THE UNIVERSITY OF CHICAGO

FIRST-PRINCIPLES SIMULATIONS OF QUANTUM ELECTRON TRANSPORT IN TWO-DIMENSIONAL SEMICONDUCTOR NANODEVICES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
WUSHI DONG

CHICAGO, ILLINOIS
AUGUST 2019

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my advisor Prof. Peter Littlewood for the opportunity and for the freedom he gave me in my Ph.D study and related research. I am thankful for his trust, patience, and guidance. I am deeply inspired by his enthusiasm for big science problems and his way to approach them. I feel blessed to have him as my advisor, my mentor and my friend.

I would also like to thank Prof. Supratik Guha, Prof. Woowon Kang and Prof. Young-Kee Kim, for serving on my advisory committee and providing valuable feedbacks. It was very challenging yet invaluable experience to present and defend my work in front of some of the world's most renowned researchers.

My gratitude goes to Prof. Jiwoong Park and Prof. Saptarshi Das, Dr. Alejandro Lopez-Bezanilla for their collaborations in our researches. It has been very fortunate for me to learn from and work with top-notch scientists in the field I study.

I would also like to acknowledge Hui Gao, Dr. Marcos Guimarães, Dr. Ryo Hanai, and Alex Edelman for helpful discussions.

Finally, I would like to express my gratitude to my family and friends, especially my parents for the forever love and support they always give me.

ABSTRACT

In this thesis, a simulation pipeline for efficient and accurate atomistic calculations of electron transport in nanoscale devices is developed. This method is based on the non-equilibrium Green's function (NEGF) formalism with tight-binding parameters of the considered materials determined from electronic structures by density-functional theory (DFT) calculations. DFT simulation is a robust technique to model nanostructures, but cannot be scaled to realistic device sizes due to heavy computational cost. We circumvented this limitation by transforming the delocalized plane-wave states into maximally localized Wannier functions (MLWFs) that serve as the localized basis for the quantum transport solver. This allows accurate modeling of device structures on a micron scale, but with atomic level accuracy.

The effectiveness of this approach is demonstrated through the investigation of nanostructures and the comparison with experimental results. Firstly, in order to validate our approach, we compared the transport results obtained by our method with that by full DFT simulation. The two methods agree very well but our method uses three orders of magnitude less time. Then we tested our transport calculations by applying it to the telescopic double wall carbon nanotube, where two nanotube of different radius overlap with each other. The obtained results are similar to the ones in literature.

We then applied our simulation pipeline to the important problem of metal-semiconductor contact. Metal-semiconductor contact is a major factor limiting the shrinking of transistor dimension to further increase device performance. Two-dimensional (2D) materials such as graphene and transition metal dichalcogenides (TMDCs) are pushing the forefront of complementary metal-oxide semiconductor (CMOS) technology beyond the Moore's law [1, 2], and show great promises for realizing atomically thin circuitry [3, 4, 5]. A fundamental challenge to their effective use remains the large resistance of electrical contacts to 2D materials for probing and harnessing their novel electronic properties [6, 7, 8]. There are generally two types of contact geometries, namely top contacts and edge contacts [6], both of which

are examined in in this thesis. Conventional 3D metallic top contacts can achieve low contact resistance with monolayer 2D materials, but cannot avoid the intrinsic problem of large electrode volume. [9, 10, 6, 11] 2D top contacts, including graphene [12, 13, 14] and recently demonstrated atomically flat metal thin films [11], can achieve both small volumes and low contact resistances of metal-semiconductor interfaces. The analysis of graphene-MoS_2 top contacts reveals that they suffer from weak van der Waals coupling to TMDCs [15] so their transfer efficiency depends largely on the contact area and is compromised dramatically below a transfer length which is typically tens of nm scale [16, 8].

On the other hand, in-plane edge contacts have the potential to achieve lower contact resistance due to stronger orbital hybridization compared to conventional top contacts. We then present full-band atomistic quantum transport simulations of the graphene/MoS_2 edge contact. We find that the potential barrier created by trapped charges decays fast with distance away from the interface, and is thus thin enough to enable efficient injection of electrons. This results in Ohmic behavior in its I-V characteristics, which agrees with experiments. Our results demonstrate the role played by trapped charges in the formation of a Schottky barrier, and how one can reduce the Schottky barrier height (SBH) by adjusting the relevant parameters of the edge contact system.

The thesis provides full details on the application of the MLWF technique to self-consistent quantum transport simulations, as implemented in our open-source software `swan` [17]. Our framework can be extended conveniently to incorporate more general nanostructure geometries as well as electron-phonon interactions. Such approaches are important for understanding electron flow beyond the quantum limit and have started to draw increasing attention from the device modeling community.

CHAPTER 1

INTRODUCTION

1.1 Nanodevices

Transistor is at the heart of the third Industrial Revolution. It is today's counterpart of the steam engine in the 20th century. From the first demonstration of a point-contact transistor in 1947 by Bardeen, Shockley, and Brattain at Bell Labs, to the mass production of modern computer processors, the semiconductor industry has aggressively pushed its limits for decades. The ever growing computational power of the ever shrinking microchips has shaped our everyday life, from economy, science, healthcare, social life, or even civilization with the booming of artificial intelligence starting only in this decade.

The exponential growth of computer performance is a result of miniaturization of dimensions as described by the empirical law by Gordon Moore in 1965.[18] Moore's law states that the number of components per chip area doubles every two years. Up till today, this law is still going strong. The ability to pack hundreds of millions of logic switches onto a decreasing area has led to modern computing devices with far-reaching functionalities. Despite significant efforts in optimizing chip design, the scaling of number of transistors ultimately gave rise to modern digital technology.

Although the number of transistors per chip is going up, the clock speed[1] and thermal design power[2] are not keeping up with pace due to physical limits. To overcome these problems of ultra-scaled devices, Intel has already replaced planar MOSFETs with three-dimensional FinFETs, which could satisfy the industry requirements for the next 2-3 technology nodes. However, in the long run novel device architectures and materials will be needed to minimize

1. Clock speed means the operating speed of a computer or its microprocessor, defined as the rate at which it performs internal operations and expressed in cycles per second (megahertz).

2. The thermal design power (TDP), sometimes called thermal design point, is the maximum amount of heat generated by a computer chip or component (often a CPU, GPU or system on a chip) that the cooling system in a computer is designed to dissipate under any workload.

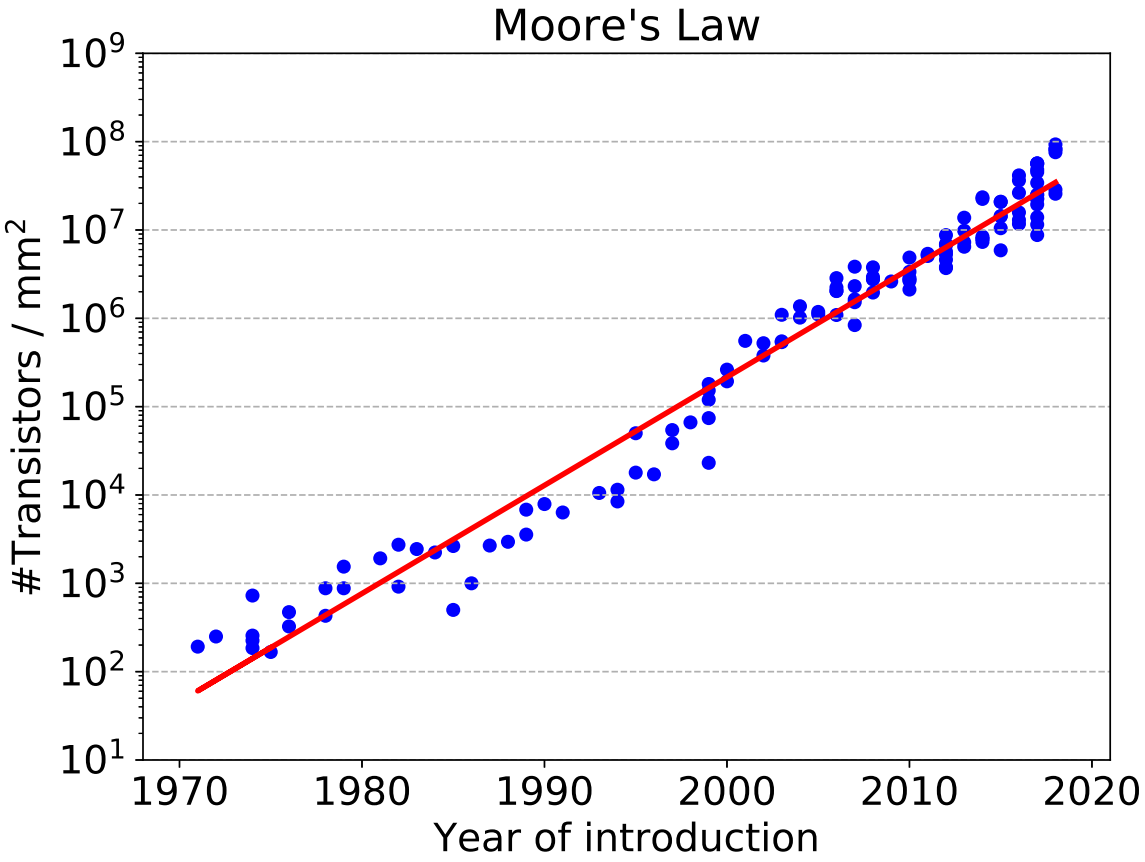


Figure 1.1: Transistor count per squared millimeter on a microprocessor obeys the Moore's law (1971-2018). Data is from Wikipedia (`https://en.wikipedia.org/wiki/Transistor_count#Microprocessors`).

current leakages, self-heating, power consumption, and maintain a good electrostatic control of the active region of the transistors at atomistic scale.

1.2 2D Materials

In recent years, 2D materials made of atomically thin structures has gained attention as an ideal candidate replacement of silicon in post-CMOS era. The discovery of graphene by Novoselov and his co-workers in 2004 [19] generated an unprecedented enthusiasm in material sciences. However, graphene lacks a band gap in its bandstructure, which hinders

its usage in digital logic applications. On the other hand, other families of 2D materials are emerging. Among them, few-layer metal-dichalcogenides (MDs), given by the chemical formula MX2, where M is a metal (Mo, W, Zr, Hf, Sn, etc.) and X a chalcogen (S, Se, Te), as well as their van der Waals crystals, hold great promises to face the challenges of present-day electronics given the existence of band gaps. Still, the integration of 2D materials into electric circuits is accompanied by several technical difficulties. For example, they suffer from a large contact resistance at the interface of bulk metal interconnects that severely limits their performance [20]. Most techniques developed for reducing the contact resistance at silicon-metal interfaces are not applicable here, due to the atomic scale of the semiconductor layer. Various computer simulations have been recently performed but they only studied the equilibrium properties of the interfaces [15], which is only one piece of the puzzle. This thesis tackles these challenges with realistic transport simulations in metals connected to 2D systems.

1.3 Computational modeling

Computational modeling plays an important role in the development of novel microchip technologies. In this sense, the semiconductor industry has also greatly benefited from its own product itself. Technology computer aided design (TCAD) integrating material simulation and device modeling has became standard in the semiconductor companies in supporting long and expensive experimental processes. Designing a new transistor strongly relies on this cost- and time-efficient method. To continue this success, we have to develop the abilities to develop next-generation computer simulation tools to accurately handle systems composed of hundreds of thousands of atoms on the atomistic, quantum-mechanical level. This is critical given the enormous design space to be explored in order to find the best active components that will hold the core of future's electronics, as the conventional silicon metal-oxide-semiconductor field-effect transistor (MOSFET) technology is approaching its

limits.

Given the device channel length as small as 1 nm [21], previous device model based mainly on effective mass approximation definitely becomes questionable. At this scale, we cannot characterize devices using a collective behavior. Atomistic effects including surface roughness, discrete dopant and impurity fluctuations will dramatically influence the device properties. Each device will be different from the rest, and variations due to imperfections will be a major challenge in controlling their operation. Therefore, more accurate atomistic models that inherently include all relevant information about the electronic structure of materials and structures should be utilized. Atomistic models can provide information about non-parabolicity, confinement level position beyond the effective mass approximation, the effects of strain in the electronic structure, as well as a more accurate distribution of charge in the device channel. Atomistic models generally take into account the full atomic structure of the system with each atom at a position either in a predefined geometry (like those given in the preceding sections), or computed within the model itself. In addition to that, atomistic description of the device in arbitrary orientations, has the advantage of being able to automatically capture the valley projections and extract the dispersions of the channels in the transport orientation. This model also automatically includes information about band coupling and mass variations as functions of quantization, although the quantitative significance of the results always depends on the parameters that are needed as an input. Compared to *ab-initio* methods, atomistic models can generally be handled with far less computational effort and allow an easy tuning of internal parameters, helping in understanding the individual physical effects.

For this purpose we have developed a simulation framework based on first-principles density-functional theory (DFT) and the Keldysh non-equilibrium Green's function (NEGF) technique[22, 23] connected through maximally localized Wannier functions (MLWFs)[24], an increasingly popular combination in the field of computational device simulations. This

thesis provides details of theoretical background and key features as well as examples for our custom-built computational pipeline. The codes implemented along with the presented examples are publicly available on GitHub.[17]

1.4 Overview

This thesis is organized as the following: Chapter 2 introduces the theory of quantum transport. Chapter 3 describes the way to calculate the electronic structures of materials using DFT methods and to extract tight-binding parameters for quantum transport simulations based on those calculations. To validate the developed simulation pipeline, we compare our results with full *ab-initio* calculation of quantum transport in vertical stacking graphene layers. We find that the two results match well, albeit our method is three orders of magnitude faster per node and has the potential of better scaling to even larger systems. In Chapter 4, We further compare our simulations of telescopic double wall carbon nanotube with literature results. We again find that the two agree well with each other. Chapter 5 and chapter 6 are dedicated to the important problem of metal-semiconductor contacts. Metal-semiconductor contact is a major factor limiting the shrinking of transistor dimension to further increase device performance. We investigated two main contact geometries for the combination of the two most popular 2D materials, namely metallic graphene and semiconducting monolayer MoS_2. In chapter 5, we calculated the transport properties of top contact structure and studied the relation between electron transfer efficiency and overlap length. In chapter 6, we performed a full self-consistent simulation of NEGF solver and Poisson solver for the edge contact geometry and find ohmic behavior in the current-voltage characteristics, which agrees with experiment. In chapter 7, We conclude our finds and suggest directions for future study.

CHAPTER 2

THEORY OF QUANTUM ELECTRON TRANSPORT SIMULATIONS

Simulation is a valuable tool for not only understanding complex phenomena but also predicting new directions for experimental investigations. Due to the complicated quantum mechanical nature of matter at the nanoscale, traditional classical or semi-classical models fall short of providing the required precision in realistic cases. Except in very few special cases, the many-body equations of quantum mechanics cannot be solved analytically. Instead, numerical solution is the way to calculate properties of systems without introducing extra parameters. Even numerical solution of the Schrödinger equation is a challenging problem, with solutions only possible for small systems. However, approximations to lower the computational cost are possible. In this chapter, we introduce the theory for various pieces of our assembled computational pipeline.

2.1 Density-Functional Theory

Ab-initio methods allow the quantitative computation of material properties without experimental input. The method commonly used within solid state physics for material simulation is DFT based on the theory by W. Kohn and L. J. Sham [25]. Remarkably, ground-state properties can be found without directly working with the many-body wavefunction; instead one works with the ground-state electron density. Formally, density functional theory is exact for the ground state. In practice a number of approximations to the functional need to be made, but the method has now achieved good accuracy for many materials. DFT has emerged as a standard tool in describing the electronic structure of materials. Its general advantage is its ability to produce accurate quantitative results, while the main disadvantage is the high computational cost. In the standard formulation it is best suited for the

description of isolated systems such as molecules and clusters or fully periodic systems such as solid crystals. However, computational study of some systems with existing frameworks is possible with additional development of computational methods. These factors lead to development of methods and formalism needed to calculate the electron transport properties of nanoscale systems.

In the following, we review the theoretical basics of DFT, covering the Hohenberg-Kohn theorem, the Kohn-Sham equations, pseudopotentials and exchange-correlation functionals. This is not intended to be a comprehensive overview as excellent tutorials including textbooks[26] are available.

The Hohenberg-Kohn (HK) theorem[27] provides the foundation for DFT: *For N electrons interacting in a static-external potential $V(\mathbf{r})$, the electronic density of the ground-state $n_0(\mathbf{r})$ minimizes the functional*

$$E[n(\mathbf{r})] = F[n(\mathbf{r})] + \int V(\mathbf{r})n(\mathbf{r})d\mathbf{r}. \tag{2.1}$$

The proof the of theorem can be found in most solid-state physics textbooks. [26]

Kohn and Sham separated the functional $F[n(\mathbf{r})]$ into three parts

$$E[n(\mathbf{r})] = T_s[n(\mathbf{r})] + \frac{1}{2}\int\int \frac{n(\mathbf{r})n(\mathbf{r'})}{|\mathbf{r}\mathbf{r'}|}d\mathbf{r}d\mathbf{r'} + E_{XC}[n(\mathbf{r})]. \tag{2.2}$$

The first term, $T_s[n(\mathbf{r})]$, is the kinetic energy of non-interacting electron gas, the second term is the Hartree (electrostatic) energy, and the last term is the exchange and correlation energy. The point of the separation is that the first two terms can be straightforwardly treated while the final term contains the complicated many-body effects. The exact form of the exchange and correlation term is unknown and approximations must be made case by case in practice in order to use this theory.

Then the Schrödinger equation for non-interacting particles can be obtained as introduced

by Kohn and Sham,

$$-\frac{1}{2}\nabla^2 + V_{KS}(\mathbf{r})\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}). \tag{2.3}$$

Here, the number of electrons in the system is fixed, enforced via a Lagrange multiplier. The previous equation provides a route to practical calculations using density functional theory and is known as the Kohn-Sham equation. Here the electron density is defined as

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\psi_i(\mathbf{r})|^2. \tag{2.4}$$

Conventional Coulomb potentials pose several computational difficulties. Since many states near atomic cores are highly localized and wavefunctions must be orthogonal, those states must rapidly oscillate to achieve orthogonality. This corresponds to a high plane-wave cutoff or very fine real-space grid spacings, which would be computationally expensive. The pseudopotential method replaces the Coulomb potential and core electrons with an effective potential. The basis of this idea is that core electrons are chemically inactive and can be treated using a constant potential.

A major issue in the practical application of DFT is that the exact form of the exchange and correlation functional is only known for simple cases such as the free electron gas. A comprehensive survey of functionals is beyond the scope of this thesis. Here I present the conceptually simple, but surprisingly accurate, local density approximation and end with some comments about more complex functionals.

Among different pseudopotentials, the local density approximation (LDA) is popular for being conceptually simple, but surprisingly accurate. This functional is derived from a homogeneous electron gas model. It depends only on the density at each point and is therefore fully local. The exchange part is known analytically, and the correlation part is known in the high and low density limits. Values between the limits are known from accurate Quantum Monte Carlo simulations.

Beyond LDA, there are many available functionals which include additional non-local information. There is a progression in terms of the trade-off between the amount of included information and computational cost, with fully local and and fully non-local representing the opposing ends of the spectrum. Generalized gradient (GGA) functional is another type of pseudopotential. It includes dependencies on the density and the gradient of the density. Modern developments include the addition of some percentage of Hartree-Fock exchange known as hybrid functionals, as well as the inclusion of van der Waals interactions in functionals.

2.2 The tight-binding approximation

2.2.1 Tight-binding formalism

At the atomistic scale, quantum mechanics of the electronic structure, crystal symmetry, atomic composition and spatial disorder become important. To use an atomistic model that can describe complicated man-made nanostructures, we choose to use a nearest neighbor model. To accurately capture electronic properties of materials, we need to extract hopping parameters from *ab-initio* results, instead of using traditional effective mass approximation. To simulate realistic structures containing tens of millions of atoms, we further have to minimize model complexities. Quantum conductance is computed using matrix form of Green's functions. To do this, we need to rely on a localized orbital representation. All these requirements speak for the choice of tight-binding (TB) formalism used in this work.

For accurate description of the electronic structures, a large number of models are based on TB method in the physics community, also known as linear combination of atomic orbitals (LCAO) in the chemistry community. Originally introduced by Bloch for the description of simple periodic structures, the method was later refined by J. C. Slater and G. F. Koster [28]. We have to look for a parameterization method that can transform intrinsically delocalized

Bloch orbitals into localized ones, in order to construct the sparse, short-ranged matrix elements of the Hamiltonian. Today, this method has a lot of applications including the one based on Wannier functions as introduced later. We first introduce the basics of TB method.

The first step of the tight binding (TB) approximation is to choose an atomic basis defined in real space as:

$$\chi_{inlm}(\mathbf{r}) = R_{inlm}(|\mathbf{r} - \mathbf{r_i}|)Y_{lm}(\theta(|\mathbf{r} - \mathbf{r_i}|), \phi(|\mathbf{r} - \mathbf{r_i}|)) \tag{2.5}$$

where $Y_{lm}(\theta, \phi)$ are the spherical harmonics. Index i, n (principal), l (angular) and m (magnetic) are the quantum number of the individual atom, and the three atomic quantum numbers respectively. R_{inlm} is the radial function and can be chosen in various ways.

The crucial step is the reduction to a very limited set of orbitals per atom, which turns this over-complete basis into a valuable approximation. The electronic properties are usually determined by a small number of orbitals near the Fermi energy. So projecting the Hamiltonian to these orbitals produces only a small error and cuts the computational cost.

The Hamiltonian of this system is given by its matrix elements in atomic basis

$$H_{inlm,i'n'l'm'} = \langle \chi_{inlm} | \hat{H} | \chi_{i'n'l'm'} \rangle \tag{2.6}$$

with each element called the hopping parameter. If this basis is non-orthogonal, this leads to non-diagonal entries in the overlap matrix

$$S_{inlm,i'n'l'm'} = \langle \chi_{inlm} | \chi_{i'n'l'm'} \rangle \tag{2.7}$$

In a non-orthogonal basis, the Schrödinger equation is expressed as:

$$\hat{H}\Psi = E\hat{S}\Psi, \tag{2.8}$$

and similarly for the Green's functions:

$$\hat{G}(E) = (E\hat{S} + i0^{+}\hat{H})^{-1}. \tag{2.9}$$

In the special case of the orthogonal TB-approximation the matrix S is set to identity and does not need to be considered further, which adds to the convenience of numerical implementation.

In general, a TB Hamiltonian can be obtained in various ways. The full many-particle Hamiltonian could be directly expressed in an atomic basis. However, that will lead to an interacting Hamiltonian that can only be solved using sophisticated techniques involving further approximations. For example, the Kohn-Sham Hamiltonian of a density-functional-theory calculation done in atomic orbitals using the SIESTA code [29] can be viewed directly as a TB-Hamiltonian. One should be aware that the Kohn-Sham theorem assigns a physical meaning only to the total energy obtained from the effective single-particle Hamiltonian and the energy of the highest occupied band. The other quantities should be treated with care.

The TB Hamiltonian are adjusted to either experiment or certain results from *ab-initio* computations. To achieve transferability of the obtained parameterization, the values can be assumed to follow a simple functional form depending on the geometry. They can be used for alternative geometries once the parameters of these functions are determined. One approach is proposed by Slater and Koster [28], who assume a dependence of the two-body hopping integral on the distance between the two involved atoms only. One excellent collection of Slater-Koster (SK) parameters for a wide range of elements was build up by Papaconstantopoulos and Mehl [30].

2.2.2 Maximally Localized Wannier Functions

The core of our proposed methodology is to use maximally-localized Wannier functions (MLWFs) for the system in considerations. They are a set of localized orbitals that resembles real atomic orbitals. They span the same Hilbert space of the Hamiltonian eigenfunctions, and allow to bridge first-principal electronic structure produced by plane-waves and lattice Green's function calculations in a coherent fashion.

There are several advantages of using MLWFs as a basis set for transport calculations [31]: (i) the MLWFs are spatially localized suitable for turning Green's functions into matrices. (ii) any eigenstate within a certain specified energy window can be exactly reproduced as a linear combination of the MLWFs by construction, and thus the accuracy of the original *ab-initio* electronic structure calculation is retained. (iii) the WF basis set can be made truly minimal and minimize the computational cost of the subsequent transport calculation. (iv) MLWFs contain information about chemical properties of the system, including bond types, coordination and electron lone pairs, and can thus be directly used as an analysis and visualization tool.

A Wannier function $\omega_{n\mathbf{R}}(\mathbf{r})$ labeled by the Bravais lattice vector $\mathbf{R}$, is usually defined via a unitary transformation of the Bloch functions $\psi_{n\mathbf{k}}(\mathbf{r})$ of the nth band:

$$\omega_{n\mathbf{R}}(\mathbf{r}) = \frac{V}{(2\pi)^3} \int_{BZ} \psi_{n\mathbf{k}}(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{R}} d^3k, \tag{2.10}$$

where V is the volume of the unit cell and the integral is performed over the entire Brillouin Zone (BZ). The MLWFs defined as above form an orthonormal basis set, and any two of them, for a given index n and different $\mathbf{R}$ and $\mathbf{R}'$, are translational images of each other. Also, MLWFs are linear combinations of Bloch functions, and span the original Hilbert space despite not representing stationary states.

The *ab-initio* eigenstates are defined modulo an arbitrary $\mathbf{R}$-dependent phase factor, so

there is no unique set of Wannier functions. Indeed, the electronic structure problem is invariant for the transformation $\psi_{n\mathbf{k}}(\mathbf{r}) \rightarrow e^{\phi_n(\mathbf{k})}\psi_{n\mathbf{k}}$. Besides this freedom in the choice of phases $\phi_n(\mathbf{k})$ for the Bloch functions, there is another gauge freedom from the fact that the many-body wavefunction is actually a Slater determinant[1]. In general, starting with a set of N Bloch functions with periodic $u_{n\mathbf{k}}$, we can construct infinite sets of MLWFs with different spatial characteristics:

$$\omega_{n\mathbf{R}}(\mathbf{r}) = \frac{V}{(2\pi)^3}\int_{BZ}\left[\sum_m U_{mn}^{(\mathbf{k})}\psi_{m\mathbf{k}}(\mathbf{r})\right]e^{-i\mathbf{k}\cdot\mathbf{R}}d^3k. \tag{2.11}$$

The unitary matrices $U^{(\mathbf{k})}$ also include the gauge freedom on phase factors as mentioned before [32].

A good choice of the set of MLWFs is the one with the narrowest spatial distribution, forming a most localized basis. Following the procedure proposed by Marzari and Vanderbilt [32], we search the particular unitary matrices $U_{mn}^{(\mathbf{k})}$ that can achieve this goal.

A quantitive measure of the spatial delocalization of MLWFs is given by a *Spread Operator* Ω, defined as the sum of the second moments of all the Wannier functions in a specific cell:

$$\Omega = \sum_n [\left\langle r^2 \right\rangle_n - \langle \mathbf{r} \rangle_n^2] \tag{2.12}$$

where the sum is over selected bands, and

$$\begin{aligned} \langle \mathbf{r} \rangle_n &= \langle \mathbf{0}n|\mathbf{r}|\mathbf{0}n\rangle\,, \\ \left\langle r^2 \right\rangle_n &= \langle \mathbf{0}n|r^2|\mathbf{0}n\rangle\,. \end{aligned} \tag{2.13}$$

Because the value of the spread Ω depends on the choice of unitary matrices $U^{(\mathbf{k})}$ we can

1. For a Slater determinant, a unitary transformation of orbitals will not change the manifold, and will not change the total energy of the system as well as the charge density.

evolve any arbitrary set of $U^{(\mathbf{k})}$ until we reach the stationarity condition:

$$\frac{\delta\Omega_{\mathbf{k}}}{\delta U^{(\mathbf{k})}} = 0 \tag{2.14}$$

We can always obtain the matrices $U^{(\mathbf{k}),ML}$ that transform the first-principles $\psi^{FP}_{n\mathbf{k}}(\mathbf{r})$ into the MLWFs according to Eq. (2.11). If we consider only $\mathbf{k}$-point mesh calculations, we can use finite differences in reciprocal space to evaluate the derivatives of Eq. (2.14). For this purpose we rewrite the expectation values $\langle\mathbf{r}\rangle$ and $\langle r^2\rangle$ as proposed by Blount [33]:

$$\langle\mathbf{0}n|\mathbf{r}|\mathbf{0}n\rangle = i\frac{1}{N}\sum_{\mathbf{k}} e^{+i\mathbf{k}\cdot\mathbf{R}}\,\langle u_{\mathbf{k}n}|\nabla_{\mathbf{k}}|u_{\mathbf{k}n}\rangle\,, \tag{2.15}$$

$$\langle\mathbf{0}n|r^2|\mathbf{0}n\rangle = \frac{1}{N}\sum_{\mathbf{k}} e^{+i\mathbf{k}\cdot\mathbf{R}}\,\langle u_{\mathbf{k}n}|\nabla^2_{\mathbf{k}}|u_{\mathbf{k}n}\rangle\,, \tag{2.16}$$

where $|u_{n\mathbf{k}}\rangle = e^{-i\mathbf{k}\cdot\mathbf{r}}\,|\psi_{n\mathbf{k}}\rangle$ is the periodic part of the Bloch function. Assuming that the BZ is sampled by a uniform $\mathbf{k}$-point mesh, and letting $\mathbf{b}$ be the vectors that connect a mesh point to its near neighbors, we can define the overlap matrix between Bloch orbitals as:

$$M^{(\mathbf{k},\mathbf{b})}_{mn} = \langle u_{m\mathbf{k}}|u_{n\mathbf{k}+\mathbf{b}}\rangle = \langle\psi_{m\mathbf{k}}|e^{-i\mathbf{b}\cdot\mathbf{r}}|\psi_{n\mathbf{k}+\mathbf{b}}\rangle\,. \tag{2.17}$$

Using the expression of the gradient in terms of finite differences and replacing $M^{(\mathbf{k},\mathbf{b})}_{mn}$ in Eqs. (2.15, 2.16), we obtain $\langle\mathbf{r}\rangle$ and $\langle r^2\rangle$ to be used in the localization procedure:

$$\langle\mathbf{r}\rangle_n = -\frac{1}{N}\sum_{\mathbf{k},\mathbf{b}} \omega_{\mathbf{b}}\,\mathbf{b}\,\mathrm{Im}\Big\{\ln M^{(\mathbf{k},\mathbf{b})}_{nn}\Big\} \tag{2.18}$$

$$\Big\langle r^2\Big\rangle_n = \frac{1}{N}\sum_{\mathbf{k},\mathbf{b}} \omega_{\mathbf{b}}\left[\left(1 - \left|M^{(\mathbf{k},\mathbf{b})}_{nn}\right|^2\right) + \left(\mathrm{Im}\Big\{\ln M^{(\mathbf{k},\mathbf{b})}_{nn}\Big\}\right)^2\right] \tag{2.19}$$

Here, $\omega_{\mathbf{b}}$ are the weights of the $\mathbf{b}$-vectors, and must satisfy the completeness condition $\sum_{\mathbf{b}} \omega_{\mathbf{b}}\mathbf{b}_\alpha\mathbf{b}_\beta = \delta_{\alpha\beta}$. Replacing the above expression into Eq. (2.12), we obtain the spread

operator as a function of the overlap matrix $M_{mn}^{(\mathbf{k},\mathbf{b})}$. In order to calculate the gradient in Eq. (2.14), we consider the first-order change in Ω stemming from an infinitesimal transformation $U_{mn}^{(\mathbf{k})} = \delta_{mn} + dW_{mn}^{(\mathbf{k})}$, where $\mathrm{d}W$ is an infinitesimal anti-unitary matrix ($\mathrm{d}W^\dagger = -\mathrm{d}W$). The gauge transformation rotates the wavefunctions according to Eq. (2.11) into $|u_{\mathbf{k}n}\rangle \rightarrow |u_{\mathbf{k}n}\rangle + \sum_m \mathrm{d}W_{mn}^{(\mathbf{b})} |u_{\mathbf{k}m}\rangle$. Following the description of ref. [32], we finally obtain the expression for the gradient of the spread functional:

$$G^{(\mathbf{k})} = \frac{\delta\Omega}{\delta W^{(\mathbf{k})}} = 4 \sum_{\mathbf{b}} \omega_{\mathbf{b}} \left(\frac{R^{(\mathbf{k},\mathbf{b})} - R^{(\mathbf{k},\mathbf{b})\dagger}}{2} - \frac{T^{(\mathbf{k},\mathbf{b})} + T^{(\mathbf{k},\mathbf{b})\dagger}}{2i} \right), \tag{2.20}$$

where

$$R_{mn}^{(\mathbf{k},\mathbf{b})} = M_{mn}^{(\mathbf{k},\mathbf{b})} M_{nn}^{(\mathbf{k},\mathbf{b})*}; \qquad T_{mn}^{(\mathbf{k},\mathbf{b})} = \frac{M_{mn}^{(\mathbf{k},\mathbf{b})}}{M_{nn}^{(\mathbf{k},\mathbf{b})}} \left[\mathrm{Im}\Big\{ \ln M_{nn}^{(\mathbf{k},\mathbf{b})} \Big\} + \mathbf{b} \cdot \langle \mathbf{r} \rangle_n \right]. \tag{2.21}$$

Note that the expression $G^{(\mathbf{k})}$ is a function of the overlap matrices $M^{(\mathbf{k},\mathbf{b})}$. Minimizing the spread functional Ω is achieved via conjugate gradient schemes or steepest descent. The procedure does not update wavefunctions, and instead only requires the overlap and unitary matrices. This step is the most computationally demanding with time complexity of $O(N^3)$ for each iteration of Wannier localization.

Wannier functions obtained using the above procedures are well-defined except for an overall phase factor. This feature can be used to check the convergence of the localization procedure. It is important to notice that only in the case of continuous BZ integrations can the truly isolated limit be recovered, because a Born-von Karman discretization of the Brillouin Zone can only make MLWFs periodic in real-space, with a superperiodicity determined by the BZ discretization. This is easily seen remembering that $\psi_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{r}}$, and $u_{n\mathbf{k}}(\mathbf{r})$ has the same periodicity as the direct lattice. Therefore the phase factors $e^{i\mathbf{k}\cdot\mathbf{r}}$ determine the superperiodicity of the $\psi_{n\mathbf{k}}$ themselves. If the $\psi_{n\mathbf{k}}$ have $\mathbf{k}$'s restricted to a uniform Monkhorst-Pack mesh, they will be periodic with a wavelength inversely propor-

tional to the spacing of the mesh. This periodicity is in turn passed onto the MLWFs. For N **k**-points along a certain direction of the BZ, the MLWFs will repeat along this direction every N cells. A sufficiently dense mesh of **k**-points guarantees that the adjacent replicas of a Wannier function are far enough so that they do not interact.

The method described above works appropriately in the case of isolated groups of bands. Isolated Bloch band refer to those that does not overlap with any other band anywhere in the BZ. On the contrary, a group of bands that are inter-connected by degeneracy but are isolated from all the other bands, is called a composite group.[32] In the case of studying quantum conductance in realistic systems we often need to calculate MLWFs for a subset of energy bands that are entangled or mixed with other bands. We are most interested in the states lying in the vicinity of the Fermi level. Since the unitary transformations $U^{(\mathbf{k})}$ mix energy bands at each **k**-point, any arbitrary choice of states inside a specified window will affect the localization properties of MLWFs, except when energy gaps effectively separate the manifold of interest from higher and lower bands. To solve this problem, Souza, Marzari, and Vanderbilt [34] introduced an additional disentanglement procedure that automatically extracts the best possible manifold for a given dimension from the states in a specified energy window. By using this method, we can deal with entangled or metallic cases of the MLWF formulation. The procedure relies on minimizing the subspace dispersion across the BZ, and effectively separate the bands of interest from the overall band structure.

In practice, we first select a desired number of bands in a predefined energy window. Then we determine the optimal subspace that can be extracted from that band structure; and finally proceed with a standard localization procedure within the selected subspace. The resulting orbitals with the small spreads have good localization properties, and allow for application of our formalism to arbitrary systems, regardless of the insulating or metallic nature of the bands. It should be noted that the MLWFs obtained in the later case are not the MLWFs of the occupied subspace with potentially poor localization properties, but are

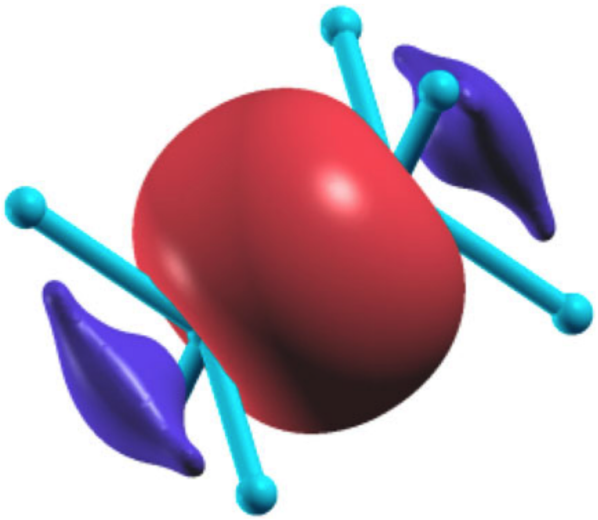


Figure 2.1: MLWFs constructed using the four valence bands of Si, which resemble the σ-bonded combinations of $sp3$ hybrid orbitals. This figure is excerpted from [24].

those of a well connected, continuous subspace that in general will have both occupied and unoccupied Bloch functions. We use the above localization procedure as implemented in the code WANNIER90 [35].

In order to calculate the transport properties according to the above prescriptions, we need to extract the matrix elements of the Hamiltonian computed on the localized MLWF basis. Assuming that a BZ sampling is fine enough to eliminate the interaction between the periodic images of MLWF, we can simply compute the tight-binding Hamiltonians $H_{ij}(\mathbf{R}) = \langle \omega_{i\mathbf{0}} | H | \omega_{i\mathbf{R}} \rangle$, from the unitary rotations $U^{(\mathbf{k})}$ obtained in the localization procedure. In the Bloch representation, we have $H_{mn}(\mathbf{k}) = \epsilon_{m\mathbf{k}} \delta_{m,n}$ by definition. When moving to the MLWF basis, we have:

$$H^{\mathrm{rot}}(\mathbf{k}) = U^{(\mathbf{k})\dagger} H(\mathbf{k}) U^{(\mathbf{k})} \tag{2.22}$$

Next, we Fourier transform $H^{\mathrm{rot}}(\mathbf{k})$ into the corresponding set of Bravais lattice vectors

$\{\mathbf{R}\}$:

$$H_{ij}(\mathbf{R}) = \frac{1}{N_k} \sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{R}} H_{ij}^{\mathrm{rot}}(\mathbf{k}). \tag{2.23}$$

By doing this, we obtained the tight-binding Hamiltonian in the real space suitable for our later transport calculations.

2.3 Quantum electron transport

In this section, we will give a brief introduction to the theory of transport in systems of mesoscopic to molecular scale. We will start by presenting the classification of transport due to the correlation between the size of the conductor and the characteristic transport lengths. Then, we will describe the Keldysh Non-Equilibrium Green's Function formalism. Then we present how to use Green's functions to calculate quantities of interest in experiments. For example, one can use the Landauer formula to calculate transport in mesoscopic systems.[36]

2.3.1 Transport regimes

The electron transport behavior at the mesoscopic level is usually classified by the interplay between system length L and its characteristic transport lengths. So in order to understand different transport regimes, we first need to define the relevant lengths: [36] An electron moving in a real crystal with impurities, lattice vibrations, and other electrons, experiences collisions that can alter its state. The time between these scattering processes is called the collision time τ_c. We define the time between two scattering processes that change the electron momentum as the momentum relaxation time τ_m. It is related to the collision time τ_c by the relation:

$$\frac{1}{\tau_m} \longrightarrow \frac{1}{\tau_m}\alpha_m,$$

where the factor α_m (lying between 0 and 1) indicates the efficiency by which an individual collision changes the electron momentum. If the electron is scattered only by a small angle,

resulting in small momentum loss and therefore small α_m, the momentum relaxation time is much longer than the collision time. The corresponding length to the momentum relaxation time is the mean free path l_m. The mean free path is the length that an electron travels before its initial momentum is destroyed. It is given by the relation

$$\ell_m = v_F \tau_m,$$

where v_F is the Fermi velocity.

Similar to τ_m, we can relate a phase relaxation time τ_φ to τ_c, with a factor α_φ describing the how a collision changes the energy and in turn the phase of the electron:

$$\frac{1}{\tau_\varphi} \longrightarrow \frac{1}{\tau_\varphi}\alpha_\varphi.$$

The phase relaxation length ℓ_φ represents the length an electron can travel with a well defined phase. Since changes in the phase information affect how the electron waves interfere, we will not expect interference effects for lengths larger than ℓ_φ. If the phase relaxation time is of the same order as the momentum relaxation time, i.e. $\tau_\varphi \sim \tau_m$, we have the following relation between τ_φ and ℓ_φ:

$$\ell_\varphi = v_F \tau_\varphi.$$

However, if the momentum relaxation time is much shorter than the phase relaxation time, $\tau_\varphi \gg \tau_m$, the trajectory of an electron over a phase relaxation time can be seen as sum of many short trajectories of length $\sim v_F \tau_m$. In this case, the relation between the phase relaxation length and the phase relaxation time becomes

$$\ell_\varphi^2 = D\tau_\varphi,$$

with the diffusion coefficient

$$D = v_F^2 \tau_m / 2.$$

The effects of collisions on transport lengths are strongly related to the kind of scattering and the scattering centers. For example, elastic scattering, caused by static impurities or interfaces etc., conserves electron energy and therefore its phase, but changes the momentum. We will ahve $\alpha_m \neq 0$ but $\alpha_\varphi = 0$. Electron-electron scattering redistributed the energy among the electrons but does not affect l_m. Indeed, any momentum lost by one electron is picked up by another, so there is no change in the net momentum, resulting in $\alpha_m = 0$ but $\alpha_\varphi \neq 0$. On the contrary, inelastic scattering arising from e.g. lattice vibrations as described by phonons, or impurities with an internal degree of freedom that allow them to change the state, lead to a change of both the energy and the momentum, $\alpha_m \neq 0$ but $\alpha_\varphi \neq 0$.

We can distinguish between several transport regimes, depending on the relation between system length L and the characteristic transport lengths.[37] We have diffusive transport if $L > \ell_m$, phase incoherent transport if $L > \ell_\varphi$, ballistic transport if $L < \ell_m$ and phase coherent transport if $L < \ell_\varphi$. Classical conductors are both in the diffusive and phase incoherent transport regime, $L \gg \ell_m, \ell_\varphi$. Usually one refers to a sample as ballistic conductor when the sample shows both ballistic and phase coherent transport behavior, $L < \ell_m, \ell_\varphi$. Usually, for most systems studied in this thesis, the electron transport length is small, suggesting the possibility of ballistic transport.[38, 39]

2.3.2 Keldysh Non-Equilibrium Green's Function formalism

In the following, we present a general formalism for systems under non-equilibrium conditions at finite temperature. For this purpose, first the ensemble average of an operator under non-equilibrium is defined. Then the contour-ordered Keldysh non-equilibrium Green's function (NEGF) formalism is introduced and the kinetics equations for the Keldysh Green's functions are presented. Finally, the relationship between the introduced formalism and quantities of

interest in experiments are described.

Non-equilibrium Ensemble Average

We employ the standard device for obtaining a non-equilibrium state. At time t_0, prior to which the system is assumed to be in thermodynamic equilibrium with a reservoir, the system is exposed to a disturbance $\hat{H}^{ext}$ to the Hamiltonian. Then total Hamiltonian can be written as:

$$\hat{H}(t) = \hat{H}_0 + \hat{H}^{int} + \hat{H}^{ext} = \hat{H} + \hat{H}^{ext} \tag{2.24}$$

where $\hat{H}^{ext} = 0$ for $t < t_0$. One is not limited to using the statistical equilibrium state at times prior to $t0$ as the initial condition. Non-equilibrium statistical mechanics aims to calculate expectation values $\hat{O}_H(t)$ of physical observables for times $t > t_0$. Given the density operator $\hat{\rho}$, the average of any operator $\hat{O}$ is then defined as

$$< \hat{O}_H(t) >= Tr[\hat{\rho}\hat{O}_H(t)], \tag{2.25}$$

where $\hat{O}_H(t)$ is an operator in the Heisenberg picture. The Green's function is defined as

$$G(\mathbf{r}, t, \mathbf{r'}, t') = -\frac{i}{\hbar} < T_t\{\hat{\psi}_H(\mathbf{r}, t)\hat{\psi}^{\dagger}_H(\mathbf{r'}, t')\} >, \tag{2.26}$$

where $\hat{\psi}_H$ is the field operator in the Heisenberg picture evolving with the Hamiltonian $\hat{H}$ as defined in Eq. (2.24), and the bracket $< ... >$ is the statistical average with the density operator defined in Eq. (2.25).

Contour-Ordered Green's Function

In order to describe non-equilibrium phenomena using Green?s functions, we can work with contour-ordering operators instead of the time- ordering operators. Under equilibrium con-

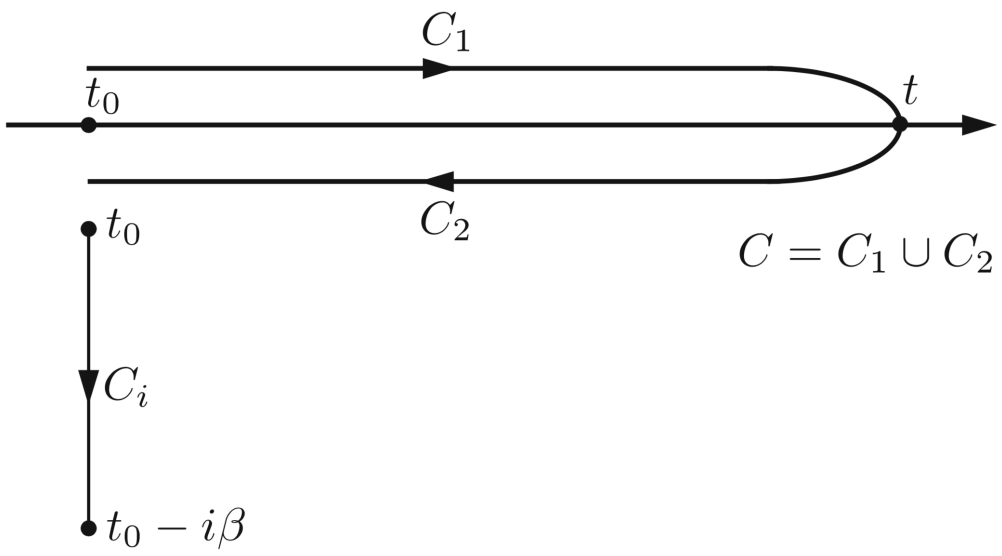


Figure 2.2: The Keldysh contour C^* runs from t_0 to t back to t_0 and to $t_0 - i\beta$.

dition, the contour-ordered method gives the same results as the time-ordered method. Non-equilibrium theory is based upon this contour technique. We show the Keldysh contour in Fig. 2.2.

We define four different contour-ordered Green's functions depending on where the start and end time reside:

$$G(t,t') = \begin{cases} G^{>}(t,t') = -i\hbar^{-1} < \hat{\psi}_H(t)\hat{\psi}_H^{\dagger}(t') >, t \in C_2, t' \in C_1 \\ G^{<}(t,t') = +i\hbar^{-1} < \hat{\psi}_H^{\dagger}(t)\hat{\psi}_H(t') >, t \in C_1, t' \in C_2 \\ G_T(t,t') = -i\hbar^{-1} < T_t\{\hat{\psi}_H(t)\hat{\psi}_H^{\dagger}(t')\} >, t, t' \in C_1 \\ G_{\tilde{T}}(t,t') = -i\hbar^{-1} < T_{\tilde{t}}\{\hat{\psi}_H(t)\hat{\psi}_H^{\dagger}(t')\} >, t, t' \in C_2, \end{cases} \tag{2.27}$$

where $G^{>}$, $G^{<}$, G_T, $G_{\tilde{T}}$ are *greater*, *lesser*, *time-ordered*, and *anti-time-ordered* Green's function respectively. Because, as one can prove, $G_T + G_{\tilde{T}} = G^{>} + G^{<}$, there are only three linearly independent Green's functions. For convenience, one usually define two more Green's

functions the retarded and advance Green's functions, denoted by G^R and G^A respectively. And we also have $G^R - G^A = G^> - G^<$.

For electrons or fermions in general, the above real-time Green's functions can be Fourier transformed to energy space, resulting in:

$$\begin{aligned} G^<(\mathbf{k}, E) &= +2\pi i f(E_\mathbf{k})\delta(E - E_\mathbf{k}) \\ G^>(\mathbf{k}, E) &= +2\pi i[1 - f(E_\mathbf{k})]\delta(E - E_\mathbf{k}) \\ G^R(\mathbf{k}, E) &= \frac{1}{E - E_\mathbf{k} + i\eta} \\ G^A(\mathbf{k}, E) &= \frac{1}{E - E_\mathbf{k} - i\eta}, \end{aligned} \tag{2.28}$$

where η is a small positive number 0^+.

Transport calculations

One advantage of using Green's functions is the relative ease with which they can be calculated, compared to a direct numerical solution of the Schrödinger equation. In particular, an efficient decimation method described in Appendix A is available to effectively represent the effect of the leads [40] on the device region. Also, the recursive Green's function (RGF) method [41], as detailed in Appendix B, is a useful tool to compute the Green's functions for the evaluating the transmission coefficients in the Landauer's formula. The method is very reliable, computationally efficient, and can be applied to arbitrary geometries. Next, we will introduce the calculation of the transmission function using the scattering matrix S as well as the Green's functions.

Since we are dealing with coherent transport, we can characterize the conductor at each energy by an $\mathcal{S}$-matrix relating the outgoing wave amplitudes to the incoming wave amplitudes at the different leads:

$$\psi^{OUT} = S\psi^{IN}, \tag{2.29}$$

where the dimension of the $\mathcal{S}$-matrix is $M_T(E)$.

The quantum mechanical transmission is defined as the ratio between the amplitude of the outgoing wave and the one of the incoming wave. Therefore one can express the transmission between the mode m in the lead p and the mode n in the lead q by the elements of the $\mathcal{S}$-matrix:

$$T_{mn} = |S_{mn}|^2 . \tag{2.30}$$

As for the transmission between the lead p and q, we obtain:

$$T_{pq} = \sum_{m \in p} \sum_{n \in q} T_{mn}. \tag{2.31}$$

It is worth mentioning that the current carried by a scattered wave is proportional to the square of the wave function multiplied by the velocity. The $\mathcal{S}$-matrix can be defined in terms of the current amplitude, which can be computed by the wave amplitude multiplied by the square root of the velocity. We will see later that it is convenient to normalize the current flux and thus define a normalized scattering matrix $\mathcal{S}^N$:

$$\mathcal{S}^N_{nm} = \sqrt{\frac{v_m}{v_n}} \mathcal{S}_{nm}. \tag{2.32}$$

To understand the physical meaning of Green's functions as the response at any point due to an excitation at any other, we consider the response function $\Psi(x)$ related to an excitation $f(x)$ by a differential operator $\mathcal{D}$:

$$\mathcal{D}\Psi(x) = f(x). \tag{2.33}$$

The general solution for this inhomogeneous differential equation is

$$\mathcal{D}\Psi(x) = \Psi_0(x) + \int dx' G(x, x') f(x'), \tag{2.34}$$

with the Green's function defined by

$$\mathcal{D}G(x,x') = \delta(x-x') \Longrightarrow G = \mathcal{D}^{-1}. \tag{2.35}$$

To obtain solutions for the Green's function, we need to specify boundary conditions. We can consider the simple case of an infinite one-dimensional wire with a constant potential energy U_0. We can write Eq.(2.33) in terms of the Hamiltonian operator as:

$$\left[E - \mathcal{H}\right]\Psi(x) = f(x), \text{ with } \mathcal{H} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + U_0, \tag{2.36}$$

where Ψ is the wavefunction, and f an excitation. Hence we get for the Green's function

$$G = \left[E - U_0 + \frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\right]^{-1}, \tag{2.37}$$

i.e.

$$\left(E - U_0 + \frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\right)G(x,x') = \delta(x-x'), \tag{2.38}$$

which is similar to the Schrödinger equation except for $\delta(x-x')$

$$\left(E - U_0 + \frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\right)\Psi(x) = 0. \tag{2.39}$$

Considering this similarity, one can think of the Green's function $G(x,x')$ as the wavefunction at x produced by a unity excitation at x'. When an excitation occurs at x', one expects two waves with the amplitudes A^+ andA^-, moving outward from x', as described in the left plot of Fig. 2.3.

We can write

$$G(x,x') = \begin{cases} A^+ e^{ik(x-x')} & \text{for } x > x' \\ A^- e^{-ik(x-x')} & \text{for } x > x' \end{cases} \tag{2.40}$$

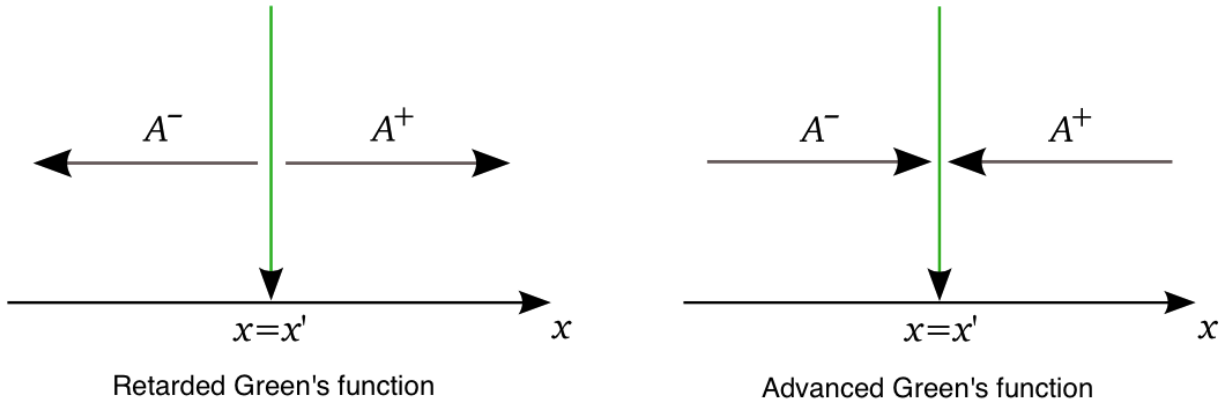


Figure 2.3: Retarded and advanced Green's functions for an infinite one dimensional wire

with $k = \sqrt{2m(E-U_0)}/\hbar$. For any A^+, A^-, this solution satisfies Eq.(2.38), but only for $x \neq x'$. Therefore we will define A^+, A^- using the boundary conditions for $x = x'$, that are:

$$[G(x,x')]_{x=x'+} = [G(x,x')]_{x=x'-} \tag{2.41}$$

$$\left[\frac{\partial G(x,x')}{\partial x}\right]_{x=x'+} - \left[\frac{\partial G(x,x')}{\partial x}\right]_{x=x'-} = \frac{2m}{\hbar}. \tag{2.42}$$

We obtain for the wave amplitudes

$$A^+ = A^- = -\frac{i}{\hbar v}, \tag{2.43}$$

where $v = \frac{\hbar k}{m}$. Hence one of the solutions which satisfies Eq.(2.38) is the retarded Green's function:

$$G^R(x,x') = -\frac{i}{\hbar v} e^{ik|x-x'|}, \tag{2.44}$$

corresponding to outgoing waves that originate at the point of excitation, as shown in the left plot of Fig. 2.3. The other solution which satisfies Eq.(2.38) is the advanced Green's

function:

$$G^R(x,x') = -\frac{i}{\hbar v} e^{ik|x-x'|}, \tag{2.45}$$

corresponding to incoming waves that disappear at the point of excitation as shown in the right plot of Fig. 2.3. A way to incorporate the boundary conditions directly into Eq.(2.38) is the addition of an infinitesimal imaginary part to the energy:

$$\left(E - U_0 + \frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2 + i\eta}\right) G^R(x,x') = \delta(x-x'). \tag{2.46}$$

The only suitable solution to the above equation is the retarded Green's function, since the solution must be bounded and the imaginary part can make the advanced Green's function divergent. Similarly, the advanced Green's function is the only adequate solution to the equation

$$\left(E - U_0 + \frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2 - i\eta}\right) G^A(x,x') = \delta(x-x'). \tag{2.47}$$

In general, the retarded and the advanced Green's functions are defined as

$$G^R(x,x') = [E - H + i\eta]^{-1}, \tag{2.48}$$

$$G^A(x,x') = [E - H - i\eta]^{-1}, \tag{2.49}$$

where $\eta \to 0^+$. It is important to mentioned that the advanced Green's function is the Hermitian conjugate of the retarded Green's function, and they can be related by:

$$G^A = G^{R\dagger}. \tag{2.50}$$

From now on, if it is not explicitly stated, we will generally refer to the retarded Green's function as the "Green's function".

Next we need to find the relation between the Green's functions and the $\mathcal{S}$-matrix. For this, we consider a conductor connected to a set of single modeled leads. The points $x_p = 0$ and $x_q = 0$ represent the interfaces between the lead p and the conductor and between the lead q and the conductor respectively. G_{qp} is the Green's function of the region between $x_p = 0$ and $x_q = 0$. A unit excitation at $x_p = 0$ will give rise to two waves, one with the amplitude A^- propagating away from the conductor and one with the amplitude A^+ propagating through the conductor and being partially transmitted to the every one of the leads.

We can write the Green's function in terms of the wave amplitudes and of the normalized scattering matrix $\mathcal{S}^N$:

$$G_{qp} = \delta_{qp} A_p^- + \mathcal{S}_{qp}^N A_p^+. \tag{2.51}$$

Using Eq.(2.32) and Eq.(2.43), we obtain the desired equation for the $\mathcal{S}$-matrix by using the Green's function:

$$\mathcal{S}_{qp} = -\delta_{qp} + i\hbar\sqrt{v_q v_p} G_{qp}, \tag{2.52}$$

where $v_q = \hbar k_q / m$.

We have all relations we need to calculate the transmission and therefore current flow now, but still there is one more problem to solve, i.e. we must invert $[EH + i\eta]$ numerically for the Green's function given as

$$G(x, x') = [E - H + i\eta]^{-1}. \tag{2.53}$$

To solve this problem, we have to partition the Green's function of the system.

Within the Landauer's formula, the system is composed of a central device connected to

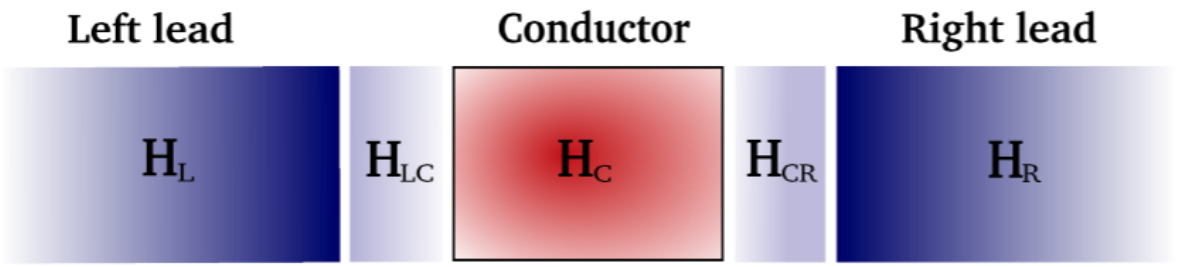


Figure 2.4: Schematic setup for the system Lead-Conductor-Lead

leads as shown in Fig. 2.4, and its Hamiltonian can therefore be divided as:

$$H = \begin{pmatrix} H_L & H_{LC} & 0 \\ H_{LC}^{\dagger} & H_C & H_{CR} \\ 0 & H_{CR}^{\dagger} & H_R \end{pmatrix} \tag{2.54}$$

where $H_{L,R}$and H_C are the Hamiltonians of the leads and of the conductor, respectively. H_{LC} is the coupling matrix between the left lead and the conductor, H_{RC} the coupling matrix between the right lead and the conductor. A direct inversion $[E - H + i\eta]^{-1}$ to obtain the Green's function is computational very expensive. We thus go around this by partitioning the Green's function Eq.(2.48) into submatrices as follows:

$$\begin{pmatrix} G_L & G_{LC} & G_{LR} \\ G_{CL} & H_C & G_{CR} \\ G_{RL} & G_{RC} & G_R \end{pmatrix} = \begin{pmatrix} \epsilon - H_L & H_{LC} & 0 \\ H_{LC}^{\dagger} & \epsilon - H_C & H_{CR} \\ 0 & H_{CR}^{\dagger} & \epsilon - H_R \end{pmatrix}^{-1} \tag{2.55}$$

with $\epsilon = (E + i\eta)\mathbb{1}$, and the Green's functions corresponding to the isolated leads and the

conductor:

$$g_L = (\epsilon - H_L)^{-1} \text{ for the left lead} \tag{2.56}$$

$$g_R = (\epsilon - H_R)^{-1} \text{ for the right lead} \tag{2.57}$$

$$g_C = (\epsilon - H_C)^{-1} \text{ for the conductor} \tag{2.58}$$

We are interested in the Green's function of the conductor with attached leads $G_C = (\epsilon H_{\text{eff}})^{-1}$. H_{eff} is the Hamiltonian we obtain if we decimate the leads in $(\epsilon - H)$, and include them in the conductor. First we decimate the right lead, using mathematical tools from linear algebra [36, 37]:

$$\left(\begin{array}{c|c|c} \epsilon - H_L & H_{LC} & 0 \\ \hline H_{LC}^\dagger & \epsilon - H_C & H_{CR} \\ \hline 0 & H_{CR}^\dagger & \epsilon - H_R \end{array}\right) \rightarrow \left(\begin{array}{c|c} \epsilon - H_L & H_{LC} \\ \hline H_{LC}^\dagger & \epsilon - H_C - H_{CR} \underbrace{(\epsilon - H_R)^{-1}}_{g_R} H_{CR}^\dagger \end{array}\right),$$

then the left:

$$\left(\begin{array}{c|c} \epsilon - H_L & H_{LC} \\ \hline H_{LC}^\dagger & \epsilon - H_C - H_{CR} g_R H_{CR}^\dagger \end{array}\right) \rightarrow \epsilon - H_C - H_{CR} g_R H_{CR}^\dagger - H_{LC}^\dagger \underbrace{(\epsilon - H_L)^{-1}}_{g_L} H_{LC}.$$

And finally we obtain for the effective Green's function of the central region

$$\begin{aligned} G_C = (\epsilon - H_{\text{eff}})^{-1} &= (\epsilon - H_C - H_{CR} g_R H_{CR}^\dagger - H_{LC}^\dagger g_L H_{LC})^{-1} \\ &= (\epsilon - H_C - \Sigma_R \Sigma_L)^{-1}, \end{aligned} \tag{2.59}$$

where $\Sigma_R = H_{CR} g_R H_{CR}^\dagger$ and $\Sigma_L = H_{LC} g_L H_{LC}^\dagger$ are the self-energy terms due to the leads. They can be seen as effective Hamiltonians arising from the interaction of the conductor with the leads. They represent exactly the effect of the leads on the conductor. In the Eq.(2.59),

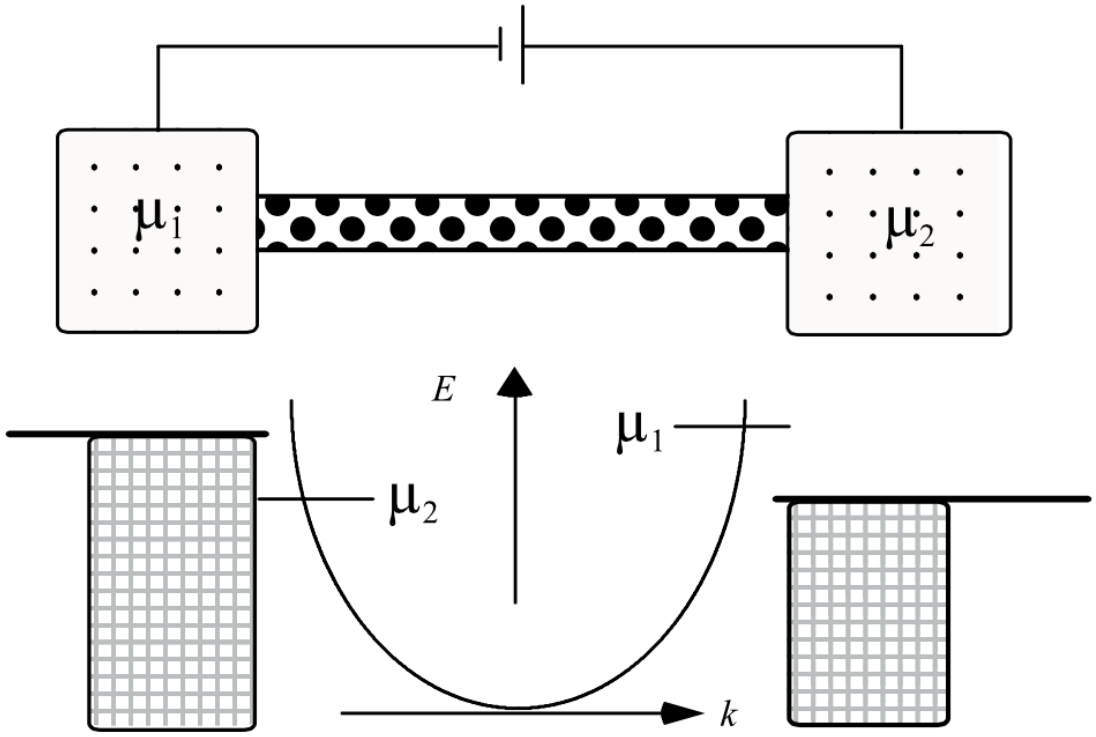


Figure 2.5: Schematic description of a ballistic device with a ballistic conductor between two bulk contacts settings the external chemical potential (taken from Ref. [42]).

all matrices are finite except for the Green's functions of the isolated leads g_L and g_R. In Appendix A, we show that one can calculate the self-energy terms of the leads using the surface Green's function instead of directly using g_L and g_R.

Calculation of transport properties

Having the mathematical framework of Keldysh formalism, one still needs a method to calculate the transport properties numerically. We consider the system of a ballistic conductor between two reflectionless bulk contacts (see Fig. 2.5). The contacts have the chemical potentials μ_1 and μ_2, with $\mu_1 - \mu_2 \to 0^+$.

The decay width matrices $\Gamma_L = i[\Sigma_L^r \Sigma_L^a]$ and $\Gamma_R = i[\Sigma_R^r \Sigma_R^a]$ represent the injection rates,

namely the strength of the coupling of the leads to the central region, where the advanced self-energy $\Sigma^A_{L(R)}$ is the Hermitian conjugate of the retarded self-energy $\Sigma^R_{L(R)}$. We obtain also a relation for the density of states (DOS):

$$N(E) = -\frac{1}{\pi}\,\mathrm{Im}[\mathrm{Tr}\, G^R_C(E)]. \tag{2.60}$$

In the above, we have assumed a truly one-dimensional chain of principal layers, which is physical only for systems having a definite quasi-one-dimensional character, e.g. nanotubes or quantum wires. The extension to higher dimensions is straightforward using Bloch functions in the directions perpendicular to the transport axis. The introduction of the concept of principal layer suggests that the system can be described by an infinite set of $k_\perp$ along the direction perpendicular to the layer, while $k_\parallel$ are still good quantum numbers for the problem. The above procedure effectively decouples the 3D system to a set of non-interacting 1D linear-chains, one for each $k_\parallel$ [43, 44]. We can then sum over the k-point samplings to evaluate the quantum conductance:

$$T(\omega) = \sum_{k_\parallel} \omega_{k_\parallel} T_{k_\parallel}(\omega), \tag{2.61}$$

where $\omega_{k_\parallel}$ are the weights of the different $k_\parallel$ in the irreducible BZ.

The Landauer-Buttiker formalism [45] has proved to be a very useful tool in describing the transport in mesoscopic systems. Once we have the Green's function, we can use it to obtain the elements of the $\mathcal{S}$-matrix with Eq.(2.52), from which the transmission can be calculated with Eq.(2.31). We can obtain the transmission by using the Green's function and the coupling of the conductor to the leads [36, 46]:

$$T(E) = \mathrm{Tr}[\Gamma_L G^R_C \Gamma_R G^A_C]. \tag{2.62}$$

Classically the conductance is given by

$$G = \frac{\sigma W}{L}, \tag{2.63}$$

where the conductivity σ is a material constant, W and L the width and the length of the conductor respectively. We expect the conductance to grow indefinitely by decreasing the length of the conductor, but experiments show an upper limit G_C of the conductance for conductor lengths much smaller than the mean free path, $L \ll \ell_m$. This is surprising because a ballistic conductor should have zero resistance. As already mentioned, we assume contacts to be reflectionless, which means that electrons can go in and out the conductor region without any scattering. Furthermore, we assume leads to have infinite length. In the contacts, the current is carried by infinitely many transverse modes, while in the conductor only a few independent conducting modes are available, the number of modes in the energy window $|\mu_1 - \mu_2|$ being energy independent, $M(E) = M$, since $\mu_1 - \mu_2 \to 0^+$.[36] This requires a redistribution of the current at the interface, which leads to the observed resistance $(G_C)^{-1}$. In order to calculate G_C, we simplify the problem to be a conductor with only one conducting mode. The number of electronic states per mode in the energy window $|\mu_1 - \mu_2|$ is

$$N_S = \frac{dn}{dE}(\mu_1 - \mu_2), \tag{2.64}$$

where $\frac{dn}{dE}$is the density of states. The current carried by a single electronic state is

$$I_S = e\frac{1}{L}v_k, \tag{2.65}$$

with the group velocity $v_k = \frac{1}{\hbar}\frac{\partial E}{\partial k}$. The current carried by a mode can be calculated by multiplying the current in one state by the number of states:

$$I = e\frac{v_k}{L}\left(\frac{dn}{dE}\right)(\mu_1 - \mu_2)\,. \tag{2.66}$$

Writing the density of states in terms of group velocity:

$$\frac{dn}{dE} = \frac{\partial n}{\partial k}\frac{\partial k}{\partial E} = \frac{L}{2\pi}\frac{1}{\hbar v_k}, \tag{2.67}$$

the current becomes

$$I_M = \frac{e}{h}\left(\mu_1 \mu_2\right). \tag{2.68}$$

Since the number of modes is independent of energy, the total current for a conductor with more than one conducting mode is obtained by multiplying the number of modes M by the current per mode:

$$I = MI_M = \frac{e}{h}M\left(\mu_1 \mu_2\right). \tag{2.69}$$

Considering the voltage bias through the chemical potential difference

$$e\Delta V = \Delta\mu, \tag{2.70}$$

we can express the conductance of the system as

$$G_C = \frac{2e^2}{h}M = G_0 M \approx M\frac{1}{12.9}mS, \tag{2.71}$$

where spin degeneracy is included by the factor 2. This suggests that the conductance through a ballistic conductor between two reflectionless contacts is quantized in units of G_0, regardless of the conductor. G_0 is called conductance quantum. The resistance of a ballistic conductor is thus given by

$$(G_C)^{-1} = \frac{(G_0)^{-1}}{M} \approx \frac{12.9k\Omega}{M} = \frac{R_K}{2M}, \tag{2.72}$$

where R_K is called the von Klitzing constant.

Finally, let's consider the case of a conductor with M modes. We define T_{ij} as the

probability of an electron of being transmitted from the ith channel of the left lead in the jth channel in the right lead. We can calculate T_{ij} using the method introduced before. We can calculate the conductance associated to the ith channel

$$G_i = \frac{2e^2}{h} \sum_j T_{ij}, \tag{2.73}$$

with the sum over all final states. The total conductance is then

$$G = \frac{2e^2}{h} \sum_{ij} T_{ij}, \tag{2.74}$$

We can write the Landauer formula also in terms of the average probability T that an electron injected from the left lead will transfer to the right lead:

$$G = \frac{2e^2}{h} M\bar{T}. \tag{2.75}$$

Then we can use the Landauer's formula to give a relationship between the chemical potential of the reservoirs and the currents flowing through the leads:

$$I_p = \frac{-2e^2}{h} \sum_q \int dE T_{pq}(E) \left[f_p(E) - f_q(E) \right], \tag{2.76}$$

where p, q label the different leads, $f_p(E)$ is the Fermi-Dirac distribution for reservoir p, and T_{pq} are the transmission coefficients for electrons to go from lead q to lead p.

2.4 Transport simulation pipeline

In this section, we summarize our simulation pipeline based on the methods discussed above. Figure 2.6 depicts the block diagram of our simulation pipeline. In order to accurately describe the electronic structures of considered materials, we first use density-functional-theory

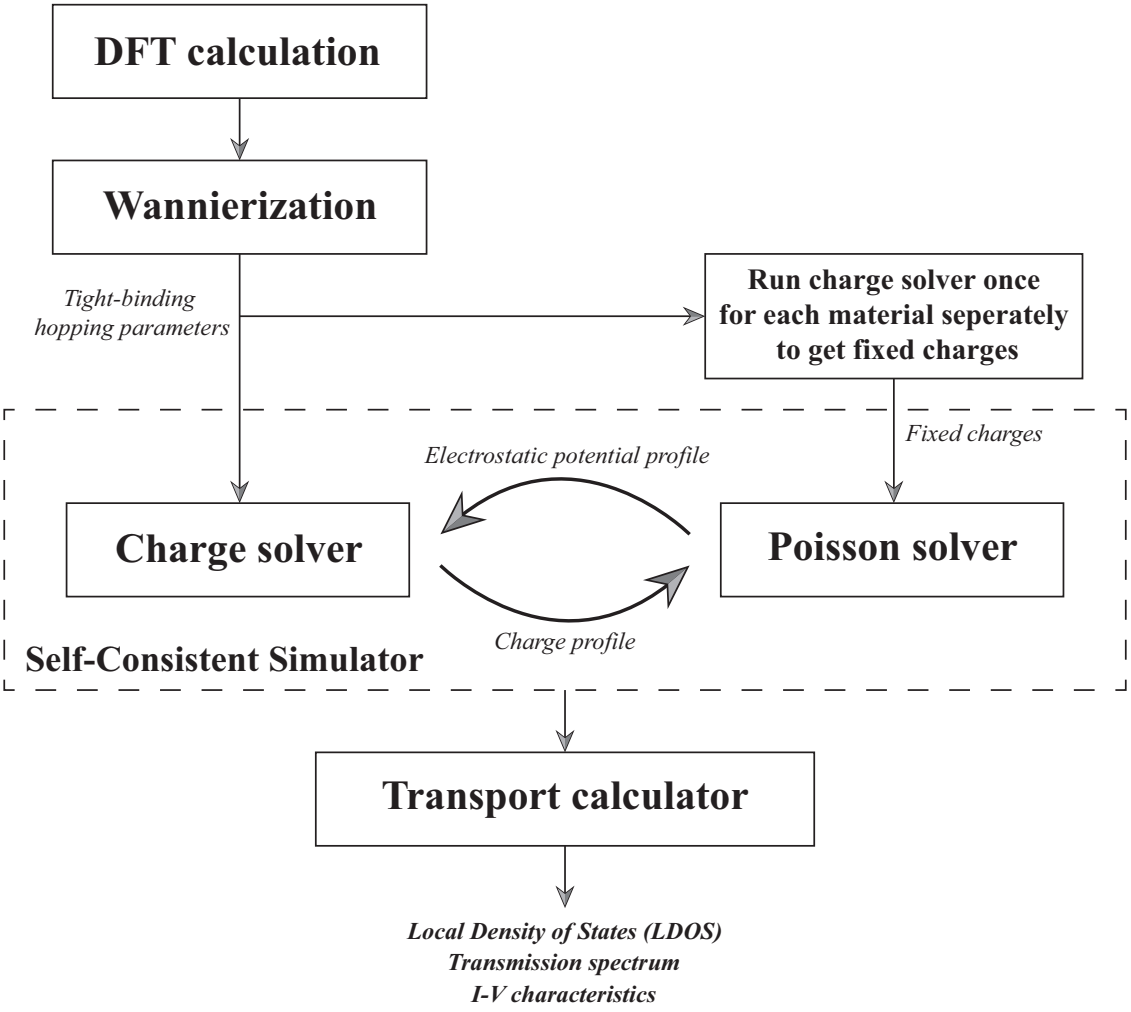


Figure 2.6: Schematic plot of our simulation pipeline. **Bold** indicates the main procedures. *Italic* shows the flowing data between different solvers. And **bold italic** at the bottom represents the output.

(DFT) framework QUANTUM ESPRESSO [47] to calculate their band structures. We then extract tight-binding hopping parameters by the Wannier technique as implemented in the code `wannier90` [48]. Those parameters allow us to self-consistently solve the electrostatics of the edge contact system using our custom-built open-source software `swan` [17] based on the Keldysh Non-equilibrium Green's Function formalism. The Poisson solver compute the electrostatic potential, which are part of the onsite energies of the system Hamiltonian. Thus the Poisson solver is coupled with the charge solver. This turns it into a non-linear equation and requires the usage of Newton-Raphson iteration method. The details are described in Appendix C. To use our software, we first run the charge solver once separately for each material in order to obtain their fixed charges. Next, using the obtained tight-binding parameters and fixed charges as inputs, we run self-consistent simulations for each bias voltage. After the simulation achieves convergence, we obtain the charge and electrostatic potential profile. Finally, we calculate the local density of states and transmission spectrum using the Landauer-Buttiker formalism, and plot the tunneling currents with biases to generate the I-V characteristic for our simulated edge contact device.

Our computational pipeline can predict the properties of realistic nano-devices based on first-principles. It allows the comparison of the performance limits of transistors based on difference novel materials for both the channel and lead, when experimental data is not available or when it is hard to separate the effects related to fabrication process and the intrinsic features of the material. Our simulation approach can give insight into the underlying microscopic physics and help design experiments. In the following chapters, we first validate this custom-built computational pipeline on ideal systems and then apply it to realistic 2D nanodevices.

CHAPTER 3

QUANTUM TRANSPORT IN VERTICALLY STACKING GRAPHENE LAYERS

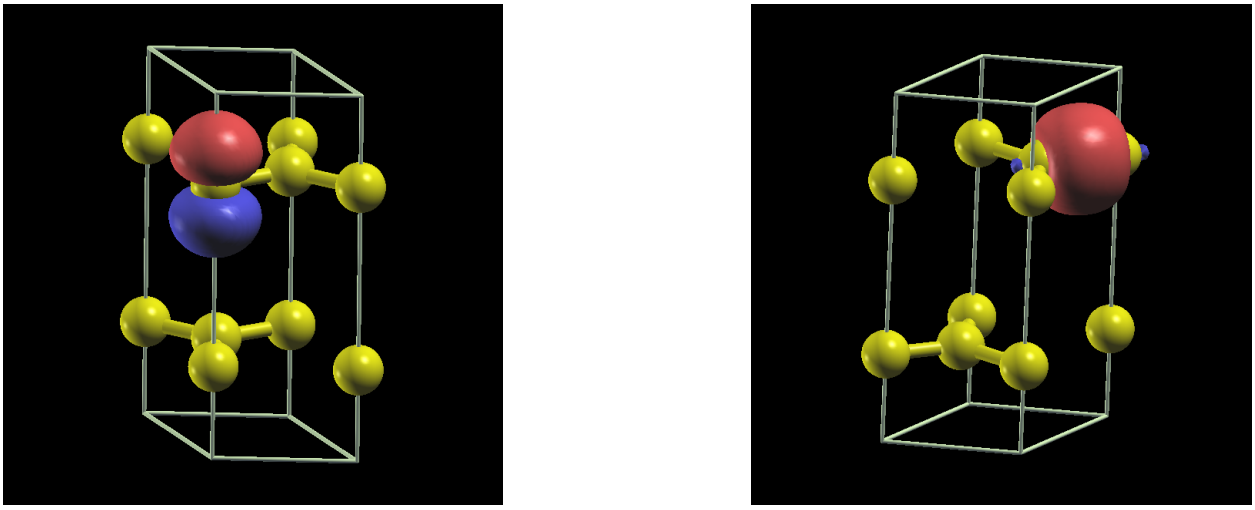
In order to validate our custom-built simulation pipeline, we apply it to the system of vertical AB-stacking of graphene layers.

3.1 Simulation setup

Consider a system containing replica stacking AB graphene layers. We try to calculate the electrons transport along the addition direction, and assume periodicity along the two perpendicular directions. In this case, we have simplified the problem by using the same unit cells for both the conductor and the lead. Thus the single layer H_{00} and the coupling H_{01} matrices are the only necessary inputs. The aim of this calculation is to examine the accuracy of our method on a simple system compare it with the one obtained by a full-DFT calculation, as implemented in the PWCOND code within the QUANTUM ESPRESSO software[49].

3.2 Results and discussions

Following the procedure described in the last chapter, we first try to construct the ML-WFs for the considered system. We perform the initial band-structure calculations using the PWSCF code [49]. A kinetic-energy cutoff of 30 Ry is used for the plane-wave expansion of the valence wavefunctions. The core-valence interaction is described by means of norm-conserving pseudopotentials in separable Kleinman-Bylander form [50]. We obtain the self-consistent ground state using a $16 \times 16 \times 16$ Monkhorst-Pack mesh of k-points [51] and a fictitious Fermi smearing [32] of 0.02 Ry for the Brillouin-zone integration. Then, we freeze the self-consistent potential and perform a non-self-consistent calculation on a



(a) p_z-type MLWF **(b)** σ-type MLWF

Figure 3.1: Isosurface contours of MLWF in graphite (red for positive value and blue for negative).

uniform $6 \times 6 \times 6$ grid of k-points. At each k-point we calculate the first 20 bands. The required overlap matrices and projections are calculated using the post-processing routine PW2WANNIER90, supplied with the PWSCF distribution. Projections onto atom-centered sp^2 and p_z functions are used to construct the initial guess, and WANNIER90 is used to obtain the MLWF. The gauge-dependent and gauge-independent spreads converge to machine precision in 300 and 70 steps, respectively. The resulting MLWF are a set of six symmetry-related bond-centerd σ-orbitals and four atom-centered p_z-orbitals, shown in Fig. 3.1 (as plotted with the XCRYSDEN package [52]).

In Fig. 3.2 we show the band structure of graphite obtained using Wannier interpolation and compare it to the band structure obtained from a full first-principles calculation. Within the inner energy window the two agree well. We further obtained the MLWFs using only p_z projections, and the reproduced bands also agree well with the DFT ones.

After testing the correctness of the set of tight-binding parameters extracted from the electronic structure, we compared our transport calculations with a full DFT calculation as implemented in PWCOND code as part of the QUANTUM ESPRESSO package. From

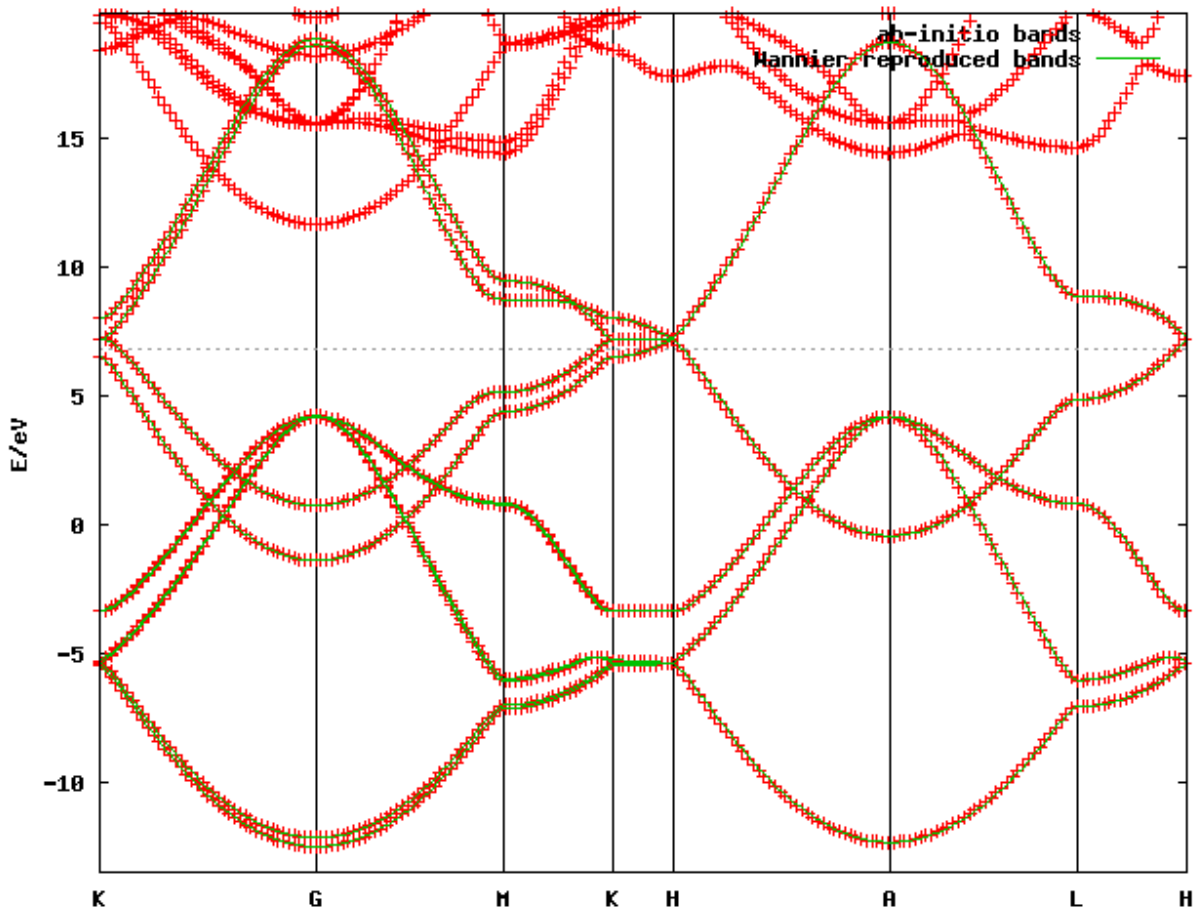


Figure 3.2: Band structure of graphite. Dotted lines: original band structure from a conventional first-principles calculation. Solid lines: Wannier-interpolated band structure. Fermi level is $\sim 6.8eV$.

Fig. 3.3, we can see that the two methods agree very well and the difference is hardly noticeable. However, our code takes three orders of magnitude less time and has better scalability because our method do not have limitation on the number of MPI ranks from the number of sampling points or plane waves used. This is especially favorable for scaling to realistic large systems to compare with experiments.

When solving the integrals over the transverse BZ, we have verified that 256×256 k values can produce converged and very accurate results. Transport computation can be still too computationally demanding, so that integration over the transverse BZ has been parallelized by means of MPI subroutines. Fig. 3.4 shows the quantum conductance and DOS near the Fermi level (shifted to 0).

3.3 Conclusions

We successfully applied our transport simulation to the simple system of stacking graphene layers. We reproduced the bandstructure by using MLWFs. The interaction matrix extracted served for the transport calculations. Our results on transport, either using regular or reduced orbitals, agree well with full DFT calculation as implemented in the PWCOND code. Having tested our methods, we are ready to apply it to larger and more realistic systems.

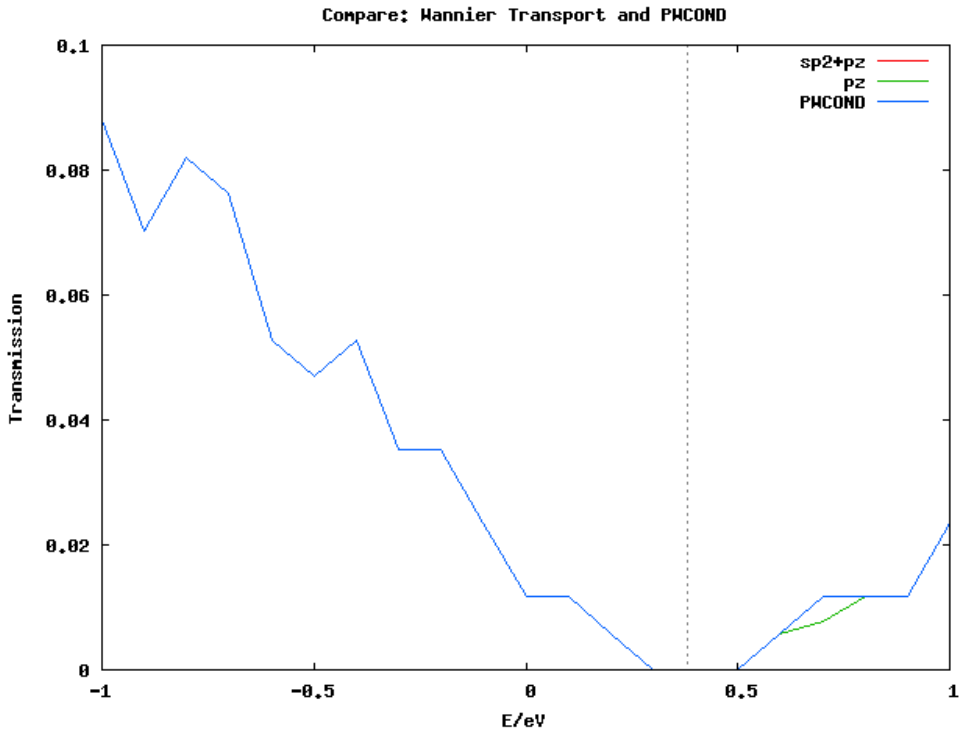


Figure 3.3: Comparison between our results using $sp^2 + p_z$ orbitals and only p_z orbitals with results of a full DFT calculation as implemented in PWCOND code as part of the QUANTUM ESPRESSO package. The differences between using full-DFT and our method, whether using sp^2 and p_z, or only p_z, are largely unnoticeable

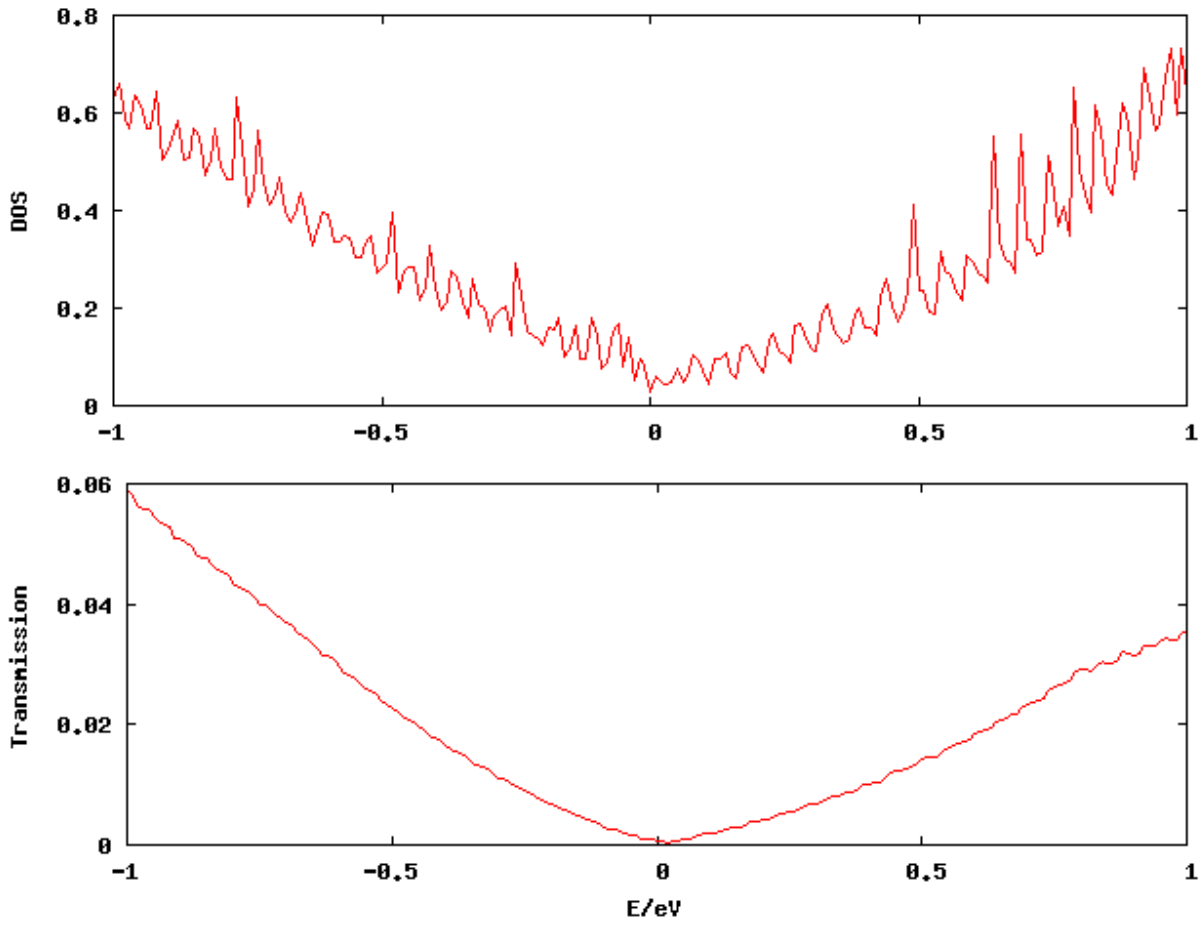


Figure 3.4: Transmission spectrum and DOS of vertical AB-stacking graphene averaged over the 256×256 k-points on the transverse BZ

CHAPTER 4

QUANTUM TRANSPORT IN TELESCOPIC DOUBLE WALL CARBON NANO-TUBE

In this chapter, we further test our simulation pipeline by applying it to the realistic system of Telescopic Double Wall Carbon Nano-Tube (TDWCNTs), and compare with literature results. This study on this system will also provide insights into similar overlapping geometries including the 2D top contact discussed in later chapters.

Theoretical and experimental examinations of the mechanical properties of double wall carbon nano-tubes revealed very low inter-shell friction forces, allowing the different shells of the two tubes to easily translate and rotate with respect to each other. This will result in TDWCNT, which consist of double rolled concentric tubes of graphene. It could be suitable for nano-electro and nano-mechanical systems. We visualize its structure in Fig. 4.1).

In a TDWCNT, the path of the electrical current along the two Single Wall Carbon Nano-Tubes (SWCNTs) is interrupted by the overlapping region and the current is forced to flow between the layers. Thus, the effect of the interlayer interaction is much evident than the one in conventional Double Wall CNTs (DWCNTs) [53]. This suggests the possibility of controlling the current by moving or rotating the layers relatively to each other. This will also resemble vertical transport systems of interest, including top contact as discussed in the next chapter, in the sense that its behavior is also controlled by the interlayer configuration. In this chapter, we will present the results of the transmission through TDWCNTs as a function of the overlap length in the following. The starting Hamiltonian is built using the tight-binding approach for the individual SWCNTs and an interlayer coupling parametrization taken from Ref. [54]. For the transmission calculation, we used the Landauer formalism as introduced.

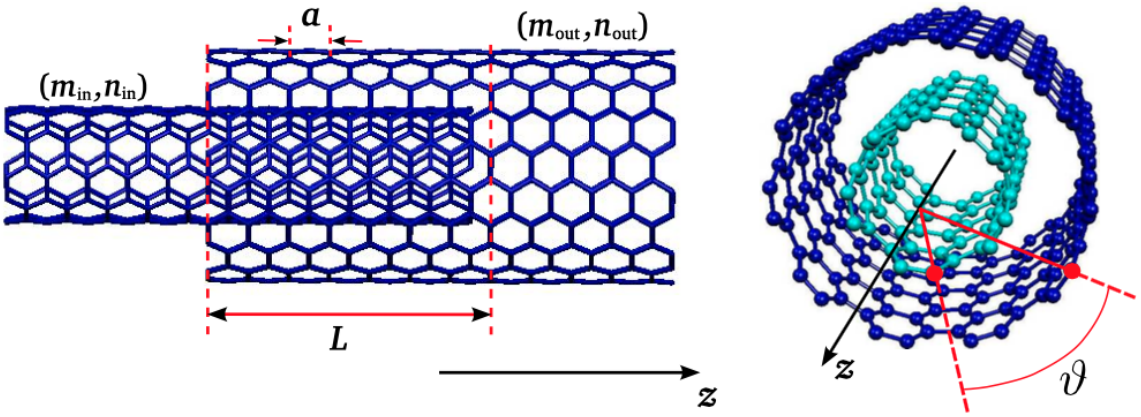


Figure 4.1: Left: Illustration of a DWCNT consisting of two commensurate SWCNTs sliding into each other over a length $L = Na + \Delta L$. a is the unit cell length. Right: Illustration of the rotation angle θ between the two SWCNTs: $\theta = 0^\circ$ is an arbitrarily taken reference angle.

4.1 Simulation setup

Consider a system containing a semi-infinite SWCNT sliding into a larger semi-infinite SWCNT as illustrated in Fig. 4.1. The variable corresponding to the overlapping length is L:

$$L = Na + \Delta L, \tag{4.1}$$

where a is the common unit cell, N is an integer, $0 < \Delta L < a$, and we have chosen θ to be zero.

In nearest neighbor tight binding models of SWCNTs, the onsite energy is set to zero $\epsilon_0 = 0$ eV and there are only nearest neighbor interactions between the carbon atoms, with the intra-shell hopping integral $\gamma_0 = 2.66$ eV. The inter-shell coupling between two atoms i, j in the different shells of a DWCNT can be parametrized as

$$\gamma_{ij} = \beta \cos\theta_{ij} \exp\left\{\frac{\Delta - d_{ij}}{\delta}\right\}, \tag{4.2}$$

with $\beta = \gamma_0/8$, $\Delta = 3.34$Å and $\delta = 0.45$Å [54]. Here d_{ij} is the distance between atoms i and j and θ_{ij} is the angle in the xy-plane. For the cutoff length d_{cutoff} of the inter-shell coupling parameter γ, we chose $d_{\text{cutoff}} = 0.97\sqrt{1.6a^2 + \Delta R^2}$, where ΔR is the actual inter-shell distance for the given tube helicities.

For the calculation of the transmission, we have conventionally divided the system in three regions: the DWCNT in the overlapping region as our "scattering region", the SWCNTs as semi-infinite leads. The transmission function and DOS will be calculated along the lines of the Landauer approach introduced in this section and thus we can compare the transmission and DOS of different overlap lengths for TDWCNTs of certain helicity.

4.2 Results and discussions

Fig. 4.2 shows the energy dependence for a zigzag (3,0)@(12,0) TDWCNT at two different L. For both overlap lengths, the transmission as a function of energy exhibits a series of minima around the Fermi level. The left panels correspond to an overlap length at which the transmission is maximal at the Fermi level, The right panels to an overlap with minimal transmission. The effects leading to the overlap dependent suppression of the transmission are low-energy effects, occurring around the Fermi level. Remarkable is also the strong suppression of the transmission in the high energy regions of the conducting bands.

The above results agree well with previous calculations in literature.[54]

4.3 Conclusions

To sum up, we further tested tested our simulation pipeline by applying it to the realistic system of Telescopic Double Wall Carbon Nano-Tube (TDWCNTs). The results agree well with literature, which again proved the accuracy of our approach. The study on the relationship between the overlap area and transport properties also provides insights into systems

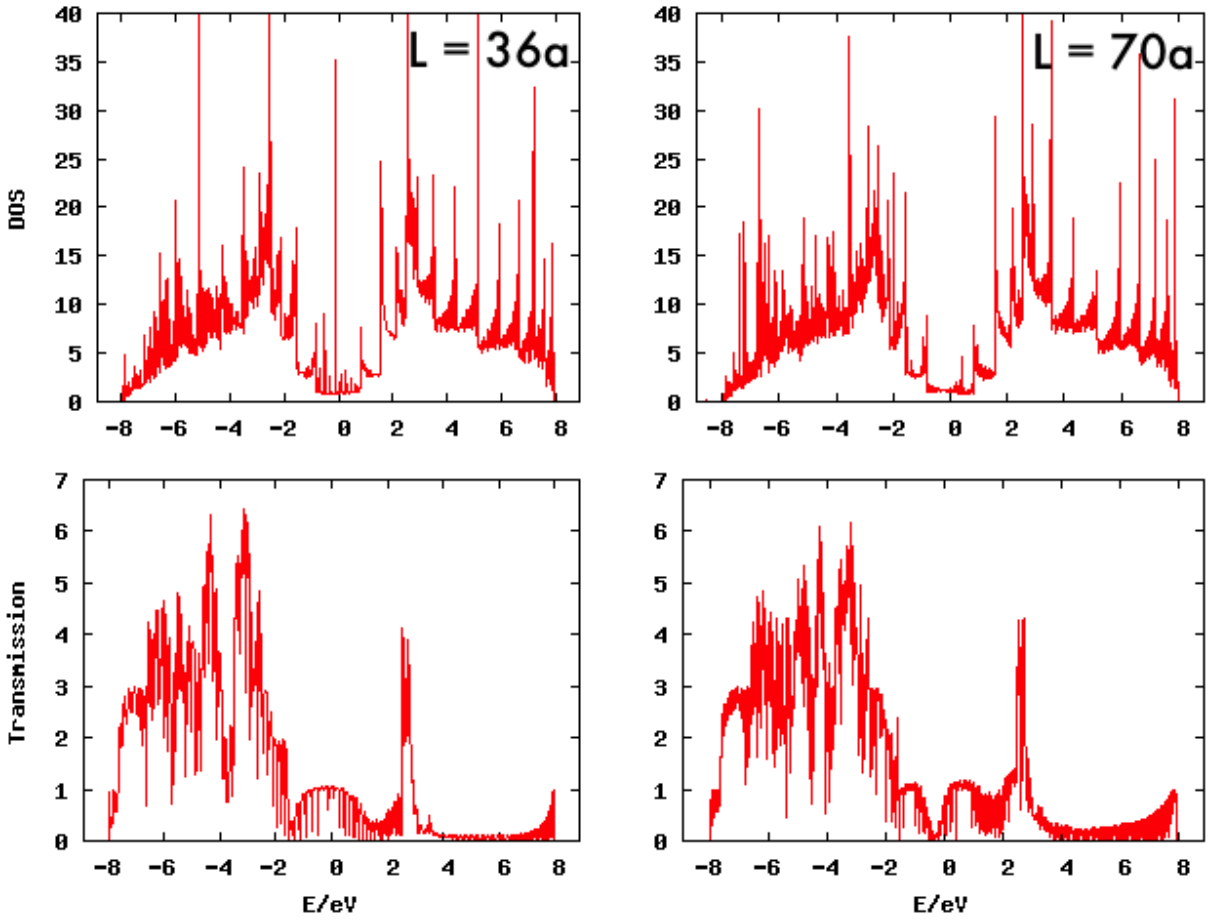


Figure 4.2: (top panels) DOS and (bottom panels) Transmission as a function of the energy for an (5,5)@(10,10) TDWCNT for $\theta = 0^{\circ}$ at (left panels) $L = 36a$ and (right panels) $L = 70a$

consisting of multiple layers e.g. the 2D top contact.

CHAPTER 5

TOP CONTACTS BETWEEN GRAPHENE AND MOS_2 MONOLAYERS

Two dimensional (2D) materials like graphene, black phosphorus, various transition metal dichalcogenides (TMDs) like MoS_2, WSe_2 and beyond are drawing significant attention as promising candidates for future electronic and optoelectronic devices which include logic transistors, radio frequency (RF) devices, light emitting diodes, solar cells and sensors of all types: chemical, biological, mechanical and thermal. The experimental results from first generation prototype 2D-devices show compelling evidence for high performance. However, as we move on to the second generation of 2D devices and start to shrink the device dimensions in order to further increase the performance (ON current, speed, sensitivity etc.), we run into fundamental scaling problems arising due to the contacts. Indeed, emphasis needs to be given on the scalability of metal-2D contacts, due to the fact that in an aggressively scaled device both channel length and contact length have to be reduced by similar factor. While length scaling reduces the channel resistance it also increases the contact resistance (RC) leading to a non-monotonic total resistance, which ultimately limits the device scaling. IBM researchers were able to scale the channel length of CNT field effect transistors (FETs) to sub-10nm regime; however, the contact lengths for such CNT-FETs could not be scaled beyond 200nm since contact resistance started to dominate the device performance. Therefore, it is important and timely to determine the ultimate scalability of contacts to 2D materials at an early stage since these are also being considered as alternative materials for future electronic and optoelectronic devices. One possible advantage that the 2D materials might have over 1D materials like CNTs is that, the 2D materials have one more degree of dimensionality. From a simple qualitative mode mis-match argument, it is easier to couple a three dimensional (3D) contact to a two-dimensional (2D) nanosheet than it is to a one-dimensional (1D) nanotube or nanowire. Note that the mode mis-match is responsible for

carriers back scattering at a metal-semiconductor contact interface and ultimately give rise to a finite contact resistance.

In this example study, we investigate the quantum transport of monolayer MoS_2 contacted with metallic graphene lead on top. Quantum conductance averaged over transverse (y-axis) Brillouin zone is computed and compared to band structure of both layers. Correspondence is found where the two sets of bands cross. We also tried to find optimize contact length by studying its relationship with quantum conductance.

5.1 Simulation setup

The general approach employed for simulating transport through an interface between a metal and a 2D semiconductor is detailed in Chapter 2. Here, a brief summary is presented along with the applied numerical parameters.

MoS_2 monolayer

Ab-initio calculation for monolayer MoS_2 is performed using the Quantum Espresso code. The in-plane lattice constant for the relaxed structure is $a_{MoS_2} = 3.186\mathring{A}$. Hopping parameters for monolayer MoS2 is extracted by using the Maximally Localized Wannier Functions (MLWF) method. As initial projections $\mathbf{Mo} : l = 2, \mathbf{S} : l = 1$ are supplied to WANNIER90. The order of nearest neighbor coupling is chosen to be three to well reproduce the band-structure. The MLWF-interpolated bands are compared to *ab-initio* bands, as shown in Fig. 5.2. We can see that chosen third-order hopping parameters can very well reproduce the bands obtained directly from the *ab-initio* code.

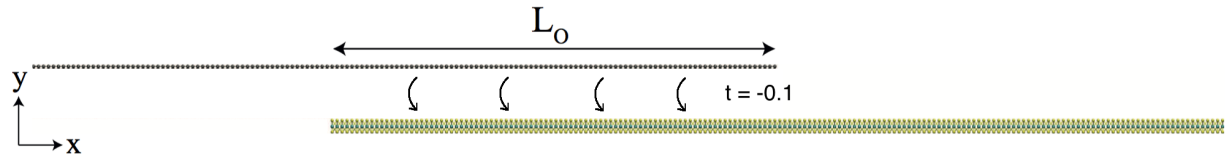


Figure 5.1: Atomic geometry of the modeled graphene(top)-MoS_2(bottom) vertical heterostructure. The vertical hopping is assumed only between the C atom of graphene and top S atom of the MoS_2 monolayer. The hopping parameter is chosen to be t = -0.1.

Graphene

For graphene, we use typical nearest-neighbor hopping parameters of -2.8 eV. The in-plane lattice constant is set to the same as MoS_2, i.e. a 29.5% extensive strain. Further improvement could be made by doing *ab-initio* calculation for a supercell containing different number of unit cells for MoS_2 and graphene, so that only a small strain needs to be added.

Graphene-MoS_2 heterostructure

Atomic geometry of the modeled graphene-MoS_2 vertical heterostructure is shown in Fig. 5.1. We have manually chosen the vertical hopping parameter to be t=-0.1, which is comparable to similar calculations in literature. We also assume that vertical hopping only happens between the C atom of graphene and top S atom of the MoS_2 monolayer.

Transport

For the transport calculation, we use the Landauer's method with Fisher-Lee relationship

$$T(k_y, E) = Tr\left[\Gamma_L(k_y, E)G^R(k_y, E)\Gamma_R(k_y, E)G^A(k_y, E)\right]. \tag{5.1}$$

A rectangular principal layer containing 5 unit cells in the transport direction x and 3 unit cells in the transverse direction y is constructed, so that couplings between different

principal layers are limited to nearest-neighbor. Then the principal layer is repeated along the transport direction, with the number of repetitions determined by the overlap length L_O, see Fig. 5.1. The transmission is averaged over the 1st Brillouin zone and sampled with 256 points in the k_y direction at $k_x = 0$,

$$T(E) = \frac{1}{2\pi} \int dk_y T(E, k_y). \tag{5.2}$$

5.2 Results and discussions

5.2.1 Quantum conductance

First, the quantum conductance of the metal-semiconductor vertical heterostructure is calculated without applying any external electrostatic potential. In Fig. 5.3, the conductance is calculated for the interface area with 10 repetitions of principal layers, i.e. an overlap length of 15.7 nm.

Further, we compared the result for transmission with the bandstructure of both graphene and monolayer, see the upper plot in Fig. 5.3. We find that transport is restrained within the energy gap and mainly happens in the energy range where the bands cross, with small amounts everywhere else. The reason is that both energy and momentum can only be conserved at those crosses, where same electron states exist in both layers. No phonons are needed in such transport, resulting in a large current compared to everywhere else.

5.2.2 The effect of transfer length

As a next step, we would like to understand the influence of the metal-semiconductor interface's area on the transmission. We then repeat the same procedure for different overlap lengths, and compare their conductances. The result is shown in Fig. 5.4.

We find that the two conductance peaks first increase fast with overlap length and then

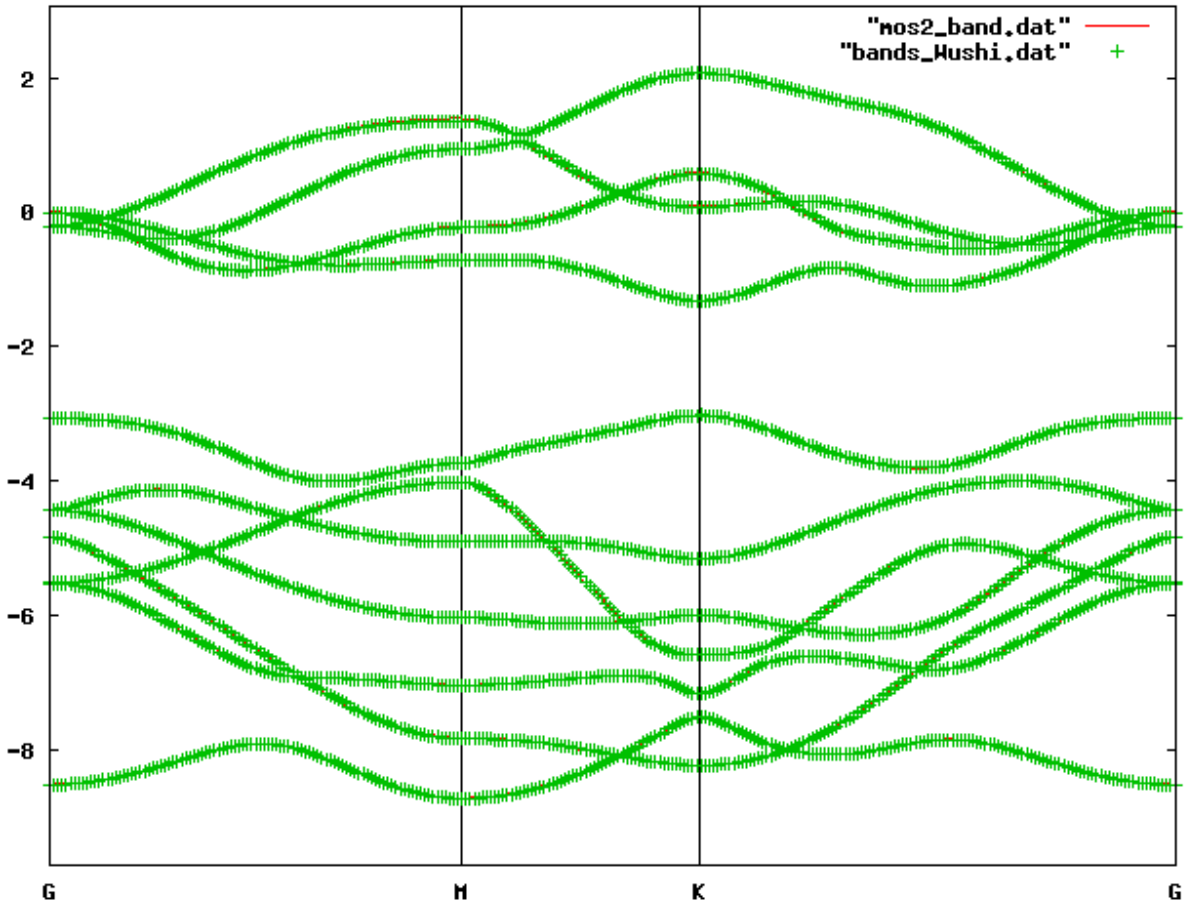


Figure 5.2: MLWF-interpolated band structure of single-layer MoS_2 (green) compared to *ab-initio* bands (red). We can see that chosen third-order hopping parameters can very well reproduce the bands obtained directly from the *ab-initio* code.

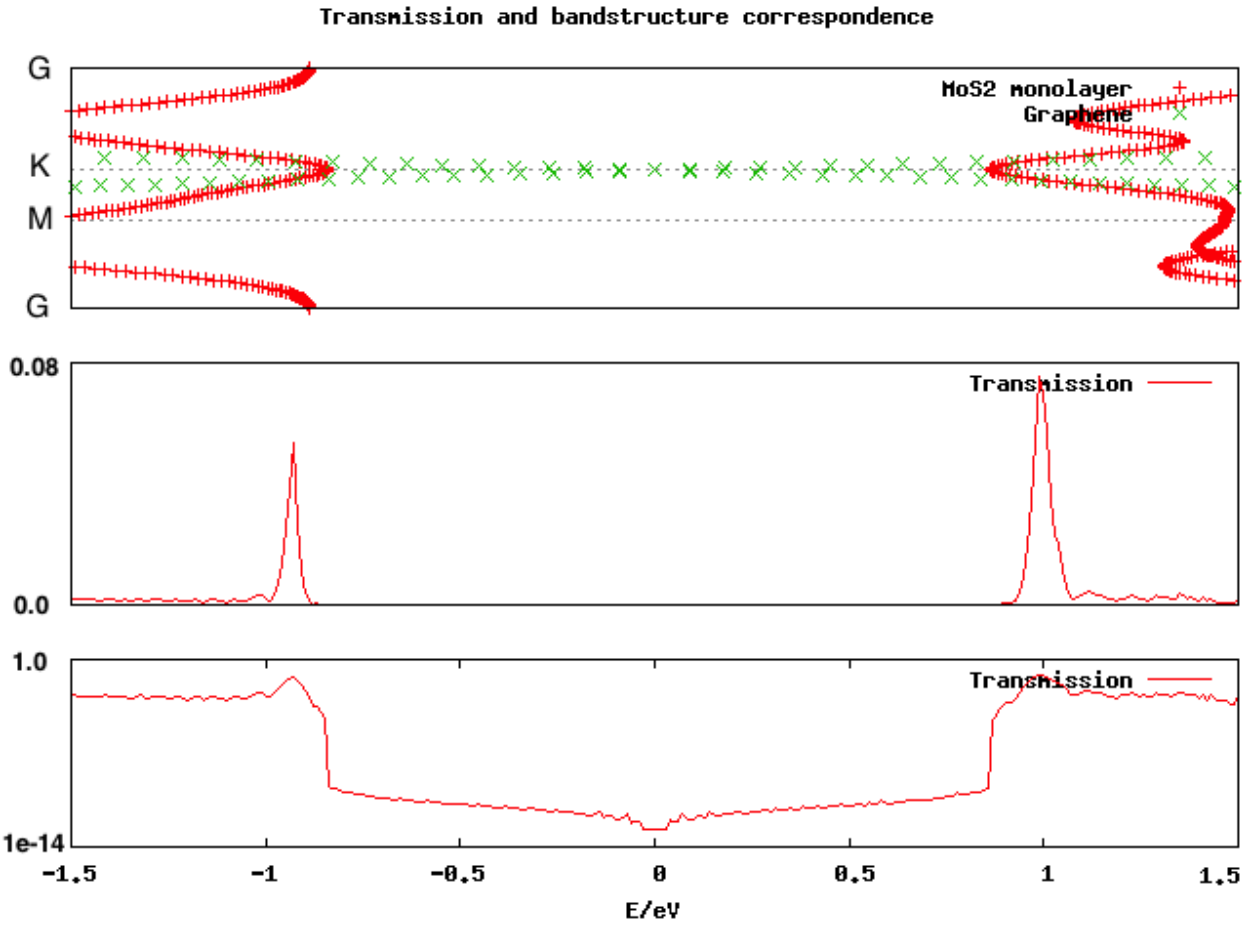


Figure 5.3: (a) Upper plot: bandstructure of both graphene (green) and MoS_2 monolayer (red). (b) Middle plot: transmission spectrum for the graphene-MoS_2 vertical heterostructure with a overlap length of 15.7 nm. (c) Lower plot: same transmission spectrum plotted on a log scale.

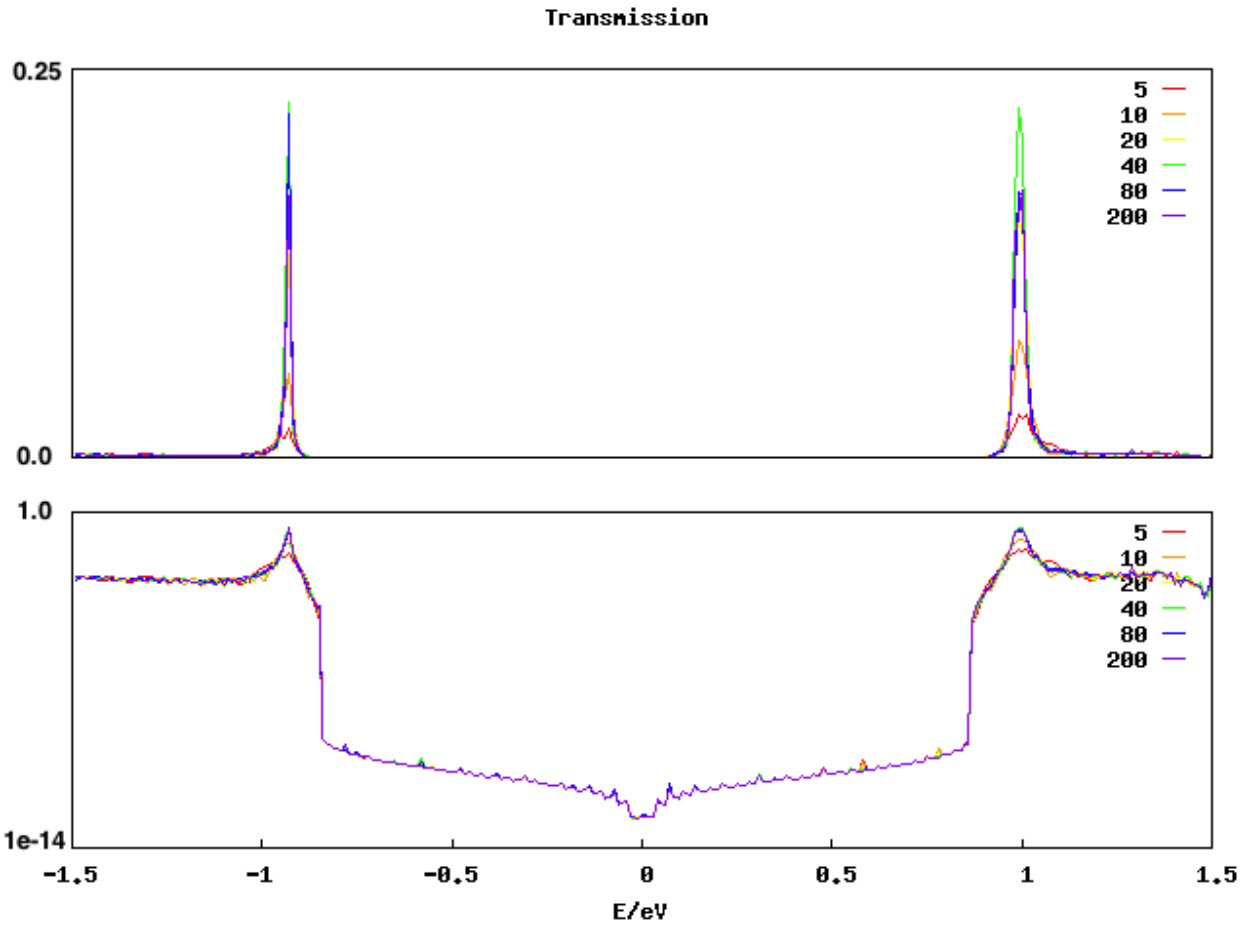


Figure 5.4: Transmission spectrum for different overlap lengths. The upper plot is on linear scale and the lower plot is on log scale. The legend shows the number of repetitions.

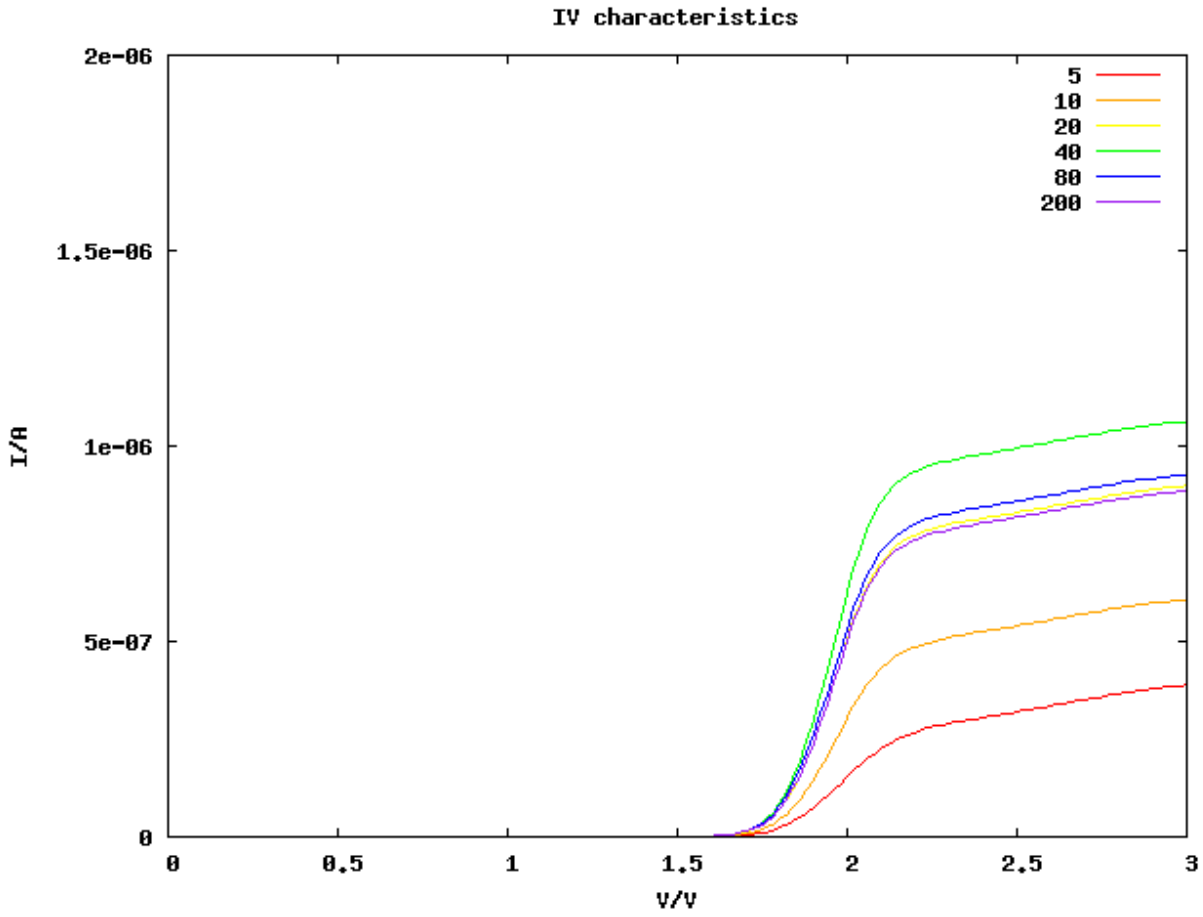


Figure 5.5: IV characteristics for different overlap lengths.

decrease gradually after repetition goes to 40, i.e. an overlap length of 62.8 nm. The peak also becomes sharper with the increase of system length. We then extract the current from transmission spectrum as shown in Fig. 5.5.

We can see that the contact resistance first decrease to a minimum value and then gradually increase with overlap length, which is similar to the case of a classic model of resistor network.

5.3 Conclusions

We calculated the quantum conductance for the graphene-MoS_2 vertical heterostructure with different overlap lengths. We find that transport mainly happens in the energy range where the bands cross. The reason is that same electron states exist at those crosses for both layers, thus transport can happen between those states with few assists from scattering, resulting in a large current compared to everywhere else. We also extract the vertical current from transmission spectrum and find similar changes with respect to overlap lengths as shown in the classic model of resistor networks. We find the optimal overlap length that results in a minimum contact resistance to be approximately 62.8 nm. As for next steps, it would be interesting to explore how the spectral current flows vertically as a function of position. In addition, the effect of phonon scattering on the electron transport in top contacts are largely unknown and we reserve these for future study.

CHAPTER 6

OHMIC EDGE CONTACTS BETWEEN TWO-DIMENSIONAL MATERIALS

6.1 Introduction

This chapter is taken from a published paper.[55]

Two-dimensional (2D) materials such as graphene and transition metal dichalcogenides (TMDCs) are pushing the forefront of complementary metal-oxide semiconductor (CMOS) technology beyond the Moore's law [1, 2], and show great promise for realizing atomically thin circuitry [3, 4, 5]. A fundamental challenge to their effective use remains the large resistance of electrical contacts to 2D materials for probing and harnessing their novel electronic properties [6, 7, 8]. There are generally two types of contact geometries, namely top contacts and edge contacts [6]. Conventional 3D metallic top contacts can achieve low contact resistance with monolayer 2D materials, but cannot avoid the intrinsic problem of large electrode volume. [9, 10, 6, 11] 2D top contacts, including graphene [12, 13, 14] and recently demonstrated atomically flat metal thin films [11], can achieve both small volumes and low contact resistances of metal-semiconductor interfaces, but they suffer from weak van der Waals coupling to TMDCs [15]. Their transfer efficiency depends largely on the contact area and is compromised dramatically below a transfer length which is typically tens of nm scale [16, 8]. In contrast, 2D edge contacts are formed by joining atomically thin metal electrodes and semiconductors laterally in a single plane. They offer the possibilities for high-quality contacts to 2D materials despite minimal contact area defined by their atomic thickness as shown by both simulations [15] and recent experimental successes [16, 56]. Among them, the graphene-MoS_2 system considered in this paper is particularly promising for a low-resistance 2D edge contact [16, 57, 56]. According to the Schottky-Mott rule, the combination of a low-work-function metallic graphene electrode and a typical n-type [58] semiconducting

monolayer MoS_2 channel naturally leads to a small SBH. Moreover, the overall system is stable under working conditions and resistant to phase transitions induced by adsorbates. While improving experimental techniques makes more tests feasible, a better quantitative understanding of the electronic structure and transport properties is still critical for improving the design of 2D edge contacts. In 2016, Yu et al [59] suggested a highly non-localized carrier redistribution and strong reduction of Fermi level pinning in 2D systems based on a semi-classical macroscopic model. In 2017, Chen et al [60] performed first-principles studies based on density functional theory (DFT) on the morphologies of the graphene/MoS_2 lateral junction and proposed several stable interface configurations. Sun et al [61] performed similar studies and tried to calculate the transport efficiencies, but did not reproduce the linear I-V characteristics observed in experiments, possibly due to the lack of doping in the semiconductor region.

In order to better model the graphene/MoS_2 interface at the atomic scale and quantitatively calculate the charge transfer properties, we introduced a custom-built self-consistent quantum transport solver based on the Keldysh Nonequilibrium Green's Function formalism [62, 63] and Maximally Localized Wannier functions (MLWFs) [24]. Such a method can efficiently solve the local electrostatics and electron transport with first-principles accuracy at a minimal cost of tight-binding calculations. It enables the inclusion of large areas of both materials, which is necessary in order to allow for a long screening length for charged interfacial states and thus to have equilibrium conditions near the edge of the central device region. This is also a necessary condition for the decimation technique to account for the effects of the semi-infinite leads. We find that trapped interface states lead to a potential barrier, which is however small enough that we find Ohmic behavior at room temperature and high enough doping levels. We successfully reproduced the linear current-voltage (I-V) characteristics with a resistivity value of approximately 30 $k\Omega \cdot \mu m$ close to that observed in experiments [56] at room temperature, which is a first for 2D edge contact systems. At lower

temperatures, We observe increasing non-linearity as a result of reduced thermalization. In the following, we calculate the band structures of graphene and monolayer MoS_2, and extract their tight-binding parameters using MLWFs. Based on these parameters, we use our custom-built quantum transport solver to generate the electrostatic potential self-consistently with the inhomogeneous charge densities induced by band bending, and by local impurity states. We confirm the validity of our solver by comparing the converged electrostatic potential profile to the analytical predictions from Thomas-Fermi screening theory, beyond angstrom distances from the contact region. Finally, we calculate the transport properties based on the Keldysh formalism and discuss how to further improve device performance.

6.2 Results and discussion

Wannier functions can accurately and efficiently capture delicate electronic structures. We used them to extract the tight-binding parameters of both materials after obtaining the band structures in the DFT framework. Figure 6.1 compares the band structures obtained with DFT and with the MLWF Hamiltonian for both monolayer graphene and MoS_2. From the plot, we can see that the Wannier projections work so well that the differences between the two bands are largely unnoticeable. We also compare the total Density of States (DOS) and the Projected Density of States (PDOS) reproduced by the MLWF orbitals for both materials. Instead of using all the orbitals in a unit cell, we choose only those contributing to the DOS near the Fermi level for Wannier projections. This can minimize the sizes of matrices used in our calculations and further reduce computational cost. The extracted hopping parameters then serve as basis for the transport simulations.

The geometry of the graphene-MoS_2 edge contact is sketched in Figure 6.2a and 6.2b. We assume periodicity of the device in the y direction, at a level of a few unit cells which is long enough to approximately match the lattice constants. As a result, the Hamiltonian shows a k_y dependence and can be decomposed into three components according to the Bloch's

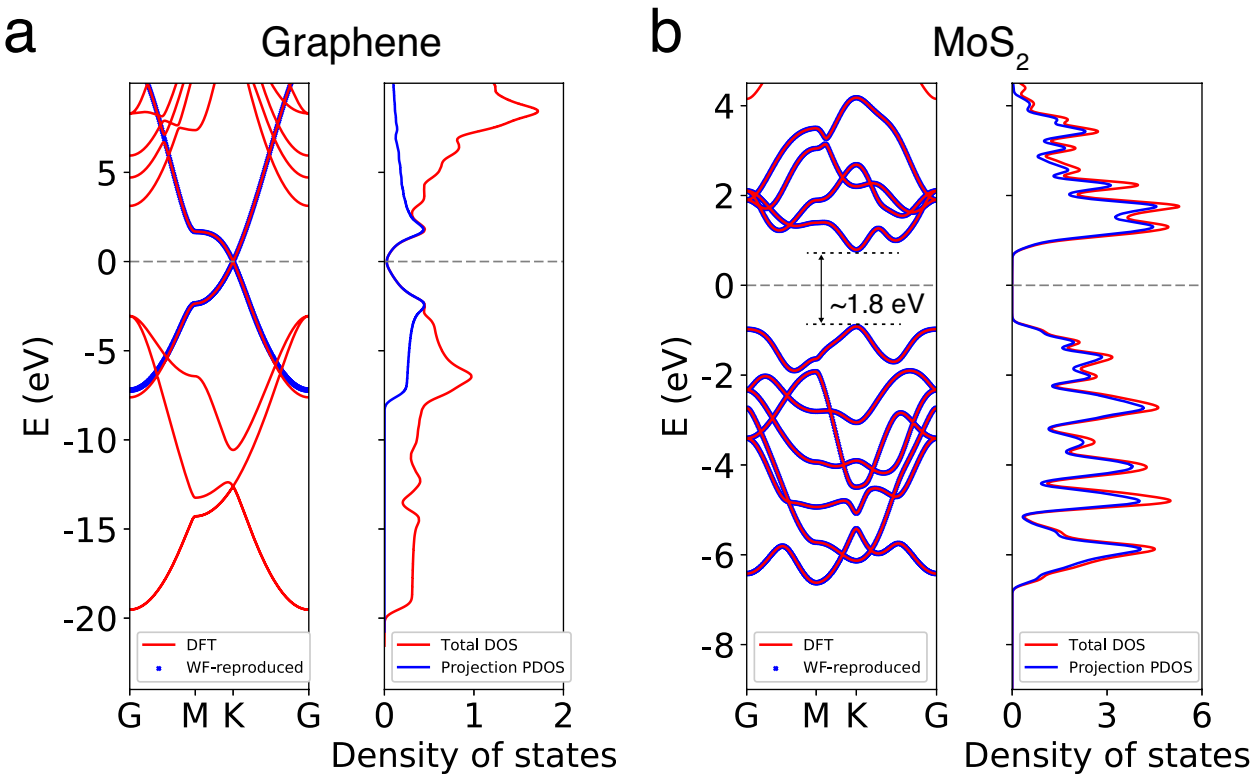


Figure 6.1: (a) Left: Graphene band structures obtained with DFT and with MLWF Hamiltonian. Right: Graphene total DOS and its PDOS reproduced by the MLWF orbitals. (b) The same plots for monolayer MoS_2. The produced band gap of approximately 1.8 eV is very close to the experimental one. The zero energy is set to the Fermi level (dashed line) for both materials.

theorem as

$$H(k_y) = H_0 + H_- e^{-\mathrm{i}k_y\Delta} + H_+ e^{\mathrm{i}k_y\Delta}, \tag{6.1}$$

where H_0 are the interactions within a strip of width, $H_\pm$ are the interactions with a neighbor strip along the $+y$ or $-y$ direction, and Δ is the width of the supercell. To make second-nearest-neighbor interactions negligible, we choose Δ to be exactly 4 times the width of the graphene unit cell, and approximately 3 times that of the MoS_2 unit cell, resulting in a lattice mismatch of only 4.2%. We do not change the lattice constant of graphene because it has a much larger Young's modulus [64]. For the interface, we consider the predominant zigzag edge of graphene [65] and MoS_2 [66] as shown in Figure 6.2a and choose a structure motivated by *ab-initio* calculations. According to Chen et al [60], the configuration chosen in our study has the lowest formation energies among other alternative geometries. We adjust the Fermi level of both materials to match the induced doping by gate voltage. For the details of the parameters used in this study and their effects on the simulation results, please refer to our supporting information. Figure 6.2c illustrates the band alignments of the edge contact device simulated in this paper. Having established the atomic geometry and band alignments of the junction, we now investigate the electrostatics and charge transfer effects at the boundary.

To evaluate the tunneling barrier, we calculate charge densities self-consistently with electrostatic potential for the edge contact device. We use the Nonequilibrium Green's Function technique based on the Keldysh formalism to calculate the charge densities, and the non-linear Newton-Raphson technique to solve for the electrostatic potential from the Poisson equation. We performed simulations for different source-drain biases at a MoS2 doping level of $4 \times 10^{14} cm^{-2}$ and at room temperature of $T = 293K$. The converged potential and charge profiles are shown in Figure 6.3a and Figure 6.3b respectively. We can see that the electrostatic potential reaches equilibrium at both edges of our simulated region. Applied source-drain biases shift the potential level in the two electrodes and in turn

modify the net charge profile. In the inset of Figure 6.3b, we enlarge the plot for the MoS_2 side to better show how the net charge distribution adapts to the external biases. Here we safely ignored the electron-phonon scattering, due to the short channel length of 2D edge contacts. In order to check the validity of the electrostatics obtained from our self-consistent simulations, we compare the converged potential profile with the analytical predictions of a quasi-1D Thomas-Fermi screening potential, and the result is shown in the inset of Figure 6.3a. We find that the two results agree well beyond angstrom distances from the contact region, where the charge densities can stay low enough for the Thomas-Fermi theory to work well. This confirms the accuracy of our method. The derivations of the quasi-1D Thomas-Fermi screening potential are given in the supporting information.

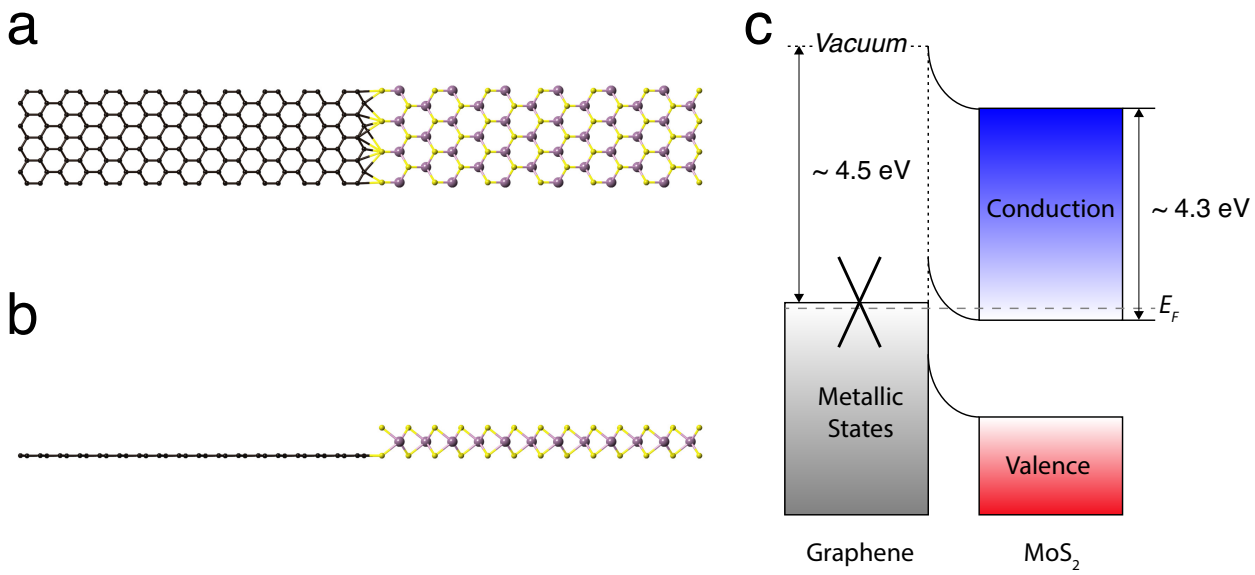


Figure 6.2: (a) Top, and (b) side views of the simulated edge contact device region. (c) Schematic illustration of its band alignments. The work function of graphene is about 4.5 eV and the electron affinity of monolayer MoS_2 is 4.3 eV [60]. E_F stands for Fermi energy.

Using the converged electrostatic profiles, we further examine the quantum transport properties of the graphene-MoS_2 edge contact by calculating its Local Density of States (LDOS) and transmission spectrum using the Landauer-Buttiker formalism. The transmis-

sion coefficients are determined by the equation:

$$T = Tr[G^R \Gamma_L G^A \Gamma_R] \tag{6.2}$$

where Γ_L and Γ_R are the linewidth functions that describe the coupling between the scattering region and the two leads on the left and right. In Figure 6.3c, We show the two results under room temperature and zero bias side by side.

From the results in Figure 6.3, we find that the electrostatic potential is screened by the electrons and decays fast as one goes away from the interface, which allows charge carriers to tunnel through the boundary efficiently. When the metallic graphene contacts with monolayer MoS_2 contacts, free electrons will flow from the graphene side to the MoS_2 side since the work function of p-type graphene is smaller than that of n-type MoS_2. When the charge redistribution reaches equilibrium, graphene is positively charged whereas MoS_2 monolayer is negatively charged near the interface region, in which a built-in electric dipole is induced. [67] In addition, trapped charges at the interface produce a monopole, which we find to be substantial. Such electric fields can shift the energy bands of the MoS_2 monolayer upward. However, we find that the barrier is efficiently screened by the free charges and becomes thin enough for the electrons to tunnel through. From Figure 6.3c, we can see the two-fold effects of the interfacial bonding: The trapped charges at the interface form a thin potential barrier, which is screened effectively allows electrons to go through; The overlap states serve as a bridge inside the barrier further assisting with charge transfer. As a result, for given parameter settings, no Schottky barrier is present in this case, and we observe sufficient transmission near the Fermi level, indicating the ohmic nature of the graphene-MoS_2 interface. From our simulations, we find that one of the main factors controlling the Schottky barrier height is the MoS_2 doping level. A higher doping level makes the conduction band minimum shift downward, resulting in a smaller barrier and larger tunneling current. On the contrary, a lower doping level raises the conduction band minimum up, even to a point where

it will stop the edge contact from being ohmic. The ohmic behavior in our simulation requires doping level one order of magnitude larger than the one reported by experiments. One possible explanation is that we only assume electron hopping between graphene supercells and sulfur atoms immediately next to the interface. By considering interactions of longer range, one can make the overlap states couple better with the conduction band. This could potentially improve transport efficiency and lower the needed doping level to achieve ohmic behavior in our simulation. Also, although we find that relative perpendicular positions of the two materials hardly affect the transmission efficiency, other boundary configurations including different edge types (armchair or zigzag) and interface roughness could still alter the positions of the overlap states, and in turn change the tunneling current. We leave a systematic examination of the above factors to future work.

We further check the reliability of our interface modeling. By integrating the boundary DOS over energy within the band gap of the MoS_2 interior states, we find the number of the interfacial states in our system to be about $n \approx 9.4$ states nm^{-1}. This is close to the full DFT simulation results reported by Chen *et al* [60] ranging from 6.3 to 8.3 states nm^{-1}, and therefore proves the accuracy of our modeling method.

To explore more on the ohmic behavior of graphene-MoS_2 edge contact, we calculate the source-drain currents under different biases and temperatures. The electric current can be calculated as:

$$I(V) = \frac{2e}{h} \int T(E, V_L, V_R)[f_L(E, V_L) f_R(E, V_R)]dE \tag{6.3}$$

where $T(E, V_L, V_R)$ is the transmission coefficient given by equation (6.2), $V = V_L V_R$ is the bias voltage, and $f_{L/R}(E, V_{L/R})$ is the Fermi distribution function of the left/right lead. In Figure 6.4a, we find that at a high MoS_2 doping level of $4 \times 10^{14}\ cm^{-2}$, the I-V curves of the graphene-MoS_2 edge contact show linear characteristics at room temperature, with ohmic behavior maintained down to temperatures at least as low as 50 K. Moreover, we run the same calculation for a MoS_2 doping level of $2 \times 10^{14}\ cm^{-2}$, and find non-linear I-V behaviors

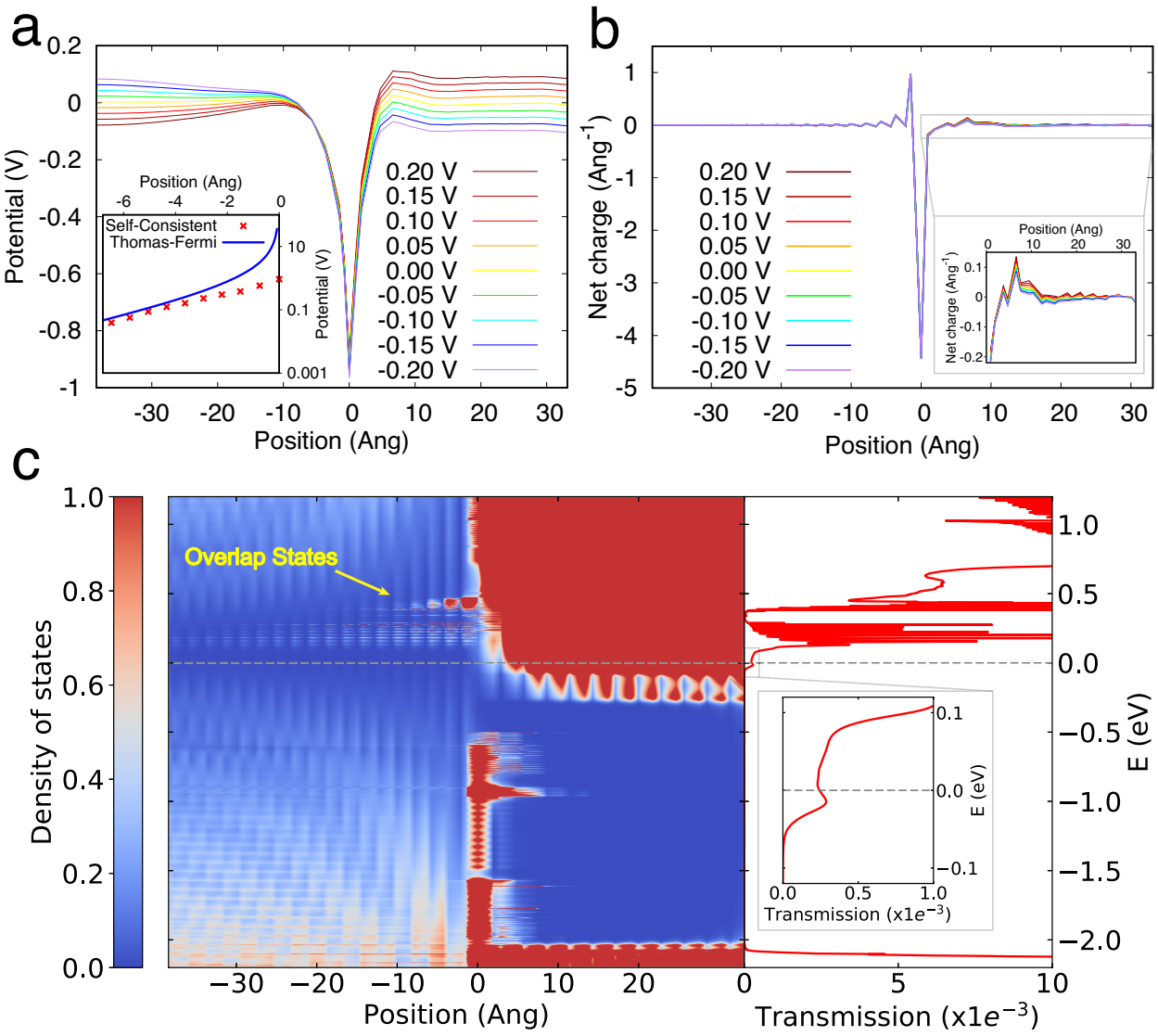


Figure 6.3: (a) Converged electrostatic potential profiles under different biases. Inset: Comparison of the magnitude between 1D Thomas-Fermi screening (solid line) and our self-consistent simulation (dots) on a log scale at zero bias. (b) Converged net charge profiles under different biases. Inset: The same results enlarged for the MoS_2 side to show the difference under different biases. (c) LDOS and transmission spectrum of the simulated graphene-MoS_2 edge contact at zero bias. Inset: Transmission spectrum near the Fermi level at $E = 0$.

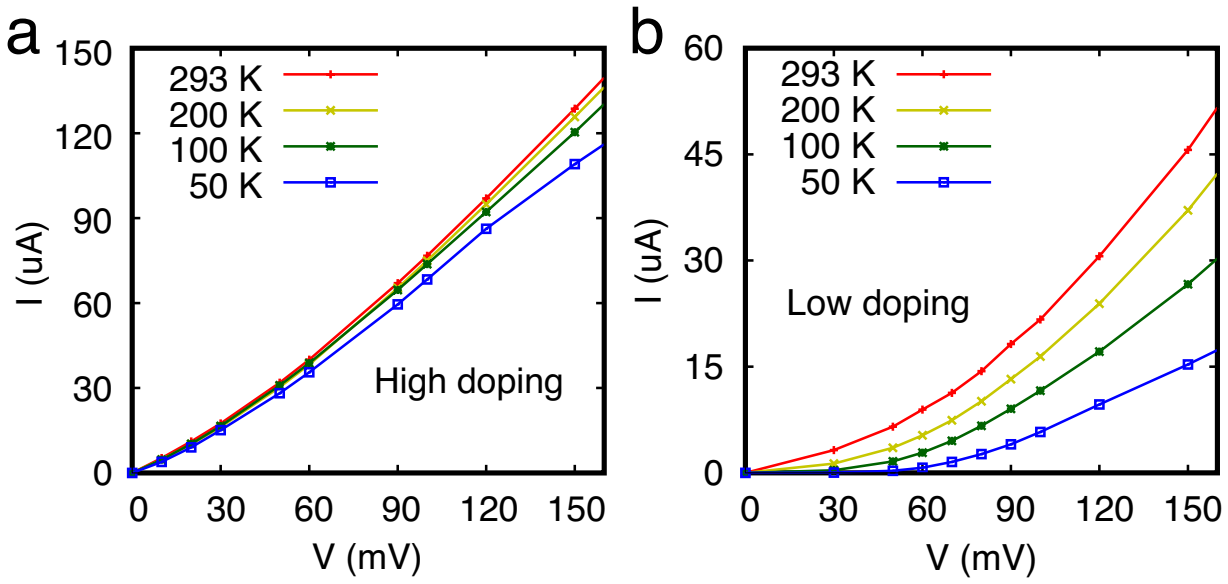


Figure 6.4: I-V characteristics under different temperatures for MoS_2 doping levels of (a) $4 \times 10^{14}\ cm^{-2}$, and (b) $2 \times 10^{14}\ cm^{-2}$.

due to the existence of a large Schottky barrier (see Figure 3b of Supporting Information) at low carrier densities, as shown in Figure 6.4b. This barrier also leads to smaller currents at lower temperatures because reduced thermalization makes it harder for the electrons to go across the interface, which also agrees with experiment. [56]

From Figure 6.3c, we notice that the resonant levels from graphene edge states can assist in the carrier injection across the interface in the range of 0.1 to 0.5 eV. One could potentially take advantage of these edge states by adding more n-type doping to graphene and effectively moving down its resonant levels closer to the Fermi level. We verified our prediction by performing simulations with the above changes, and indeed obtained larger transmission values at the Fermi level, as shown in Figure 4 of our Supporting Information. This leads to a lower resistance with a greater source-drain current under the same bias. Therefore, we propose that the usage of an n-doped graphene electrode could be a further improvement to the present edge contact design.

6.3 Band structure calculation and Wannierization

We performed the DFT calculations as implemented in the package QUANTUM ESPRESSO. We used a plane wave basis set, ultrasoft pseudopotential, and Perdew-Burke-Ernzerhof (PBE) generalized gradient approximation (GGA) exchange-correlation functional, which produces a band gap very close to the experimental one ($E_{g,exp} = 1.8$ eV) in the trigonal-prismatic form of single-layer MoS_2 [68]. The plane wave cutoff was 100 Ry for wavefunctions and 400 Ry for the charge density. A 15 Å interlayer distance was used to eliminate interlayer interaction. The momentum space was sampled on a $36 \times 36 \times 1$ Monkhorst-Pack k-point grid for graphene and $25 \times 25 \times 1$ for MoS_2. Spin-orbit coupling is neglected. The simulated cell is optimized until the atomic forces decrease to values less than 10^{-3} a.u.. The convergence criterion is set to less than 10^{-6} eV total energy difference between two subsequent iterations.

We then used the `wannier90` code[48] to determine the Maximally Localized Wannier Function basis for extracting the tight-binding hopping parameters in the system Hamiltonian. Instead of using all the orbitals in the unit cell, we choose only those contributing to the Density of States (DOS) near the Fermi level for Wannier projections. This can minimize the size of matrices used in our calculation and further reduce computational cost. We use one atomic p_z orbital for each carbon atom to reproduce the Dirac cones of graphene. In order to capture the seven highest valence bands and four lowest conduction bands of single-layer MoS_2, we use all three p-like Wannier functions centered on each sulfur atom and all five d-like projections on each molybdenum. We included tight-binding parameters up to three orders of nearest neighbors to reproduce the band structure.

6.4 Effects of different parameters on the simulation results

Here, we show the effects of different model parameters the simulation results, including the local density of states (LDOS) and transmission spectrum. This also demonstrates the

robustness of our model. We study below the following four parameters: MoS_2 doping, graphene doping, interfacial hopping strength, and temperature. The LDOS is evaluated using the the equation:

$$LDOS(E) = \frac{1}{2\pi}[G^r(\Gamma_L + \Gamma_R)G^a] \tag{6.4}$$

6.4.1 MoS_2 doping

We match the doping level induced by gate voltages by adjusting the Fermi level of MoS_2. For the device used in our manuscript, we used a MoS_2 doping of $4 \times 10^{14}\ cm^{-2}$. Here in Figure 6.5, we compare the LDOS and transmission spectrum for several more doping levels of MoS_2 for room temperature, zero bias and hopping strength of $t_0 = -1.0\ eV$. We find that the main effect of the MoS_2 doping concentration is shifting the bands of MoS_2. A higher doping level makes its conduction band minimum move downward, and leads to a larger transmission around the Fermi level.

6.4.2 Graphene doping

The Fermi level of graphene is lowered to account for an unavoidable p-type doping of approximately $10^{12}\ cm^{-2}$ resulted from the manufacturing process in the lab. In our paper, we suggest that the usage of n-doped graphene could potentially improve transport efficiency. The related simulation results are shown in Figure 6.6. From the plots, we can see that the transmission in Figure 6.6b for the edge contact device using an n-doped graphene lead is significantly larger than the one in Figure 6.6a using a p-doped graphene lead, especially around the Fermi level. The reason is that the edge states of n-doped graphene are closer to the Fermi level compared to p-doped graphene. This helps the electrons near that energy range to tunnel through and leads to a lager transmission at the Fermi level.

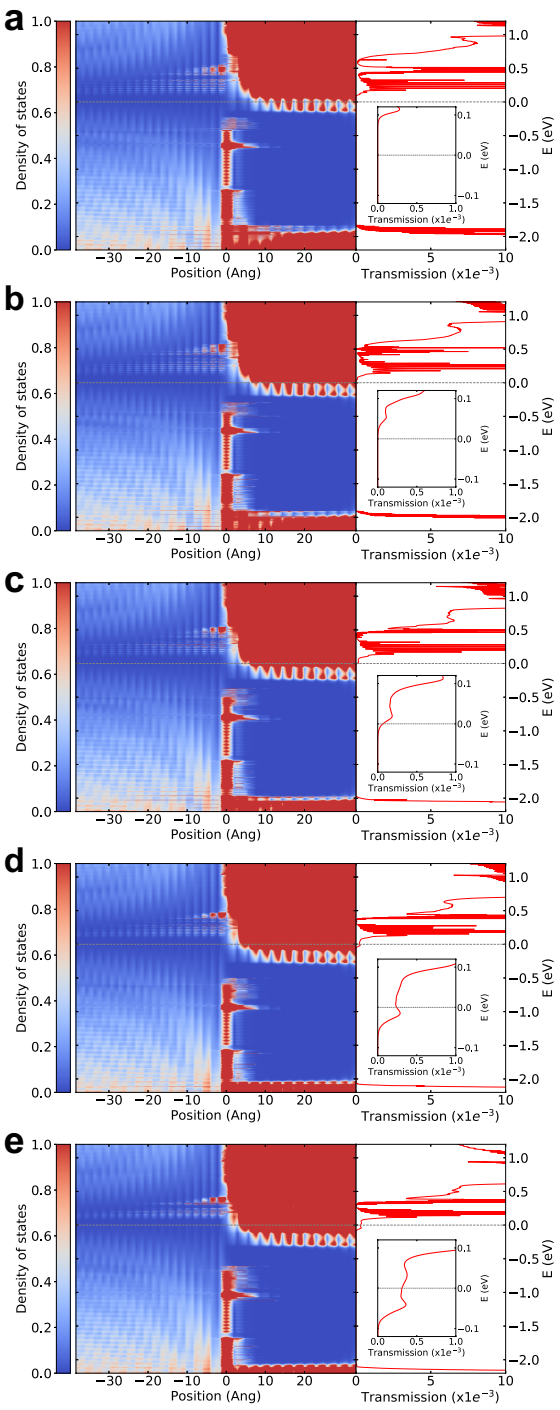


Figure 6.5: LDOS and transmission spectrum for different MoS_2 doping levels: (a) $1 \times 10^{14}\ cm^{-2}$ (b) $2 \times 10^{14}\ cm^{-2}$ (c) $3 \times 10^{14}\ cm^{-2}$ (d) $4 \times 10^{14}\ cm^{-2}$ (e) $5 \times 10^{14}\ cm^{-2}$. Inset: Transmission spectrum near the Fermi level at E = 0. ($t_0 = -1.0\ eV$, $T = 293\ K$)

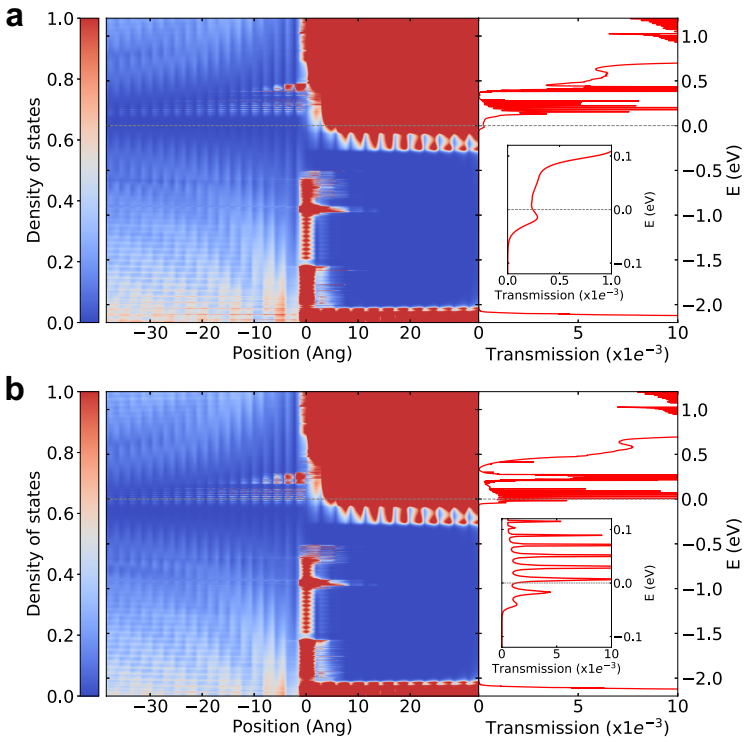


Figure 6.6: LDOS and transmission spectrum for p-doped and n-doped graphene leads: (a) p-doped (b) n-doped. Inset: Transmission spectrum near the Fermi level at $E = 0$. Notice the transmission scale used in (b) is an order of magnitude larger than that used in (a). (MoS_2 doping $= 4 \times 10^{14}\ cm^{-2}$, $t_0 = -1.0\ eV$, $T = 293\ K$)

6.4.3 Interfacial hopping strength

We consider interfacial interactions only between the p_z-like MLWFs of graphene edge carbon atoms and the p_x, p_y-like MLWFs of MoS_2 edge sulfur atoms. The hopping parameter is modeled by an exponential dependence on the distance as $t = t_0 e^{-(r r_0)}$, where r is the distance between two MLWF centers and t_0 is the interaction strength at a distance of $r_0 = 1.5$ Å, which is assumed to be the shortest distance between the graphene and the MoS_2 region. We manually set t_0 to be comparable to the strength of the covalent bonds formed at the interface. Comparison of the number of interfacial states in our simulation with the first-principles one proves the accuracy of our interface modeling method. From Figure 6.7, we can see that interfacial hopping strength mainly affects the magnitude of the transmission. Larger interfacial hopping strengths result in larger transmission coefficients.

6.4.4 Temperature

In Figure 6.8, we compare the LDOS and the transmission spectrum at different simulation temperatures. We find that the temperature affects the thermalization of electrons but does not largely alter the transmission spectrum. The reason why we observe smaller currents at lower temperatures is because the decreased thermalization makes it harder for electrons near the Fermi level to overcome the potential barrier. This can also be seen from the calculations of electric current in equation (3) of our manuscript.

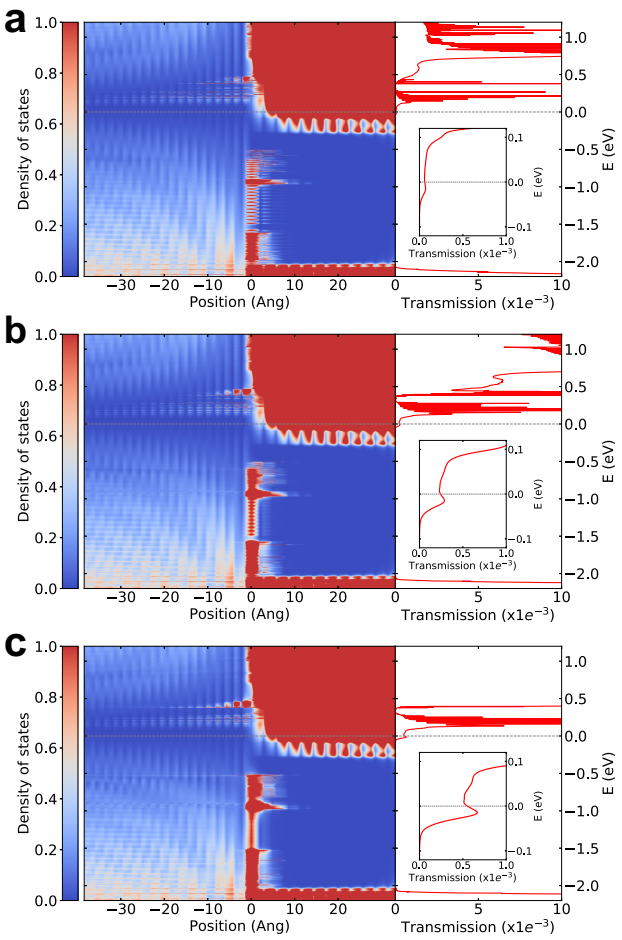


Figure 6.7: LDOS and transmission spectrum for different interface hopping strengths t_0: (a) $t_0 = -0.5\ eV$ (b) $t_0 = -1.0\ eV$ (c) $t_0 = -1.5\ eV$. Inset: Transmission spectrum near the Fermi level at E = 0. (MoS_2 doping $= 4 \times 10^{14}$ cm^{-2}, $T = 293\ K$)

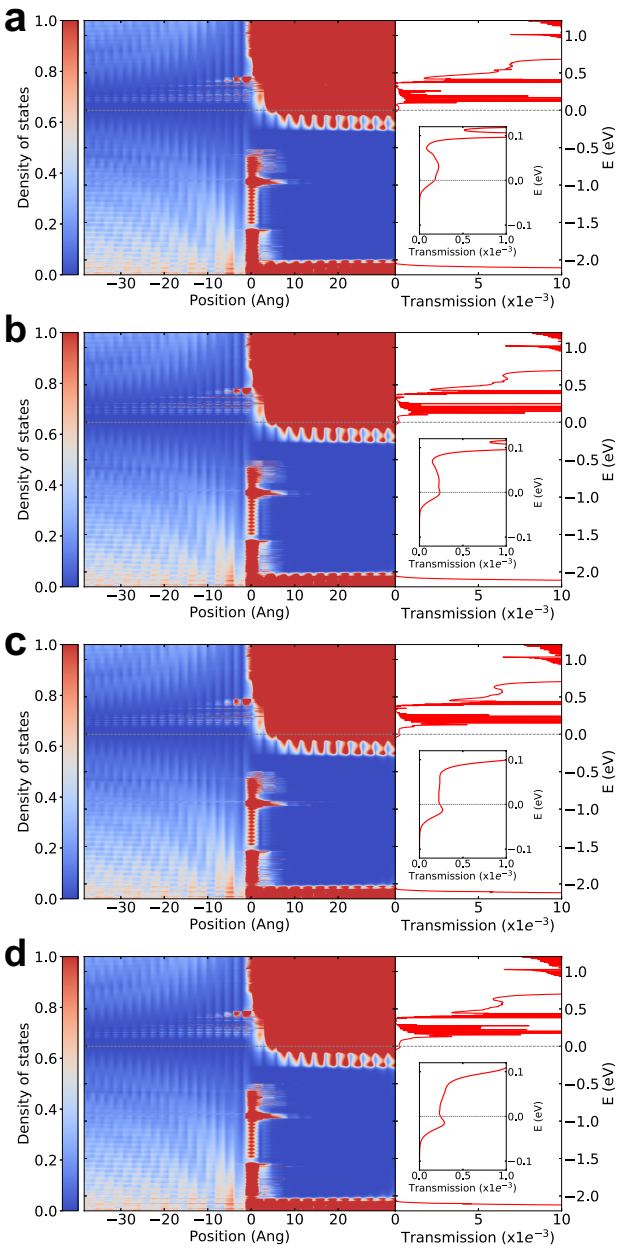


Figure 6.8: LDOS and transmission spectrum for different temperatures T: (a) $T = 50\ K$ (b) $T = 100\ K$ (c) $T = 200\ K$ (d) $T = 293\ K$. Inset: Transmission spectrum near the Fermi level at E = 0. (MoS_2 doping $= 4 \times 10^{14}\ cm^{-2}$, $t_0 = -1.0\ eV$)

6.4.5 Large bias

We have further explored the I-V characteristics of the edge contact for a larger source-drain bias range beyond available experimental data. We show the results in Figure 6.9. We find that the current curves begin to show structure due to features in the density of states. For example, we see some non-monotonic dependence between current on bias and temperature. This is related to features in the density of states near the edge of the integral window in the Landauer-Buttiker formula, i.e. $-\frac{V}{2}$ to $\frac{V}{2}$. For example, the current in Figure 6.9a at V = 0.6 V under 50 K is larger than the ones under 100 K and 200 K. This is due to the dip in the density of states around 0.3 eV as shown in Figure 6.5d. Under a low temperature of 50 K, the current at 0.6 V is less affected by that dip and therefore larger. So the monotonic dependence is a result of large variations in the DOS and different smearing for different temperatures.

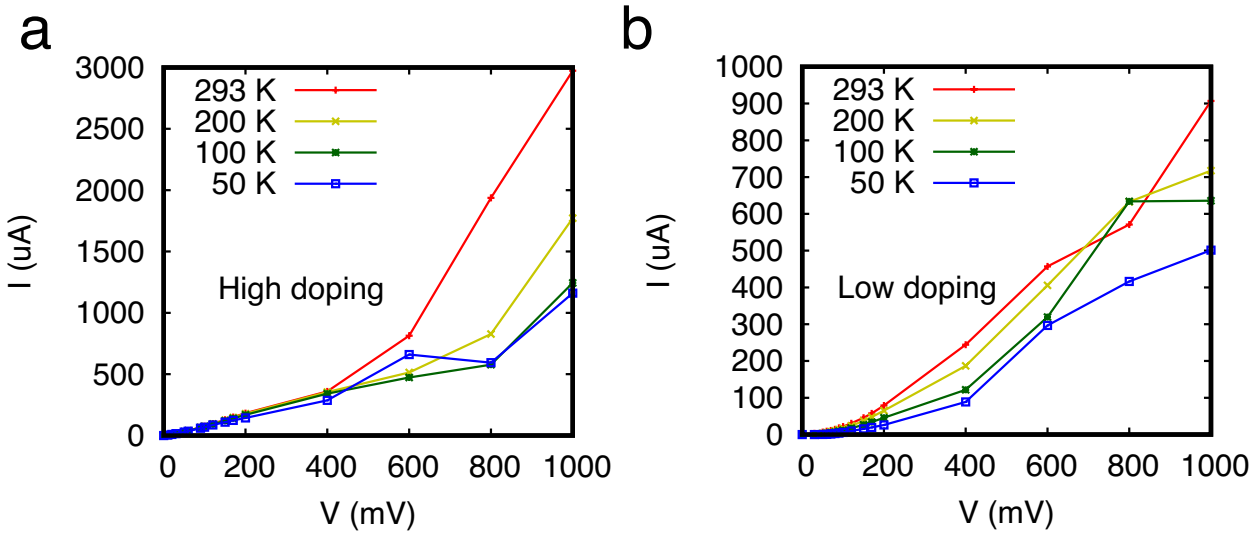


Figure 6.9: I-V characteristics under different temperatures and large bias for MoS_2 doping levels of (a) $4 \times 10^{14}\ cm^{-2}$, and (b) $2 \times 10^{14}\ cm^{-2}$.

6.5 Conclusions

In conclusion, we have developed a computational pipeline to study the electrostatic and transport properties of the 2D graphene-MoS_2 edge contact, and proposed a possible explanation for its ohmic behavior observed in experiments. By applying the custom-built quantum transport simulation scheme, we obtain the charge density profile self-consistently with the electrostatic potential profile of the device. We find that the potential barrier decays fast away from the interface and is thin enough for the electron to tunnel through efficiently. Our results are consistent with both analytical Thomas-Fermi screening theory and the experimental measurements. Because our methods can be scaled effectively to large systems, but maintain the fidelity of *ab-initio* band structures, they can be used to efficiently predict the electrostatic and transport properties for nanostructures, including those with complex geometries. These findings could have broad implications in the design and fabrication of metal-semiconductor junction for realizing low-resistance contacts.

CHAPTER 7

CONCLUSIONS AND OUTLOOK

In the course of this work, we developed a first treatment that includes a nonlinear poisson solver with atomistic accuracy which can treat systems of physical dimensions corresponding to actual devices. Some light has been shed on the electronic structure and quantum transport properties of several nano structures including the metal-semiconductor contacts, which is critical for further shrinking the dimensions of semiconductor devices. This thesis explained the theoretical methods and implementation details relevant to the developed efficient pipeline for quantum electron transport simulations.

In chapter 4, we compared the transport results obtained by our method with that by full DFT simulation. We find that our methods agree very well with the DFT method but uses three orders of magnitude less time. Then in chapter 5 we tested our transport calculations by applying it to the telescopic double wall carbon nanotube, where two nanotube of different radius overlap with each other. The obtained results also match those of literature.

In chapter 6, we applied our simulation pipeline to metal-semiconductor top contact. We choose the materials of the most interests to the two dimensional material community, namely metallic graphene leads and MoS_2 semiconducting channel. We find that the transfer efficiency depends largely on the contact area and is compromised dramatically below a transfer length which is typically tens of nm scale.

In chapter 7, on the other hand, we investigated in-plane edge contacts, which have the potential to achieve lower contact resistance due to stronger orbital hybridization compared to conventional top contacts. We then present full-band atomistic quantum transport simulations of the graphene/MoS_2 edge contact. We find that the potential barrier created by trapped charges decays fast with distance away from the interface, and is thus thin enough to enable efficient injection of electrons. This results in Ohmic behavior in its I-V characteristics, which agrees with experiments. Our results demonstrate the role played by trapped

charges in the formation of a Schottky barrier, and how one can reduce the Schottky barrier height (SBH) by adjusting the relevant parameters of the edge contact system.

Our framework can be extended conveniently to incorporate more general nanostructure geometries. For example, a full 3D solution of the electrostatics will also lead to better modeling of the electrical potential. Furthermore, better *ab-initio* calculations can be conveniently added to our methods to further improve their accuracy. Extending the method to properly model bulk 3D materials or 1D systems requires only minor modifications of the available codes.

One potential improvement on top of the established pipeline is the inclusion of electron-phonon interaction to represent the effect of heat on the electron transport. We have calculated the electron-phonon coupling matrices for graphene and MoS_2. Fig 7.1 and 7.2 show the couplings for both graphene and MoS_2. We have also included the Frölich interaction for LO phonons to properly treat the divergence at the long-wavelength limit for MoS_2.

We also transform the electron-phonon coupling matrices to real space using the Wannier technique [69] similar to the ones introduced in chapter 3, but with emphasis on phonons instead of electrons. The advantage of this is the convenience of including the effects of heat as represented by the quasi-particle phonons in the same real space as the electron transport calculation. Fig 7.3 and 7.4 show the comparisons of *ab-initio* phonon dispersion with the ones reproduced in real-space using the Wannier technique, for both graphene and MoS_2. From the figures, we can confirm the accuracy of this method.

A complete treatment of both electron and phonon transport would work well especially at nanometers of scale, where heat cannot dissipate fast enough given a limited volume. These can be incorporated directly into our current established pipeline to further improve simulation accuracy.

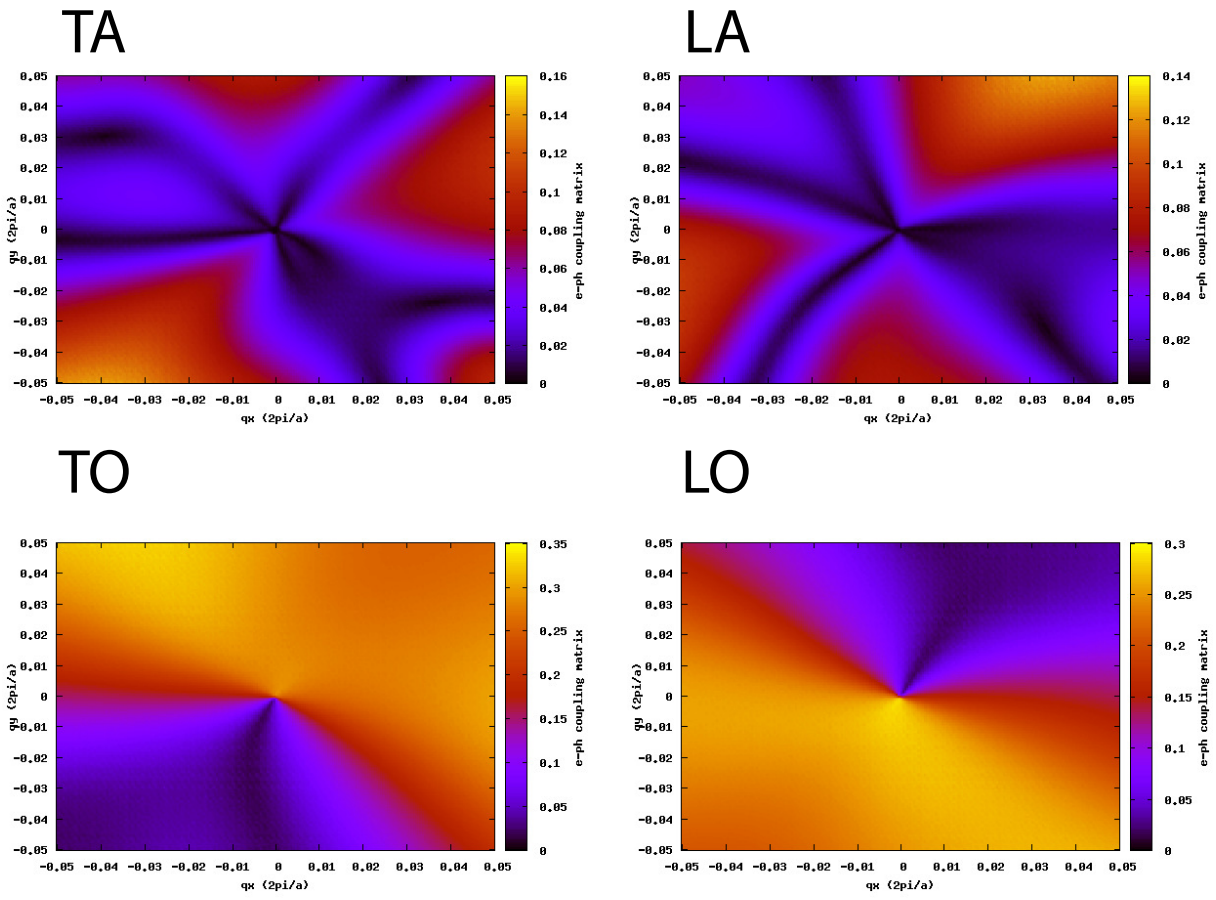


Figure 7.1: Electron-phonon coupling constants of all phonon branches for monolayer graphene in the reciprocal space.

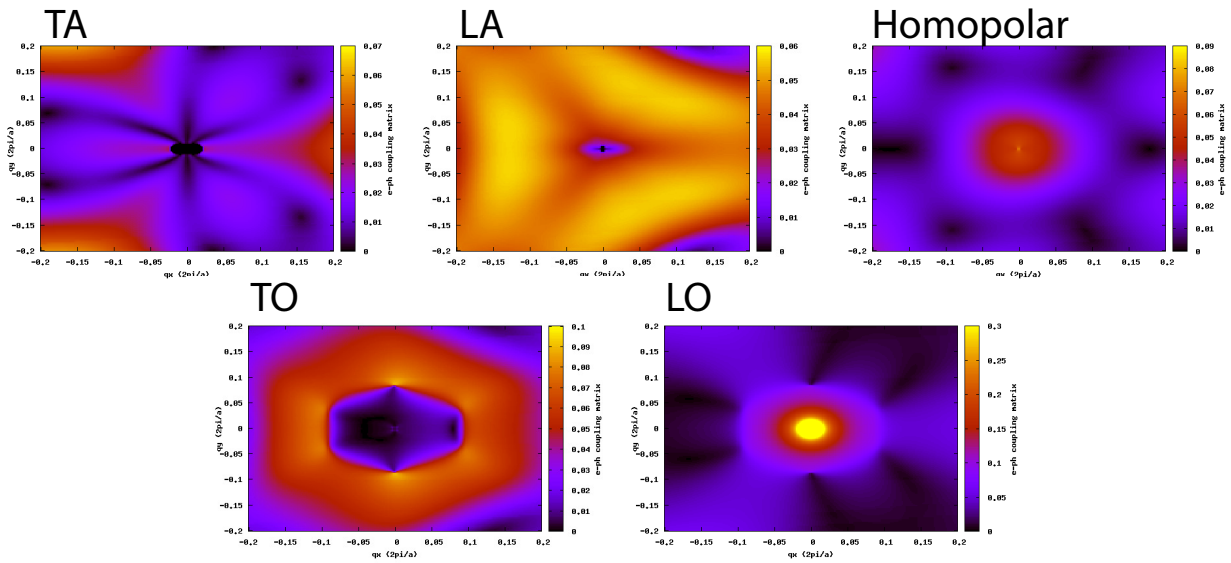


Figure 7.2: Electron-phonon coupling constants of all phonon branches for monolayer MoS_2 in the reciprocal space.

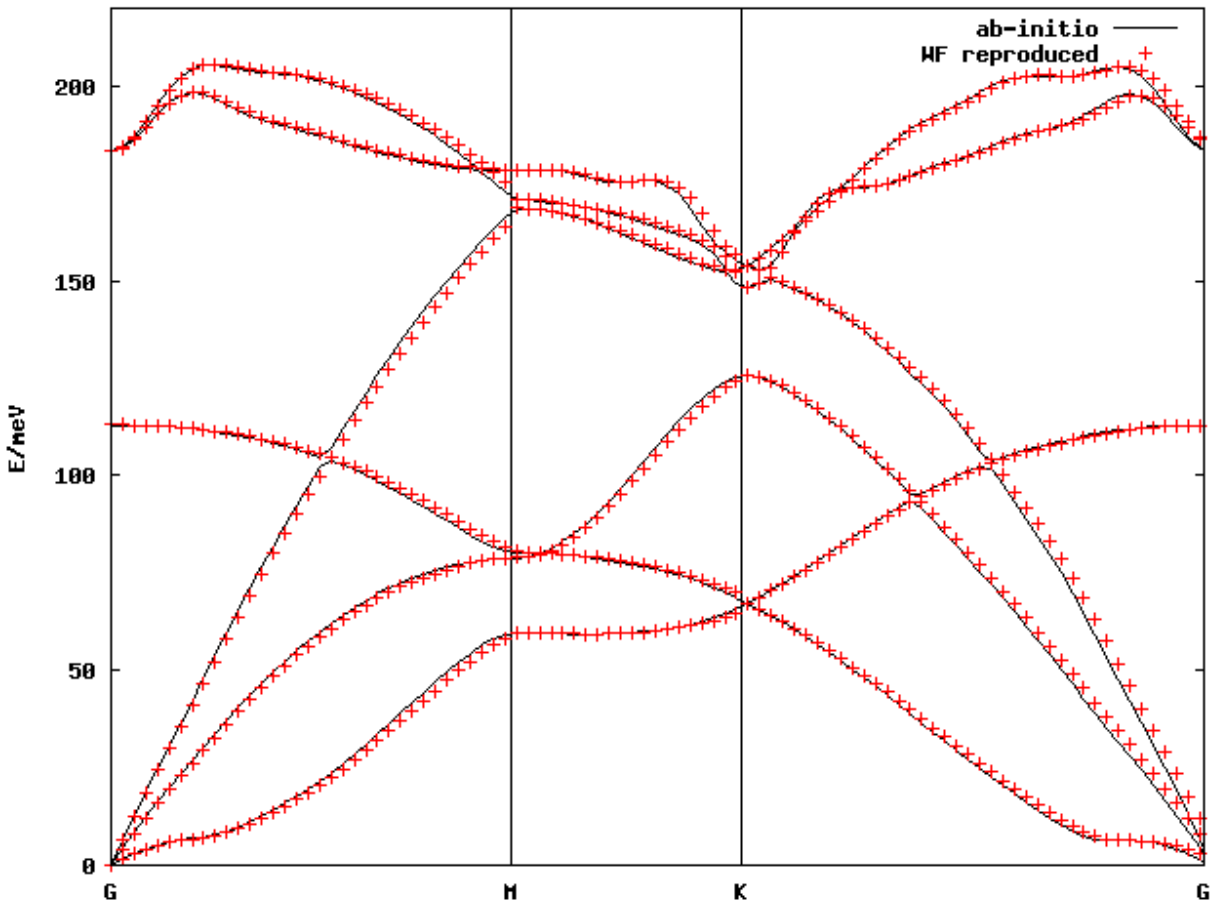


Figure 7.3: Phonon dispersion of monolayer graphene using both *ab-initio* simulation and the real-space Wannier technique.

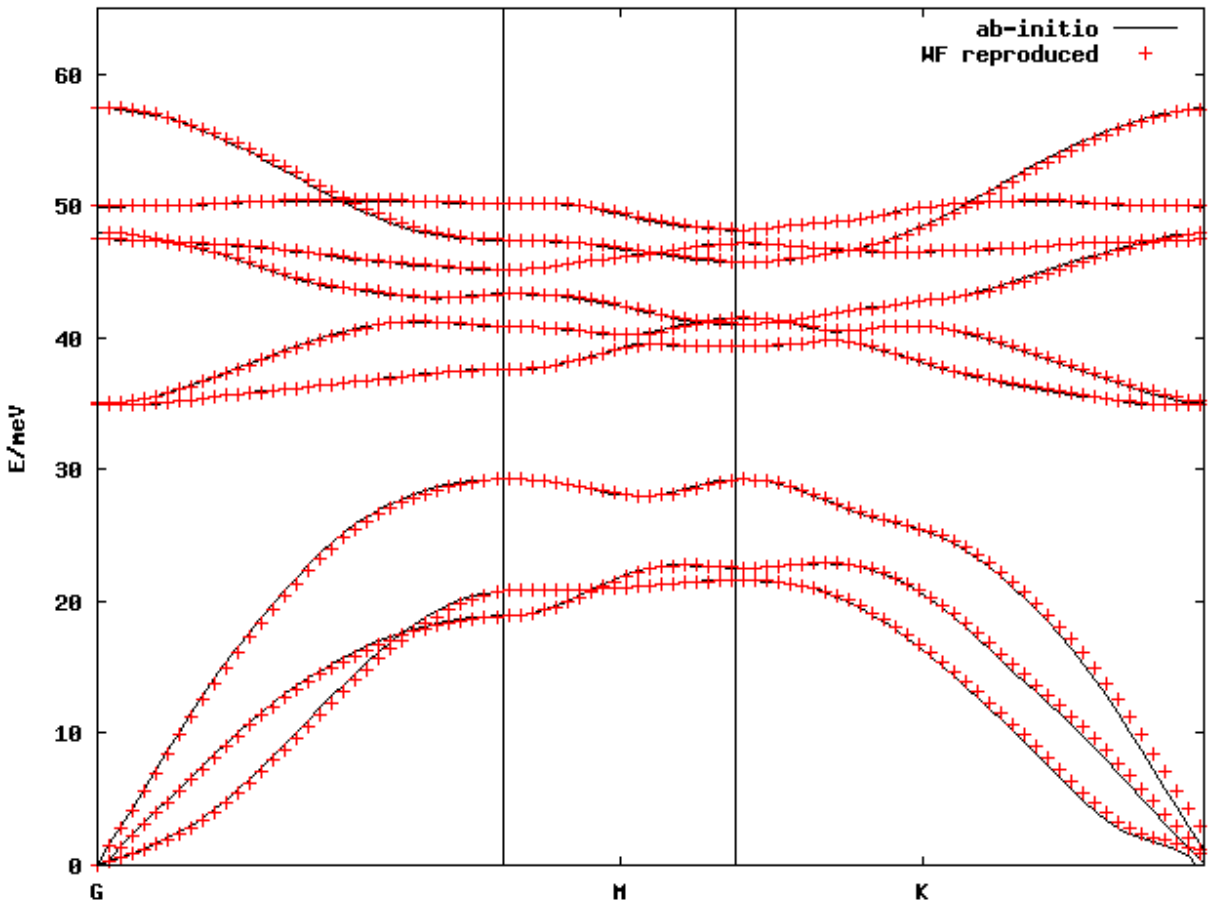


Figure 7.4: Phonon dispersion of monolayer MoS_2 using both *ab-initio* simulation and the real-space Wannier technique.

APPENDIX A

DECIMATION TECHNIQUE FOR CALCULATING SELF-ENERGIES OF SEMI-INFINITE LEADS

As shown in Fig. A.1, one can split the layered structure into left contact marked by L, central device region marked by C, and right contact marked by R. The device corresponds to the region where one solves the transport equations and the contacts are the metallic regions connected to the device. While the device region consists of only N layers, the matrix equation corresponding to Eq. (2.55) is infinite due to the semi-infinite contacts. It is shown next that the influence of the semi-infinite contacts can be added in a recursive fashion into the device region, where the semi-infinite contacts only affect the edge layers 1 and N of the device region.[40, 70, 71, 72]. The Sancho-Rubio method [40] for calculating the surface Green's functions will be described here, which will then be used to obtain the self energies of the semi-infinite leads. Any solid with a surface can be described by a semi-infinitely stacking principal layers with nearest-neighbor interactions. [43] If the bulk periodicity on the surface plane is preserved to the surface, then $\boldsymbol{k}$ is a good quantum number. We can build Bloch-state orbitals for each atomic orbital ϕ_α along the direction of any plane, e.g. the λth atomic plane of the nth principal layer. Take m orbitals per atom, and assume each

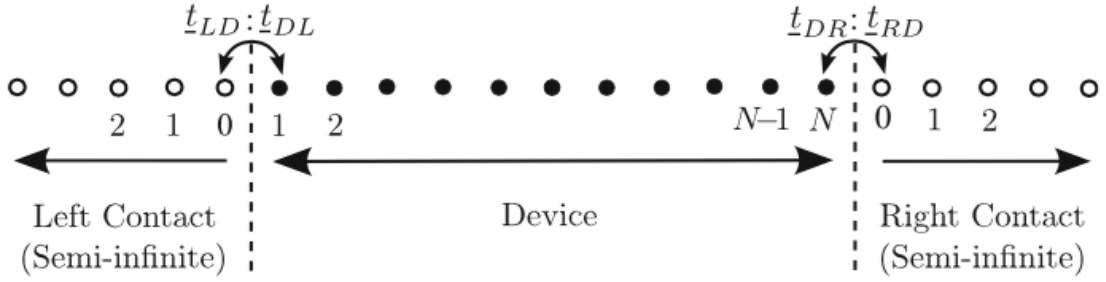


Figure A.1: 1D model including the central device region, and left and right contacts

principal layer is composed of l atomic planes. Then one can form Bloch states for each principal layer:

$$\Psi_n(\boldsymbol{k}_\parallel) = \begin{pmatrix} \varphi_n^{11}(\boldsymbol{k}_\parallel) \\ \vdots \\ \varphi_n^{\lambda\alpha}(\boldsymbol{k}_\parallel) \\ \vdots \\ \varphi_n^{lm}(\boldsymbol{k}_\parallel) \end{pmatrix} \tag{A.1}$$

where

$$\varphi_n^{\lambda\alpha}(\boldsymbol{k}_\parallel) = \frac{1}{\sqrt{N_\parallel}} \sum_{\boldsymbol{R}_\parallel} e^{(i\boldsymbol{k}_\parallel \boldsymbol{R}_\parallel)} \phi_n^{\lambda\alpha}(\boldsymbol{R}_\parallel) \tag{A.2}$$

and $N_\parallel$,and $\boldsymbol{R}_\parallel$ denote the numbers of atoms, and lattice vectors in an atomic plane.

Taking matrix elements of $(\omega - H)G(\omega) = \mathbb{1}$ between the Bloch states, one has the usual chain for each $\boldsymbol{k}_\parallel$:

$$\begin{aligned} (\omega - H_{00})G_{00} &= \mathbb{1} + H_{01}G_{10} \\ (\omega - H_{00})G_{10} &= H_{01}^\dagger G_{00} + H_{01}G_{20} \\ &\vdots \\ (\omega - H_{00})G_{n0} &= H_{01}^\dagger G_{n-1,0} + H_{01}G_{n+1,0} \end{aligned} \tag{A.3}$$

where $n = 0$ denotes the surface principal layer and the matrices

$$\begin{aligned} H_{nn'}(\boldsymbol{k}_\parallel) &=< \Psi_n(\boldsymbol{k}_\parallel)|H|\Psi'_n(\boldsymbol{k}_\parallel) >, \\ G_{nn'}(\omega, \boldsymbol{k}_\parallel) &=< \Psi_n(\boldsymbol{k}_\parallel)|G(\omega)|\Psi'_n(\boldsymbol{k}_\parallel) > \end{aligned} \tag{A.4}$$

and $\mathbb{1}$ (the unit matrix) are of rank $l \times m$. In Eq.(A.3) we have simplified the problem by using an ideal surface, i.e. $H_{00} = H_{11} = \ldots$ and $H_{01} = H_{12} = \ldots$ We can now discuss the method of effective layers.

From the general term in Eq.(A.3), one has

$$G_{0n}(\omega) = (\omega - H_{00})^{-1}(H_{01}^{\dagger} G_{n-1,0} + H_{01}(\omega - H_{00})^{-1} H_{01} G_{20}, \tag{A.5}$$

Put $n = 1$ into this equation and put the result into the first equation of the chain (A.3). This yields

$$[\omega - H_{00} - H_{01}(\omega - H_{00})^{-1} H_{01}^{\dagger}] G_{00} = I + H_{01}(\omega - H_{00})^{-1} H_{01} G_{20}, \tag{A.6}$$

which relates G_{00}, to G_{20}. Similarly, if we consider the general equation of the chain, Eq.(A.5) and replace $G_{n-1,0}$ and $G_{n+1,0}$, after Eq.(A.3), we obtain

$$\begin{aligned} &[\omega - H_{00} - H_{01}(\omega - H_{00})^{-1} H_{01}^{\dagger} - H_{01}^{\dagger}(\omega - H_{00})^{-1} H_{01}] G_{n0} \\ &= H_{01}^{\dagger}(\omega - H_{00})^{-1}) H_{01}^{\dagger} G_{n-2,0} + H_{01}(\omega - H_{00})^{-1} H_{01} G_{n+2,0} \qquad (n \geqslant 2). \end{aligned}$$

Nearest neighbors have disappeared in Eqs. (A.6) and (A.7). These equations can be rearranged more compactly as

$$\begin{aligned} &(\omega - \varepsilon_{1s}) G_{00} = I + \alpha_1 G_{20} \\ &(\omega - \varepsilon_1) G_{n0} = \beta_1 G_{n-2,0} + \alpha_1 G_{n+2,0} \qquad (n \geqslant 2) \\ &(\omega - \varepsilon_1) G_{nn} = I + \beta_1 G_{n-2,n} + \alpha_1 G_{n+2,n} \end{aligned} \tag{A.7}$$

with

$$\begin{aligned} \alpha_1 &= H_{01}(\omega - H_{00})^{-1} H_{01} \\ \beta_1 &= H_{01}^{\dagger}(\omega - H_{00})^{-1} H_{01}^{\dagger} \\ \varepsilon_{1s} &= H_{00} + H_{01}(\omega - H_{00})^{-1} H_{01}^{\dagger} \\ \varepsilon_1 &= H_{00} + H_{01}(\omega - H_{00})^{-1} H_{01}^{\dagger} + H_{01}^{\dagger}(\omega - H_{00})^{-1} H_{01}. \end{aligned} \tag{A.8}$$

We then consider the subset formed by taking only even n values in Eq.(A.7), i.e.,

$$\begin{aligned}
(\omega - \varepsilon_{1s})G_{00} &= I + \alpha_1 G_{20} \\
(\omega - \varepsilon_1)G_{2n,0} &= \beta_1 G_{2(n-1),0} + \alpha_1 G_{2(n+1),0} \\
(\omega - \varepsilon_1)G_{2n,2n} &= I + \beta_1 G_{2(n-1),2n} + \alpha_1 G_{2(n+1),2n}.
\end{aligned} \tag{A.9}$$

These equations define a chain which couples the Green's function matrix elements with even indices only, $G_{2n,0}$, via effective nearest-neighbor interactions given by the first two Eqs. of (A.8), and with effective zeroth-order matrix elements that are different for both the surface (ε_{1s}) and the inner layers (ε_1). Equations (A.8) define an effective Hamiltonian describing a chain of effective layers of lattice constant $2a$, twice the original one. Each effective layer contains implicitly the effect of its nearest neighbors in the original chain through the use of equation (A.5).

Except for the different zeroth-order matrix elements, $\varepsilon_{1s} \neq \varepsilon_1$, equations (A.9) are similar to equations (A.3). So equations from (A.3) to (A.9) can be repeated if we start from (A.9). By repeating the argument i times, we have the iterative sequence

$$\begin{aligned}
\alpha_i &= \alpha_{i-1}(\omega - \varepsilon_{i-1})^{-1}\alpha_{i-1} \\
\beta_i &= \beta_{i-1}(\omega - \varepsilon_{i-1})^{-1}\beta_{i-1} \\
\varepsilon_i &= \varepsilon_{i-1} + \alpha_{i-1}(\omega - \varepsilon_{i-1})^{-1}\beta_{i-1} + \beta_{i-1}(\omega - \epsilon_{i-1})^{-1}\alpha_{i-1} \\
\varepsilon_i^s &= \varepsilon_{i-1}^s + \alpha_{i-1}(\omega - \varepsilon_{i-1})^{-1}\beta_{i-1},
\end{aligned} \tag{A.10}$$

starting with $\varepsilon_0 = H_{00}$, $\alpha_0 = H_{01}$ and $\beta_0 = H_{01}^{\dagger}$. Eqs. (A.10) define an effective Hamiltonian for a chain of lattice constant $2^i a$ with nearest-neighbor couplings α_i and β_i and zeroth-order

Hamiltonian matrix elements ε_i and ε_i^s. After i iterations

$$\begin{aligned}(\omega-\varepsilon_i^s)G_{00} &= I+\alpha_i G_{2^i n,0}\\ (\omega-\varepsilon_i)G_{2^i n,0} &= \beta_i G_{2^i(n-1),0}+\alpha_i G_{2^i(n+1),0} \qquad (n\geqslant 1).\end{aligned} \tag{A.11}$$

Each layer of the ith chain contains the effect of the nearest neighbors of the previous chain $(i-1)$ implicitly. After ν iterations, the zeroth layer is equivalent to the original zeroth layer coupled to 2^ν layers, while any inner layer has $2^{n+1}-1$ layers of the original chain. The iteration is to be repeated until α_ν, and β_ν are small enough. Then we have $\varepsilon_\nu \approx \varepsilon_{\nu-1}, \varepsilon_\nu^s \approx \varepsilon_{\nu-1}^s$ and

$$\begin{aligned}(\omega-\varepsilon_\nu^s)G_{00} &\approx I\\ (\omega-\varepsilon_\nu)G_{2^\nu n,2^\nu n} &\approx I \qquad (n\geqslant 1).\end{aligned} \tag{A.12}$$

as well as an good approximation for G_{00},

$$G_{00}(\omega)\approx(\omega-\varepsilon_\nu^s)^{-1}. \tag{A.13}$$

And the Green's function for the edge layer, i.e. for the zeroth layer of the complementary chain, can be calculated by exchanging α_i and β_i:

$$\bar{G}_{00}(\omega)\approx(\omega-\bar{\varepsilon}_\nu^s)^{-1} \tag{A.14}$$

where $\bar{\varepsilon}_\nu^s$ is obtained by iterating

$$\bar{\varepsilon}_{i+1}^s=\bar{\varepsilon}_{i+1}^s+\beta_i(\omega-\varepsilon_i)^{-1}\alpha_i, \tag{A.15}$$

starting with $\varepsilon_0=\bar{\varepsilon}_0^s=H_{00}, \alpha_0=H_{01}$ and $\beta_0=H_{01}^\dagger$, as before, until $\varepsilon_\nu^s\approx\bar{\varepsilon}_{\nu-1}^s$. The terms G_{00} and $\bar{G}_{00}$ represents the surface Green's functions of g_R and g_L respectively. It is

worth pointing out that the iterative procedure presented above is exact in the sense that no interactions are ignored in a nearest-neighbor chain.

APPENDIX B

RECURSIVE GREEN'S FUNCTION TECHNIQUE

We now deal with the Green's functions of the central device appearing in Eq. (2.59), which will be solved very efficiently by the recursive Green's function (RGF) method. We begin by introducing two equivalent Dyson formulas for calculating an exact Green's function (for a derivation of these formulas, see [36]),

$$G = G^{(0)} + G^{(0)}VG, \tag{B.1}$$

$$G = G^{(0)} + GVG^{(0)}, \tag{B.2}$$

where $G^{(0)}$ represents the "unperturbed" Green's function and V the perturbation. We use these expressions to obtain recursive relations for the exact Green's function of a quasi-one-dimensional system coupled with the leads. The basic idea is to partition the system into independent regions and associate these regions with "unperturbed" Green's functions $G^{(0)}$. The hoping matrix elements connecting those parts are then partially built into the perturbation V from our selections. By choosing the connecting matrix elements and applying Eqs. (B.1) and (B.2) carefully, we can slice by slice obtain the full Green's function G.

Our presentation is specialized to the case of two-probe conductance as depicted in Fig. 2.4. It can also be conveniently extended to multi-probe systems. It is necessary to first derive several intermediate results before obtaining expressions for the exact Green's function. We first run the recurrence from left to right, resulting in a series of Green's functions G^L. At every step, Eq. (B.1) is employed using a different choice for G^0 and V. We then repeat the procedure from right to left, generating another series of functions G^R. Finally, we join these two families to obtain the exact Green's functions for the whole system. By doing this, we used all parts and connecting matrix elements exactly once.

The system is split into N thin slices, each one with a maximum of M sites or cells, as

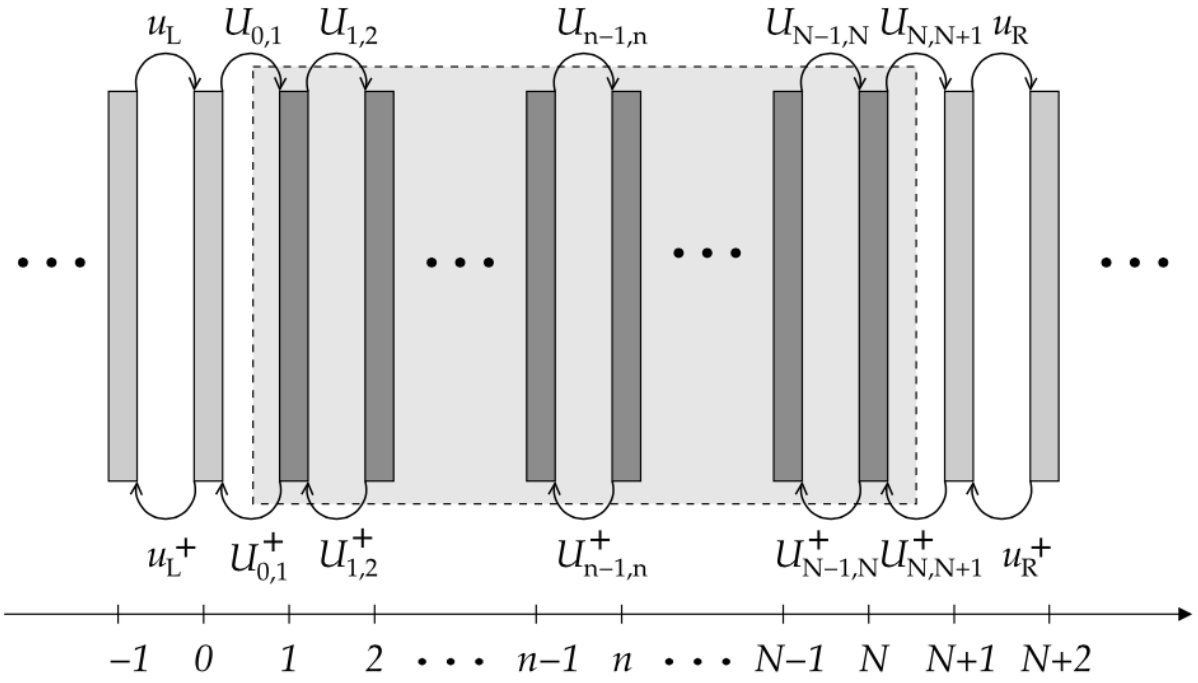


Figure B.1: Slicing scheme. The central rectangle containing the dark strips (slices) represents the conductor (taken from Ref. [73]).

show in Fig. B.1. The slices with numbers lower than 1 or larger than N represent the left and right lead regions, respectively. The corresponding retarded surface Green's functions (when the leads are decoupled from the system) are denoted by $g_L(E)$ and $g_R(E)$, as noted earlier. These Green's functions are computed separately and before the recurrence procedure. The retarded Green's function of the isolated nth slice in the system, $g_n(E) = (E - h_n + i0^+)^{-1}$, does not need to be individually evaluated before the recursive calculations. Here, h_n denotes the Hamiltonian of the isolated nth slice.

Neighboring slices within the sample are connected to each other through the matrices $U_{n-1,n}$ (left to right) and $[U_{n-1,n}]^\dagger \equiv U_{n,n-1}$ (right to left), with $n = 1, \ldots, N$. The first and last slices in the system are connected to their nearest neighboring slices in the leads through the coupling matrices $U_{0,1}$ and $U_{N,N+1}$. The matrix elements of these matrices are the tight-binding hopping amplitudes connecting sites at different slices.

Here, we assumed that the matrices U only connect nearest-neighbor slices. For tight-binding models that include next-nearest hopping terms, one can still use this algorithm by doubling the "width" of the unit slices, which slows down the computation by a factor 2^3. We use subscripts to denote longitudinal spatial indices (except for g_L, g_R, and g_n). Thus, $G_{n,m}(E)$ is the matrix Green's function connecting the n and m slices. Sites indices are shown as a pair of variables: $G_{n,m}(j, j')$ denotes the Green's function connecting site j in the nth slice to site j' in the mth slice. From now on, we will drop the energy variable E (since scattering is assumed elastic, E is conserved throughout the system).

Coupling with leads For the two-probe system considered, the central region is coupled to a left lead L and to a right lead R. We now show how to built the Green's function that describes this coupling.

The first step is to incorporate the $n = 1$ slice to the left contact surface Green's function (A.14), which we write as g_L. We also introduce the kets $|0\rangle$ and $|1\rangle$ representing the states where electrons are found in slices $n = 0$ and $n = 1$ respectively. The "unperturbed" Green's

function in this case is $G^{(0)} = |0\rangle\, g_L\, \langle 0| + |1\rangle\, g_1\, \langle 1|$, while $V = |0\rangle\, U_{0,1}\, \langle 1| + |1\rangle\, U_{1,0}\, \langle 0|$ is the perturbation that couples the $n = 1$ slice to the left lead. Then, using Eq. (B.1), we have

$$\begin{aligned}\langle 1|G^L|1\rangle &= \langle 1|G^{(0)}|1\rangle + \sum_{m,m'} \langle 1|G^{(0)}|m\rangle\, \langle m|V|m'\rangle\, \langle m'|G^L|1\rangle \\ &= \langle 1|G^{(0)}|1\rangle + \langle 1|G^{(0)}|1\rangle\, \langle 1|V|0\rangle\, \langle 0|G^L|1\rangle\end{aligned}$$

and

$$\begin{aligned}\langle 0|G^L|1\rangle &= \langle 0|G^{(0)}|1\rangle + \sum_{m,m'} \langle 0|G^{(0)}|m\rangle\, \langle m|V|m'\rangle\, \langle m|G^L|1\rangle \\ &= \langle 0|G^{(0)}|0\rangle\, \langle 0|V|1\rangle\, \langle 1|G^L|1\rangle\,.\end{aligned}$$

Using the more compact notation $\langle n|G^L|m\rangle = G^L_{n,m}$, we drop the bras and kets and can rewrite these equation as

$$G^L_{1,1} = g_1 + g_1 U_{1,0} G^L_{0,1} \tag{B.3}$$

and

$$G^L_{0,1} = g_L U_{0,1} G^L_{1,1} \tag{B.4}$$

Therefore,

$$G^L_{1,1} = (I - g_1 U_{1,0} g_L U_{0,1})^{-1} g_1. \tag{B.5}$$

Now, since $g_1 = (E - h_1)^{-1}$, we can write

$$G^L_{1,1} = (E - h_1 - U_{1,0} g_L U_{0,1})^{-1}. \tag{B.6}$$

Notice that this Green's function takes care the coupling of the first slice with the left lead, but does not have information on the rest of the system or the right lead.

It is important to mention that we have safely neglected the infinitesimal imaginary part, since the self-energy term has its own finite imaginary part.

We proceed in a similar fashion in order to couple the last slice to the right lead. With $G^{(0)} = g_R + g_N$ and $V = U_{N,N+1}$, we have

$$G^R_{N,N} = g_N + g_N U_{N,N+1} G^R_{N+1,N} \tag{B.7}$$

and

$$G^R_{N+1,N} = g_R U_{N+1,N} G^R_{N,N}. \tag{B.8}$$

Therefore,

$$G^R_{N,N} = (I - g_N U_{N,N+1} g_R U_{N+1,N})^{-1} g_N. \tag{B.9}$$

Again, since $g_N = (E - h_N)^{-1}$, we can write

$$G^R_{N,N} = (E - h_N - U_{N,N+1} g_R U_{N+1,N})^{-1}. \tag{B.10}$$

The Green's function $G^L_{1,1}$(or$G^R_{N,N}$) describes all single-electron processes that begin and end that on the $n = 1$ (or $n = N$) slice, taking into consideration all possible number of incursions in and out of the left (or right) lead. It does not yet take into consideration incursions into the bulk of the system.

Left Green's functions With $G^L_{1,1}$, we can examine the next successive $N-1$ left Green's functions by using a recurrence formula similar to Eq. (B.6). To derive this, we choose $G^{(0)} = G^L_{n-1,n-1}$ and $V = U_{n-1,n} + U_{n,n-1}$. Using the Dyson's equation (B.1), we obtain

$$G^L_{n,n} = (I - g_n U_{n,n-1} G^L_{n-1,n-1} U_{n-1,n})^{-1} g_n, \tag{B.11}$$

with $n = 2, \ldots, N$. Using $g_n = (E - h_n)^{-1}$, we obtain

$$G^L_{n,n} = (E - h_n - U_{n,n-1} G^L_{n-1,n-1} U_{n-1,n})^{-1}. \tag{B.12}$$

This formula is accompanied by another, which connects the left-most slice (the surface slice of the left lead) with the nth one,

$$G^L_{0,n} = G^L_{0,n-1} U_{n-1,n} G^L_{n,n}. \tag{B.13}$$

Note that N inversions are necessary to reach the Nth slice. Each inversion requires $O(M^3)$ operations. Thus, the time complexity of the calculation scales as $O(NM^3)$.

Right Green's functions The right Green's functions are similar to the left ones. Using Eq. (B.10) and starting from the Nth slice, we find that

$$G^R_{n,n} = (I - g_n U_{n,n+1} G^R_{n+1,n+1} U_{n+1,n})^{-1} g_n, \tag{B.14}$$

with $n = N-1, \ldots, 1$. Substituting $g_n = (E - h_n)^{-1}$, we obtain

$$G^R_{n,n} = (E - h_n - U_{n,n+1} G^R_{n+1,n+1} U_{n+1,n})^{-1}. \tag{B.15}$$

Also,

$$G^R_{N+1,n} = G^R_{N+1,n+1} U_{n+1,n} G^R_{n,n}. \tag{B.16}$$

Again, N additional inversions have to be done in order to reach slice the first slice ($n = 1$), with time complexity of $O(NM^3)$.

Full Green's functions Assuming one reaches the n slice by either a left or right sweep ($1 < n < N$), in order to obtain the exact full Green's function of the system, we again use Eq. (B.1) assuming $G^{(0)} = g_n + G^L_{n-1,n-1} + G^R_{n+1,n+1}$, with $V = U_{n-1,n} + U_{n,n-1} + U_{n,n+1} + U_{n+1,n}$. We find

$$G_{n,n} = g_n + g_n(U_{n,n-1} G_{n-1,n} + U_{n,n+1} G_{n+1,n}, \tag{B.17}$$

$$G_{n-1,n} = G^L_{n-1,n-1} U_{n01,n} G_{n,n}, \tag{B.18}$$

and

$$G_{n+1,n} = G_R n+1, n+1 U_{n+1,n} G_{n,n}. \tag{B.19}$$

Thus,

$$G_{n,n} = [I - g_n(U_{n,n-1} G^L_{n-1,n-1} U_{n-1,n} + U_{n,n+1} G^R_{n+1,n+1} U_{n+1,n})]^{-1} g_n, \tag{B.20}$$

and since $g_n = (E - h_n)^(- 1)$, we obtain

$$G_{n,n} = (E - h_n - U_{n,n-1} G^L_{n-1,n-1} U_{n-1,n} - U_{n,n+1} G^R_{n+1,n+1} U_{n+1,n})^{-1}, \tag{B.21}$$

together with

$$G_{0,n} = G^L_{0,n-1} U_{n-1,n} G_{n,n} \tag{B.22}$$

and

$$G_{N+1,n} = G^R_{N+1,n+1} U_{n+1,n} G_{n,n}. \tag{B.23}$$

Note that in order to compute $G_{n,n}$ and $G_{N+1,n}$, we need to keep track of $G^L_{n,n}$ and $G^R_{n,n}$ (obtained recursively from Eqs. (B.12) and (B.15), respectively), as well as $G^L_{0,n}$ and $G^R_{N+1,n}$ (which follow from Eqs. (B.13) and (B.16), respectively). In order to obtain $G_{n-1,n}$ and $G_{n,n+1}$, we can apply Dyson's equation again to a situation where only the nth slice is decoupled, yielding

$$G_{n,n+1} = G_{n,n} U_{n,n+1} G^R_{n+1,n+1}, \tag{B.24}$$

while

$$G_{n-1,n} = G^L_{n-1,n-1} U_{n-1,n} G_{n,n}. \tag{B.25}$$

These equations are useful for calculating the local current distribution.

When computing the exact Green's, we note that we have used each matrix $U_{n,n'}$ only once. Similarly, at each step, an isolated Hamiltonian h_n was also used once. Thus, at the end of the calculation, all hoping parameters and local potentials of the underlying tight-binding model have been used exactly once.

An alternative way to compute full Green's functions is to close the left (or right) sweep with a connection to the right (left) lead: (This is useful if only transmission and reflection matrices are required.) 1. Left sweep, we use Eq. (B.22) to write

$$G_{0,N+1} = G^L_{0,N} U_{N+1,N} G_{N+1,N+1}, \tag{B.26}$$

which is complemented by

$$G_{N+1,N+1} = (g_R^{-1} - U_{N+1,N} G^L_{N,N} U_{N,N+1})^{-1} \tag{B.27}$$

obtained from Eq. (B.21).

2. Right sweep, we use Eqs. (B.23 and (B.22) to obtain

$$G_{N+1,0} = G^R_{N+1,1} U_{1,0} G_{0,0} \tag{B.28}$$

and

$$G_{0,0} = (g_L^{-1} - U_{0,1} G^R_{1,1} U_{1,0})^{-1}, \tag{B.29}$$

respectively. Eqs. (B.26) and (B.29) can be used to calculate the left-to-right transmission and left reflection matrices respectively, while Eqs. (B.28) and (B.27) lead to the right-to-left transmission and the right reflection matrices. For systems with inversion symmetry, we expect $G_{00} = G_{N+1,N+1}$ and $G_{0,N+1} = G_{N+1,0}$ and therefore only one sweep (left or right)

would be necessary for evaluating the whole scattering matrix. For symmetric leads,

$$[G^R_{0,N+1}]^\dagger = G^A_{N+1,0} \tag{B.30}$$

and also only one sweep is necessary. Any local observable (such as the local density of states or the local current flux), requires $G_{0,N+1}$ as well as $G_{n,n}$ for all $n = 1, \ldots, N$.

The recurrence relations shown above need input information. Specifically, one needs to define the Green's functions of the leads (g_L and g_R), the Hamiltonian of the isolated slices ($h_n, n = 1, \ldots, N$) in the device region, and the hopping between slices (the U matrices) between the two regions, before the calculation of the Green's functions.

APPENDIX C

QUASI-1D POISSON SOLVER

Restate our task in this section: assuming the electron density distribution n is known throughout the entire device, we solve Poisson's equation for electric potential profile ϕ:

$$\nabla^2 V(x) = -\frac{1}{\epsilon}(-n + N_D^+ N_A^-) \tag{C.1}$$

where

Poisson's equation is a simple second-order partial differential equation (PDE), but in this case what complicates the solution are the 2-dimensional nature of the problem and spatially varying material composition ϵ. One common method to solve PDE in this case is via discretization methods and seeking solutions on a grid.

In this section, we present our simple Poisson solver and how it is coupled with transport equations. We start with the device grid and proper boundary conditions used. We point out Poisson's equation is to be solved self-consistently with transport equations, and in such case more efficient algorithms can be deployed. By introducing the use of quasi-Fermi levels as a bridge, transport equations and Poisson's equation obtain a damping factor in the self-consistent loop. We look into the details of this coupled scheme and its implementation. In the end, the Poisson's equation are solved using Newton-Raphson iterating technique.

To solve a PDE, boundary conditions have to be properly set. Boundary conditions are grid nodes on edges, so we have to make sure not only the grid completely covers the region of our simulation interest, but also it terminates at places where boundary condition is known to us.

At the metal-semiconductor contact interface we use the Neumann boundary condition:

$$\frac{dV}{dr_\perp} = 0, \tag{C.2}$$

with $r_\perp$ being the direction perpendicular to the interface. Zero field here allows the potential to float to whatever value it needs to maintain charge neutrality. Although the source/drain extension regions are heavily doped semiconductors, they are not perfect metals and in reality can still have electric field within them under high drain bias. The Neumann boundary condition forces the boundary and only the boundary to assume the role of a metal by having no electric field, thus eliminating drifting current at that point. Under non-equilibrium conditions, electrons are supplied from source solely by diffusive current, which the diffusion constant is that of semiconductor body.

In Gummel's scheme, equations are solved iteratively starting from an initial guess. Each iterative loop takes previous solution as input and solves for a new set of solution. If the new and old solutions become very similar to each other, we declare such solution as converged. However, if the new and old does not get close to each other, or in worst case depart from each other, we call such unwanted situation divergence.

The transport equation is solved in one step for exact solution, so the burden of damping is placed at Poisson's equation. In order to archive stable convergence in Poisson's equation, we choose to exploit the concept of quasi-Fermi level due to its exponential form. The reason for this will become clear shortly.

The quasi-Fermi level for electrons is defined as

$$F = -qV + k_B T \mathcal{F}_{1/2}^{-1}(\frac{n}{N_C}) \tag{C.3}$$

In terms of quasi-Fermi level, the electron density thus becomes

$$n = N_C \mathcal{F}_{1/2}^{-1}(\frac{F + qV}{k_B T}) \tag{C.4}$$

In non-degenerate limit, the Fermi-Dirac integral can be simplified

$$n = N_C \exp(\frac{F + qV}{k_B T}) \tag{C.5}$$

Transport equation uses electrostatic potential of previous iterative loop to solve for electron density, and we denote this potential V_{old}. Poisson's equation uses the electron density from transport equation to solve for a new electrostatic potential, and we denote this potential V_{new}. Substitute the quasi-Fermi level obtained from transport equation into the electron density term in Poisson's equation, we get

$$n = N_C \exp(\frac{q(V_{new} V_{old})}{k_B T}) \tag{C.6}$$

The above equation only holds true if V_{new} and V_{old} are equal, and that is when the solution converges and become exact. Since V_{new} is to be determined by the Poisson's equation, it gives an extra degree of freedom to the solution.

Now, the Poisson's equation has became non-linear, and to solve it we need to use Newton-Raphson method. Newton-Raphson method is a simple and powerful iterative technique to determine the solution of an equation. After we discretize the Poisson equation, the system of equation contain N unknowns. We need to find its derivative, then the tangent line interpolating from one guess to another, and repeat until solution converges. In our Poisson system of equations, this derivative term has become a full matrix called Jacobian matrix. Jacobian is a term familiar to vector calculus denoting the matrix of all first-order partial derivatives of a vector-valued function. The Jacobian matrix can then be formed by taking derivative with respect to all other nodes. Notice each equation centered at a specific node contains five unique variables, so the resulting Jacobian matrix contains five diagonal terms. The matrix is very sparse, so we can take advantage of this to save computational memory. We denote the Jacobian matrix element containing the derivative of node α with respect to

node β as

$$f_{\alpha,\beta} = \frac{\partial f_\alpha}{\partial V_\beta} \tag{C.7}$$

Thus, we obtained the Newton-Raphson iterative equation with Taylor's expansion correction to first order

$$f_\alpha(V_{new}) \approx f_\alpha(V_{old}) + f_{\alpha,\beta}(V_{old})\Delta V_\beta = 0 \tag{C.8}$$

In case of divergence, we can use the strategy suggested by Brown and Lindsay [74] to ensure stable convergence. Usually, such divergent situation only occurs at the beginning of the simulation, when the starting several guesses can be far away from true solution. Once guess becomes near the true solution, it is the nature of Newton's method that a fast and stable convergence can be achieved.

APPENDIX D

DERIVATION OF THOMAS-FERMI APPROXIMATION FOR QUASI-1D SYSTEMS

The derivation begins with the equation for the screened potential energy from an impurity charge distribution $\lambda_i(x)$

$$\nabla^2 V(x) = 4\pi e \left[\lambda_i(x) + \lambda_s(x)\right] A \tag{D.1}$$

where $\lambda_s(x)$ is the screening charge and A is the cross section area. We assume $A = 1$ in the following. The Thomas-Fermi theory approximates the local electron density $n(x)$ as a free-particle system

$$n(x) = \frac{k_F(x)}{\pi} \tag{D.2}$$

where the Fermi wave vector k_F is now a local quantity. It can also be determined by the condition that the chemical potential μ is independent of position:

$$\frac{k_F^2(x)}{2m} = E_F(x) = \mu V(x) \tag{D.3}$$

We write the screening charge as the difference between $n(x)$ and the equilibrium charge density n_0

$$\lambda_s(x) = -e\left[n(x) n_0(x)\right] \tag{D.4}$$

and the above approximations result the equation

$$\nabla^2 V(x) = 4\pi e \left[\lambda_i(x) + e n_0 e n_0 \sqrt{1 \frac{V(x)}{E_F}}\right] \tag{D.5}$$

Assuming $V/E_F << 1$, we can expand the root as $\sqrt{1V(x)/E_F} \approx 1V/2E_F$ to obtain the equation

$$(\nabla^2 q_{TF}^2) V(x) = 4\pi e \lambda_i(x) \tag{D.6}$$

$$q_{TF}^2 = \frac{2\pi e^2 n_0}{E_F} \tag{D.7}$$

where q_{TF} is the Thomas-Fermi screening wave vector. This equation may be solved in 1D Fourier transform space to give

$$V(x) = -4\pi e \int \frac{dq}{2\pi} \frac{\lambda_i(q)}{q^2 + q_{TF}^2} e^{iqx} \tag{D.8}$$

In the case of edge contact, the interfacial trap charge acts as the impurity residing around $x = 0$, and can be represented as $\lambda_i(q) = Q_i$. We can then evaluate the integral and obtain an analytical result

$$V(x) = -\frac{2\pi e Q}{q_{TF}} e^{-q_{TF}|x|} \tag{D.9}$$

The interactions declines rapidly at large distances because of the exponential dependence $e^{-q_{TF}x}$ and thus allows efficient electron tunneling through the interface.

REFERENCES

[1] AH Castro Neto, Francisco Guinea, Nuno MR Peres, Kostya S Novoselov, and Andre K Geim. The electronic properties of graphene. *Reviews of modern physics*, 81(1):109, 2009.

[2] KS Novoselov, A Mishchenko, A Carvalho, and AH Castro Neto. 2d materials and van der waals heterostructures. *Science*, 353(6298):aac9439, 2016.

[3] Mark P Levendorf, Cheol-Joo Kim, Lola Brown, Pinshane Y Huang, Robin W Havener, David A Muller, and Jiwoong Park. Graphene and boron nitride lateral heterostructures for atomically thin circuitry. *Nature*, 488(7413):627, 2012.

[4] Gianluca Fiori, Francesco Bonaccorso, Giuseppe Iannaccone, Tomás Palacios, Daniel Neumaier, Alan Seabaugh, Sanjay K Banerjee, and Luigi Colombo. Electronics based on two-dimensional materials. *Nature nanotechnology*, 9(10):768, 2014.

[5] Xidong Duan, Chen Wang, Anlian Pan, Ruqin Yu, and Xiangfeng Duan. Two-dimensional transition metal dichalcogenides as atomically thin semiconductors: opportunities and challenges. *Chemical Society Reviews*, 44(24):8859–8876, 2015.

[6] Adrien Allain, Jiahao Kang, Kaustav Banerjee, and Andras Kis. Electrical contacts to two-dimensional semiconductors. *Nature Materials*, 14(12):1195, 2015.

[7] Yang Xu, Cheng Cheng, Sichao Du, Jianyi Yang, Bin Yu, Jack Luo, Wenyan Yin, Erping Li, Shurong Dong, Peide Ye, and Xiangfeng Duan. Contacts between two-and three-dimensional materials: ohmic, schottky, and p–n heterojunctions. *ACS nano*, 10 (5):4895–4919, 2016.

[8] Daniel S Schulman, Andrew J Arnold, and Saptarshi Das. Contact engineering for 2d materials and devices. *Chemical Society Reviews*, 47(9):3037–3058, 2018.

[9] Song-Lin Li, Katsuyoshi Komatsu, Shu Nakaharai, Yen-Fu Lin, Mahito Yamamoto, Xiangfeng Duan, and Kazuhito Tsukagoshi. Thickness scaling effect on interfacial barrier and electrical contact to two-dimensional mos2 layers. *ACS nano*, 8(12):12836–12842, 2014.

[10] Wei Liu, Deblina Sarkar, Jiahao Kang, Wei Cao, and Kaustav Banerjee. Impact of contact on the operation and performance of back-gated monolayer mos2 field-effect-transistors. *ACS nano*, 9(8):7904–7912, 2015.

[11] Yuan Liu, Jian Guo, Enbo Zhu, Lei Liao, Sung-Joon Lee, Mengning Ding, Imran Shakir, Vincent Gambin, Yu Huang, and Xiangfeng Duan. Approaching the schottky–mott limit in van der waals metal–semiconductor junctions. *Nature*, 557:1, 2018.

[12] Saptarshi Das, Hong-Yan Chen, Ashish Verma Penumatcha, and Joerg Appenzeller. High performance multilayer mos2 transistors with scandium contacts. *Nano letters*, 13 (1):100–105, 2012.

[13] Saptarshi Das, Richard Gulotty, Anirudha V Sumant, and Andreas Roelofs. All two-dimensional, flexible, transparent, and thinnest thin film transistor. *Nano letters*, 14 (5):2861–2866, 2014.

[14] Yuan Liu, Hao Wu, Hung-Chieh Cheng, Sen Yang, Enbo Zhu, Qiyuan He, Mengning Ding, Dehui Li, Jian Guo, Nathan O Weiss, Yu Huang, and Xiangfeng Duan. Toward barrier free contact to molybdenum disulfide using graphene electrodes. *Nano letters*, 15(5):3030–3034, 2015.

[15] Jiahao Kang, Wei Liu, Deblina Sarkar, Debdeep Jena, and Kaustav Banerjee. Computational study of metal contacts to monolayer transition-metal dichalcogenide semiconductors. *Physical Review X*, 4(3):031005, 2014.

[16] Xi Ling, Yuxuan Lin, Qiong Ma, Ziqiang Wang, Yi Song, Lili Yu, Shengxi Huang, Wenjing Fang, Xu Zhang, Allen L Hsu, YiHsien Lee, Yimei Zhu, Lijun Wu, Ju Li, Pablo JarilloHerrero, Mildred Dresselhaus, Tomás Palacios, and Jing Kong. Parallel stitching of 2d materials. *Advanced Materials*, 28(12):2322–2329, 2016.

[17] Wushi Dong and Peter B. Littlewood. swan: An open-source c++ software for efficient nanoscale quantum transport simulations, 2019. URL `https://doi.org/10.5281/zenodo.2553787`.

[18] Gordon E Moore et al. Cramming more components onto integrated circuits, 1965.

[19] Kostya S Novoselov, Andre K Geim, Sergei V Morozov, D Jiang, Y_ Zhang, Sergey V Dubonos, Irina V Grigorieva, and Alexandr A Firsov. Electric field effect in atomically thin carbon films. *science*, 306(5696):666–669, 2004.

[20] Han Liu, Adam T Neal, and Peide D Ye. Channel length scaling of mos2 mosfets. *ACS nano*, 6(10):8563–8569, 2012.

[21] Sujay B Desai, Surabhi R Madhvapathy, Angada B Sachid, Juan Pablo Llinas, Qingxiao Wang, Geun Ho Ahn, Gregory Pitner, Moon J Kim, Jeffrey Bokor, Chenming Hu, et al. Mos2 transistors with 1-nanometer gate lengths. *Science*, 354(6308):99–102, 2016.

[22] Leonid V Keldysh et al. Diagram technique for nonequilibrium processes. *Sov. Phys. JETP*, 20(4):1018–1026, 1965.

[23] Leo P Kadanoff. *Quantum statistical mechanics*. CRC Press, 2018.

[24] Nicola Marzari, Arash A Mostofi, Jonathan R Yates, Ivo Souza, and David Vanderbilt. Maximally localized wannier functions: Theory and applications. *Reviews of Modern Physics*, 84(4):1419, 2012.

[25] Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965.

[26] Richard M Martin. *Electronic structure: basic theory and practical methods*. Cambridge university press, 2004.

[27] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136 (3B):B864, 1964.

[28] John C Slater and George F Koster. Simplified lcao method for the periodic potential problem. *Physical Review*, 94(6):1498, 1954.

[29] José M Soler, Emilio Artacho, Julian D Gale, Alberto García, Javier Junquera, Pablo Ordejón, and Daniel Sánchez-Portal. The siesta method for ab initio order-n materials simulation. *Journal of Physics: Condensed Matter*, 14(11):2745, 2002.

[30] DA Papaconstantopoulos and MJ Mehl. The slater–koster tight-binding method: a computationally efficient and accurate approach. *Journal of Physics: Condensed Matter*, 15(10):R413, 2003.

[31] Arrigo Calzolari, Nicola Marzari, Ivo Souza, and Marco Buongiorno Nardelli. Ab initio transport properties of nanostructures from maximally localized wannier functions. *Phys. Rev. B*, 69(035108), 2004.

[32] N. Marzari and D. Vanderbilt. Maximally localized generalized wannier functions for composite energy bands. *Phys. Rev. B*, 56(12847), 1997.

[33] E.Blount. Formalisms of band theory. *Solid State Physics*, 13(305), 1962.

[34] Ivo Souza, Nicola Marzari, and David Vanderbilt. Maximally localized generalized wannier functions for entangled energy bands. *Phys. Rev. B*, 65(035109), 2001.

[35] Arash A. Mostofi, Jonathan R. Yates, Young-Su Lee, Ivo Souza, David Vanderbilt, and Nicola Marzari. wannier90: A tool for obtaining maximally-localised wannier functions. *Computer Physics Communications*, 178:685–699, 2008.

[36] S. Datta. *Electronic transport in mesoscopic systems*. Cambridge University Press, Cambridge, 1995.

[37] D. K. Ferry and S. M. Goodnick. *Transport in Nanostructures*. Cambridge University Press, Cambridge, 1997.

[38] Samantha Bruzzone, Demetrio Logoteta, Gianluca Fiori, and Giuseppe Iannaccone. Vertical transport in graphene-hexagonal boron nitride heterostructure devices. *Scientific Reports*, 5(14519), 2015.

[39] G. Iannaccone, Q. Zhang, S. Bruzzone, and G. Fiori. Insights on the physics and application of off-plane quantum transport through graphene and 2d materials. *Solid-State Electronics*, 115:213–218, 2016.

[40] M. P. Lopez Sancho, J. M. Lopez Sancho, and J. Rubio. Highly convergent schemes for the calculation of bulk and surface green functions. *Journal of Physics F: Metal Physics*, 15(851), 1985.

[41] D.J. Thouless and S. Kirkpatrick. Conductivity of the disordered linear chain. *J. Phys. C*, 14(235), 1981.

[42] S. Datta. Nanoscale device modeling: the green's function method. *Superlattices and Microstructures*, 28(4), 2000.

[43] D. H. Lee and J. D. Joannopoulos. Simple scheme for surface-band calculations. i. *Phys. Rev. B*, 23(4988), 1981.

[44] D. H. Lee and J. D. Joannopoulos. Simple scheme for surface-band calculations. ii. the green's function. *Phys. Rev. B*, 23(4997), 1981.

[45] R. Landauer. Electrical resistance of disordered one-dimensional lattices. *Phil. Mag.*, 21(863), 1970.

[46] D. S. Fisher and P. A. Lee. Relation between conductivity and transmission matrix. *Physical Review B*, 23(6851), 1981.

[47] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, and et al. Quantum espresso: a modular and open-source software project for quantum simulations of materials. *Journal of physics: Condensed matter*, 21(39):395502, 2009.

[48] Arash A Mostofi, Jonathan R Yates, Young-Su Lee, Ivo Souza, David Vanderbilt, and Nicola Marzari. wannier90: A tool for obtaining maximally-localised wannier functions. *Computer physics communications*, 178(9):685–699, 2008.

[49] et al Paolo Giannozzi. Quantum espresso: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter*, 21(39), 2009. URL `http://www.quantum-espresso.org/`.

[50] Leonard Kleinman and D. M. Bylander. Efficacious form for model pseudopotentials. *Phys. Rev. Lett.*, 48(1425), 1982.

[51] H.J. Monkhorst and J.D. Pack. Special points for brillouin-zone integrations. *Phys. Rev. B*, 13(5188), 1976.

[52] Anton Kokalj. Computer graphics and graphical user interfaces as tools in simulations of matter at the atomic scale. *Computational Materials Science*, 28(2):155–168, 2003.

[53] R. Tamura, Y. Sawai, and J. Haruyama. Suppression of the pseudoantisymmetry channel in the conductance of telescoped double-wall nanotubes. *Physical Review B*, 72 (045413), 2005.

[54] S. Roche, F. Triozon, A. Rubio, and D. Mayou. Conduction mechanisms and magneto-transport in multiwalled carbon nanotubes. *Physical Review B*, 64(121401), 2001.

[55] Wushi Dong and Peter B. Littlewood. Quantum electron transport in ohmic edge contacts between two-dimensional materials. *ACS Applied Electronic Materials*, 1(6): 799–803, Jun 2019. ISSN 2637-6113. doi: 10.1021/acsaelm.9b00095. URL `http://dx.doi.org/10.1021/acsaelm.9b00095`.

[56] M.H. Guimarães, H. Gao, Y. Han, K. Kang, S. Xie, C.J. Kim, D.A. Muller, D.C. Ralph, and J. Park. Atomically thin ohmic edge contacts between two-dimensional materials. *ACS nano*, 10(6):6392–6399, 2016.

[57] Mervin Zhao, Yu Ye, Yimo Han, Yang Xia, Hanyu Zhu, Siqi Wang, Yuan Wang, David A Muller, and Xiang Zhang. Large-scale chemical assembly of atomically thin transistors and circuits. *Nature nanotechnology*, 11(11):954, 2016.

[58] Yi-Hsien Lee, Xin-Quan Zhang, Wenjing Zhang, Mu-Tung Chang, Cheng-Te Lin, Kai-Di Chang, Ya-Chu Yu, Jacob Tse-Wei Wang, Chia-Seng Chang, Lain-Jong Li, and TsungWu Lin. Synthesis of large-area mos2 atomic layers with chemical vapor deposition. *Advanced materials*, 24(17):2320–2325, 2012.

[59] Henry Yu, Alex Kutana, and Boris I Yakobson. Carrier delocalization in two-dimensional coplanar p–n junctions of graphene and metal dichalcogenides. *Nano letters*, 16(8):5032–5036, 2016.

[60] Wei Chen, Yuan Yang, Zhenyu Zhang, and Efthimios Kaxiras. Properties of in-plane graphene/mos2 heterojunctions. *2D Materials*, 4(4):045001, 2017.

[61] Jie Sun, Na Lin, Cheng Tang, Haoyuan Wang, Hao Ren, and Xian Zhao. First principles studies on electronic and transport properties of edge contact graphene-mos2 heterostructure. *Computational Materials Science*, 133:137–144, 2017.

[62] Mathieu Luisier, Andreas Schenk, Wolfgang Fichtner, and Gerhard Klimeck. Atomistic simulation of nanowires in the s p 3 d 5 s* tight-binding formalism: From boundary conditions to strain calculations. *Physical Review B*, 74(20):205323, 2006.

[63] Supriyo Datta. Nanoscale device modeling: the green's function method. *Superlattices and microstructures*, 28(4):253–278, 2000.

[64] Jin-Wu Jiang. Graphene versus mos 2: A short review. *Frontiers of Physics*, 10(3): 287–302, 2015.

[65] Qingkai Yu, Luis A Jauregui, Wei Wu, Robert Colby, Jifa Tian, Zhihua Su, Helin Cao, Zhihong Liu, Deepak Pandey, Dongguang Wei, Ting Fung Chung, Peng Peng, Nathan P. Guisinger, Eric A. Stach, Jiming Bao, Shin-Shem Pei, and Yong P. Chen. Control and characterization of individual grains and grain boundaries in graphene grown by chemical vapour deposition. *Nature materials*, 10(6):443, 2011.

[66] Jeppe V Lauritsen, Jakob Kibsgaard, Stig Helveg, Henrik Topsøe, Bjerne S Clausen, Erik Lægsgaard, and Flemming Besenbacher. Size-dependent structure of mos 2 nanocrystals. *Nature nanotechnology*, 2(1):53, 2007.

[67] NC Bristowe, Philippe Ghosez, Peter B Littlewood, and Emilio Artacho. The origin of two-dimensional electron gases at oxide interfaces: insights from theory. *Journal of Physics: Condensed Matter*, 26(14):143201, 2014.

[68] Leitao Liu, S Bala Kumar, Yijian Ouyang, and Jing Guo. Performance limits of monolayer transition metal dichalcogenide transistors. *IEEE Transactions on Electron Devices*, 58(9):3042–3047, 2011.

[69] Feliciano Giustino. Electron-phonon interactions from first principles. *Reviews of Modern Physics*, 89(1):015003, 2017.

[70] C. Caroli, R. Combescot, D. Lederer, P. Nozieres, and D. Saint-James. A direct calculation of the tunnelling current. ii. free electron description. *J. Phys. C Solid State Phys.*, 4(16):2598–2610, 1971.

[71] A. Jauho, Wingreen, N.S., and Y. Meir. Time-dependent transport in interacting and noninteracting resonant-tunneling systems. *Phys. Rev. B*, 50(8):5528–5544, 1994.

[72] R. Lake, G. Klimeck, R.C. Bowen, and D. Jovanovic. Single and multiband modeling of quantum electron transport through layered semiconductor devices. *J. Appl. Phys.*, 81(12):7845–7869, 1997.

[73] C. H. Lewenkopf and E. R. Mucciolo. The recursive green's function method for graphene. *J. Comp. Elec.*, 12(2):203–231, 2013.

[74] GW Brown and BW Lindsay. The numerical solution of poisson's equation for two-dimensional semiconductor devices. *Solid-State Electronics*, 19(12):991–992, 1976.