contacts not only improve the DC performance by increasing the on-current, but also

improve the AC performance by decreasing the parasitic capacitance between the contacts, which leads to higher operation speed. The intrinsic delay indicates how fast a transistor can switch. In Fig. 5-12, the intrinsic delays of a GNR FET and a Si MOSFET are compared. The intrinsic delay of the Si MOSFET is projected by ITRS to year 2020 and the channel length will be shrunk down to 14 nm [41]. The delay data ($C_g V / I$) in the ITRS roadmap are computed using the total gate capacitance C_g that includes the parasitic capacitance, as shown by the dashed line with circles in Fig. 5-12. In order to perform a fair comparison to the intrinsic delay of the nominal SG GNR FETs (which does not treat the parasitic capacitance), we also compute the intrinsic delay of Si MOSFETs by using the ideal gate capacitance values provided by the ITRS roadmap instead of the total gate capacitance values, as shown by the dashed line. The results indicate that the intrinsic delay of GNR FET is approximately a factor of 3 smaller than that of Si MOSFETs. The major reason is that the GNR has a larger band-structure limited velocity for the first subband, which contributes dominantly to the conduction in low voltage operation.

Comparison to Si MOSFET

A comprehensive study on the scaling characteristics of GNR SBFET is performed using self-consistent atomistic simulations. Because both GNR and CNT are one-dimensional nanostructures derived from graphene (with a bandgap created by quantum confinement in the width or circumferential direction), the scaling characteristics of GNR SBFETs show similarities with CNT SBFETs. Potentially better control of channel material properties than CNT FETs, however, could be achieved. The bandgap of the GNR channel strongly depends on its width, which significantly affects

the on-current and off-current. The ambipolar I - V characteristics cannot be suppressed by engineering the Schottky barrier height when the gate insulator is thin (below about 10 nm). Reducing the gate insulator thickness and the contact size results in thinner Schottky barriers, and therefore, larger on-current. The intrinsic speed of the GNR-FETs is several times faster than Si FETs due to its large carrier velocity and near ballistic transport, and reducing parasitics is essential for benefiting from the fast intrinsic speed. As a new type of transistor, GNR-FET, however, is still governed by transistor electrostatics and quantum effects, which imposes a similar ultimate scaling limit as for Si MOSFETs [75].

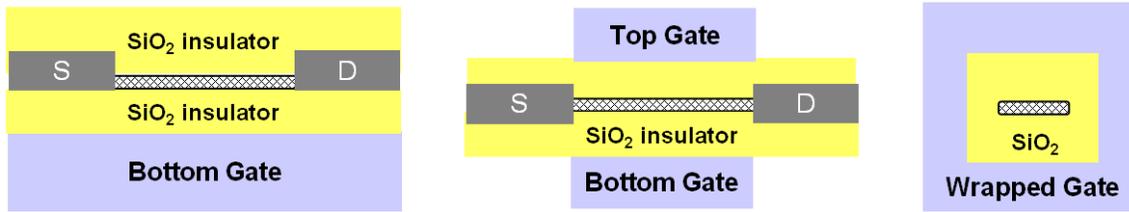


Figure 5-1. Simulated device structures with different gate geometries (a) single gate (SG), (b) double gate (DG), and (c) wrapped around gate (WG) graphene nanoribbon (GNR) FET. The channel materials are GNRs. Source and drain contacts are metals and the Schottky barrier height is a half band gap of the channel material.

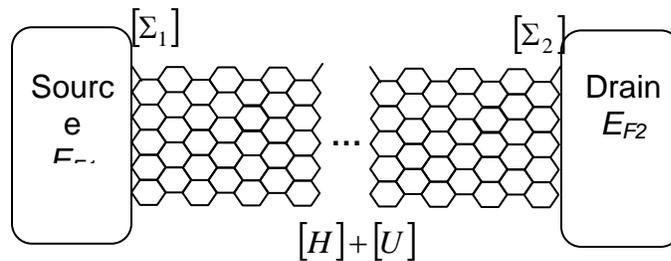


Figure 5-2. The schematic sketch of an armchair edge $N=12$ GNR channel with the source and drain contacts. The quantities used in the NEGF formalism are also shown.

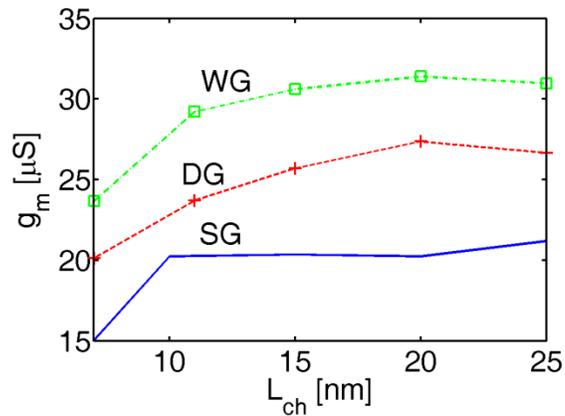


Figure 5-3. The gate transconductance vs. channel length. Solid line is for SG, dashed line with crosses is for DG, and dash-dot line with squares is for WG GNRFET. A better gate control results in larger transconductance.

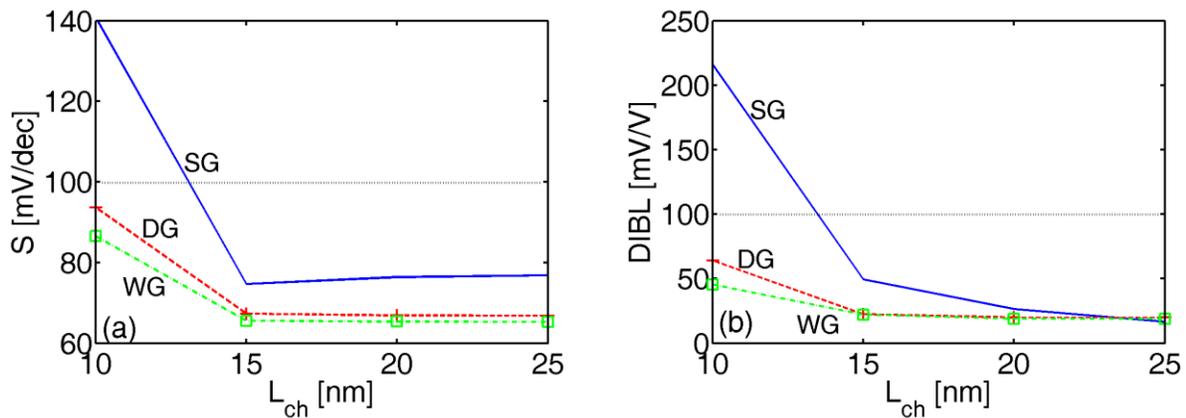


Figure 5-4. Channel length dependence of (a) S and (b) DIBL. Solid line is for SG, dashed line with crosses is for DG, and dash-dot line with squares is for WG GNRFET. The multiple gate geometries can extend the scaling down of channel length below 10 nm.

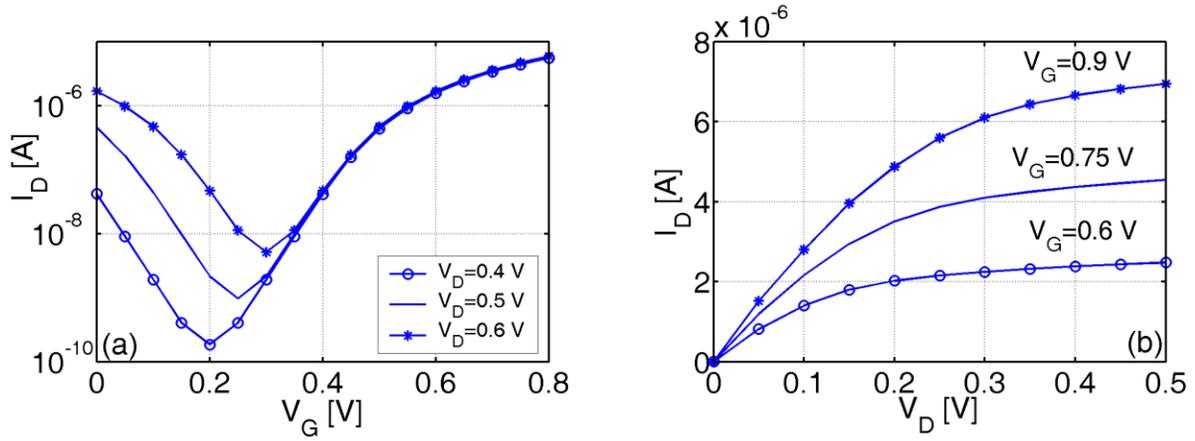


Figure 5-5. I - V characteristics of the single gate GNFET (a) The I_D vs. V_G characteristics for nominal single gate device at different source-drain voltages. (b) The I_D vs. V_D characteristics for nominal single gate device at different gate voltages.

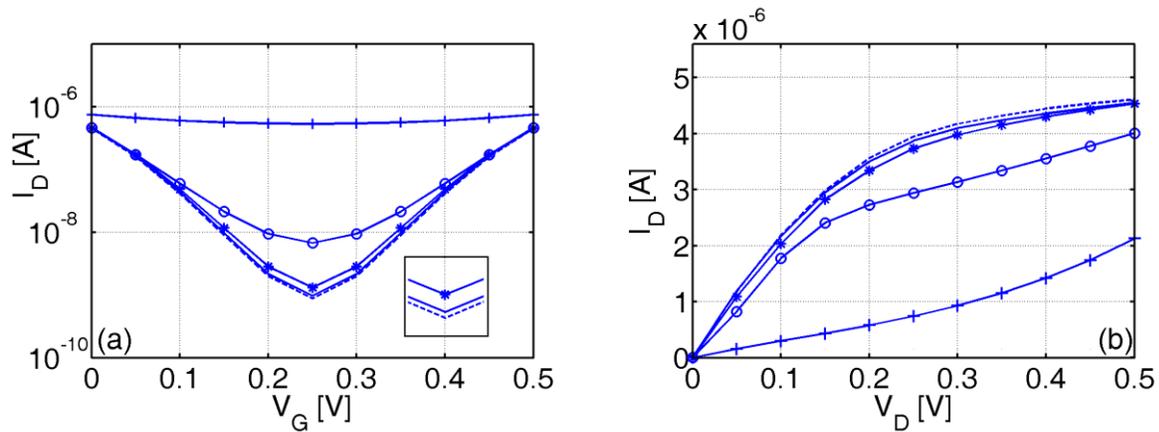


Figure 5-6. I - V characteristics of GNFETs with different channel lengths (a) The I_D vs. V_G characteristics for single gate device with different channel lengths at $V_D=0.5$ V. (b) The I_D vs. V_D characteristics for single gate device with different channel lengths at $V_G=0.75$ V (The curve with crosses is for $L_{ch}=5$ nm; the curve with circles is for $L_{ch}=10$ nm; the curve with stars is for $L_{ch}=15$ nm; the curve without marks is for $L_{ch}=20$ nm; the dashed one is for $L_{ch}=25$ nm).

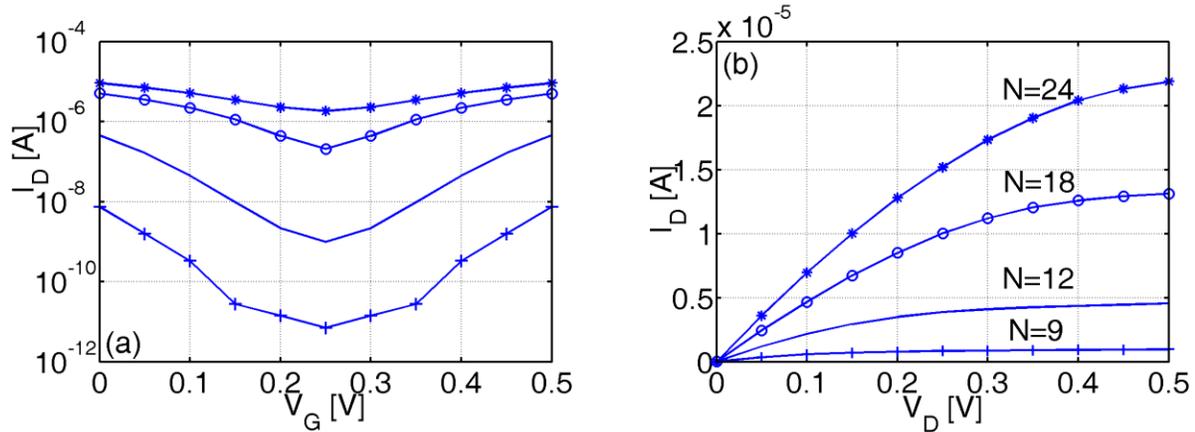


Figure 5-7. I - V characteristics of GNRFETs with different channel widths (a) The I_D vs. V_G characteristics for single gate device with different channel widths at $V_D=0.5$ V. (b) The I_D vs. V_D characteristics for single gate device with different channel widths at $V_G=0.75$ V. (The curves with crosses, circles, stars and without marks are for $N=9$, $N=18$, $N=24$ and $N=12$ armchair edge GNR respectively. The channel length is 20 nm in all cases).

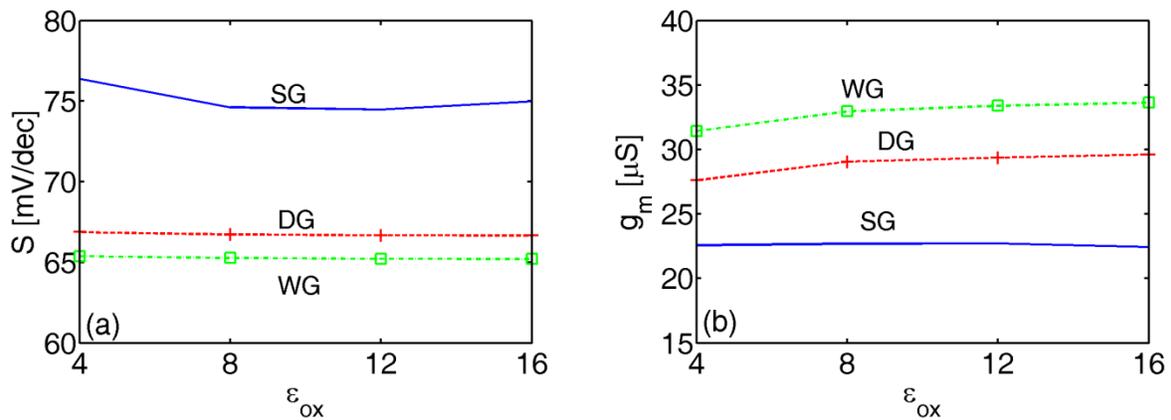


Figure 5-8. Gate oxide dielectric constant dependence of (a) S and (b) g_m . Solid line is for SG, dashed line with crosses is for DG, and dash-dot line with squares is for WG GNRFET.

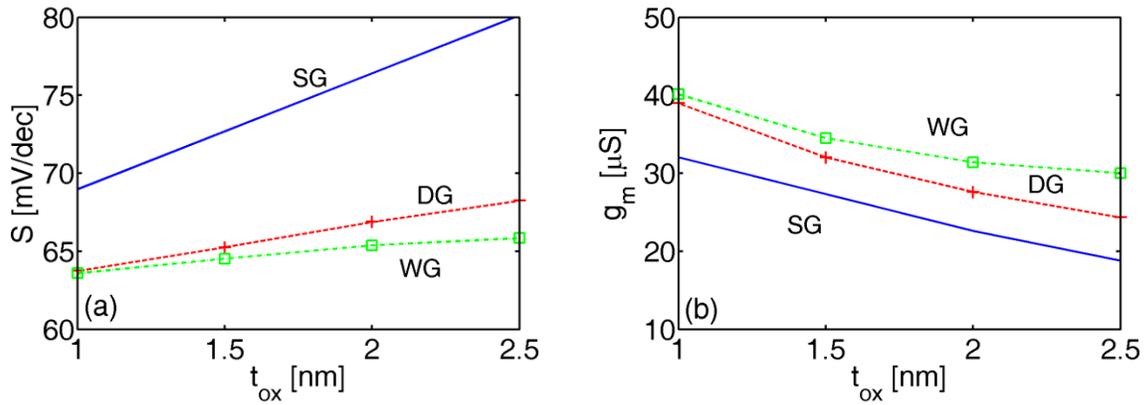


Figure 5-9. Gate oxide thickness dependence of (a) S and (b) g_m . Solid line is for SG, dashed line with crosses is for DG, and dash-dot line with squares is for WG GNRFET. A better gate control results in smaller subthreshold swing and larger transconductance.

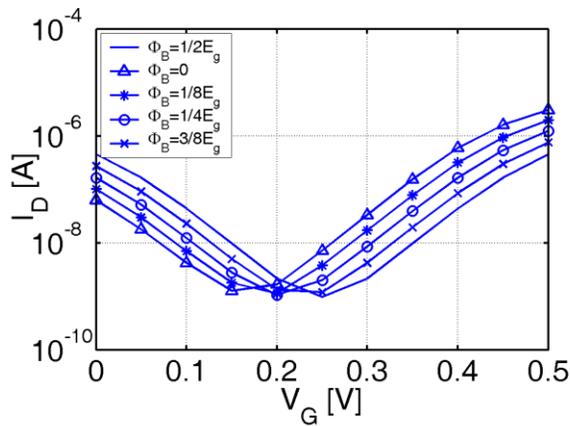


Figure 5-10. The I_D vs. V_G characteristics for single gate device with different Schottky barrier heights (The curves with triangles, stars, circles, crosses and without marks are for Schottky barrier height equal to 0, $1/8 E_g$, $1/4 E_g$, $3/8 E_g$ and $1/2 E_g$ respectively).

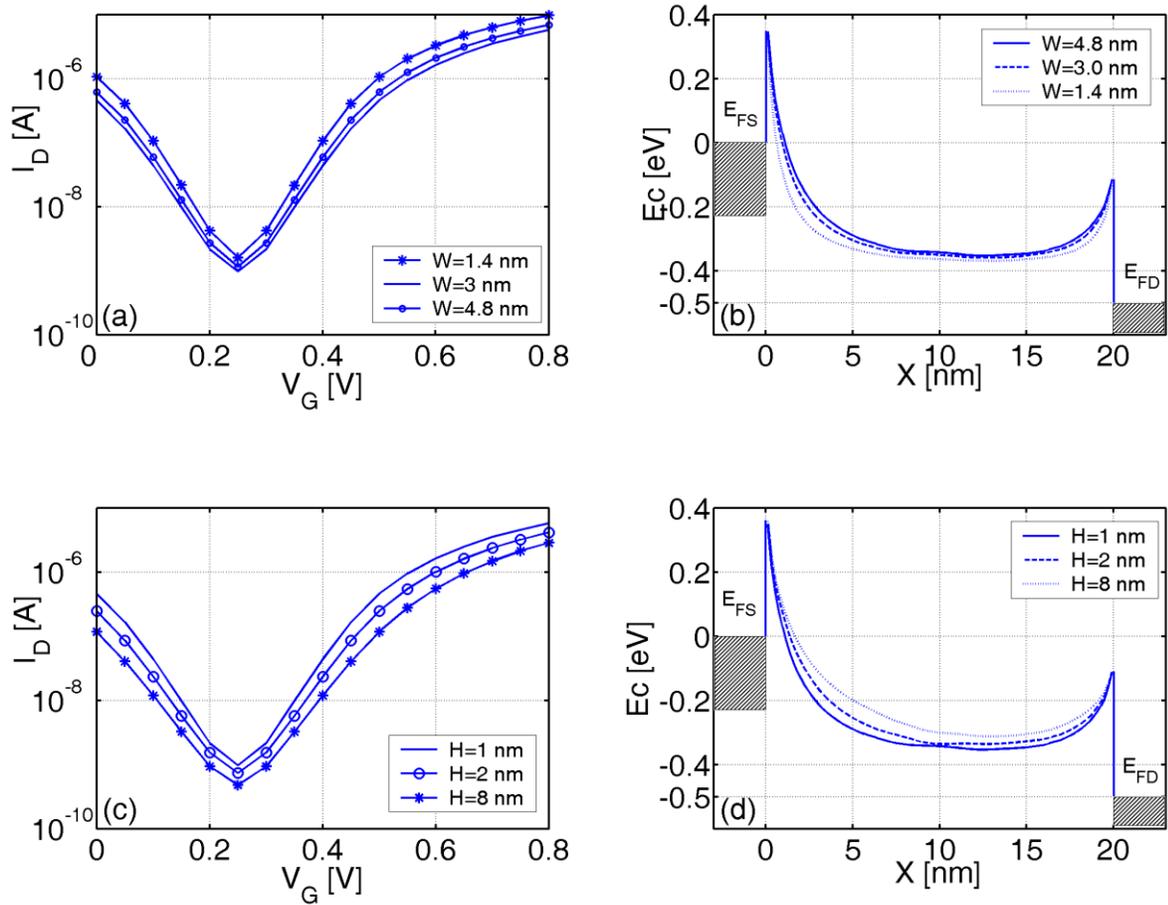


Figure 5-11. Effects of contact width and height (a) The I_D vs. V_G characteristics for single gate device with contact width equal to 1.4 nm, 3 nm, and 4.8 nm, respectively. (b) The conduction band profiles in transport direction for source/drain contact width equal to 4.8 nm, 3 nm and 1.4 nm. E_{FS} and E_{FD} indicate source and drain Fermi levels respectively. (c) The I_D vs. V_G characteristics for single gate device with contact height equal to 1 nm, 2 nm and 8 nm, respectively. (d) The conduction band profiles in transport direction for source/drain contact height equal to 1 nm, 2nm and 8 nm.

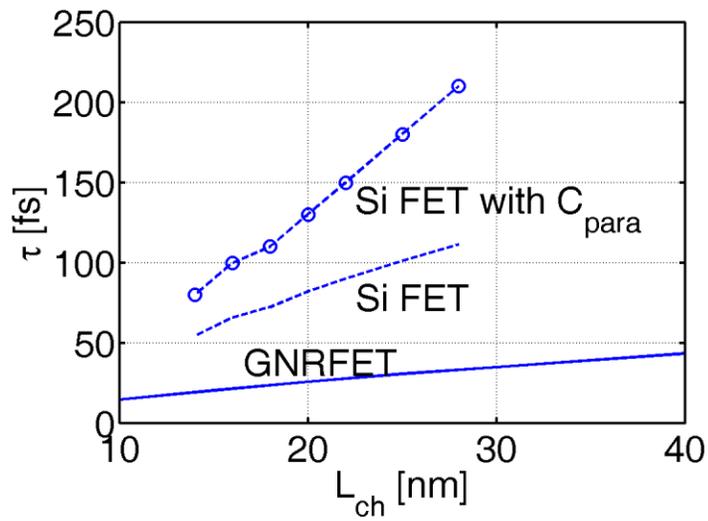


Figure 5-12. The intrinsic delay vs. channel length for a GNRFET (the solid line) and a Si MOSFET (the dashed line). The line with circles is the ITRS' long-term projection on transistor delay with parasitic capacitance and the dashed line is recomputed by using ideal gate capacitance without parasitic capacitance for a fair comparison of the intrinsic delay.

NOTES ON FABRY-PEROT SIMULATIONS

Self-consistent electrostatics is captured by a simple MOS capacitor model. The bias dependent quantum capacitance of GNR channel is taken care of. The conductance peaks happen at the wave vectors that give $2n\pi$ phase shift for a round trip in the channel. From $E-k$ the energies of this wave vectors can be found and related to the gate voltages by equating Fermi energy to this carrier energy. At these gate voltages conductance has the peak values. This simple argument is only true when only one subband is conducting in channel.

To obtain the whole conductance-vs.-gate-voltage curve, the transmission as a function of carrier energy is computed using the scattering matrix method as described below. Subband-dependent parameters are used and they are the contact reflection coefficients, initial phases and transmission scaling factors. The voltages for conductance peaks for single subband transport exactly match those derived from the simple method above. When there are multi-subbands in conduction, the total transmission is the summation of the transmissions from each subbands. The number of peaks in total transmission curve is less than the sum of the peak numbers of all subbands because some small peaks are washed away in the summation.

Modeling of Channels: The source, drain and channel are modeled by three scattering matrices respectively,

$$S(E)_i, i = 1,2,3 \tag{A-1}$$

Ballistic transport is assumed in channel. The notations follow Ref. [57].

$\sqrt{v_{out}}\psi_{out} = S(E)_2\sqrt{v_{in}}\psi_{in}$, note that it is the flux amplitude (wave function amplitude times square root of velocity) that are related with scattering matrix elements

$$\begin{bmatrix} D^+ \\ S^- \end{bmatrix} = \begin{bmatrix} t & r' \\ r & t' \end{bmatrix} \begin{bmatrix} S^+ \\ D^- \end{bmatrix} \quad (\text{A-2})$$

where $r'=[0]$, $r=[0]$, and $t=t'$

$$t = \begin{bmatrix} \exp(i\varphi_1) & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & \exp(i\varphi_N) \end{bmatrix} \quad (\text{A-3})$$

N is the number of modes in the channel and the scattering matrices are of a size of $2N$ by $2N$. The number of modes N and phase shift φ_i are determined as follows. The number of modes is equal to branches crossed by the horizontal line Phase is equal to channel length times the wave vector k at the energy.

Modeling of Scatterers: Intra-mode reflection is treated as follow,

$$S_i \approx 1 + i[a]$$

$$[a] = \begin{bmatrix} a_0 & a_1 \\ a_1 & a_0 \end{bmatrix}$$

$$[a_0] = [0], [a_1] = \begin{bmatrix} ar & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & ar \end{bmatrix} \quad \text{A-4}$$

S_i is exactly unitary by using $S_i = \exp(i[a])$. This will ensure that an incident wave is partially transmitted to an outgoing wave of the same mode on the other port and partially reflected to the same mode at the same port.

Inter-mode scattering is treated as,

$$S_i \approx 1 + i[b], [b] = \begin{bmatrix} br & \dots & br \\ \dots & \dots & \dots \\ br & \dots & br \end{bmatrix}$$

$$S_i = \exp(i[b])$$

$$\beta = \frac{\exp(i2Nb) - 1}{2N} \quad \text{A-5}$$

S_i has a size of $2N$ by $2N$. The physical meaning is that the an incident wave of mode v is transmitted to a outgoing wave on v at the other port with $1-(2N-1)|\beta|^2$ probability, and $|\beta|^2$ to other $N-1$ outgoing waves at the other port and $|\beta|^2$ to N reflected waves at the same port.

The final form of the scattering matrices of two contacts: $S_i = \exp(i[a] + i[b])$, larger $[a]$ elements will enhance the sharpness of the resonance peaks.

Mode-dependent parameters: Initial phases: $\varphi_i' = k_i(E)L + \varphi_{i0}$, which can be used to shift the FP resonance peaks in subband i . Reflection amplitudes: $[a]$ and $[b]$ in contact S matrices. Transmission scaling factors $\{\alpha_i\}$: $T_i = \alpha_i T_{i0} = \alpha_i \sum_{j=1}^N |t_{ij}|^2 = \alpha_i [tt^+]_{ii}$.

The effects of these parameters on the conductance characteristics are shown in Fig. A-3.

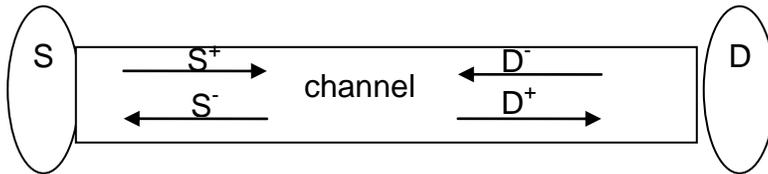


Figure A-1. The schematics of the multi-mode conduction in GNR channel. The plus sign indicates the right-going modes whereas the minus sign indicates the left-going modes.

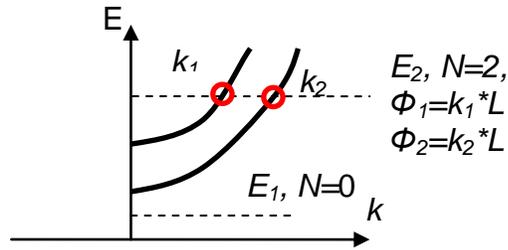


Figure A-2. Schematics of the E - k with positive group velocity for phase calculation.

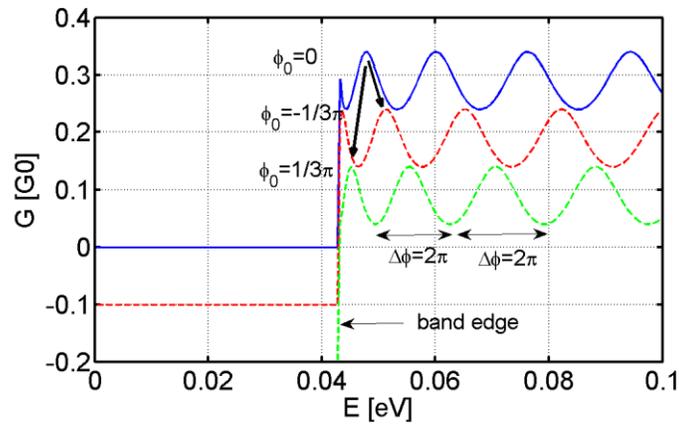


Figure A-3. Conductance as a function of Fermi energy for different set of mode parameters

LIST OF REFERENCES

- [1] S. Iijima, "Helical microtubules of graphitic carbon," *Nature*, vol. 354, no. 6348, pp. 56-58, Nov. 1991.
- [2] K. S. Novoselov et al., "Electric field effect in atomically thin carbon films," *Science*, vol. 306, no. 5696, pp. 666-669, Oct. 2004.
- [3] Y. Zhang, Y.-W. Tan, H. L. Stormer, and P. Kim, "Experimental observation of the quantum Hall effect and Berry's phase in graphene," *Nature*, vol. 438, no. 7065, pp. 201-204, Nov. 2005.
- [4] R. Van Noorden, "Moving towards a graphene world," *Nature*, vol. 442, no. 7100, pp. 228-229, Jul. 2006.
- [5] R. Saito, M. S. Dresselhaus, and G. Dresselhaus, *Physical properties of carbon nanotubes*. Imperial College Press, 1998.
- [6] R. A. Jishi, D. Inomata, K. Nakao, and M. S. D. and G. Dresselhaus, "Electronic and lattice properties of carbon nanotubes," *Journal of the Physical Society of Japan*, vol. 63, pp. 2252-2260, 1994.
- [7] J. W. G. Wilder, L. C. Venema, A. G. Rinzler, R. E. Smalley, and C. Dekker, "Electronic structure of atomically resolved carbon nanotubes," *Nature*, vol. 391, no. 6662, pp. 59-62, Jan. 1998.
- [8] H. Suzuura and T. Ando, "Phonons and electron-phonon scattering in carbon nanotubes," *Physical Review B*, vol. 65, no. 23, p. 235412, May. 2002.
- [9] D. Mann, A. Javey, J. Kong, Q. Wang, and H. Dai, "Ballistic transport in metallic nanotubes with reliable pd ohmic contacts," *Nano Letters*, vol. 3, no. 11, pp. 1541-1544, Nov. 2003.
- [10] Z. Yao, C. L. Kane, and C. Dekker, "High-field electrical transport in single-wall carbon nanotubes," *Physical Review Letters*, vol. 84, no. 13, p. 2941, Mar. 2000.
- [11] R. Martel, T. Schmidt, H. R. Shea, T. Hertel, and P. Avouris, "Single- and multi-wall carbon nanotube field-effect transistors," *Applied Physics Letters*, vol. 73, no. 17, pp. 2447-2449, Oct. 1998.
- [12] S. J. Tans, A. R. M. Verschueren, and C. Dekker, "Room-temperature transistor based on a single carbon nanotube," *Nature*, vol. 393, no. 6680, pp. 49-52, May. 1998.
- [13] A. Javey et al., "High-field quasiballistic transport in short carbon nanotubes," *Physical Review Letters*, vol. 92, no. 10, p. 106804, Mar. 2004.

- [14] A. Javey, R. Tu, D. B. Farmer, J. Guo, R. G. Gordon, and H. Dai, "High performance n-type carbon nanotube field-effect transistors with chemically doped contacts," *Nano Letters*, vol. 5, no. 2, pp. 345-348, Feb. 2005.
- [15] A. Javey et al., "Carbon nanotube field-effect transistors with integrated ohmic contacts and high- κ gate dielectrics," *Nano Letters*, vol. 4, no. 3, pp. 447-450, Mar. 2004.
- [16] Y. Lu, S. Bangsaruntip, X. Wang, L. Zhang, Y. Nishi, and H. Dai, "DNA functionalization of carbon nanotubes for ultrathin atomic layer deposition of high κ dielectrics for nanotube transistors with 60 mV/decade switching," *Journal of the American Chemical Society*, vol. 128, no. 11, pp. 3518-3519, Mar. 2006.
- [17] P. L. McEuen and M. S. Fuhrer, Hongkun Park, "Single-walled carbon nanotube electronics," *Nanotechnology, IEEE Transactions on*, vol. 1, no. 1, pp. 78-85, 2002.
- [18] S. J. Kang et al., "High-performance electronics using dense, perfectly aligned arrays of single-walled carbon nanotubes," *Nat Nano*, vol. 2, no. 4, pp. 230-236, Apr. 2007.
- [19] G. Zhang et al., "Selective etching of metallic carbon nanotubes by gas-phase reaction," *Science*, vol. 314, no. 5801, pp. 974-977, Nov. 2006.
- [20] C.-L. Chen, V. Agarwal, S. Sonkusale, and M. R. Dokmeci, "The heterogeneous integration of single-walled carbon nanotubes onto complementary metal oxide semiconductor circuitry for sensing applications," *Nanotechnology*, vol. 20, no. 22, p. 225302, 2009.
- [21] M. S. Haque et al., "On-chip deposition of carbon nanotubes using CMOS microhotplates," *Nanotechnology*, vol. 19, no. 2, p. 025607, 2008.
- [22] Z. Chen, Y.-M. Lin, M. J. Rooks, and P. Avouris, "Graphene nano-ribbon electronics," *Physica E: Low-dimensional Systems and Nanostructures*, vol. 40, no. 2, pp. 228-232, Dec. 2007.
- [23] X. Wang et al., "N-Doping of graphene through electrothermal reactions with ammonia," *Science*, vol. 324, no. 5928, pp. 768-771, May. 2009.
- [24] L. Jiao, L. Zhang, X. Wang, G. Diankov, and H. Dai, "Narrow graphene nanoribbons from carbon nanotubes," *Nature*, vol. 458, no. 7240, pp. 877-880, Apr. 2009.
- [25] J. Heyd, G. E. Scuseria, and M. Ernzerhof, "Hybrid functionals based on a screened Coulomb potential," *The Journal of Chemical Physics*, vol. 118, no. 18, p. 8207, 2003.
- [26] S. Datta and M. S. Lundstrom, with A. Rahman, Jing Guo, "Theory of ballistic nanotransistors," *Electron Devices, IEEE Transactions on*, vol. 50, no. 9, pp. 1853-1864, 2003.

- [27] H.-S. P. Wong, with Jie Deng, “A compact SPICE model for carbon-nanotube field-effect transistors including nonidealities and its application—Part II: full device model and circuit performance benchmarking,” *Electron Devices, IEEE Transactions on*, vol. 54, no. 12, pp. 3195-3205, 2007.
- [28] C. Berger et al., “Electronic confinement and coherence in patterned epitaxial graphene,” *Science*, vol. 312, no. 5777, pp. 1191-1196, May. 2006.
- [29] X. Li, X. Wang, L. Zhang, S. Lee, and H. Dai, “Chemically derived, ultrasmooth graphene nanoribbon semiconductors,” *Science*, vol. 319, no. 5867, pp. 1229-1232, Feb. 2008.
- [30] Y.-M. Lin, V. Perebeinos, Z. Chen, and P. Avouris, “Electrical observation of subband formation in graphene nanoribbons,” *Physical Review B (Condensed Matter and Materials Physics)*, vol. 78, no. 16, pp. 161409-4, Oct. 2008.
- [31] X. Wang, Y. Ouyang, X. Li, H. Wang, J. Guo, and H. Dai, “Room-Temperature all-semiconducting sub-10-nm graphene nanoribbon field-effect transistors,” *Physical Review Letters*, vol. 100, no. 20, pp. 206803-4, May. 2008.
- [32] Y. Ouyang, X. Wang, H. Dai, and J. Guo, “Carrier scattering in graphene nanoribbon field-effect transistors,” *Applied Physics Letters*, vol. 92, no. 24, pp. 243124-3, Jun. 2008.
- [33] D. Gunlycke, H. M. Lawler, and C. T. White, “Room-temperature ballistic transport in narrow graphene strips,” *Physical Review B (Condensed Matter and Materials Physics)*, vol. 75, no. 8, pp. 085418-5, Feb. 2007.
- [34] Y. Ouyang, Y. Yoon, J. K. Fodor, and J. Guo, “Comparison of performance limits for carbon nanoribbon and carbon nanotube transistors,” *Applied Physics Letters*, vol. 89, no. 20, pp. 203107-3, Nov. 2006.
- [35] Z. Chen, J. Appenzeller, J. Knoch, Y.-ming Lin, and P. Avouris, “The role of metal–nanotube contact in the performance of carbon nanotube field-effect transistors,” *Nano Letters*, vol. 5, no. 7, pp. 1497-1502, Jul. 2005.
- [36] A. Javey, R. Tu, D. B. Farmer, J. Guo, R. G. Gordon, and H. Dai, “High performance n-type carbon nanotube field-effect transistors with chemically doped contacts,” *Nano Letters*, vol. 5, no. 2, pp. 345-348, Feb. 2005.
- [37] D. V. Kosynkin et al., “Longitudinal unzipping of carbon nanotubes to form graphene nanoribbons,” *Nature*, vol. 458, no. 7240, pp. 872-876, Apr. 2009.
- [38] K. Nakada, M. Fujita, G. Dresselhaus, and M. S. Dresselhaus, “Edge state in graphene ribbons: Nanometer size effect and edge shape dependence,” *Physical Review B*, vol. 54, no. 24, p. 17954, Dec. 1996.

- [39] Y.-W. Son, M. L. Cohen, and S. G. Louie, "Energy gaps in graphene nanoribbons," *Physical Review Letters*, vol. 97, no. 21, pp. 216803-4, Nov. 2006.
- [40] A. Gruneis et al., "Tight-binding description of the quasiparticle dispersion of graphite and few-layer graphene," *Physical Review B (Condensed Matter and Materials Physics)*, vol. 78, no. 20, pp. 205425-16, Nov. 2008.
- [41] "ITRS, International Technology Roadmap for Semiconductors, www.itrs.net."
- [42] J.-H. Chen, C. Jang, S. Xiao, M. Ishigami, and M. S. Fuhrer, "Intrinsic and extrinsic performance limits of graphene devices on SiO₂," *Nat Nano*, vol. 3, no. 4, pp. 206-209, Apr. 2008.
- [43] Y. Sui and J. Appenzeller, "Screening and interlayer coupling in multilayer graphene field-effect transistors," *Nano Letters*, vol. 9, no. 8, pp. 2973-2977, 2009.
- [44] J. Guo, Y. Yoon, and Y. Ouyang, "Gate electrostatics and quantum capacitance of graphene nanoribbons," *Nano Letters*, vol. 7, no. 7, pp. 1935-1940, Jul. 2007.
- [45] N. Neophytou, A. Paul, M. S. Lundstrom, and G. Klimeck, "Bandstructure effects in silicon nanowire electron transport," *Electron Devices, IEEE Transactions on*, vol. 55, no. 6, pp. 1286-1297, 2008.
- [46] T. Fang, A. Konar, H. Xing, and D. Jena, "Mobility in semiconducting graphene nanoribbons: Phonon, impurity, and edge roughness scattering," *Physical Review B*, vol. 78, no. 20, p. 205403, Nov. 2008.
- [47] A. Betti, G. Fiori, G. Iannaccone, and Y. Mao, "Physical insights on graphene nanoribbon mobility through atomistic simulations," in *Electron Devices Meeting (IEDM), 2009 IEEE International*, 2009, pp. 1-4.
- [48] S. Datta, *Quantum transport*. Cambridge University Press, 2005.
- [49] S. Sanvito, C. J. Lambert, J. H. Jefferson, and A. M. Bratkovsky, "General Green's-function formalism for transport calculations with spd Hamiltonians and giant magnetoresistance in Co- and Ni-based magnetic multilayers," *Physical Review B*, vol. 59, no. 18, p. 11936, May. 1999.
- [50] P. Ordejon, E. Artacho, and J. M. Soler, "Self-consistent order-N density-functional calculations for very large systems," *Physical Review B*, vol. 53, no. 16, p. R10441, Apr. 1996.
- [51] D. A. Abanin, P. A. Lee, and L. S. Levitov, "Spin-Filtered edge states and quantum hall effect in graphene," *Physical Review Letters*, vol. 96, no. 17, pp. 176803-4, May. 2006.

- [52] C. Adessi, S. Roche, and X. Blase, “Reduced backscattering in potassium-doped nanotubes: Ab initio and semiempirical simulations,” *Physical Review B (Condensed Matter and Materials Physics)*, vol. 73, no. 12, pp. 125414-5, Mar. 2006.
- [53] J. P. Robinson, H. Schomerus, L. Oroszlany, and V. I. Fal’ko, “Adsorbate-Limited conductivity of graphene,” *Physical Review Letters*, vol. 101, no. 19, pp. 196803-4, Nov. 2008.
- [54] X. Wang et al., “Graphene nanoribbon with smooth edges as quantum wires,” Manual in preparation.
- [55] D. B. Farmer et al., “Chemical doping and electron–hole conduction asymmetry in graphene devices,” *Nano Letters*, vol. 9, no. 1, pp. 388-392, Jan. 2009.
- [56] S. Datta, *Quantum Transport: Atom to Transistor*. Cambridge, UK: Cambridge University Press, 2005.
- [57] M. Cahay, M. McLennan, and S. Datta, “Conductance of an array of elastic scatterers: A scattering-matrix approach,” *Physical Review B*, vol. 37, no. 17, p. 10125, Jun. 1988.
- [58] J. M. B. Lopes dos Santos, N. M. R. Peres, and A. H. Castro Neto, “Graphene bilayer with a twist: electronic structure,” *Physical Review Letters*, vol. 99, no. 25, p. 256802, Dec. 2007.
- [59] D. Zhang and E. Polizzi, “Efficient modeling techniques for atomistic-based electronic density calculations,” *Journal of Computational Electronics*, vol. 7, no. 3, pp. 427-431, 2008.
- [60] M. Y. Han, J. C. Brant, and P. Kim, “Electron transport in disordered graphene nanoribbons,” *Physical Review Letters*, vol. 104, no. 5, p. 056801, Feb. 2010.
- [61] M. Y. Han, B. Ozyilmaz, Y. Zhang, and P. Kim, “Energy band-gap engineering of graphene nanoribbons,” *Physical Review Letters*, vol. 98, no. 20, p. 206805, May. 2007.
- [62] W. Liang, M. Bockrath, D. Bozovic, J. H. Hafner, M. Tinkham, and H. Park, “Fabry - Perot interference in a nanotube electron waveguide,” *Nature*, vol. 411, no. 6838, pp. 665-669, Jun. 2001.
- [63] M. I. Katsnelson, K. S. Novoselov, and A. K. Geim, “Chiral tunnelling and the Klein paradox in graphene,” *Nat Phys*, vol. 2, no. 9, pp. 620-625, 2006.
- [64] T. Dürkop, B. M. Kim, and M. S. Fuhrer, “Properties and applications of high-mobility semiconducting nanotubes,” *Journal of Physics: Condensed Matter*, vol. 16, no. 18, p. R553-R580, 2004.
- [65] A. Javey et al., “Self-Aligned ballistic molecular transistors and electrically parallel nanotube arrays,” *Nano Letters*, vol. 4, no. 7, pp. 1319-1322, Jul. 2004.

- [66] B. Obradovic et al., "Analysis of graphene nanoribbons as a channel material for field-effect transistors," *Applied Physics Letters*, vol. 88, no. 14, p. 142102, 2006.
- [67] Y. Ouyang, Y. Yoon, J. K. Fodor, and J. Guo, "Comparison of performance limits for carbon nanoribbon and carbon nanotube transistors," *Applied Physics Letters*, vol. 89, no. 20, pp. 203107-3, Nov. 2006.
- [68] P. Avouris, J. Appenzeller, V. Derycke, R. Martel, and S. Wind, "Carbon nanotube electronics," in *Electron Devices Meeting, 2002. IEDM '02. Digest. International, 2002*, pp. 281-284.
- [69] J. Tersoff, "Schottky barrier heights and the continuum of gap states," *Physical Review Letters*, vol. 52, no. 12, p. 1054, Mar. 1984.
- [70] R. Chau et al., "Benchmarking nanotechnology for high-performance and low-power logic transistor applications," *Nanotechnology, IEEE Transactions on*, vol. 4, no. 2, pp. 153-158, 2005.
- [71] S. Luryi, "Quantum capacitance devices," *Applied Physics Letters*, vol. 52, no. 6, p. 501, 1988.
- [72] S. Datta and M. Lundstrom, with Jing Guo, "A numerical study of scaling issues for Schottky-barrier carbon nanotube transistors," *Electron Devices, IEEE Transactions on*, vol. 51, no. 2, pp. 172-177, 2004.
- [73] S. Heinze, M. Radosavljević, J. Tersoff, and P. Avouris, "Unexpected scaling of the performance of carbon nanotube Schottky-barrier transistors," *Physical Review B*, vol. 68, no. 23, p. 235418, Dec. 2003.
- [74] D. L. John, L. C. Castro, J. Clifford, and D. L. Pulfrey, "Electrostatics of coaxial Schottky-barrier nanotube field-effect transistors," *Nanotechnology, IEEE Transactions on*, vol. 2, no. 3, pp. 175-180, 2003.
- [75] M. Lundstrom, *Nanoscale Transistors: Device Physics, Modeling and Simulation*. New York: Springer, 2006.

BIOGRAPHICAL SKETCH

Yijian Ouyang was born in Wuhan, Hubei Province, China in 1982. He received his B. S. degree in electrical engineering in 2005 from Zhejiang University, China. In 2005, he started his Ph. D study in Department of Electrical and Computer Engineering at University of Florida. His doctoral research focuses on device physics, and numerical modeling of carbon based nanoscale devices for electronics applications.