Title
Quantum Transport and Phase-Field Modeling for Next-Generation Logic Devices

Author
Smith, Samuel Justin

Publication Date
2017

Peer reviewed|Thesis/dissertation

Quantum Transport and Phase-Field Modeling for Next-Generation Logic Devices

by

Samuel Justin Smith

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Sayeef Salahuddin, Chair
Professor Jeffrey Bokor
Professor Jeffrey Neaton

Summer 2017

Quantum Transport and Phase-Field Modeling for Next-Generation Logic Devices

Abstract

Quantum Transport and Phase-Field Modeling for Next-Generation Logic Devices

by

Samuel Justin Smith

Doctor of Philosophy in Engineering – Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Sayeef Salahuddin, Chair

Modeling of semiconductor devices plays an important role in determining which future technologies are most promising for the semiconductor industry as well as optimizing the performance and better understanding the underlying physics of existing devices. This thesis focuses on the design, development, and use of software to study transport in low-dimensional materials and explores the physics of negative capacitance in ferroelectrics.

Quantum transport simulation is used to examine the properties of graphene nanoribbons in geometries that can be fabricated through bottom-up chemical synthesis. The chevron graphene nanoribbon is shown to have an electronic structure analogous to traditional semiconductor superlattices. It is shown how this property could be utilized to create a new type of device, which exhibits both negative differential resistance and steep-slope (< 60 mV/decade) switching for low-power electronics applications. We discuss BerkeleyNano3D, a new quantum transport simulator based on the non-equilibrium Greens function (NEGF) formalism, which is capable of efficient three-dimensional device simulation on large computing clusters.

Finally, the phenomenon of ferroelectric negative capacitance is examined through the lens of phase-field simulations based on the time-dependent Ginzburg-Landau equation. This phenomenon has been previously predicted as a means to enable energy-efficient steep-slope device with minimal modification to existing transistor processes. Simulation results from three-dimensional phase-field modeling provide new insight into the underlying mechanisms for negative capacitance and give far better agreement with experiment than previously studied single domain models.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank Prof. Sayeef Salahuddin for the opportunity to perform research in his group and learn from him. Working with him has helped me grow enormously as a researcher and as a professional. I would like to thank Profs. Jeffrey Bokor, Vivek Subramanian, and Jeffrey Neaton for agreeing to serve on my qualifying exam committee. I would like to thank Prof. Bokor especially for his advice and suggestions on the portion of my work related to transport in graphene nanoribbons, which was also guided by Profs. Felix Fischer and Michael Crommie. Within the graphene supergroup, I would like to acknowledge many useful discussions with Juan-Pablo Llinás, Kyunghoon Lee, Shuang Wu, Chen Chen, Zahra Pedramrazi, Daniel Rizzo, Rebecca Durr, and Ting Cao.

Within the Salahuddin Group, I would like to acknowledge many useful discussions and collaborative works with the modeling and simulation subgroup members: Youngki Yoon, Khalid Ashraf, Kartik Ganapathi, Varun Mishra, and Sangwook Kim. I would also like to recognize Asif Khan and Korok Chatterjee for their expertise in negative capacitance and helping shape those portions of this thesis. I would also like to acknowledge the other members of the Salahuddin Group with whom I have worked extensively over the past six years: Debanjan Bhowmik, Dominic Labanowski, Justin Wong, Niklas Roschewsky, Zhongyuan Lu, James Clarkson, Ava Tan, Mohammad Esmaeilii, Long You, Golnaz Karbasian, Praveen Gowtham, Charles Henri Lambert, Ajay Yadav, Sangwan Kim, and Daewoong Kwon. I would also like to thank Charlotte Jones and Shirley Salanio for their help with administrative matters on countless occasions.

During my time at UC Berkeley, I have taken several classes that have shaped the direction my research in a significant way. Math 211 with Prof. James Demmel taught me how to think about approaching problems in numerical linear algebra, and my project for that course forms the basis for the chapter in this manuscript on the HSC algorithm. Computer Science 267, also with Prof. Demmel, gave me great insight into how to write efficient scientific simulation software to target modern parallel architectures. Mechanical Engineering 280A with Prof. Tarek Zohdi gave me an appreciation for the mathematical beauty that underlies the finite element method. Electrical Engineering 231 (now 230B) with Prof. Vivek Subramanian taught me the fundamentals of semiconductor devices in interesting and exciting way.

I would like to thank Uygar Avci, Dmitri Nikonov, and Ian Young of the Components Research division at Intel Corporation for their guidance in improving the phase-field simulator code during my Summer 2014 internship in Hillsboro, OR.

I would also like to thank my many friends and colleagues I have met throughout graduate school, of whom there are too many to name. My parents Craig and Keith Ann Smith have always been there and supported me throughout my life, along with my grandfather Jack Matheny, my brothers Benjamin and Brian Smith, and my sisters Kelly Adkins and Kerry Rapp. Finally, I would like to thank Stacie Ong for her love and support.

Chapter 1

Introduction

1.1 Overcoming Boltzmann limit for switching

The power dissipated by a transistor in an electronic circuit is $P = fCV^2$, where f is the clock frequency, C is the capacitance, and V is the supply voltage. In modern integrated circuits, the capacitance is dominated by metallic interconnect lines and the surrounding interlayer dielectrics. Frequency is set by the logic depth and propagation delays and has remained mostly constant since the early 2000s. As a result, much of the work on reducing power consumption in modern metal-oxide-semiconductor field-effect transistor (MOSFET) devices has focused on reducing the logic supply voltage V_{DD}. Reducing this value significantly below 0.7 V has proven to be difficult as it is constrained by the thermal nature of the switching mechanism of a MOSFET.

For any device which works by modulating an electrostatic barrier that prevents a thermal population of carriers from crossing the barrier, the ability to turn off the device and prevent charge flow is inherently limited by the thermal tail of the charge carrier distribution [125]. In the limit of Boltzmann statistics, the for every factor of $\ln(10)kT$ (approximately 60 mV at 300 K) the barrier is lowered, current will increase by a factor of 10. This means that to have an ON-OFF ratio of 10^6, a supply voltage of no less than 360 mV is required before taking into account noise margin and electrostatic concerns.

Not only is supply voltage reduction limited by the nature of the thermal distribution of carriers, it is also limited by electrostatics. The gate sees a capacitive divider network (see Fig. 1.2) with the oxide capacitance C_{ox} in series with the semiconductor depletion capacitance C_{dm}, with the channel surface potential ψ_s at the semiconductor surface between those two effective capacitances.

The full expression for the subthreshold current of a MOSFET taking into account this capacitive divider effect is given by Taur and Ning as [125]:

$$I_{ds} = \mu_{eff} C_{ox} \frac{W}{L} (m-1) \left(\frac{kT}{q}\right) e^{-q(V_{gs}-V_t)/mkT} \left(1 - e^{qV_{ds}/kT}\right) \tag{1.1}$$

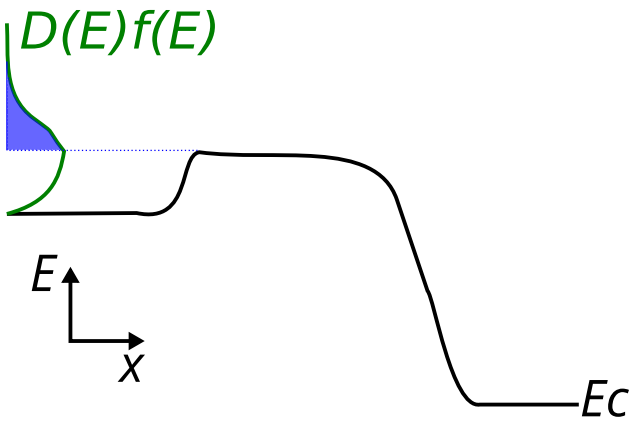


Figure 1.1: Leakage current (blue) over the top of the barrier in an n-channel MOSFET. The conduction band is indicated by E_c. The carrier density of states $D(E)$ is multiplied by the Fermi-Dirac distribution $f(E)$ to give the total population of carriers that can cross the barrier. The height of the source-side injection barrier is modulated by the gate voltage and controls how much of the thermal distribution of carriers at the source can cross into the channel.

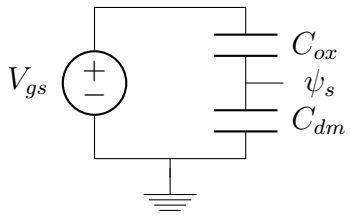


Figure 1.2: MOSFET gate stack series capacitor circuit model. The channel potential ψ_s is between the oxide capacitance C_{ox} and the semiconductor depletion capacitance C_{dm}.

where

$$m \equiv \Delta V_{gs}/\Delta\psi_s = \left(1 + \frac{C_{dm}}{C_{ox}}\right) \tag{1.2}$$

The subthreshold swing is then given by [125]:

$$S = \left(\frac{d\log_{10} I_{ds}}{dV_{gs}}\right)^{-1} = \ln(10)\frac{mkT}{q} = 60\,\text{mV}\left(1 + \frac{C_{dm}}{C_{ox}}\right) \tag{1.3}$$

Substantial work towards optimizing the electrostatics has been done and modern FinFET [56, 101] devices can get reasonably close to this 60 mV/decade limit. Proposals to go beyond this limit primarily involve devices which are not based on thermal switching mechanisms and include tunnel field-effect transistors (TFETs) [60], MEMS relays [108], and nanomagnetic logic [13]. In this thesis, we will examine two other technologies through the lens of computational modeling. The first, the superlattice FET [48, 80], works by filtering

high energy electrons above the source side injection barrier using a superlattice that has finite bandwidth for its conduction band. The second, the negative capacitance FET [116] uses a ferroelectric gate oxide that can behave as "negative capacitor" and thereby passively amplify the value of ψ_s the device sees.

1.2 Device modeling with the Non-Equilibrium Green's Function formalism

The Non-Equilibium Green's function (NEGF) formalism is a method for describing transport in a quantum mechanical system that interacts with its surrounding environment [28, 29, 30, 3]. In contrast to solving the Schrödinger equation for an isolated system, NEGF allows contacts with open boundary conditions to be incorporated in a natural way.

In contrast with semiclassical models, such as drift-diffusion, hydrodynamic and energy-transport models [50], deterministic solvers for the Boltzmann transport equation [58], and Monte Carlo methods [62, 39], the description of the system used by NEGF can model phenomena, such as tunneling and quantum confinement, that are purely quantum mechanical in origin in a realistic way, without resorting to empirical models or approximations.

A schematic of how a device is modeled using NEGF is shown in 1.3. For purely ballistic quantum transport, a central scattering region represented by a Hamiltonian, H, interacts with contact reservoirs described by self-energies Σ_L and Σ_R. Additionally, NEGF also allows interactions with other scattering processes, such as phonons or surface roughness, to be naturally incorporated into simulations via additional scattering self-energy terms Σ.

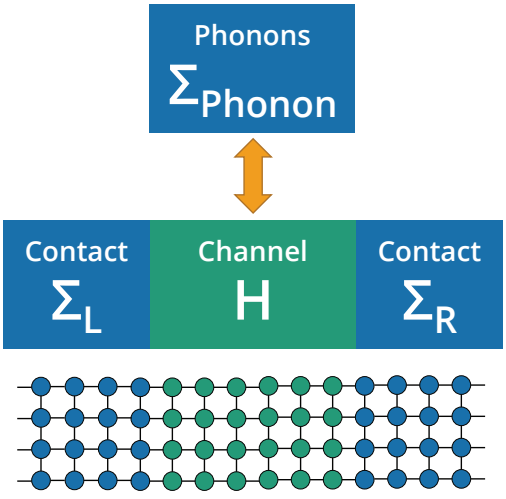


Figure 1.3: Schematic of a device within the NEGF formalism. A central region described by a Hamiltonian H is connected to semi-infinite contacts described by self-energies Σ_L and Σ_R. Interaction with scattering processes is described by self-energy Σ_{Phonon}.

Typically, the Hamiltonian and all other quantities are expressed in a set of localized tight-binding basis functions, such as those given by Slater and Koster [121]. While this is not a requirement of the NEGF formalism, it is typically required for computational efficiency.

Key NEGF equations

To model devices with NEGF, the retarded Green's function G^r, the electron correlation function G^n, and the hole correlation function G^p typically have to be calculated [3]. The retarded Green's function can be computed by:

$$G^r = [EI - H - \Sigma + i\eta]^{-1} \tag{1.4}$$

where E is energy, I is the identity matrix, Σ is the self-energy as a result of the left and right semi-infinite contacts and scattering processes, and $i\eta$ is a small imaginary number.

The electron and hole correlation functions can be computed as:

$$G^n = G\Sigma^{in}G^\dagger \tag{1.5}$$

$$G^p = G\Sigma^{out}G^\dagger \tag{1.6}$$

where Σ^{in} and Σ^{out} are the in-scattering and out-scattering self energies respectively. Typically G^r, G^n, and G^p are not computed directly with full matrix inversion and multiplication, but rather through an efficient algorithm such as the recursive Green's function (RGF) method [79]. Once the Green's functions have been calculated, the density of states can be calculated from the diagonal of G^r, and the electron and hole charge densities can be calculated through the diagonals or G^n and G^p respectively. For ballistic transport, the transmission can be calculated as:

$$T(E) = \mathrm{Tr}\left[\Gamma_L G^r \Gamma_R G^{r\dagger}\right] \tag{1.7}$$

where

$$\Gamma_L = -2\mathrm{Im}[\Sigma_L(E)] \tag{1.8}$$

$$\Gamma_R = -2\mathrm{Im}[\Sigma_R(E)] \tag{1.9}$$

Current can be calculated from the transmission coefficient and the difference between the value of the Fermi-Dirac distributions at the left and right contacts at each energy point and then integrating over energy:

$$J = \frac{ie}{2\pi\hbar}\int_{E_{\min}}^{E_{\max}} T(E)[f_L(E) - f_R(E)] \tag{1.10}$$

Self-consistent field method

Typically, we are not simply interested in solving the NEGF equations by themselves. To model a device, we need to self-consistently solve for the electrostatic potential from Poisson's equation with the charge and current from NEGF. The details of this procedure as well as a complete description of BerkeleyNano3D transport simulator is discussed in Chapter 2. In Chapters 4 and 4 we will discuss the application of the NEGF method to graphene nanoribbons and carbon nanotubes.

1.3 Ferroelectric modeling

Landau models for ferroelectric materials

The second half of this thesis focuses on phase-field simulations of ferroelectric materials. Ferroelectric materials are those materials which have a spontaneous electric polarization [91, 112]. Analogous to how a ferromagnet can be switched by means of an external magnetic field, ferroelectric polarization can be switched by an external electric field. When the external field is removed, there is still a remnant polarization. This behavior is depicted by the hysteresis loop shown in Fig. 1.4.

The hysteresis loop and the relationship between polarization, electric field, and free energy can be modeled through Landau theory [20]. We can write an expression for the free energy F of a ferroelectric as a function of external electric field E and polarization P as:

$$F = \alpha P^2 + \beta P^4 + \gamma P^6 - EP \tag{1.11}$$

where α (< 0), β (usually > 0), and γ (> 0) are the Landau coefficients for the ferroelectric. The Landau coefficients for different ferroelectrics are typically fit with thermodynamic data. The free energy relationship is plotted for an arbitrary ferroelectric in Fig. 1.5.

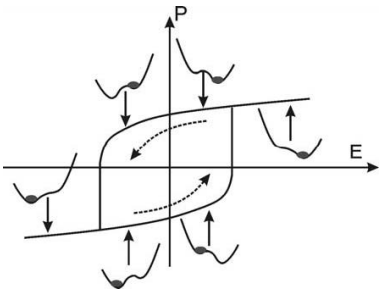


Figure 1.4: Relationship between polarization and electric field for a ferroelectric. The hysteretic effect can be used to make memory devices. Figure from Chandra and Littlewood[20].

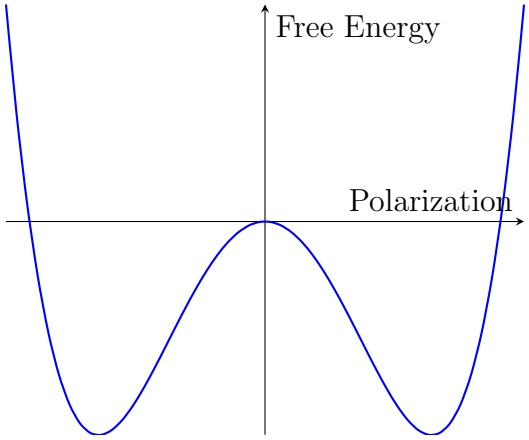


Figure 1.5: Free energy of a ferroelectric with respect to polarization with no electric field present.

The equilibrium polarization for a given electric field can be computed by setting the derivative of free energy with respect to polarization equal to zero:

$$\frac{dF}{dP} = 0 = 2\alpha P + 4\beta P^3 + 6\gamma P^5 - E \tag{1.12}$$

If we ignore the P^5 term, which is inconsequential for our basic analysis, we find that this equation can have up to three possible solutions. We can visually see from Fig. 1.5 that these solutions at zero field will be $\pm P_S$, where P_S the spontaneous polarization, and $P = 0$. The solution at $P = 0$ is, however, dynamically unstable to infinitesimal perturbations because the second derivative is negative. This is a point we will return to later.

The dynamics of ferroelectric can be modeled through the Landau-Khalatnikov equation [81]:

$$-\rho\frac{dP}{dt} = \frac{dF}{dP} = 2\alpha P + 4\beta P^3 + 6\gamma P^5 - E \tag{1.13}$$

where ρ is a parameter that has units of resistance and sets the timescale over which free energy is minimized in the ferroelectric. This equation represents a gradient descent of the ferroelectric free energy landscape over time. Again, we see that the system will stabilize when $dF/dP = 0$. The final value of polarization will subject to the initial conditions. By varying the electric field and finding the steady state condition, we can trace out hysteresis loops similar to the one previously shown in Fig. 1.4.

Negative capacitance

The key idea behind the theory negative capacitance is that the unstable equilibrium of the system near $P = 0$ can be stabilized under some circumstances [116]. We know from basic physics that a capacitor with charge Q and capacitance C has energy $F = Q^2/2C$. Now

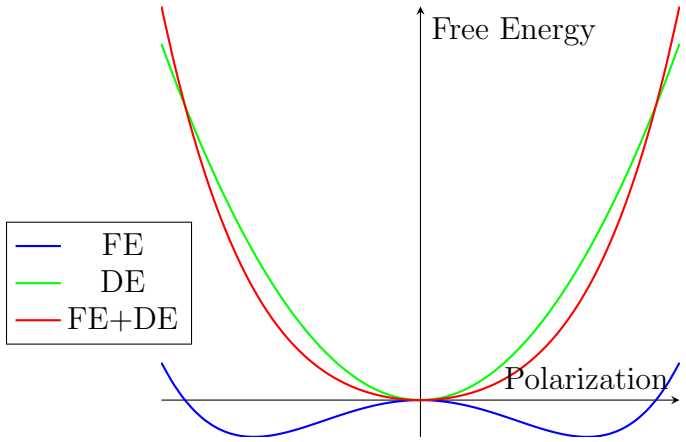


Figure 1.6: Free energy of a combination of a ferroelectric (FE) and a dielectric (DE) capacitor. The unstable free energy extremum at $P = 0$ has been stabilized by the addition of the dielectric capacitor.

suppose we connect a standard dielectric capacitor in series with the ferroelectric capacitor subject to the boundary condition that the two will have the same charge (polarization).

The free energy of the combination of the ferroelectric (FE) and dielectric (DE) capacitors is now given by:

$$F = \left(\alpha_{\mathrm{FE}} + \frac{1}{2C_{\mathrm{DE}}}\right) P^2 + \beta P^4 + \gamma P^6 - EP \tag{1.14}$$

The combined free energy landscape is plotted in Fig. 1.6 for a suitably chosen capacitor value. As we can see visually, the previously unstable point $P = 0$ has been stabilized by the addition of the dielectric capacitor and that the combined system has higher capacitance than the dielectric alone.

We can relate the capacitance of a ferroelectric or dielectric to its free energy through the following expression:

$$C = \left(\frac{\partial^2 F}{\partial P^2}\right)^{-1} \tag{1.15}$$

It was shown by Salahuddin and Datta [116] that so long as the total capacitance of the system is positive, the ferroelectric can be stabilized in the region of negative curvature (and hence negative capacitance) near $P = 0$ and that the hysteresis will disappear. If we go back to circuit of Fig. 1.2 and the expression for subthreshold swing in a MOSFET (Eqn. 1.3) and plug in a negative value of C_{ox} by using a ferroelectric as the gate oxide, we see that we can achieve a subthreshold swing less than 60 mV/decade at room temperature. In this case, the surface potential ψ_s will be higher than the external gate voltage V_{gs}. A schematic for a negative capacitance field-effect transistor (NC-FET) is shown in Fig. 1.7.

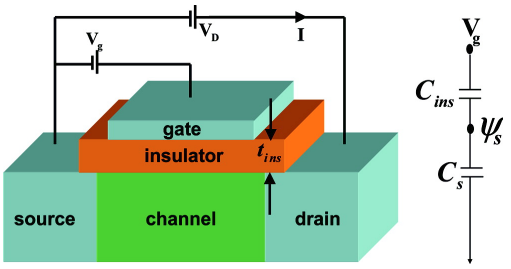


Figure 1.7: Negative capacitance FET. The surface potential ψ_s is amplified when the insulator capacitance C_{ins} is negative. This can be achieved by matching the ferroelectric to the semiconductor capacitance C_s. Image from Salahuddin and Datta [116].

Phase-field modeling of ferroelectrics

In the past few sections we have considered only a very basic model for a ferroelectric. In that model, the ferroelectric's free energy is a function of only a single value of polarization. Real ferroelectrics, much like ferromagnets, are comprised of many domains that have different orientations of polarization [91]. While single domain models are capable of capturing much of the physics of ferroelectrics, multidomain models offer a more realistic, complete picture of ferroelectric behavior at the cost of significantly more complexity.

The most common model for multidomain ferroelectricity is the phase-field model [88, 21]. In a phase-field model, a continuum field is used to represent how an order parameter, polarization in our case, varies over space [109]. Interfaces are never explicitly tracked, but are taken into account through continuous variation of the order parameter field. In this thesis, we will use a phase-field simulator, fully described in Chapter 6, to study negative capacitance. An example of polarization throughout a sample calculated by solving the time-dependent Ginzburg-Landau equation, the continuum equivalent of the Landau-Khalatnikov equation, is shown in Fig.1.8. In this example, there is a mixture of domains oriented up and down with respect to the z-axis.

Effect of multidomain ferroelectricity on negative capacitance

The effect of multidomain ferroelectricity on negative capacitance remains an open problem [17, 145, 83]. The key question is whether a ferroelectric can truly be stabilized in a state near zero or whether it will split into a mixture of possibly more stable up and down domains. In the latter case, can this multidomain state still behave as a negative capacitor for transistor application purposes? We will explore these questions in Chapters 7 and 8.

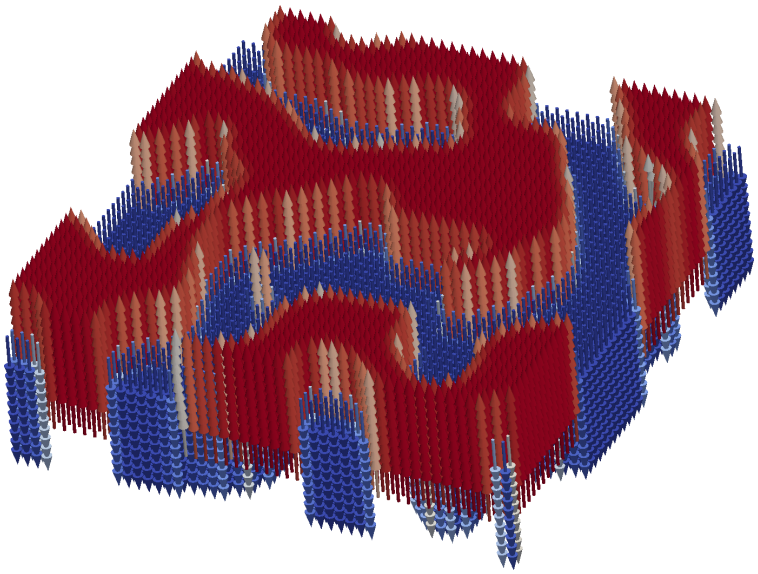


Figure 1.8: Phase-field simulation output for a ferroelectric with domains oriented along the z-axis.

1.4 Outline of the thesis

The thesis is divided into two major parts. The first part, Chapters 2-5, discusses the development of a quantum transport simulator and the applications of that simulator to modeling devices based on graphene nanoribbons. The second part of the thesis, Chapters 6-8, describes work on our group's phase-field simulator and how the phase-field method can be used to model phenomena related to negative capacitance. The chapters can be summarized as follows:

Chapter 2 gives an overview of the development and implementation of BerkeleyNano3D. The finite element-based electrostatics solver, the methods used for self-consistency, the transport solver, the parallelization methods, the implementation of phonon scattering, and the Lua-based scripting interface are described in detail. This chapter is intended to serve as a general reference for the organization of the source code for BerkeleyNano3D.

Chapter 3 covers the implementation of the hierarchical Schur Complement (HSC) method for computationally efficient transport simulation of large devices. This method uses the nested dissection method for partitioning the device Hamilitonian matrix. Benchmarks of our implementation demonstrate the usefulness of this algorithm for large problems at the expense of significant overhead for smaller problems.

Chapter 4 discusses simulation of carbon nanotube and graphene nanoribbon devices with quantum transport. Here we discuss our transport simulation results for devices of these carbon-based materials with particular emphasis on simulating new graphene nanoribbon-based heterostructure devices that can be fabricated with bottom-up chemical synthesis techniques.

Chapter 5 is about chevron-type graphene nanoribbons, which can be made though bottom-up chemical synthesis and have interesting electronic structures similar to superlattices made from traditional semiconductors. Quantum transport simulations show that transistors made from these ribbons could potentially exhibit negative differential resistance and steep-slope behavior, even in the presence of optical phonon scattering.

Chapter 6 covers the development and implementation of the phase-field simulator for modeling ferroelectric materials. This simulator models microstructure evolution in ferroelectric materials taking into account local Landau-Devonshire free energy, domain wall energy, and long-range electrostatic and elastic interactions.

Chapter 7 discusses the stability of negative capacitance within the the time-dependent Ginzburg-Landau equation framework with both analytical results and phase-field simulations of a dielectric capacitor in series with a ferroelectric. The implications of this stability analysis for transistor applications is discussed.

Chapter 8 demonstrates that negative capacitance can be observed for a ferroelectric capacitor in series with a resistor based on a nucleation and growth mechanism within the ferroelectric. This is demonstrated with the phase-field simulator, using an implicit differential equation solver that can self-consistently compute the ferroelectric charge and current.

Chapter 9 contains the conclusion and suggestions for future work.

Chapter 2

BerkeleyNano3D Transport Simulator

2.1 Introduction

As the minimum feature size of devices continues to be scaled into the sub-10 nm regime, quantum transport is rapidly becoming a necessity for accurate modeling of commercial semiconductor devices. Unlike traditional methods of semiclassical semiconductor device modeling, such as the drift-diffusion or hydrodynamic approaches, quantum transport is able to accurately capture phenomena, such as tunneling, that are of a purely quantum mechanical origin. This is important because the key operating principles of several emerging devices, such as the tunnel field-effect transistor (TFET) and superlattice FET, can only be modeled in a rigorous manner through the application of quantum mechanics [3, 30].

The ability to accurately capture the effects of quantum mechanics on device performance, however, comes at a cost. Whereas workstation-class machines are capable of simulating modern three-dimensional devices such as FinFETs with semiclassical transport methods, rigorous atomistic quantum transport calculations require extraordinary computational resources, such as a large cluster or supercomputer. As such, quantum transport simulations incorporating an atomistic Hamiltonian and all the scattering mechanisms relevant in the current generation of commercial MOSFET devices is beyond the scope of problems which can be solved by current transport simulators. But as computing resources grow exponentially and devices shrink in accordance with Moore's law, simulation of devices relevant to commercial products is becoming a more tractable problem.

In this chapter, we will describe the development of BerkeleyNano3D, a new quantum transport simulator based on the non-equilibrium Green's function (NEGF) formalism [30] written to solve several of the perceived deficits in currently available transport codes. While our research group already had transport simulation code written in the C programming language that had been used to solve a number of practical problems from simulating MoS_2 transistors to TFETs [138, 42, 41, 43, 96, 97], the code would have been difficult to extend to solve new problems without extensive amounts of restructuring and rewriting. Several other academic simulation codes exist as well, such as NEMO5 from Purdue University[123],

NanoTCAD ViDES from the University of Pisa [38], and OMEN from ETH Zurich [93]. For various reasons including licensing, availability, and documentation, none of these options was deemed suitable for running the transport calculations to be described later in this thesis, so a new simulator had to be developed.

BerekleyNano3D is written in C++ and is designed according to the principles of object-oriented programming to allow for easy extensibility and maintenance of the code base. The Eigen [51] library is used for linear algebra. This library was chosen as it offers far greater readability of code compared with directly calling BLAS/LAPACK functions. There is no loss of performance when Eigen is linked against an optimized BLAS/LAPACK implementation, such as the Intel Math Kernel Library (MKL) or OpenBLAS. The simulator can be used through a Lua-based scripting interface.

This chapter gives an overview of the structure of the BerkeleyNano3D code, and will discuss the implementation of several key features including: the Hamiltonian and lattice description, the NEGF transport solver based on the recursive Green's function (RGF) algorithm, the finite-element electrostatics solver, the parallel computing interface, the self-consistency solver, and finally the bindings to the Lua programming language for scripting. Additional documentation can be found in the Doxygen comments within the code itself.

2.2 Code overview

The code for the BerkeleyNano3D is organized among several C++ classes with source code within the root directory. The headers are located in the `Nano3DInclude` directory. Various test cases are located in the `Tests` directory, though these have been largely superseded by test cases implemented as Lua scripts. The `TranScriptInterface` directory contains the code for the Lua interface, including the wrapper functions the Sol2 Lua interface directly binds to the scripting interface. The other major directory is the `Util` directory, which contains various Python and Octave/MATLAB helper scripts for post-processing the simulator output.

Atomic lattice

The atomic lattice and the Hamiltonian are implemented within the `RgfLattice` class. The data structure for the storage is optimized for the recursive Green's function algorithm [79], which works with the device as a layered structure with interaction terms of the Hamiltonian present each layer and between adjacent layers. As a result, the Hamiltonian can be stored in a fairly sparse representation as a block tridiagonal matrix. A `vector` of complex double precision matrices is used to store the blocks of the Hamiltonian. Several `vector`s of integers store the indices of the diagonal, upper diagonal, lower diagonal blocks of the Hamiltonian. The same indices can be repeated multiple times to save storage space when Hamiltonian blocks along the device are identical as is the case when a local basis set is used and the device is made of the same material throughout. An additional index `vector` is used when

performing k-space calculations to indicate the blocks to be added to the diagonal blocks after being multiplied by $\exp(ik \cdot r)$ along with Hermitian conjugates scaled by $\exp(-ik \cdot r)$. Two additional indices are used to indicate the periodic hopping blocks at the left and right contacts, which are used in the computation of the surface Green's function. The size of these blocks must be the same as the size of the first and last blocks of the main diagonal respectively.

The Cartesian coordinates of the lattice are also stored in the `RgfLattice` class. The `orbital_mapping` array stores the index of the atomic coordinate each row/column in the full Hamiltonian corresponds to. This is to allow multiple orbitals to be centered around the same atomic site. Finally, the `RgfLattice` class has a method to read the lattice from binary or ASCII files. Code to generate example lattices can be found in the `Tests` directory.

Transport solver

The `RgfTransportSolver` class implements the core of the recursive Green's function (RGF) algorithm for ballistic transport. The solver is initialized with an `RgfLattice` and an options `struct` for the surface Green's function calculation. The method `SolveCoherent` evaluates the transmission coefficient, electron charge, hole charge, and current at a particular energy. The left and right contact Fermi levels, f_L and f_R, along with the electrostatic potential ϕ must also be supplied as input. If a k-space calculation is being performed a value of $k \cdot r$ must also be supplied. The electrostatic potential is typically provided directly as an input for a non-self-consistent field calculation or is computed iterative until self-consistency is achieved by the self-consistency solver, to be described later.

Contact self-energy

Several options for calculating the contact self-energies, Σ_L and Σ_R, are available. There is a simple iterative scheme used mainly for debugging purposes. The faster Sancho-Rubio iteration scheme [118] is generally preferred as it has a much higher convergence rate and should be used in most circumstances. An implementation of the Umerski algorithm [130] is also available. This algorithm solves for the surface Green's function is deterministic time by finding the eigendecomposition of a $2n \times 2n$ matrix, where n is the size of the contact block. This algorithm unfortunately has the drawback that the hopping matrix must be full-rank, a condition not satisfied for most realistic geometries beyond the simple square/cubic lattices used in an effective mass approximation.

Electrostatics solver

Electrostatics in BerkeleyNano3D are handled with a finite element Poisson solver implemented in the `PoissonFEM` class. The abstract class `FemMesh` is implemented by the `FemTriMesh` class for two-dimensional triangular meshes and the `FemTetMesh` class for three-dimensional tetrahedral meshes.

Finite element form of Poisson's equation

The relationship between charge and potential is given by Poisson's equation:

$$\nabla \cdot \epsilon \nabla \phi = -\rho \tag{2.1}$$

where ϵ is the local dielectric constant, ϕ is the electrostatic potential, and ρ is the charge density. While this equation accurately describes the electrostatics of a semiconductor device, it is not in a form that can be directly solved numerically.

In earlier versions of the transport simulator, Poisson's equation was solved using the finite difference method [100]. In this method, derivatives are handled through taking successive differences between points of a field defined on a finite grid. For example, second derivative of potential in Poisson's equation in one dimension is written as:

$$\frac{\partial^2 \phi(i)}{\partial x^2} = \frac{\phi(i-1) - 2\phi(i) + \phi(i+1)}{h^2}, \tag{2.2}$$

where h is the grid spacing. This method allows for regions with different dielectric constants and can be easily extended to three dimensions as is described in Nagel [100]. The finite difference method, while fairly easy to implement, fails to be computationally efficient for systems with irregular geometry, particularly in three dimensions. The finite difference method typically requires a rectilinear (or at least regular) grid. Grids with variable spacing along each axis are possible, but there is no general way to mesh the geometry with variable mesh precision at arbitrary points in the 3D geometry as is possible with the finite volume and finite element methods. Despite the finite difference electrostatics solver no longer being used in BerkeleyNano3D, its ability to handle regular 3D grids made it ideal for inclusion in the phase-field simulator for ferroelectrics described in Chapter 6.

In BerkeleyNano3D, the Poisson solver is implemented with the finite element method. In the finite element method (FEM), the geometry of the system is partitioned into small, simplistic volumes known as finite elements. Within these volumes, the solution of some differential equation is approximated as some continuous interpolation function over the volume. Such interpolating functions, known as shape functions, can be linear or nonlinear and are typically parameterized by values given at the nodes of the element. The element size can be varied throughout the mesh, allowing for the solution to be more precise at some points that others. This is particularly relevant for atomistic device simulation, where the solution needs to be very accurate in the channel and the immediately surrounding regions where the potential can vary over fairly short distances, but in the oxide environment far away from the active region, the solution typically varies slowly as a function of space. Because we can use different mesh densities in different regions as needed, we can get a solution with the same or better level of precision across the entire simulation domain with less computational work.

After generating a suitable mesh for the device, we write the weak (variational) form of Poisson's equation for every element [144]. To get the weak form, we multiply both sides of

Eqn. 2.2 by v, an arbitrary test function within the same approximation space as the shape functions defined over the elements, and integrate over the volume Ω:

$$-\int_{\Omega} \left(\nabla^2 \epsilon\phi\right) v \; d\Omega = \int_{\Omega} \rho v \; d\Omega. \tag{2.3}$$

Note that the dielectric constant ϵ is no longer taken to be spatially varying. This is because we can always mesh the geometry such that material dielectric constant boundaries intersect with the boundaries between elements. When the resulting system of differential equations is built, the variation in the dielectric constant will be taken into account automatically. Next we apply Green's identity to arrive at the weak form:

$$\int_{\Omega} \nabla\epsilon\phi\nabla v \; d\Omega = \int_{\Omega} \rho v \; d\Omega \tag{2.4}$$

Implementation of Poisson solver

This form of Poisson's equations is discretized over a mesh comprised of either triangle elements in 2D or tetrahedral elements in 3D using linear Lagrange shape functions. The solver is implemented in an object-oriented fashion with substantial code shared between the 2D and 3D versions of the solver. While the use of quadrilateral (2D) and hexahedral (3D) elements was considered, implementation complexity as well as the availability of open-source mesh generators was found to be significantly better for triangular and tetrahedral elements.

In the code, the `FemTetMesh` and `FemTriMesh` classes extend the `FemMesh` class. These classes have methods to find the element index corresponding to a set of Catesian coordinates (`FindElement`), convert Cartesian coordinates to barycentric coordinates for a specific element (`BarycentricCoords`), and compute a subset of elements that exist between two corner points specified by Cartesian coordinates (`Subset`). The mesh classes contain the coordinates of the mesh nodes, the mesh connectivity defining the elements, which nodes correspond to Dirichlet (constant voltage) boundary conditions, and the relative dielectric constant ϵ_r for every element in the mesh. The FEM left-hand side (LHS) matrix $\mathbf{K}$ is assembled with the `BuildFEMMatrix` method. The matrix is assembled in the compressed sparse column (CSC) format. The method `BuildFEMMatrixDirichlet` assembles a sparse matrix which must be multiplied by the right-hand side (RHS) vector and the result subtracted from the RHS vector before boundary conditions are applied. This is done to preserve the positive definite property (and therefore symmmetry) of $\mathbf{K}$ so that the conjugate gradient (CG) method can be used to solve the resulting linear system.

The `PoissonFEM` class manages the solution of the electrostatics problem for a particular device. Its constructor takes pointers to both an `FemMesh` and and `RgfLattice`. This constructor calls the matrix construction methods for the `FemMesh` and computes which mesh elements contain every atomic site in the `RgfLattice` as well as the corresponding barycentric coordinates for those sites, which are required for creating the RHS vector as will now be explained.

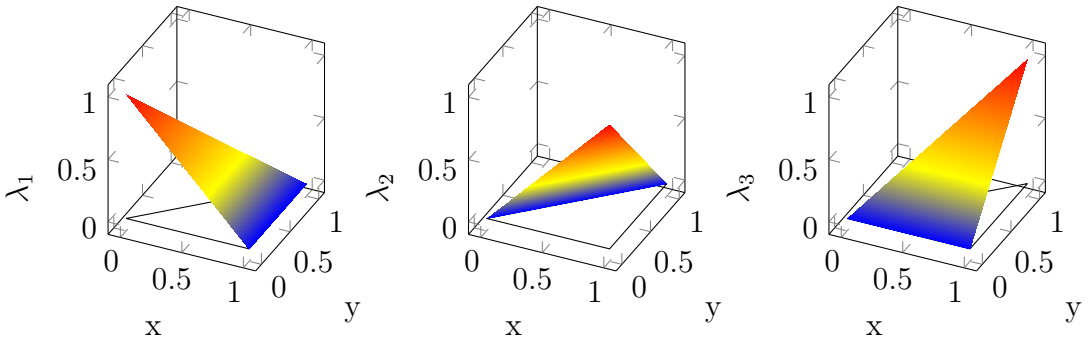


Figure 2.1: Linear shape functions over a triangle with vertices (0,0), (1,0), and (1,1). Each shape function is one at one node and zero at the other two, varying linearly in between. The barycentric coordinates $(\lambda_1, \lambda_2, \lambda_3)$ corresponding the Cartesian coordinates (x, y) are given by the values of the shape functions at each point.

Mapping between atomistic grid and FEM mesh

In order to solve Poisson's equation, we must compute the RHS vector

$$\int_\Omega \rho v \, d\Omega \tag{2.5}$$

where ρ is the charge computed from NEGF, and v is a shape function. For a triangular mesh, each element will have three shape functions, and for a tetrahedral mesh, each element will have four shape functions. Our linear Lagrange shape functions (shown in Fig.2.1 for a triangular element) are defined as being equal to one at a particular node of the element, zero at all the other nodes, and vary linearly throughout the rest of the element. Given that ρ is computed at coordinates defined by the atomic lattice that have no direct relation to the nodes of the FEM mesh, it is not immediately obvious how to evaluate this integral. An obvious solution is to simply assign all the charge to the shape function corresponding to the nearest neighbor node of the mesh or equally to all the nodes of the element containing the lattice site. While these approaches are trivial to implement and clearly conserve total charge, they are problematic in that accuracy can vary immensely given how close each atomic site is to it surrounding mesh nodes.

Another possible solution to this problem is to assume the charge ρ_i for each lattice site i with coordinates $\mathbf{x}_i$ can be represented by a Dirac delta distribution centered about that site, allowing us to write total charge as:

$$\rho(\mathbf{x}) = \sum_i \rho_i \delta(\mathbf{x} - \mathbf{x}_i) \tag{2.6}$$

In the case where each lattice site corresponds directly to a mesh node, the value of the integral is nonzero only for the shape function corresponding to node and the integrand is nonzero only exactly at the mesh node where the shape function is defined to be unity.

Unfortunately, it is difficult to generate FEM meshes where some points of the mesh are predefined at these lattice sites. One possible solution is to generate a mesh using some mesh generation software, add the extra nodes to node list, and then perform Delaunay triangulation or tetrahedralization to generate a new mesh with the updated node list. While this result of this will be a mesh with nodes at the correct locations, there are no guarantees, such as those usually provided by the initial mesh generation process, that the resulting elements will be high quality and not have extreme aspect ratios that will lead to numerical problems.

A way around custom mesh generation and the approach adopted by our electrostatics solver is to evaluate the shape functions for each element at each atomic site assuming the aforementioned Dirac delta charge distribution using a barycentric coordinate transformation. For any given point with Cartesian coordinates $\mathbf{x}$ inside a triangular (tetrahedral) element, the value for each of the three (four) shape functions is given by the barycentric coordinates $\boldsymbol{\lambda}$ corresponding to that point. Barycentric coordinates [35] represent the fraction of total mass that must be placed at each vertex of a triangle or tetrahedron for a given point to be the center of mass (or equivalently center of charge when using charge instead of mass) of that triangle. Thus to determine how to assign the charge from each atomic site to the the vertices of the surrounding element with the correct weighting, we compute the barycentric coordinates within the element for each atomic site. Because the sum of the barycentric coordinates λ_i is always equal to one, charge is conserved.

Using this coordinate transformation, we can now evaluate the integral of 2.5 for lattice site i and shape function j as:

$$\int_{\Omega} \rho_i v_j \, d\Omega = \rho_i \lambda_j(\mathbf{x}_i) \tag{2.7}$$

To quickly calculate the RHS for Poisson's equation, the element containing each site in the `RgfLattice` and the barycentric coordinates corresponding to that site for that element are precomputed when the `PoissonFEM` object is initialized.

Numerical solution of Poisson's equation

Once charge is calculated at each lattice site, the `SolvePoissonSimple` method is used to solve Poisson's equation. This method computes the RHS vector $\boldsymbol{\rho}$ from the input charge using the barycentric coordinate method previous described. Dirichlet boundary conditions, corresponding to regions of constant voltage, are applied next. Finally the resulting system of equations $\mathbf{K}\boldsymbol{\phi} = \boldsymbol{\rho}$ is solved using Eigen's conjugate gradient solver for the potential $\boldsymbol{\psi}$ [51].

The resulting solution of Poisson's equation is defined at the nodes of the FEM mesh. This full mesh potential, extracted using the `FullSolution` method, is useful in determining the behavior of the solution far away from the active region of the device. To get the potential at each lattice site, we simply evaluate the sum of the shape functions at that lattice using the barycentric coordinates previously used to map from the lattice coordinate space to the

FEM mesh. At lattice site i in element j in d dimensions with barycentric coordinates $\boldsymbol{\lambda}$, this is:

$$\phi_i = \sum_{k=0}^{d} \lambda_k \phi_{j,k}. \tag{2.8}$$

This is computed by the `LatInterpSolution` method.

Self-consistency solver

The Poisson equation is solved self-consistently with the NEGF equations as described in Datta [30] and Anantram *et al.* [3]. The procedure is outlined as follows and shown in Fig.2.2:

1. Estimate the potential ϕ
2. Use the estimate of ϕ to solve for the Green's functions.
3. Extract the electron charge from the diagonal of G^n and the hole charge from the diagonal of G^p
4. Solve for Poisson's equation for ϕ_{new} using the calculated charge.
5. If ϕ_{new} is close enough to ϕ by some metric, stop.
6. Else, estimate a new potential and return to step 2.

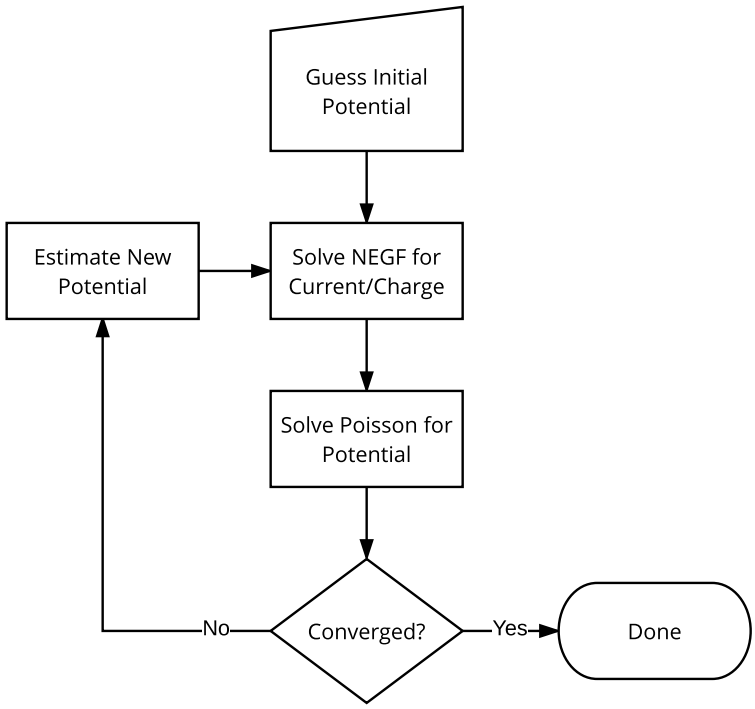


Figure 2.2: Self-consistency loop for coherent transport calculations.

Mixing schemes

In step 6, we estimate a new potential. In most cases simply taking the solution to Poisson's equation from step 4 is not sufficient. If this solution is directly used, the charge can change significantly in the next iteration, and oscillatory or divergent behavior may be observed. Finding a self-consistent solution to Poisson's equation and the NEGF equations is related to the mathematical problem of finding fixed points of a function (*i.e.*, values of x such that $f(x) = x$). In our case, we are looking for fixed points in the solution space of Poisson's equation such that:

$$\phi = \phi_{\text{Poisson}}\left(\rho_{\text{NEGF}}(\phi)\right) \tag{2.9}$$

where $\rho_{\text{NEGF}}(\phi)$ signifies computing charge from NEGF for a given potential and $\phi_{\text{Poisson}}(\rho)$ signifies solving Poisson's equation for a given charge.

To get the solution to converge, it is necessary to implement some sort of mixing scheme to combine the solution of Poisson's equation with the solutions at previous iterations. The simplest of these methods is the relaxed Picard method. In this scheme, sometimes called linear mixing, the new potential is estimated as a combination of the solution of Poisson's equation, ϕ_{Poisson}, from step 4 and the potential from the previous iteration, ϕ_i. This is:

$$\phi_{i+1} = \omega\phi_{\text{Poisson}} + (1-\omega)\phi_i \tag{2.10}$$

for some relaxation parameter ω such that $0 \leq \omega \leq 1$. Values of ω are highly problem-dependent and can range from 0.001 to 0.3. If the value of ω is too small, it may take many iterations to converge to the solution, but if the value is large, the solution may diverge.

While not described in this thesis, BerkeleyNano3D has several potential mixing schemes implemented including: gradient descent, the relaxed secant method, and Pulay mixing (DIIS) [110]. The solution method can be selected using the `outer_relax_method` option in `SelfConsistencyOptions` section of the Lua input script.

Nonlinear predictor-corrector

The most dramatic improvements in convergence can be obtained using a physically based approximation of the Jacobian of the system, such as the predictor-corrector scheme of Trellakis *et al.* [129]. In our implementation, we assume that charge will vary exponentially according to Maxwellian statistics based on some change $\Delta\phi$ in the potential:

$$\rho = \rho_n \exp(q\beta\Delta\phi) + \rho_p \exp(-q\beta\Delta\phi) \tag{2.11}$$

where ρ_n and ρ_p are the electron and hole charge respectively from NEGF and $\beta = 1/k_B T$. Using the first two terms of the Taylor series expansion, we write $\exp(x) = 1 + x$. We substitute this into Poisson's equation and make the substitution $\hat{\phi} = \phi + \Delta\phi$:

$$-\nabla \cdot \epsilon\nabla(\phi + \Delta\phi) = \rho_n \exp(q\beta\Delta\phi) + \rho_p \exp(-q\beta\Delta\phi) \tag{2.12}$$

$$-\nabla \cdot \epsilon\nabla(\phi + \Delta\phi) = \rho_n(1 + q\beta\Delta\phi) + \rho_p(1 - q\beta\Delta\phi) \tag{2.13}$$

$$-\nabla \cdot \epsilon\nabla\hat{\phi} = \rho_n(1 + q\beta\hat{\phi} - q\beta\phi) + \rho_n(1 - q\beta\hat{\phi} + q\beta\phi) \tag{2.14}$$

$$-\nabla \cdot \epsilon\nabla\hat{\phi} + \mathrm{diag}(q\beta\rho_p - q\beta\rho_n)\hat{\phi} = \rho_n(1 - q\beta\phi) + \rho_n(1 + q\beta\phi) \tag{2.15}$$

$$\left(-\nabla \cdot \epsilon\nabla + \mathrm{diag}(q\beta\rho_p - q\beta\rho_n)\right)\hat{\phi} = \rho_n(1 - q\beta\phi) + \rho_n(1 + q\beta\phi), \tag{2.16}$$

where operations are elementwise unless otherwise noted and diag indicates the construction of a diagonal matrix with the elements of a vector. BerkeleyNano3D solves the weak form of this this linearized form of the nonlinear Poisson equation and uses the new estimated potential $\hat{\phi}$ as a guess for the next self-consistency iteration. This is done in one of two ways. The `SolvePoissonLinearized` method of `PoissonFEM` simply calculates the solution of the linearized equation. The `SolvePoissonNonlinear` method recomputes the charge density using the semiclassical estimate of 2.11 and re-linearizes the equation about the new value of ϕ. Usually, under-relaxation is used for $\hat{\phi}$ to prevent the potential from changing too much between iterations and diverging. This iterative solution method requires many conjugate gradient solves per self-consistency iteration, but this is typically not a problem as the conjugate gradient Poisson solver typically takes orders of magnitude less time than the NEGF solver. For extremely large FEM meshes, this may be a bottleneck as the electrostatics solver is limited to shared-memory parallelism on a single compute node rather than the NEGF solver, which can be run in parallel on a large number of compute nodes.

Phonon scattering

Phonon scattering is implemented according to the method of Nikonov *et al.* [106]. We consider scattering processes where the in/out-scattering functions take the form:

$$\Sigma^{in} = D\left(r_1, r_2, E\right)\left(n_q + 1\right) G^n\left(r_1, r_2, E + \hbar\omega\right) \tag{2.17}$$

$$+ D^*\left(r_1, r_2, E\right)\left(n_q\right) G^n\left(r_1, r_2, E - \hbar\omega\right) \tag{2.18}$$

$$\Sigma^{out} = D^*\left(r_1, r_2, E\right)\left(n_q + 1\right) G^n\left(r_1, r_2, E - \hbar\omega\right) \tag{2.19}$$

$$+ D\left(r_1, r_2, E\right)\left(n_q\right) G^n\left(r_1, r_2, E + \hbar\omega\right), \tag{2.20}$$

where the first term corresponds to emission of a phonon and the second term corresponds to absorption of a phonon, D is the electron-phonon coupling constant, $n_q = 1/(\exp(\hbar\omega/k_BT) - 1)$ is the Bose factor, and $\hbar\omega$ is the phonon energy. In our implementation, we consider only the case when scattering in local in space (*i.e.*, $r_1 = r_2$) and consider only the diagonal terms of G^n and G^p.

In BerkeleyNano3D, scattering processes are described by objects implementing the `ScatteringProcess` interface. Currently, the `AcousticPhonon` and `OpticalPhonon` classes have been implemented. All `ScatteringProcess` classes implement the `SigmaInScattering` and `SigmaOutScattering` methods to compute the in/out-scattering diagonal matrices. While elastic scattering with $\hbar\omega \approx 0$ is assumed for acoustic phonons, computing the optical phonon scattering terms require $G^{n/p}$ at $E \pm \hbar\omega$, energy points, which may not be locally stored on the compute node evaluating transport at energy E. The parallel data transfer process to ensure that all compute nodes have access to $G^{n/p}$ at the relevant energies is detailed in a later section.

The effect of phonon scattering is computed self-consistently in a manner similar to charge-potential self-consistency:

1. Solve for G, G^n, and G^p assuming ballistic transport using the RGF algorithm.
2. Compute new values for Σ^{in} and Σ^{out}.
3. Solve for G, G^n, and G^p using new values for Σ^{in} and Σ^{out}.
4. If G^n and G^p are sufficiently close to their previous values by some metric across all energy points, done.
5. Else, go to step 2.

Unlike the charge-potential self-consistency, no mixing scheme is required to achieve convergence in most cases. Typically several iterations are required to achieve current continuity throughout the device, which is physically required. When performing fully self-consistent calculations, note that two self-consistency loops are now required: an outer loop for charge-potential and an inner loop for phonon scattering. The RGF algorithm to compute G, G^n, and G^p must be run significantly many more times than in the case of strictly ballistic transport for each bias point. As a result, self-consistent simulations with phonon scattering are

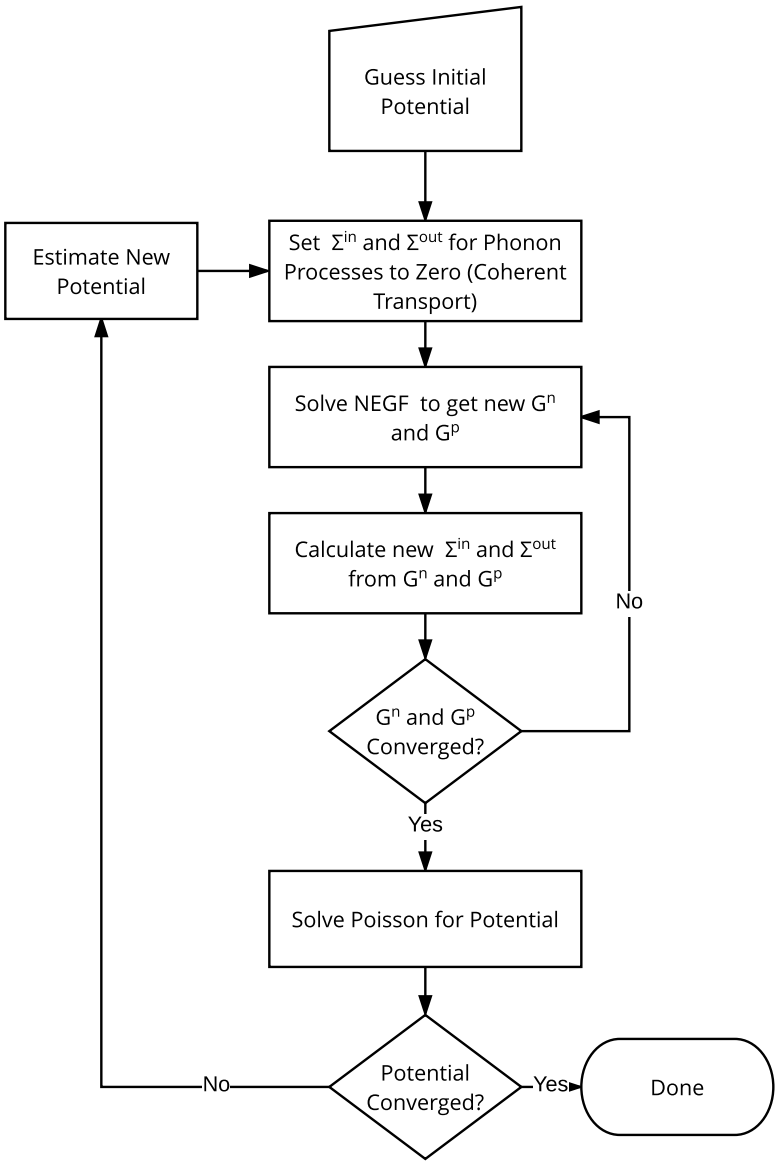


Figure 2.3: Nested self-consistency loop for phonon scattering within electrostatic self-consistency loop.

significantly more computationally expensive. Additionally, the hole charge can be computed without explicitly solving for G^p in the case of ballistic transport, so extra work has to be done solving for G^p when computing dissipative transport.

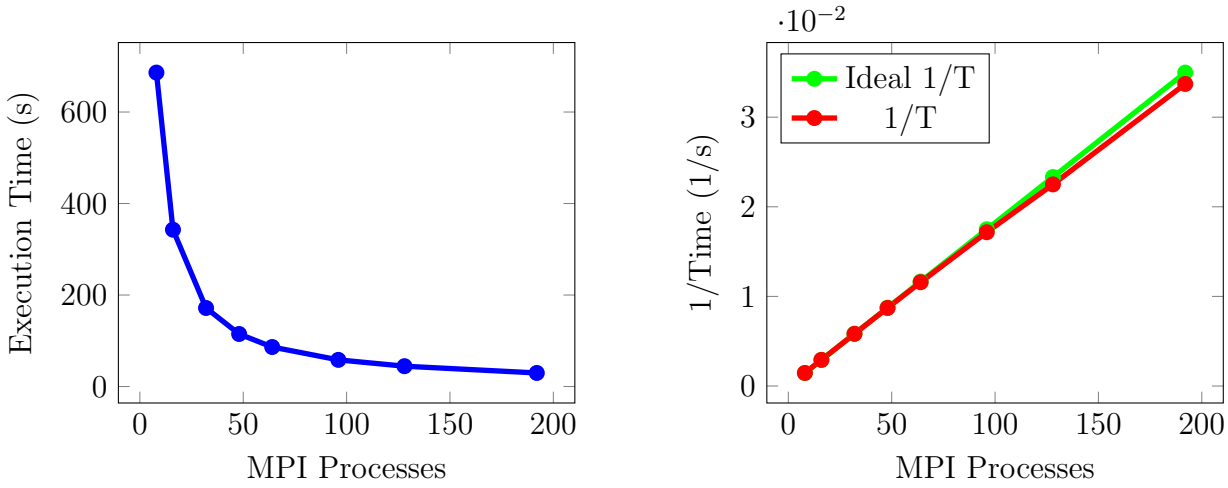


Figure 2.4: Runtime with respect to the number of MPI processes shown on both linear and inverse scales.

Parallel computing interface

BerkeleyNano3D is parallelized with a hybrid of OpenMP for shared memory parallelism and MPI for distributed memory parallelism. This hybrid implementation allows the simulator to take advantage of large parallel systems with a reduced memory footprint and faster interprocess communication within each individual compute node.

Parallel functionality is handled almost exclusively through the `ParallelIntegrator` class. This class contains the `SolveParallel` and analogous `SolveCoherentParallel` methods which divide the workload among the different processes and threads. When running a coherent transport simulation, the workload is embarrassingly parallel with respect to energy and k-points. That is, the required calculations are completely separable, and processes evaluating transport at one energy point never need to communicate with other processes until the computation is done and the results aggregated.

penMP tasks is dependent on the system OpenMP implementation. Testing found that there is little performance difference in using only MPI with a number of MPI processes equal to the total number of CPU cores and using a number of MPI processes equal to the number of compute nodes and an OpenMP thread count (environment variable `OMP_NUM_THREADS`) equal to the number of cores per compute node, but that the hybrid MPI+OpenMP implementation uses less memory. In some cases it may be advantageous to use more processes on recent processors that have Hyper-Threading or an equivalent technology.

Fig. 2.4 shows how the performance of the simulator scales on our group's cluster up to 192 CPU cores for solving a graphene nanoribbon problem at 2302 energy points with 20 self-consistency iterations. Nearly ideal linear scaling is observed due to the embarrassingly parallel nature of the problem, and scaling is expected to extend far beyond 192 cores.

The parallel scaling is excellent and close to linear as a result of the aforementioned

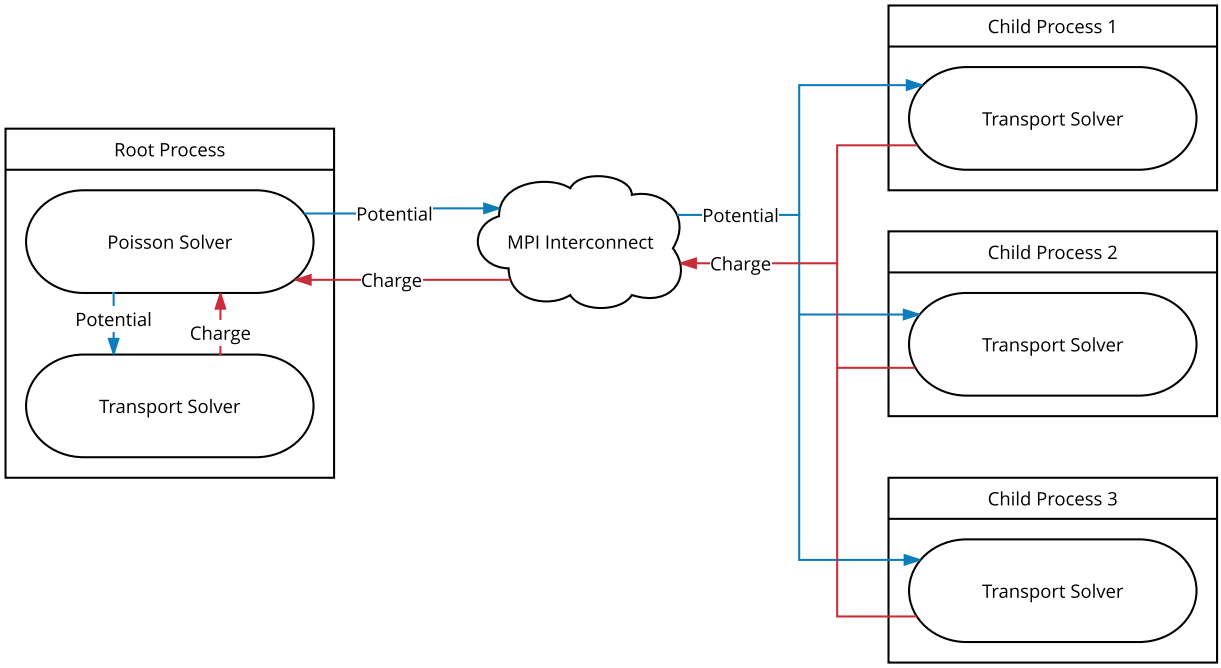


Figure 2.5: Dataflow for parallel computation of coherent transport. Both the Poisson solver and a transport solver run on the root process. During each iteration, every process sends the root process it charge computed at their respective energy quadrature points. The root process aggregates this data, solves Poisson's equation, and then broadcasts the updated potential to every process.

embarrassingly parallel nature of the calculations. The only communication among the processes occurs immediately before and immediately after the RGF transport algorithm is run at all energy points. Before, the root process sends the potential from the electrostatics solver to all the nodes. After, the summed charge is sent to the root process for the electrostatics calculation. In this case, each process does not send charge to the root, but a more efficient `MPI_Reduce` call is used to minimize the communication. Communication is also required when exporting quantities, such as charge and density of states, to output files for post-processing.

Phonon parallelization

Parallelization with phonon scattering is more complicated than the case for ballistic transport because processes evaluating transport at energies $E \pm \hbar\omega$ require the values of G^n and G^p at energy E at every iteration of the phonon self-consistency loop. Parallelization of coherent transport requires only simple MPI calls implementing either all-to-one or one-to-all data transfer, while efficiently parallelizing a calculation with phonon scattering requires a

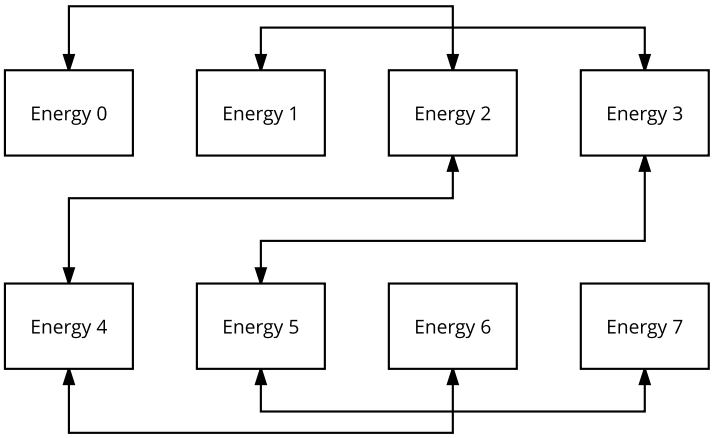


Figure 2.6: Dataflow between processes during self-consistent computation of the Green's functions with phonon scattering. Here, each process holds one energy point with spacing ΔE between energy points, and the phonon energy $\hbar\omega = 2\Delta E$. Therefore, each process in this case must communicate with its second nearest neighbors.

more intricate communication topology.

An initial implementation involved sending the all the values at every energy point of G^n and G^p to every process from every process using `MPI_Gather` calls. This approach obviously does not scale as the required memory at every node is now proportional to the number of energy points and does not reduce with the number of processes. Additionally, communication, rather than computation, is often a bottleneck in parallel computing and should be avoided if unnecessary.

In the current implementation, each process precomputes the processes it must send data to and which it must receive data from based on the how the energy grid is distributed among processes, which energy points it owns, and the optical phonon energy $\hbar\omega$. This is somewhat simplified by realizing that for two energy points separated by $\hbar\omega$ the information flow is necessarily bidirectional. The values for G^n and G^p are transmitted using the asynchronous, non-blocking MPI calls (`MPI_Irecv` and `MPI_Isend`). This allows the data to be sent efficiently without having to worry about ordering the communication for synchronous blocking calls in a careful manner to avoid deadlocks. This implementation is reasonably efficient and performs well in practice, but could be further improved.

A possible future improvement would be to minimize communication through rearranging which processes solve for which energy points instead of simply evenly subdividing the range of energy points. One reasonable approach to do this would be a greedy algorithm that initially assigns one energy point to a process and continues to add energy points that the current set of energy points must communicate with until that process is full and then continues to the next process. Such a scheme could yield a significant reduction in communication in cases where $\hbar\omega$ is large and data locality is unlikely for a large number of

processes.

2.3 Generating inputs for the transport simulator

Lattice and Hamiltonian generation

BerkeleyNano3D does not internally contain code for generating atomic lattices beyond simple rectangular lattices with an effective mass approximation, but rather specifies a format for reading lattices generated by some external script that can be read by the constructor for `RgfLattice`. Example lattice/Hamiltonian scripts for InAs (Python, developed by Charles Zhang) and graphene nanoribbons (Octave/MATLAB) are provided in the `Tests` directory of the source code.

These scripts generate the (x, y, z) coordinates of the atomic sites which are automatically mapped into the Poisson solver, the upper, lower, and main diagonal blocks of the Hamiltonian used for the RGF algorithm, the Hamiltonian blocks for the contacts, and various other mapping arrays that identify which orbitals in the Hamiltonian are associated with which atoms and which matrix indices correspond to which Hamiltonian blocks.

Finite element mesh generation

BerkeleyNano3D also does not generate the finite element meshes used by the electrostatics solver. While it is possible to use any meshing software that can generate tetrahedral meshes, the SALOME platform [113] was used to generate the meshes for the graphene nanoribbon (GNR) devices described later in this thesis. SALOME was chosen because of its excellent scripting capabilities with Python, integrated CAD module, and the possibility of using it in the future to generate quadrilateral and hexahedral meshes.

Geometry is created using the CAD/solid geometry module of SALOME. In this module, different solid regions are used to represent different materials of the device with different dielectric constants (*e.g.*, field oxide, high-κ oxide, channel, etc.). Typically metallic contacts where Dirichlet boundary conditions will be applied are modeled as surfaces. An example of creating geometry for a GNR device with SALOME is shown in Fig. 2.7.

Meshing the device is done in the mesh module of SALOME. Meshing rules are typically applied to constrain the size of mesh elements based on what region of the geometry they belong to. For example, the mesh in the channel region usually needs to be finer than the mesh of the field oxide far away from the active region of the device. BerkeleyNano3D reads meshes in the .dat format SALOME can output. Typically several files are required: a global mesh, submeshes to define which elements belong to the different material regions, and node lists which describe which nodes correspond to fixed potential Dirichlet boundary conditions. Example Python scripts for geometry and mesh creation for SALOME are provided under the `Tests` directory of the BerkeleyNano3D source directory. An example of a meshing a GNR device is shown in Fig. 2.8.

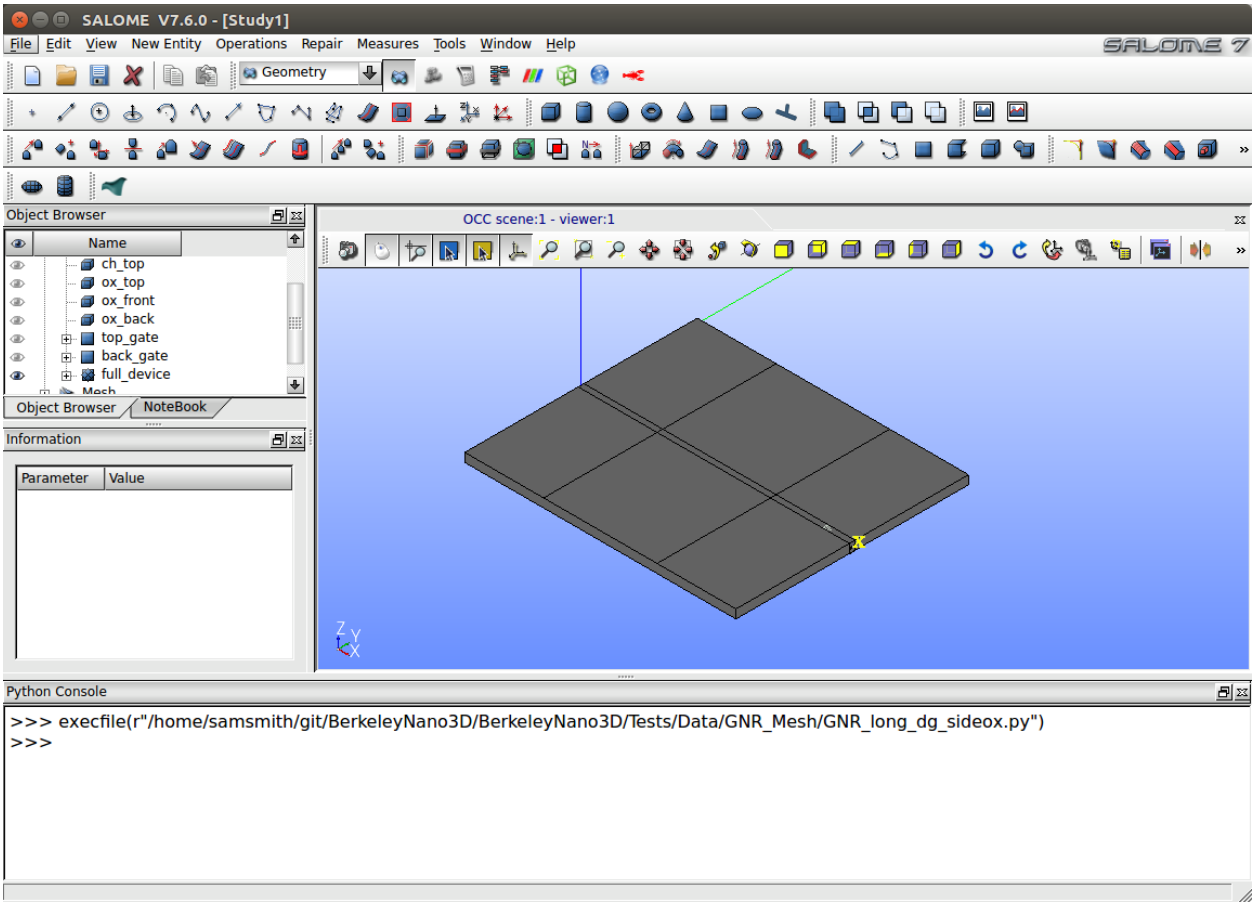


Figure 2.7: Solid geometry model for double-gate GNR transistor in SALOME.

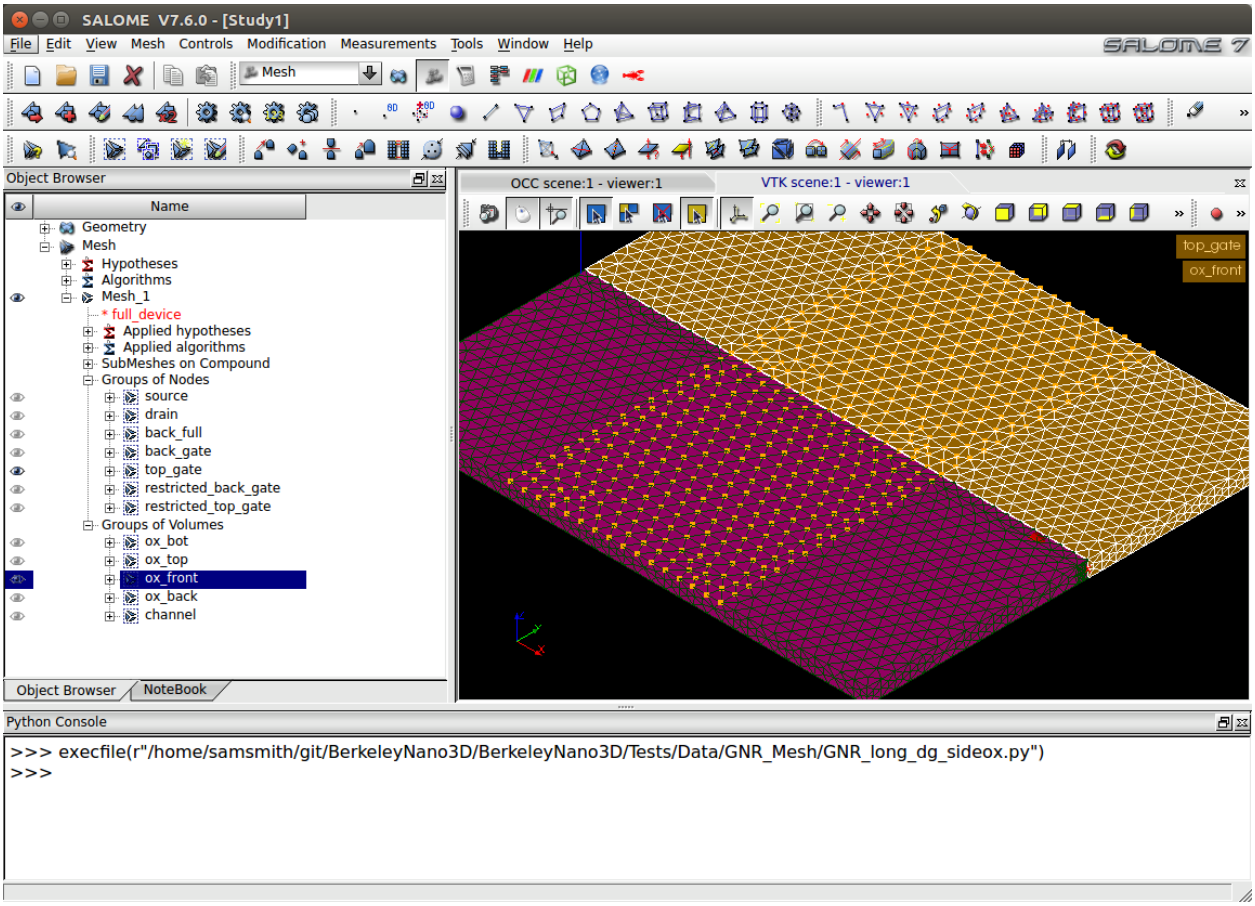


Figure 2.8: Meshing a double-gate GNR transistor in SALOME.

Lua scripting

The primary way of running simulations in BerkeleyNano3D is through the Lua interface known as Transport Script or simply TranScript. The Lua programming language was chosen to implement the scripting interface for several reasons:

1. Portability. The Lua interpreter is written in ANSI C and can easily be compiled on virtually any platform.

2. Ease of Embedding. The Lua language is easy to embed in large applications written in C/C++.

3. Concise mapping to C++ code. The Sol2 Lua API wrapper (Available from `https://github.com/ThePhD/sol2`) for the C++ programming language allows mapping a C++ functions and data structures to a Lua function with very little glue code.

Another major transport simulator, NanoTCAD ViDES [37] uses Python as its scripting interface. We strongly considered this, but ultimately found working with Lua to be easier. The NEMO5 simulator [123] uses Boost.Spirit, but we did not consider this approach as we did not want to create our own domain-specific language and sacrifice the functionality and library support that a general-purpose language, such as Lua or Python, provides.

The functions accessible to the scripting interface are listed in `TranScript.h`. While an exhaustive description of the approximately 50 functions accessible within Lua from BerkeleyNano3D is beyond the scope of this chapter, there are several important concepts behind the implementation.

All interaction with the transport simulator occurs through a `SolverContext` object. Instances of this object contain pointers to an `RgfLattice`, an `RgfTransportSolver`, a `ParallelIntegrator`, and a `SelfConsistency`. While multiple `SolverContext` objects may be created in one script, most simulations will require only one. After creating the context object, individual components of the context may be initialized through the various setup methods for those objects. For example, to setup the `SelfConsistency` solver, the `SetupSelfConsistency` function is called with a table containing the doping passed along with the solver context.

Providing arrays of information to the C++ code, such as the doping or an initial guess of the electrostatic potential, can be done in several ways. The first is to pass a Lua table (numerically indexed from 1) to the simulator. Binary and text files, as well as linear combinations of two or more files, can also be passed as is seen in some of the examples. Configuring the finite element mesh is also done through passing a specially formatted table to `SetupFemMesh`.

After the simulation is setup, performing a self-consistent simulation at a given bias point is done by calling `SolveSelfConsistent`. There is no direct way to run a simulation at multiple bias points. This is instead handled by adding `for` loops to the Lua code and calling `UpdateContactVoltage` to change the contact voltages. This is done to give the user

maximum flexibility in terms of when to write output files and perform post-processing tasks between solving each bias point.

Several examples of Lua scripts for running simulations are provided with the BerkeleyNano3D source code and should be used as references along with the documented C++ source itself when running new simulations.

Processing transport simulator output

The transport simulator generates several types of output files. When and if output files are written is specified by the user in the Lua script for a particular simulation. The Lua functions `WriteSingleTransportPoint` (single quadrature point from an `RgfTransportSolver`) and `WriteAllQuadPoints` (all quadrature points from a `ParallelIntegrator`) write the electron charge, hole charge, density of states, or current at each layer depending on which quantity is specified. This output is in either a raw binary or ASCII format or a binary or ASCII VTK format. Binary formats are greatly preferred, especially for large problems and when manual inspection of text output is not required. ASCII formats are recommended only for very small outputs or in rare cases when data must be transferred from a big-endian CPU architecture, such as some versions of ARM and POWER, to a little-endian CPU architecture, such as x86_64. VTK output files are automatically byte-swapped to the big-endian format required by the VTK specification, though this byte swapping may need to be disabled if porting the simulator to a big-endian architecture.

Additional Lua functions include: `WriteEnergyGrid`, `WriteTransmissionAllQuadPoints`, `WriteCurrentAllQuadPoints` and `WriteQuadratureWeights`. These functions output ASCII text files, which are typically small in size as they output only a single number per quadrature point. The `WriteElectrostaticPotential` function writes the electrostatic potential. This can be for the entire FEM mesh or interpolated at each atomic site depending on the options.

Examples in Python and Octave/MATLAB for post-processing binary data from the simulator are found in the `Util` directory of the code. VTK files are typically handled with either ParaView or VisIt.

Chapter 3

Implementation of the Hierarchical Schur Complement Algorithm

In this chapter, we discuss the our implementation of the Hierarchical Schur Complement (HSC) algorithm originally published by Hetmaniuk *et al.* [54] in 2013. This algorithm was implemented in an earlier version of BerkeleyNano3D and was found to be efficient for certain types of problems. However, support for this algorithm was removed from later versions of the transport simulator due to lack of applicability to most practical transport problems.

In quantum transport simulation within the NEGF formalism, the HSC algorithm is used to calculate the retarded Green's function G^r and the electron correlation function G^n. The fundamental underlying mathematical problem is the computation of selected entries in the inverse of a large sparse matrix with a sparsity pattern similar to the device Hamiltonian. We compare the performance of the HSC algorithm with the recursive Green's function [79] algorithm and show that it has asymptotically better performance, though with substantial overhead.

3.1 Efficient algorithms for quantum transport

Numerical implementations of NEGF typically involve substantial amounts of dense linear algebra. The primary mathematical problem in solving NEGF equations is the computation of the retarded Green's function G^r and the electron correlation function G^n as defined by the equations:

$$\begin{aligned} A &= [EI - H - \Sigma_L - \Sigma_R] \\ AG^r &= I \\ AG^n &= \Sigma^{in}(G^r)^\dagger \end{aligned}$$

where E is a scalar energy, I is the identity matrix, H is the device Hamiltonian, and Σ_L and Σ_R are the self-energies for the left and right contacts. The matrix Σ^{in} describes the

scattering in the device and is typically block diagonal. For coherent transport problems examined in this study only the blocks corresponding to the left and right contacts of the device are nonzero.

Fortunately, we do not need the entire inverse of the matrix A to calculate quantities of physical interest. Typically only blocks along the diagonal are needed. Efficient algorithms, such as the Recursive Green's Function (RGF) method [79] solve for these entires in $\mathcal{O}(ln^3)$ time where l is the length of the device and n is the dimension of the matrix for each layer of the device. The primary problem with the RGF algorithm is that it is fundamentally sequential and limits the ability to parallelize device simulations. For years this issue has been largely ignored because typical simulations involve solving the problem over a large number of values of energy E, and the problem is embarrassingly parallel for calculating separate energy points. With access to modern supercomputers with symmetric multiprocessing (SMP) nodes that have 16 or more cores, it would be advantageous to have an algorithm that can be efficiently parallelized so that larger systems can be studied with more resources available for computation at a single energy point.

In 2007 an algorithm called Fast Inverse using Nested Dissection (FIND) was developed and applied to semiconductor transport [85]. This work was an important first step towards the realization of an efficient direct method for transport. It unfortunately used very large separator regions in its implementation of the nested dissection partitioning of the semiconductor lattice and imposed other geometric restriction on the lattice.

Recently, a new algorithm to selectively invert sparse matrices called the Hierarchical Schur Complement (HSC) method has appeared [89, 90]. This method uses an arbitrary hierarchical partitioning of the graph representing the sparse matrix, usually with nested dissection (ND). At a given level of the partition hierarchy, all subproblems are independent and can thus be solved in parallel. An implementation of this algorithm specifically applied to NEGF simulation to calculate G^r has been developed [107]. It was later extended to the calculation of G^n and to use preexisting software for graph partitioning [54].

3.2 Nested dissection

A key ingredient in the HSC algorithm is a good hierarchical partitioning of the graph representing connectivity in the semiconductor device Hamiltonian. A well-known method to generate such a hierarchical partitioning is nested dissection. The basic idea of nested dissection is to partition a graph into three parts S, L, and R, where L and R are both connected to the separator region S and there are no vertices between L and R. For some applications, such as load balancing in parallel computation, it is ideal that L and R be approximately the same size and that S be small. Nested dissection can be applied recursively to a graph to generate a hierarchy of separator regions, where the resultant hierarchy can be represented as a binary tree. This is illustrated in Fig. 3.1.

For our implementation of lattice partitioning, we used the METIS package [68]. It is somewhat difficult to actually extract the hierarchical partitioning directly from METIS as

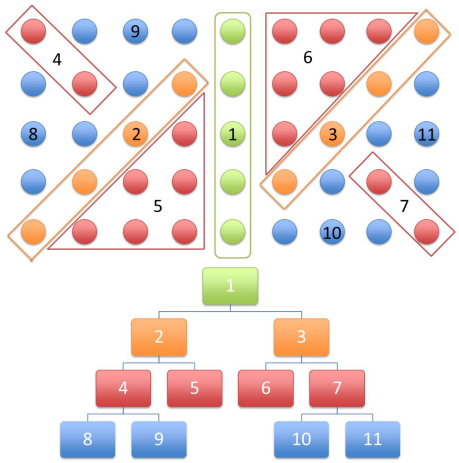


Figure 3.1: Example of a hierarchical partitioning for a rectangular lattice. Figure from Hetmaniuk, *et al.* [54].

the function `Metis_NodeND` returns only a fill-reducing ordering for a sparse matrix intended to be used by a Cholesky factorization routine. Direct correspondence with the authors of Hetmaniuk, *et al.* [54] indicated that their results were generated through modification of the source code to the package to output the partitions. After going through the METIS source code it was determined that no modification was actually necessary if the routine `Metis_NodeNDP` was called instead. This routine is an extra function designed to be used by the parallel version of METIS called ParMETIS and returns the partition sizes in addition to the ordering. As the ordering is a list of vertices in each partition from post-order traversal of the partition tree, it is possible to fully reconstruct all information about the partitioning with knowledge of the partition sizes. Partitioning for a 50×20 lattice with our transport simulator is shown in Fig. 3.2.

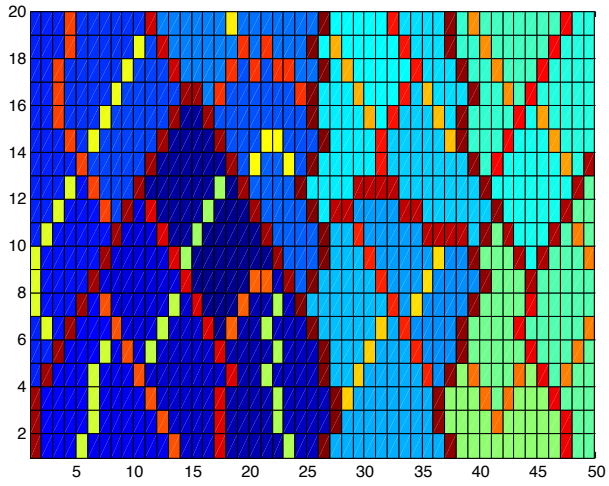


Figure 3.2: Lattice colored by partition number for a 50×20 lattice.

3.3 Hierarchical Schur complement algorithm

Computation of G^r

After we have constructed the nested dissection ordering for the graph, we partition the matrix into regions corresponding to the left and right branches of the tree and the separator region. This results in a matrix with zeros in the 21 and 12 blocks (connections between the left and right partitions). Here we only show the algorithm for a single separator rather than a full tree. The full details of the algorithm are discussed in Hetmaniuk, *et al.* [54].

$$A = \begin{bmatrix} A_{LL} & 0 & A_{LS} \\ 0 & A_{RR} & A_{RS} \\ A_{LS}^T & A_{RS}^T & A_{SS} \end{bmatrix}$$

This form of the matrix can be factorized into LDL^T:

$$A = LDL^T = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ A_{LS}^T A_{LL}^{-1} & A_{RS}^T A_{RR}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{LL} & 0 & 0 \\ 0 & A_{RR} & 0 \\ 0 & 0 & \hat{A}_{SS} \end{bmatrix} \begin{bmatrix} I & 0 & A_{LL}^{-1} A_{LS} \\ 0 & I & A_{RR}^{-1} A_{RS} \\ 0 & 0 & I \end{bmatrix}$$

The term $\hat{A}_{SS}$ is the Schur complement and is defined as:

$$\hat{A}_{SS} \equiv A_{SS} - A_{LS}^T A_{LL}^{-1} A_{LS} - A_{RS}^T A_{RR}^{-1} A_{RS}$$

We now wish to calculate $G^r \equiv A^{-1}$:

$$G^r = (I - L^T) G^r + D^{-1} L^{-1}$$

Substituting these values we get:

$$G^r = \begin{bmatrix} A_{LL}^{-1}A_{LS}G_{SL}^r & A_{LL}^{-1}A_{LS}G_{SR}^r & A_{LL}^{-1}A_{LS}G_{SS}^r \\ A_{RR}^{-1}A_{LS}G_{SL}^r & A_{RR}^{-1}A_{LS}G_{SR}^r & A_{RR}^{-1}A_{LS}G_{SS}^r \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} A_{LL}^{-1} & 0 & 0 \\ 0 & A_{RR}^{-1} & 0 \\ 0 & 0 & \hat{A}_{SS}^{-1} \end{bmatrix} \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -A_{LS}^T A_{LL}^{-1} & -A_{RS}^T A_{RR}^{-1} & I \end{bmatrix}$$

Thus the separator and off-diagonal blocks are calculated as:

$$\begin{aligned} G_{SS}^r &= \left(\hat{A}_{SS}\right)^{-1} \\ G_{LS}^r &= -A_{LL}^{-1}A_{LS}G_{SS}^r \\ G_{RS}^r &= -A_{RR}^{-1}A_{RS}G_{SS}^r \end{aligned}$$

The diagonal blocks can be calculated completely independently of each other. This is where the additional parallelism comes from.

$$\begin{aligned} G_{LL}^r &= A_{LL}^{-1} - A_{LL}^{-1}A_{LS}(G_{LS}^r)^T = A_{LL}^{-1} + A_{LL}^{-1}A_{LS}G_{SS}^r A_{LS}^T A_{LL}^{-1} \\ G_{RR}^r &= A_{RR}^{-1} - A_{RR}^{-1}A_{RS}(G_{RS}^r)^T = A_{RR}^{-1} + A_{RR}^{-1}A_{RS}G_{SS}^r A_{RS}^T A_{RR}^{-1} \end{aligned}$$

Implementing this method recursively, we can solve for more entries of G^r. This method will not generate the full inverse of G^r, but rather only the entires for blocks which interact within the same separator region. Fortunately, these are the only parts of the inverse of significant physical interest.

Complexity analysis

For an $n \times n = N$ grid directly inverting the matrix costs $\mathcal{O}(N^3) = \mathcal{O}(n^6)$. To analyze the complexity of the HSC algorithm we will assume the separator region is of size $\sqrt{m}$ for some partition of size m. If the calculation at this level is dominated by matrix inversion and multiplication of matrices the size of the separator region, the cost is $\mathcal{O}(m^{3/2})$. We split the problem into two subproblems of size $m/2$. This yields a recurrence:

$$T(n) = 2T\left(\frac{m}{2}\right) + m^{3/2} \tag{3.1}$$

Solving the recurrence using the master theorem [27] gives a complexity of $\mathcal{O}(m^{3/2})$. Taking $m = N = n^2$, constructing selective entires in the inverse of G^r for an $n \times n$ device is $\mathcal{O}(n^3)$. This is asymptotically better than the $\mathcal{O}(n^4)$ complexity for RGF which requires n inversions of $n \times n$ matrices and several multiplications of the same size.

Computation of G^n

To calculate G^n, we multiply our partial solution of G^r by Σ^{in}, a known quantity which can be calculated from the self energy matrices Σ_R and Σ_L and the Fermi level at each contact, which is a function of device bias conditions. Some of the entries of G^n can be calculated using a similar algorithm to the one for G^r.

3.4 Implementation

To implement the HSC algorithm several features had to be implemented in our C++ transport simulator code. The interface to METIS for partitioning was previously described. To store the blocks of the sparse matrix for each partition a data structure called `BlockSparse` was developed. This structure contains a sparse matrix stored in the CSR or CSC format where each entry P_{ij} is a pointer to a dense matrix storing the entries for the matrix elements between block i and block j. This is similar to data structures such as block sparse row and block sparse column, but unlike some implementations of these structures, the block sizes can be arbitrary. This is necessary because the different partitions may be different sizes. As a result of each block having a different size, operations on the structure other than set and get methods for individual blocks can become somewhat complicated. The current implementation of this structure is adequate for the implementation of HSC, but substantial performance enhancements, especially for the multiplication of Σ^{in} with G^r in the computation of of G^n could be made.

The implementation of the HSC algorithm follows the description given the paper by Hetmaniuk, *et al.* [54]. When the transport routine is called, the matrix $A \equiv (G^r)^{-1} = EI - H - \Sigma_L - \Sigma_R$ is constructed. EI is diagonal and H can be constructed from the lattice connectivity graph. The contact self-energies Σ_L and Σ_R represent the effect of an infinite lead on either side of the device and are calculated with the Sancho-Rubio algorithm [118]. For certain Hamiltonians, the calculation of the self-energy matrices is a bottleneck and in some cases is actually slower than the transport calculation itself. As a result this calculation is not included in any of the benchmarks in this chapter.

Currently, the code calculates both G^r and G^n. Results have been verified against the recursive Green's function algorithm and match everywhere except very close to the poles of the Green's function, where results are generally suspect. This is not typically an issue in practice. The calculation of G^r has been parallelized with OpenMP. The code can crash or produce erroneous results unless writes to the `BlockSparse` structure are performed as atomic operations. This limitation on multithreading reduces performace. Possible ways to rewrite the data structure to avoid race conditions and the requirement for atomic operations are currently being explored. The other factor limiting parallelism is that near the top of the partition tree there are only a small number of blocks that must be inverted, and these are unfortunately the largest blocks that must be inverted. It may be possible to use a multithreaded LAPACK implementation such as Intel MKL or OpenBLAS to get around this issue.

The code was tested with a few model problems to verify correctness. Fig. 3.3 shows a contour plot the density of states for the device with 50×20 lattice sites with a simple effective mass Hamiltonian whose partitioning was shown above as a function of position (along the x and y axes) and energy (along the z axis). The density of states is a function of the diagonal of G^r. The transmission coefficient for transport in this device as a function of energy is shown in Fig. 3.4. The transmission coefficient is calculated from the first diagonal block of G^r.

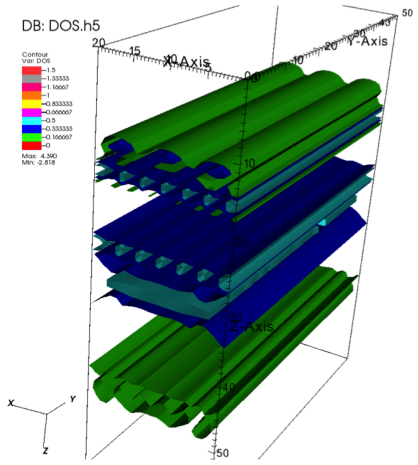


Figure 3.3: Density of states for the 50×20 device as a function of space (x and y axes) and energy (z-axis).

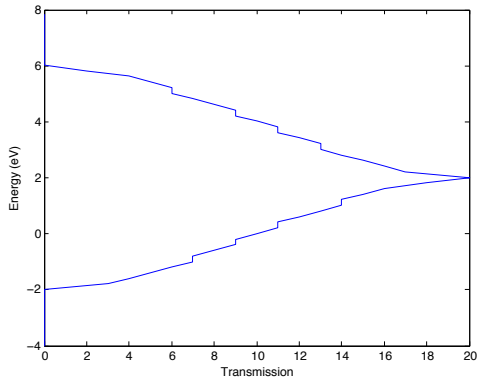


Figure 3.4: Zero bias transmission coefficient for the 50×20 device as a function and energy

3.5 Benchmarks

The code was benchmarked on an 3.5 GHz Intel Core i5 Ivy Bridge workstation running Ubuntu Linux 14.04 LTS. The first benchmark compared the performance of calculating G^r at a single energy point for an $n \times n$ device. For the largest size of 300×300, RGF took 38.179 s, HSC took 16.372 s and HSC with 4 OpenMP threads took 11.926 s. Full results are shown in Fig. 3.5. This indicates that the parallel scaling is suboptimal, but even without parallelism, the algorithm is faster. If the aforementioned data race concurrency issue could be resolved, this performance could improve.

When computing G^r and G^n the performance gap began to close as a result of a particularly inefficient implementation of the multiplication by Σ^{in}. This matrix is block diagonal in the original ordering, but this property is unfortunately destroyed when the rows are

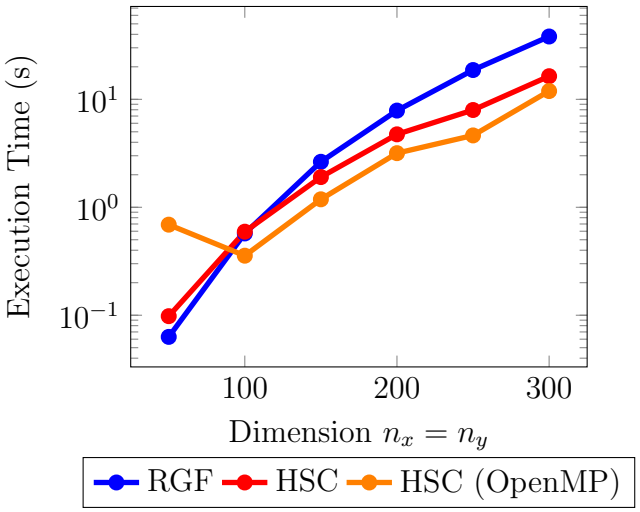


Figure 3.5: Performance of RGF, HSC, and HSC with OpenMP as a function of device size for the computation of G^r

permuted to the nested dissection ordering. A partial redesign of the data structure should enable this operation to be much faster. Due to this issue, no attempt has yet been made to parallelize the computation of G^n. For the largest size of 300×300, RGF took 122.916 s and HSC took 85.406 s. Full results are shown in Fig. 3.6.

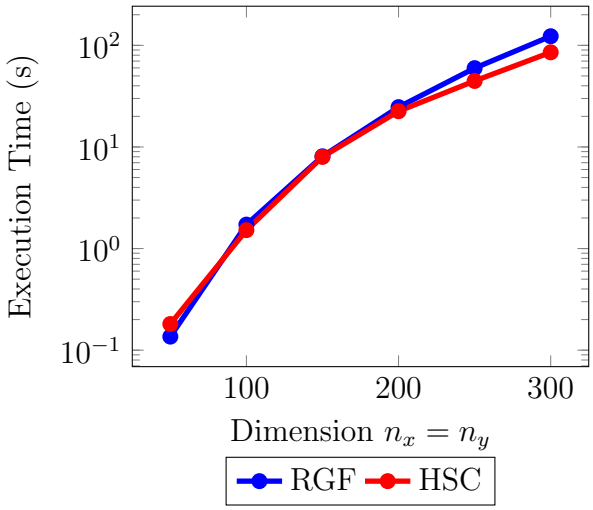


Figure 3.6: Performance of RGF and HSC as a function of device size for the computation of G^r and G^n

An additional degree of freedom over which performance could be optimized was the `TreeReduce` (TR) parameter. The value of this parameter merges the bottom TR levels

of the partition tree into a single partition. This results in trading off more floating point operations for the ability to do dense linear algebra on larger blocks resulting in less overhead for the algorithm. For the above benchmarks, $TR = 3$. Another series of benchmarks was done for TR between 1 and 8 for the 300×300 device for the calculation of only G^r with 4 OpenMP threads. For $TR = 1$ (not shown in the graph), it took 135.097 s. The optimal value was found to be $TR = 5$ with a calculation time of 6.529 s. This value of TR corresponds to a minimum block size of around 32. The value of $TR = 3$ in the previous benchmarks took 12.233 s. The optimal block size is likely to vary as a function of the structure of the device Hamiltonian as well as the machine cache size and hierarchy. Full results are shown in Fig. 3.7.

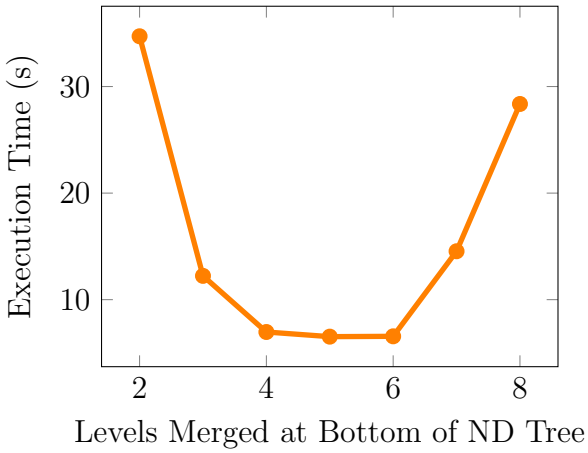


Figure 3.7: Performance of the HSC algorithm as a function of the `TreeReduce` parameter. For this simulation $n_x = n_y = 300$, and 4 OpenMP threads are used.

3.6 Future work and other methods

There are several possible directions for future work on our implementation of the HSC algorithm. The most obvious is to enhance the performance of the `BlockSparse` data structure and resolve issues with blocking in the OpenMP implementation. It would also be interesting to compare the quality of the nested dissection partitioning generated by METIS with those generated by other packages, such as PT-SCOTCH [24]. There is a new algorithm called PSelInv which extends the original SelInv algorithm to work with distributed memory parallelism [63]. Recently an approximate algorithm has been reported which uses a technique called model order reduction [59]. Such an algorithm may give substantial performance increases over a direct method without sacrificing too much accuracy.

Chapter 4

Electronic Structure and Transport Properties of Graphene Nanonribbons and Carbon Nanotubes

In this chapter we will discuss how graphene nanoribbon (GNR) and carbon nanotube (CNT) devices are simulated in BerkeleyNano3D and explore some benchmark simulations. We will begin by discussing the tight-binding model for graphene and carbon-based materials. Next, we will provide some simulations of GNR and CNT MOSFET devices using the transport simulator we described in the previous chapter. Finally, we will discuss transport through a GNR heterojunction structure.

4.1 Making one-dimensional carbon: bottom-up synthesis of graphene nanoribbons

Since graphene was first isolated by Novoselov in 2005 [104], graphene and graphene based materials have been proposed for a number of applications [128]. In this chapter, we will focus mostly on graphene nanoribbons fabricated through bottom-up chemical synthesis. The first major work on the synthesis of GNRs was that of Cai *et al.*, who synthesized $n = 7$ armchair GNRs (AGNRs) along with the chevron-type GNRs, which are the topic of the next chapter [15]. These ribbons are highly attractive for research because the chemical synthesis allows for the creation of atomically precise ribbons. Such ribbons, unlike those made from etching bulk graphene, can be made with precise edges and can possibly avoid the problems with edge roughness that have been theoretically predicted for GNR transistors [139].

Beyond making traditional devices like MOSFETs out of new materials, chemically assembled graphene nanoribbons also offer the possibility for entirely new types of devices. The assembled ribbons need not be straight semiconducting AGNRs, but the ribbons can take on new geometries previously impossible to construct in any semiconducting system at the nanometer length scale. Take for instance the 7-13 AGNR heterostructure synthesized

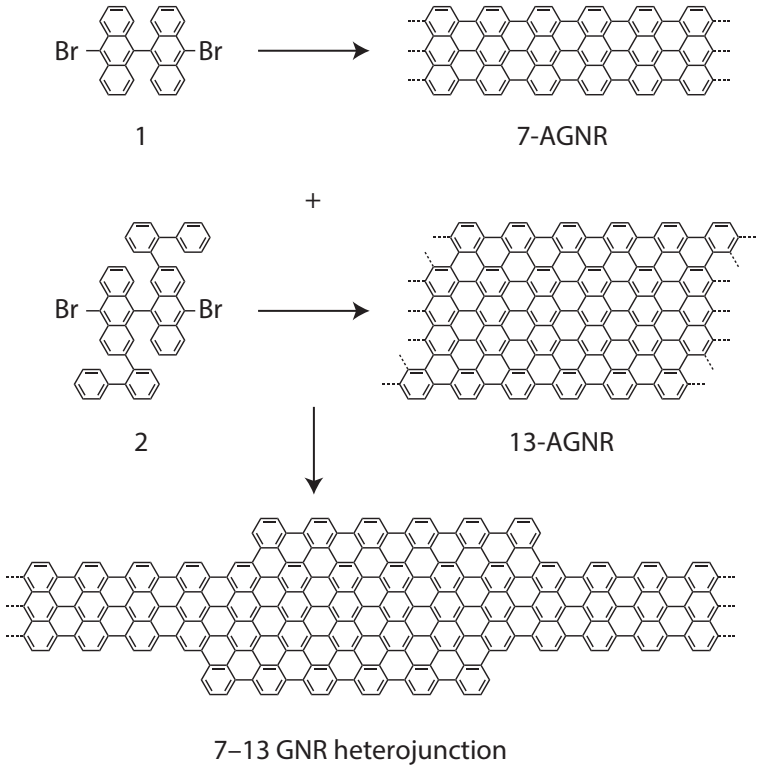


Figure 4.1: Synthesis of a 7-13 AGNR heterojunction. Figure taken from Chen *et al.*. [22].

by Chen *et al.* [22], whose synthesis is shown in 4.1 and an scanning-tunneling microscopy image of which is shown in Fig. 4.2. By varying the width of the ribbon, and hence the quantum confinement, semiconductor heterostructures with bandgap differences at ultra-short length scales can be created. This could have immense impact for technologies like heterojunction TFETs [74] and superlattice FETs [48]. Later in this chapter, we will analyze the transmission and density of states of this 7-13 heterojunction using our simulator.

Other interesting ribbons have also been created. Chen *et al.* also synthesized $n = 13$ AGNRs, which we will use as the basis for our MOSFET simulations later in this chapter [23]. Ruffieux *et al.* synthesized GNRs with zig-zag edges, theoretically predicted to be antiferromagnetic [114]. Doping through the synthesis process is also possible. Cai *et al.* synthesized heterojunctions of pristine and nitrogen-doped chevron GNRs [16]. Nguyen *et al.* made $n = 13$ sulfur-doped AGNRs [102], and Cloke *et al.* fabricated boron-doped AGNRs [26].

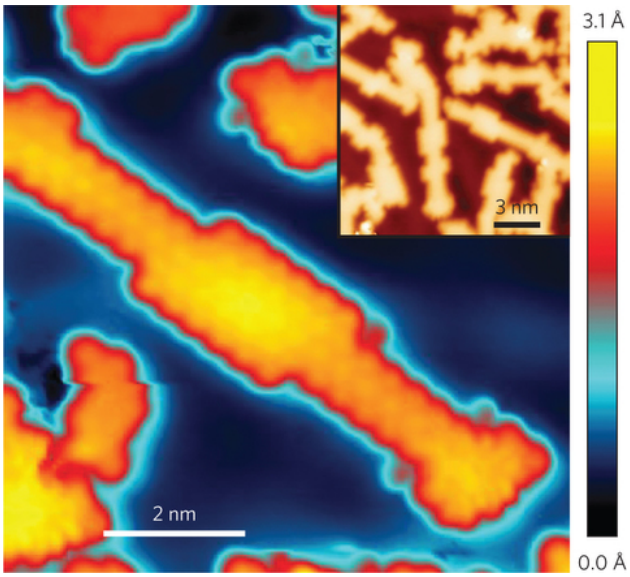


Figure 4.2: Scanning-tunneling microscopy image of a 7-13 AGNR heterojunction [22]. Figure taken from Chen *et al.*.

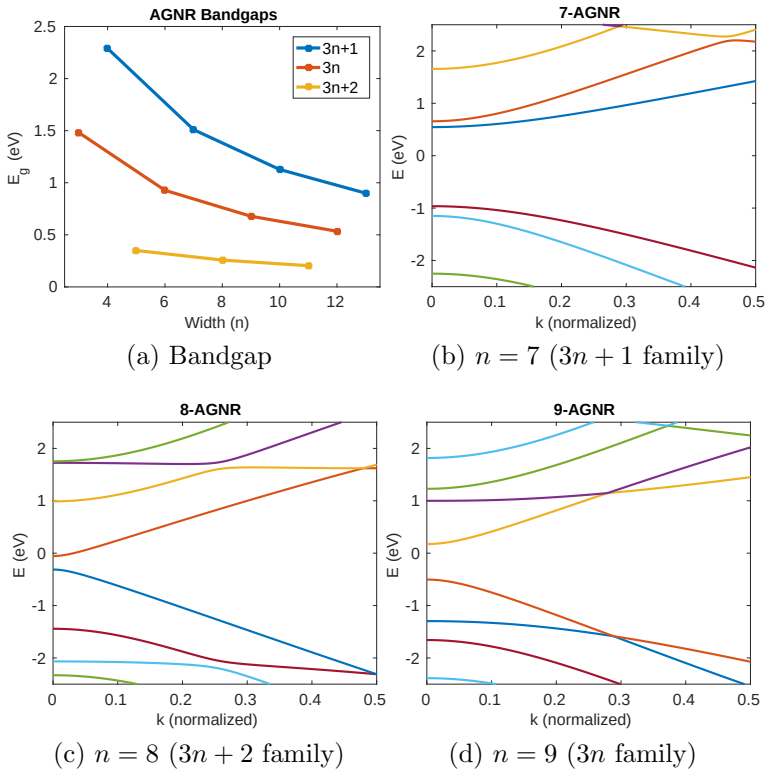
4.2 Tight-binding models for graphene nanoribbons

We model the electronic structure of graphene nanoribbons and carbon nanotubes in this chapter using tight-binding models. While *ab-initio* methods such as density functional theory with the GW correction can provide an excellent description of the electronic structure of graphene nanoribbons and can account for many-body effects [122, 136], these methods are very computationally intensive and the computed electronic structure is very hard to directly use in transport calculations.

The simplest description for the Hamiltonian of a CNT or a GNR is to assume a single p_z orbital for each carbon atom with a constant overlap term $-t_0$ with neighboring atoms and an on-site energy of zero, where $t_0 = 2.7\,\text{eV}$. This model was originally published seventy years ago in 1947 by Wallace [132], and works remarkably well for predicting many of the properties of bulk graphene, CNTs, and GNRs. It predicts the Dirac cones near the K points in bulk graphene, as well as the semiconducting nature of some GNRs and CNTs.

Despite the many successes of this description, it has several problems. It typically underestimates the bandgaps of GNRs and in some cases predicts that some semiconducting ribbons are metallic with zero band gap. One example of this is the $3n+2$ family of armchair edge graphene nanoribbons (n is number of carbon atoms along the width of the ribbon). This is because the model, originally derived for bulk graphene, does not take into account the change in electronic structure when the π network is terminated by hydrogen atoms at the edge of the ribbon.

To get around the shortcomings of the p_z description of graphene for our transport simulations, we have also implemented the tight-binding model of Boykin *et al.* for our



(a) Bandgap (b) $n = 7$ ($3n + 1$ family)

(c) $n = 8$ ($3n + 2$ family) (d) $n = 9$ ($3n$ family)

Figure 4.3: Electronic structure calculated from the p/d tight-binding model of Boykin *et al.* for several armchair-edge GNRs [14]. AGNRs can be categorized based on their width modulo 3, with the $3n + 2$ family having the lowest bandgap, the $3n$ family having the next highest bandgap, and the $3n + 1$ family having the highest bandgap.

calculations [14]. This model was fit from *ab-initio* calculations and takes into account hydrogen atoms on the edges of armchair GNRs (AGNRs) and also adds d_{yz} and d_{zx} orbitals to the p_z description of carbon. The Hamiltonian is parameterized using Slater-Koster overlap integrals [121], which are very common in NEGF device simulation. The computed bandstructures of three example graphene nanoribbons, one from each of families of GNRs categorized by the width of the ribbon modulo 3, are shown in Fig. 4.3 along with the computed bandgaps for a number of different widths. As is expected, wider ribbons have lower band gaps and the $3n + 2$ family is not metallic.

This increased accuracy comes at a cost however. Using three orbitals per atom instead of one consequently triples the dimensions of the Hamiltonian and increases the cost of doing calculations by about a factor of more 30. This factor comes from the $\mathcal{O}(n^3)$ scaling of the matrix inversion and multiplication operations that dominate the computation in the RGF algorithm as well as the need to explicitly consider the extra hydrogen atoms on the edge.

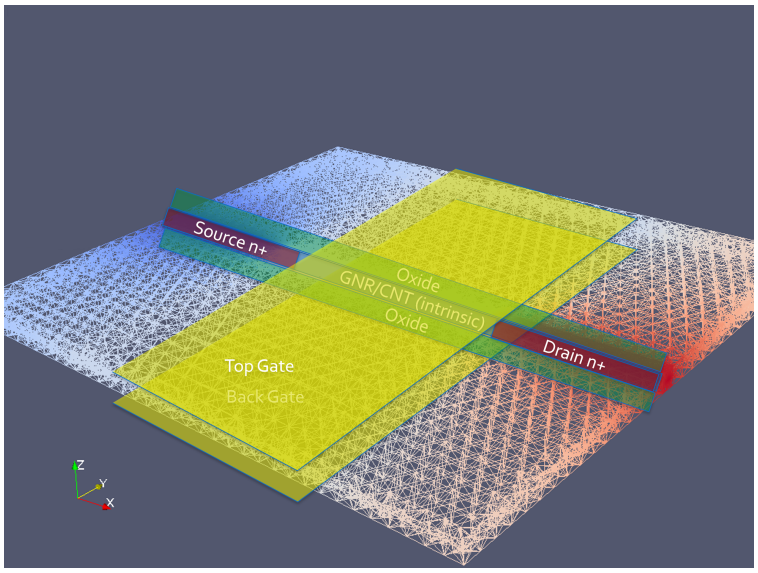


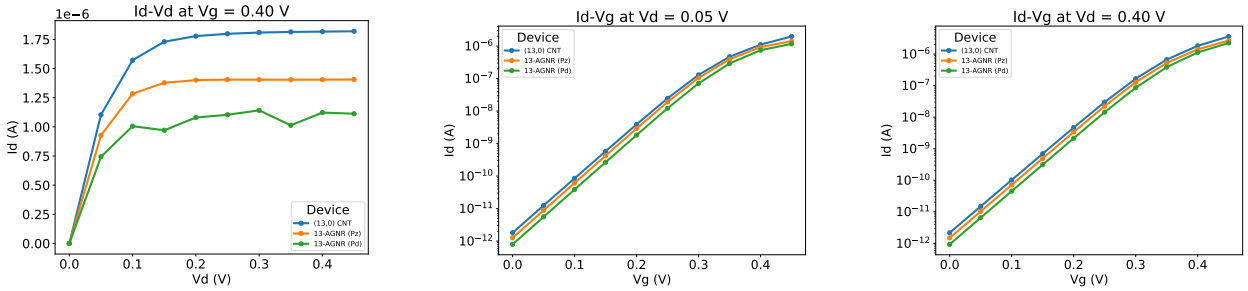
Figure 4.4: Wireframe of tetrahedral finite element mesh for calculating electrostatics for the GNR and CNT MOSFET devices generated by SALOME. Our finite element Poisson solver allows us to handle fringing fields as a result of the gate far away from the active region of the device.

4.3 MOSFET devices

To benchmark our quantum transport simulator and its ability to model graphene nanoribbon and carbon nanotube devices, we preformed several benchmark MOSFET simulations. These simulations largely replicate what has been published in literature [38, 37, 139, 52]. We simulate a dual-gate n-channel MOSFET with a channel length of 15 nm, simulated GNR/CNT length of 70 nm, source-drain donor doping of 5×10^{12} /cm^2, an intrinsic channel, and effective oxide thickness of 1 nm for both the top and bottom gates. The simulation was performed using two different channels: a 13-AGNR, and a (13,0) CNT (26 carbon atoms around circumference). Both the single p_z and p/d tight-binding models are used for the GNR, but only the p_z model is used for the CNT. The 13-AGNR was chosen because it was recently experimentally synthesized and had a low enough bandgap (0.9 eV in the p/d model) to be useful for a MOSFET. Electron-phonon scattering was not taken into account for these calculations.

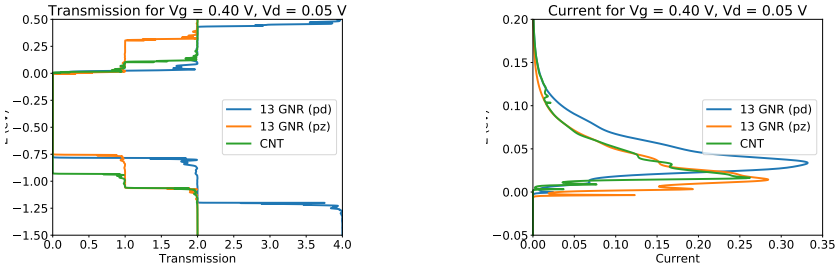
The wireframe for our mesh generated with SALOME [113] is shown in Fig. 4.4. Slight modifications to the mesh had to be made for the channel region to accommodate the different dimensions of the GNR compare with the CNT. The gate is extended beyond the channel on both sides by 30 nm to model the fringing fields that are a consequence of the gate extension beyond the channel in many experimental devices, such as those fabricated by Llinás *et al.* [92].

The $I_{ds} - V_{ds}$ and $I_{ds} - V_{gs}$ curves for the device are shown in Fig. 4.5. The ON-current



(a) $I_{ds} - V_{ds}$ for $V_{gs} = 0.40\,\mathrm{V}$ (b) $I_{ds} - V_{gs}$ for $V_{ds} = 0.05\,\mathrm{V}$ (c) $I_{ds} - V_{gs}$ for $V_{ds} = 0.40\,\mathrm{V}$

Figure 4.5: $I - V$ characteristics for GNR-based device using the p_z tight binding model.



(a) (b)

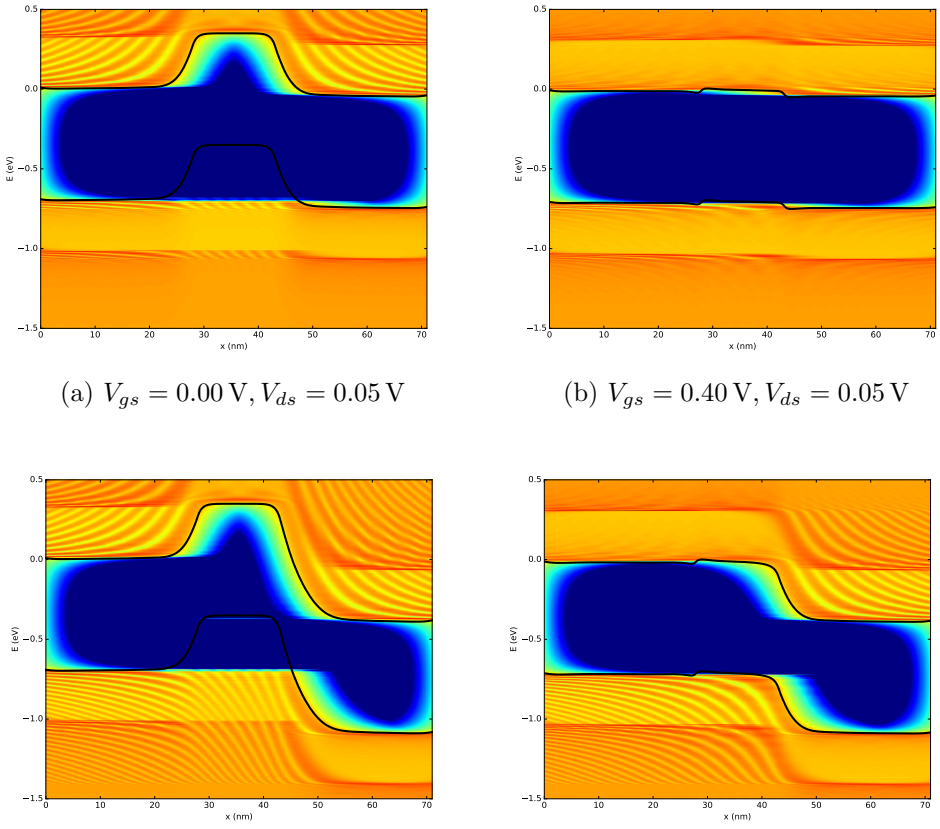
Figure 4.6: Comparison of transmission and current in the ON-state at $V_{gs} = 0.40\,\mathrm{V}, V_{ds} = 0.05\,\mathrm{V}$.

at $V_{gs} = 0.40\,\mathrm{V}, V_{ds} = 0.05\,\mathrm{V}$ is $0.94\,\mu\mathrm{A}$ for the 13-AGNR (single p_z), $0.75\,\mu\mathrm{A}$ for the 13-AGNR (p/d), and $1.11\,\mu\mathrm{A}$ for the (13,0) CNT (single p_z). At $V_{gs} = 0.40\,\mathrm{V}, V_{ds} = 0.40\,\mathrm{V}$ the ON-current is $1.40\,\mu\mathrm{A}$ for the 13-AGNR (single p_z), $1.12\,\mu\mathrm{A}$ for the 13-AGNR (p/d), and $1.179\,\mu\mathrm{A}$ for the (13,0) CNT.

Clearly, the p/d Hamiltonian gives less current than the single p_z Hamiltonian. This is because the bandgap for the p_z model is $0.7\,\mathrm{eV}$ compared with $0.9\,\mathrm{eV}$ for the p/d model. The CNT also has around a $0.9\,\mathrm{eV}$ bandgap, but it gives the highest current because its conduction band has two degenerate valleys, unlike the AGNR which only has a single conduction band valley. This can be seen very clearly when looking at the transmission at $V_{gs} = 0.40\,\mathrm{V}, V_{ds} = 0.05\,\mathrm{V}$ plotted in Fig. 4.6.

The local density of states as a function of position along the device and energy in the ON and OFF states for two different values of V_{ds} for the 13-AGNR with the single p_z basis set. These figures illustrate standard ballistic MOSFET operation and should be compared with the similar plots in the next chapter for the chevron-type GNR transistor where the density of states above the conduction band edge has abrupt gaps that are a result of the superlattice-like electronic structure.

Though it is beyond the scope of this chapter, we also were also able to simulate a *pin*



(a) $V_{gs} = 0.00\,\mathrm{V}, V_{ds} = 0.05\,\mathrm{V}$ (b) $V_{gs} = 0.40\,\mathrm{V}, V_{ds} = 0.05\,\mathrm{V}$

(c) $V_{gs} = 0.00\,\mathrm{V}, V_{ds} = 0.40\,\mathrm{V}$ (d) $V_{gs} = 0.40\,\mathrm{V}, V_{ds} = 0.40\,\mathrm{V}$

Figure 4.7: Density of states plotted as a function of position and energy in a 13-AGNR MOSFET at four biasing conditions. The conduction and valence band edges are indicated by solid black lines. The colormap is based on a logarithmic scale.

tunnel field-effect transistor (TFET) based on the 13-AGNR. The results of this simulation were consistent with those of Kim *et al.* [74]. In our ballistic calculation, this device was able to acheive 5.1 mv/decade subthreshold swing, though, as we will see in the next chapter, phonon scattering can significantly degrade steep-slope behavior.

4.4 Transmission through a 7-13 AGNR heterojunction

After simulating basic FET structures, we turned to modeling the 7-13 AGNR heterojunction originally synthesized by Chen *et al.* [22]. This ribbon is shown in Fig. 4.1. Using our NEGF solver with open boundary conditions and the p/d, we computed the transmission coefficient

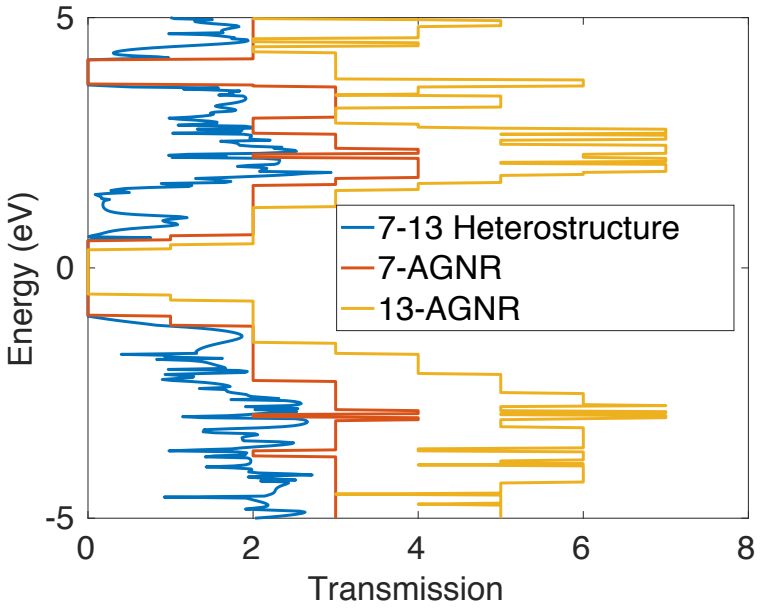


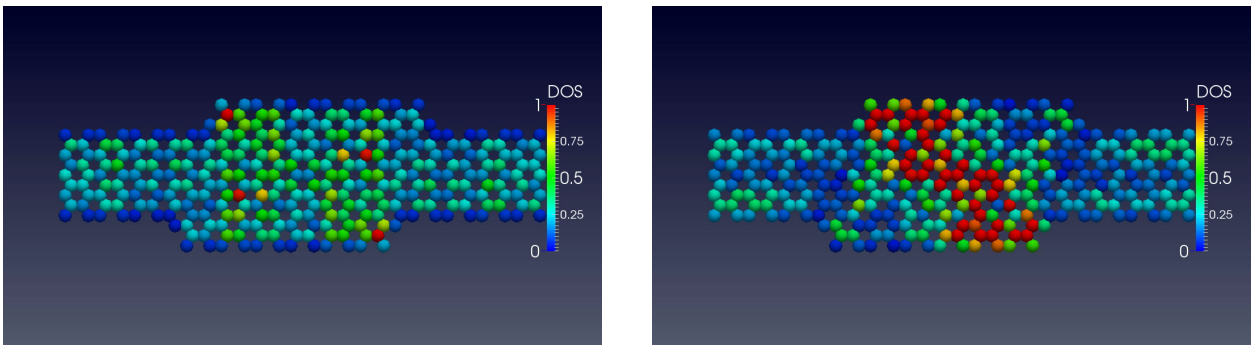
Figure 4.8: Transmission as a function of energy for the 7-13 AGNR heterojunction. The transmission is plotted for the parent 7-AGNR and 13-AGNR structures for comparison. Above the conduction band edge, there is a region for which the transmission coefficient drops dramatically. This could be used to build a device that exhibits negative differential resistance.

as a function of energy. This is shown in Fig. 4.8 along with the transmission coefficients for the parent 7-AGNR and 13-AGNR ribbons.

Of note, is that above the conduction band edge, the transmission coefficient does not monotonically increase. Transmission peaks occur when bulk 7-AGNR modes align with the allowed energies for the quantum dot-like 13-AGNR region. This property could be used to create a device that exhibits negative differential resistance (NDR) when transmission is cut off by increasing bias. Fig. 4.9 shows the local density of states at two different energies, one in which the density of states is fairly uniform across the ribbon and another where it is highly localized within the 13-AGNR region. In the next chapter, NDR is demonstrated for another GNR heterostructure, the periodic chevron GNR.

4.5 Conclusion

In this chapter, we have demonstrated the ability of the BerkeleyNano3D to simulate MOSFET devices based on graphene nanoribbons and carbon nanotubes using two different electronic structure models for GNRs. Our results for these devices are consistent with those found in literature. We have also laid the foundation for simulating interesting transport properties, including negative differential resistance, in assembled GNR devices.



(a) Delocalized state at 2.2 eV

(b) Localized state at 3.8 eV

Figure 4.9: Density of states for the 7-13 AGNR heterojunction as calculated through NEGF with open boundary conditions.

Chapter 5

Transport Properties of Chevron Graphene Nanoribbon Field-Effect Transistors

5.1 Introduction

In this chapter, we show that recently fabricated Chevron-type graphene nanoribbons (CGNRs) act as a monolithic superlattice structure. This is enabled by the large periodic unit cells with regions of different effective bandgap in these nanoribbons, resulting in minibands and gaps in the density of states above the conduction band edge. Quantum transport calculations based on the non-equilibrium Green's function formalism reveal that a negative differential resistance (NDR) is expected to manifest in these nanoribbons. Due to the relatively low density of states, such NDR behavior can also be modulated with a gate electric field. We show that a sub-thermal subthreshold swing ($< kT/q$) can potentially be obtained in a three terminal configuration, even in the presence of optical phonon scattering.

In 1970, L.Esaki and R. Tsu predicted [34] that in an appropriately made superlattice, it should be possible to obtain very narrow width bands, which could then lead to negative differential resistance. The remarkable property of these superlattices is in the fact that, unlike in Esaki diodes, this negative differential resistance does not need any tunneling, rather it comes from the direct conduction of electrons. Nonetheless, significant difficulty in synthesizing atomically precise, epitaxial heterostructures has made it very challenging to realize such superlattice structures [33, 25, 61, 12, 49, 69, 134]. Much work has been done on modeling graphene nanoribbon heterostructures and superlattices which could exhibit NDR [36, 86, 126, 119, 120, 115, 47]. Other work has been done on steep-slope devices based on GNR and CNT heterojunctions [74, 141]. Gnani *et al.* showed how superlattices could be used in a III-V nanowire FET to achieve steep-slope behavior by using the superlattice gap to filter high energy electrons in the OFF state [48]. Here, we show that the recently synthesized chevron nanoribbons [15] provide a natural, monolithic material system where narrow-width

energy bands and negative differential resistance (NDR) can be achieved. Our atomistic calculations predict that the NDR behavior should manifest at room temperature along with sub-thermal steepness (<60 mV/decade at room temperature). Such NDR behavior could lead to completely novel devices for next generation electronics.

Unlike a graphene sheet, a narrow strip of graphene, often called a graphene nanoribbon (GNR), can provide a sizable bandgap. As a result, GNRs could lead to devices with good ON/OFF ratio at the nanoscale. However, a number of studies have also shown the deleterious effect of edge roughness on the device performance [37, 139]. Recent breakthroughs in bottom-up chemical synthesis can produce GNRs with atomistically pristine edge states and overcome this shortcoming [15]. In fact, a recent experimental work demonstrated working transistors with 9- and 13-AGNRs made with these techniques [92]. The methods used to synthesize these ribbons can also be used to generate complex periodic structures beyond simply armchair and zigzag nanoribbons [16, 22]. In this work, we will consider one of those structures, the chevron graphene nanoribbon (CGNR).

Fig. 5.1(a) shows both the structure of the 6-9 CGNR originally fabricated by Cai *et al.* and the band structure calculated with a p_z orbital-based tight-binding method [15]. A key feature of the band structure is the presence of minibands with regions of forbidden energy above the conduction band edge, such as those seen in superlattices of III-V semiconductors. Analogous to III-V superlattices, the CGNR contains regions of different effective bandgap. When we look at the CGNR in Fig. 5.1(a), we see that its narrowest segment is 6 carbon atoms across and its widest segment is 9 carbon atoms across, with both segments having armchair-type edges. Using a p_z-basis set (GW [136]), the bandgap, E_g, of a 6-AGNR is 1.33 eV (2.7 eV), and the bandgap of a 9-AGNR is 0.95 eV (2.0 eV). However, given the very short length scale over which the width changes in our structure ($\sim 1\,\mathrm{nm}$), one would not expect the system to behave as though the local effective potential oscillates between the bulk values of E_g for the isolated AGNRs. In fact, our chevron structure has an overall bandgap of 1.59 eV. This value is consistent with the 1.62 eV bandgap from LDA DFT calculations, but significantly smaller than the 3.74 eV value from calculations incorporating the GW correction [133]. Both LDA and GW calculations show the presence of minibands and gaps above the conduction band edge [133].

5.2 Simulation Methods and Details

The simulated device, illustrated in Fig. 5.2, has a double gate GNR-on-insulator structure. Like a typical MOSFET, our superlattice field-effect transistor (SLFET) can be turned ON and OFF with a gate voltage at low drain biases. Operation differs from a MOSFET in two key ways. The first is that the device shows NDR with respect to the drain voltage. At some value of V_{ds} determined by the width of the first miniband, I_{ds} decreases substantially when the conduction band at the source becomes aligned with the superlattice gap at the drain. At higher drain bias, current increases again when the conduction band at the source is aligned with the second miniband at the drain. The second feature of the SLFET is that

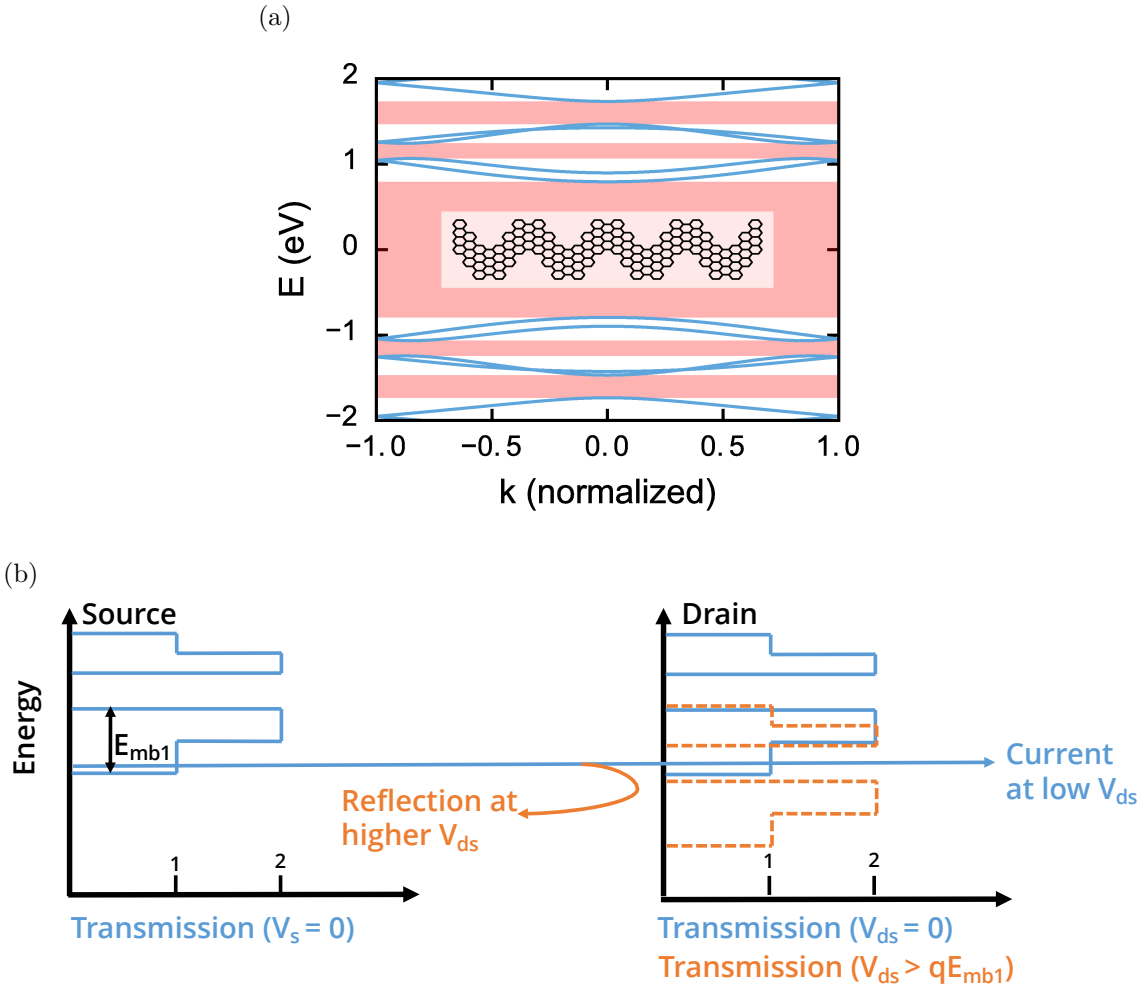


Figure 5.1: (a) Band structure of a chevron GNR based on a p_z orbital basis set. The width (and thus quantum confinement) varies across the unit cell, giving a superlattice-like band structure. Forbidden energies are highlighted in red. The bandgap of the ribbon is 1.59 eV, the first conduction band has a bandwidth of 0.272 eV, and the first gap between minibands is 0.178 eV. Inset: Chevron GNR structure. (b) When a gate voltage turns the device ON, current conduction occurs at low values of V_{ds} where the first miniband at the source is aligned with the first miniband at the drain. As the drain voltage is increased beyond qE_{mb1}, the bandwidth of the first miniband, transmission is cut off and the device exhibits NDR.

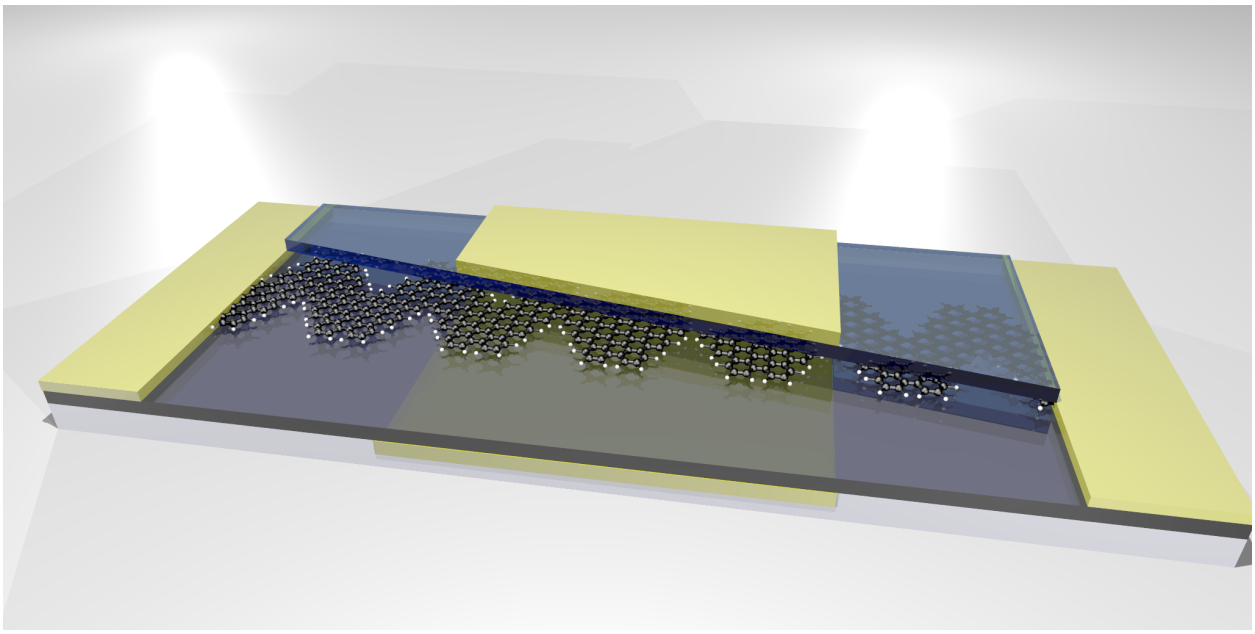


Figure 5.2: Artistic rendering of double-gate CGNR on insulator FET. Parts of the top gate and oxide region have been cut away so that the channel is visible. The device has a doped source/drain and an intrinsic channel.

the superlattice gap at the drain filters out higher energy electrons from the first miniband at the source when a source-drain bias is applied. This cuts off the higher energy portion of the thermionic tail at the source, which would contribute to leakage current in a traditional MOSFET. This filtering does not, however, affect the low-energy electrons, which carry most of the ON state current as they are in the window where the first minibands at the source and drain overlap. Transport in an SLFET is entirely intraband like a MOSFET, whereas a TFET relies on band-to-band tunneling. This could possibly allow higher ON current than a TFET. Fig. 5.1(b) illustrates the how the device exhibits negative differential resistance. At low drain biases, conduction will occur within the first miniband. When the drain voltage is increased to the point where the gap above the first miniband at the drain is aligned with the source Fermi level, conduction is cut off. With even higher drain voltage, band-to-band tunneling from the first miniband to the second miniband becomes possible, and current increases again.

The CGNR used in our simulation has a width of 1.9 nm. The simulation domain is approximately 70 nm long, and the gate has a length of 15 nm. The source and drain are doped with $N_D = 1.0 \times 10^{13}$ cm^{-2} donors. An effective oxide thickness of 1.0 nm is used for both the top and bottom gates. The gate contacts are extended 30 nm perpendicular to the channel to capture fringing gate fields. While our simulation uses an effective doping density to align the source and drain Fermi levels to the CGNR conduction band, this could be achieved in an experiment through electrostatics alone. The parameters for the simulation

Parameter	**Value**
Source/Drain doping	$1.0 \times 10^{13}\ \mathrm{cm}^{-2}$
Channel length	15 nm
Gate contact width ($\perp$)	30 nm
CGNR length	70 nm
Effective oxide thickness	1 nm
p_z hopping energy	2.7 eV
Acoustic phonon coupling, $D_{\omega,AP}$	$0.01\,\mathrm{eV}^2$
Optical phonon coupling, $D_{\omega,OP}$	$0.01\,\mathrm{eV}^2$
Optical phonon energy, $\hbar\omega_{OP}$	196 eV

Table 5.1: Simulation parameters for chevron-GNR FET simulation.

are listed in Table 5.1.

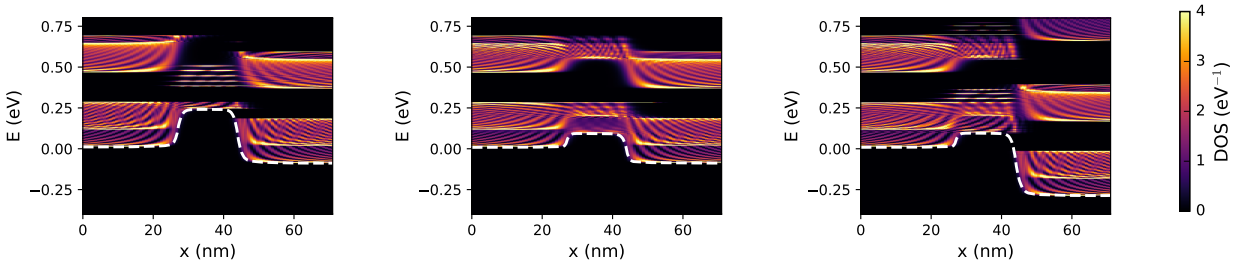
Simulations are performed using the non-equilibrium Green's function (NEGF) formalism [30]. A simple p_z-basis Hamiltonian with hopping parameter $t_0 = 2.7\,\mathrm{eV}$ is used. Charge and current are calculated with the recursive Green's function algorithm [79], and contact self-energy is computed with the Sancho-Rubio iteration scheme [118].

The NEGF equations are solved self-consistently with the nonlinear Poisson equation in three dimensions using the predictor-corrector scheme described by Trellakis *et al.* [129]. The geometry is modeled using a tetrahedral finite element mesh generated with the SALOME package [113]. The solution of the discretized Poisson equation is performed using the conjugate gradient solver from the Eigen library [51].

Acoustic and optical phonon scattering are taken in account with the according to the method described by Pal *et al.* [106], which has been used previously to calculate the effect of electron-phonon scattering in CNT and GNR devices [76, 105, 140, 142]. Because we lack an exact values for the electron-phonon coupling constants, we use approximate values similar to those used by Koswatta *et al.* for CNTs [76]. Our values are $D_{\omega,\mathrm{AP}} = 0.01\,\mathrm{eV}^2$, $D_{\omega,\mathrm{OP}} = 0.01\,\mathrm{eV}^2$, and $\hbar\omega_{\mathrm{OP}} = 196\,\mathrm{eV}$.

5.3 Results

Initially, a ballistic transport calculation excluding phonon scattering was run to establish a benchmark for the performance of the CGNR devices. Fig. 5.3 shows the density of states for several bias points. Fig. 5.3b shows peak current case when a large enough drain bias has been applied to generate enough splitting between the source and drain Fermi levels to allow significant current to flow, but not a high enough bias to move the first miniband outside of the current conduction window. For higher bias as in Fig. 5.3c, intraband conduction from the first miniband is completely cut off. As the drain bias is further increased, current can only flow due to a band-to-band tunneling from the first miniband at the source to the second miniband at the drain. Note that, due to the minibands, there will be regions



(a) $V_{gs} = 0.55\,\mathrm{V}, V_{ds} = 0.10\,\mathrm{V}$ (b) $V_{gs} = 0.70\,\mathrm{V}, V_{ds} = 0.10\,\mathrm{V}$ (c) $V_{gs} = 0.70\,\mathrm{V}, V_{ds} = 0.30\,\mathrm{V}$

Figure 5.3: Local density of states as a function of position and energy from a coherent transport calculation.

of operation for both gate and drain voltages where direct current flow is abruptly turned on or off, as the overlap between source minibands and drain minibands is modified. This leads to a steep subthreshold swing (< 60 mV/decade at room temperature) in the $I_{ds} - V_{gs}$ characteristic and a negative differential resistance in the $I_{ds} - V_{ds}$ characteristic.

Fig. 5.4(a) shows I_d vs. V_{gs} for several values of V_{ds}. While steep slope behavior is exhibited at some point for all values of V_{ds}, the highest ON current is obverved for $V_{ds} = 0.1\,\mathrm{V}$. At this drain bias, an ON current of 168 nA is achieved at a gate bias of $V_{gs} = 0.75\,\mathrm{V}$. In the steep slope region of this curve, the subthreshold swing is 6 mV/decade when averaged over five orders of magnitude of I_d. With gate work function engineering and additional device optimization, it should be possible to achieve reasonable ON current with a low supply voltage. The origin of the steep-slope behavior can be understood from Fig. 5.5. In the OFF state shown in Fig. 5.5a, the superlattice gap at the drain prevents leakage current from flowing over the source-side injection barrier. The states near the top of the barrier are seen to decay rapidly in the drain region. Fig. 5.5b shows the ON state, in which low-energy electrons, which make up virtually all of the ON current, can flow unimpeded from source to drain.

The $I_d - V_{ds}$ curves are shown in Fig. 5.4(b). When $V_{gs} = 0.7\,\mathrm{V}$, we see an increase in current up to $V_{ds} = 0.10\,\mathrm{V}$. As the drain bias is further increased, we see a decrease in current as the drain miniband goes out of alignment with the source miniband. The current beigins to pick up again as the second miniband at the drain starts to come in alignment with the source miniband again. The peak-to-valley ratio (PVR) at this gate voltage is 4.88×10^3.

Next, the effect of phonon scattering is incorporated into the calculations. The Green's functions are computed self-consistently with acoustic and optical phonon scattering using the converged potential from the coherent calculations. The local density of states calculated for the CGNR MOSFET in the presence of phonon scattering is shown in Fig. 5.5 for several biasing conditions. As is visually apparent when comparing with Fig. 5.3, phonon

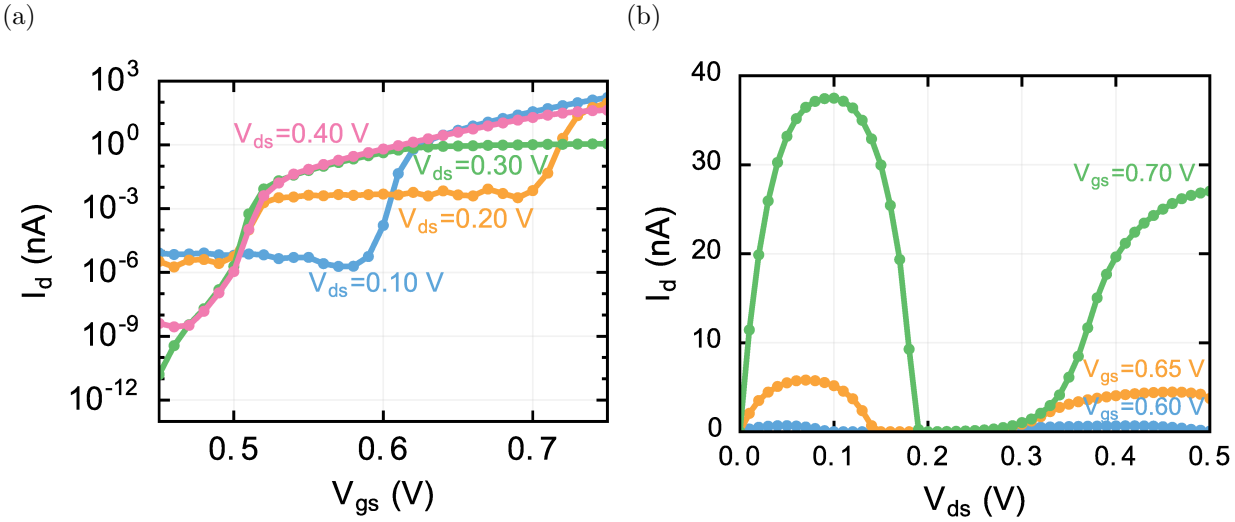
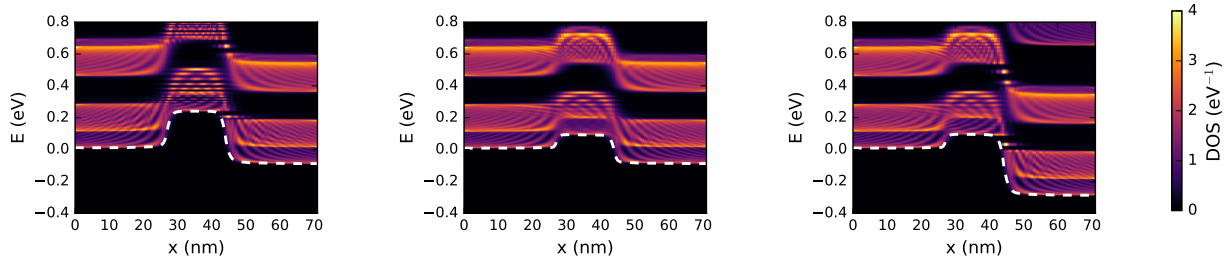


Figure 5.4: I-V curves from a coherent transport calculation. $I_d - V_{gs}$ for different values of V_{ds}. Steep-slope behavior is observed with a subtheshold swing of around 6 mV/decade over five orders of magnitude around $V_{gs} = 0.6\,\mathrm{V}$ when $V_{ds} = 0.1\,\mathrm{V}$. NDR is evident in that the peak value of I_d is lower for higher V_{ds}. (b) $I_d - V_{ds}$ for different values of V_{gs}. When $V_{gs} = 0.7\,\mathrm{V}$, a PVR of 4.88×10^3 is achieved.

scattering broadens the local density of states significantly. For the OFF state shown in Fig. 5.5a, states above the conduction band edge in the channel appear as a result of optical phonon scattering. In the ON state shown in Fig. 5.5b, the aforementioned states appear above the conduction band edge, but these no longer significantly affect device performance because ballistic transport from source to drain at the top of the barrier is permitted at this bias condition. At the higher drain bias shown in Fig. 5.5c, current is higher relative to the ballistic calculation because optical phonon scattering couples the first miniband in the channel to the first miniband in the drain at the channel-drain junction.

Figure 5.6 shows the energy-resolved current spectrum in the OFF state at $V_{gs} = 0.55\,\mathrm{V}, V_{ds} = 0.10\,\mathrm{V}$ (the same bias point as Fig. 5.5(a)). In the OFF-state, subthreshold leakage current is suppressed because of the lack of states at the drain at the energy of the top of the barrier. However, optical phonon scattering permits some carriers to be scattered down into the first miniband at the drain. Nevertheless, the steep-slope behavior is preserved, despite being deteriorated relative to 6 mV/decade subthreshold swing we observe in ballistic calculations. Previously, Yoon and Salahuddin [142] studied the effects of phonon-assisted band-to-band tunneling in armchair graphene nanoribbon tunnel transistors. That work concluded that while subthreshold swing in those devices was degraded by electron-phonon scattering, it was still possible to achieve a subthreshold swing of 34 mV/decade, far below the Boltzmann limit, consistent with the degradation in performance we see in our device when the effects of the electron-phonon interaction are taken into account.



(a) $V_{gs} = 0.55\,\mathrm{V}, V_{ds} = 0.10\,\mathrm{V}$ (b) $V_{gs} = 0.70\,\mathrm{V}, V_{ds} = 0.10\,\mathrm{V}$ (c) $V_{gs} = 0.70\,\mathrm{V}, V_{ds} = 0.30\,\mathrm{V}$

Figure 5.5: Local density of states as a function of position and energy in the presence of phonon scattering. (a) OFF state. States above the conduction band edge in the channel are a result of the optical phonon interaction. (b) ON state. Current flows as it does in the coherent transport case but with some degradation due to acoustic phonon scattering. (c) ON state for a higher value of V_{ds}. Current is higher than before as optical phonon scattering couples the first miniband in the channel to states in the drain.

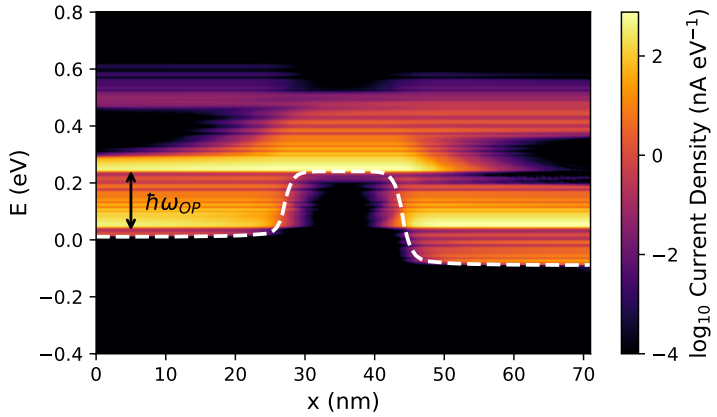


Figure 5.6: Energy-resolved current spectrum at $V_{gs} = 0.55\,\mathrm{V}, V_{ds} = 0.10\,\mathrm{V}$. Optical phonon scattering allows coupling between minibands that would not be coupled in purely ballistic transport, thereby degrading both subthreshold swing and negative differential resistance. The colormap is based on a logarithmic scale.

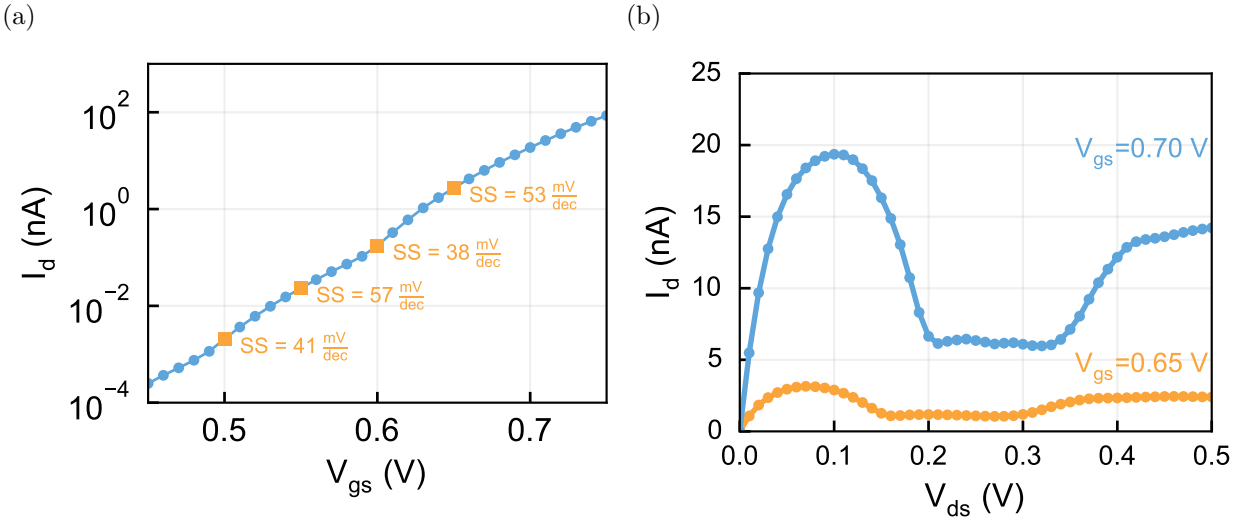


Figure 5.7: *I-V* curves from a calculation with acoustic and optical phonon scattering. (a) $I_{ds} - V_{gs}$ for $V_{ds} = 0.1\,\mathrm{V}$. Steep-slope behavior is observed with a an subthreshold swing of around of 49 mV/decade over four decades of current between $V_{gs} = 0.48\,\mathrm{V}$ and $V_{gs} = 0.68\,\mathrm{V}$ with a minimum value of 38 mV/decade around $V_{gs} = 0.60\,\mathrm{V}$. When $V_{gs} = 0.7\,\mathrm{V}$, a peak-to-valley ratio of 3.15 is achieved.

Fig. 5.7(a) shows I_{ds} vs. V_{gs} for $V_{ds} = 0.1\,\mathrm{V}$. As in the coherent transport calculation, the highest ON current is observed for $V_{ds} = 0.1\,\mathrm{V}$. At this drain bias, an ON current of 85 nA is achieved at a gate bias of $V_{gs} = 0.75\,\mathrm{V}$. This is degraded relative to the ballistic calculation primarily as a result of acoustic phonon scattering. The subthreshold swing is 49 mV/decade when averaged over four orders of magnitude of I_{ds} between $V_{gs} = 0.48\,\mathrm{V}$ and $V_{gs} = 0.68\,\mathrm{V}$ at $V_{ds} = 0.1\,\mathrm{V}$ with a minimum value of 38 mV/decade around $V_{gs} = 0.60\,\mathrm{V}$.

The $I_{ds} - V_{ds}$ curves are shown in Fig. 5.7(b). As with the ballistic calculation, When $V_{gs} = 0.7\,\mathrm{V}$, we see an increase in current up to $V_{ds} = 0.10\,\mathrm{V}$. The peak-to-valley ratio (PVR) at this gate voltage is 3.15. At $V_{gs} = 0.65\,\mathrm{V}$, the calculated PVR is 2.89. While these values are far worse than those predicted in the calculation without optical phonon scattering, it should still be possible to make a useful device with PVR ≈ 3.

5.4 Conclusion

Previous work on III-V MOSFETs with a superlattice source filter by Lam *et al.* [80] found that it should be possible to achieve $I_{ON} = 0.81\,\mathrm{mA\,\mu m^{-1}}$ with $SS = 20.9\,\mathrm{mV/decade}$ at $V_{ds} = 0.6\,\mathrm{V}$, though this calculation did not take into account the effects of phonon scattering. Assuming a device with multiple ribbons in parallel with a pitch of 5 nm, our device achieves $I_{ON} = 0.017\,\mathrm{mA\,\mu m^{-1}}$. While this value is much smaller, our device operates $V_{ds} = 0.1\,\mathrm{V}$.

Because of the NDR effect, higher drain biases unfortunately result in lower current.

In summary, we have shown that CGNR devices can exhibit both steep-slope subthreshold behavior and negative differential resistance. Both properties are the result of the superlattice-like electronic structure of the ribbon. CGNR SLFETs could be promising for a number of applications ranging from low-power logic transistors to high speed oscillators. A major obstacle to building a real device is making contacts with appropriate Schottky barrier heights to be able to match the band alignment conditions achieved in this work through doping. The performance of a real device could also be impacted by edge roughness, which we have not considered here, though the ability to synthesize ribbons with virtually no defects may minimize this effects. Additional optimization is also likely necessary to make a functioning device. DFT+GW calculations predict a much higher bandgap for the CGNR in vacuum than the tight-binding model used in this work. While surface screening may reduce the bandgap, a wider ribbon with a narrower bandgap may be required. Co-optimization of the bandgap with the bandwidths of the minibands and the gaps between minibands is also a necessary topic for future work.

Chapter 6

Phase-Field Simulator

6.1 Introduction

The phase-field method describes the time evolution of systems where two or more phases exist simultaneously [109]. The distinguishing feature of phase-field models compared to other methods for modeling multiphase systems is that phase-field models use a continuum field to track the local value of the order parameter. At a phase boundary, the order parameter varies continuously. The advantage to this is that the interface does not have to be tracked explicitly, thereby allowing the model to be easily implemented without having to store complex information about the geometry of the interface. Phase-field models can take into account many contributions to the free energy of a system and minimize total free energy through a gradient descent of the free energy functional with respect to the order parameter field.

When the phase-field method is applied to ferroelectrics, the order parameter is the polarization P, typically specified to be a vector with $\vec{P} = (P_x, P_y, P_z)$ [21].

The time evolution of the system is given the time-dependent Ginzburg-Landau (TDGL) equation [21]:

$$\frac{\delta \vec{P}}{\delta t} = -L\frac{\delta F}{\delta \vec{P}} = -L\frac{\delta F_{local} + \delta F_{wall} + \delta F_{elec} + F_{elas}}{\delta \vec{P}} \tag{6.1}$$

This equation is the gradient flow of the Ginzburg-Landau free energy functional for ferroelectrics, which includes the local Landau-Devonshire energy, a gradient term corresponding to domain walls, a long-range electrostatic interaction, and a long range elastic interaction. In this chapter, we will discuss each of the free energy terms in detail. In some phase-field literature, the TDGL is sometimes called the Allen-Cahn equation [109].

6.2 Code overview

The phase-field simulator is based on the original code developed by Khalid Ashraf, which is described in great detail in his dissertation [9]. Unlike the previous version of the simulator,

support for large-scale parallel (MPI) computing has been removed due to the difficulty in maintaining this part. While this is useful for simulating large micron scale ferroelectric capacitors, capacitors with dimensions in the 10-100 nm size range can be easily simulated on standard multicore workstations with adequate speed. Despite the removal of support for MPI-based parallel computing, several new features have been added:

1. Complete rewrite of code in object-oriented C++11 with the Eigen template library [51].
2. Paralution algebraic multigrid and Krylov subspace solvers, which can run on multicore CPUs and GPUs [78].
3. Support for Sundials CVODE and IDA time-steppers [55].
4. JavaScript Object Notation (JSON) based input file.
5. Circuit-based models for negative capacitance with external resistors and capacitors.
6. New finite difference Poisson solver with support for embedding a semiclassical model for semiconductor charge for negative capacitance field-effect transistor (NC-FET) electrostatics.
7. Two defect models: fixed polarization and reduced coercive field.
8. Randomized Voronoi tessellation for generating random initial conditions and defect positions.

The key operation of the phase-field simulator is shown in figure 6.1. First the simulator reads the inputs from a material parameters file, a simulation configuration file, and an optional initial polarization file to continue from a previous simulation. After the initial conditions have been defined, the simulator evaluates the time derivative of polarization dP/dt as a result of several forcing terms including: local energy, domain wall energy, electrostatics, and elasticity. When dP/dt has been computed, a time-stepper integrates the TDGL and computes a new value of polarization. The force computation and time-stepping continues in a loop until the conclusion of the simulation. During this process, several outputs can be generated at the request of the user: polarization, electrostatic potential, charge, stress, and strain. Additionally, the mean polarization along each axis as well as the contact voltages (if enabled) are written to a log file at each time-step. We will now describe each part of the simulator in greater detail. Further details of the original implementation can be found in the work of Ashraf *et al.* [7, 8, 9].

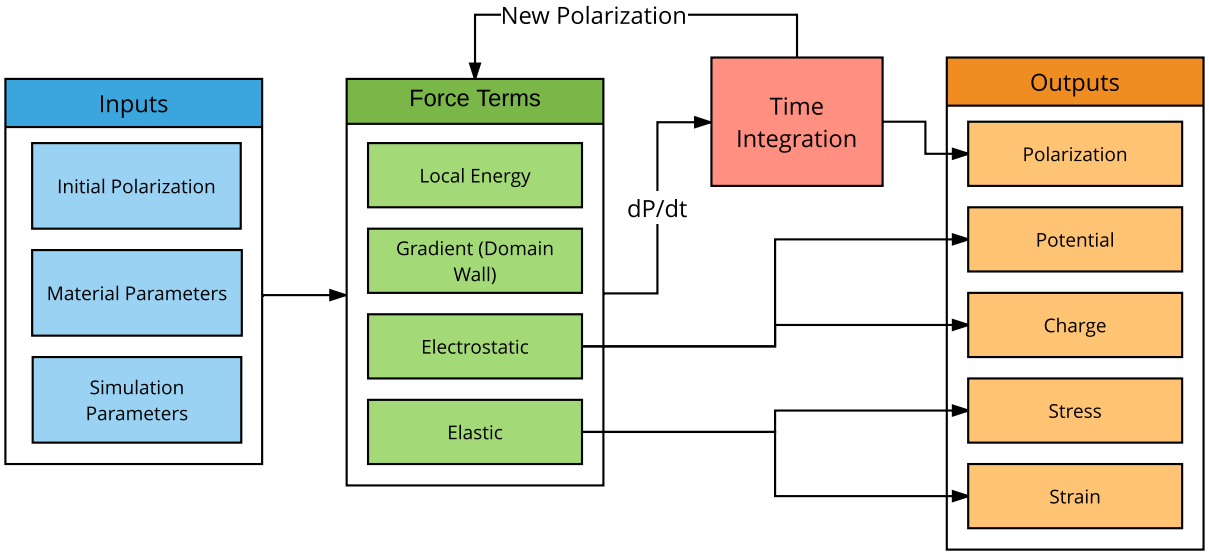


Figure 6.1: Self-consistency loop for coherent transport calculations.

Polarization field and initial conditions

The polarization field is stored in a `VectorField` object, which specifies the vector polarization on a finite difference grid with regular spacing along the x, y, and z-axes.

The initial conditions for the system are specified in the `ICSetup` portion of the input file. Various patterns can be specified, such as stripe domains and a centered region with a different polarization than the surrounding region. Random patterns can be specified such that every point takes on a random polarization. After some time for stable domains to form, these random patterns can be considered analogous to those in a ferroelectric cooled below its Curie temperature before any field has been applied. Additionally, initial domains can be generated based on a Voronoi tessellation of random points within the simulation domain. This feature will be discussed in greater detail in a later section.

Another option to specify the initial conditions is to restart from an existing polarization file. This is useful to continue from an existing simulation that has crashed or to run several simulations with different conditions after some period of common conditions. Custom initial conditions can also be created using Octave/MATLAB or another scripting language by writing a VTK file of the same format that the phase-field simulator writes. Full documentation of the possible initial conditions is given in the `ICSetup` section of `SolverConfig.h`.

Time-stepping

Given, the polarization at time t, $P(t)$, the force terms are evaluated and a `VectorField` with the values of dP/dt at every point is generated by summing the contributions from all

the individual force terms. This information is passed to a time-stepper, which may then request computation of dP/dt at additional partial time-steps in order to produce a more accurate estimate of $P(t+\Delta t)$ where Δt is a fixed time-step size. Support for several different time-steppers is provided at compile time via preprocessor options.

The first time-stepper is a second-order Verlet algorithm based on the one implemented by Ashraf [9]. This method typically produces far more accurate results than simple forward-Euler based methods, while remaining fairly simple to implement. There are, however, several issues with this method. The first is that the solver requires using a fixed time-step size. For times when polarization is changing rapidly, very short time-steps have to be used for the solution to be stable. These time-steps are small relative to what may actually be required for stability during other times in the simulation, so significant extra work may be required.

To get around the limitations of our simple algorithm, we implemented an interface to the Sundials package package from Lawrence Livermore National Laboratories [55]. The CVODE solver supports variable-order, variable-step methods to solve stiff and non-stiff ordinary differential equation systems in the form $y' = f(t, y)$. This solver is the preferred option for most problems. We also support the Sundials IDA solver which solves implicit systems in the form $y' = f(t, y, y')$. This is necessary for accurately solving for the dynamics of a ferroelectric in series with a resistor, whose voltage depends on current, which is $\int d\Omega \; dP_z/dt$, the volume integral of the rate change in polarization over time.

Landau local energy

The local energy term is solely a function of the local polarization $\vec{P} = (P_x, P_y, P_z)$. The free energy is expressed as a polynomial function of the components of $\vec{P}$. The free energy expression including up to fourth-order terms is:

$$\begin{aligned} F_{local} = \int_\Omega [&\alpha_1 \left(P_x^2 + P_y^2 + P_z^2\right) \\ &+ \alpha_{11} \left(P_x^4 + P_y^4 + P_z^4\right) \\ &+ \alpha_{12} \left(P_x^2 P_y^2 + P_y^2 P_z^2 + P_z^2 P_x^2\right) + ...] \; d\Omega \end{aligned} \tag{6.2}$$

where the α values are Landau-Devonshire free energy coefficients. Typically the the values for the α coefficients fit from experimental thermodynamic data or *ab-initio* simulation results. Values for several common ferroelectrics are found in the appendix of Rabe, *et al.* [112]. The phase-field simulator support polynomial expressions for local free energy up to eighth-order, which are specified in the material input file.

This expression can simply be plugged into the time-dependent Ginzburg-Landau equation, and we can write the the time derivative of the x-component of polarization at grid point i considering terms up to fourth order as:

$$\frac{\partial P_{i,x}}{\partial t} = -L \frac{\partial F}{\partial P_{i,x}} = 2\alpha_1 P_{i,x} + 4\alpha_{11} P_{i,x}^3 + 2\alpha_{12} P_{i,x}(P_{i,y}^2 + P_{i,z}^2) + ... \tag{6.3}$$

with similar expressions for the the y and z-components.

Gradient (domain wall) energy

The energy of a domain wall between regions of differing polarization is the result of dipole-dipole interaction and strain. This energy can be approximated as an isotropic domain wall energy constant G_{11} multiplied by the sum of the gradients in every direction for every component of $\vec{P}$. The free energy can be written as:

$$\begin{aligned} F_{wall} = \int_{\Omega} G_{11}(&P'_{x,x} + P'_{x,y} + P'_{x,z} + P'_{y,x} + P'_{y,y} \\ &+ P'_{y,z} + P'_{z,x} + P'_{z,y} + P'_{z,z})\, d\Omega \end{aligned} \tag{6.4}$$

Its associated contribution to the TDGL is:

$$\frac{\delta F_{grad}}{\delta P_i} = G_{11}\nabla^2 P_i \tag{6.5}$$

The current implementation computes this expression on the finite difference polarization grid and adds this contribution to the total forcing term at each time-step. Currently, there is only support for isotropic domain wall energy, though adding anisotropic domain wall energy could be done in the future with little difficulty should the need for it arise.

Electrostatics

The electrostatic contribution to total free energy and contribution to the TDGL at grid point i can be expressed as:

$$F_{elec} = \int_{\Omega} -\vec{E} \cdot \vec{P}\, d\Omega \tag{6.6}$$

$$-\frac{\delta F_{elec}}{\delta \vec{P}_i} = \vec{E}_i \tag{6.7}$$

The phase-field simulator provides two implementations of this term. The first, `SimpleEField`, simply assumes a uniform electric field at every point in the simulation domain. This field can be set to any orientation, and a number of waveform options are available. This method is fast and provides a reasonable approximation for the electric field in cases where an externally applied field dominates contributions to the electrostatic interactions in the system.

The second implementation, `ElectrostaticForce`, uses a finite difference Poisson solver to calculate the potential and electric field at every point in the simulation domain. As discussed in Chapter 2, the Poisson equation is written as:

$$\nabla \cdot \epsilon \nabla \phi = -\rho \tag{6.8}$$

and can be solved with either the finite element or finite difference method. The charge ρ is calculated from the ferroelectric polarization using:

$$\rho = \nabla \cdot \vec{P} \tag{6.9}$$

Charged domain walls can influence the electric field even at long distances and must therefore be taken into account with a full solution of Poisson's equation.

The Poisson solver, implemented in the `FDPoisson3D` class, is a modified version of the finite difference Poisson solver that was used in our quantum transport simulator before development switched to a finite element-based solver as was described in Chapter 2. The discretized version of Poisson's equation is solve using an algebraic multigrid solver from the Paralution package [78]. Multigrid methods can be used to solve linear systems derived from the discretization of elliptic partial differential equations, such as Poisson's equation, in $\mathcal{O}(n)$ time complexity [32]. The Poisson solver is interfaced with the `ElectrostaticForce` class that defines the solver geometry and computes the final force term for the time-stepper.

The contact voltages at the top and bottom of the ferroelectric capacitor are incorporated as Dirichlet boundary conditions. The top contact can be underlapped in order to simulate the fringing fields that occur when a small area top contact is used with a back contact covering the entire area of ferroelectric. At the sides of the capacitor, both periodic and Neumann (isolated) boundary conditions can be applied.

In order to simulate negative capacitance devices with a ferroelectric epitaxially grown on a dielectric, the Poisson solver has an option to simulate an additional dielectric region between the ferroelectric region and the bottom contact. There is an also option to simulate this region as a semiconductor with charge computed from a semiclassical expression that assumes parabolic conduction and valence bands and a two-dimensional semiconductor (constant density of states within the bands).

External resistors and capacitors

One of the key features of the phase-field simulator is the ability to simulate the response of a multidomain ferroelectric capacitor in series with an external resistor or capacitor. This is useful for comparison with several experimental demonstrations of negative capacitance. The formulation for a series capacitor is covered in Chapter 7 and the formulation for a series resistor is covered in Chapter 8. In both cases, the effect of the external circuit is taken into account by setting the contact voltages in the electrostatics solver to values consistent with the voltage across the ferroelectric capacitor in the circuit model. When an external resistor is used, only the Sundials IDA time-stepper can be used because it is an implicit ODE solver which it provides an estimate of dP/dt. This is required because the voltage across the resistor is a function of the current, which is $-dP/dt$. There are no-restrictions on which time-stepper can be used in the case of an external capacitor.

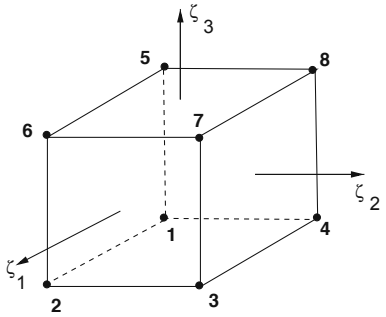


Figure 6.2: Hexahedral finite element with 8 nodes. Graphic taken from Zohdi [144].

Elasticity

Overview

Long-range elastic interactions can affect how domains evolve over the course of a simulation and which orientations are preferred for domains. In our simulation, we compute the effect of the elastic interaction according to the following process:

1. Compute the load vector for the body stress forces as a result of local polarization $\vec{P}(\mathbf{x})$.

2. Solve for the strain developed as a result of those stress forces by solving a linear elasticity problem.

3. Compute the contribution to the TDGL as a result of local strain.

We employ the finite element method (FEM) to solve the linear elasticity problem [144]. Unlike the FEM electrostatics solver described in Chapter 2 for the transport simulator, we do not use an externally generated mesh, but rather use a regular mesh of hexahedral ("brick" or "cube") elements. This is because ultimately we must take polarization values from the regularly spaced, rectilinear finite difference grid on which polarization is defined. In our model each, finite difference grid point is represented by a single hexahedral element with eight nodes. Because of the extra nodes on the boundaries of the final elements, the finite element grid has $(n_x+1)\times(n_y+1)\times(n_z+1)$ nodes instead of $n_x \times n_y \times n_z$ grid points, where n_x, n_y, and n_z are the dimensions of the finite difference polarization grid.

Weak form and discretization

First we formulate the weak form of the equation for for linear elasticity according the the procedure of Zohdi [144]. We begin with the strong form of the equation of motion (a form

of Newton's second law) and the physical constitutive law:

$$\nabla \cdot \boldsymbol{\sigma} + \mathbf{F} = \mathbf{0} \tag{6.10}$$

$$\boldsymbol{\sigma} = \mathbf{E} : \nabla \mathbf{u} \tag{6.11}$$

where $\mathbf{E}$, the stiffness tensor can be spatially variable. The body force per unit volume is given by $\mathbf{F} = \mathbf{F}(x, y, z)$, the stress tensor is $\boldsymbol{\sigma}$, and the displacement is $\mathbf{u}$. This linear equation is valid in the limit of a small, infinitesimal deformation. In our simulation, the stiffness tensor is constant across space because we only solve this equation in ferroelectric, but the simulator could be extended to allow simulating the substrate and surrounding media. The stiffness tensor is calculated from the Young's modulus E and the Poisson ratio ν.

We multiply by a smooth test function $\mathbf{v}$ that is at least once differentiable and integrate over the domain:

$$\int_\Omega (\nabla \cdot \boldsymbol{\sigma} + \mathbf{F})\mathbf{v} \, d\Omega = \int_\Omega \mathbf{r} \cdot \mathbf{v} \, d\Omega = 0 \tag{6.12}$$

Taking the condition that this hold for every $\mathbf{v}$ where $\mathbf{v}|_{\Gamma_u} = 0$, where Γ_u is the region of the domain over which Dirichlet boundary conditions (fixed displacement) apply, this allows us to conclude that the residual $\mathbf{r} = 0$. This is the standard condition applied in Galerkin finite element methods. Using the product rule of differentiation for vector valued functions, integrating with the divergence theorem, and substituting the constitutive law gives the final version of the weak form:

$$\int_\Omega \nabla \mathbf{v} : \mathbf{E} : \nabla \mathbf{u} \, d\Omega = \int_\Omega \mathbf{F} \cdot \mathbf{v} \, d\Omega \tag{6.13}$$

For $\mathbf{u}$ satisfying the clamped (Dirichlet) boundary conditions Γ_u. Fixed boundary conditions consistent with the fractional strain specified in the input file are applied on the bottom surface of the ferroelectric, though this can be modified through the `IsDirichletBC` method of `ElasticForce`.

Next, we discretize the weak form of this equation with finite elements. We begin by assuming the media is isotropic. This allows us to reduce the number of elements in the stiffness tensor from 81 independent elements down to just 6. This allows us to write Eq.6.13 as:

$$\int_\Omega ([\mathbf{D}]\{\mathbf{v}\})^T [\mathbf{E}]([\mathbf{D}]\{\mathbf{u}\}) \, d\Omega = \int_\Omega \{\mathbf{v}\}^T \{\mathbf{F}\} \, d\Omega \tag{6.14}$$

where $\{\mathbf{u}\}$, $\{\mathbf{v}\}$, and $\{\mathbf{F}\}$ are three component vector-valued functions and:

$$[\mathbf{D}] = \begin{bmatrix} \frac{\partial}{\partial x_1} & 0 & 0 \\ 0 & \frac{\partial}{\partial x_2} & 0 \\ 0 & 0 & \frac{\partial}{\partial x_3} \\ \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_1} & 0 \\ 0 & \frac{\partial}{\partial x_3} & \frac{\partial}{\partial x_2} \\ \frac{\partial}{\partial x_3} & 0 & \frac{\partial}{\partial x_1} \end{bmatrix} \tag{6.15}$$

We will now express $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$ in some finite element basis.

As $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$ describe three component vector fields, we cannot directly assign a and b coefficients to multiply the shape functions as with scalar field problems, such as the electrostatic Poisson problem described in Chapter 2. Define the $3 \times 3N$ matrix of shape functions $[\boldsymbol{\phi}]$ as:

$$[\boldsymbol{\phi}] = \begin{bmatrix} \phi_1 & \cdots & \phi_N & 0 & \cdots & 0 & \cdots & 0 & \\ 0 & \cdots & 0 & \phi_1 & \cdots & \phi_N & 0 & \cdots & 0 \\ 0 & \cdots & & 0 & \cdots & 0 & \phi_1 & \cdots & \phi_N \end{bmatrix} \tag{6.16}$$

Where the rows correspond to the first second and third component of some vector formed from the linear combination of these shape functions. Accordingly we express $\{\mathbf{v}\}$ and $\{\mathbf{u}\}$ as:

$$\{\mathbf{v}\} = [\boldsymbol{\phi}]\{b\}$$
$$\{\mathbf{u}\} = [\boldsymbol{\phi}]\{a\}$$

The weak form can now be written as:

$$\int_\Omega ([\mathbf{D}][\boldsymbol{\Phi}]\{b\})^T[\mathbf{E}]([\mathbf{D}][\boldsymbol{\Phi}]\{a\})\, d\Omega = \int_\Omega [\boldsymbol{\Phi}]\{b\}^T\{\mathbf{F}\}\, d\Omega \tag{6.17}$$

We substitute the 3D trilinear Lagrange shape functions over the hexahedral volume shown in Fig. 6.2:

$$\hat{\phi}_1 = \frac{1}{8}(1-\zeta_1)(1-\zeta_2)(1-\zeta_3)$$
$$\hat{\phi}_2 = \frac{1}{8}(1+\zeta_1)(1-\zeta_2)(1-\zeta_3)$$
$$\hat{\phi}_3 = \frac{1}{8}(1+\zeta_1)(1+\zeta_2)(1-\zeta_3)$$
$$\hat{\phi}_4 = \frac{1}{8}(1-\zeta_1)(1+\zeta_2)(1-\zeta_3)$$
$$\hat{\phi}_5 = \frac{1}{8}(1-\zeta_1)(1-\zeta_2)(1+\zeta_3)$$
$$\hat{\phi}_6 = \frac{1}{8}(1+\zeta_1)(1-\zeta_2)(1+\zeta_3)$$
$$\hat{\phi}_7 = \frac{1}{8}(1+\zeta_1)(1+\zeta_2)(1+\zeta_3)$$
$$\hat{\phi}_8 = \frac{1}{8}(1-\zeta_1)(1+\zeta_2)(1+\zeta_3)$$

Taking our final finite element weak form and noting that the $\{b\}$ vector is arbitrary by construction, we can thus get a simplified matrix form:

$$\{b\}^T\{[K]\{a\} - \{R\}\} = 0 \tag{6.18}$$
$$[K]\{a\} = \{R\} \tag{6.19}$$

$$[K] \equiv \int_{\Omega} ([\mathbf{D}][\Phi])^T [\mathbf{E}]([\mathbf{D}][\mathbf{\Phi}])\, d\Omega \tag{6.20}$$

$$\{R\} \equiv \int_{\Omega} [\mathbf{\Phi}]^T \{\mathbf{F}\}\, d\Omega \tag{6.21}$$

This gives a general form for the stiffness matrix and right hand side vector, but to explicitly construct these we must perform global/local transformations. We define our coordinate mapping as follows:

$$x_i = \sum_{j=1}^{8} \chi_{ij} \hat{\phi}_j \tag{6.22}$$

where χ_{ij} is the i coordinate for node j in this element. So that we can do our calculations over the master element and the $\hat{\phi}$ functions rather than in x space over the ϕ functions, we need the Jacobian J for the mapping, which we take as the determinant of the deformation gradient F (not to be confused with the force $\mathbf{F}$):

$$J \equiv |F| = \begin{vmatrix} \frac{\partial x_1}{\partial \zeta_1} & \frac{\partial x_1}{\partial \zeta_2} & \frac{\partial x_1}{\partial \zeta_3} \\ \frac{\partial x_2}{\partial \zeta_1} & \frac{\partial x_2}{\partial \zeta_2} & \frac{\partial x_2}{\partial \zeta_3} \\ \frac{\partial x_3}{\partial \zeta_1} & \frac{\partial x_3}{\partial \zeta_2} & \frac{\partial x_3}{\partial \zeta_3} \end{vmatrix} \tag{6.23}$$

The 24×24 blocks of the stiffness matrix for a particular element volume ω can now be written as:

$$K_{ij} = \int_{\omega} ([\hat{\mathbf{D}}][\hat{\mathbf{\Phi}}][F]^{-1})^T [\mathbf{E}]([\hat{\mathbf{D}}][\hat{\mathbf{\Phi}}][F]^{-1})\, |F|\ d\omega \tag{6.24}$$

where

$$[\hat{\phi}] = \begin{bmatrix} \hat{\phi}_1 & \cdots & \hat{\phi}_8 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \hat{\phi}_1 & \cdots & \hat{\phi}_8 & 0 & \cdots & 0 \\ 0 & \cdots & & 0 & \cdots & 0 & \hat{\phi}_1 & \cdots & \hat{\phi}_8 \end{bmatrix} \tag{6.25}$$

and

$$[\hat{D}] = \begin{bmatrix} \frac{\partial}{\partial \zeta_1} & 0 & 0 \\ 0 & \frac{\partial}{\partial \zeta_2} & 0 \\ 0 & 0 & \frac{\partial}{\partial \zeta_3} \\ \frac{\partial}{\partial \zeta_2} & \frac{\partial}{\partial \zeta_1} & 0 \\ 0 & \frac{\partial}{\partial \zeta_3} & \frac{\partial}{\partial \zeta_2} \\ \frac{\partial}{\partial \zeta_3} & 0 & \frac{\partial}{\partial \zeta_1} \end{bmatrix} \tag{6.26}$$

To construct the full sparse matrix from the local element matrices, the elements of the matrix are added to the global rows/columns corresponding to the global node numbers for the eight nodes of the matrix. Note that for each node there are three components. Note that this is an "add" operation and not a "set" operation as each node may be a member of multiple elements.

Load vector

The load vector $\{b\}$ is constructed in a somewhat analogous way.

$$b_i = \int_\omega [\hat{\mathbf{\Phi}}]^T [\mathbf{F}] \, |F| \; d\omega \tag{6.27}$$

Here, the body force is computed by computing the fractional strain for an isolated element with polarization $\mathbf{P}$ and then back-computing the force that would lead to that strain [9]. The diagonal components of the local strain tensor are given by:

$$\epsilon^0_{ii} = Q_{11} P_i^2 + Q_{12}(P_j^2 + P_k^2) \tag{6.28}$$

The off-diagonal components are given by:

$$\epsilon^0_{ij} = Q_{44} P_i P_j \tag{6.29}$$

After the local strain tensor has been computed, the body force is calculated by multiplying the local stiffness matrix 6.24 with this strain:

$$\mathbf{F} = K\epsilon \tag{6.30}$$

This the load vector is then computed using this force and equation 6.27.

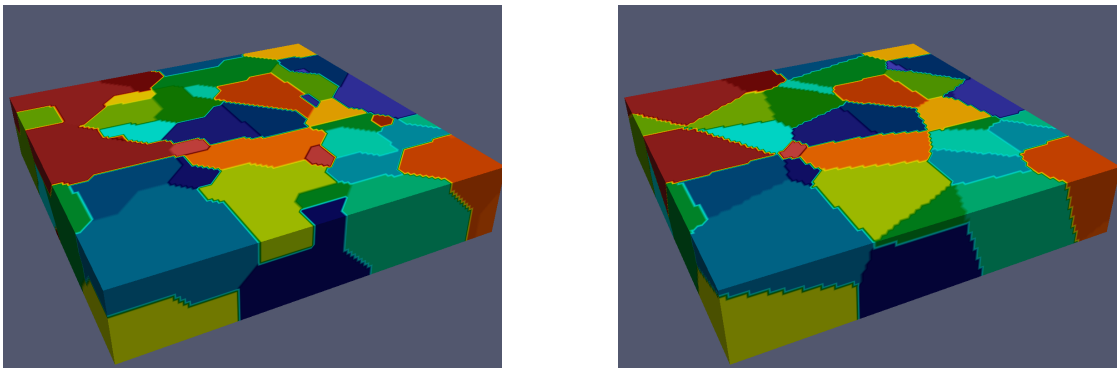
Force term

The matrix K is assembled once when the constructor of `ElasticForce` is called. At each iteration, the load vector is recomputed based on the current value of $\mathbf{P}$. After the load vector is assembled, the resulting system of linear equations is solved using the conjugate gradient solver with an incomplete Cholesky preconditioner from the Paralution library [78]. The force term for the TDGL is then evaluated from the local strain and electrostriction relations as described in Li *et al.* [88].

Defects

Three possible types of defects are implemented in the phase-field simulator. The first is for pinned defects where polarization remain fixed at a particular value and orientation throughout the simulation. Defects can also be constrained to a particular axis without the norm of the polarization vector being fixed, thus allowing the ferroelectric to switch only along a path through zero polarization. The final type of defect is a low-coercivity defect where the defect region will switch at a lower field than the surrounding regions. This is useful for simulating domain nucleation from a fixed region. Low-coercivity defects are implemented by reducing the value of α_1 in the local free energy density expression to some fraction of its bulk value (*e.g.*, 20% is used in the simulations in Chapter 8).

Defect regions can be specified either with with Cartesian coordinates or generated randomly to achieve a particular density of defects with a given average defect size. The random generation of defects is handled by the Voronoi tesselation module, described in the next section.



(a) L_1 (Manhattan distance) (b) L_2 (Euclidean distance)

Figure 6.3: Voronoi tessellations generated from the same set of seed points with L_1 and L_2 distance metrics.

Voronoi tessellation

A Voronoi tesselation of a set of n seed points is a partitioning of an area or volume into n regions with each region defined by being closer to one of those seed points than any of the other seed points by some metric [31, 11]. The phase-field simulator can use Voronoi tessellations for generating initial conditions with randomized contiguous domain-like regions that have the same polarization. Voronoi tessellations can also be used to specify defect regions. When a volume concentration of defects is specified, randomized cells of the tessellation are added to the defect region until enough regions are selected to get the specified concentration of defects. The Voronoi tessellation code supports both L_2 (Euclidean distance, $d = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}$) and L_1 (Manhattan distance, $d = |\Delta x| + |\Delta y| + |\Delta z|$) metrics. Tessellations based on the L_1 metric tend to have domain walls with 45°, 90°, and 180° intersections, which may be relevant for modeling tetragonal ferroelectrics like the perovskite family. Distribution of the angles formed at cell boundaries tends to be a bit more random when using the L_2 metric. Examples of random Voronoi tessellations generated by the simulator are shown in Fig. 6.3. The number of points n used in the tessellation is inversely proportional to the mean size of the individual regions. The mean volume of each cell can be approximated as $\bar{V}_{cell} = V_{total}/n$. As with all randomized initial conditions in the simulator, the seed for the pseudorandom number generator can be specified in the input file so that a particular random pattern can be exactly reproduced for future calculations.

6.3 Input and output

Input files

The input to the phase-field simulator is handled with JSON (JavaScript Object Notation) files, which are parsed using RapidJSON [137]. Typically, two JSON files are required to run a simulation: a material definition file and a simulation parameters file. A material definition file has three sections:

1. `Landau`- Coefficients for the local Landau-Devonshire expansion for free energy.
2. `Elastic`- Elastic coefficients for computing the mechanical contribution to free energy.
3. `Electrical`- Electrical properties of the material used in the electrostatics calculation and to normalize units throughout the simulation.

Full documentation of the parameters in the material definition file is provided in `Material.h`. The simulation parameters file has a number of blocks, which map directly to the structures in `SolverConfig.h`:

1. `IO`- Sets when output files are written, whether output is in ASCII or binary format.
2. `Forces`- Enables and disables different contributions to free energy as well as enabling defects.
3. `ICSetup`- Defines initial conditions for the simulation.
4. `Geometry`- Defines the size of the simulation domain as well as the boundary conditions.
5. `ElectrostaticSolver`- Configures the finite difference Poisson solver and the geometry of regions outside the ferroelectric region.
6. `ElasticSolver`- Configures fractional strain applied to the ferroelectric in the elasticity calculation.
7. `TopContactSweep`- Array containing segments of a piecewise-linear voltage sweep applied to the top contact of the electrostatic solver. When using a series capacitor or resistor, this defines the total voltage across the series combination of that element and the ferroelectric.
8. `Defects`- Defines the locations and types of defects present in the simulation.
9. `SimpleEField`- Configures the `SimpleEField` force term for a time-varying, spatially uniform electric field.
10. `Time`- Defines end time for the simulation as well as the time-step size.

VTK output

As with the BerkeleyNano3D quantum transport simulator, the default output format of the phase-field simulator is the legacy VTK file format, which can be opened using VisIt or ParaView. Which outputs are generated is specified in the `IO` block of the input file along with after how many iterations to generate each output. Polarization, stress, and strain are written as vector fields over the simulation domain, while charge and potential are written as scalar fields.

Chapter 7

Negative Capacitance in Phase-Field Models

The use of a ferroelectric in the gate stack of a field-effect transistor has been proposed as a way to improve on the thermodynamically imposed 60 mV/decade limit on the subthreshold swing of any device based on thermal injection of carriers over an electrostatically modulated barrier [117]. In this chapter, we explore how the formation of ferroelectric domains affects negative capacitance using continuum phase field models based on the time-dependent Landau-Ginzburg equation. We show that for a ferroelectric in series with a capacitor that the formation of domains can inhibit negative capacitance. This may significantly affect the stability of transistors using a ferroelectric for enhanced subthreshold performance and result in unwanted hysteretic behavior. These results primarily apply to larger, thicker ferroelectrics where the physics of continuum phase field models remain valid. At smaller length scales, such as those relevant to the thin-film dielectrics used in modern MOSFET devices, the applicability of the phase-field method remains an open question.

States near $P = 0$ are generally forbidden in a free ferroelectric below the Curie temperature as this corresponds to the high symmetry dielectric phase. However, recent experimental results have shown negative capacitance (NC), indicating that the interactions between ferroelectric (FE) and dielectric (DE) layers are indeed sufficient to stabilize negative capacitance [45, 4, 70]. The single-domain, uniform FE polarization model for negative capacitance has been extensively studied, and there have been several proposals for new devices based on the underlying theory [64, 40]. Some literature has indicated that the negative capacitance states may be unstable with respect to the formation of domains based on simple analytical models [98, 17], but there has been no study of the effect of domain formation on negative capacitance with a fully three-dimensional phase-field model. By coupling such a model to the dynamics of an external capacitor, we will show that the formation of domains can lead to a substantial suppression of negative capacitance and demonstrate how this instability can be used to explain the hysteretic behavior observed in many NC-FET devices, particularly those with large external FE capacitors connected in series with the gate stack.

7.1 Theory of ferroelectric-dielectric series combinations

Time-Dependent Ginzburg-Landau Dynamics

For the purposes of deriving a simplified model for the stability of polarization in a multidomain ferroelectric, suppose that the ferroelectric is compressively strained such that the polarization is energetically preferred to be along the z-axis (normal to the ferroelectric-dielectric interface) such that we can consider $P = P_z$. To describe the dynamics of a one-dimensional ferroelectric capacitor with spatially inhomogeneous polarization $P(x)$, we use the time-dependent Ginzburg-Landau (TDGL) equation [21]:

$$\frac{\delta P}{\delta t} = -L\frac{\delta F}{\delta P} = -L\frac{\delta F_{local} + \delta F_{wall} + \delta F_{elec}}{\delta P} \tag{7.1}$$

This equation is the gradient flow of the Ginzburg-Landau free energy functional for ferroelectrics, which includes the local Landau-Devonshire energy and a gradient term corresponding to domain walls. Mathematically, these terms of the free energy functional can be expressed for a ferroelectric in one dimension as:

$$F_{local} = \int \left[\alpha \left(P(x)\right)^2 + \beta \left(P(x)\right)^4 + \gamma \left(P(x)\right)^6\right] dx \tag{7.2}$$

$$F_{wall} = \int G_{wall} \left(\frac{dP}{dx}\right)^2 dx \tag{7.3}$$

$$F_{elec} = \int -E(x) \cdot P(x)\, dx \tag{7.4}$$

where α, β, and γ are the Landau-Devonshire free energy coefficients and G_{wall} is related to the domain wall energy density, $E(x)$ is the external electric field as a function of position, and L is a term by which the time scale is normalized. The F_{elec} term of the free energy corresponds to the interaction with an externally applied electric field, which will be derived in the next section for a series combination of a ferroelectric and dielectric capacitor.

Free energy of a ferroelectric-dielectric series combination

To model negative capacitance, we couple a dielectric (DE) capacitor to our ferroelectric (FE) through a circuit model that assumes equal charge across the DE and the FE:

$$V_{DE} = \frac{Q_{DE}}{C_{DE}} = \frac{-\langle P_{FE}\rangle}{C_{DE}} \tag{7.5}$$

$$V_{FE} = V_{app} - V_{DE} = V_{app} + \frac{\langle P_{FE}\rangle}{C_{DE}} \tag{7.6}$$

$$E_{FE} = -\frac{V_{FE}}{t_{FE}}, \tag{7.7}$$

where $\langle P_{FE} \rangle$ is the mean polarization about the z-axis in the ferroelectric, C_{DE} is the capacitance per unit area of the series DE capacitor, V_{DE} is the voltage across the DE capacitor, V_{app} is the voltage applied to the series combination of the DE and the FE capacitors, V_{FE} is the voltage across the FE, E_{FE} is the electric field across the ferroelectric, and t_{FE} is the ferroelectric thickness.

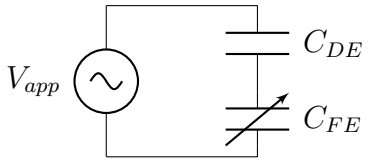


Figure 7.1: Series Capacitor Circuit Model.

7.2 Stability analysis

Here we will perform a simplified analysis of the stability of a ferroelectric where polarization is allowed to vary throughout the volume of the ferroelectric using a discretized version of the Ginzburg-Landau free energy functional. We begin by discretizing the previous expression for free energy in one dimension. Suppose a ferroelectric of thickness t_{FE} is divided into n identical regions with polarization P_i. Also suppose that the ferroelectric is connected in series with a dielectric subject to the condition that $Q_{DE} = -\langle P \rangle$. This condition where the polarization can be simply averaged is equivalent to having a metal between the FE and DE layers.

We can write the areal free energy density of the ferroelectric as:

$$F = F_{local} + F_{elec} \tag{7.8}$$

$$F_{local} = \frac{t_{FE}}{n} \sum_{i=1}^{n} \left(\alpha P_i^2 + \beta P_i^4 \right) \tag{7.9}$$

$$F_{elec} = \frac{t_{FE}}{n} \sum_{i=1}^{n} \left(-E \cdot P_i \right) = \frac{V_{FE}}{n} \sum_{i=1}^{n} P_i \tag{7.10}$$

We are ignoring higher order terms in F_{local}, but the analysis is similar if they are included. When a voltage V_{app} is applied we can calculate the voltage across the ferroelectric from Kirchoff's voltage law as $V_{app} = V_{DE} + V_{FE}$. The voltage across the dielectric with capacitance C_{DE} is calculated as:

$$V_{DE} = \frac{Q_{DE}}{C_{DE}} = \frac{-\langle P \rangle}{C_{DE}} = -\frac{1}{nC_{DE}} \sum_{i=1}^{n} P_i \tag{7.11}$$

With knowledge of V_{app} and V_{DE} we can write F_{elec}:

$$F_{elec} = \frac{V_{FE}}{n}\sum_{i=1}^{n} P_i = \frac{V_{app} - V_{DE}}{n}\sum_{i=1}^{n} P_i \tag{7.12}$$

Making one final substitution, we get:

$$F_{elec} = \frac{V_{app} + \frac{1}{nC_{DE}}\sum_{i=1}^{n} P_i}{n}\sum_{i=1}^{n} P_i \tag{7.13}$$

The presence of the double sum is expected as the metallic nature of the contacts to the ferroelectric means that the appearance of charge at one point in the ferroelectric changes the dielectric voltage which in turn affects the externally applied electric field at all points in the ferroelectric.

Similarly to the second derivative of free energy test for stability in the simple single domain picture, we can assess the stability of a state by looking at the eigenvalues of the Hessian H for the free energy. If all the eigenvalues of H are positive, the system is stable. If there are any negative eigenvalues, the system is unstable. We can write H for a system homogeneous polarization:

$$H = \begin{bmatrix} \frac{\partial^2 F}{\partial P_1^2} & \frac{\partial^2 F}{\partial P_1 \partial P_2} & \cdots & \frac{\partial^2 F}{\partial P_1 \partial P_n} \\ \frac{\partial^2 F}{\partial P_1 \partial P_2} & \frac{\partial^2 F}{\partial P_2^2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial^2 F}{\partial P_n \partial P_1} & \cdots & \cdots & \frac{\partial^2 F}{\partial P_n^2} \end{bmatrix} = \begin{bmatrix} a+b & b & \cdots & b \\ b & a+b & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ b & \cdots & \cdots & a+b \end{bmatrix} \tag{7.14}$$

where $a = (2\alpha + 12\beta P^2)\frac{t_{FE}}{2}$ and $b = \frac{1}{nC_{DE}}$. Each row in this matrix is cyclically rotated one element from the previous row. This is therefore a circulant matrix, which can be diagonalized with the discrete Fourier transform [32].

The eigenvalues of this matrix are given can be computed as:

$$\lambda_j = a + b\left(\sum_{k=0}^{n-1} \omega_j^k\right) \tag{7.15}$$

where the ω_j's are the n^{th} roots of unity. For $j \neq 0$ (multiplicity $n-1$), we can calculate λ_j by summing the series as:

$$\lambda_j = a + b\frac{\omega_j^n - 1}{\omega_j - 1} = a + b(0) = a = (2\alpha + 12\beta P^2)\frac{t_{FE}}{2}. \tag{7.16}$$

When $j = 0$, $\omega_j = 1$ and we get:

$$\lambda_0 = a + b\left(\sum_{k=0}^{n-1} 1^k\right) = a + nb = (2\alpha + 12\beta P^2)\frac{t_{FE}}{2} + \frac{1}{C_{DE}} \tag{7.17}$$

with the eigenvector corresponding to the zero frequency (constant) Fourier component :

$$v_0 = \frac{1}{\sqrt{n}}[1\ 1\ \cdots\ 1]^T. \tag{7.18}$$

Let us now consider the stability of homogeneous negative capacitance in the case where $P = 0$. We find that we have $n-1$ eigenvalues $\lambda_j = 2\alpha$. Given that α is negative for any ferroelectric, we conclude that all of these $n-1$ eigenvalues correspond to an instability in the ferroelectric. In fact, these eigenvalues will always be negative for any value of P corresponding to a region of negative curvature in the free energy landscape where negative capacitance is expected to occur. The value of λ_0 is positive only when $t_{FE} < t_{crit}$ where

$$t_{crit} = -\frac{1}{2\alpha C_{DE}}. \tag{7.19}$$

This is exactly the same as the stability criterion derived by Salahuddin and Datta in their original paper on negative capacitance [116]. Indeed, it corresponds to the eigenvector in the ferroelectric configuration space corresponding to a homogeneous polarization. With no additional coupling terms, the free energy minimum in the original single domain theory of negative capacitance is actually an saddle point with respect to the formation of a multidomain state. Looking at the other $n-1$ eigenvalues, we find that local stability everywhere in the ferroelectric is a requirement for the stability of any state with uniform polarization across the entire ferroelectric.

Adding domain wall energy

We now add a domain wall energy term of the form:

$$F_{grad} = G_{DW} \sum_i \left(P_i - P_{i+1}\right)^2 \tag{7.20}$$

This is simply the discretized form of the previously given continuum domain wall energy expression. Assuming periodic boundary conditions to preserve the circulant property of H, we can write the new Hessian as the original Hessian plus a tridiagonal correction (with two extra entries in the corners for the periodicity):

$$H = H_0 + 2G_{DW} \begin{bmatrix} 2 & -1 & & -1 \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ -1 & & -1 & 2 \end{bmatrix}. \tag{7.21}$$

The new eigenvalues of this matrix can be computed as:

$$\lambda_j = \lambda_{j0} + 4G_{DW}\left[1 - \cos\left(2\pi\frac{j}{n}\right)\right]. \tag{7.22}$$

This has the physical interpretation that the presence of a domain wall energy term can stabilize previously unstable states with respect to perturbation of high spatial frequency. However, we find that the eigenvalue shift quadratically goes to zero for small values of j corresponding to long wavelength polarization waves. Hence, we would expect any ephemeral negative capacitance state to rapidly transition to an inhomogeneous state with some minimum domain size determined by the magnitude of G_{DW} in this simplified model.

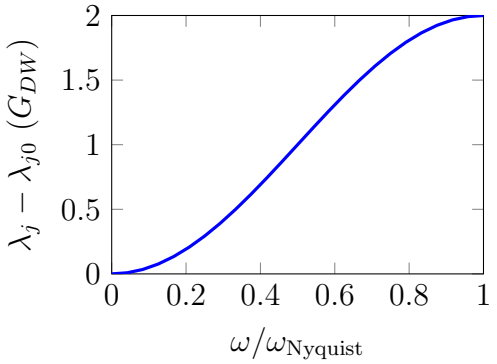


Figure 7.2: Eigenvalue shift from domain wall energy in terms of domain wall energy constant G_{DW} as function of spatial frequency ω (arbitrary units).

7.3 Phase-Field Method

The phase-field method [88, 21] is a popular way to simulate the behavior of ferroelectric materials subject to a plethora of complex boundary conditions in three dimensions. Phase-field methods minimize a functional of free energy by integration of the Time-Dependent Ginzburg-Landau (TDGL) equation:

$$\frac{\delta \vec{P}}{\delta t} = -L\frac{\delta F}{\delta \vec{P}} = -L\frac{\delta F_{local} + \delta F_{wall} + \delta F_{elec} + \delta F_{elas}}{\delta \vec{P}} \tag{7.23}$$

In this chapter, our description of free energy includes the local Landau-Devonshire energy and a gradient term corresponding to domain walls. The electrostatic interaction is added through a 3D finite difference Poisson solver, which is capable of handling arbitrary contact geometries and boundaries between the ferroelectric and external dielectrics. The nonlocal elastic interaction is handled through a finite-element elasticity solver. Details of the simulator can be found in Chapter 6.

Our simulations were done with the perovskite ferroelectric $PbTiO_3$ (PTO) using parameters taken from literature [111]. The ferroelectric region in the simulations was approximately 60×60×15 nm. Periodic boundary conditions were used to eliminate edge effects, but the

small size of the simulation domain effectively suppresses the type of very long-range interactions found in larger capacitors. To achieve negative capacitance, we couple a DE capacitor through a circuit model similar to the one previously used to derive the one-dimensional stability criterion:

$$V_{ox} = \frac{\langle P_{z,FE} \rangle}{C_{ox}} \tag{7.24}$$

$$V_{FE} = V_{app} - V_{ox}, \tag{7.25}$$

where $\langle P_{z,FE} \rangle$ is the mean polarization about the z-axis in the ferroelectric, C_{ox} is the capacitance per unit area of the series dielectric capacitor, V_{ox} is the voltage across the DE capacitor, V_{app} is the voltage applied to the series combination of the PE capacitor and the FE, and V_{FE} is the voltage across the FE. The value of V_{FE} is computed at every time step in the phase-field simulation and is used to set the Dirichlet boundary condition for electrostatic potential at the metal contacts at the top and bottom of the simulation domain. This model better corresponds to a metal-ferroelectric-metal-insulator-semiconductor gate stack used in some NC-FET transistors than previous phase-field simulations of ferroelectric-paraelectric interfaces [6], in which the materials are directly in contact with each other. In the direct contact case, the depolarization field from the uncompensated interface can lead to the formation of flux-closure quadrants [1, 135], which was recently demonstrated experimentally [124]. These vortex-like structures do not result in significant negative capacitance because near the interface, polarization is oriented parallel to the interface, greatly reducing the coupling between layers.

In some experimental setups used to study negative capacitance, an external capacitor is wired in series with the ferroelectric rather. In this case, the capacitor may not initially have the same charge as the ferroelectric. This does not change any results beyond a simple shift in the which voltage/charge corresponds to zero in the model. This can be trivially shown for some initial difference in charge between the capacitor and ferroelectric P_0:

$$V'_{ox} = \frac{\langle P_{z,FE} \rangle}{C_{ox}} - \frac{P_0}{C_{ox}} \tag{7.26}$$

$$V_{FE} = V_{app} - V_0 - V_{ox} \tag{7.27}$$

$$V_{FE} = V'_{app} - V_{ox}, \tag{7.28}$$

with $V'_{app} = V_{app} - V_0 = V_{app} - P_0/C_{ox}$.

In our simulations the external capacitance is taken to be approximately 12 μF/cm^2, a value which should be possible through high-κ dielectrics in modern CMOS processes. This gives a critical thickness for negative capacitance of 25 nm for PTO in series, below which Landau theory predicts that the spontaneous polarization in the ferroelectric should vanish and the ferroelectric should be stabilized in a state where it exhibits negative capacitance [117].

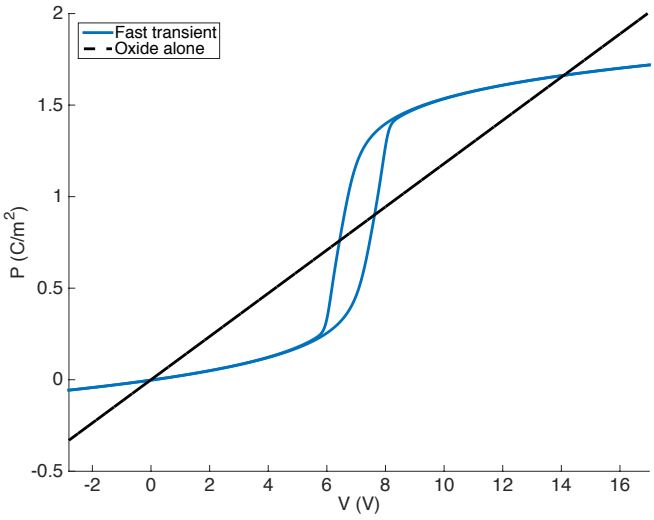


Figure 7.3: Hysteresis loop with fast sweep rate. Clear enhancement of capacitance over the series capacitor alone can be seen.

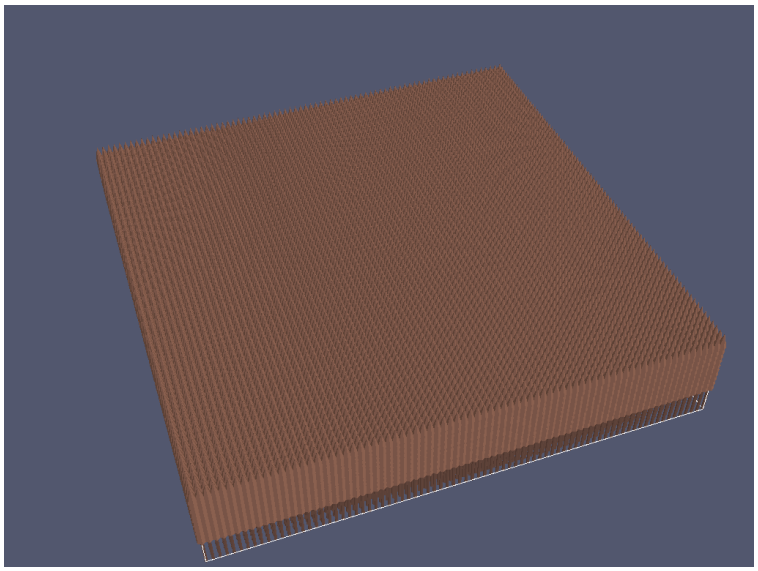


Figure 7.4: Single domain state near $P = 0$. This state is unstable and disappears if the voltage ramp rate is not fast enough.

7.4 Fast Transient

When the sweep rate is very fast relative to the formation of domains through the amplification of polarization waves in the system, negative capacitance can be clearly observed. This can be shown by noting that the series combination of the ferroelectric and paraelectric has a higher polarization than the ferroelectric alone. In this case, the ferroelectric switches coherently as a single domain. During the switching transient the polarization of the system takes on a uniform value near $P = 0$ as shown is Fig. 7.4. It is these states which directly correspond the single domain theory of negative capacitance. The small hysteresis in Fig. 7.3 can be explained by the lag of polarization behind voltage from the Landau-Khalatnikov dynamics.

7.5 Defects

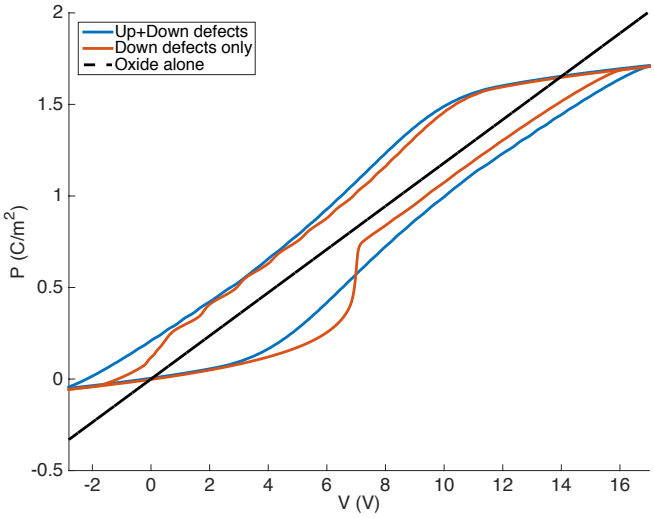


Figure 7.5: Hysteresis loops in the presence of pinned defects. When pinned defects are only of a single polarity and the voltage sweep rate is fast enough, a small region of apparent negative capacitance can be observed.

When small pinned defects are added to the system, these defects serve as nucleation centers for the formation of domains as illustrated in Fig. 7.6. When the ferroelectric switches through domain wall motion as in this case, the negative capacitance behavior is no longer observed and the hysteresis loop falls below the line indicating the series dielectric capacitance. To achieve observable capacitance enhancement, a large region of the ferroelectric must simultaneously transit the negative curvature region of the free energy landscape. In domain wall switching, only a small region near/within the domain walls is in this region.

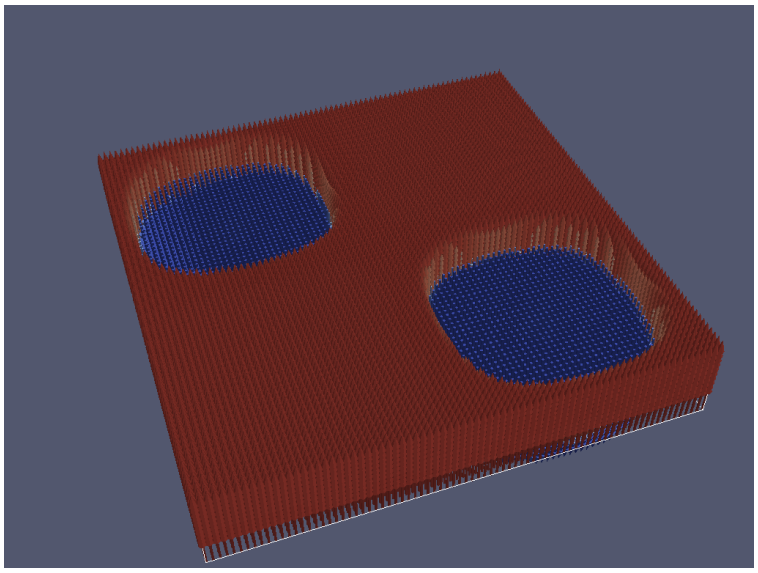


Figure 7.6: Intermediate state from defect-mediated domain nucleation

In the case where all defects are pinned along the same direction, we can observe negative capacitance over a short segment of the hysteresis loop where the defects cannot serve as nucleation centers for domains of the opposite polarity.

7.6 Breakdown through Spinodal Decomposition

Even without defects, it is still possible for domains to form. It was previously noted that when a single domain ferroelectric enters a region with where the single domain Landau-Khalatnikov model predicts negative capacitance, the system may rapidly break into domains through a barrier-free process similar to spinodal decomposition [5]. This type of spontaneous breakdown is seen in our phase-field models when the voltage sweep is conducted slowly enough such that infinitesimal inhomogeneities in the polarization amplified into a large amplitude polarization waves as predicted by the previous stability analysis. In this case, negative capacitance ceases to be observed as the capacitance of the ferroelectric is again dominated by localized domain wall motion instead of a global, coherent switching process through $P = 0$. An intermediate state generated by the spinodal breakdown mechanism for a long-wavelength polarization wave is shown in Fig. 7.8. In this case, absolute voltage amplification, corresponding to a negative voltage across the ferroelectric, does not occur. However, it is clear from Fig. 7.7 that in during the spinodal transition dQ/dV is higher than for the dielectric alone, giving differential voltage amplification.

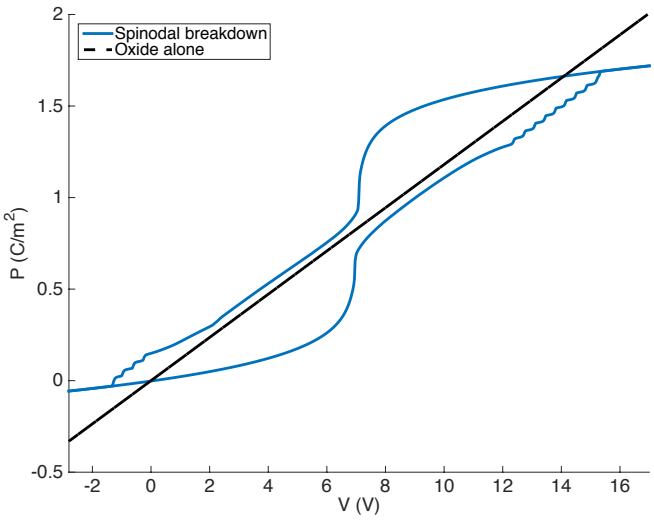


Figure 7.7: Hysteresis loop resulting from domain formation through spinodal breakdown without defects. The steep transition regions are accompanied by a sudden decrease in the free energy of the system, resulting in irreversibility.

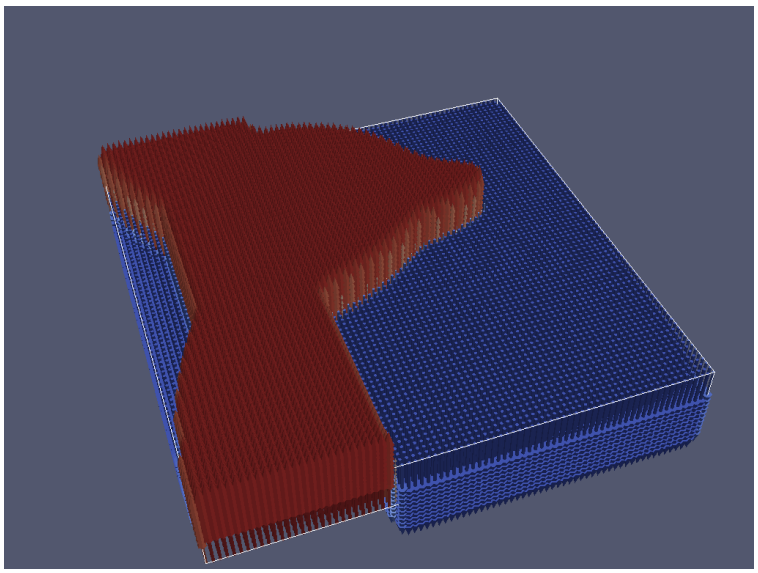


Figure 7.8: Stabilization of a long wavelength polarization waves as domains from spinodal breakdown

7.7 Implications for NC-FET devices

In order to have low-hysteresis enhancement of capacitance for MOSFET operation, we must remain close to the single domain state as in our fast transient case. In the spinodal breakdown case, we observe regions of the hysteresis loop that show a rapid increase in charge for a small change in potential. It is very important not to confuse these transitions with those predicted by a single domain theory of negative capacitance even though they may indicate a sharp transition. In the single domain model, the free energy varies continuously with the polarization. When the ferroelectric breaks into domains via the spinodal mechanism, there is a substantial, immediate change in free energy because the multidomain state with absolute local polarization $|P| \approx P_S$ is strongly energetically preferable over the single domain state where $|P| \ll P_S$ absent a huge domain wall energy, where P_S is the spontaneous polarization. The domain wall energy densities required to stabilize the negative capacitance states would have to be much higher than those required to stabilize a simple capacitor in a single domain state because of the large energetic penalty of being so far from the stable spontaneous polarization.

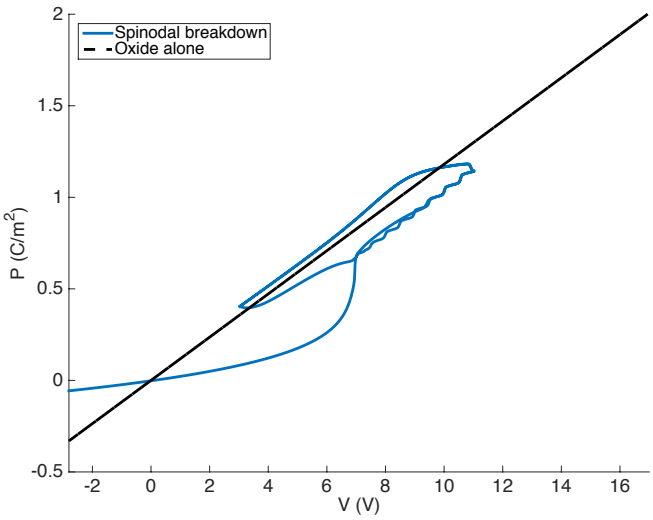


Figure 7.9: Hysteresis loop from incomplete poling after spinodal breakdown

The discontinuity in free energy at this transition implies thermodynamic irreversibility, meaning that to return to the single domain state require traversing a different trajectory in the configuration space of the system. This invariably leads to hysteretic behavior as in Fig. 7.9, where the system underwent spinodal decomposition and the voltage was swept such that the system remained in the multidomain state. The implication for transistors is that if ferroelectric breaks into domains through the proposed spontaneous, barrier-free mechanism, it may be possible to turn the device on with a subthreshold slope steeper than

60 mV/decade when this irreversible transition occurs, but it will not be possible to simply turn the device off by slightly decreasing the voltage.

7.8 Conclusion

Negative capacitance can be suppressed through the formation of domains. To observe negative capacitance experimentally, it is thus critical to have very high quality films free of defects which promote domain nucleation or a very thin film where the dynamics of the ferroelectric are more locally coupled to the dielectric than is predicted from our model where charge is averaged across the ferroelectric. More experimental work is needed to understand the kinetics of the formation of domains through the barrier-free spinodal decomposition mechanism. If such a mechanism were to be experimentally confirmed to be rapid, it would be highly challenging to stabilize a large-area, thick ferroelectric in the negative capacitance region such that a less than 60 mV/decade subthreshold swing could be achieved without significant hysteresis in an NC-FET.

Finally, it may be possible to achieve low-hysteresis negative capacitance in other materials more recently under investigation for NC-FET applications, such as ferroelectric $HfZrO_2$, which has a lower spontaneous polarization and a smaller energetic barrier for switching, than the PTO we examined in this chapter. $HfZrO_2$ can also be grown at very small thicknesses with atomic layer deposition (ALD) directly on silicon, which may alleviate the "averaging" effect of the metal plate boundary conditions we have considered.

Chapter 8

Multidomain Phase-Field Modeling of Negative Capacitance Switching Transients

In this chapter, we demonstrate that continuum phase-field simulations show how a multidomain ferroelectric capacitor in series with a resistor can exhibit a transient response in which the ferroelectric behaves as a negative capacitor. We show that accelerating domain growth leads to negative capacitance and this happens even when there is no initial switching of the domains. The observed behavior is in close agreement with experimental results of negative capacitance transients seen recently in a number of ferroelectric material systems.

8.1 Introduction

In previous chapters, we discussed how negative capacitance (NC) has been proposed as a way to overcome the Boltzmann limit for subthreshold swing in a field-effect transistor [116]. Recently, transients corresponding to a negative capacitance have been observed in an experimental system by connecting a resistor in series with a ferroelectric capacitor [72, 75, 57]. Other experiments have shown clear evidence of an NC effect from ferroelectric layers in a ferroelectric-dielectric superlattice as measured by small signal capacitance [71, 44]. Another recent study correlated similar experimental results with atomistic Monte-Carlo simulations for such superlattices [145]. In addition, a number of recent papers have demonstrated < 60 mV/decade subthreshold swing in transistors at room temperature [143, 82, 84, 73, 67, 66, 103, 95].

In this chapter, we address the question of how multidomain ferroelectricity relates to the viability of negative capacitance [18, 145], specifically analyzing the direct measurement of negative capacitance transients that were observed in single crystal ferroelectric capacitors [72]. The experimental set up is shown in Fig. 8.1 where a voltage source is connected to a series combination of a resistor and a ferroelectric capacitor. Experimental measurements

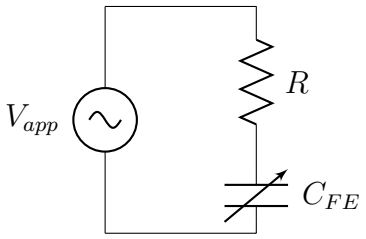
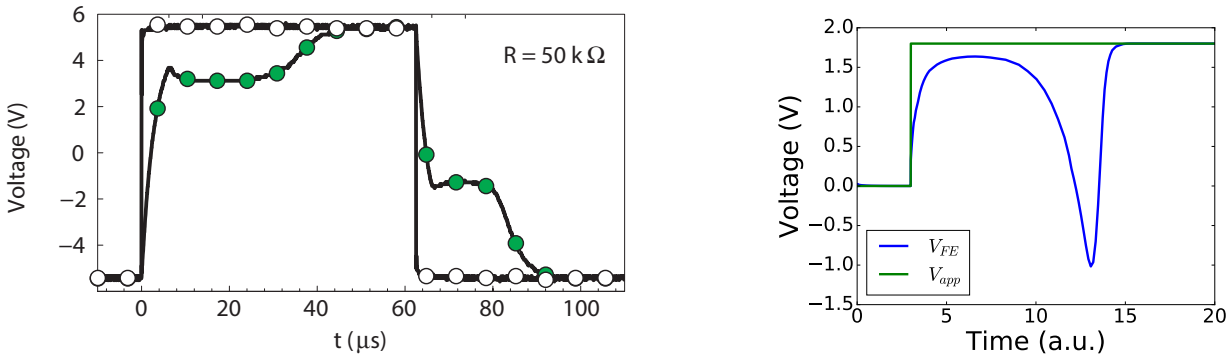


Figure 8.1: Series Resistor Circuit Model. Here the capacitor C_{FE} is a ferroelectric capacitor.



(a) Experiment (b) Single Domain Model

Figure 8.2: (a) The experimental data from Khan *et al.* [72]. (b) Landau-Khalatnikov model for a ferroelectric in series with a resistor. While the $dV_{FE}/dt < 0$ behavior as seen in the experiment is replicated, the long flat region of the transient is not present. Additionally, the voltage across the ferroelectric can become negative resulting in the voltage across the resistor becoming larger than the source voltage V_{app}.

[72] show that when a voltage is applied to this series combination, the voltage across the ferroelectric,V_F decreases during some parts of the switching transient, even though the charge, Q_F is itself increasing. This leads to a situation where $dQ_F/dV_F < 0$, exhibiting negative capacitance. Fig. 8.2(a) shows a typical experimental trace of this phenomenon.

Our objective here is go beyond the traditional practice of using a single-domain Landau-Khalatnikov equation [81] and investigate how the multidomain nature of the ferroelectric is expected to affect the negative capacitance behavior. For comparison, Fig. 8.2(b) shows a single-domain calculation corresponding to experiment. While the single-domain model definitively predicts the negative capacitance behavior qualitatively, the lack of similarity between the experimental and modeling results indicates that the multidomain nature of the capacitors is playing a significant role in the observed behavior. In addition, given that the capacitors used in the experiment have at least a 20 μm $\times$20 μm footprint area, a multidomain model is more realistic. Motivated by these observations, we have used a three-dimensional phase field model to analyze the experiment.

8.2 Model

Time-Dependent Ginzburg-Landau Dynamics

To describe the dynamics of a ferroelectric capacitor with spatially inhomogeneous polarization $\vec{P}$, we use the time-dependent Ginzburg-Landau (TDGL) equation [21]:

$$\frac{\delta \vec{P}}{\delta t} = -L\frac{\delta F}{\delta \vec{P}} = -L\frac{\delta F_{local} + \delta F_{wall} + \delta F_{elec}}{\delta \vec{P}} \tag{8.1}$$

This equation is the gradient flow of the Ginzburg-Landau free energy functional for ferroelectrics, which includes the local Landau-Devonshire energy and a gradient term corresponding to domain walls. Mathematically, these terms of the free energy functional can be expressed for a ferroelectric as:

$$\begin{aligned} F_{local} = \int_\Omega [&\alpha_1 \left(P_x^2 + P_y^2 + P_z^2\right) \\ &+ \alpha_{11} \left(P_x^4 + P_y^4 + P_z^4\right) \\ &+ \alpha_{12} \left(P_x^2 P_y^2 + P_y^2 P_z^2 + P_z^2 P_x^2\right) + ...] \, d\Omega \end{aligned} \tag{8.2}$$

$$\begin{aligned} F_{wall} = \int_\Omega G_{11}(&P'_{x,x} + P'_{x,y} + P'_{x,z} + P'_{y,x} + P'_{y,y} \\ &+ P'_{y,z} + P'_{z,x} + P'_{z,y} + P'_{z,z}) \, d\Omega \end{aligned} \tag{8.3}$$

$$F_{elec} = \int_\Omega -\vec{E} \cdot \vec{P} \, d\Omega \tag{8.4}$$

where the α values are Landau-Devonshire free energy coefficients. The full expression of local free energy density is taken from [112] and includes terms up sixth order for polarization in $PbTiO_3$. G_{11} is related to the domain wall energy density, $\vec{E}$ is the electric field, and L is a parameter with units of resistance by which the time scale is normalized. The F_{elec} term of the free energy corresponds to the interaction with an externally applied electric field, which will be derived in the next section for a series combination of a ferroelectric capacitor and a resistor. For the purposes of this work, we have ignored long-range elastic interactions as they are not relevant to the physics that give negative capacitance transients, but they could be included in a future work focusing on more quantitative predictions.

When a resistor is added to the simulation, the equations must be reformulated into a differential-algebraic equation where $\frac{\delta P}{\delta t}$ is now implicit. *This is a key point to remember when simulating such a system and has significant impact on overall behavior of the simulation as described in the following.* In simulations without an external resistor, the electric field can be calculated by solving Poisson's equation within the ferroelectric with the voltages at the top and bottom contacts set by the specified external voltage. When the resistor is added, the ferroelectric contact voltage becomes a function of both the external source voltage and the current flowing through the circuit. The voltage across the ferroelectric capacitor shown

in Fig. 8.1 can be calculated as:

$$V_{FE} = V_{app} - RI_{FE} = -R \int_{\Omega} \frac{dP_z}{dt}\, d\Omega \tag{8.5}$$

for a ferroelectric where switching occurs along the z-axis. It can now be seen that the *local* rate of change in polarization anywhere in the system can affect the contact voltage, which affects the *global* dynamics of the ferroelectric. To solve the resulting implicit form of the TDGL equation in which the ferroelectric current and voltage are self-consistent, we used the IDA solver from the Sundials package [55].

Defects are added to the simulation in one of two different ways. In the first method, defects are represented as regions where the polarization is fixed to a particular orientation and is not allowed to change throughout the simulation. Such defects resemble 'pinned defects' and the switching of the ferroelectric happens simply by domain wall growth from these defects. The second method is to model defects as regions with a lower coercive field than the rest of the ferroelectric. This is done by reducing the the value of the Landau coefficient α_1 to some fraction of its bulk value. For this case, the defect regions will switch first when a field is applied and domains will grow from those defect regions. For this work, a value of $\alpha_{1,\text{defect}} = 0.2\alpha_{1,\text{bulk}}$ was chosen. The main difference between the two methods is that in the first case, there is no initial switching of the domains, rather the applied voltage results in the growth of oppositely polarized domains from the pinned defects; by contrast for the second method, there is an initial switching of the domains at the defect points, and subsequently there is growth of those switched domains.

8.3 Results and Discussion

Our simulated structure consists of a $PbTiO_3$ ferroelectric bar that is 75 nm long, 30 nm wide, and 15 nm thick. Initially, the polarization is oriented in the upward direction everywhere. A 3 nm × 3 nm × 3 nm defect is included at the center of the end of the bar as is illustrated in Fig. 8.3(a). A voltage step with a magnitude of 2.3 V is applied to switch the ferroelectric. The voltage across the ferroelectric is shown in Fig. 8.4.

For both the low-coercivity defect as well as the pinned defect, an NC transient is seen that resembles what is seen experimentally. The transient appears when the domain growth rate is accelerating. As can be seen in Fig. 8.3(b), the domain surface area, and hence growth rate, is proportional to the square of the radius the domain has propagated from the defect. Assuming a constant domain wall velocity, it is apparent that domain growth will accelerate. During this accelerating growth, the current through (and the voltage across) the resistor will increase, thus the voltage across the ferroelectric capacitor will decrease. This decrease in voltage can be interpreted as a negative capacitance. A secondary effect of the reduced voltage across the ferroelectric is that the switching takes place at a slightly slower rate due to the lower electric field.

As a greater volume of the ferroelectric switches to the orientation of the new domain, the effective domain wall area saturates as in Fig. 8.3(c). At this point voltage will remain

relatively constant across the ferroelectric. Eventually, all of the ferroelectric will be switched to the final polarization orientation as in Fig. 8.3(d), and the current will go to zero with ferroelectric voltage now equal to the source voltage.

It is interesting to note that both the pinned and the low-coercivity defect models give qualitatively similar results and that the microscopic physics of the domain nucleation center is fairly unimportant in determining the overall switching characteristics. Therefore, the "correct" defect model would be difficult to determine directly from looking at ferroelectric switching data. What really matters in our simulations is the rate at which the domains grow after the initial nucleation.

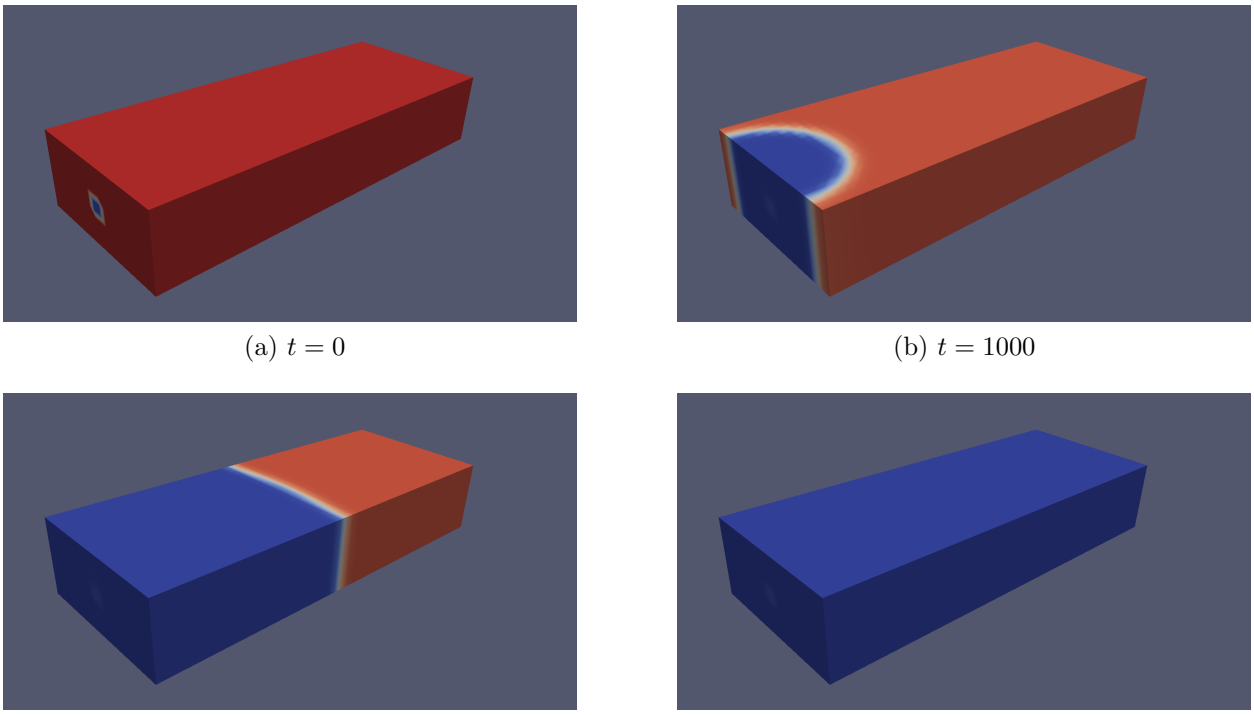
To compare our defect-mediated switching model with previous theoretical work, the simulation was also run with defects turned off so that the ferroelectric behaves as a single domain in accordance with the Landau-Khalatnikov equation. As seen in Fig. 8.4(c), the behavior of the voltage is the same as a simple one-dimensional model as shown in Fig. 8.2(b).

While the bar geometry model with a single defect allows the concept of negative capacitance through accelerating domain growth to be easily visualized, it is clearly not a physical model for a large-area, thin-film ferroelectric capacitor. To demonstrate that the physics of our negative capacitance model is not limited to such a contrived example, we ran a large area simulation for a 300 nm × 300 nm × 15 nm ferroelectric capacitor. Within the capacitor, 15 fixed polarization defects, each occupying a 1.5 nm × 1.5 nm × 1.5 nm volume, were placed at random locations.

The voltage across the ferroelectric capacitor, V_{FE}, is shown in Fig. 8.4(d) and the domain structure at different points in time is shown in Fig. 8.5. As with the simplified geometry, when the rate of domain growth is accelerating, the negative capacitance transient is evident. With randomized defects, the qualitative features of the experimental NC transient are clearly observed. There is a clear region where $dQ/dV < 0$, an extended region where the voltage does not change too much indicating a mostly constant domain growth rate, and finally a region where the ferroelectric voltage smoothly slopes up the its final value.

8.4 Conclusion

In conclusion, we have presented three-dimensional phase field simulations for a ferroelectric capacitor connected in series with a resistor. We address two specific questions: (i) Is it possible to obtain a negative capacitance behavior within a multidomain model and (ii) is domain switching necessary for a negative capacitance behavior? Our results show that the multidomain model gives excellent agreement with the experimental data. In addition, we show that even when domains grow from pinned defects, a negative capacitance behavior is expected, indicating, domain switching is not a precondition for negative capacitance behavior. Our results show that the negative capacitance behavior in a multidomain model originates from the acceleration of the domain growth. Within the picture of an activation model, where certain external stimulus is necessary to go over the activation barrier, such



(a) $t = 0$ (b) $t = 1000$

(c) $t = 2500$ (d) $t = 5000$

Figure 8.3: Internal state of the ferroelectric with a pinned defect during switching. (a) Initial state before any voltage is applied. (b) During the NC transient, the rate of domain growth is accelerating because of the increasing surface area of the domain. This results in an increasing switching current and therefore an increasing voltage dropped across the resistor. (c) When the domain growth rate has stabilized, the switching current remains approximately constant. (d) Switching is complete, and the ferroelectric polarization has a uniform orientation. Red indicates a domain with polarization along the z-axis pointing up, while blue indicates down.

acceleration is identifiable with the kinetic energy gained while 'rolling down the hill' from the activation barrier, in close resemblance to a double-well energy landscape that signifies a two-level system, such as a ferroelectric.

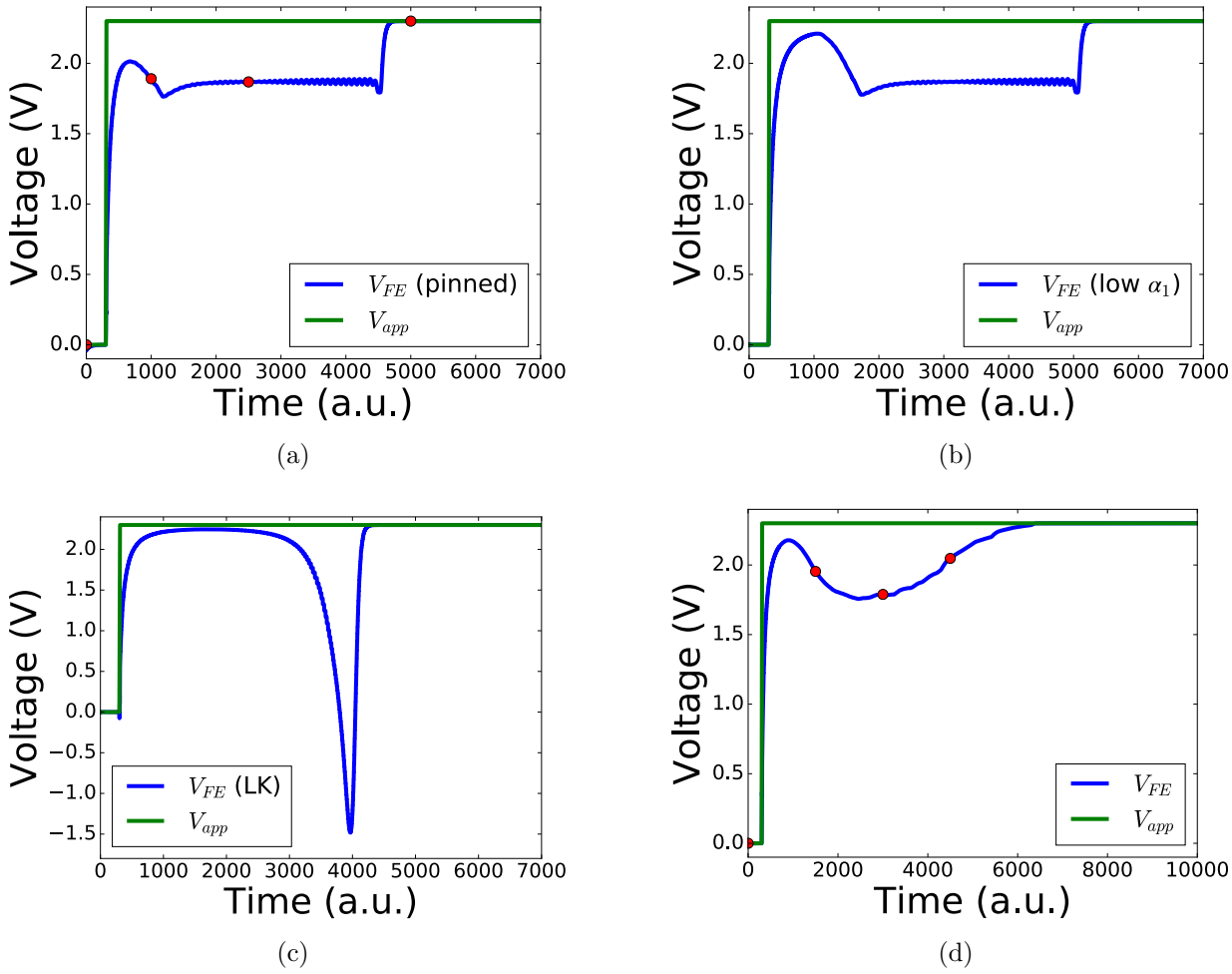
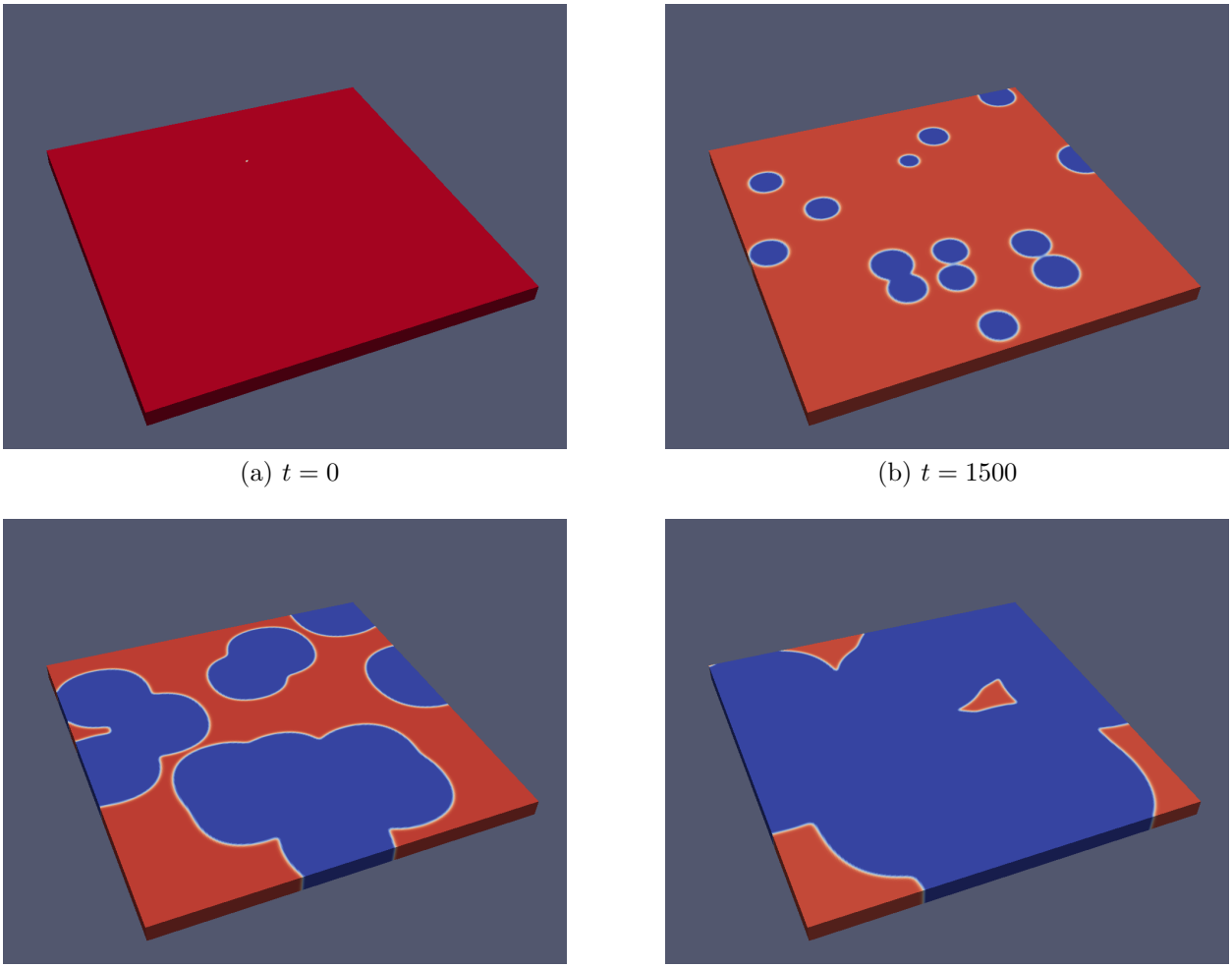


Figure 8.4: (a-c) Comparison of different defect models for the bar geometry ferroelectric with a single defect with the Landau-Khalatnikov solution. (a) The low-coercivity defect has $\alpha_1 = 0.2\alpha_{1,\text{bulk}}$. Markers along the curve correspond to the time-steps shown in Fig. 8.3. (b) The pinned defect has orientation rotated 180° to that of the bulk. Because the difference between initial switching current and saturation switching current for the pinned defect case is smaller than the difference between the initial nucleation current and saturation switching current for the low-coercivity case, the pinned defect case exhibits a smaller NC transient in terms of voltage magnitude. Markers along the curve indicate the time-steps shown in Fig. 8.3. (c) The intrinsic, no defect case is for a ferroelectric which switches uniformly as a single domain. This is equivalent to the solution of the Landau-Khalatnikov (LK) equation. (d) For a large area capacitor with 15 randomly positioned pinned defects, the results are qualitatively very similar to experiment. Markers along the curve correspond to the time-steps shown in Fig. 8.5.



(a) $t = 0$ (b) $t = 1500$

(c) $t = 3000$ (d) $t = 4500$

Figure 8.5: Switching a large-area capacitor with random defects. (a) 15 defects are randomly placed in the ferroelectric oriented down. These defects are mostly inside the ferroelectric and are therefore occluded because only surface polarization is shown. (b) Negative capacitance is observed as the growth rate increases. (c) The domain growth rate has saturated, and the NC transient stops. (d) As more of the volume of the ferroelectric switches, the effective domain growth rate slows down and voltage rises across the ferroelectric as in a dielectric capacitor.

Chapter 9

Conclusion and Future Work

This thesis covered the development of two simulators capable of solving problems relevant to the field of semiconductor devices. The first was BerkeleyNano3D, a quantum transport simulator based on the non-equilibrium Green's function formalism, built from scratch for simulating low-dimensional semiconductor devices. The second was a phase-field simulator, extended from our group's previous work, capable of simulating negative capacitance within the framework of the time-depenedent Ginzburg-Landau equation.

We have explored two phenomena that could be utilized in producing transistors that could overcome the Boltzmann limit for subthreshold swing of 60 mV/decade: superlattice energy filtering and ferroelectric negative capacitance. Our chevron GNR simulations are, to the best of our knowledge, the first results on steep-slope superlattice FET devices that rigorously take into account the effects of phonon scattering. The phase-field models for negative capacitance go far beyond the limits of the single-domain theory that has so far been the standard for modeling this phenomenon.

9.1 Summary of key results

The key results of this thesis are summarized as follows:

1. The quantum transport simulator BerkeleyNano3D provides an environment to study both ballistic and dissipative transport in low-dimensional material systems, such as graphene nanoribbons and carbon nanotubes. Simulations can be efficiently scaled to large parallel systems. It features a finite-element based electrostatics solver that solves the nonlinear Poisson equation for rapid convergence. The simulator was built over the course of several years and provides a foundation for future work.

2. The hierarchical Schur complement (HSC) algorithm can be used to solve for the retarded Green's function G^r and the electron correlation function G^n with fewer operations than the standard recursive Green's function algorithm. Compared with the implementation in literature [54], our version required no modification to the external

graph partitioning library, but we ultimately found that this algorithm was relatively impractical for simulating the low-dimensional structures that we were interested in.

3. Our simulator is capable of simulating CNT- and GNR-based devices with the ability to handle a wide range of geometries, including GNR heterojunctions. These heterojunctions may show interesting transport properties including negative differential resistance and tunneling effects.

4. Chevron graphene nanoribbons (CGNRs), assembled with bottom-up chemical synthesis techniques, are a monolithic superlattice that can be utilized in several interesting ways because of the presence of minibands above the conduction band edge. Ballistic calculations reveal that field-effect transistors fabricated from CGNRs could have a subthreshold swing as low as 6 mV/decade. These devices could also exhibit negative differential resistance with respect to the drain voltage with a peak-to-valley ratio (PVR) in excess of 4800.

5. The performance of CGNR devices is, however, degraded upon the inclusion of optical phonon scattering. Dissipative transport calculations reveal that it should be possible to achieve a subthreshold swing less than 60 mV/decade along with a PVR ≈ 3 even in the presence of phonon scattering.

6. While CGNR devices can exhibit steep-slope switching, the negative differential resistance property significantly limits the maximum possible logic voltage and the ability to use CGNR superlattice FETs as a drop-in replacement for traditional silicon-based CMOS technology.

7. A phase-field simulator, based on the previous work of K. Ashraf [7], has been extensively developed with new features to support simulation of ferroelectric negative capacitance. These features include a modular electrostatics solver with support for coupled external resistors and capacitors, interfaces to fast linear solvers and time-steppers, several defect models, and flexible initial condition generators including Voronoi tessellation.

8. Phase-field models predict that negative capacitance states with uniform polarization near $P = 0$, which are predicted to be stabilized by an external dielectric capacitor in a single domain model, are unstable and will spontaneously undergo a transition into a multidomain state via a mechanism similar to spinodal decomposition. This was originally predicted by Cano and Jiménez [17] using a one-dimensional ferroelectric model. Our work demonstrates this effect in a three-dimensional simulation, derives the stability criterion for a discretized model, and predicts a modified hysteresis loop. The validity of these results may not extend to very thin ferroelectric films grown directly on semiconductors because of the continuum nature of phase-field models.

9. Experimentally observed negative capacitance transients for a ferroelectric in series with a resistor can be explained through domain growth in the ferroelectric with a

variable domain growth rate. This model agrees well with the experimental voltage transient waveform.

9.2 Future work

Quantum transport simulator development

Several features could be added to BerkeleyNano3D to make it more suitable for solving even larger problems:

Adaptive integration

The most useful improvement to the BerkeleyNano3D would arguably support for adaptive integration. The current implementation uses a simple Riemann sum to add up current and charge over a grid of energy points. This is simple to implement and easy to parallelize. Quadrature rules, such as Gauss-Kronrod quadrature [77] give some estimate of the accuracy of the quadrature and can suggest when the number of quadrature points over an interval should be refined to converge to some tolerance. Some changes to the code to enable this would be necessary. Many sections of code, particularly in the `ParallelIntegrator` class assume that the positions of the quadrature point are known before starting the transport calculation. Therefore, some synchronization and distribution must developed to allow new energy points to be dispatched to the MPI processes to refine certain intervals where the error bounds must be improved.

Parallel linear algebra and efficient algorithms

Currently parallelism in BerkeleyNano3D is limited by the extent to which the problem can be subdivided into independent energy and k-points. While multithreaded linear algebra libraries, such as OpenBLAS and Intel MKL allow the use of an entire compute node to solve to solve an individual quadrature point, very large problems, such as full simulations of silicon FinFETs likely will require significantly more memory than what will fit in a single compute node. In this case, the linear algebra itself must be done with a library such as ScaLAPACK, which supports MPI parallel communication. However, rewriting the code to support this would likely be a long and arduous undertaking.

Additionally, better algorithms for computing transport other than the recursive Green's function method could greatly enhance performance and scalability. We discussed the hierarchical Schur complement (HSC) algorithm [54] in Chapter 3. Approximate techniques, such as model order reduction [59] may also enable more efficient transport calculations.

Electrostatics parallelization

One area of the code lacking significantly in parallelization is the electrostatics solver, which is executed only on the root process. The main computational task in this part of the simulator is the solution of a linear system $Ku = b$ where K is a sparse, positive definite matrix. There are two common algorithms used to solve problems of this type [32]. The first is conjugate gradient (CG), an iterative Krylov subspace method whose performance depends on fast sparse matrix-vector multiplication. The second is Cholesky decomposition (LDL^T factorization), a direct method which factors the matrix into a lower triangular, a diagonal, and an upper triangular matrix. Currently the CG solver from the Eigen library [51] is used because Eigen is already being used, and this reduces the number of external dependencies. This solver offers reasonable speed, especially when used with OpenMP, but it can still only be run on a single MPI process. This is adequate for FEM problems of size we are currently solving ($\sim 10^4$ to $\sim 10^5$ unknowns), but may become a bottleneck for very large ($\sim 10^7$ unknowns) FEM problems with very dense meshes. In this case, it may be necessary to use a fully parallel sparse solver, such as the Kyrlov subspace solvers provided by PETSc [10] or Trilinos [53], or a direct parallel solver, such as SuperLU [87] or MUMPS [2].

First-principles electronic structure

The simulations on graphene nanoribbons in this thesis were done with a p_z tight-binding Hamiltonian. Some of the armchair GNR and heterostructure results were validated against the more sophisticated tight-binding model of Boykin *et al.*, which adds two extra d-orbitals to give more accurate estimates of the bandgaps of AGNRs [14]. While these tight-binding models are excellent for simulating carbon-only GNRs, doping can only be simulated by adding a background charge to the Poisson solver that results in the equilibrium Fermi level in the ribbon floating to a level consistent with the doping.

GNRs with site-specific substitutional doping have been synthesized by Cloke *et al.* with boron and nitrogen dopants [26]. A possible way to rigorously model these ribbons would be to use *ab-initio* calculations and extract a tight-binding Hamiltonian in a localized basis set. Approaches to this include directly fitting Slater-Koster overlap integrals to a density functional theory (DFT) bandstructure as done by Zahid *et al.* for MoS_2, or using a tool like Wannier90 [99] to extract a maximally-localized Wannier function basis for wavefunctions from a DFT program, such as Quantum Espresso [46]. Caution should be used with this approach, however, as the DFT bandgaps for graphene nanoribbons are usually significantly smaller than those calculated when the GW correction is added [122, 136].

Optimization of GNR heterojunctions

The chevron GNR we studied in Chapter 5 was chosen as a topic for research because it had actually been experimentally synthesized. Going forward, the choice of graphene nanoribbon

to use for superlattice FET structures will likely have to be optimized before a working device exhibiting negative differential resistance can be fabricated. While this ribbon has a bandgap of 1.59 eV in our tight-binding calculations, DFT+GW calculations predict that it should have a much higher bandgap of around 3.74 eV [133].

In practice, ribbons will have to chosen from a complex design space considering the following criteria:

1. Bandgap. Ribbons will likely need to have a small bandgap (< 2 eV) in order for a useful FET device to be made.

2. Miniband width. The first miniband must be sufficiently wide to get a reasonable range of voltage operation without being too wide to prevent NDR within that voltage range.

3. Miniband gap width. The gap between minibands should be wide enough that band-to-band tunneling from the first to second minibands does not substantially impact the performance of the device.

4. Feasibility of synthesis. Some ribbons are more easily synthesized than others.

5. Ability to form heterojunctions. If it were possible to form a heterojunction such that high energy electrons were only filtered at the source, it may be possible to make devices closer to those described by Lam *et al.* [80] that exhibit steep-slope behavior without the negative differential resistance that makes our devices impractical as drop-in replacements for current CMOS technology.

Topological properties of GNR heterojunctions

Recent work by Cao *et al.* studied the topology of junctions between AGNRs of different widths [19]. This work found that depending on the alignment of the AGNR segments at the junction, topological edge states can be present dependent on whether the two sides of the junction are topologically equivalent or inequivalent. This could have interesting applications both for spintronic devices and optimizing transport in the FET-like device described in this thesis.

Phase-field simulator development

Several enhancements to the phase-field simulation code would greatly enhance the ability to simulate ferroelectric negative capacitance devices.

GPU support

The linear solvers used by the elasticity and electrostatics modules of the simulator are from the Paralution package [78]. Currently, we are only using the multicore CPU solvers

parallelized through OpenMP, but Paralution also offers CUDA/OpenCL solvers that can run on graphics processing units (GPUs). This was never fully explored due to lack of available hardware, but it may be possible to get significant performance improvements for large problems by switching to GPU-based solvers.

Finite element implementation with adaptive mesh refinement

Efficient phase-field simulations of very large ferroelectric systems may be possible through using the finite element method (FEM). A key advantage of using FEM over finite difference solvers is the ability to have lower mesh density far away from regions of interest. This is somewhat problematic for our ferroelectric simulations because the domain walls move during the simulation and the regions of interest change over time. A possible solution to this would be to use adaptive mesh refinement. A finite element implementation of phase-field simulation for ferroelectrics has previously been done by the Ferret application [94] based on the open-source MOOSE multiphysics package [127].

Full-device negative capacitance FET simulations

Currently the phase-field simulator only simulates what happens within the ferroelectric with any degree of sophistication. It should be possible to model a negative capacitance field-effect transistor (NC-FET) by coupling the phase-field model to a semiconductor transport solver through a common electrostatics solver. Some preliminary attempt was made to do this through coupling a simplified quantum transport simulator, but ultimately the performance of such a model was not good enough to warrant further work. Another approach that was mentioned in the electrostatics solver section of this chapter was to approximate the equilibrium charge in the semiconductor using just the density of states and the potential. While this is more computationally efficient, it can only simulate equilibrium conditions.

The best way to accomplish this would be to couple the electrostatics solver with a drift-diffusion or hydrodynamic transport solver such as those described by Vasileska and Goodnick [131]. This could provide a realistic model for an NC-FET that takes into account the multidomain nature of the ferroelectric in a way only limited by the validity of the phase-field model. This could be done through extension of an existing simulator, such as those sold by Sentaurus or Silvaco, or through writing a new, simplified simulation code. It may also be possible to build a drift-diffusion or hydrodynamic solver by extending an open-source computational fluid dynamics (CFD) package, such as OpenFOAM [65], to solve semiconductor transport equations, which are very similar in form to CFD equations.

Bibliography

[1] Pablo Aguado-Puente and Javier Junquera. "Ferromagneticlike closure domains in ferroelectric ultrathin films: First-principles simulations". In: *Physical Review Letters* 100.17 (2008), pp. 1–4.

[2] Patrick R Amestoy, Iain S Duff, and Jacko Koster. "MUMPS: a general purpose distributed memory sparse solver". In: ().

[3] MP Anantram, Mark S Lundstrom, and Dmitri E Nikonov. "Modeling of nanoscale devices". In: *Proceedings of the IEEE* 96.9 (2008), pp. 1511–1550.

[4] Daniel J R Appleby et al. "Experimental observation of negative capacitance in ferroelectrics at room temperature". In: *Nano Letters* 14.7 (2014), pp. 3864–3868. ISSN: 15306992. DOI: 10.1021/nl5017255.

[5] A Artemev and A Roytburd. "Spinodal single-domain polydomain transition and P–E hysteresis in thin ferroelectric films". In: *Acta Materialia* 58.3 (2010), pp. 1004–1015.

[6] A. Artemev et al. "Thermodynamic analysis and phase field modeling of domain structures in bilayer ferroelectric thin films". In: *Journal of Applied Physics* 103.7 (2008). ISSN: 00218979.

[7] Khalid Ashraf and Sayeef Salahuddin. "Phase field model of domain dynamics in micron scale, ultrathin ferroelectric films: Application for multiferroic bismuth ferrite". In: *Journal of Applied Physics* 112.7 (2012). ISSN: 00218979. DOI: 10.1063/1.4754800.

[8] Khalid Ashraf, Samuel Smith, and Sayeef Salahuddin. "Electric field induced magnetic switching at room temperature: Switching speed, device scaling and switching energy". In: *Technical Digest - International Electron Devices Meeting, IEDM* (2012), pp. 601–604. ISSN: 01631918. DOI: 10.1109/IEDM.2012.6479109.

[9] Md Khalid Ashraf. "Switching dynamics of thin film ferroelectric devices-a massively parallel phase field study". PhD thesis. University of California, Berkeley, 2013.

[10] Satish Balay et al. "Efficient Management of Parallelism in Object Oriented Numerical Software Libraries". In: *Modern Software Tools in Scientific Computing*. Ed. by E. Arge, A. M. Bruaset, and H. P. Langtangen. Birkhäuser Press, 1997, pp. 163–202.

[11] Fabrice Barbe et al. "Intergranular and intragranular behavior of polycrystalline aggregates. Part 1: FE model". In: *International journal of plasticity* 17.4 (2001), pp. 513–536.

[12] G Bernstein and DK Ferry. "Negative differential conductivity in lateral surface superlattices". In: *Journal of Vacuum Science & Technology B: Microelectronics Processing and Phenomena* 5.4 (1987), pp. 964–966.

[13] Debanjan Bhowmik, Long You, and Sayeef Salahuddin. "Spin Hall effect clocking of nanomagnetic logic without a magnetic field". In: *Nature nanotechnology* 9.1 (2014), pp. 59–63.

[14] Timothy B Boykin et al. "Accurate six-band nearest-neighbor tight-binding model for the π-bands of bulk graphene and graphene nanoribbons". In: *Journal of Applied Physics* 109.10 (2011), p. 104304.

[15] Jinming Cai et al. "Atomically precise bottom-up fabrication of graphene nanoribbons". In: *Nature* 466.7305 (2010), pp. 470–473.

[16] Jinming Cai et al. "Graphene nanoribbon heterojunctions". In: *Nature nanotechnology* 9.11 (2014), pp. 896–900.

[17] A. Cano and D. Jiménez. "Multidomain ferroelectricity as a limiting factor for voltage amplification in ferroelectric field-effect transistors". In: *Applied Physics Letters* 97.13 (2010), pp. 48–51. ISSN: 00036951.

[18] A Cano and D Jimenez. "Multidomain ferroelectricity as a limiting factor for voltage amplification in ferroelectric field-effect transistors". In: *Applied Physics Letters* 97.13 (2010), p. 133509.

[19] Ting Cao, Fangzhou Zhao, and Steven G Louie. "Topological Phases in Graphene Nanoribbons: Junction States, Spin Centers and Quantum Spin Chains". In: *arXiv preprint arXiv:1702.02674* (2017).

[20] Premi Chandra and Peter B Littlewood. "A Landau primer for ferroelectrics". In: *Physics of ferroelectrics*. Springer, 2007, pp. 69–116.

[21] Long Qing Chen. "Phase-Field Models for Microstructure Evolution". In: *Annual Review of Materials Research* 32.1 (2002), pp. 113–140. ISSN: 1531-7331. DOI: 10.1146/annurev.matsci.32.112001.132041. URL: http://www.annualreviews.org/doi/abs/10.1146/annurev.matsci.32.112001.132041.

[22] Yen-Chia Chen et al. "Molecular bandgap engineering of bottom-up synthesized graphene nanoribbon heterojunctions". In: *Nature nanotechnology* 10.2 (2015), pp. 156–160.

[23] Yen-Chia Chen et al. "Tuning the band gap of graphene nanoribbons synthesized from molecular precursors". In: *ACS nano* 7.7 (2013), pp. 6123–6128.

[24] Cédric Chevalier and François Pellegrini. "PT-Scotch: A tool for efficient parallel graph ordering". In: *Parallel computing* 34.6 (2008), pp. 318–331.

[25] KK Choi et al. "Periodic negative conductance by sequential resonant tunneling through an expanding high-field superlattice domain". In: *Physical Review B* 35.8 (1987), p. 4172.

[26] Ryan R Cloke et al. "Site-specific substitutional boron doping of semiconducting armchair graphene nanoribbons". In: *Journal of the American Chemical Society* 137.28 (2015), pp. 8872–8875.

[27] Thomas H Cormen. *Introduction to algorithms*. MIT press, 2009.

[28] Supriyo Datta. *Electronic transport in mesoscopic systems*. Cambridge university press, 1997.

[29] Supriyo Datta. "Nanoscale device modeling: the Greens function method". In: *Superlattices and microstructures* 28.4 (2000), pp. 253–278.

[30] Supriyo Datta. *Quantum transport: atom to transistor*. Cambridge University Press, 2005.

[31] Mark De Berg et al. "Computational geometry". In: *Computational geometry*. Springer, 2000, pp. 1–17.

[32] James W Demmel. *Applied numerical linear algebra*. SIAM, 1997.

[33] L Esaki and LL Chang. "New Transport Phenomenon in a Semiconductor" Superlattice"". In: *Physical Review Letters* 33.8 (1974), p. 495.

[34] Leo Esaki and Ray Tsu. "Superlattice and negative differential conductivity in semiconductors". In: *IBM Journal of Research and Development* 14.1 (1970), pp. 61–65.

[35] Carlos A. Felippa. *The Linear Tetrahedron*. URL: `http://www.colorado.edu/engineering/CAS/courses.d/AFEM.d/AFEM.Ch09.d/AFEM.Ch09.pdf`.

[36] Gerson J Ferreira et al. "Low-bias negative differential resistance in graphene nanoribbon superlattices". In: *Physical Review B* 84.12 (2011), p. 125453.

[37] Gianluca Fiori and Giuseppe Iannaccone. "Simulation of graphene nanoribbon field-effect transistors". In: *IEEE Electron Device Letters* 28.8 (2007), pp. 760–762.

[38] Gianluca Fiori, Giuseppe Iannaccone, and Gerhard Klimeck. "A three-dimensional simulation study of the performance of carbon nanotube field-effect transistors with doped reservoirs and realistic geometry". In: *IEEE Transactions on electron devices* 53.8 (2006), pp. 1782–1788.

[39] Massimo V Fischetti and Steven E Laux. "Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects". In: *Physical Review B* 38.14 (1988), p. 9721.

[40] David J. Frank et al. "The quantum metal ferroelectric field-effect transistor". In: *IEEE Transactions on Electron Devices* 61.6 (2014), pp. 2145–2153. ISSN: 00189383. DOI: `10.1109/TED.2014.2314652`.

[41] Kartik Ganapathi and Sayeef Salahuddin. "Heterojunction vertical band-to-band tunneling transistors for steep subthreshold swing and high on current". In: *IEEE Electron Device Letters* 32.5 (2011), pp. 689–691.

[42] Kartik Ganapathi, Youngki Yoon, and Sayeef Salahuddin. "Analysis of InAs vertical and lateral band-to-band tunneling transistors: Leveraging vertical tunneling for improved performance". In: *Applied Physics Letters* 97.3 (2010), p. 033504.

[43] Kartik Ganapathi et al. "Ballistic IV characteristics of short-channel graphene field-effect transistors: Analysis and optimization for analog and RF applications". In: *IEEE Transactions on Electron Devices* 60.3 (2013), pp. 958–964.

[44] Weiwei Gao et al. "Room-temperature negative capacitance in a ferroelectric–dielectric superlattice heterostructure". In: *Nano Letters* 14.10 (2014), pp. 5814–5819.

[45] Weiwei Gao et al. "Room temperature negative capacitance in a ferroelectric-dielectric superlattice heterostructure Room temperature negative capacitance in a ferroelectric-dielectric superlattice heterostructure". In: *Nano Letters* 14 (2014), pp. 5814–5819. ISSN: 1530-6992. DOI: 10.1021/nl502691u.

[46] Paolo Giannozzi et al. "QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials". In: *Journal of physics: Condensed matter* 21.39 (2009), p. 395502.

[47] Eduardo Costa Girao, Eduardo Cruz-Silva, and Vincent Meunier. "Electronic transport properties of assembled carbon nanoribbons". In: *ACS nano* 6.7 (2012), pp. 6483–6491.

[48] E Gnani et al. "Performance limits of superlattice-based steep-slope nanowire FETs". In: *Electron Devices Meeting (IEDM), 2011 IEEE International*. IEEE. 2011, pp. 5–1.

[49] HT Grahn et al. "Electric-field domains in semiconductor superlattices: a novel system for tunneling between 2D systems". In: *Physical review letters* 67.12 (1991), p. 1618.

[50] Tibor Grasser et al. "A review of hydrodynamic and energy-transport models for semiconductor device simulation". In: *Proceedings of the IEEE* 91.2 (2003), pp. 251–274.

[51] Gaël Guennebaud and Benoît Jacob. *Eigen v3*. http://eigen.tuxfamily.org. 2010.

[52] Jing Guo. "Modeling of graphene nanoribbon devices". In: *Nanoscale* 4.18 (2012), pp. 5538–5548.

[53] Michael Heroux et al. *An Overview of Trilinos*. Tech. rep. SAND2003-2927. Sandia National Laboratories, 2003.

[54] U Hetmaniuk, Y Zhao, and MP Anantram. "A nested dissection approach to modeling transport in nanodevices: Algorithms and applications". In: *International Journal for Numerical Methods in Engineering* 95.7 (2013), pp. 587–607.

[55] Alan C Hindmarsh et al. "SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers". In: *ACM Transactions on Mathematical Software (TOMS)* 31.3 (2005), pp. 363–396.

[56] Digh Hisamoto et al. "FinFET-a self-aligned double-gate MOSFET scalable to 20 nm". In: *IEEE Transactions on Electron Devices* 47.12 (2000), pp. 2320–2325.

[57] Michael Hoffmann et al. "Direct observation of negative capacitance in polycrystalline ferroelectric HfO2". In: *Advanced Functional Materials* 26.47 (2016), pp. 8643–8649.

[58] S-M Hong and C Jungemann. "A fully coupled scheme for a Boltzmann-Poisson equation solver based on a spherical harmonics expansion". In: *Journal of computational electronics* 8.3 (2009), pp. 225–241.

[59] Jun Z Huang et al. "Model Order Reduction for Quantum Transport Simulation of Band-To-Band Tunneling Devices". In: (2014).

[60] Adrian M Ionescu and Heike Riel. "Tunnel field-effect transistors as energy-efficient electronic switches". In: *Nature* 479.7373 (2011), pp. 329–337.

[61] K Ismail et al. "Negative transconductance and negative differential resistance in a grid-gate modulation-doped field-effect transistor". In: *Applied physics letters* 54.5 (1989), pp. 460–462.

[62] Carlo Jacoboni and Lino Reggiani. "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials". In: *Reviews of Modern Physics* 55.3 (1983), p. 645.

[63] Mathias Jacquelin, Lin Lin, and Chao Yang. "PSelInv–A Distributed Memory Parallel Algorithm for Selected Inversion: the Symmetric Case". In: *arXiv preprint arXiv:1404.0447* (2014).

[64] Ankit Jain and Muhammad Ashraful Alam. "Proposal of a Hysteresis-Free Zero Subthreshold Swing Field-Effect Transistor". In: *IEEE Transactions on Electron Devices* 61.10 (2014), pp. 3546–3552. ISSN: 0018-9383. DOI: 10.1109/TED.2014.2347968. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6894595.

[65] Hrvoje Jasak, Aleksandar Jemcov, Zeljko Tukovic, et al. "OpenFOAM: A C++ library for complex physics simulations". In:

[66] Jaesung Jo and Changhwan Shin. "Negative capacitance field effect transistor with hysteresis-free sub-60-mV/decade switching". In: *IEEE Electron Device Letters* 37.3 (2016), pp. 245–248.

[67] Jaesung Jo et al. "Negative capacitance in organic/ferroelectric capacitor to implement steep switching MOS devices". In: *Nano letters* 15.7 (2015), pp. 4553–4556.

[68] George Karypis and Vipin Kumar. "A fast and high quality multilevel scheme for partitioning irregular graphs". In: *SIAM Journal on scientific Computing* 20.1 (1998), pp. 359–392.

[69] J Kastrup et al. "Multistability of the current-voltage characteristics in doped GaAs-AlAs superlattices". In: *Applied physics letters* 65.14 (1994), pp. 1808–1810.

[70] Asif Islam Khan et al. "Experimental evidence of ferroelectric negative capacitance in nanoscale heterostructures". In: *Applied Physics Letters* 99.11 (2011), pp. 27–30. ISSN: 00036951.

[71] Asif Islam Khan et al. "Experimental evidence of ferroelectric negative capacitance in nanoscale heterostructures". In: *Applied Physics Letters* 99.11 (2011), p. 113501.

[72] Asif Islam Khan et al. "Negative capacitance in a ferroelectric capacitor". In: *Nature Materials* 14.2 (2015), pp. 182–186.

[73] Asif Islam Khan et al. "Negative capacitance in short-channel FinFETs externally connected to an epitaxial ferroelectric capacitor". In: *IEEE Electron Device Letters* 37.1 (2016), pp. 111–114.

[74] SungGeun Kim et al. "Computational study of heterojunction graphene nanoribbon tunneling transistors with pd orbital tight-binding method". In: *Applied Physics Letters* 104.24 (2014), p. 243113.

[75] Masaharu Kobayashi et al. "Experimental study on polarization-limited operation speed of negative capacitance FET with ferroelectric HfO 2". In: *Electron Devices Meeting (IEDM), 2016 IEEE International*. IEEE. 2016, pp. 12–3.

[76] Siyuranga O Koswatta et al. "Nonequilibrium Green's function treatment of phonon scattering in carbon-nanotube transistors". In: *IEEE Transactions on Electron Devices* 54.9 (2007), pp. 2339–2351.

[77] Aleksandr Semenovich Kronrod. "Nodes and weights of quadrature formulas". In: (1965).

[78] PARALUTION Labs. *PARALUTION v1.0.0*. `http://www.paralution.com/`. 2015.

[79] Roger Lake et al. "Single and multiband modeling of quantum electron transport through layered semiconductor devices". In: *Journal of Applied Physics* 81.12 (1997), pp. 7845–7869.

[80] Kai-Tak Lam, Yee-Chia Yeo, and Gengchiau Liang. "Performance comparison of III-V MOSFETs with source filter for electron energy". In: *Electron Devices Meeting (IEDM), 2012 IEEE International*. IEEE. 2012, pp. 17–6.

[81] LD Landau and IM Khalatnikov. "On the anomalous absorption of sound near a second order phase transition point". In: *Dokl. Akad. Nauk SSSR*. Vol. 96. 1954, pp. 469–472.

[82] MH Lee et al. "Physical thickness 1. x nm ferroelectric HfZrOx negative capacitance FETs". In: *Electron Devices Meeting (IEDM), 2016 IEEE International*. IEEE. 2016, pp. 12–1.

[83] P Lenarczyk and M Luisier. "Physical modeling of ferroelectric field-effect transistors in the negative capacitance regime". In: *Simulation of Semiconductor Processes and Devices (SISPAD), 2016 International Conference on*. IEEE. 2016, pp. 311–314.

[84] Kai-Shin Li et al. "Sub-60mV-swing negative-capacitance FinFET without hysteresis". In: *Electron Devices Meeting (IEDM), 2015 IEEE International*. IEEE. 2015, pp. 22–6.

[85] S Li, S Ahmed, and E Darve. "Fast inverse using nested dissection for NEGF". In: *Journal of Computational Electronics* 6.1-3 (2007), pp. 187–190.

[86] Suchun Li et al. "Low-bias negative differential resistance effect in armchair graphene nanoribbon junctions". In: *Applied Physics Letters* 106.1 (2015), 2_1.

[87] Xiaoye S Li and James W Demmel. "SuperLU_DIST: A scalable distributed-memory sparse direct solver for unsymmetric linear systems". In: *ACM Transactions on Mathematical Software (TOMS)* 29.2 (2003), pp. 110–140.

[88] Y. L. Li et al. "Phase-field model of domain structures in ferroelectric thin films". In: *Applied Physics Letters* 78.24 (2001), pp. 3878–3880. ISSN: 00036951. DOI: 10.1063/1.1377855.

[89] Lin Lin et al. "Fast algorithm for extracting the diagonal of the inverse matrix with application to the electronic structure analysis of metallic systems". In: *Communications in Mathematical Sciences* 7.3 (2009), pp. 755–777.

[90] Lin Lin et al. "SelInv—An Algorithm for Selected Inversion of a Sparse Symmetric Matrix". In: *ACM Transactions on Mathematical Software (TOMS)* 37.4 (2011), p. 40.

[91] Malcolm E Lines and Alastair M Glass. *Principles and applications of ferroelectrics and related materials*. Oxford university press, 1977.

[92] Juan Pablo Llinás et al. "Short-Channel Field Effect Transistors with 9-Atom and 13-Atom wide Graphene Nanoribbons". In: *arXiv preprint arXiv:1605.06730* (2016).

[93] Mathieu Luisier and Gerhard Klimeck. "OMEN an atomistic and full-band quantum transport simulator for post-CMOS nanodevices". In: *Nanotechnology, 2008. NANO'08. 8th IEEE Conference on*. IEEE. 2008, pp. 354–357.

[94] John Mangeri et al. "Topological phase transformations and intrinsic size effects in ferroelectric nanoparticles". In: *Nanoscale* (2017).

[95] Felicia A McGuire et al. "Sub-60 mV/decade switching in 2D negative capacitance field-effect transistors with integrated ferroelectric polymer". In: *Applied Physics Letters* 109.9 (2016), p. 093101.

[96] Varun Mishra et al. "Dependence of intrinsic performance of transition metal dichalcogenide transistors on materials and number of layers at the 5 nm channel-length limit". In: *Electron Devices Meeting (IEDM), 2013 IEEE International*. IEEE. 2013, pp. 5–6.

[97] Varun Mishra et al. "Screening in ultrashort (5 nm) channel MoS 2 transistors: A full-band quantum transport study". In: *IEEE Transactions on Electron Devices* 62.8 (2015), pp. 2457–2463.

[98] I. B. Misirlioglu. "Stability of a ferroelectric phase with electrical domains in multilayers". In: *Applied Physics Letters* 94.17 (2009), pp. 44–47. ISSN: 00036951. DOI: 10.1063/1.3129872.

[99] Arash A Mostofi et al. "wannier90: A tool for obtaining maximally-localised Wannier functions". In: *Computer physics communications* 178.9 (2008), pp. 685–699.

[100] James R Nagel. "Solving the Generalized Poisson Equation with the Finite-Difference Method". In: (2012).

[101] S Natarajan et al. "A 14nm logic technology featuring 2 nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm 2 SRAM cell size". In: *Electron Devices Meeting (IEDM), 2014 IEEE International*. IEEE. 2014, pp. 3–7.

[102] Giang D Nguyen et al. "Bottom-Up Synthesis of N= 13 Sulfur-Doped Graphene Nanoribbons". In: *The Journal of Physical Chemistry C* 120.5 (2016), pp. 2684–2687.

[103] Amirhasan Nourbakhsh et al. "Subthreshold swing improvement in MoS 2 transistors by the negative-capacitance effect in a ferroelectric Al-doped-HfO 2/HfO 2 gate dielectric stack". In: *Nanoscale* 9.18 (2017), pp. 6122–6127.

[104] Kostya S Novoselov et al. "Two-dimensional gas of massless Dirac fermions in graphene". In: *nature* 438.7065 (2005), pp. 197–200.

[105] Yijian Ouyang et al. "Carrier scattering in graphene nanoribbon field-effect transistors". In: *Applied Physics Letters* 92.24 (2008), p. 243124.

[106] Himadri Pal, Dmitri Nikonov, and George Bourianoff. *Scattering in NEGF: Made simple*. 2009.

[107] Dan Erik Petersen et al. "A hybrid method for the parallel computation of Greens functions". In: *Journal of Computational Physics* 228.14 (2009), pp. 5020–5039.

[108] Vincent Pott et al. "Mechanical computing redux: Relays for integrated circuit applications". In: *Proceedings of the IEEE* 98.12 (2010), pp. 2076–2094.

[109] Nikolas Provatas and Ken Elder. *Phase-field methods in materials science and engineering*. John Wiley & Sons, 2011.

[110] Peter Pulay. "Improved SCF convergence acceleration". In: *Journal of Computational Chemistry* 3.4 (1982), pp. 556–560.

[111] K. M Rabe et al. *Physics of Ferroelectrics*. Vol. 105. 2007. 2007. ISBN: 978-3-540-34590-9. DOI: 10.1007/978-3-540-34591-6. URL: http://www.springerlink.com/index/10.1007/978-3-540-34591-6.

[112] Karin M Rabe, Charles H Ahn, and Jean-Marc Triscone. *Physics of Ferroelectrics: a Modern Perspective*. Vol. 105. Springer Science & Business Media, 2007.

[113] Andre Ribes and Christian Caremoli. "Salome platform component model for numerical simulation". In: *Computer Software and Applications Conference, 2007. COMPSAC 2007. 31st Annual International*. Vol. 2. IEEE. 2007, pp. 553–564.

[114] Pascal Ruffieux et al. "On-surface synthesis of graphene nanoribbons with zigzag edge topology". In: *Nature* 531.7595 (2016), pp. 489–492.

[115] Gobinda Saha, Atanu Kumar Saha, and ABM Harun-ur Rashid. "Double quantum well resonant tunneling negative differential resistance device design using graphene nanoribbons". In: *Nanotechnology (IEEE-NANO), 2015 IEEE 15th International Conference on*. IEEE. 2015, pp. 440–443.

[116] Sayeef Salahuddin and Supriyo Datta. "Use of negative capacitance to provide voltage amplification for low power nanoscale devices". In: *Nano Letters* 8.2 (2008), pp. 405–410.

[117] Sayeef Salahuddin and Supriyo Datta. "Use of negative capacitance to provide voltage amplification for low power nanoscale devices". In: *Nano Letters* 8.2 (2008), pp. 405–410. ISSN: 15306984. DOI: 10.1021/nl071804g.

[118] MP Lopez Sancho et al. "Highly convergent schemes for the calculation of bulk and surface Green functions". In: *Journal of Physics F: Metal Physics* 15.4 (1985), p. 851.

[119] H Sevinçli, M Topsakal, and S Ciraci. "Superlattice structures of graphene-based armchair nanoribbons". In: *Physical Review B* 78.24 (2008), p. 245402.

[120] M Sharifi, E Akhoundi, and H Esmaili. "Negative differential resistance in new structures based on graphene nanoribbons". In: *Journal of Computational Electronics* (2016), pp. 1–9.

[121] John C Slater and George F Koster. "Simplified LCAO method for the periodic potential problem". In: *Physical Review* 94.6 (1954), p. 1498.

[122] Young-Woo Son, Marvin L Cohen, and Steven G Louie. "Energy gaps in graphene nanoribbons". In: *Physical review letters* 97.21 (2006), p. 216803.

[123] Sebastian Steiger et al. "NEMO5: A parallel multiscale nanoelectronics modeling tool". In: (2011).

[124] Y. L. Tang et al. "Observation of a periodic array of flux-closure quadrants in strained ferroelectric PbTiO3 films". In: *Science* (2015). ISSN: 0036-8075. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1259869.

[125] Yuan Taur and Tak H Ning. *Fundamentals of modern VLSI devices*. Cambridge university press, 2013.

[126] Hansen Teong et al. "Shape effects in graphene nanoribbon resonant tunneling diodes: A computational study". In: *Journal of Applied Physics* 105.8 (2009), p. 084317.

[127] Michael R Tonks et al. "An object-oriented finite element framework for multiphysics phase field simulations". In: *Computational Materials Science* 51.1 (2012), pp. 20–29.

[128] Luis EF Foa Torres, Stephan Roche, and Jean-Christophe Charlier. *Introduction to graphene-based nanomaterials: from electronic structure to quantum transport*. Cambridge University Press, 2014.

[129] A Trellakis et al. "Iteration scheme for the solution of the two-dimensional Schrödinger-Poisson equations in quantum structures". In: *Journal of Applied Physics* 81.12 (1997), pp. 7880–7884.

[130] A Umerski. "Closed-form solutions to surface Green's functions". In: *Physical Review B* 55.8 (1997), p. 5266.

[131] Dragica Vasileska and Stephen M Goodnick. "Computational electronics". In: *Synthesis Lectures on Computational Electromagnetics* 1.1 (2005), pp. 1–216.

[132] Philip Richard Wallace. "The band theory of graphite". In: *Physical Review* 71.9 (1947), p. 622.

[133] Shudong Wang and Jinlan Wang. "Quasiparticle energies and optical excitations in chevron-type graphene nanoribbon". In: *The Journal of Physical Chemistry C* 116.18 (2012), pp. 10193–10197.

[134] AC Warren et al. "Surface superlattice formation in silicon inversion layers using 0.2-μm period grating-gate electrodes". In: *IEEE electron device letters* 6.6 (1985), pp. 294–296.

[135] C M Wu et al. "Controllability of vortex domain structure in ferroelectric nanodot: fruitful domain patterns and transformation paths." In: *Scientific reports* 4 (2014), p. 3946. ISSN: 2045-2322. DOI: 10.1038/srep03946. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3912473\&tool=pmcentrez\&rendertype=abstract.

[136] Li Yang et al. "Quasiparticle energies and band gaps in graphene nanoribbons". In: *Physical Review Letters* 99.18 (2007), p. 186801.

[137] Milo Yip. *RapidJSON: A fast JSON parser/generator for C++ with both SAX/DOM style API*. https://github.com/miloyip/rapidjson. 2015.

[138] Youngki Yoon, Kartik Ganapathi, and Sayeef Salahuddin. "How good can monolayer MoS2 transistors be?" In: *Nano letters* 11.9 (2011), pp. 3768–3773.

[139] Youngki Yoon and Jing Guo. "Effect of edge roughness in graphene nanoribbon transistors". In: *Applied Physics Letters* 91.7 (2007), p. 073103.

[140] Youngki Yoon, Dmitri E Nikonov, and Sayeef Salahuddin. "Role of phonon scattering in graphene nanoribbon transistors: Nonequilibrium Greens function method with real space approach". In: *Applied Physics Letters* 98.20 (2011), p. 203503.

[141] Youngki Yoon and Sayeef Salahuddin. "Barrier-free tunneling in a carbon heterojunction transistor". In: *Applied Physics Letters* 97.3 (2010), p. 033102.

[142] Youngki Yoon and Sayeef Salahuddin. "Dissipative transport in rough edge graphene nanoribbon tunnel transistors". In: *Applied Physics Letters* 101.26 (2012), p. 263501.

[143] Jiuren Zhou et al. "Ferroelectric HfZrO x Ge and GeSn PMOSFETs with Sub-60 mV/decade subthreshold swing, negligible hysteresis, and improved I ds". In: *Electron Devices Meeting (IEDM), 2016 IEEE International*. IEEE. 2016, pp. 12–2.

[144] Tarek I Zohdi. *A Finite Element Primer for Beginners: The Basics*. Springer, 2015.

[145] Pavlo Zubko et al. "Negative capacitance in multidomain ferroelectric superlattices". In: *Nature* (2016).