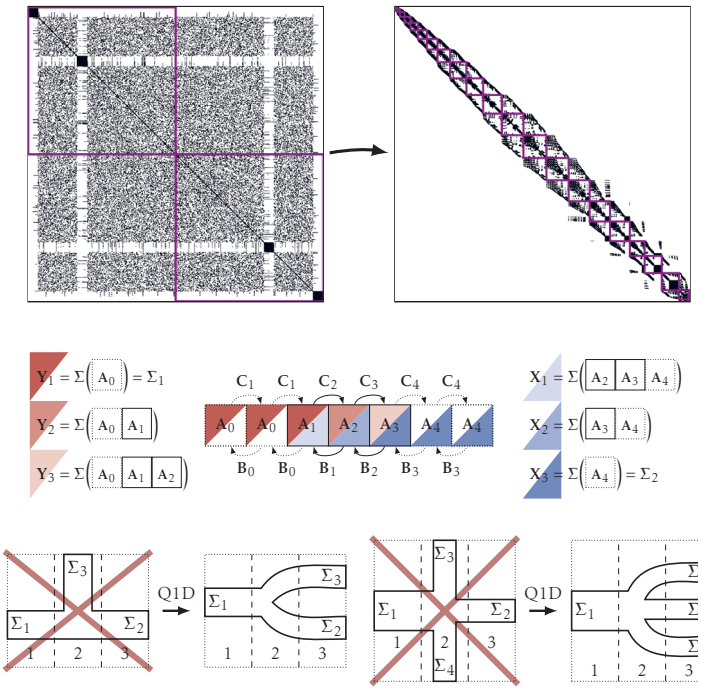
Computational Tools and Studies of Graphene Nanostructures

Papior, Nick Rübner

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record



Computational Tools and Studies of Graphene Nanostructures

Nick Rübner Papior
PhD Thesis January 2016

DTU Nanotech
Department of Micro- and Nanotechnology

Computational Tools and Studies of Graphene Nanostructures

Nick Rübner Papior

January 29, 2016

Thesis submitted for the degree of Doctor of Philosophy

DTU — Nanotech
Center for Nanostructured Graphene
2012–2016

Preface

This work is submitted in candidacy for the Ph.D. degree at the Technical University of Denmark (DTU). The work has been carried out between August 2012 and January 2016 at the Department of Micro and Nanotechnology. The project has been funded by the Center for Nanostructured Graphene (CNG), which is sponsored by the Danish Research Foundation, Project DNRF58.

First and foremost I would like to thank my supervisor Prof. Mads Brandbyge. He has persistently and continuously planted seeds of ideas, of which only a fraction have had time to grow and spur research. His enthusiasm for research and acknowledgement of students *own* prioritisations and *their* research interests is truly inspiring.

Through my Ph.D. I have had the pleasure of collaborating with many people. At Barcelona ICN2 I worked with the group of Prof. Pablo Ordejón, and in particular Nicolás Lorente. Here I came in contact with Georg Huhs whom have helped debugging and testing the new TranSIESTA implementation. Additionally, I have visited DiPC in San Sebastian, with Res. Prof. Thomas Frederiksen, Prof. Daniel Sánchez-Portal, Dr. Aran García-Lekue, Mads Engelund and Pedro B. Mendonça. In particular, Mads and Pedro have been helpful in debugging and providing comments on the implementation.

A huge thanks to Prof. Alberto García for his endurance with my many mails and discussion topics. It has truly been a pleasure collaborating with him. A sincere gratitude to Alberto and the other SIESTA developers for inviting me as lecturer for a SIESTA/TranSIESTA workshop in Tel-Aviv, 2014.

In conjunction with my SIESTA developments I was invited for a CECAM ESL workshop during the summer 2015 in Lausanne. I would like to thank the ESL community for inspiring conversations and their belief in my ability to contribute.

I would also express my gratitude to all current and former members of the groups headed by Mads Brandbyge and Prof. Antti-Pekka Jauho. Particularly, M. L. N. Palsgaard, T. Gunst, D. Stradi, J. T. Falkenberg, K. W. Jacobsen and J. T. Lü, with whom I have contributed to the scientific community via publications.

A big thanks to the people who, willingly, helped in proofreading this manuscript; Kaare W. Jacobsen, Mattias L. N. Palsgaard and Rasmus B. Christensen.

A special thanks to Rasmus B. Christensen for enduring me and my computer oddities while sharing office for the major part of my time at Nanotech.

To my family and my wife, Sidsel Rübner Papior, for her love, support and encouragement which defines me as a scientist and as a person.

...thank you.
Nick Rübner Papior, 2016

Abstract

Nano-electronics industry has during the past decade decreased feature sizes to roughly 10 nm. Such feature sizes are at the quantum limit, requiring a description at the quantum mechanical level. Parallel to the experimental work reside the theoretical tools used to investigate and understand such systems. These theoretical tools still have a high computational requirement. Thus more efficient algorithms are needed to perform studies on even larger systems. Although the gap between the theoretical tools and the experimental setups are reduced, there is *still* a gap, and the used theoretical methods require revised algorithms.

Furthermore, the advent of 2D materials may prove prominent in future nano-electronics for electronic and heat transport devices. Such materials include the Nobel Prize winning material, *graphene* which has unique properties.

The main focus of the work presented in this thesis has been to introduce extensions to the non-equilibrium Green function code TranSIESTA which will help reduce the gap between experimental and theoretical studies. One main achievement in this work is a truly $N_{\mathfrak{C}} \geq 1$ electrode implementation. Another contribution is a general $N_{\mathfrak{C}} \geq 1$ tight-binding transport code which enables not only electronic transport but also phonon transport. This tight-binding code includes features such as, bond-currents, transmission projections and bias-interpolations. For both codes the inversion algorithm for calculating the Green function is revised. We implement a high performing block-tri-diagonal algorithm for calculating the Green function and spectral density function — even for $N_{\mathfrak{C}} > 2$. This is accomplished by bandwidth reduction schemes that increase the quasi-1D interpretation.

We also present a new gating method capable of calculating gating effects. A graphite example is used to highlight the importance of the quantum capacitance that is evident in low density of states systems. Additionally the gating method was used in non-equilibrium to study the gate-bias dependence on graphene nano-constrictions. This indicated a pinning effect arising due to differences in coupling strength between the device and the two electrodes.

Two studies are presented using the non-equilibrium method with $N_{\mathfrak{C}} = 3$. First, graphene T-junctions are studied to uncover potential interconnects in future graphene based devices. This T-junction is studied under two non-equilibrium situations. Our second study is a graphene Scanning Tunnelling Microscopy setup where the inelastic current is calculated. We show, that the experimental inelastic signal may be fully recovered using density functional theory and non-equilibrium Green function techniques.

Resumé

I det seneste årti er komponenterne anvendt i nanoelektronik industrien blevet mindre og mindre, og de har nu nået dimensioner helt ned til omkring 10 nm. Komponenter af sådanne dimensioner kræver en beskrivelse af kvantemekanisk karakter. Udover den store eksperimentelle indsats, der gøres for at beskrive sådanne komponenter, er der behov for teoretiske værktøjer til at analysere og forstå dem. Disse teoretiske værktøjer stiller stadig store krav til beregningskapaciteten og der er derfor behov for mere effektive algoritmer og implementeringer. Selv om forskellene mellem de teoretiske modeller og de faktiske eksperimentelle undersøgelser er blevet reduceret, findes der stadig et gab, og de teoretiske modeller kræver derfor reviderede algoritmer.

Derudover kan forekomsten af 2D materialer være fremtræden i fremtidig nanoelektronik med henblik på elektroniske og transport af varme egenskaber. Et sådant materiale er f.eks. det Nobel pris vindende materiale grafen med unikke egenskaber.

Denne afhandling har stort fokus på at introducere udvidelser til uligevægts Green funktion koden, TranSIESTA. Dette vil hjælpe til at mindske gabet mellem eksperimentelle og teoretiske studier. Et af de vigtigste resultater af arbejdet er udviklingen af en generel $N_\mathfrak{c} \geq 1$ kode. Et andet bidrag er en generel $N_\mathfrak{c} \geq 1$ "tight-binding" transport kode, som gør det muligt at beregne både elektron og fonon transport. Denne "tight-binding" kode omfatter analyse metoder såsom, bånd-strømme, transmission projektioner og spændings-interpolationer. For begge koder er inversions algoritmen til at beregne Green funktionen revideret. Vi har implementeret en blok-tri-diagonal algoritme for effektivt at beregne Green og spektral funktionen — selv for $N_\mathfrak{c} > 2$. Dette er opnået ved en båndbredde reduktion, der øger kvasi-1D formen.

Vi præsenterer også en ny gate metode til beregning af effekter. En beregning på grafen/grafit er anvendt som eksempel til at fremhæve betydningen af kvante kapacitansen, der er prominent for systemer med lav densitet af tilstande. Derudover er gating metoden anvendt i uligevægt for at studere gate-skævhed afhængighed grafen nano-forsnævringer. Dette viste en "pinning" effekt, der opstod på grund af forskelle i koblings styrken mellem systemet og de to elektroder.

Yderlige to beregninger er præsenteret, der begge bruger uligevægts metoden med $N_\mathfrak{c} = 3$. I den første beregning er et grafen T-kryds undersøgt for at afdække potentielle forbindelser i fremtidige grafen baserede systemer. Dette T-kryds er undersøgt i to uligevægt situationer. Den anden beregning omhandler et grafen STM setup hvor den uelastiske strøm er beregnet. Vi viser, at det eksperimentelle uelastiske signal kan blive fuldstændigt genproduceret ved hjælp af tæthedsfunktionalteori og uligevægts Green funktion teknikker.

Contents

List of Contributions

The published/submitted articles amount to the following articles, in chronological order. They can all be found in App. D. Throughout the thesis the citation of published articles will be in italics. In the bibliography section the published articles are printed in bold.

[**134**] *N. L. Schneider, N. Nel, N. P. Andersen, J. T. L, M. Brandbyge, J. Krger, and R. Berndt. "Spectroscopy of transmission resonances through a C60 junction". In:* Journal of Physics: Condensed Matter *27.1 (2015), p. 015001.* DOI: *10.1088/0953-8984/27/1/015001*

[**171**] *M. L. N. Palsgaard, N. P. Andersen, and M. Brandbyge. "Unravelling the role of inelastic tunneling into pristine and defected graphene". In:* Physical Review B *91.12 (2015).* DOI: *10.1103/PhysRevB.91.121403*

[**168**] *J. Lagoute, F. Joucken, V. Repain, Y. Tison, C. Chacon, A. Bellec, Y. Girard, R. Sporken, E. H. Conrad, F. Ducastelle, M. Palsgaard, N. P. Andersen, M. Brandbyge, and S. Rousset. "Giant tunnel-electron injection in nitrogen-doped graphene". In:* Physical Review B *91.12 (2015), pp. 1–5.* DOI: *10.1103/PhysRevB.91.125442*

[**140**] *N. Papior, T. Gunst, D. Stradi, and M. Brandbyge. "Manipulating the voltage drop in graphene nanojunctions using a gate potential". In:* Phys. Chem. Chem. Phys. *18 (2016), pp. 1025–1031.* DOI: *10.1039/C5CP04613K*

[**159**] *K. W. Jacobsen, J. T. Falkenberg, N. Papior, P. Bggild, A.-P. Jauho, and M. Brandbyge. "All-graphene edge contacts: Electrical resistance of graphene T-junctions". In:* Carbon *(2016).* DOI: *10.1016/j.carbon.2016.01.084*

A list of conference contributions

- *DFT-NEGF calculations of gated graphene nano-structures*, N. Papior, T. Gunst, D. Stradi, M. Brandbyge; Graphene 2014, Toulouse, France
- *Re-implementation of* TranSIESTA *and N-terminal NEGF*, N. Papior, M. Brandbyge; Trends in Nano Technology 2014, Barcelona, Spain. **David Prize award**, best theory poster.
- *Electrostatic gating of 2D nanostructures & re-implementation of* TranSIESTA, N. Papior, T. Gunst, M. Brandbyge; QuantumHagen 2014, Copenhagen, Denmark
- *Multiple electrode ($N_{\mathfrak{e}} \geq 1$) support in the DFT+NEGF code* TranSIESTA, N. Papior, M. Brandbyge; Psi-K 2015, San-Sebastian, Spain

- *Field effect transistor devices based on two-dimensional heterostructures investigated by large-scale DFT+NEGF simulations* TranSIESTA, D. Stradi, N. Papior, M. Brandbyge; Psi-K 2015, San-Sebastian, Spain
- *Electronic and Transport Properties of 4-Terminal Crossed Graphene Nanoribbon Devices*, P. B. Mendonça, N. Papior, M. Engelund, A. García-Lekue, T. Frederiksen, D. Sánchez-Portal; Psi-K 2015, San-Sebastian, Spain
- *The Electronic Structure Library: a community-driven common library project for software development in electronic structure, et. al.*; Psi-K 2015, San-Sebastian, Spain
- *Electron Transport in Crossed Graphene Nanoribbon Devices: 4-Terminal* ab-initio *Simulations*, P. B. Mendonça, N. Papior, M. Engelund, A. García-Lekue, T. Frederiksen, D. Sánchez-Portal; Open Quantum Systems Computational Methods 2015, Hong Kong

I have been invited lecturer for a workshop (SIESTA/TranSIESTA *Tutorial*) in Tel-Aviv, 2014. Here I gave a series of lectures on using TranSIESTA and made a tutorial for users.

I have also been invited as contributing developer to the Electronic Structure Library. In 2015 a dedicated workshop took place in Lausanne, Switzerland. Here I developed the Lua-hook library, `flook`, which enables Lua-scripting directly in fortran.

List of Figures

Chapter 1

Introduction

The current technological era is expanding its horizons and applications in an exponentially increasing stream of development. Most importantly the advancements made within semi-conductor technology have driven forth a multitude of research areas as the handling and analysis of massive amounts of data become accessible on common work-station computers [1].

The workhorse of computers[1] are transistors which enable the control of current, as the name also implies. A transistor is the combination of the transconductance *and* the transresistance of the device. It is not only defined by one of them, but both, hence *transistor* [2]. Ever since the discovery of the transistor in 1947 it has decreased from a mm-scale device to a nm-scale device, more than 6 orders of magnitude in 6 decades. This has lead to the formulation of Moore's law which states that the number of transistors per unit area would double around every 2nd year yielding an exponential increase in transistor density. Remarkably transistors are still based on silicon technology due to its unique features [3]. This advancement is a self perpetuating cycle as the increased computer power allows the investigation of more complex and/or complete physical models. This aids in the understanding of the physics governing even smaller features, yielding to the development of smaller physical devices in a feedback loop. See Figure 1.1 for the currently used technology which was projected from 2007.

Feature sizes in current transistors are on the order of 10 nm which correspond to roughly 40 atoms of silicon. A perplexing thought, that our current technological

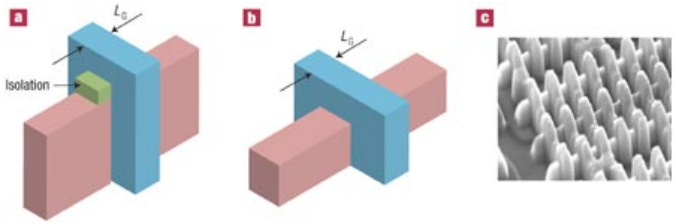


Figure 1.1 | Advances within the tri-gate non-planar transistors implemented in Intel processing units since 2007. The transformation from 2D planar structures to 3D structural transistors allowed further increasing the density of transistors. From [3].

[1] Computers encompass; mobile phones, televisions, control boards for electricity, computer workstations up to HPC computers, etc.

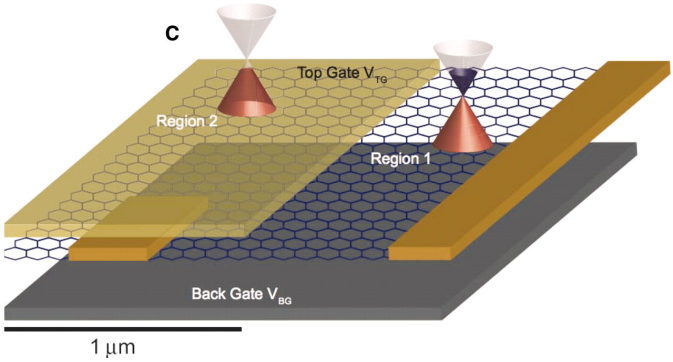


Figure 1.2 | Experimental setup for gate controlled *pn*-junction displaying the quantum-Hall effect. Quantised conductances are measured via gate controlling the graphene sheet in separate regions. The back gate controls the entire sheet, while the top only controls the left side of the sheet allowing a *pn*-junction. From [28].

world relies on cross sections of only 40 atoms! These feature sizes, although seemingly *extremely* small, are still just touching the limit of the computational complexity involved when simulating a full device using quantum mechanical tools.

From a theoretical point of view we need tools and methods to study realistic devices. Within this thesis we focus on the density functional theory [4, 5] which greatly simplifies the quantum mechanical N-particle problem. Current density functional theory codes can handle thousands of atoms [6–9]. However, density functional codes does typically not handle non-equilibrium situations such as that experienced in a transistor with an applied bias. To handle non-equilibrium the Green function techniques are used [10–18]. These techniques are typically heavier than equivalent calculations using density functional theory. As such, research in *both* experimental downscaling *and* computationally performing algorithms are crucial.

1.1 Will silicon remain dominant?

The versatility and shear abundance of silicon makes it a still prominent candidate to carry the advancements into the near future, and perhaps longer [3]. However, several new materials are emerging as candidates to replace silicon; carbon nanotubes for small-scale 1D-like devices, graphene (also pure carbon) and other 2D-materials [19–23]. These, however still face a large range of problems before they will be able to replace silicon. Yet their possible influence have spurred a massive amount of research, including the work presented in this thesis.

The 2D nature of graphene has some interesting features which are outlined in Refs. [24, 25]. A particular feature of graphene is the tuneability of the Fermi level due to doping [26, 27] allowing a range of doping corresponding to roughly 1 eV. So far, control of graphene has been proven in various studies [28–34]. Figure 1.2 shows the experimental setup conducted by Williams *et al.* [28], where a *pn*-junction is realised using two gates, one gating the entire graphene sheet and another gating only part of the graphene sheet. Due to the independent gating a *pn*-junction is formed, as indicated by the two differently filled Dirac-cones. This device yields quantum conductance channels exhibiting a step-wise increment of conductance. A needed feature in graphene

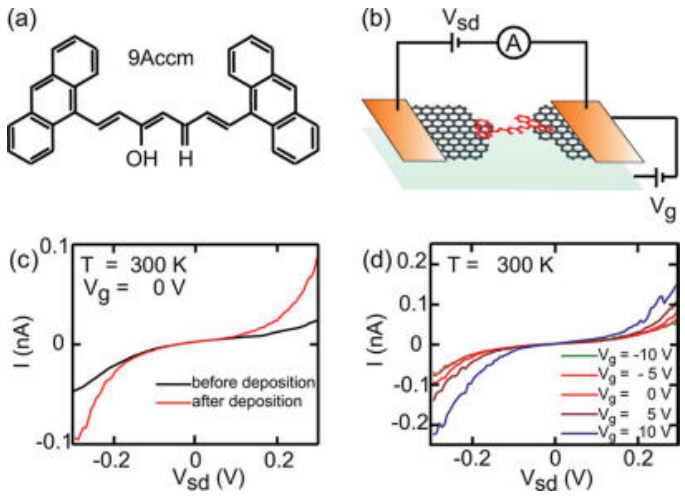


Figure 1.3 | Molecular junction connecting two pristine graphene sheets. a) the molecule bridged, b) the device setup with gate controlling the entire device, c) signature of molecule signal from before and after molecule deposition, and d) the device features for varying gate potentials. From [26].

for transistor operations is a band-gap. I.e. for the *on* state the device needs to be tuned/gated out of the band-gap (non-vanishing density of states at the Fermi-level) while the *off* state needs to be well in-side the band-gap (vanishing density of states at the Fermi-level)[2] [22]. Several suggestions for band-gap engineering exists [35, 36]. Generally, a confinement of the electrons generate *particle-in-a-box* like states which inherently creates gapped energy levels with larger spread for narrower widths [37]. Graphene constrictions can be used while derivatives of connecting two pristine graphene sheets with, for instance, molecules have already been realised [26, 38]. In Figure 1.3 a molecular junction is shown. a) and b) shows the molecule and experimental setup, respectively. c) and d) presents the measurements indicating an adsorbed molecule in the junction with and without gating potentials.

Theoretical models describing graphene are well studied, since they only involve single specie parameterisations [39, 40]. It is ideal to study using tight-binding models and empirical potential models for carbon. In these empirical models the scalability of even the simplest calculations are extreme and one can easily calculate properties of graphene systems exceeding 10,000 atoms. However, investigating graphene in chemical environments such as adsorption of different chemical species requires *parameter free* methods. Here lies the strength of density functional theory and extensions, such as Green function techniques. However, they are heavily limited by the implemented algorithms and calculations rarely exceed 1,000 atoms due to time and memory restraints on the hardware.

[2] In this regard research is focusing on a multitude of different 2D-materials as *graphene alternatives*.

1.2 Contacting < 3D materials

An interface to connect to external devices, such as bulk copper wires or gold contacts, is required for using 1D/2D materials as intrinsic elements in electronics [41–43]. Studies have shown that for graphene contacting metal, a 1D contact along the edge is favourable [43]. However, we still lack a complete understanding of why this is. This also reflects itself in STM experiments and gate controlled devices [29–31, 42, 44–46]. However, conducting theoretical calculations to obtain current characteristics poses difficulties for STM investigations of graphene and other low dimensional materials. An obstacle is that the majority of non-equilibrium techniques can not simulate devices connected to more than 2 electrodes. Hence, a tri-gated device becomes difficult to describe using common codes. Sometimes, tri-gates can be simulated using phenomenological models which describe the gate as a perturbation to the electrostatic environment [47–49]. Yet generally methods for conducting truly $N_{\mathfrak{e}}$-terminal non-equilibrium calculations have been scarce, and as far as the author is aware, no free academic code can handle full non-equilibrium with $N_{\mathfrak{e}}$ terminals. Reports do exist of NEGF $N_{\mathfrak{e}}$-terminal support such as [14, 50, 51].

1.3 Outline of thesis

From this thesis perspective, two major issues exist; 1) the capability of non-equilibrium Green function techniques is in dire need of expansion such that even larger and more realistic systems can be analysed, 2) a generic $N_{\mathfrak{e}}$-terminal Green function technique would complement the experimental advancements and allow more complex geometric systems.

In Chapter 2 an introduction to density functional theory is given. The most basic concepts are presented and the Hessian matrix, periodicity and the 1D chain band structure problems are introduced. Furthermore, an introduction to SIESTA is given.

The Green function method is presented in Chapter 3. The concept of density of states via Green functions and the self-energy is introduced. Bloch expansion of Green functions are also covered.

Chapter 4 is covering the full re-implementation of TranSIESTA starting from the $N_{\mathfrak{e}}$ electrode Green function equations under equilibrium and non-equilibrium. The details of the implemented algorithms are explained and several comparisons and peculiarities of the method and algorithms are emphasised.

Chapter 5 covers TBtrans which enables the calculation of transport properties. The inversion algorithm is, again, explained and several new techniques are covered, such as interpolation, bond-currents, eigenstate transmission projections.

A new gate model is introduced in Chapter 6 where we show the importance of correctly describing graphene with a gate-model and charging effects. In particular, the electrostatics are behaving differently under such a gate-model.

T-junctions are investigated in Chapter 7. Here we make use of the $N_{\mathfrak{e}} > 2$ capabilities of TranSIESTA to investigate how to make 3D interconnects in graphene 2D devices.

Lastly, in Chapter 8 the inelastic conductance of STM on graphene is covered. Full DFT+NEGF calculations are used to reproduce experimental findings in pristine graphene and nitrogen doped graphene.

Chapter 2

Density functional theory

In this Chapter we will introduce the basic underlying methodology used throughout this thesis, namely the mean-field theory: density functional theory (DFT). For the past several decades the increase in computational power and ever more complex computer codes have made DFT codes a common method in a physicists toolbox for understanding complex systems. For more introductory material we refer to [52–55] which covers many important aspects of condensed matter theory and DFT schemes. An example 1D chain will be used to explain certain important aspects and it will be re-used in the following Chapter.

Lastly, a dedicated section on SIESTA [6] introduces the local basis set used throughout the thesis.

2.1 Density functional theory

Density functional theory (DFT) is one the most frequently used methods to describe complex physics in a wide variety of environments. DFT is a parameter free method giving it a predictive power, without relying on explicit fitting parameters. However, it is based on ground state theory which will be made clear through this section.

The generic problem in the quantum mechanical picture is the solution of the time independent Schrödinger equation

$$\mathbf{H}\Psi(\mathbf{r},\mathbf{R}) = E\Psi(\mathbf{r},\mathbf{R}). \tag{2.1}$$

Here $\mathbf{r} = \{\mathbf{r}_1, \mathbf{r}_2, \ldots\}$ (electron positions) and $\mathbf{R} = \{\mathbf{R}_1, \mathbf{R}_2, \ldots\}$ (nuclei positions). The Hamiltonian *describes* the system having terms originating from different levels of interaction, the Hamiltonian can be written as a sum of kinetic and potential contributions

$$\mathbf{H} = \mathbf{T} + \mathbf{V}, \tag{2.2}$$

$$= \mathbf{T}_e + \mathbf{T}_n + \mathbf{V}_{ee} + \mathbf{V}_{en} + \mathbf{V}_{nn}, \tag{2.3}$$

where subscript e and n denotes electron and nuclei, respectively, hence $\mathbf{V}_{en}$ describes the potential contribution of electron-nuclei interactions. The contributions can be written

as

$$\mathbf{T}_e = -\sum_i \frac{\hbar^2}{2m_i}\nabla_i \qquad \mathbf{V}_{ee} = +\frac{e^2}{2}\sum_i\sum_j \frac{1}{|\mathbf{r}_j - \mathbf{r}_i|}$$
$$\mathbf{T}_n = -\sum_\alpha \frac{\hbar^2}{2M_\alpha}\nabla_\alpha \quad ; \quad \mathbf{V}_{en} = -\frac{e^2}{2}\sum_i\sum_\alpha \frac{1}{|\mathbf{R}_\alpha - \mathbf{r}_i|} \tag{2.4}$$
$$\mathbf{V}_{nn} = +\frac{e^2}{2}\sum_\alpha\sum_\beta \frac{1}{|\mathbf{R}_\beta - \mathbf{R}_\alpha|}$$

with i, j denoting electronic indices and α, β nuclei indices. The 1/2 in the $\mathbf{V}$ terms removes double counting terms.

In principle the exact wavefunction can be found by solving the Schrödinger equation including the terms in Eq. (2.4). However, the complete wavefunctions cannot be calculated using conventional methods for anything but the simplest systems involving very few particles. This is because the Schrödinger equation includes $3N_i + 3N_\alpha$ coupled equations, with N_i, N_α being the number of electrons and atoms, respectively.

A first approximation in reducing the complexity, was the scheme of an *adiabatic* separation of electrons and nuclei proposed by Born and Oppenheimer [56]. The mass of the heavy nuclei are so large (even for Hydrogen) that the electrons will always see the atoms in a static configuration, $M_\alpha \approx 1836 m_i$. This allows us to write the electrons Schrödinger equation as

$$\mathbf{H}_e = \mathbf{T}_e + \mathbf{V}_{ee} + \mathbf{V}_{en} \tag{2.5}$$
$$\mathbf{H}_e\psi(\mathbf{r},\mathbf{R})\chi(\mathbf{R}) = E\psi(\mathbf{r},\mathbf{R})\chi(\mathbf{R}), \tag{2.6}$$

where now the nuclei wavefunctions χ are separated from the electronic wavefunctions ψ. Note that ψ still depends parametrically on the atomic coordinates due to the environment[1].

Secondly, as the name implies, density functional theory is based on the *density* rather than the exact wavefunctions. The initial theorems of DFT originated in 1964 by Hohenberg and Kohn [4] which has since been the basis of a plethora of successful DFT codes. A thorough and now old, but still applicable review can be found in [57] while a recent, short and concise summation of its influence over the past decades can be found in [58].

The energy associated with a trial electronic density ρ can be given as

$$E[\rho] = F[\rho] + \int \mathrm{d}\mathbf{r}\rho(\mathbf{r})V_{\text{ext}}(\mathbf{r}), \tag{2.7}$$
$$F[\rho] = T[\rho] + E_{ee}[\rho] = T[\rho] + E_H[\rho] + E_Q[\rho] \tag{2.8}$$

where $V_{\text{ext}}(\mathbf{r})$ is the external potential and $F[\rho]$ the electrons kinetic and electron-electron interaction energy (also known as the Hohenberg-Kohn functional), both of which are defined via ρ. The electronic interaction energy is split up in the classical Coulomb (Hartree) energy and a non-classical part denoted by Q. The Hohenberg and Kohn theorems states that the electronic density univocally determines the external potential, hence if $V_{\text{ext}} \leftarrow \rho$ and $V_{\text{ext}} \leftarrow \rho'$ then $\rho \equiv \rho'$. Further, the variational principle for the

[1] Equivalently χ depends on electron positions, $\mathbf{r}$, but in an adiabatic approximation with an average density of electrons, hence an explicit dependence of electron positions is *too* specific.

ground state density enables us to find the ground state energy E_0 by minimising the energy functional $E[\rho]$

$$E_0[\rho] = \min_{\int_\nu \rho' \mathrm{d}\mathbf{r} \to N} E[\rho']. \tag{2.9}$$

We note that the variational principle can be used to estimate an excited state through $E_0[\rho']_{\int \rho' = N+1} - E_0[\rho]_{\int \rho = N}$.

At this point we know how to find the ground state by trying different densities until you find the density with the lowest energy. However, this is unfeasible due to the magnitude of the problem arising from a many-body wavefunction. To solve this problem Kohn and Sham proposed the following [5]. They suggested to calculate the kinetic energy of a non-interacting reference system which recreates the density in the interacting frame

$$\rho(\mathbf{r}) \equiv \rho_{KS}(\mathbf{r}) = \sum_i |\psi_i^{KS}(\mathbf{r})|^2 f_i, \tag{2.10}$$

$$T_{KS}[\rho] = -\frac{1}{2} \sum_i \langle \psi_i^{KS} | \nabla^2 | \psi_i^{KS} \rangle \tag{2.11}$$

where we call the non-interacting orbitals, or single-particle wavefunctions, ψ_i^{KS} the Kohn-Sham orbitals and f_i the distribution function of orbital i. f_i is here the Fermi-function. Note that the physical interpretation of the Kohn-Sham orbitals has been debated since their formulation, yet their influence and appealing physical interpretation *is* used throughout [59]. In the following we will disregard the super script KS and assume them Kohn-Sham orbitals unless otherwise stated. By defining Eq. (2.11) in a non-interacting frame we neglect the many-body effects contribution to the kinetic energy through the two so-called *exchange* and *correlation* energies

$$E_x[\rho] + E_c[\rho] = T[\rho] - T_{KS}[\rho] + E_Q[\rho], \tag{2.12}$$

where we also absorb the non-classical part from Eq. (2.8). Typically the exchange and correlation energies are referred by a single energy functional $E_{xc}[\rho]$ resulting in the final expression for the Hohenberg-Kohn energy functional

$$F[\rho] = T_{KS}[\rho] + E_H[\rho] + E_{xc}[\rho]. \tag{2.13}$$

If the exact exchange-correlation functional was known, so would the ground state energy and density. The Kohn-Sham equations can then be written as the following eigenvalue problem

$$\left[-\frac{1}{2}\nabla^2 + V_{\text{eff}}(\mathbf{r}) \right] \psi_i = \varepsilon_i \psi_i. \tag{2.14}$$

The effective potiental, $V_{\text{eff}}(\mathbf{r})$, has the terms corresponding to the Hartree and the exchange-correlation terms. The Kohn-Sham equations is a self contained set of equations where the density from ψ_i are needed for calculating the effective potential. Hence the equation can be solved by an iterative approach. This short introduction to DFT can be extended by looking into the multitude of papers and text books on such codes [52–55].

From Eq. (2.13) it is important to note that we have simply moved several unknowns into a single unknown, the exchange-correlation functional. The exact form of the exchange-correlation is unknown and attempts to find the *best* (true) form is long, hard,

and most importantly, still on-going [60]. It is thus outside the scope of this thesis to go into a discussion of exchange and correlation functionals. Suffice to acknowledge that an omnipotent form does not currently exist and each form has their own weakness' and strength's. The open-source project LibXC [61] tries to collect the published functionals in a single library and is considered the *de facto* library for exchange-correlation functionals. Throughout this thesis we rely on the simple and widely used form of the local density approximation (LDA) PZ [62], or the generalised gradient approximation (GGA) as suggested by PBE [63].

2.1.1 Forces and the Hessian matrix

From the equilibrium density of a given atomic configuration we can calculate the forces acting on the atoms via the Hellmann-Feynman theorem [64]. This further enables DFT to minimise forces of a given geometry by rearranging the atoms. Eventually the equilibrium positions are reached via the minimisation of the forces exerted on the atoms. Nearly all studies start by this minimisation of forces to ensure that it is indeed the equilibrium positions that is being studied.

Consequently one can calculate the force constant matrix, also known as the Hessian matrix

$$\mathcal{K} = \frac{\partial^2 E(\mathbf{R})}{\partial \mathbf{R}' \partial \mathbf{R}''} = -\frac{\partial \mathbf{F}(\mathbf{R})}{\partial \mathbf{R}''}. \tag{2.15}$$

The calculation of the Hessian matrix can be cumbersome due to the double differential of the energy functional. Two ways of calculation the Hessian are; 1) the atomic displacement method also known as *frozen phonons*, or 2) via linear response theory which is a derivative of the Hellmann-Feynman theorem. For an excellent review of the methods see Baroni *et al*. [65].

The simplest method to implement is the frozen phonons. To correctly describe the Hessian matrix one requires a displacement of each atom in every Cartesian[2] direction along the positive and negative direction[3] for the direct calculation of the Hessian

$$\mathcal{K}_{Ia,Jb} = \frac{\mathbf{F}_{Jb}(\mathbf{R} + \delta\mathbf{R}_{Ia}) - \mathbf{F}_{Jb}(\mathbf{R} - \delta\mathbf{R}_{Ia})}{2\delta\mathbf{R}_{Ia}}, \tag{2.16}$$

where a and b belong to the Cartesian directions $\{x, y, z\}$. One chooses $\delta\mathbf{R}_{Ia}$ so small that the harmonic approximation holds. Hence for light species $\delta\mathbf{R}_{Ia}$ must be relatively small compared to heavier atoms, however, the displacements must not be so small that the egg box effect[4] becomes important. Typically a value around 0.02 Å is a good compromise of small $\delta\mathbf{R}$ to capture the harmonic approximation while large enough to disregard the egg box effect, if the mesh is defined by a large enough cutoff. The dynamical matrix is defined as the mass scaled Hessian matrix

$$\mathbf{D}_{Ia,Jb} = \mathcal{K}_{Ia,Jb} \frac{1}{\sqrt{M_I M_J}} \tag{2.17}$$

and the eigenvalues (ω_i) and eigenvectors (v_i) of the dynamical matrix are the corresponding phonon spectrum and eigenmodes of the system in the harmonic approximation.

[2] It *need* not be a Cartesian direction, any orthogonal coordinate basis can be chosen, for example, estimated phonon modes.

[3] One could do with one direction but it is advised to do both to capture symmetries.

[4] The egg-box effect arises due to finite grids where each element volume is too large.

For calculating the entire Hessian, Eq. (2.15), a total of $6N_A$ displacements are necessary which makes the frozen phonon method extremely cumbersome for large systems.

The linear response method is much faster as it is a perturbative method only relying on the equilibrium position. Further the linear response theory enables the implicit unfolding of $\mathbf{q}$-points reducing the computationally complexity enormously. For further details on the implementation of the linear response theory in SIESTA see [66].

2.1.2 Periodicity

An important aspect of periodic calculations is that one can expand periodic potentials using Bloch's theorem. In this section we take advantage of the reciprocal space with wave vectors $\mathbf{k}$ defined in the Brillouin zone.

For each unit cell defined by the lattice vectors $\mathbf{a}_i$, for $i \in \{1,2,3\}$ there exists a reciprocal lattice vector $\mathbf{b}_i$ such that $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$. The reciprocity can be related through the volume relation

$$\mathcal{V}_\mathbf{a} = \frac{(2\pi)^3}{\mathcal{V}_\mathbf{b}}. \tag{2.18}$$

By choosing $\mathcal{V}_\mathbf{a} \to \infty$, we get $\mathcal{V}_\mathbf{b} \to 0$. In that limit there is no need to describe periodicity as the system under investigation can be adequately sampled using a single $\mathbf{k} = \mathbf{0}$ point. This special point is called the Γ-point. The reciprocal space can be used to expand wavefunctions and other periodic quantities in terms of Bloch states. Whence a wavefunction known in a unit cell is simultaneously known in a translated unit cell via the wave vector components

$$\psi_\mathbf{k}(\mathbf{r} + \mathbf{T}) = e^{\mathrm{i}\mathbf{k}\cdot\mathbf{T}} u(\mathbf{r}) \tag{2.19}$$

where $\mathbf{T} = i\mathbf{a}_1 + j\mathbf{a}_2 + k\mathbf{a}_3$ defines an integer translation of the unit cell lattice vectors. Note that we have added a subscript $\mathbf{k}$ which is bounded in the interval $\mathbf{k} \in]-\frac{1}{2};\frac{1}{2}]\mathbf{b}$ for the first Brillouin zone. The wave vector $\mathbf{k}_i$ describes a resulting periodicity in the ith direction via

$$n\mathbf{k}_i \cdot \mathbf{a}_i = 2\pi m \quad \text{, for } \{n,m\} \in \mathbb{N}, \tag{2.20}$$

hence for $\mathbf{k} \to 0$ the periodicity tends towards an infinite periodicity, i.e. the periodicity of a single unit cell. While for $\mathbf{k}_i = \mathbf{b}_i/2$ (the Brillouin zone boundary) a periodicity of the wave function spans two unit cells. Note that the allowed values of $\mathbf{k}$ are constrained to be fractions of $\mathbf{b}$.

At this point it is instructive to define the Hamiltonian for a certain $\mathbf{k}$

$$\mathbf{H}_\mathbf{k} = \mathbf{H}e^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}} \tag{2.21}$$

or alternatively as

$$\mathbf{H}_\mathbf{k} = \mathbf{H}e^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}, \tag{2.22}$$

with $\mathbf{R}$, $\mathbf{r}$ being the lattice vector and interatomic distance, respectively. Both methods of constructing the Hamiltonian are valid, the former is easier than the latter as one need not determine the distance between atoms.

The energy dispersion in the first Brillouin zone creates a band-structure for the crystal. For more details see [52, 54, 55].

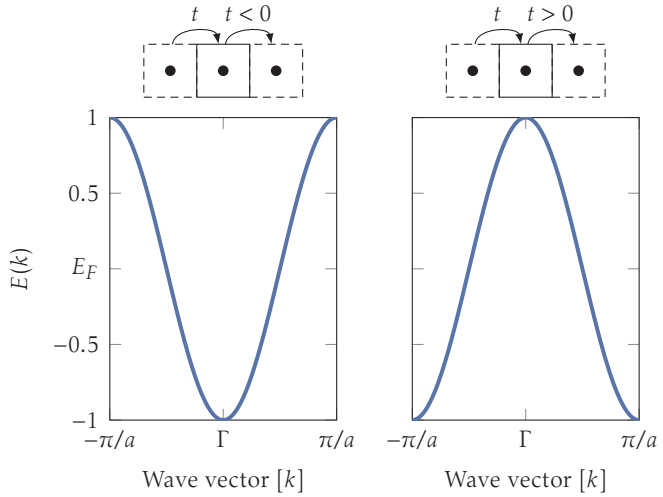


Figure 2.1 | One orbital system shown on top of band-structures with full lines the system, and dashed lines the nearest neighbour interactions. Band structure of 1D chain with hopping integral $t = \mp 1/2$ showing the $2t\cos(ka)$ dispersion.

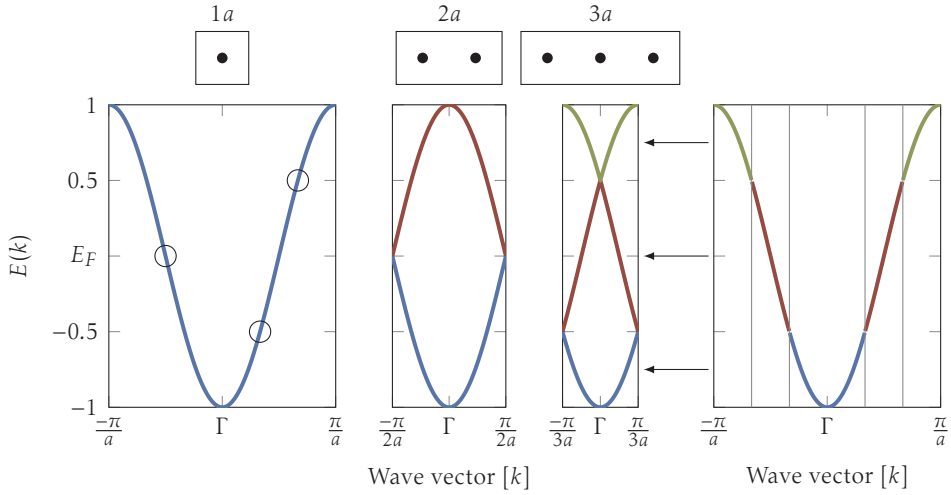


Figure 2.2 | Band structure of 1D chain with increasing number of orbitals $N = \{1, 2, 3\}$. For larger cells the reciprocal cell scales inversely as indicated by the size of the k-axis. Far right; a down folding of the $N = 1$ band into the first Brillouin zone of $N = 3$. This is indicated by vertical slices of the Brillouin zone and translating them into the primary zone.

In the following we will consider a 1D chain with nearest neighbour interactions, and since it is 1D there is only one **k** component. For completeness sake we write the energy dispersion in the reciprocal cell which can easily be derived to be

$$E(k) = 2t\cos(ka). \tag{2.23}$$

In Figure 2.1 the band structure of a 1D chain is shown with different signs of the hopping parameter. The dispersion has maximum/minimum at the band-edges and the Γ point. Note that at any given k there is only one eigenvalue.

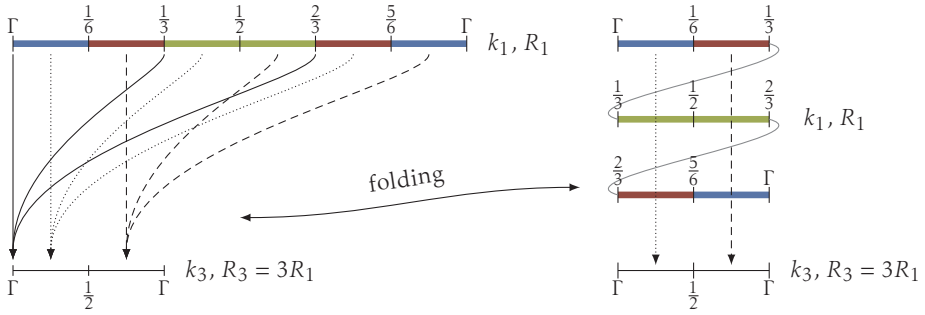


Figure 2.3 | Band folding method shown in reciprocal space. Can be compared to Figure 2.2 for the $N = 3$ case. The left shows the two reciprocal cells compared against each other, R_1 and R_3. Each type of line represents an equivalence of k-points between the two different cells. The right shows the same equivalence scheme by slicing the Brillouin zone into 3 pieces. In this scheme a k_3 point encompass all vertically crossed points in the k_1 Brillouin zone. For any $R_N = NR_1$ there will be N splits with each band folding taking a point from each split section. This is also indicated by the colouring of the Brillouin zone parts.

By creating a unit cell with an integer number of more orbitals we find a band folding as seen in Figure 2.2. With an N times larger system there exists N bands due to the folding as seen by comparing the right-most figure with the 3 left figures. The $N = 1$ system has 1 band, for $N = 2$ there are two orbitals, i.e. two bands, while $N = 3$ has 3 bands. The arrows indicate the band folding due to increasing the system size, see e.g. [52]. The figure also emphasises Eq. (2.18) by decreasing the Brillouin zone for increasing N.

From Figure 2.3 the explicit k points that are down folded can be easily extracted by splitting the smaller unit cell (larger reciprocal cell) into n parts and adding each of the crossing k points in the smaller unit cell combined into a single k point in the larger unit cell (smaller reciprocal cell).

Using the above folding description one can *unfold* the Hamiltonian for a 1D chain of size R to any size nR $(\{\}^n)$ at any k_n, index n to denote k_n with respect to the nR unit cell. Again only the primary Hamiltonian $(\{\}^1)$ for the system of size R for n different k_1 values is needed

$$\mathbf{H}^n_{k_n} = \frac{1}{n} \sum_{\substack{i \\ k_i = k_n + 2\pi \frac{i-1}{nR}}}^{n} \mathbf{H}^1_{k_i} \cdot \begin{bmatrix} \mathbf{I} & e^{-\mathrm{i}k_i R} & \cdots & e^{-\mathrm{i}nk_i R} \\ e^{\mathrm{i}k_i R} & \mathbf{I} & \cdots & e^{-\mathrm{i}(n-1)k_i R} \\ \vdots & \vdots & \ddots & \vdots \\ e^{\mathrm{i}nk_i R} & e^{\mathrm{i}(n-1)k_i R} & \cdots & \mathbf{I} \end{bmatrix}. \tag{2.24}$$

This expansion involves several important aspects, first recognise that $k_N = 2\pi f/(NR)$ with $f \in]-\frac{1}{2}; \frac{1}{2}]$.

1. $k_i = k_N + 2\pi \frac{i-1}{NR}$ consists of two terms; 1) k_N is equal to the equivalent k_1 point as $k_1 = k_N NR/(NR)$[5], 2) $2\pi(i-1)/(NR)$ are the expansion k_1 points from the primary cell downfolded into the larger unit cell (smaller reciprocal cell) and are equivalent to the distance between marked points/crossing lines shown in Figure 2.3,

[5] Note that for generality a proper conversion is performed if non-parallel unit cells and/or different lattice vector orderings are used.

2. $\mathbf{H}^1_{k_i}$ the Hamiltonian for the ith downfolded k_1 point in the primary cell,
3. Matrix factors $e^{\mathrm{i}sk_iR}$ with $s \in \{-N, -N+1, \cdots, N\}$ are the translation factors of the Bloch states according to Eq. (2.19), note that one need not the R factor as it is redundant: $e^{\mathrm{i}sk_iR} = e^{\mathrm{i}s2\pi f_i}$, with $f_i = (f + i - 1)/N$. Every matrix element is a sub-matrix of dimensions equal to $\mathbf{H}^1$ and the $\cdot$ refers to element wise multiplications in the matrix elements

From the above recipe, any Hamiltonian based on the primary unit cell Hamiltonian can be calculated accurately and efficiently. Note, that it is not limited to expansions in one direction, but in *all* directions, simply by expanding the sum and matrix sizes in the order of unfolding. We will return to this unfolding in Sec. 3.3.

2.2 SIESTA

SIESTA is one of the large variety of DFT codes that solves the Kohn-Sham equations Eq. (2.14) [6]. The underlying methodology requires a substantial effort to fully grasp, hence a short introduction of the method is outlined here.

The basis set only utilise the valence electrons by absorbing the core electrons in an effective potential using norm-conserving pseudo potentials. This choice of describing the core electrons via effective potentials drastically reduces storage and computational requirements and is widely used in numerous popular DFT codes [7–9, 16, 67, 68]. The basis set is based on a linear combination of atomic orbitals (LCAO)

$$\phi_{\alpha lmn}(\mathbf{r}) = \phi_{\alpha ln}(\mathbf{r} - \mathbf{r}_\alpha) Y_{lm}\left(\frac{\mathbf{r} - \mathbf{r}_\alpha}{|\mathbf{r} - \mathbf{r}_\alpha|}\right) \tag{2.25}$$

where α is an atomic index, l, m the angular momentum and n the multiple basis index for orbitals with same angular momentum but differing radial dependence, these multiple similar orbitals are referred to as *multiple-*ζ basis functions[6]. Y_{lm} is the corresponding spherical harmonic which is a collection of orthogonal polynomials. The choice of basis orbitals thus equal that of free atoms in vacuum. The atomic orbitals are confined *on* the atom with a decaying tail into vacuum. This natural decay limits the range and creates a sparse Hamiltonian with non-zero elements only where two basis orbitals overlap. The use of LCAO orbitals have several advantages; 1) the Hamiltonian becomes sparse and a natural partitioning of the elements can be performed, 2) a sparse Hamiltonian can enable order-N methods due to the fast internal data structures of a sparse matrix, 3) the accuracy can be controlled by *i*) increasing the radial dependence, *ii*) adding additional ζ orbitals for the same angular momentum with differing radial dependence and/or *iii*) adding polarisation orbitals to account for the deformation induced by bond formation, here we denote polarisation orbitals as $\zeta^{\pm}$. The disadvantage of this choice is that the accuracy is not a simple parameter but a multitude of parameters, including the ones mentioned above.

The LCAO basis is a non-orthogonal basis set resulting in a *so called* overlap matrix

$$\mathbf{S}_{\mu\nu} = \langle \phi_\nu | \phi_\mu \rangle, \tag{2.26}$$

where for simplicity the complex indices for the ϕ orbitals have been reduced to single Greek letters. The overlap matrix defines the dual basis of the basis orbitals by $|\phi^\nu\rangle =$

[6] The basis functions are named; single-ζ, double-ζ, etc.

$|\phi_\mu\rangle \mathbf{S}^{-1}_{\mu\nu}$, hence $\langle\phi^\nu|\phi_\mu\rangle = \delta^\nu_\mu$. The Kohn-Sham orbitals are then expanded in terms of the atomic orbitals by

$$\psi_i(\mathbf{r}) = \sum_\mu \phi_\mu(\mathbf{r}) c_{\mu i}, \tag{2.27}$$

$$c_{\mu i} = \langle\phi^\mu|\psi_i\rangle = \mathbf{S}^{-1}_{\nu\mu}\langle\phi_\mu|\psi_i\rangle, \tag{2.28}$$

where we have used that the overlap matrix is Hermitian. The above equations finally describes the non-orthogonal Kohn-Sham eigenvalue problem with coefficients $c_{\nu i}$ being the generalised eigenvector of $\mathbf{H}$ and $\mathbf{S}$

$$\sum_\nu \mathbf{H}_{\mu\nu} c_{\nu i} = \varepsilon_i \sum_\nu \mathbf{S}_{\mu\nu} c_{\nu i}. \tag{2.29}$$

This also changes the calculation of the density slightly via

$$\rho(\mathbf{r}) = f_i \sum_i |\psi_i(\mathbf{r})|^2 = f_i \sum_i \sum_{\nu\mu} c^*_{\nu i} \phi^*_\nu(\mathbf{r}) \phi_\mu(\mathbf{r}) c_{\mu i} \tag{2.30}$$

$$\int \mathrm{d}\mathbf{r}\rho(\mathbf{r}) = N = f_i \sum_i \sum_{\nu\mu} c^*_{\nu i} \mathbf{S}_{\nu\mu} c_{\mu i}, \tag{2.31}$$

from which we define the *density matrix*

$$\boldsymbol{\rho}_{\nu\mu} \equiv f_i \sum_i c^*_{\nu i} c_{\mu i}, \tag{2.32}$$

which can express the density integral Eq. (2.31) in matrix form

$$\int \mathrm{d}\mathbf{r}\rho(\mathbf{r}) = N = \mathrm{Tr}[\boldsymbol{\rho}\mathbf{S}]. \tag{2.33}$$

Note that the density matrix is distribution normalised to N via f_i. Although the density matrix is a dense matrix the only required elements are those defined where the overlap matrix is non-zero, thus leaving the density matrix in a sparse format as well. The indexed density of atomic orbitals are the equivalent Mulliken populations [69] which conveniently divides the charge amongst atoms. Importantly it is difficult to fully associate charges to explicit atoms. Many different schemes exists which divides the atomic charges in various ways [70, 71].

2.2.1 Electrostatics

The Hartree electrostatic potential, Eq. (2.13), for the electrons is determined from the difference between the self-consistent electron density $\rho(\mathbf{r})$ and the neutral atom density[7] $\rho^{\mathrm{atom}}(\mathbf{r})$ which we denote $\delta\rho(\mathbf{r}) = \rho(\mathbf{r}) - \rho^{\mathrm{atom}}(\mathbf{r})$. The Poisson equation solved in SIESTA can then be written as

$$\nabla^2 \delta V(\mathbf{r}) = -\delta\rho(\mathbf{r})/\epsilon_0. \tag{2.34}$$

SIESTA uses periodic boundary conditions in all directions and thus enables the use of a 3D Fourier transform for solving the Poisson equation. Further details may be found

[7] Determined from the basis orbitals and the neutral atom's valence population.

in [6]. However, this adds some peculiarities for certain systems. For charge neutral systems ($\int \delta\rho(\mathbf{r}) = 0$) the Hartree potential is easily calculated, $\nabla^2 \delta V(\mathbf{r}) = \nabla^2 \delta V_0(\mathbf{r}) = -\delta\rho(\mathbf{r})/\epsilon_0$. For charged systems the potential is obtained by adding a constant term which corresponds to a uniform charge background that compensates the excess charge, i.e. $\delta V(\mathbf{r}) = \delta V_0(\mathbf{r}) + \delta V_{\text{back}}$. This is a consequence of the Fourier transform which has a zero wave vector component corresponding to the uniform background. For slab calculations a vacuum is required to electrostatically decouple its periodic replica, and using a slab dipole correction may be a requirement [72]. This is because the periodic Fourier transforms does not impose Dirichlet[8] boundary conditions at infinity. However, for LCAO DFT vacuum is inexpensive, contrary to plane wave codes.

[8] Dirichlet: Fixed boundary potential. Neumann: Fixed potential derivative at the boundary.

Chapter 3

Green function techniques

Green function[1] is a general mathematical tool which we here use in a quantum mechanical picture. In this Chapter we will highlight a useful abstraction of the quantum mechanical equations through the usage of Green function techniques.

This thesis use Green functions to describe non-equilibrium effects via, either a self-consistent mean-field approach or from by quantifying the quantum transport properties, or both. Its usage as a replacement for diagonalisation routines is outlined with respect to the 1D chain introduced in the previous Chapter. The 1D chain will be used to show how one can calculate the density of states using the Green function and compare this to the eigenvalue problem shown in the previous Chapter. Also, the concept of a *self-energy* is introduced and its importance and usefulness is again highlighted for the 1D chain.

This Chapter will not introduce the Green function technique of calculating the transport problem. For this we refer to other texts, such as [55, 74, 75].

3.1 Single particle Green function

We start with the unperturbed retarded single particle Green function ($\mathbf{G}^0$)

$$\Big[(E+\mathrm{i}\eta)\mathbf{I}-\mathbf{H}_\mathbf{k}\Big]\mathbf{G}^0_\mathbf{k}(E)=\mathbf{I}, \quad \text{with } \eta\to 0^+, \tag{3.1}$$

with $\mathbf{k}$ denoting the periodicity in the Brillouin zone. The 0 index is used to denote the unperturbed Green function. Importantly the identity matrix, $\mathbf{I}$, is used for an orthogonal basis set while the overlap matrix, $\mathbf{S}$, is used for a non-orthogonal basis set. Either representation is valid while in the remaining section we will use an orthogonal representation without loss of generality. We can rewrite Eq. (3.1) in terms of the eigenstates of the Hamiltonian as

$$\mathbf{G}^0_\mathbf{k}(E)=\sum_i^N \frac{|\psi_{\mathbf{k},i}\rangle\langle\psi_{\mathbf{k},i}|}{E+\mathrm{i}\eta-\varepsilon_{\mathbf{k},i}}, \tag{3.2}$$

where Eq. (3.1) can easily be asserted. We will omit the total number of orbitals N in the following for clarity. It is instructive to split the Green function into the real and

[1] The community has still not converged on *Green* vs. *Green's* function [73]. I prefer Green.

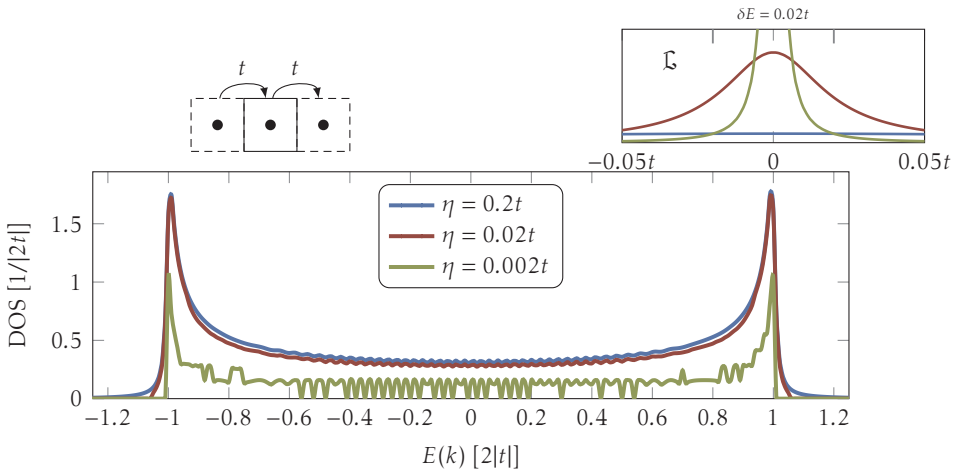


Figure 3.1 | DOS of linear chain (above left) calculated using different η values for fixed $\mathbf{k}$ and E spacing. δE is shown in top right zoom of $\mathfrak{L}$. The wiggles appear due to inadequate linear integration of $\mathfrak{L}_{\mathbf{k},i}(E)$. This is because of a lower η. This is clear in the $\eta = 0.002t$ plot.

imaginary part

$$\mathfrak{Re}\,\mathbf{G}^0_{\mathbf{k}}(E) = \sum_i |\psi_{\mathbf{k},i}|^2 \mathfrak{L}_{\mathbf{k},i}(E)(E - \varepsilon_{\mathbf{k},i})/\eta \tag{3.3}$$

$$\mathfrak{Im}\,\mathbf{G}^0_{\mathbf{k}}(E) = -\sum_i |\psi_{\mathbf{k},i}|^2 \mathfrak{L}_{\mathbf{k},i}(E) \tag{3.4}$$

$$\mathfrak{L}_{\mathbf{k},i}(E) \equiv \frac{\eta}{(E - \varepsilon_{\mathbf{k},i})^2 + \eta^2}, \tag{3.5}$$

where for $\eta \to 0^+$ we find an infinitely narrow Lorentzian ($\mathfrak{L}_{\mathbf{k},i}(E)$), a δ-function, with an area of π and full width half maximum 2η. Hence the density of states ($|\psi_{\mathbf{k},i}|^2$) can be expressed as

$$\mathrm{DOS}(E) = -\frac{1}{\pi}\,\mathrm{Tr}[\mathfrak{Im}\,\mathbf{G}^0_{\mathbf{k}}(E)] \tag{3.6}$$

$$\mathrm{LDOS}(E) = -\frac{1}{\pi}\,\mathfrak{Im}\,\mathbf{G}^0_{\mathbf{k},\{ii\}}(E), \quad \text{local at site } i. \tag{3.7}$$

Note that for non-orthogonal basis sets a similar expression to Eq. (2.33) is used. The density of states yields the number of states at energy E with non-negligible contributions from eigenstates with eigenenergies $|E - \varepsilon| \lesssim 2\eta$. As $\eta \to 0$ the density of states turns into delta functions at the eigenenergies. Note that integrating the DOS in the full energy spectrum will *always* be a normalised quantity, N, as can be inferred from the identity $\mathbf{I} = \sum_i |\psi_i\rangle\langle\psi_i|$. While the total occupied orbitals can be obtained by integrating a product of the Fermi function, $n_F(E)$, and DOS(E).

It is instructive to calculate the energy resolved DOS for a 1D periodic chain using the single particle Green function method shown above. The system setup is shown in Figure 3.1 top-left with hopping parameter t between the sites. We calculate the DOS by integrating the Brillouin zone in a fine mesh of 400 points as well as an energy grid spaced by $\delta E = 0.02t$. In Figure 3.1 the DOS is shown for three values of η, for a high

$\eta > \delta E$ it is relatively smooth as would be expected if the Lorentzians are fully described by discretised integrals. Furthermore we clearly find the van-Hove singularities for $\nabla_{\mathbf{k}} \to \mathbf{0}$ at the band onset and band offset. For a slightly lower $\eta = \delta E$ it almost does not change. Yet, for even lower $\eta < \delta E$ the DOS becomes non-recognisable and the general 1D DOS cannot be recovered. The above discussion for δE also applies to the discretisation of $\mathbf{k}$ via the eigenvalue difference for $\delta\mathbf{k}$. This emphasise three points which we suggests to be simultaneously fulfilled to adequately describe the DOS via Green functions

1. choose $\eta \geq \delta E$ to correctly capture band features in energy space,
2. then choose $\delta\mathbf{k}$ such that $\eta \geq \delta E|_{\mathbf{k}+\delta\mathbf{k}}$ to correctly capture band features in reciprocal space.
3. Choose η as low as possible to correctly describe the Lorentzian features which would otherwise get too broadened.

Here $\delta E|_{\mathbf{k}+\delta\mathbf{k}}$ is the maximum change in energy due to a shift in wavenumber for any $\mathbf{k}$. It should be recognised that a multitude of post-processing utilities for calculating density of states using *only* the band-structure calculations rely on Eqs. (3.2) and (3.6) and hence the above arguments should also be considered for seemingly non-Green function calculations.

To recapitulate, the density of states can be a) performed using pristine Green function techniques, or b) using diagonalisation of the Hamiltonian equation and post-process the DOS according to Eq. (3.4) from the bandstructure. The latter is typically preferred as you need only a Brillouin zone discretisation while the former requires *both* a Brillouin and an energy space discretisation. Furthermore the latter is guaranteed to capture all states as you explicitly calculate the eigenstates. Hence the majority of DFT codes employ the latter for periodic calculations.

The following section describes how the Green function method can be used without performing a double integral in a similar context via the use of self-energies.

3.2 Bulk Green function

The discussion in Sec. 3.1 suggests that Green functions are not suitable for periodic systems as it requires a double integral (E and $\mathbf{k}$) for estimating a correct DOS[2]. However, the Green function technique possesses other methods to circumvent such difficulties. In particular the term *self energy* is important. We will in the following omit the energy dependence and only highlight its dependence when deemed necessary.

We will start by introducing a perturbation to the single particle Green function and define the perturbed Green function ($\mathbf{G}$)

$$[(E + \mathrm{i}\eta)\mathbf{I} - \mathbf{H} - \mathbf{V}]\mathbf{G} = \mathbf{I}, \tag{3.8}$$

with $\mathbf{V}$ being the perturbation. Rewriting Eq. (3.1) and Eq. (3.8)

$$[(E + \mathrm{i}\eta)\mathbf{I} - \mathbf{H} - \mathbf{V}]\mathbf{G} = \mathbf{I} \tag{3.9}$$

$$[(E + \mathrm{i}\eta)\mathbf{I} - \mathbf{H}]\mathbf{G} = \mathbf{I} + \mathbf{V}\mathbf{G} \tag{3.10}$$

$$\mathbf{G} = \mathbf{G}^0 + \mathbf{G}^0\mathbf{V}\mathbf{G}, \tag{3.11}$$

[2]Compared to the diagonalisation method which only requires a single integral.

which is known as the *Dyson* equation. The perturbation $\mathbf{V}$ can be of any kind, local impurity, neighbouring impurity, neighbouring cell, infinite bulk etc. To illustrate its usefulness we first describe a 2 orbital system and then extend it to our 1D chain.

$$\mathbf{H} = \begin{bmatrix} \epsilon & 0 \\ 0 & \epsilon \end{bmatrix} \quad ; \quad \mathbf{V} = \begin{bmatrix} 0 & t \\ t & 0 \end{bmatrix}. \tag{3.12}$$

We first consider the perturbation of $\mathbf{H}_{\{1,1\}}$ due to the neighbouring site $\mathbf{H}_{\{2,2\}}$

$$\mathbf{G}_{\{1,1\}} = \mathbf{G}^0_{\{1,1\}} + \mathbf{G}^0_{\{1,1\}}\mathbf{V}_{\{1,2\}}\mathbf{G}_{\{2,1\}}, \tag{3.13}$$

from the matrix product in Eq. (3.11) we take $\mathbf{G}_{\{2,1\}} = \mathbf{G}^0_{\{2,2\}}\mathbf{V}_{\{2,1\}}\mathbf{G}_{\{1,1\}}$ and insert

$$\mathbf{G}_{\{1,1\}} = \mathbf{G}^0_{\{1,1\}} + \mathbf{G}^0_{\{1,1\}}\mathbf{V}_{\{1,2\}}\mathbf{G}^0_{\{2,2\}}\mathbf{V}_{\{2,1\}}\mathbf{G}_{\{1,1\}} \tag{3.14}$$

$$\mathbf{G}_{\{1,1\}} = \left[E + \mathrm{i}\eta - \epsilon - \mathbf{V}_{\{1,2\}}\mathbf{G}^0_{\{2,2\}}\mathbf{V}_{\{2,1\}}\right]^{-1} \tag{3.15}$$

$$\mathbf{G}_{\{1,1\}} \equiv \left[E + \mathrm{i}\eta - \epsilon - \Sigma(E)\right]^{-1} \tag{3.16}$$

Here $\mathbf{V}_{\{1,2\}}\mathbf{G}^0_{\{2,2\}}\mathbf{V}_{\{2,1\}}$ is typically referred to as the self energy $\Sigma(E)$ with an energy dependence due to $\mathbf{G}^0_{\{2,2\}}$. The self energy can be understood in terms of two aspects, 1) the eigenstate energies (ϵ) are re-normalised and, 2) the life-time of the energy levels are broadened from the imaginary part of the self energy. An important aspect of the self energy is that the range of the normalisation terms correspond to the range of $\mathbf{V}$ that couples the external perturbation into the system of consideration, hence the coupling of the self energy must not go beyond the perturbed Green function.

The above example highlights an important feature of the Green function method, namely

> The solution of a system under an external perturbation *only* requires the self energy of the perturbation.
>
> An infinite system can be solved using a one-particle basis and the self energy on the one-particle due to the remaining infinite system.

The last paragraph might seem contradictory in that we still need to calculate the self energy of an infinite system? However, we will now show that it is, relatively, easy to calculate the self energy of an infinite perturbation for certain infinite systems.

We start by showing the steps of calculating the surface self energy corresponding to a particle connected to a bulk in *one* direction (hence surface). An infinite system can be split into two surfaces, hence calculating the *left* and *right* surface self energy is equivalent to one infinite chain. The Hamiltonian of a *right* semi infinite system

$$\mathbf{H} = \begin{bmatrix} \epsilon & \mathbf{0} \\ \mathbf{0} & \ddots \end{bmatrix} \quad ; \quad \mathbf{V} = \begin{bmatrix} 0 & t & \mathbf{0} \\ t & \ddots & \ddots \\ \mathbf{0} & \ddots & \ddots \end{bmatrix}. \tag{3.17}$$

We quickly repeat the calculations done in Eq. (3.13)

$$\mathbf{G}_{\{1,1\}} = \mathbf{G}^0_{\{1,1\}} + \mathbf{G}^0_{\{1,1\}}\mathbf{V}_{\{1,2\}}\mathbf{G}_{\{2,1\}} \tag{3.18}$$
$$\mathbf{G}_{\{2,1\}} = \mathbf{G}^0_{\{1,1\}}\mathbf{V}_{\{2,1\}}\mathbf{G}_{\{1,1\}} + \mathbf{G}^0_{\{1,1\}}\mathbf{V}_{\{2,3\}}\mathbf{G}_{\{3,1\}} \tag{3.19}$$
$$\vdots \tag{3.20}$$
$$\mathbf{G}_{\{n,1\}} = \mathbf{G}^0_{\{1,1\}}\mathbf{V}_{\{n,n-1\}}\mathbf{G}_{\{n-1,1\}} + \mathbf{G}^0_{\{1,1\}}\mathbf{V}_{\{n,n+1\}}\mathbf{G}_{\{n+1,1\}}. \tag{3.21}$$

Where $\mathbf{V}_{\{1,2\}}$ determines the size of the so-called *primary layer*. I.e. the size of the primary layer only couples to one neighbouring layer in each direction of the bulk part. The above recursive equations enables the calculation of the surface self energy [76] which can be written as

$$\mathbf{G}_{\{1,1\}} = \mathbf{G}^0_{\{1,1\}} + \mathbf{G}^0_{\{1,1\}}\mathbf{V}_{\{1,2\}}\mathbf{T}\mathbf{G}_{\{1,1\}} \tag{3.22}$$
$$\mathbf{G}_{\{1,1\}} = \left[E + \mathrm{i}\eta - \mathbf{H}_{\{1,1\}} - \mathbf{V}_{\{1,2\}}\mathbf{T}\right]^{-1}, \tag{3.23}$$

where we define the self energy as $\Sigma_{\{\infty,1\}} \equiv \mathbf{V}_{\{1,2\}}\mathbf{T}$. The self energy can be obtained through the following recursive equations which are terminated when the difference between two iteration's surface self energy goes to 0.

$$\Sigma_{\{1,1\}}(E) = \mathbf{V}_{\{1,2\}}\left[E + \mathrm{i}\eta - \mathbf{H}_{\{1,1\}}\right]\mathbf{V}_{\{2,1\}} \tag{3.24a}$$
$$\Sigma_{\{i,1\}}(E) = \mathbf{V}_{\{1,2\}}\left[E + \mathrm{i}\eta - \mathbf{H}_{\{1,1\}} - \Sigma_{\{i-1,1\}}(E)\right]\mathbf{V}_{\{2,1\}} \tag{3.24b}$$

and for large i the self energy will converge and the recursive scheme terminates. $\Sigma_{\{i,1\}}$ will then be the surface self-energy which rescales the surface Hamiltonian due to the underlying semi-infinite bulk. Although typically unused, the index i can be thought of as the range of the electrostatic potential by the bulk. The recursive scheme in Eqs. (3.24) is a slow method for calculating the self energy. It should *never* be the method of choice for semi-infinite bulk. We will return to this point after showing a much faster converging series for the surface self energy [76]. For this method it is convenient to use both the left and right self energies

$$\alpha_0 = \mathbf{V}_{\{1,2\}} = \beta_0^\dagger = \mathbf{V}^\dagger_{\{2,1\}} \tag{3.25a}$$
$$\alpha_i = \alpha_{i-1}\left[E + \mathrm{i}\eta - \mathbf{H}_{\{1,1\}} - \Sigma^L_{\{i-1,1\}} - \Sigma^R_{\{i-1,1\}}\right]^{-1}\alpha_{i-1} \tag{3.25b}$$
$$\beta_i = \beta_{i-1}\left[E + \mathrm{i}\eta - \mathbf{H}_{\{1,1\}} - \Sigma^L_{\{i-1,1\}} - \Sigma^R_{\{i-1,1\}}\right]^{-1}\beta_{i-1} \tag{3.25c}$$
$$\Sigma^L_{\{i,1\}}(E) = \Sigma^L_{\{i-1,1\}}(E) + \alpha_{i-1}\left[E + \mathrm{i}\eta - \mathbf{H}_{\{1,1\}} - \Sigma^L_{\{i-1,1\}} - \Sigma^R_{\{i-1,1\}}\right]^{-1}\beta_{i-1} \tag{3.25d}$$
$$\Sigma^R_{\{i,1\}}(E) = \Sigma^R_{\{i-1,1\}}(E) + \beta_{i-1}\left[E + \mathrm{i}\eta - \mathbf{H}_{\{1,1\}} - \Sigma^L_{\{i-1,1\}} - \Sigma^R_{\{i-1,1\}}\right]^{-1}\alpha_{i-1} \tag{3.25e}$$

for $i > 0$ and $\Sigma^{L/R}_{\{0,1\}} = \mathbf{0}$. Eqs. (3.25) converges with exponential scaling 2^i, i.e. the self energy describes a re-normalisation from the nearest 2^i layers as each iteration doubles the amount of interacting layers. For further details see [76].

Eqs. (3.24) and (3.25) can be used to describe the bulk properties of a linear chain by calculating both the left and right surface self energy and then calculate the bulk Green function. The DOS and the number of iterations required for the calculation of

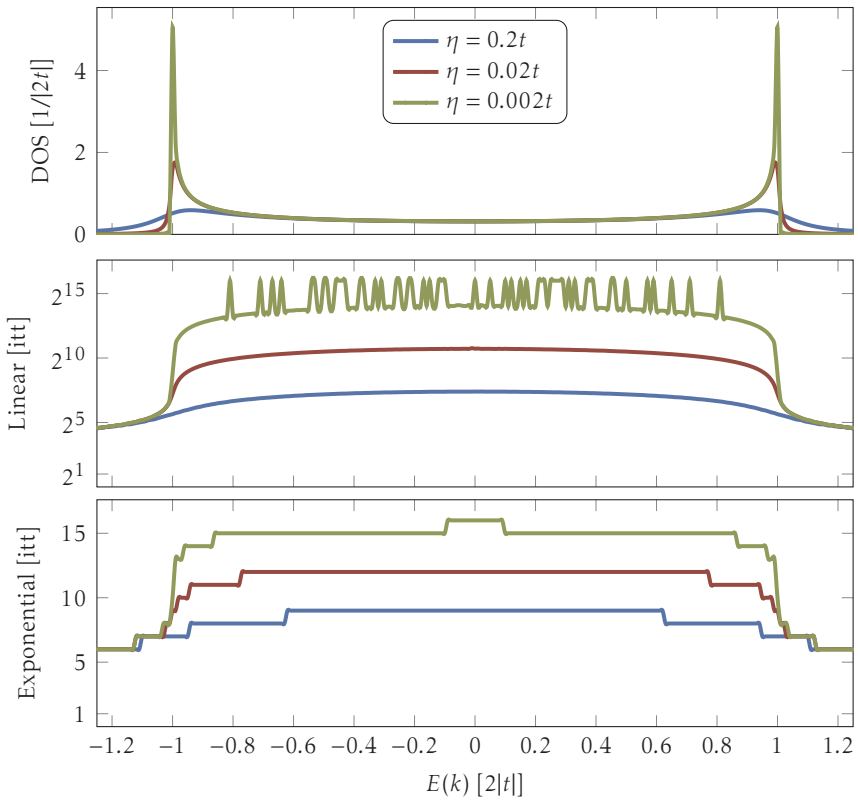


Figure 3.2 | Top) density of states from self energy perturbation via two semi infinite bulk parts. Middle) number of iterations for linear recursion algorithm, Eq. (3.24). y-axis in base 2. Bottom) iterations for the exponential recursion algorithm, Eq. (3.25). The linear uses $\sim 2^{15} = 32768$ iterations at $E = 0$ while the recursive converges in ~ 15 iterations. Clearly [76] converges *much* faster while they both yield the same DOS.

the self energy for the two methods are shown in Figure 3.2. The top plot shows the DOS from the bulk Green function. In the middle plot we see the number of iterations for the slow method with y-axis in base 2. The last plot shows the number of iterations required for the fast method [76]. By having the linear in base 2 they become comparable as they converge in roughly the same number of layers. By comparing Figure 3.1 to Figure 3.2 it is clearly seen that the self energy method captures the DOS perfectly at any energy, unlike the Brillouin zone integration which requires an extremely fine $\delta\mathbf{k}$ for a smooth DOS. This is because all $\mathbf{k}$ points along the semi infinite direction is taken into account and hence it corresponds to a smooth Brillouin zone integration along that direction. Note how a larger η reduces the number of interacting layers. For $\eta \to \infty$ this corresponds to a constant self energy and essentially a constant DOS smeared out from $E = -\infty$ to ∞.

Recapitulating, the surface self energy describes the *correction* on a primary unperturbed layer via a re-normalisation of the energy spectrum due to the correction from *all* particles in the semi infinite direction. In other words, it is equivalent to perturbing the primary layer by all Bloch states in the semi infinite direction and thus the $\mathbf{k}$ resolution along that direction becomes obsolete. It is remarkable that the perturbed Green function defined in the primary layer can now express the DOS for a semi-infinite region.

Currently the recursive method by Sancho *et al*. is the most frequently used due to

its simplicity, ease of use and strict convergence control. Incidentally it also converges relatively fast, even for long range potentials in the bulk $\lesssim 20$ iterations seem to be enough for reasonable η values. However, other methods which calculates the surface self energy in a one-step process are available [10, 77–81]. The earlier methods involve the inversion of a possible singular matrix ($\mathbf{H}_{\{1,2\}}$) leaving them unusable for certain systems [10, 77, 78, 80] while a later method solved the singularity problem [81]. They all utilise a quadratic eigenvalue problem of the form similar to this (here shown from [81])

$$\mathbf{K}_{\{i,j\}} = (E + \mathrm{i}\eta)\mathbf{S}_{\{i,j\}} - \mathbf{H}_{\{i,j\}} \tag{3.26}$$

$$\begin{bmatrix} \mathbf{K}_{\{1,1\}} & \mathbf{K}_{\{2,1\}} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{\Phi}_{+,n} = e^{\mathrm{i}\mathbf{k}_n} \begin{bmatrix} \mathbf{K}_{\{1,2\}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \mathbf{\Phi}_{+,n}. \tag{3.27}$$

While the calculation of the self energy does take a substantial amount of time one can with benefit pre-calculate the self energy and store it to be used later at any time. For this reason the used method in this thesis is the Sancho *et al*. recursive method and storing them on disk for reducing calculations.

3.3 Bloch expansion of the self energy

From Sec. 3.2 a direct way of coupling a system to an infinite perturbation has been presented. Yet, the recursive scheme Eqs. (3.25) are still requiring a decent number of inversions making it a substantial part of any Green function calculation. The infinite perturbations will typically describe truly bulk parts such as metals with transverse periodicities that couples an integral periodic bulk system to a much larger system. Here we, again, utilise Bloch's theorem to reduce computational overhead. As explained in Sec. 2.1.2 the periodicity can easily save computations by using a reduced unit-cell and then unfolding it using Eq. (2.24). Particularly for the self energy calculation where several inversions are made this will greatly increase performance. For clarity we re-write the self energy unfolding here

$$\Sigma^n_{k_n} = \frac{1}{n} \sum_{\substack{i \\ k_i = k_n + 2\pi \frac{i-1}{nR}}}^{n} \Sigma^1_{k_i} \cdot \begin{bmatrix} \mathbf{I} & e^{-\mathrm{i}k_i R} & \cdots & e^{-\mathrm{i}nk_i R} \\ e^{\mathrm{i}k_i R} & \mathbf{I} & \cdots & e^{-\mathrm{i}(n-1)k_i R} \\ \vdots & \vdots & \ddots & \vdots \\ e^{\mathrm{i}nk_i R} & e^{\mathrm{i}(n-1)k_i R} & \cdots & \mathbf{I} \end{bmatrix}, \tag{3.28}$$

where the variables are equivalently explained as for Eq. (2.24). To the author's knowledge this is only implemented in TranSIESTA while other codes could greatly benefit from this simple abstraction of the self energy calculation, even in the direct calculation of the self-energies. This will drastically increase throughput in calculations as one will always be able to reduce the bulk self energy calculations to the minimal bulk periodic cell.

Chapter 4

TranSIESTA

In this Chapter we cover the full re-implementation of TranSIESTA which is a DFT+NEGF code. The non-equilibrium Green function (NEGF) method, on which TranSIESTA is based, is one theoretical model which has successfully described experimental findings within many areas. It is a widely used method in the community and several codes based on DFT also implements this approach. A selection of NEGF codes based on SIESTA can be found in these Refs. [10–12, 81, 82]. Other NEGF codes, see Refs. [15, 16, 50, 51] are based on other DFT software. Importantly, the NEGF method is not constrained to DFT. In particular it is heavily used in tight-binding studies [39, 83, 84]. Also parameterisations of DFT, such as Refs. [14, 85] implement DFT+NEGF. Altogether, only few codes handles more than two electrodes [14, 15, 50, 51].

We will introduce the basics of TranSIESTA, starting from calculating the density matrix using Green functions. For multi-electrode calculations these equations are carefully written and we highlight important considerations when conducting multi-electrode calculations. Optimisations of the equilibrium contour as well as the different schemes involving non-equilibrium density averaging will also be covered. A large part of this Chapter is devoted to the algorithms and the performance enhancements achieved in this re-implementation. We emphasise the algorithm in context of $N_{\mathfrak{e}} > 2$ calculations which complicates the algorithms.

4.1 Density matrix from Green function

The following discussion borrows techniques discussed in Secs. 2.2 and 3.1. Symmetries will be implicitly used in the context of time-reversal symmetry and all relevant symmetries may be found in App. A. In the following we consider the Kohn-Sham orbitals as mean-field one-electron orbitals used to describe non-equilibrium effects. The E_{xc} functional is assumed adequate for describing such non-equilibrium effects in conjunction with the Kohn-Sham orbitals. We limit this section to not consider many-body effects and/or electron-phonon couplings. Also, the equations are easily generalised for co-linear spin. Note that some expansions of the mean-field approximation may be described using additional self-energy terms.

For clarity and later use, an explicit $\mathbf{k}$ dependence is added to the equations while $\mathfrak{e}$ refers to an electrode index. Furthermore all equations are written generically with $N_{\mathfrak{e}}$ electrodes to clarify specifics related for any number of electrodes $N_{\mathfrak{e}} \geq 1$. The following

expressions are heavily used throughout the remainder of the thesis

$$\mathbf{G}_{\mathbf{k}}(z) = \left[z\mathbf{S}_{\mathbf{k}} - \mathbf{H}_{\mathbf{k}} - \sum_{\mathfrak{e}} \Sigma_{\mathfrak{e},\mathbf{k}}(z-\mu_{\mathfrak{e}})\right]^{-1} \quad \text{, with } z \equiv \epsilon + i\eta. \tag{4.1}$$

$$\mathbf{\Gamma}_{\mathfrak{e},\mathbf{k}}(z) = i\left(\Sigma_{\mathfrak{e},\mathbf{k}}(z-\mu_{\mathfrak{e}}) - \Sigma^{\dagger}_{\mathfrak{e},\mathbf{k}}(z-\mu_{\mathfrak{e}})\right) \tag{4.2}$$

$$\mathcal{A}_{\mathfrak{e},\mathbf{k}}(z) = \mathbf{G}_{\mathbf{k}}(z)\mathbf{\Gamma}_{\mathfrak{e},\mathbf{k}}(z)\mathbf{G}^{\dagger}_{\mathbf{k}}(z) \tag{4.3}$$

$$\widetilde{\mathcal{A}}_{\mathfrak{e},\mathbf{k}}(z) = \mathbf{G}^{\dagger}_{\mathbf{k}}(z)\mathbf{\Gamma}_{\mathfrak{e},\mathbf{k}}(z)\mathbf{G}_{\mathbf{k}}(z) \tag{4.4}$$

$$\rho = \frac{1}{2\pi}\iint_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\mathrm{d}\epsilon \sum_{\mathfrak{e}} \mathcal{A}_{\mathfrak{e},\mathbf{k}}(z) n_{F,\mathfrak{e}}(\epsilon) e^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}}, \tag{4.5}$$

with the usual quantities. $\Sigma_{\mathfrak{e}}$ and $\mathcal{A}_{\mathfrak{e}}(\widetilde{\mathcal{A}}_{\mathfrak{e}})$ are the self energy and spectral function (time-reversed) of electrode $\mathfrak{e}$, respectively with associated scattering matrix. ρ is the non-equilibrium density matrix. BZ denotes an integration over the Brillouin zone with $\mathbf{k}$ being the Brillouin zone $\mathbf{k}$-point. ϵ and η are the energy and 0^+ small positive constant, respectively. Eq. (4.5) is the standard form of the density matrix for equilibrium *and* non-equilibrium. The chemical potential is denoted μ, and the temperature as kT. A combined quantity for the chemical potential and the associated temperature is denoted $\varsigma \equiv \{\mu, kT\}$. We will freely denote a Fermi distribution by $n_{F,\varsigma}$ as well as $n_{F,\mathfrak{e}}$ where the latter implicitly refers to ς belonging to the electrode $\mathfrak{e}$. Lastly, the so called "transport direction" is omitted in the following as it can only be defined with respect to a plane. For $N_{\mathfrak{e}} > 2$ or misaligned electrodes no single transport direction can thus be defined. As such the terminology for NEGF calculations should be given with respect to the semi-infinite directions. This is now the convention in TranSIESTA.

TranSIESTA also calculates the *energy density matrix* to enable force calculations via NEGF [16, 86]

$$\mathcal{E} = \frac{1}{2\pi}\iint_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\mathrm{d}\epsilon\, \epsilon \sum_{\mathfrak{e}} \mathcal{A}_{\mathfrak{e},\mathbf{k}}(z) n_{F,\mathfrak{e}}(\epsilon) e^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}}. \tag{4.6}$$

In the following any derivation of the density matrix equally holds for the energy density matrix with the additional ϵ multiplication in the integral. It is stressed that the forces in a steady-state transport problem can be defined equivalently to a closed system at equilibrium [87].

Remark that the definition of the Hamiltonian in the Brillouin zone is that of Eq. (2.21). This choice has several advantages:

- reduces several computations as the phases can be pre-calculate for each auxiliary cell instead of phases determined by inter-atomic distances,
- the inter-atomic distances are not needed so long as you know which Hamiltonian element belongs to which auxiliary super cell,
- you allow *any* unit cell, even a semi-infinite direction which is not orthogonal to the periodic directions: $\mathbf{R}_{\infty} \cdot \mathbf{R} \neq 0$,
- periodicity of the device region can exist while a semi-infinite direction has the same periodic direction.

We note that the calculation of the self energies are typically pre-calculated and stored on disk to reduce overhead of re-calculating the same quantity for each self

consistent cycle. Further, we use Eq. (3.28) to considerably speed up the calculation of the surface self energy, as discussed in Sec. 3.3. Hence the calculation of the self energies will in nearly all cases be a small fraction of the total computation time and focus will be on algorithms regarding the Green function calculation.

4.1.1 Equilibrium (EGF)

By assuming all chemical potentials equal in Eq. (4.5) we rewrite

$$\rho = \frac{1}{2\pi}\iint_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\mathrm{d}\epsilon\, \mathbf{G}_{\mathbf{k}} \sum_{\mathfrak{c}} \mathbf{\Gamma}_{\mathfrak{c},\mathbf{k}} \mathbf{G}_{\mathbf{k}}^{\dagger} e^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}} n_F(\epsilon) \tag{4.7}$$

$$\rho = \frac{i}{2\pi}\iint_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\mathrm{d}\epsilon\, \mathbf{G}_{\mathbf{k}} \left[\Sigma_{\mathbf{k}} - \Sigma_{\mathbf{k}}^{\dagger}\right] \mathbf{G}_{\mathbf{k}}^{\dagger} e^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}} n_F(\epsilon), \tag{4.8}$$

by introducing Eq. (4.1)

$$\rho = \frac{i}{2\pi}\iint_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\mathrm{d}\epsilon\, \mathbf{G}_{\mathbf{k}} \left[\mathbf{G}_{\mathbf{k}}^{\dagger,-1} - \mathbf{G}_{\mathbf{k}}^{-1} + 2i\eta\mathbf{S}_{\mathbf{k}}\right] \mathbf{G}_{\mathbf{k}}^{\dagger} e^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}} n_F(\epsilon) \tag{4.9}$$

$$\rho = \frac{i}{2\pi}\iint_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\mathrm{d}\epsilon \left[\mathbf{G}_{\mathbf{k}} - \mathbf{G}_{\mathbf{k}}^{\dagger} + 2i\eta\mathbf{G}_{\mathbf{k}}\mathbf{S}_{\mathbf{k}}\mathbf{G}_{\mathbf{k}}^{\dagger}\right] e^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}} n_F(\epsilon) \tag{4.10}$$

$$\approxeq \frac{i}{2\pi}\iint_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\mathrm{d}\epsilon \left[\mathbf{G}_{\mathbf{k}} - \mathbf{G}_{\mathbf{k}}^{\dagger}\right] e^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}} n_F(\epsilon) \equiv \rho_{\mathrm{eq}}, \tag{4.11}$$

only for the Γ point does Eq. (4.11) reduce to $i(\mathbf{G}_{\Gamma} - \mathbf{G}_{\Gamma}^{\dagger}) = -2\,\mathrm{Im}\,\mathbf{G}_{\Gamma}$. We define Eq. (4.11) as the equilibrium density ρ_{eq} belonging to the chemical potential n_F.

4.1.2 Non-equilibrium (NEGF)

Non-equilibrium arises due to differences between the electrode electronic distributions via $\varsigma_{\mathfrak{c}} \neq \varsigma_{\mathfrak{c}'}$, either via the chemical potential or via temperature differences. The NEGF part is the most complex part of the equations as it requires the Green function *and* a triple matrix product. We start from Eq. (4.5) and a suitably chosen 0 term

$$\rho = \frac{1}{2\pi}\iint_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\mathrm{d}\epsilon\, \mathbf{G}_{\mathbf{k}} \sum_{\mathfrak{c}} \mathbf{\Gamma}_{\mathfrak{c},\mathbf{k}} n_{F,\mathfrak{c}}(\epsilon) \mathbf{G}_{\mathbf{k}}^{\dagger} e^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}} \tag{4.12}$$

$$(1-1)\mathbf{G}_{\mathbf{k}} \sum_{\mathfrak{c}'\neq\mathfrak{c}} \mathbf{\Gamma}_{\mathfrak{c}',\mathbf{k}} n_{F,\mathfrak{c}}(\epsilon) \mathbf{G}_{\mathbf{k}}^{\dagger} e^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}} \quad \text{, note the Fermi function.} \tag{4.13}$$

Adding Eq. (4.12) and Eq. (4.13) and reducing yields

$$\rho^{\mathfrak{c}} = \rho_{\mathrm{eq}}^{\mathfrak{c}} + \frac{1}{2\pi}\sum_{\mathfrak{c}'\neq\mathfrak{c}}\iint_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\mathrm{d}\epsilon\, \mathbf{G}_{\mathbf{k}} \mathbf{\Gamma}_{\mathfrak{c}',\mathbf{k}} \mathbf{G}_{\mathbf{k}}^{\dagger} e^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}} \left[n_{F,\mathfrak{c}'}(\epsilon) - n_{F,\mathfrak{c}}(\epsilon)\right], \tag{4.14}$$

$$\rho^{\mathfrak{c}} \equiv \rho_{\mathrm{eq}}^{\mathfrak{c}} + \sum_{\mathfrak{c}'\neq\mathfrak{c}} \Delta_{\mathfrak{c}'}^{\mathfrak{c}} \equiv \rho_{\mathrm{eq}}^{\mathfrak{c}} + \rho_{\mathrm{neq}}^{\mathfrak{c}}, \tag{4.15}$$

where we call $\Delta_{\mathfrak{c}'}^{\mathfrak{c}}$ a non-equilibrium correction term for the equilibrium density of electrode $\mathfrak{c}$ due to electrode $\mathfrak{c}'$. This reduce the non-equilibrium correction term integrals to be confined in the bias window with respect to the different Fermi distributions. Eq. (4.15) deserves a few comments. It can be expressed equivalently for all electrodes $\mathfrak{c}$, and thus one finds $N_{\mathfrak{c}}$ different expressions for the same quantity $\rho = \rho^{\mathfrak{c}} = \rho^{\mathfrak{c}'}$ as shown

in [15]. If any two or more electrodes have the same Fermi distribution, $\varsigma_{\mathfrak{e}} = \varsigma_{\mathfrak{e}'}$, we have $\boldsymbol{\rho}^{\mathfrak{e}}_{\mathrm{eq}} = \boldsymbol{\rho}^{\mathfrak{e}'}_{\mathrm{eq}}$ and $\boldsymbol{\Delta}^{\mathfrak{e}}_{\mathfrak{e}'} = 0$, whence we can reduce $N_{\mathfrak{e}}$ to N_{ς} different expressions. So for any $N_{\mathfrak{e}} > 2$ electrodes with 2 different Fermi distributions we only have 2 equations made of different terms (although they are mathematically equivalent). The number of correction terms for $\mathfrak{e}$ is equal to the number of electrodes with $\varsigma_{\mathfrak{e}} \neq \varsigma_{\mathfrak{e}'}$. Equivalently, and more rigorously, Eq. (4.15) can be written as

$$\boldsymbol{\rho}^{\varsigma} = \boldsymbol{\rho}^{\varsigma}_{\mathrm{eq}} + \frac{1}{2\pi} \sum_{\mathfrak{e}|\varsigma_{\mathfrak{e}}\neq\varsigma} \iint_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\mathrm{d}\epsilon\, \mathbf{G}_{\mathbf{k}} \boldsymbol{\Gamma}_{\mathfrak{e},\mathbf{k}} \mathbf{G}^{\dagger}_{\mathbf{k}} e^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}} \Big[n_{F,\varsigma_{\mathfrak{e}}}(\epsilon) - n_{F,\varsigma}(\epsilon) \Big] = \boldsymbol{\rho}^{\varsigma}_{\mathrm{eq}} + \sum_{\mathfrak{e}|\varsigma_{\mathfrak{e}}\neq\varsigma} \boldsymbol{\Delta}^{\varsigma}_{\mathfrak{e}}, \quad (4.16)$$

where $\mathfrak{e}|\varsigma_{\mathfrak{e}} \neq \varsigma$ are electrodes with Fermi distributions different from ς. Eq. (4.15) is equivalent to Eq. (4.16) where the former have possible duplicates and the latter does not.

Different implementations either calculate one or all $\boldsymbol{\rho}^{\varsigma}_{\mathrm{eq}}$. The former is straight forward while the latter requires an averaging of the quantities. For best convergence properties the latter method is preferred.

4.2 Discretised integration

A problem with integrating Eqs. (4.11) and (4.16) is that the precision of the integral is determined by the fineness of the energy contour. The discreteness of the integration has to resolve the possibly very narrow eigenstates along the real axis.

To circumvent the meticulous and tedious calculation of the EGF along the energy real axis, the integration variable, ϵ, is changed to a complex energy, $z = \epsilon + i\eta$, which moves the Green function far into the complex plane. The advantage is that the Green function is quickly varying near the poles/the real axis, and slowly varying far from the poles/the real axis. Hence we obtain a smoothly varying Green function which allows numerically accurate quadrature methods instead of fine grid Newton-Cotes integration. Consequently the residue theorem is applied which requires the function to be analytic. For the Green function we know it has poles on the real axis (the eigenvalues of its inverse) whereas the Fermi function n_F has poles as

$$\operatorname{Res} n_F(z_\nu) = kT \quad , \text{with } z_\nu = ikT\pi(2\nu+1) \quad , \nu \in \mathbb{N} \quad (4.17)$$

$$\oint \mathrm{d}z \Big[\mathbf{G}_{\mathbf{k}}(z) - \mathbf{G}^{\dagger}_{\mathbf{k}}(z) \Big] n_F(z) = -2\pi i k T \sum_{z_\nu} \Big[\mathbf{G}_{\mathbf{k}}(z_\nu) - \mathbf{G}^{\dagger}_{\mathbf{k}}(z_\nu) \Big], \quad (4.18)$$

To calculate the real axis integral one divides the enclosed contour into the real axis and *everything else*. An example of two different, but equivalent, enclosed contours are shown in Figure 4.1. Note that $n_F(\epsilon) \to 0$ for $\epsilon \gg E_F$ which leverages the need for a fully enclosed contour. It is stressed that the choice of contour in the complex plane is purely a mathematical choice. The careful reader will also notice that the residue theorem applied to $\mathbf{G}_{\mathbf{k}}(z) - \mathbf{G}^{\dagger}_{\mathbf{k}}(z)$ indeed is equivalent to two residue theorems. One for the positive part of the imaginary coordinate system for integrating $\mathbf{G}_{\mathbf{k}}(z)$, and the second for the negative part of the imaginary coordinate system for integrating $\mathbf{G}^{\dagger}_{\mathbf{k}}(z)$. This is indicated in Figure 4.1 with $\mathcal{R}^{+/-}$, a) and b), respectively. The imaginary part of the contour line integral $\mathcal{L}/\mathcal{S}$ should be chosen large enough so that the Green function indeed is smooth. A higher number of poles increases the distance to the real axis. Thus

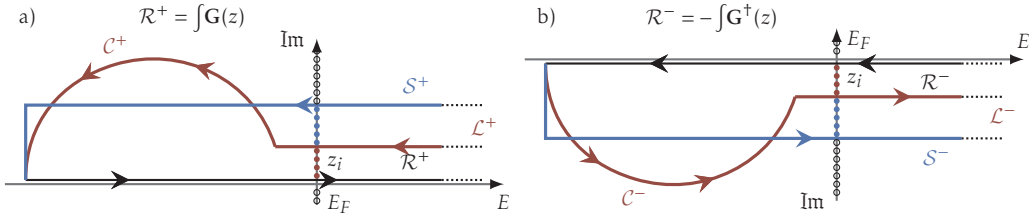


Figure 4.1 | Two enclosing contours in the complex plane. The red contour, $\mathcal{C}$, is a circle contour while the square contour, $\mathcal{S}$, is shown by a blue line. Both are mathematically equivalent. The arrows indicate the direction of the contour integration for the residue theorem. a) is the integration of the retarded Green function while b) is the integration of the advanced Green function. Note the sign change of the advanced Green function which results in the opposite contour direction.

one should take care of the number of poles used in the calculation as the imaginary part is solely determined by the temperature, see Eq. (4.17). Advisably the imaginary energy should be above 2.5 eV. The previous TranSIESTA implementation, and all available implementations denote the number of poles explicitly [10–12, 88], to the author's knowledge. However, we argue that a rigorous choice should be the imaginary energy which uniquely defines the number of poles for the integral[1]. This rigorous choice ensures that the line integral is far from the real-axis, irrespective of the electronic temperature, and particularly so for non-equilibrium calculations with only electronic temperature differences.

In TranSIESTA the line-integral $\mathcal{L}$ can be a Gauss-Fermi quadrature with pre-calculated abscissa and weights or any other standard method with weight function $w(x) = 1$.

Most importantly *must* the lower bound (end/start point of $\mathcal{C}^{+/-}$) of the contour integration be well below the lowest eigenvalue of the system as the Green function fans out when increasing the complex energy. If the real axis energy is not low enough some density will be neglected in the residue theorem and the density cannot be fully encapsulated, a simple check is to analyse the electrode band-structure and use the band bottom as a reference.

4.2.1 Optimising the energy integral

As indicated in Sec. 4.2 several choices of the contour can be made and selecting an optimum equilibrium contour *is* difficult [11, 13, 16, 50]. TranSIESTA implements several different methods, from Newton-Cotes to advanced quadrature methods, using Legendre polynomials or Tanh-Sinh quadrature [89]. Furthermore both circle and square contours are possible as shown in Figure 4.1. The continued fraction suggested by Ozaki [16, 90] is also implemented for full capability of standard methods described in the literature. The strength of the continued fraction method is a *single* convergence parameter: number of poles. The continued fraction integration is based on the continued

[1] For 100 K 47 poles are required to have an imaginary part above 2.5 eV while for 1000 K only 5 are required.

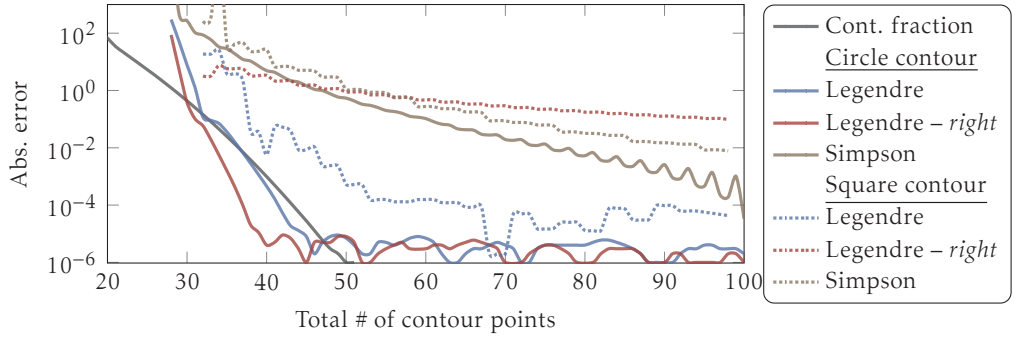


Figure 4.2 | Absolute free energy error, calculated using different Gaussian quadrature methods. Comparison of the continued fraction method and the square and circle contour using variants of integration methods. Regular Legendre, *right* Legendre and Simpson quadrature methods. It is seen that the circle *right* scheme performs better than the regular Legendre quadrature.

fraction expansion of the Fermi function

$$\frac{1}{1+e^{x}} = \frac{1}{2} - \cfrac{x/4}{1+\cfrac{x^2/4}{3+\cfrac{x^2/4}{5+\ddots}}}, \tag{4.19}$$

where the mathematical extraction of the poles and weights consist of a diagonalisation and the method can be found in [90]. Contrary to the continued fraction method, the other enclosing contours methods have at least three convergence parameters, number of poles (z_ν), points on line $\mathcal{L}^{+}$, points on circle/square $\mathcal{C}^{+}/\mathcal{S}^{+}$. A rule-of-thumb is that a higher number of poles reduces the required number of points on the $\mathcal{C}$ and $\mathcal{L}$ contours as the Green function becomes more smooth. TranSIESTA defaults to a Gauss-Legendre for the circle contour $\mathcal{C}$ and a Gauss-Fermi quadrature for the line integral $\mathcal{L}$.

Another optimisation of the residue theorem can be realised by examining Figure 4.1. It is evident that the two contours $\mathcal{C}^{+/-}$ for the retarded and advanced Green function in practice is a connected circle (or square). Hence they can be considered as *one* integration $\mathcal{C}^{+}+\mathcal{C}^{-}$, far from any poles on the real axis the Green function is analytic. A quadrature on the combined path with an even number[2] of abscissa, N, will have $N/2$ abscissas on $\mathcal{C}^{+}$ and the complex conjugate of the same points on $\mathcal{C}^{-}$. This choice has two advantages; 1) the contour at the band bottom end have fewer points, and 2) points will chunk closer to the $\mathcal{L}^{+/-}$ contour. This small "trick" allows one to slightly reduce the number of equilibrium contour points without loss of accuracy. We call this optimisation the *right* scheme.

To show the convergence properties of the equilibrium contour we have investigated a 2 electrode gold system which is connected via a gold 1D chain (112 atoms) and calculated the free energy as a function of contour points. The calculated free energy at 300 energy points is the reference[3], as such the reference is itself. This study is difficult

[2] When defining any N_i contour points on a single $\mathcal{C}^{+/-}$ path the combined points will be $2N_i = N$ which is always even.

[3] As the convergence path is non-deterministic and the "correct" value cannot be found we resort to this reference.

to extrapolate to arbitrary systems, yet it is thought to be indicative for the convergence properties of the different quadrature methods. The results are seen in Figure 4.2. x-axis has total number of quadrature points while y-axis has the absolute error. The number of poles are 16 and we use 10 $\mathcal{L}^+$ points. Both the circle and square contours are presented using both the standard and the *right* scheme as well as the continued fraction method. Note that the numerical accuracy limits the error to 10^{-6}. Clearly, the regular Newton-Cotes method using Simpson's rule is a poor choice and the convergence is extremely slow. The advanced quadrature methods converges considerably faster and using the *right* scheme generally improves the integration further. The circle contour outperforms the square which is related to the more smooth integration path of the circle contour. Further the right scheme circle contour exhibits better convergence properties compared to the continued fraction method. The square method cannot be easily applied to a *right* scheme due to the piece wise parameterisation of the contour, this is clearly visible in the poor convergence of said method. Lastly, the continued fraction scheme also converges fast and indeed is a very powerful method given that it only has one convergence parameter.

For all but the continued fraction method one should move the contour as far into the imaginary plane as possible as the Green function becomes increasingly smooth and thus the required number of circle contour points is reduced. By decreasing the required imaginary energy for the poles (fewer poles) the graphs in Figure 4.2 will tend to shift right, while for increasing the required imaginary energy (more poles) the graphs will shift to the left. Hence for another set of parameters for the circle/square contour it may perform equally well as the continued fraction method. Especially in this case where the initial number of points on the $\mathcal{L}$ contour is relatively high.

4.2.2 Estimating the exact ρ

TranSIESTA uses Eq. (4.16), calculates all ρ^ς and weighs them for improved convergence. Importantly using Eq. (4.16) instead of (4.15) reduces memory requirements substantially for $N_\mathfrak{c} > N_\varsigma \geq 2$.

Similar to the original implementation TranSIESTA estimate the density by a unity weight scheme

$$\rho = \int_{\mathrm{BZ}} \mathrm{d}\mathbf{k} \sum_\varsigma w_{\varsigma\mathbf{k}} \rho^\varsigma_{\mathbf{k}}, \tag{4.20}$$

with $\sum_\varsigma w_{\varsigma\mathbf{k}} = 1$. The variance of $\rho^\varsigma_{\mathrm{eq}}$ is assumed zero so that the error in ρ originates only from the non-equilibrium correction terms $\Delta^\varsigma_\mathfrak{c}$. From Eq. (4.20) there are two choices regarding the weights. They can be correlated[4] in $\mathbf{k}$-space which removes the $\mathbf{k}$ dependence and the easier

$$\rho = \sum_\varsigma w_\varsigma \rho^\varsigma \tag{4.21}$$

can be used. Or they are un-correlated and Eq. (4.20) is used by setting the co-variance term of $\mathrm{Cov}[\Delta^\varsigma_{\mathfrak{c},\mathbf{k}}, \Delta^\varsigma_{\mathfrak{c},\mathbf{k}'}] = 0$. We proceed the following derivations with the correlated weights as uncorrelated weights are straightforwardly deduced.

The choice of weights are rationalised by the following argumentation. Assume that the numerical integration is given by a stochastic variable $\tilde{\Delta}$ with mean value Δ and the

[4] We name it *correlated* because we do not know the co-variance term $\mathrm{Cov}[\Delta^\varsigma_{\mathfrak{c},\mathbf{k}}; \Delta^\varsigma_{\mathfrak{c},\mathbf{k}'}] = ?$.

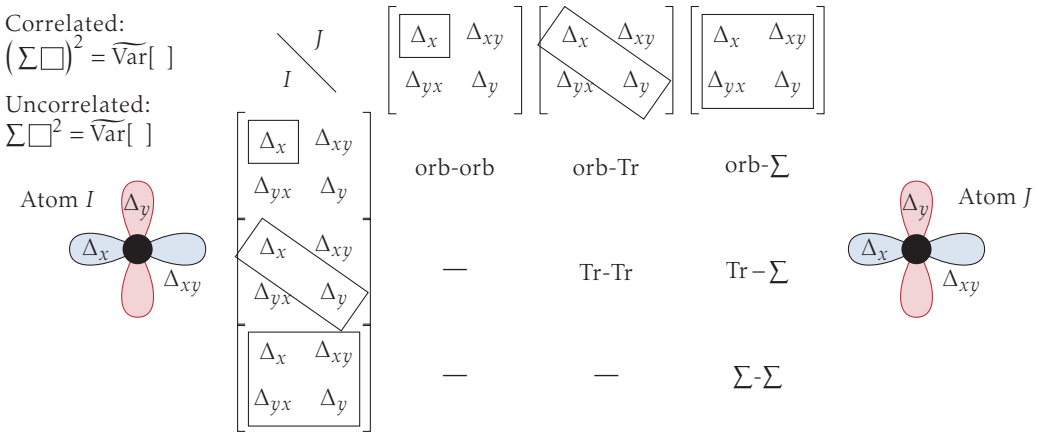


Figure 4.3 | Diagram for different choices of weights. The orbitals on atom I are shown in the left part (left of density matrix column). Atom J orbitals are shown in the right (and top row of density matrix). The weights are created using different parts of the sub-matrix indicated by boxes. The variance is estimated using the boxed elements shown in the weight-matrix. Example; orb-Tr estimates the variance through the geometric mean between the Ith atom's orbital and the trace of the orbitals on atom J. Additionally the boxed elements may be used in a correlated or uncorrelated scheme as indicated in the top-left corner.

standard deviation proportional to the magnitude of the variable itself, i.e. $\mathrm{Var}[\tilde{\Delta}] = \Delta^2$. This choice defines the variance of the density as

$$\mathrm{Var}[\boldsymbol{\rho}] = \sum_{\varsigma} w_{\varsigma}^2 \Big(\sum_{\mathfrak{e}|\varsigma_{\mathfrak{e}}\neq\varsigma} \mathrm{Var}[\Delta_{\mathfrak{e}}^{\varsigma}] \Big) = \sum_{\varsigma} w_{\varsigma}^2 \Big(\sum_{\mathfrak{e}|\varsigma_{\mathfrak{e}}\neq\varsigma} (\Delta_{\mathfrak{e}}^{\varsigma})^2 \Big) \equiv \sum_{\varsigma} w_{\varsigma}^2 \widetilde{\mathrm{Var}}[\boldsymbol{\rho}_{\mathrm{neq}}^{\varsigma}], \tag{4.22}$$

note the last definition of $\widetilde{\mathrm{Var}}$. The weights that minimise the variance are found to be

$$w_{\varsigma} = \prod_{\varsigma'\neq\varsigma} \Big(\sum_{\mathfrak{e}|\varsigma_{\mathfrak{e}}\neq\varsigma'} (\Delta_{\mathfrak{e}}^{\varsigma'})^2 \Big) \Big/ \Big\{ \sum_{\varsigma'} \prod_{\varsigma''\neq\varsigma'} \Big(\sum_{\mathfrak{e}|\varsigma_{\mathfrak{e}}\neq\varsigma''} (\Delta_{\mathfrak{e}}^{\varsigma''})^2 \Big) \Big\} \tag{4.23}$$

$$= \prod_{\varsigma'\neq\varsigma} \widetilde{\mathrm{Var}}[\boldsymbol{\rho}_{\mathrm{neq}}^{\varsigma'}] \Big/ \Big\{ \sum_{\varsigma'} \prod_{\varsigma''\neq\varsigma'} \widetilde{\mathrm{Var}}[\boldsymbol{\rho}_{\mathrm{neq}}^{\varsigma''}] \Big\}. \tag{4.24}$$

The physical interpretation is straightforwardly understood in a two electrode setup; if $\Delta_{\mathfrak{e}}^{\varsigma'} > \Delta_{\mathfrak{e}}^{\varsigma}$, the density must be contained in the equilibrium contour of ς, as the correction is small. As the equilibrium density is assumed precise it yields $w_{\varsigma} > w_{\varsigma'}$. Note that Eq. (4.23) is different from the expression used in Saha *et al*. [15].

After defining the variance and the optimum weights one has to apply the weights. The original implementation uses an orbital based partitioning [11]. This can be reasoned by assuming that each orbital is invariant to the local environment of other orbitals on the same atom. E.g. if an STM tip s orbital couples to the p_z orbital, the p_z filling will largely be unrelated to the p_x orbital population. Other variants exists such as Li *et al*. who uses the geometric mean of the orbital trace for each atom to enhance the description of bound states [13]. This assumes that the coupling to either the left/right chemical potential must be an orbital averaged quantity on each atom. Altogether the best choice of weight *may* depend on the environment.

In TranSIESTA a multitude of different methods can be chosen as it is unclear which choice is the best. Figure 4.3 shows the different choices available. Two atoms, I/J, each with two orbitals $p_{x/y}$ and their associated density sub-matrices are indicated on the left and top of the matrix-structure. The orb-orb bases the weight on the individual orbitals, orb-Tr is based on the geometric mean between the orbital of I and the trace of J. Tr-Tr is using the same weight on all orbitals connecting two atoms using the geometric mean of both atoms. Similarly for the $\sum$-$\sum$ weight with the geometric mean of the full sum of the sub-matrices. As indicated top-left, all methods can be used in a correlated and uncorrelated version where either the $\sum$ is taken before or after squaring the Δ^ς contributions. Again, one method may prove good in certain environments, however, seemingly the prevailing method is the orb-orb method which was already the standard method [11].

A numerically upper bound of the error in the final density for the given parameters is given by the maximum difference, $\mathrm{e}_{\max}$, between the different estimations, while we define the weighted error, e_w, as the difference between the estimated density and the individual densities

$$\mathrm{e}_{\max} = \max\left[\rho^{\varsigma'} - \rho^{\varsigma}; \rho^{\varsigma''} - \rho^{\varsigma}; \rho^{\varsigma''} - \rho^{\varsigma'}; \dots\right] \tag{4.25}$$

$$\mathrm{e}_w = \max\left[\rho^{\varsigma} - \rho; \rho^{\varsigma'} - \rho; \dots\right]. \tag{4.26}$$

Typically $\mathrm{e}_{\max} = 2\mathrm{e}_w$ for 2 electrode systems with no vacuum gaps. Surely the error increases for large systems as well as higher applied bias and thus the terminology "error" does not fully apply.

4.3 Electrostatics in NEGF

NEGF calculations require certain boundary conditions which are needed to correctly describe the electronic structure close to said boundaries. TranSIESTA uses the same method for solving the Poisson equation (Eq. (2.34)) as SIESTA. Although one cannot force strict boundary conditions such as Neumann or Dirichlet boundary conditions, the overall accuracy using a 3D Fourier transform has proven reliable and fast.

When performing N_ς electrode calculations one should start by considering the same concepts as standard two-terminal calculations. The electrodes are coupled to a bulk part and hence the electrostatics at the device-electrode boundary *must* be continuous and bulk-like. In the following discussion pay attention to the difference between the terms *additional electrode* and *buffer* layers. To ensure the device-electrode boundary one adds a suitable amount of *additional electrode* layers before any perturbations of the atomic structure takes place. For metals the typical amount of *additional electrode* layers need not be more than the electrode size, while for semi-conductors long range electrostatics require many more layers. Additionally the electrode must also enforce bulk electrostatics and hence be connected to bulk on both sides of the semi-infinite direction. The *additional electrode* layers ensure this towards the device region. For the other direction it can be more difficult. In typical two electrode setups with the same bulk electrode on the left and right, assuring a bulk electrostatics is easily accompanied by making the cell periodic such that the electrodes are terminated by each other, i.e. connected to bulk. For two electrodes of differing periodicity the common method is to terminate each electrode by *buffer* layers consisting of the electrode replicated a

suitable number of times. Note that these *buffer* layers do not participate in the actual NEGF calculation and are merely used for enhancing the initial guess of the Hamiltonian and retaining more stable electrostatics when calculating the Hamiltonian in the SCF cycle. One should ensure possible dangling bonds in the *buffer* layers to be sufficiently terminated to not introduce long range effects. Such as Hydrogen adsorption or other chemical species. In addition to the *buffer* layers a sufficient vacuum region is needed to decouple neighbouring super cells.

Generally these steps needs to be carefully followed

1. Setup all electrodes needed
    a) Assert the range of the basis orbitals to correctly create a structure with *only* principal layer interactions in the semi infinite direction
    b) Perform bulk calculation with high number of **k** points (> 50 is advisable) in the semi infinite direction
    c) If symmetry allows, create the *smallest* unit cell in the transverse directions, internal subsequent expansion via Eq. (2.24)
2. Construct device geometry
    a) Add few(many) *additional electrode* layers for metallic(non-metallic) electrodes to assert continuous electrostatics towards the bulk
    b) If the electrode is non-periodic in the device, add a reasonable amount of *buffer* layers on the back side of the electrode emulating the bulk for continuous electrostatics
    c) Add junction geometry to the device
3. (*optional*) the initial guess of the non-equilibrium Hartree potential may be user-defined to improve convergence

Additionally, for junctions which emulate single contacts (STM for instance) one should carefully ensure that the phase coherent transmission is reduced to an absolute minimum as discussed in [91]. This can be accomplished by extending the system in the transverse directions.

The boundary conditions for the electrostatic potential, as inferred from the above discussion, is a very important aspect of Green function calculations. For equilibrium calculations one must ensure the correct boundary conditions, and let the self-consistency converge the electrostatics. For non-equilibrium calculations one must enforce the correct boundary conditions *and*, preferentially, supply an initial guess for the final Hartree potential. In equilibrium and non-equilibrium one fixes the Hartree potential to that of a constant plane such that the potential indeed is fixed. This is a necessity as the internal potential is not fixed at a certain energy. For the initial guess of the Hartree potential there are two cases encountered in TranSIESTA; 1) a single transport direction along a unit-cell vector and 2) all other cases.

Considering the first case a common method is the use of the linear ramp which is the solution for the electric field between two infinite plate capacitors [10, 11, 16, 88]. This enhances the convergence rate and has proven as a stable method even for systems where the final potential profile can change drastically with different applied biases [92].

In TranSIESTA two different linear ramps are implemented:

$$V_H(x) \leftarrow V_H(x) + \begin{cases} \mu_L & \text{, for } x < x_{LC} \\ \mu_R & \text{, for } x_{CR} < x \\ (\mu_R - \mu_L)\frac{x - x_{LC}}{x_{CR} - x_{LC}} + \mu_L & \text{, else,} \end{cases} \tag{4.27}$$

where x_{LC}, x_{CR} are the transport direction coordinates that intersect a plane between the left electrode and the device region, and the right electrode and the device region, respectively. Eq. (4.27) assumes the electrodes to be bulk, whereas the same equation for non-bulk electrodes would be simpler

$$V_H(x) \leftarrow V_H(x) + V\left(\frac{1}{2} - \frac{x}{L_x}\right), \tag{4.28}$$

with L_x being the length of the cell vector along the transport direction. Remark that only Eq. (4.28) was implemented in the older version of TranSIESTA as well as only **c** axis could be used for the semi-infinite direction. This restriction has been relieved and *any* vector can be used, even non-orthogonal cell vectors.

For the second case (arbitrary alignments of the semi-infinite directions) an initial potential profile cannot be easily defined and hence the simplest and most reasonable approximation is to enforce the boundary conditions on the electrodes only. This can be expressed as

$$V_H(\mathbf{r}) \leftarrow V_H(\mathbf{r}) + \sum_{\mathfrak{e}} \begin{cases} \mu_{\mathfrak{e}} & \text{, for } \mathbf{r} \in \mathbf{r}_{\mathfrak{e}} \\ 0 & \text{, for } \mathbf{r} \notin \mathbf{r}_{\mathfrak{e}} \end{cases}, \tag{4.29}$$

where $\mathbf{r}_{\mathfrak{e}}$ denotes a region of the simulation cell that fully encapsulates the electrostatics of the electrode. This forces the correct boundary condition in the electrode regions while it does not infer any conditions on the device region. As this is sub-optimal in terms of estimating an initial potential profile TranSIESTA implements a generic interface for using a user-supplied potential profile as a guess on the final profile

$$V_H(\mathbf{r}) \leftarrow V_H(\mathbf{r}) + V_H^G(\mathbf{r}), \tag{4.30}$$

with $V_H^G(\mathbf{r})$ being the user-defined guess. This guess must enforce Eq. (4.29) for the electrode regions. Generally, if the electrode chemical potentials are a linear parameterisation of the applied bias, so will $V_H^G(\mathbf{r})$ be, which means that the user only needs one calculation of $V_H^G(\mathbf{r})$ for one applied bias, V^G, and any subsequent guess can be calculated as

$$V_H(\mathbf{r}) \leftarrow V_H(\mathbf{r}) + \frac{V_H^G(\mathbf{r})}{V^G} V, \tag{4.31}$$

where V is the applied bias for the current simulation.

In Figure 4.4 an example of the convergence property for a $N_{\mathfrak{e}} = 6$ electrode device is shown. The system consists of 3 linear gold chains and 3 linear Carbon chains connected via a middle junction, see b). At the end of each linear chain is an electrode (colors) with buffer atoms (tainted purple) attached to disconnect it from the asymmetric supercell image. In a) the charge difference (y) from charge equilibrium is shown as a function of SCF iterations for the equilibrium calculation (dashed) and for two different guesses for the electrostatic potential. One using Eq. (4.29) and the other an elaborate multigrid (MG) solver for the Poisson equation for the boxed electrode chemical potentials. For

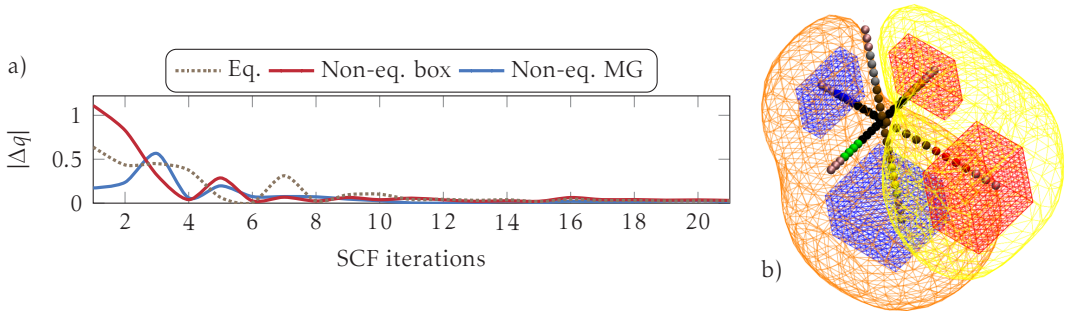


Figure 4.4 | a) Charge conservation with respect to # of SCF iterations for a $N_{\mathfrak{c}} = 6$ device (b). Both equilibrium and two different initial guesses of the Hartree potential $V_H(r)$ are shown. The dashed curve is for equilibrium while the full lines are an electrode box guess and a guess based on a multigrid (MG) solution. Providing a better guess improves convergence. b) also shows the initial MG guess for 4 iso-values $\pm V/2$ and $\pm V/10$. The box guess only uses the two iso-values $\pm V/2$.

the electrode boundary condition the initial charge fluctuation is more than $1e$ while the MG method is roughly $1/4e$ which indicates the importance of an accurate initial guess. Subsequent SCF iterations have the electrode boundary condition more fluctuating as opposed to the MG method with smaller fluctuations providing smoother and more efficient convergence. In b) the initial guess for the MG solution is seen for 4 different iso-values, $\pm V/2$ (coloured boxes, Eq. (4.29)) and $\pm V/10$ (irregular surfaces).

The multigrid[5] solver has been developed uniquely for the purpose of solving the Poisson equation as initial guesses for TranSIESTA. It can solve a mesh grid with Dirichlet, Neumann and/or periodic boundary conditions chosen in any of the cell directions.

4.4 Algorithms for inverting a matrix

In the preceding sections the governing equations for NEGF calculations have been established. A short discussion on the essential requirements for correctly performing NEGF calculations has also been discussed. Here the importance of the chosen algorithms are discussed.

Implementations of Green function techniques have received an increasing amount of attention through the development of efficient and alternative routines for inverting a matrix [10, 12, 15, 16, 50, 51, 76, 82, 93–96]. The challenging aspect of NEGF techniques is inversion of the matrix Eq. (4.1) and, in particular, a fast triple matrix product Eq. (2.32). The EGF and NEGF routines require two different algorithms. Generally, NEGF calculations are notoriously heavy in terms of memory which calls for optimum memory usage. TranSIESTA tries to never have any unused memory allocated at memory peak time. This enables a *very* efficient memory usage without altering the underlying algorithms and data-distributions. Most importantly, the needed quantities are the elements in the Green/spectral function which corresponds to the sparse elements in the density matrix Eq. (4.16), hence only a subset of the full matrix is needed.

A common method used for fast inversion algorithms is the recursive Green function method [93] which is widely used [12, 96, 97]. For truly 1D systems it is an efficient and

[5] Development by Nick R. Papior at: `http://github.com/zerothi/multigrid`

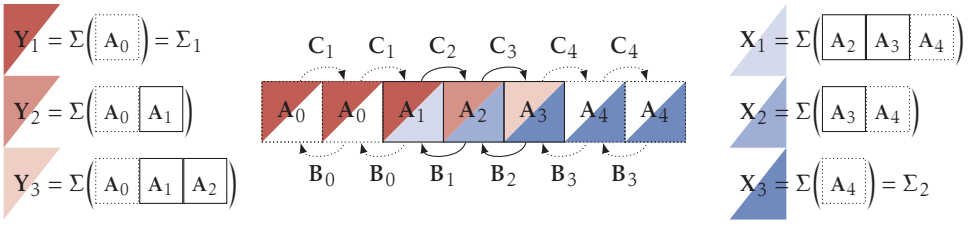


Figure 4.5 | Block-tri-diagonal inversion algorithm shown in terms of the Hamiltonian elements of a 2-electrode system indicated by $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. The surface self-energies are calculated from the bulk electrodes $\mathbf{A}_0$ and $\mathbf{A}_4$. Subsequently the self-energies are propagated through the system, allowing one to calculate the exact Green function in any block of the infinite matrix.

easily implemented method and even for quasi-1D it is a highly used and fast method. Other methods involve sparse inversion algorithms such as [98–103]. These sparse inversion algorithms also exist for efficiently calculating the spectral function [101, 104].

In this section we will cover the implemented inversion algorithms in TranSIESTA and compare them to already available methods. An emphasis on speed and optimal library choices will be emphasised and the standard libraries BLAS, LAPACK and its routines will be used throughout. In the following the spectral function (triple product) can be schematised as block matrix products where only the black parts are required

$$\mathcal{A}_\mathfrak{c}(z) = \mathbf{G}(z)\mathbf{\Gamma}_\mathfrak{c}(z)\mathbf{G}^\dagger(z) = \blacksquare\square\blacksquare\square^\dagger\blacksquare = \square\blacksquare\square\blacksquare^\dagger. \tag{4.32}$$

Finally a perspective on future $N_\mathfrak{c}$ inversion algorithms are discussed.

4.4.1 Block-tri-diagonal inversion

The block-tri-diagonal (BTD) inversion follows the recursive scheme [93] given by the following equations [105] and which can be illustrated as shown in Figure 4.5

$$\mathbf{G}^{-1} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{C}_2 & 0 & \cdots & \\ \mathbf{B}_1 & \mathbf{A}_2 & \mathbf{C}_3 & 0 & \cdots \\ 0 & \mathbf{B}_2 & \ddots & \ddots & 0 \\ \vdots & 0 & \ddots & \ddots & \mathbf{C}_p \\ & \vdots & 0 & \mathbf{B}_{p-1} & \mathbf{A}_p \end{pmatrix} \qquad \begin{aligned} \widetilde{\mathbf{X}}_n &= \left[\mathbf{A}_{n+1} - \mathbf{X}_{n+1}\right]^{-1}\mathbf{B}_n \quad , \mathbf{X}_p = 0 \\ \mathbf{X}_n &= \mathbf{C}_{n+1}\widetilde{\mathbf{X}}_n \\ \widetilde{\mathbf{Y}}_n &= \left[\mathbf{A}_{n-1} - \mathbf{Y}_{n-1}\right]^{-1}\mathbf{C}_n \quad , \mathbf{Y}_1 = 0 \\ \mathbf{Y}_n &= \mathbf{B}_{n-1}\widetilde{\mathbf{Y}}_n \end{aligned} \tag{4.33}$$

Where $\mathbf{A}_i$, $\mathbf{B}_i$ and $\mathbf{C}_i$ corresponds to the non-zero elements of $\mathbf{S} - \mathbf{H} - \sum_\mathfrak{c}\Sigma_\mathfrak{c}$. $\mathbf{X}_i/\mathbf{Y}_i$ can be thought of as the downfolded self-energies as also indicated in Figure 4.5, for instance $\mathbf{Y}_2$ corresponds to a *left* surface self-energy of an infinite bulk part connecting to $\mathbf{A}_1$. Note that only for a Left/Right terminal system with strict ordering of atoms will $\mathbf{A}_1 = \mathbf{S}_{\{1,1\}} - \mathbf{H}_{\{1,1\}} - \Sigma_{\text{Left}}$ and $\mathbf{A}_p = \mathbf{S}_{\{p,p\}} - \mathbf{H}_{\{p,p\}} - \Sigma_{\text{Right}}$. Generally this is not a requirement[6] and $\Sigma_\mathfrak{c}$ can exist in any $\mathbf{A}_i$, $\mathbf{B}_i$ or $\mathbf{C}_i$. A detailed implementation of the Green function calculation can be seen in Alg. 4.6 EGF.

Calculating $\widetilde{\mathbf{X}}/\widetilde{\mathbf{Y}}$ matrices, Alg. 4.6 EGF) line 5 and 6, can be performed in the standard 2-step process

[6] $\Sigma_\mathfrak{c}$ will only be split into two consecutive blocks as it is a dense matrix.

EGF

1: allocate $\mathbf{M}$ and $\mathbf{N}$ with p blocks
2: **for all** $\varsigma \in N_\varsigma$ **do**
3: initialise all $\mathbf{A}_i$, $\mathbf{B}_i$, $\mathbf{C}_i$ in $\mathbf{M}$
4: **for all** $i \in \{1,\dots,p-1\}$ **do**
5: $\mathbf{N}_{i+1,i} \leftarrow \widetilde{\mathbf{Y}}_{i+1}$
6: $\mathbf{N}_{p-i,p-i+1} \leftarrow \widetilde{\mathbf{X}}_{p-i}$
7: **end for**
8: **for all** $i \in \{1,\dots,p\}$ **do**
9: $\mathbf{N}_{i,i} \leftarrow \mathbf{G}_{i,i}$
10: **end for**
11: **for all** $i \in \{1,\dots,p-1\}$ **do**
12: $\mathbf{N}_{i+1,i} \leftarrow \mathbf{G}_{i+1,i}$
13: $\mathbf{N}_{p-i,p-i+1} \leftarrow \mathbf{G}_{p-i,p-i+1}$
14: **end for**
15: $\rho^{\varsigma}_{\mathrm{eq}} \leftarrow \mathbf{N}$
16: **end for**

NEGF

1: allocate $\mathbf{M}$ with p blocks + padding
2: allocate $\mathbf{N}$ with p blocks
3: initialise all $\mathbf{A}_i$, $\mathbf{B}_i$, $\mathbf{C}_i$ in $\mathbf{M}$
4: **for all** $i \in \{1,\dots,p\}$ **do**
5: **for all** $\mathfrak{c} \in \{1,\dots,N_\mathfrak{c}\}$ **do**
6: **if** any(column($\mathfrak{c}$)) $\in \mathbf{M}_{i,i}$ **then**
7: calc(i) $\leftarrow$ **True**
8: **else**
9: calc(i) $\leftarrow$ **False**
10: **end if**
11: **end for**
12: **end for**
13: **for all** $i \in \{1,\dots,p-1\}$ **do**
14: $\mathbf{N}_{i+1,i} \leftarrow \widetilde{\mathbf{Y}}_{i+1}$
15: $\mathbf{N}_{p-i,p-i+1} \leftarrow \widetilde{\mathbf{X}}_{p-i}$
16: **end for**
17: **for all** $i \in$ calc = **True do**
18: $\mathbf{N}_{i,i} \leftarrow \mathbf{G}_{i,i}$ ▷ Only columns needed
19: **end for**
20: **for all** $\mathfrak{c} \in \{1,\dots,N_\mathfrak{c}\}$ **do**
21: $\mathbf{M}_{i,i} \leftarrow \mathbf{G}_{i,i}\mathbf{\Gamma}_\mathfrak{c}\mathbf{G}^\dagger_{i,i}$
22: **for all** $\{m,n\} \in \{1,\dots,p\}$ **do**
23: **if** $m+n < 2i$ **then**
24: $\mathbf{M}_{m,n} \leftarrow \widetilde{\mathbf{Y}}_{m+1}\cdots\widetilde{\mathbf{Y}}_i\mathbf{G}_{i,i}\mathbf{\Gamma}_\mathfrak{c}\mathbf{G}^\dagger_{i,i}\widetilde{\mathbf{Y}}^\dagger_i\cdots\widetilde{\mathbf{Y}}^\dagger_{n+1}$
25: **else**
26: $\mathbf{M}_{m,n} \leftarrow \widetilde{\mathbf{X}}_{m-1}\cdots\widetilde{\mathbf{X}}_i\mathbf{G}_{i,i}\mathbf{\Gamma}_\mathfrak{c}\mathbf{G}^\dagger_{i,i}\widetilde{\mathbf{X}}^\dagger_i\cdots\widetilde{\mathbf{X}}^\dagger_{n-1}$
27: **end if**
28: $\mathbf{M}_{m,n} \leftarrow (-1)^{m+n}\mathbf{M}_{m,n}$
29: **end for**
30: $\Delta^{\varsigma\neq\varsigma_\mathfrak{c}}_\mathfrak{c} \leftarrow \mathbf{M}$
31: **end for**

Algorithm 4.6 | The algorithm used for inverting a generic BTD matrix as well as columns for calculating the spectral function in NEGF calculations. The EGF algorithm is a direct recursive algorithm [105], while the NEGF algorithm is based on the same recursive scheme which can easily be employed after an initial triple-matrix product of the scattering state in the diagonal entry. Note that there is no memory overhead of the NEGF algorithm.

1. Calculate inverse
2. Matrix multiplication (`gemm`)

or by using a single step process considering $\mathbf{AXC} = \mathbf{IC}$ as the solution of a set of linear equations. One saves two matrix-multiplication operations per block by applying the latter method. Solving $\mathbf{G}^{-1}\mathbf{G} = \mathbf{I}$, Alg. 4.6 EGF) line 9, 12 and 13, follows the iterative solution of these equations

$$\mathbf{G}_{n,n} = [\mathbf{A}_n - \mathbf{X}_n - \mathbf{Y}_n]^{-1} \tag{4.34}$$

$$\mathbf{G}_{m,n} = -\widetilde{\mathbf{X}}_{m-1}\mathbf{G}_{m-1,n} \quad \text{, for } m > n \tag{4.35}$$

$$\mathbf{G}_{m,n} = -\widetilde{\mathbf{Y}}_{m+1}\mathbf{G}_{m+1,n} \quad \text{, for } m < n. \tag{4.36}$$

The equilibrium energy points are easily calculated and further optimisation could be to limit Eqs. (4.34)–(4.36) for only the sparse elements. This is particularly so for the

off-diagonal blocks which could be reduced by approximately 50% or more for large systems.

For the non-equilibrium integral the algorithm with the optimal performance is not so straight forwardly contemplated. Here we need the triple matrix product Eq. (4.32) for all $N_\mathfrak{e}$ electrodes at any given energy point. There are two possible strategies. The first is straightforward and involves calculating the Green function column following Eqs. (4.35) and (4.36) for the columns where the scattering matrix exists. Subsequently the spectral function can be calculated using (assuming $\mathbf{\Gamma}_\mathfrak{e}$ lives in the first block), as Eq. (4.32)

$$\mathcal{A}_{\mathfrak{e},m,n} = \mathbf{G}_{m,1}\mathbf{\Gamma}_\mathfrak{e}\mathbf{G}_{n,1}^\dagger. \tag{4.37}$$

Assuming that $\mathbf{G}_{1,1}$ is calculated the calculation of $\mathcal{A}_\mathfrak{e}$ for the BTD blocks involves $5p-3$ matrix multiplications; $p-1$ for calculating the $\mathbf{G}$ column, and $4p-2$ for calculating the triple-product for all blocks.

The second strategy is a propagation of the spectral function using $\widetilde{\mathbf{X}}/\widetilde{\mathbf{Y}}$ which has the same complexity as the Green function algorithm[7], with an additional 2 matrix multiplications. Essentially this makes the NEGF contour points roughly as computationally, or less, demanding as the EGF contour points. Rewriting Eq. (4.37) yields

$$\mathcal{A}_{\mathfrak{e},m,n} = (-1)^{m+n}\widetilde{\mathbf{X}}_{m-1}\cdots\widetilde{\mathbf{X}}_2\widetilde{\mathbf{X}}_1\mathbf{G}_{1,1}\mathbf{\Gamma}_\mathfrak{e}\mathbf{G}_{1,1}^\dagger\widetilde{\mathbf{X}}_1^\dagger\widetilde{\mathbf{X}}_2^\dagger\cdots\widetilde{\mathbf{X}}_{n-1}^\dagger \quad \text{, for } m+n>2 \tag{4.38}$$

while generally

$$\mathcal{A}_{\mathfrak{e},m,n} = (-1)^{m+n}\widetilde{\mathbf{Y}}_{m+1}\cdots\widetilde{\mathbf{Y}}_i\mathbf{G}_{i,i}\mathbf{\Gamma}_\mathfrak{e}\mathbf{G}_{i,i}^\dagger\widetilde{\mathbf{Y}}_i^\dagger\cdots\widetilde{\mathbf{Y}}_{n+1}^\dagger \quad \text{, for } m+n<2i. \tag{4.39a}$$

$$\mathcal{A}_{\mathfrak{e},m,n} = (-1)^{m+n}\widetilde{\mathbf{X}}_{m-1}\cdots\widetilde{\mathbf{X}}_i\mathbf{G}_{i,i}\mathbf{\Gamma}_\mathfrak{e}\mathbf{G}_{i,i}^\dagger\widetilde{\mathbf{X}}_i^\dagger\cdots\widetilde{\mathbf{X}}_{n-1}^\dagger \quad \text{, for } m+n>2i \tag{4.39b}$$

Note that we define $\widetilde{\mathbf{X}}_{j<i}\cdots\widetilde{\mathbf{X}}_i = \widetilde{\mathbf{Y}}_{j>i}\cdots\widetilde{\mathbf{Y}}_i \equiv 1$. Importantly Eqs. (4.39) are recursive equations as implemented in lines 22 – 29. Assuming that $\mathbf{G}_{i,i}$ is calculated the calculation of $\mathcal{A}_\mathfrak{e}$ for the BTD blocks involves $3p-1$ matrix multiplications; $3p-3$ for all blocks besides $\mathcal{A}_{\mathfrak{e},i,i}$ and 2 for the $\mathcal{A}_{\mathfrak{e},i,i}$ block. Note that this complexity is equivalent to calculating $\mathbf{G}$ in an equilibrium calculation (with an additional 2 matrix multiplications).

Consequently if the spectral function is needed Eqs. (4.39) is minimally 7/5 times faster ($p=2$) and up to 5/3 times faster ($p\to\infty$) than Eq. (4.37). Note that these speedups only apply if all blocks and scattering matrices have the same size. If $\mathbf{\Gamma}_\mathfrak{e}$ has more rows/columns than any other block the speedup becomes much larger, and if $\mathbf{\Gamma}_\mathfrak{e}$ has fewer rows/columns than any other block the first strategy may be better. In TranSIESTA both are implemented and for $N_\mathfrak{e}\geq 4$ the column method is used as the quasi-1D system forces blocks which generally may be much larger than the scattering matrices. However, the chosen algorithm can be changed at will.

Instead of re-calculating the Green function for each spectral function for each energy point we pre-calculate all $\widetilde{\mathbf{X}}_i$ and $\widetilde{\mathbf{Y}}_i$, further we also calculate $\mathbf{G}_{i,i}$ for all columns where all $\mathbf{\Gamma}_\mathfrak{e}$ are non-zero, Alg. 4.6 NEGF) lines 13 – 19. Hence only a small subset of $\mathbf{G}_{i,i}$ are needed, while generally all $\widetilde{\mathbf{X}}_i$ and $\widetilde{\mathbf{Y}}_i$ are needed. These 3 quantities are stored in one of two BTD matrices. The subsequent calculation of all electrode spectral functions are straightforward following Eqs. (4.39) or Eq. (4.37) in the second BTD matrix, without additional memory usage.

[7] For $p>2N_\mathfrak{e}$ it is most certainly *less* complex, while for $p>N_\mathfrak{e}$ it may be less complex than calculating the full Green function. This is because one need not calculate any diagonal block of the Green function where any $\mathbf{\Gamma}_\mathfrak{e}$ does not live.

4.4.1.1 Orbital pivoting for minimizing bandwidth

The performance of the BTD algorithm is determined solely by the bandwidth of the Hamiltonian and overlap matrices matrices, i.e. the size of the $\mathbf{A}_n$ blocks. The bandwidth can be understood as an expression of the quasi 1D size

$$B(\mathbf{M}) = \max\Big(|i-j| \Big| \mathbf{M}_{ij} \neq 0\Big). \tag{4.40}$$

Several different methods for minimising the bandwidth have already been investigated, for a brief overview see [106, 107]. In SIESTA the sparse matrix, and hence bandwidth, is dependent on the input sequence of the atoms. However, this sparsity pattern will rarely have the minimum bandwidth. To minimise the matrix bandwidth, and heavily increase performance, 5 different pivoting methods have been implemented. 1) connectivity graph based on the Hamiltonian non zero elements (basis orbital range)[8], 2) peripheral connectivity graph based on a longest-path solution before a connectivity graph between end-points, 3) Cuthill-Mckee [108], 4) Gibbs-Poole-Stockmeyer [109], and 5) generalised Gibbs-Poole-Stockmeyer [110]. The first 2 are developed by the author and exhibits a good bandwidth reduction of the matrix for a majority of systems. The latter 3 methods are heavily used in interaction graphs with few nodal points which may be the reason for their, sometimes, poor bandwidth reduction capability in atomic structure calculations with many orbitals per atom, hence many nodal points per orbital. In TranSIESTA both the orbital interaction graph and the atomic interaction graph can be used as a mean to minimise the bandwidth. The atomic interaction graph is *much* faster and unsurprisingly yields similar results as the orbital graph. In TranSIESTA this can be chosen by the user.

Pivoting becomes increasingly important when considering $N_{\mathfrak{E}} > 2$ electrodes as the quasi 1D block-partitioning becomes less obvious. In Figure 4.7 we illustrate the naive block partition for $N_{\mathfrak{E}} = \{2, 3, 4\}$ together with an improved partitioning. For $N_{\mathfrak{E}} = 2$ (a) the naive *is* a good partitioning. The naive $N_{\mathfrak{E}} = 3$ (b) problem will create a big block for $n = 2$ which will heavily impact performance. However, by grouping two of the electrodes the quasi-1D problem will be much improved. The grouping should be chosen to minimise all block sizes, *e.g.* if the self-energy bandwidth of $B(\Sigma_2) > B(\Sigma_1) + B(\Sigma_3)$ then the Σ_1 and Σ_2 branches should swap places in Figure 4.7b. Similarly for $N_{\mathfrak{E}} = 4$ (c) two groups occur for both ends of the quasi 1D matrix. The grouping of electrodes can easily be generalised for any $N_{\mathfrak{E}}$ electrodes. Specifically if one electrode is much larger than the other (d) it can be beneficial to group the remaining electrodes.

Pivoting complicates the triple product Eq. (4.32) due to partitioning of the scattering matrix with respect to the Green function. However, each block $\mathbf{A}_n$ can be sorted such that $\mathbf{\Gamma}_{\mathfrak{E}}$ becomes consecutive in memory for optimal performance. This will maximally split $\mathbf{\Gamma}_{\mathfrak{E}}$ into two blocks. We stress that having more blocks far outweighs the benefit of $\mathbf{\Gamma}_{\mathfrak{E}}$ being fully consecutive.

An example of the importance of pivoting is shown in Figure 4.8 where a graphene nano-ribbon system is connected to either 3(top) or 4(bottom) electrodes. The figure in the far right shows the system explored with the heavily coloured atoms terminated by semi-infinite electrodes. The top purple electrode is only an electrode in the 4-electrode setup. The total number of atoms is 2,400 atoms resulting in 21,600 orbitals. When the atomic indices are randomised, except the electrodes, the sparsity pattern comes out as the left figures. The 4 electrodes can be seen as the four dense sub-blocks. The left

[8] Similar to the *Contact block reduction formalism* [106].

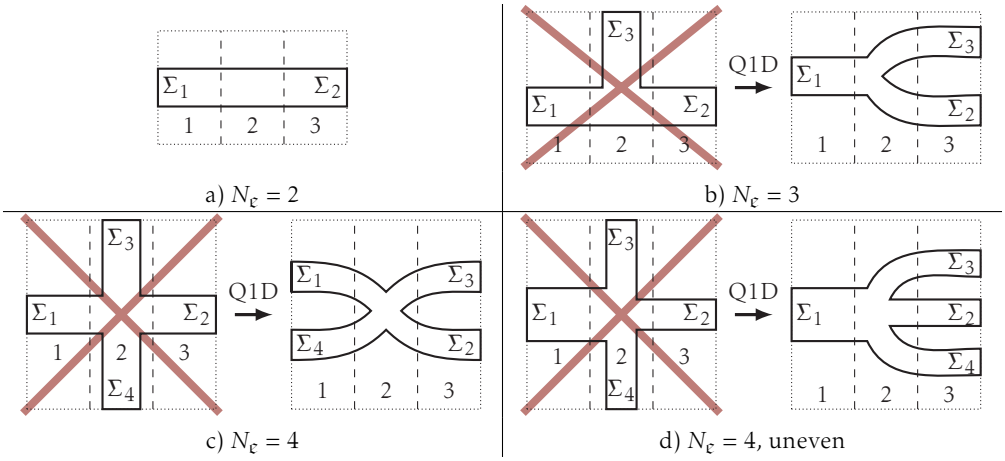


Figure 4.7 | Quasi 1D (Q1D) partitioning into 3 parts indicated by dashed lines, for varying number of electrodes. The dotted lines denote the cell boundary and the fully drawn lines encompass the atoms/Hamiltonian elements. Σ_i are the self-energies that couple the device to the semi-infinite electrodes. The crossed illustrations are the naive partitioning of the quasi-1D system (the naive partitioning for $N_\mathfrak{C} = 2$ *is* the best partitioning), whereas a better quasi-1D pivoting is also shown for $N_\mathfrak{C} > 2$. Note that in *any* of the systems shown, the ordering of the self-energies can be swapped *at will* as also highlighted in d).

figure can only be BTD partitioned in 2 blocks as indicated by the square lines dividing the matrix in two blocks. The sparsity pattern is then analysed using 20 different[9] pivoting algorithms and the two optimum pivoting results are shown in the middle figure. The resulting BTD format effectively removes roughly 87% zero elements from the computation. The 3-electrode system is partitioned best using a connectivity graph starting from the blue electrode, while the 4-electrode systems exhibits the best pivoting using the GPS algorithm [109]. After pivoting, the matrix can be split into 29 blocks. To see other pivoting matrix structures for the same system, see App. B.

4.4.2 Parallelisation

The inversion methods are *embarrassingly* parallelised across energy points, meaning that each MPI processor handles one energy point on the contour. As the matrices dealt with in TranSIESTA can become of GB size this parallelisation choice will quickly reach the memory limit, mainly dependent on the width of the electrodes. A method to further increase system sizes is to use a hybrid parallelisation strategy which is a combination of MPI and OpenMP. We have applied OpenMP 3.1 threading to the SIESTA code[10] and on top of this TranSIESTA is also fully OpenMP 3.1 threaded in all routines. Thus instead of using N_{tot} processors and reaching the memory limit, one can use N_{tot}/N_T processors and N_T threads per processor to reduce TranSIESTA memory requirement by a factor of roughly N_T and still obtain a good scaling. In Figure 4.9 the threading performance for TranSIESTA is shown for different hardware architectures using both OpenBLAS(left column) and MKL(right column). The x axis shows the number of threads used for the

[9] The combination of orbital/atomic interaction graphs, number of electrodes and all different methods + the reversed counter parts, yields 20 different schemes.

[10] SIESTA threading has only been implemented a few places, with priority on grid operations, as the SIESTA code is huge. However, using a threaded BLAS/LAPACK enables threaded diagonalisation.

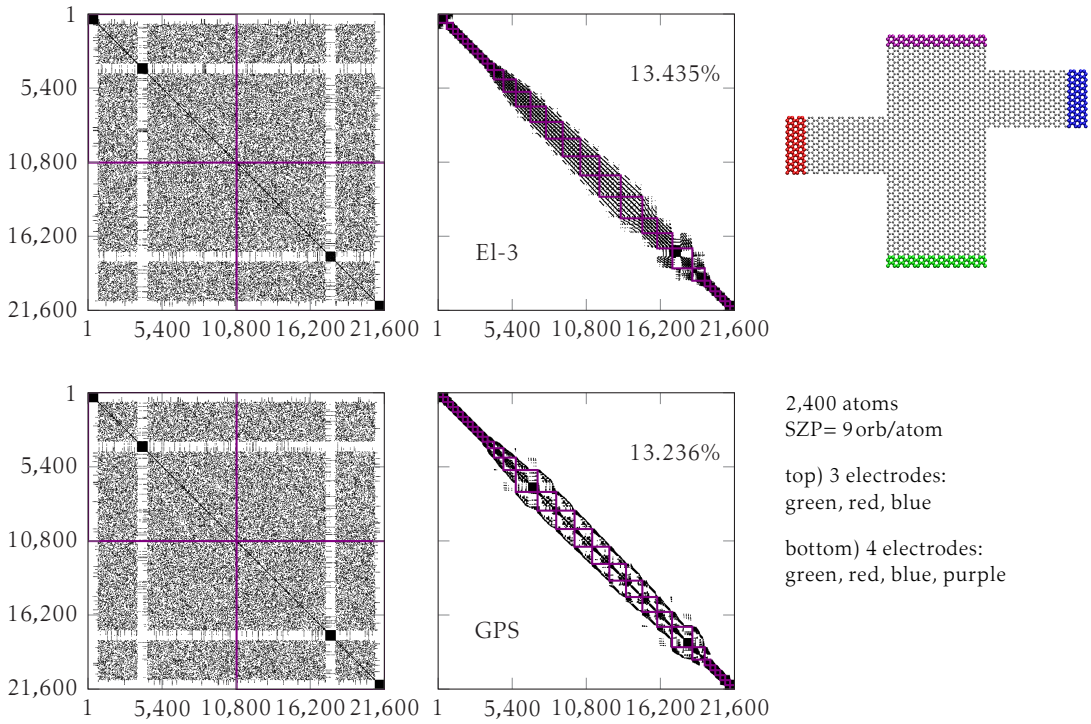


Figure 4.8 | Optimal pivoting algorithms for a 3- and 4-electrode system, shown in top-right. Graphene ribbons connected at various points along a wider ribbon. Coloured atoms are electrodes. In the left column the randomised sparsity pattern with non-zero elements are shown for 3/4 electrodes (top/bottom). To the right of the randomised pattern are the optimal pivoted matrices which are seen to cluster elements along the diagonal. The 3-electrode system has an optimal pivoting by considering the connectivity graph starting from electrode 3 (blue). The 4-electrode system further improves the pivoting by using the GPS algorithm. Both systems removes $\sim 87\%$ of the zero elements. Each BTD matrix thus only occupy ~ 0.9 GB as opposed to ~ 7.0 GB.

respective architecture. A single node is used and hence MPI-communication can be considered negligible for hardware with multiple sockets. In our implementation an increase of MPI processors will not affect the threading performance, however additional MPI processors would increase MPI communication time, thus favouring threading for large number of MPI processors. Our system consists of 11 BTD blocks with an average block size of $N_B = 830$ using the **k**-space version of the BTD method with a bias, we expect this system to represent a typical medium sized system of 9130 orbitals. We overall find a good scaling which is seen in the top plots (y-axis speedup). In the bottom plots we show the parallel fraction f, expressed from Amdahl's law

$$f(N_T) = \frac{1 - T(N_T)/T(1)}{1 - 1/N_T}, \tag{4.41}$$

$$T(N_T) = \text{execution time} = \frac{f}{N_T} T(1) + (1 - f) T(1) \tag{4.42}$$

where f is the parallel fraction of the program. As is seen in the bottom plot it retains a parallel fraction of about 95% across all N_T. In this context it is important to note that

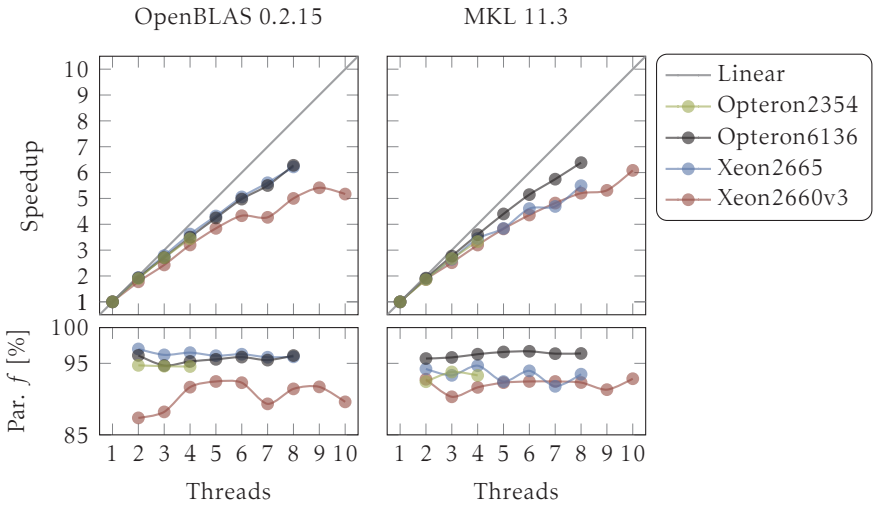


Figure 4.9 | Threading performance for TranSIESTA using different hardware architectures on a single node using either GCC and OpenBLAS(left column) or Intel and MKL(right column). The test system have 11 blocks with roughly 830 orbitals per BTD block. As the threading performance primarily stems from the threaded BLAS library one can see that the threading reaches a limit due to the relatively small blocks. The bottom row shows that the calculated parallel fraction of the TranSIESTA code. It is found to be approximately 95% dependent on library and compiler.

threading is optimal for $N_B/N_T \ggg 1$ which is one reason for the limiting speedup for large thread-counts in the example. We note that advanced CPU extensions (vectorisation) of the Intel processors further leverages the threading performance for 2660v3. Only for even wider problem sizes will the threading limitation be relieved for high performance processors. This aptly demonstrates when hybrid parallelisation can be of greater benefit and when one should revert to pure MPI parallelism.

4.4.3 Direct inversion and sparse inversion

The BTD method discussed in Sec. 4.4.1 may generally be the best inversion algorithm, however TranSIESTA also implements 2 other inversion algorithms. This enables the user to choose and test the performance of either of the methods in a single code.

4.4.3.1 Direct inversion

The most simple algorithm for inversion is the direct inversion of a matrix using dense linear algebra [11, 12, 16, 111]. The EGF calculation is simple as we invert the full matrix and retain the density matrix elements.

For NEGF Eq. (4.32) is used and one immediately recognise that only the columns where $\mathbf{\Gamma}_\varepsilon$ lives are needed. Instead of recalculating the different columns for different electrodes we can simultaneously calculate all required columns by solving a set of linear equations with the inverse Green function on the LHS and the block diagonal identity

matrix for all columns where $\mathbf{\Gamma}_\mathfrak{c}$ lives on the RHS

$$\mathbf{\Gamma}_L(z) = \blacksquare\square \quad , \quad \mathbf{\Gamma}_R(z) = \square\blacksquare \tag{4.43}$$

$$\mathbf{G}^{-1}\widetilde{\mathbf{G}} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \tag{4.44}$$

the subsequent triple matrix product can be carried out using regular dense BLAS routines.

The direct inversion algorithm is by far the slowest method and its use is discouraged other than for testing purposes.

4.4.3.2 Sparse inversion — MUMPS

Instead of calculating the full Green function one can calculate selected entries of an inverse matrix [99, 100, 102, 112, 113]. In TranSIESTA the MUMPS method is employed [99, 100, 113]. For the EGF part this has the advantage of *only* calculating the required elements. MUMPS performs best for very high sparsity which means that for small systems it performs poorly. As the different methods are embarrassingly parallelised across energy points and not across orbitals this method will appear as a slower method than it is in a fully parallelised code. Future work on TranSIESTA may implement parallel MUMPS for better scalability.

The difficulty using MUMPS stem from calculating dense parts as required by Eqs. (4.15) and (4.32). Computing dense parts using the MUMPS method is sub-optimal and thus the performance for NEGF calculations is poor compared to the BTD and just slightly faster than the dense LAPACK.

A comparison of the three methods is shown in Figure 4.10. A pristine graphene system is used with electrodes of 48 atoms (single-$\zeta^{\pm}$; 9 orbitals), corresponding to 2x6, zig-zag by arm-chair respectively, square electrode. a) shows the speedup vs. the old TranSIESTA 3.2 for different lengths and the three implemented algorithms. Both EGF(left) and NEGF(right) are compared. Note that we cannot go higher than ≈ 576 atoms due to the high requirement of memory in 3.2. b) shows the timing for the same calculation but comparing the advanced methods against the LAPACK implementation, for both EGF and NEGF. The speedup seen in a) EGF and NEGF for the LAPACK method arise only due to sparse matrices and algorithm optimisations besides the LAPACK routines. MUMPS yields up to 15 times faster for the largest matrices. The BTD method excels in this setup as the 2D (quasi-1D) matrix bandwidth becomes very narrow. With almost 50 and 100 times faster for EGF, NEGF respectively the performance is much faster for even these relatively small systems of 576 atoms. Note that MUMPS scales poorly for NEGF due to clustering of columns.

4.5 Example memory usage and timing

The previous sections have covered the specific implementation, algorithms and performance enhancements. Here an example is covered to showcase the scalability in memory consumption and speed. It is imperative that the reader knows the limitation of such information, i.e. no direct transferability to other systems of similar atomic sizes,

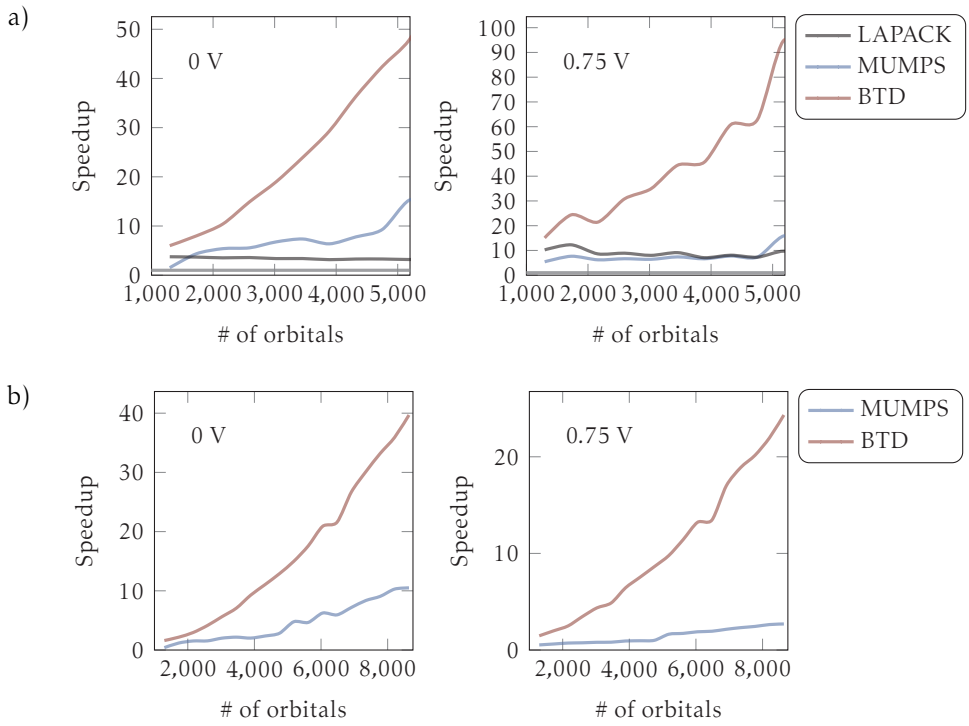


Figure 4.10 | a) Comparison of the currently implemented TranSIESTA against the previous v3.2. The speedup of the LAPACK method are optimisations related to data layouts etc. and is pretty stable around ×10 for non-equilibrium. For non-equilibrium a speedup of ∼ 100 at 5,616 orbitals is seen, this corresponds to the memory limit of v3.2 in this setup. b) Comparison between the different inversion methods in the current implementation. For non-equilibrium at 8,640 orbitals the intrinsic LAPACK speedup of ×10 yields a speedup of roughly ×250 compared to v3.2.

etc. However, it can be used as a guideline to estimate memory requirement for $N_\mathfrak{c} = 2$ systems.

The example system is a bulk Gold system using a single-$\zeta^\pm$ basis set with varying electrode[11] and device size. The basis range is roughly 4 Å. A full initialisation and memory peak read-out is performed while simultaneously a timing of the density matrix calculation using a full NEGF scheme with $V \neq 0$. Tests are performed for full parallelisation with each core performing 1 EGF and 1 NEGF calculation. The CPU type is a modern 20 core Xeon2660v3. There is no threading involved in these tests.

Results are shown in Figure 4.11 where the left contour shows the total memory usage while the right shows the SCF timing for 1 EGF and 1 NEGF calculation. On top of each plot the corresponding colorbars are shown. On the x-axis the total number of orbitals (Green function dimension) is shown, while the y-axis is the dense matrix size of the electrode self-energies. A mark at 576 orbitals indicates a rather typical electrode size corresponding to an 4×4×4 atom electrode. For tiny electrodes one can easily reach 30,000 orbitals with extremely fast SCF times, while generally such narrow electrodes are rarely used.

[11] The careful reader will notice that an electrode size may be reached by either a repetition in 1 direction or a combination of 2 directions, here we show the version with the highest memory usage.

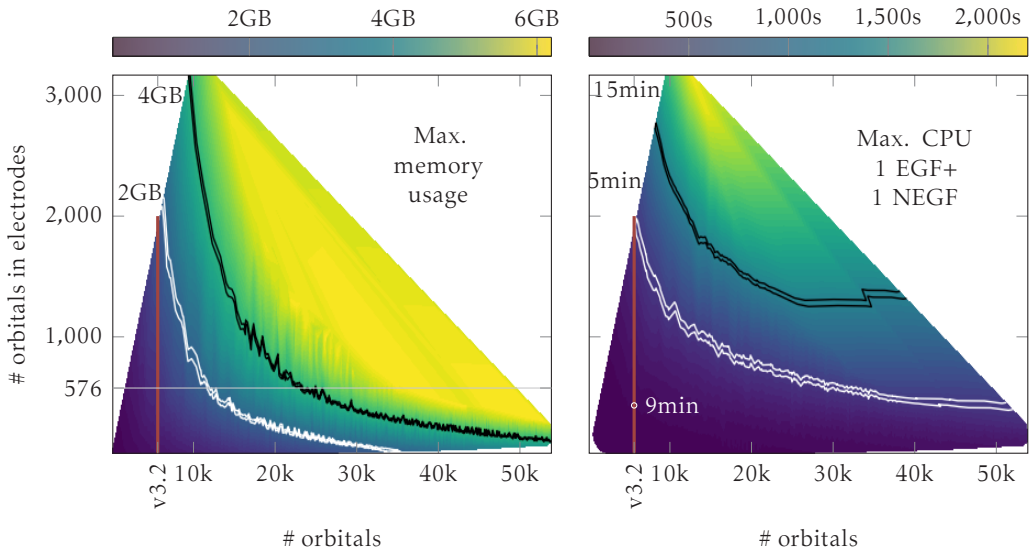


Figure 4.11 | Test bulk system of 9 orbitals per atom. The x-axis corresponds to the total number of orbitals (electrodes plus device). The y-axis is the number of orbitals in each of the self-energies, i.e. dense electrodes. Left plot is the maximum memory usage of TranSIESTA up to 6GB/core. The black highlighted contour line is the 4GB mark and the white contour line is the 2GB mark. An extra highlighted electrode size of 576 orbitals, corresponding to 4×4×4 atoms, indicates that such a typical electrode allows 1,200 atoms with only 2GB available per core and up to 2,500 atoms with 4GB available per core. Right plot shows the corresponding timings for 1 EGF and 1 NEGF per core. Timing is roughly per SCF timing in TranSIESTA. I.e. calculating density via Green function techniques. The black contour line corresponds to 15 minutes per SCF, while the white contour is for 5 minutes per SCF.

Two memory usages are highlighted via additional contour lines for the two typical hardware limits, 2GB and 4GB per core. Correspondingly if one uses the 576 size electrode one can reach system sizes of 11,000 orbitals (1,200 atoms) with 2GB per core and 23,000 orbitals (2,500 atoms) with 4GB per core. Similarly for SCF timings below 15 minutes one can e.g. calculate 1 EGF and 1 NEGF SCF cycle with electrode sizes of 1,600 orbitals and a device size of 20,000 orbitals. Note that for these matrix sizes, TranSIESTA is the memory limiting part of the calculation. Whence, using threading can double (2 threads per process) or quadruple (4 threads per process) the achievable system sizes.

The electrode size can be reached in combinations of either one direction repetition or two directions repetitions. Indeed the memory usage will change drastically if one of the directions is 1D like, for instance graphene and other 2D materials.

4.6 Summary

In this Chapter a full description of non-equilibrium calculations has been outlined from the initial density matrix to the different algorithms available for inverting matrices. We investigated the different available contour integration techniques for the equilibrium density with the continued fraction and two simple enclosed contours. In the latter methods a reduction of needed contour points were shown by selecting the *right* part of

the quadrature abscissa. It was shown that TranSIESTA employs an averaged density matrix calculation for $V \neq 0$ which better approximates the *exact* non-equilibrium density. A consequence of the electrode electronic structure distributions for non-equilibrium equations was that one may reduce the complexity of $N_\mathfrak{e} > N_\varsigma \geq 2$ calculations by reducing from $N_\mathfrak{e}$ equivalent densities to N_ς. This reduction in complexity becomes increasingly important for large scale calculations, for both memory and throughput considerations. We showed that a multitude of different weighting schemes for $N_\varsigma > 1$ may be used. Both **k** resolved density matrix weighting and Brillouin zone integrated density matrix weighting are available, as well as choices regarding orbital vs. atomic weighting schemes.

The 3 implemented inversion algorithms were covered and compared. For the investigated systems the BTD algorithm proved more efficient than using the MUMPS or the direct inversion algorithms. For an example graphene system the BTD is up to 250 times faster than the previous TranSIESTA implementation for ~ 8000 orbitals. The BTD algorithm was thoroughly explained to highlight some selected design choices for the underlying data structures and in particular the 2 different ways of computing the spectral density matrices. Either via column product or a propagation method. For $N_\mathfrak{e} = 2$ the propagation method will in almost all cases be faster than the column product. Several different bandwidth reduction methods where presented and an example for a complex system with $N_\mathfrak{e} = \{3, 4\}$ electrodes where shown. In the shown example the BTD algorithm could reduce memory usage by $\sim 87\%$.

Importantly the implemented algorithms were shown to encompass a threading parallelisation on top of the *embarrassingly* parallelisation across energy points. This provides TranSIESTA to handle even bigger systems as threading reduces memory consumption by $\approx N_T$ with N_T being number of threads. An example for a medium sized system showed a scalable code with a parallel fraction of around 95%. Finally we showed TranSIESTA memory usage for a bulk system with varying electrode and device size.

Future implementations may use the parallel MUMPS method [99, 100] or the massively parallel PEXSI algorithm [102]. An extension of the parallelisms could also be made by using more recent methodologies of MPI, such as remote-memory-access in one-sided communications. Furthermore we expect the BTD algorithm to be expanded into a branching algorithm which deals with matrices in a more general way, an example of such a branching algorithm will be apparent from the implemented inversion algorithm explained in Sec. 5.2.

Chapter 5

Tight binding transport calculations

In this Chapter we cover the implementation details of the ballistic transport equations and in particular expand on the algorithm choices concerning $N_{\mathfrak{e}} \geq 1$. The general transport formalism for multi-terminal cases are covered in various articles and books which we refer to for additional information. Both the Landauer-Büttiker formalism and Lippmann-Schwinger equation, among others, are covered in [55, 74, 75, 114–118].

It is important to consider transport calculations as a separate entity compared to self-consistent calculations as other optimisations will be important. Furthermore, other quantities such as transmission eigenvalues, projections, DOS, spectral DOS and bond-currents may be important.

While TBtrans is primarily intended as a back-end for TranSIESTA we emphasise the generality of TBtrans which is a generalised code for performing tight-binding transport calculations. Hence no prior knowledge to SIESTA/TranSIESTA is required. Also TBtrans exists in a variant, PHtrans, which enables phonon transport calculations. As such, TBtrans can be used without any SIESTA/TranSIESTA interaction. In this Chapter we will, however, not cover the tight-binding usage. See App. C for an example and more details. The recently released GOLLUM code [119] is another code capable of custom transport tight-binding calculations as well as multi-terminal support.

5.1 General calculated quantities

The transport equations governing $N_{\mathfrak{e}}$ electrodes can be derived as also done in [15, 85, 114, 119, 120]. The bias, V, will in the following be an implicit dependency of variables via the non-equilibrium Hamiltonian and the self-energies. For non-self consistent Hamiltonians the bias dependency must be implemented by hand.

One quantity calculated is the density of states which has been formulated in Eqs. (3.6) and (3.7) for an orthogonal basis set. However, the direct conversion to a non-orthogonal basis set can be performed by using Eq. (2.33) which yields the DOS and spectral DOS using the Mulliken population separation, [69]

$$\rho(\epsilon) = -\frac{1}{\pi}\operatorname{Im}\operatorname{Tr}[\mathbf{G}(\epsilon)\mathbf{S}], \tag{5.1a}$$

$$\rho^{\mathfrak{e}}(\epsilon) = \frac{1}{2\pi}\operatorname{Re}\operatorname{Tr}[\boldsymbol{\mathcal{A}}_{\mathfrak{e}}(\epsilon)\mathbf{S}], \tag{5.1b}$$

while the local DOS on orbital ν can be expressed as

$$\rho_\nu(\epsilon) = -\frac{1}{\pi}\mathrm{Im}[\mathbf{G}(\epsilon)\mathbf{S}]_\nu, \tag{5.1c}$$

$$\rho^{\mathfrak{e}}_\nu(\epsilon) = \frac{1}{2\pi}\mathrm{Re}[\mathcal{A}_{\mathfrak{e}}(\epsilon)\mathbf{S}]_\nu. \tag{5.1d}$$

Importantly the overlap matrix $\mathbf{S}$ infers some problems for estimating the exact Mulliken DOS. Special care must be taken for the orbitals on the boundary of the device region. For these orbitals ν a slight error is made by disregarding the shared population with orbitals outside the defined region. Typically this error can be neglected while one should note the importance of this deficiency.

The transmission functions can be calculated using the scattering matrix formalism [10], we will implicitly assume energy dependence on all quantities,

$$s_{\mathfrak{e}\mathfrak{e}'} = -\delta_{\mathfrak{e}\mathfrak{e}'} + \mathrm{i}\mathbf{\Gamma}^{1/2}_{\mathfrak{e}}\mathbf{G}\mathbf{\Gamma}^{1/2}_{\mathfrak{e}'}, \tag{5.2}$$

where $s_{\mathfrak{e}\mathfrak{e}'}$ is an element in the scattering matrix with indices corresponding to the electrodes. The transmission between any two electrodes is $\mathcal{T}_{\mathfrak{e}\to\mathfrak{e}'} = s^\dagger_{\mathfrak{e}\mathfrak{e}'}s_{\mathfrak{e}\mathfrak{e}'}$. For $\mathfrak{e} = \mathfrak{e}'$ it is the reflection, hence $\mathcal{R}_{\mathfrak{e}} \equiv \mathcal{T}_{\mathfrak{e}\to\mathfrak{e}}$. It is instructive to write up $\mathcal{T}_{\mathfrak{e}\to\mathfrak{e}'}$, the definition of $\mathcal{T}_{\mathfrak{e}}$ and $\mathcal{R}_{\mathfrak{e}}$

$$\mathcal{T}_{\mathfrak{e}\to\mathfrak{e}'} = \mathrm{Tr}\left[\mathbf{\Gamma}_{\mathfrak{e}'}\mathbf{G}\mathbf{\Gamma}_{\mathfrak{e}}\mathbf{G}^\dagger\right] \quad \text{, for } \mathfrak{e} \neq \mathfrak{e}' \tag{5.3}$$

$$\mathcal{T}_{\mathfrak{e}} \equiv \sum_{\mathfrak{e}'\neq\mathfrak{e}} \mathcal{T}_{\mathfrak{e}\to\mathfrak{e}'} \tag{5.4}$$

$$\mathcal{R}_{\mathfrak{e}} = \mathbf{1} - \mathcal{T}_{\mathfrak{e}} = \mathbf{1} - \left\{\mathrm{i}\,\mathrm{Tr}\left[(\mathbf{G}-\mathbf{G}^\dagger)\mathbf{\Gamma}_{\mathfrak{e}}\right] - \mathrm{Tr}[\mathbf{\Gamma}_{\mathfrak{e}}\mathbf{G}\mathbf{\Gamma}_{\mathfrak{e}}\mathbf{G}^\dagger]\right\}, \tag{5.5}$$

where $\mathcal{T}_{\mathfrak{e}\to\mathfrak{e}'}$ is the usual calculated quantity. A rule-of-thumb to remember the sequence of matrix multiplications for the $\mathfrak{e}\to\mathfrak{e}'$ is that the spectral function is the origin of the electron and the remaining scattering matrix is the absorbing electrode. $\mathcal{T}_{\mathfrak{e}}$ is the total transmission out of an electrode. The reflection is conveniently written in 2 terms; $\mathbf{1}$ is the number of open channels in electrode $\mathfrak{e}$ at the given energy (bulk transmission), the last term consists of yet another two terms. The first is the total transmission out of electrode $\mathfrak{e}$ plus $\mathrm{Tr}[\mathbf{\Gamma}_{\mathfrak{e}}\mathbf{G}\mathbf{\Gamma}_{\mathfrak{e}}\mathbf{G}^\dagger]$, the last term is a correction term[1], see Eq. (A.22). Eq. (5.5) display an important, and often overlooked detail. In transport calculations with $N_{\mathfrak{e}} = 2$, one *can* calculate the transmission using only the diagonal Green function *and* only one scattering matrix. We stress that the quantities calculated may have numerical deficiencies as $\mathrm{Tr}[(\mathbf{G}-\mathbf{G}^\dagger)\mathbf{\Gamma}_{\mathfrak{e}}]$ and $\mathrm{Tr}[\mathbf{\Gamma}_{\mathfrak{e}}\mathbf{G}\mathbf{\Gamma}_{\mathfrak{e}}\mathbf{G}^\dagger]$ may both be numerically large values which leads to inaccuracies when the transmission is orders of magnitudes smaller than the reflection[2].

In App. A.1 symmetry arguments can be made for $\mathcal{T}_{\mathbf{k},\mathfrak{e}\to\mathfrak{e}'}(\epsilon) = \mathcal{T}_{-\mathbf{k},\mathfrak{e}'\to\mathfrak{e}}(\epsilon) = \mathcal{T}_{\mathbf{k},\mathfrak{e}'\to\mathfrak{e}}(\epsilon)$ which is implicitly assumed in the following[3]. The conductance can be expressed in terms of the transmission as

$$G_{\mathfrak{e}\to\mathfrak{e}'}(\epsilon) = G_0\int_{\mathrm{BZ}}\mathrm{d}\mathbf{k}\mathcal{T}_{\mathbf{k},\mathfrak{e}\to\mathfrak{e}'}(\epsilon) = \frac{e^2}{h}\int_{\mathrm{BZ}}\mathrm{d}\mathbf{k}\mathcal{T}_{\mathbf{k},\mathfrak{e}\to\mathfrak{e}'}(\epsilon) = \frac{\mathrm{d}I}{\mathrm{d}V}, \tag{5.6}$$

[1] We are currently unaware of any physical meaning of $\mathrm{Tr}[\mathbf{\Gamma}_{\mathfrak{e}}\mathbf{G}\mathbf{\Gamma}_{\mathfrak{e}}\mathbf{G}^\dagger]$ why we resort to *correction term*.

[2] Typically this is a problem for systems with large bulk transmissions. If the number of incoming channels is only a couple of orders larger than the transmission the numerical precision is adequate.

[3] Even though $\mathcal{T}_{\mathbf{k},\mathfrak{e}\to\mathfrak{e}'}(\epsilon) = \mathcal{T}_{\mathbf{k},\mathfrak{e}'\to\mathfrak{e}}(\epsilon)$ it is important to keep the terminology as this equality only holds for $\mathbf{B} = 0$.

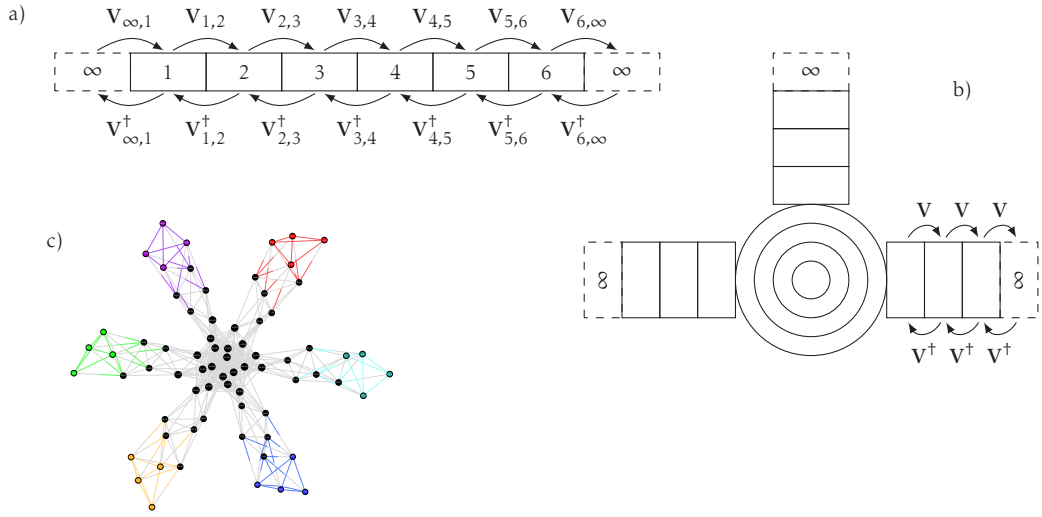


Figure 5.1 | Three different $N_\mathfrak{C}$ systems. a) simple quasi 1D system with 2 electrodes divided in 6 BTD blocks. b) 3 electrode system connected via a circular region. c) connectivity graph of a 6 terminal system (coloured atoms are electrodes), the lines are equivalent to a non-zero Hamiltonian element. TBtrans allows calculating the transport in the smallest region where *all* electrode self-energies can uniquely be defined.

where spin multiplicity is intentionally left out. The total current then becomes

$$I_{\mathfrak{C}\to\mathfrak{C}'} = \int \mathrm{d}\epsilon G_{\mathfrak{C}\to\mathfrak{C}'}(\epsilon)[n_{F,\mathfrak{C}}(\epsilon) - n_{F,\mathfrak{C}'}(\epsilon)]. \tag{5.7}$$

Simultaneously the thermal energy exchange of the electronic reservoirs can be calculated by the similar energy integral

$$H_{\mathfrak{C}\to\mathfrak{C}'} = \frac{1}{h}\iint_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\mathrm{d}\epsilon(\epsilon - \mu_\mathfrak{C})\mathcal{T}_{\mathbf{k},\mathfrak{C}\to\mathfrak{C}'}(\epsilon)[n_{F,\mathfrak{C}}(\epsilon) - n_{F,\mathfrak{C}'}(\epsilon)], \tag{5.8}$$

which generally fulfils $H_{\mathfrak{C}\to\mathfrak{C}'} \neq H_{\mathfrak{C}'\to\mathfrak{C}}$. Only in the case of a fully symmetric junction will $H_{\mathfrak{C}\to\mathfrak{C}'} = H_{\mathfrak{C}'\to\mathfrak{C}}$.

These above quantities are the easiest and simplest to calculate. In following sections we will cover more advanced features of TBtrans.

5.2 Algorithms for inverting a matrix — *again*

The calculation of the Green function is not exactly equivalent to the TranSIESTA counterpart. Mainly because if one is *only* interested in calculating the transmission function inverting the full matrix is needlessly complex [97]. However, the algorithm used for the following discussion is roughly equivalent to the algorithm presented in Sec. 4.4.1.

When calculating the transmission TBtrans only need the diagonal Green function where all self-energies can be separately defined[4]. The self-energies are "down-folded"

[4] As explained in Sec. 5.1 one can calculate the transmission via only one diagonal block, Eq. (5.5), and one scattering matrix but for $N_\mathfrak{C} \geq 3$ this is not a viable strategy.

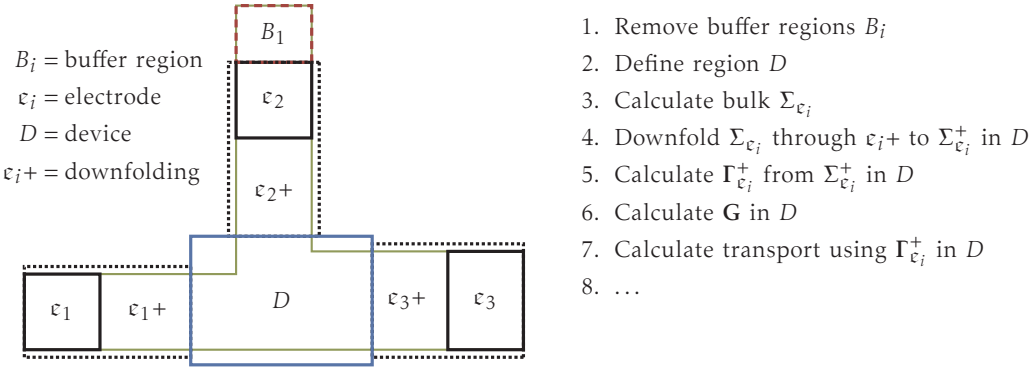


Figure 5.2 | Algorithm for calculating the Green function in TBtrans for a 3 terminal system. Electrodes are marked by thick black lines. After defining the "device" region, D, self-energies are downfolded through $\mathfrak{e}_i+$ to D. Subsequently, the system used for calculating any quantity is based on region D which means that DOS etc. can only be estimated in this region.

from each electrode via Eq. (4.33) into some region denoted the "device" region. The down-folding of the self-energies is shown schematically in Figure 4.5 and correspond to the $\widetilde{\mathbf{Y}}_i$ matrices. Additional examples are presented in Figure 5.1 where a) corresponds to a simple 2 electrode setup (directly transferable to Figure 4.5), b) a 3 electrode setup and lastly c) which is the atomic interaction graph from TranSIESTA of a 6 electrode setup. In a) the transmission can be calculated in block 2, or block 3, or blocks $\{3,4\}$ etc. For b) the minimal region required for calculating all transmission coefficients is the outermost ring of the circular connecting region. One can also add any number of the square blocks, while it cannot be reduced to any of the smaller rings as then the self-energies become entangled and their origin is lost. For c) the differently coloured atoms correspond to electrodes which all connect via a single crossing junction. This is like b) where circles can be formed from the centre and out.

The algorithm for calculating the Green function can be shown schematically in Figure 5.2. The system comprise of 3 electrodes, $\mathfrak{e}_i$ marked by thick lines, connected through the light coloured line. A *sub*-region in the system is defined as the "device" region (D in figure). This region is subsequently used for calculating all requested quantities. The buffer region B_1 indicates a needed region when carrying out SCF calculations in TranSIESTA, in TBtrans this region is automatically removed. The electrode self-energies are down-folded through $\mathfrak{e}_i+$ into D via $\widetilde{\mathbf{Y}}_i$ where the scattering matrices are then calculated from the down-folded $\Sigma^+_{\mathfrak{e}_i} = \mathbf{Y}_p$. This procedure arises naturally from the BTD matrix inversion algorithm outlined in Sec. 4.4.1 and is also identical to the linear self-energy calculation Eq. (3.24). Remark that only $\widetilde{\mathbf{Y}}_i$ are needed in the down-folding algorithm. One must ensure that regions $\{\mathfrak{e}_i+\} \notin \{\mathfrak{e}_j+\}$ for any $i \neq j$ such that the origin of the self-energy can always be tracked. A too small or ill chosen "device" region can be checked in the sparsity pattern via the connectivity graph from each electrode. Another benefit of choosing the "device" region is the possibility of carrying out extreme scale transport calculations using an efficient and scalable code. The memory footprint of TBtrans is only[5] dependent on the number of orbitals in the

[5]While internal data structures keep track of the full structure, and do contribute to the total memory footprint, they are not nearly as memory consuming as the "device" Green function.

"device" region.

The algorithm can be generalised to any $N_{\mathfrak{E}} \geq 1$ where for $N_{\mathfrak{E}} = 1$ one may use TBtrans to explore, for example, bound states or bond currents. Neither MUMPS or full matrix inversions can be used in TBtrans.

TBtrans is created with high throughput as the top priority and hence it automatically decides on algorithms for optimal performance dependent on the quantities that needs to be calculated. From Eq. (5.3) it can be deduced that calculating the transmission only requires the diagonal of the quadruple matrix product. $\mathcal{T}_{\mathfrak{E}\to\mathfrak{E}'}$ can be calculated via the spectral function $\mathcal{A}_{\mathfrak{E},\{\mathfrak{E}'\}}$ in the diagonal block where $\mathbf{\Gamma}_{\mathfrak{E}'}$ exists, denoted $\{\mathfrak{E}'\}$. Instead of taking the trace of the full matrix multiplication, one can calculate the trace via the dot-product of the vectorised matrices:

$$\mathcal{T}_{\mathfrak{E}\to\mathfrak{E}'} = \sum_{i\in\{\mathfrak{E}'\}} \mathbf{\Gamma}_{\mathfrak{E}',\{i,:\}} \cdot \mathcal{A}_{\mathfrak{E},\{:,i\}} = \mathrm{vec}(\mathbf{\Gamma}_{\mathfrak{E}'}^{T}) \cdot \mathrm{vec}(\mathcal{A}_{\mathfrak{E},\{\mathfrak{E}'\}}), \tag{5.9}$$

where vec translates the matrix to a vector. Note that for an efficient implementation one of the matrices must be transposed and hence storing the transposed scattering matrix is beneficial. If one requests the transmission eigenvalues [11, 121, 122] the full diagonal matrix from the 4th matrix multiplication in $\{\mathfrak{E}'\}$ *is* needed. One need not the entire matrix $\mathbf{\Gamma}_{\mathfrak{E}'}\mathcal{A}_{\mathfrak{E}}$ as

$$\det\left(\mathbf{\Gamma}_{\mathfrak{E}'}\mathcal{A}_{\mathfrak{E}} - \lambda\mathbf{I}\right) = \det\left(\blacksquare - \lambda\mathbf{I}\right) = \det\left(\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{A} \\ \mathbf{0} & \mathbf{0} & \mathbf{B} \\ \mathbf{0} & \mathbf{0} & \mathbf{C} \end{bmatrix} - \lambda\mathbf{I}\right) \to \det(\mathbf{C} - \lambda\mathbf{I}). \tag{5.10}$$

This reduces the complexity considerably when calculating the transmission eigenvalues while also reducing the required memory.

Due to internal data structures the calculation of the Green function DOS is a relatively expensive calculation due to the NEGF algorithm outlined in Alg. 4.6.

5.2.1 Orbital pivoting for minimising bandwidth

In Figure 5.2 the calculation is divided into $N_{\mathfrak{E}}+1$ regions, one for each electrode and one "device" region. All $N_{\mathfrak{E}}+1$ regions are regarded as separate BTD matrices and hence the bandwidth of all regions, Eq. (4.40), becomes essential for fast performing algorithms. Dividing the system into $N_{\mathfrak{E}}+1$ regions improves the quasi-1D handling in each of the regions thus further increasing performance. However, the performance is highly dependent on the proper choice of the device region. This will often make TBtrans inversion *much* faster than TranSIESTA as one can define the "device" region so small that it becomes more quasi 1D like than the equivalent full system from TranSIESTA, particularly so for $N_{\mathfrak{E}} > 2$.

5.3 Bias interpolation

Though the implementation outlined in Chapter 4 suggests that performing NEGF calculations is much faster than previously, they are still a heavy duty analysis tool. Especially for I-V characteristics which requires many bias calculations. To increase throughput, and decrease the number of full NEGF calculations a general interpolation scheme has

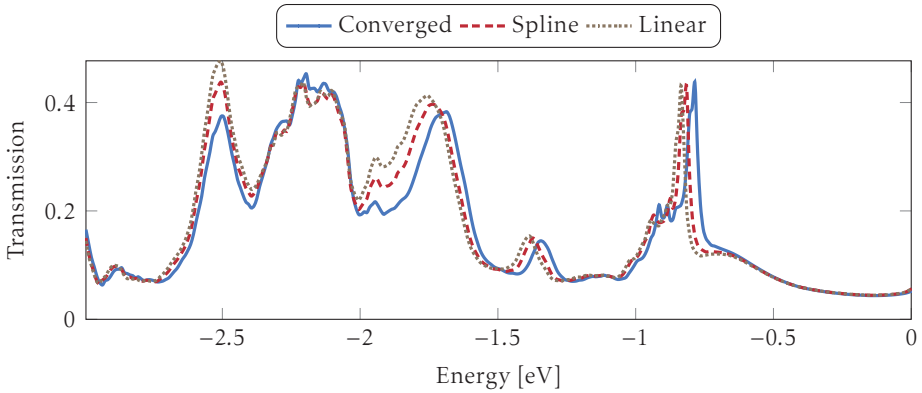


Figure 5.3 | Bias interpolation of transmission for C_{60} molecule at −1.5 V. The converged is the SCF NEGF calculation at −1.5 V. A linear interpolation using −2 V and −1 V SCF calculations. This captures the majority of features seen in the transmission plot. The spline interpolation also use the 0 V, 1 V and 2 V SCF calculations. A better agreement with the converged calculation is seen for the spline.

been implemented. Two interpolations are implemented; 1) a linear interpolation and, 2) a natural spline interpolation. The interpolated quantity is the Hamiltonian

$$\mathbf{H}_{\mathbf{k}}^{\text{int}}(V) = f(V, \mathbf{H}_{\mathbf{k}}(V_1), \mathbf{H}_{\mathbf{k}}(V_2), \dots) = f_{\mathbf{H}}(V, V_1, V_2, \dots) \tag{5.11}$$

with $\mathbf{H}^{\text{int}}(V)$ being the interpolated Hamiltonian and f the interpolation function. Currently we are only aware of the `OpenMX` code which employs the linear interpolation scheme [16]. The linear interpolation only uses the Hamiltonian from the two closest bias points while a spline interpolation is a smooth function which crosses each bias point exactly.

In Figure 5.3 an example of interpolation of the Hamiltonian for a given bias and the resulting transmission plots. The system is a C_{60} molecule slightly protruding the surface of one Copper electrode while the other Copper electrode has the form of an STM tip. The electrodes have the semi infinite direction along the 111 lattice direction. Full NEGF calculations have been performed between −2.4 V to 2.4 V with $\delta V = 0.1$ V. Interpolated transmissions for all bias calculations have been performed using 5 converged Hamiltonians taken at −2 V, −1 V, 0 V, 1 V and 2 V. From these 44 interpolations we select the interpolated transmission which deviated the most and plot the corresponding transmissions in Figure 5.3. The interpolated bias which deviated the most was −1.5 V. This corresponds roughly to a level alignment of the protruding electrode Fermi level with the highest 5 semi-degenerate HOMO levels (~ -0.75 eV). This alignment may correspond to a larger change in the electronic structure which cannot be captured by interpolating across 0.5 V ranges. For both the linear and the spline interpolation the main features are well captured. Clearly the spline interpolation has a smaller error compared to the linear interpolation scheme. We believe interpolation may be used to efficiently calculate I-V characteristics with high precision for a large variety of systems while care is needed if the electronic structure is non-linear in V.

Future interpolation algorithms which takes drastic changes in the electronic structure into consideration may prove even better than the presented interpolation schemes.

5.4 Bond currents

An increasingly used methodology for understanding currents in nanoscale structures is via the so-called "bond currents" [83, 85, 86, 123]. Bond currents are defined via the spectral density matrix for an originating electrode $\tilde{\mathcal{A}}_{\mathfrak{c},\mathbf{k}}$. Here, $\{\tilde{}\}$ denotes that the basis set is orthogonal. We omit the implicit energy dependence. The resulting (orbital) bond current can be expressed via [83, 85]

$$\mathcal{J}_{\mathfrak{c},\mathbf{k},\nu\mu} = \frac{1}{\hbar}\mathbb{Im}\left[\tilde{\mathcal{A}}_{\mathfrak{c},\mathbf{k},\nu\mu}\tilde{\mathbf{H}}_{\mathbf{k},\mu\nu} - \tilde{\mathcal{A}}_{\mathfrak{c},\mathbf{k},\mu\nu}\tilde{\mathbf{H}}_{\mathbf{k},\nu\mu}\right], \tag{5.12}$$

$$\mathcal{J}_{\mathfrak{c},\nu\mu} = \int_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\mathcal{J}_{\mathfrak{c},\mathbf{k},\nu\mu} \tag{5.13}$$

where ν and μ are orbital indices. For the Γ-point this reduces to

$$\mathcal{J}_{\mathfrak{c},\Gamma,\nu\mu} = \frac{2}{\hbar}\mathbb{Im}\left[\tilde{\mathcal{A}}_{\mathfrak{c},\Gamma,\nu\mu}\tilde{\mathbf{H}}_{\Gamma,\mu\nu}\right]. \tag{5.14}$$

Any sum of bond currents crossing a surface separating the originating electrode from the remaining device will result in the total current due to the current continuity equation.

Importantly the above bond currents are valid in an orthogonal basis set while for non-orthogonal basis sets they do not apply [83, 85]. This is related to the ambiguity in associating populations on individual atoms via Mulliken population analysis [69, 70, 124]. It then becomes an issue whether an electron is on atom I or J which is important when calculating bond currents. In order to attribute local bond currents in a non-orthogonal basis set a transformation to an orthogonal basis set is required. Typically the Löwdin transformation, [125], is used as it is straightforwardly applied by extracting $\mathbf{S}_\mathbf{k}$ from Eq. (3.1) resulting in this alternate Hamiltonian and Green function

$$\tilde{\mathbf{H}}_\mathbf{k} = \mathbf{S}_\mathbf{k}^{-1/2}\mathbf{H}_\mathbf{k}\mathbf{S}_\mathbf{k}^{-1/2} \tag{5.15}$$

$$\tilde{\mathbf{G}}_\mathbf{k} = \mathbf{S}_\mathbf{k}^{1/2}\mathbf{G}_\mathbf{k}\mathbf{S}_\mathbf{k}^{1/2}, \tag{5.16}$$

$$\tilde{\mathcal{A}}_{\mathfrak{c},\mathbf{k}} = \mathbf{S}_\mathbf{k}^{1/2}\mathbf{G}_\mathbf{k}\mathbf{\Gamma}_{\mathfrak{c},\mathbf{k}}\mathbf{G}_\mathbf{k}^\dagger\mathbf{S}_\mathbf{k}^{1/2}. \tag{5.17}$$

Unfortunately the Löwdin transformation delocalises the orbitals creating long-range interactions non-ideal for efficient block tri diagonal inversion algorithms. To the authors knowledge no strict formulation of bond currents in a non-orthogonal basis set have been achieved[124, 126]. However, through tests it is found that orbital currents calculated equivalently to Eq. (5.12) with a non-orthogonal basis set provide significant physical intuition, nonetheless. Hence, the bond currents shown later in Sec. 7.1 will be calculated in a non-orthogonal basis, well knowing that a transformation to an orthogonal basis is required for correct definition. For a non-orthogonal basis set the bond-current becomes [124, 126]

$$\mathcal{J}_{\mathfrak{c},\mathbf{k},\nu\mu} = \frac{1}{\hbar}\mathbb{Im}\left[\mathcal{A}_{\mathfrak{c},\mathbf{k},\nu\mu}(\mathbf{H}_{\mathbf{k},\mu\nu} - \epsilon\mathbf{S}_{\mathbf{k},\mu\nu}) - \mathcal{A}_{\mathfrak{c},\mathbf{k},\mu\nu}(\mathbf{H}_{\mathbf{k},\nu\mu} - \epsilon\mathbf{S}_{\mathbf{k},\nu\mu})\right], \tag{5.18}$$

In the following, the $\mathbf{k}$ point is omitted without loss of generality.

For atoms with more than one orbital a total bond current from atom I to atom J can be defined as

$$\mathcal{J}_{\mathfrak{c},IJ} = \sum_{\mu\in I}\sum_{\nu\in J}\mathcal{J}_{\mathfrak{c},\mu\nu}. \tag{5.19}$$

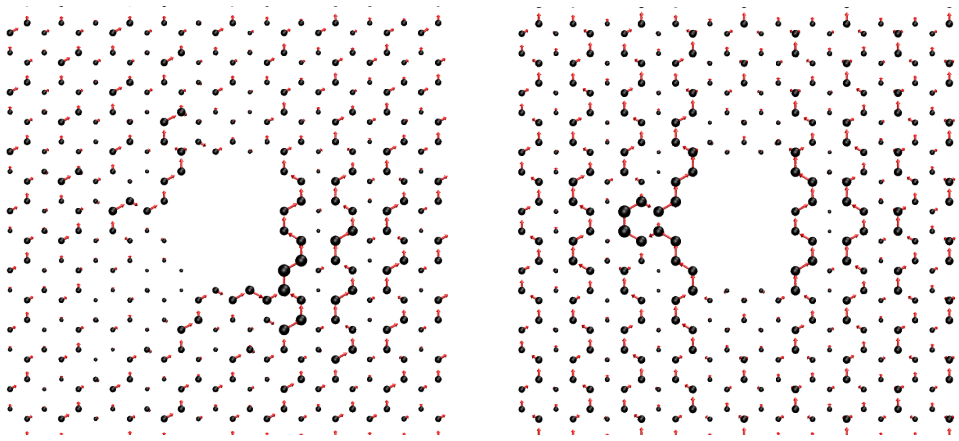


Figure 5.4 | Bond current example of (3rd nearest neighbour, [39, set D; orthogonal]) tight-binding model for periodic graphene flake with an asymmetric hole. Scattering states originate from electrode at the bottom. The left figure are bond currents calculated only using positive $\mathbf{k}$ values (thus imposing TRS). Clearly the momentum tilts the current to the right ($\mathbf{k} \geq 0$ has *right* momentum). The right figure has both positive and negative $\mathbf{k}$ (no TRS) and hence also the average bond current with the left momentum is seen. A very different picture arises although the transmissions are the same.

Further definitions include an absolute atomic current and an absolute orbital current flowing through atom I as

$$\mathcal{J}^{|a|}_{\mathfrak{e},I} \equiv \frac{1}{2} \sum_J |\mathcal{J}_{\mathfrak{e},IJ}|, \tag{5.20}$$

$$\mathcal{J}^{|o|}_{\mathfrak{e},I} \equiv \frac{1}{2} \sum_{\mu \in I} \sum_{\nu} |\mathcal{J}_{\mathfrak{e},\mu\nu}|. \tag{5.21}$$

For atoms with only one orbital they are equivalent, while in general $\mathcal{J}^{|o|}_{\mathfrak{e},I} \geq \mathcal{J}^{|a|}_{\mathfrak{e},I}$. $\mathcal{J}^{|o|}_{\mathfrak{e},I}$ may be much larger than $\mathcal{J}^{|a|}_{\mathfrak{e},I}$ as there may be outwards current and inwards current through orbitals on pair atoms. If there is a lot of circulating current between pair atoms one may get help in distinguishing this via a single atomic quantity which we will call the *activity current*, defined as the geometric mean[6] between the absolute atomic and absolute orbital current

$$\mathcal{J}^{\mathcal{A}}_{\mathfrak{e},I} \equiv \sqrt{\mathcal{J}^{|a|}_{\mathfrak{e},I} \mathcal{J}^{|o|}_{\mathfrak{e},I}}. \tag{5.22}$$

For no *excess* activity we have $\mathcal{J}^{\mathcal{A}}_{\mathfrak{e},I} = \mathcal{J}^{|a|}_{\mathfrak{e},I}$, while a larger number will correspond to an *excess* current circulating between atom I and its connecting atoms. In Sec. 7.1 we will return to the usefulness of Eq. (5.22).

For bond currents with $\mathbf{k}$ points one have to take extra care. One cannot apply time reversal symmetry (TRS) when imaging the bond currents. This can be seen in Figure 5.4 which is an example of a tight-binding calculation for a periodic graphene flake with

[6] One could define the ratio between the absolute and magnitude current, however, the absolute atomic current may often be very small, often zero.

an asymmetric hole and third nearest neighbour interactions, [39, set D; orthogonal]. The bond currents are calculated at −0.375 eV with a full Brillouin zone sampling of 41 **k**-points, carbon atom sizes are scaled according to the absolute atomic current. The scattering states originates from the bottom and exits the system at the top. In the left figure TRS is applied and only the positive $\mathbf{k} \geq 0$ points are sampled yielding a net current-momentum exhibiting the same **k** direction, i.e. a tilt towards the right. Removing TRS, the full Brillouin zone is sampled and the negative **k** points are included, as is visible in the right figure. Clearly the interpretation of the bond currents change drastically!

5.5 Eigenstate projection of transmission

Another way to analyse systems having localised eigenstates is to project the transmission onto these eigenstates [127]. Here we expand on the molecular projected self-consistent Hamiltonian (MPSH) method which importantly is *not* limited to molecules. Moreover, the following discussion only concentrates on $N_{\mathfrak{C}} = 2$ while it is also possible to use for $N_{\mathfrak{C}} \neq 2$. In this section we expand on the possibility of projecting scattering matrices onto localised eigenstates to differentiate between conducting and non-conducting eigenstates. Several important aspects covered in Sec. 2.2 are used throughout.

A multitude of nanoscale electronic calculations using NEGF simulations are based on molecular electronics [128–132]. Within this methodology the understanding of the molecular orbitals is increasingly important in for instance STM experiments where topography maps can differentiate between HOMO or LUMO molecular orbitals [128, 130, 132]. In the following a molecule will be used as an example of localised eigenstates.

We begin by denoting the entire basis set of the system as $\{S\}$. Within $\{S\}$ we can define a subset consisting of an equal or fewer basis components $\{M\}$ with $\{M\} \subseteq \{S\}$, henceforth known as the "molecule". Typically this molecule subset coincides with the device region as covered in Sec. 5.2. A diagonalisation of $\{S\}$ and $\{M\}$ provides the eigenstates for the device and the molecule

$$\mathbf{H}_{\{S\}}|S_i'\rangle = \varepsilon_i^{\{S\}}\mathbf{S}_{\{S\}}|S_i'\rangle, \tag{5.23}$$

$$\mathbf{H}_{\{M\}}|M_i'\rangle = \varepsilon_i^{\{M\}}\mathbf{S}_{\{M\}}|M_i'\rangle, \tag{5.24}$$

where $|A_i'\rangle$ are the generalised eigenvectors with associated eigenvalues $\varepsilon_i^{\{A\}}$ defined in the basis functions of the non-orthogonal basis. Note that in a non-orthogonal basis subset $\langle A_i'|\mathbf{S}_{\{A\}}|A_j'\rangle = \delta_{ij}$ still applies. To create orthogonal eigenvectors we rotate the basis set to form orthonormal projection vectors $|A_i\rangle \equiv \mathbf{S}_{\{A\}}^{1/2}|A_i'\rangle$ such that $\langle A_i|A_j\rangle = \delta_{ij}$, i.e. the Löwdin transformation [125]. We note that the eigenset $\{M\}$ will only coincide with the eigenset $\{S\}$ if the subset $\{M\}$ is fully decoupled from the complement basis set $\{C\} = \{S\} \setminus \{M\}$. In cases where the molecule is only slightly hybridised by $\{C\}$, one can assume that the molecular eigenstates are sufficiently describing a subset of the device eigenstates. The transmission through molecular orbitals can be adequately attributed close to this limit.

Lets rewrite (assuming **k** and ϵ dependencies implicit, and $\{\mathfrak{C}, \mathfrak{C}'\} = \{L, R\}$)

$$\mathcal{T}_{L\to R} = \mathrm{Tr}[\mathbf{\Gamma}_R\mathbf{G}\mathbf{\Gamma}_L\mathbf{G}^\dagger] \tag{5.25}$$

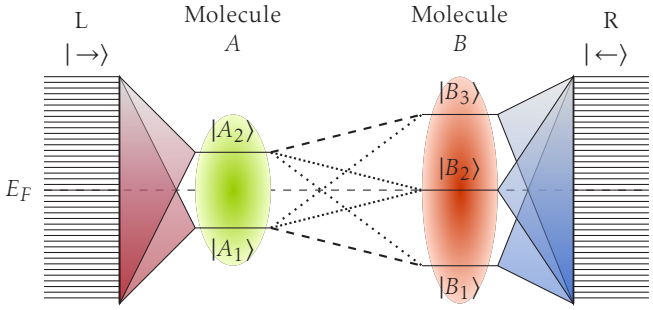


Figure 5.5 | Schematic illustration of two molecules, A (2 states) and B (3 states), coupled to 2 electrodes, L/R. One can follow each of the lines connecting molecule A and B. An electron has to scatter across the molecular states $|A_1\rangle$, $|A_2\rangle$, $|B_1\rangle$, $|B_2\rangle$ and $|B_3\rangle$ to reach electrode R.

inserting $\mathbf{1}$ on each side of the scattering states, implicit $\sum\left(|\cdot\rangle\langle\cdot|\right)$ parenthesis

$$= \mathrm{Tr}\left[\sum_j |S_j\rangle\langle S_j|\mathbf{\Gamma}_R \sum_j |S_j\rangle\langle S_j|\mathbf{G}\sum_i |S_i\rangle\langle S_i|\mathbf{\Gamma}_L \sum_i |S_i\rangle\langle S_i|\mathbf{G}^\dagger\right]. \tag{5.26}$$

Instead of using $\{S\}$, we can use $\{M\}$ which requires the scattering states to only *exist* on this subset. This requirement can be forced by down-folding the self-energies onto the subset $\{M\}$ by selecting $\{M\}$ as the "device" region. Hence by reducing the boundaries for the transport calculations we can project onto $\{M\}$ instead of $\{S\}$ and *select* for which eigenstates we calculate the transmission. Note that the choice of projectors is not immediately obvious as also discussed in [133]. An investigation of different projectors is outside the scope of this work and we will use projectors comprising those of the form in Eq. (5.26).

As an example we sketch a possible two molecule projection in Figure 5.5. It comprises of two electrodes (densely packed lines) and two molecules A and B. Molecule A consists of 2 localised eigenstates whereas molecule B consists of 3 localised eigenstates. The electron may propagate onto molecule A through either $|A_1\rangle$ or $|A_2\rangle$ and further propagate onto molecule B. Lastly it may propagate onto the right electrode through either $|B_1\rangle$, $|B_2\rangle$ or $|B_3\rangle$, or any combination of the three. These choices are schematised using the different connecting lines. One choice would be the transmission through $|A_1\rangle$ and through $|B_{\{1,2\}}\rangle$,

$$\mathcal{T}_{A_{\{1\}}B_{\{1,2\}}} = \mathrm{Tr}\left[\sum_{j=1}^{2} |B_j\rangle\langle B_j|\mathbf{\Gamma}_R \sum_{j=1}^{2} |B_j\rangle\langle B_j|\mathbf{G}|A_1\rangle\langle A_1|\mathbf{\Gamma}_L|A_1\rangle\langle A_1|\mathbf{G}^\dagger\right], \tag{5.27}$$

and by defining the symbol $|i\rangle\overset{A,\mathbf{\Gamma}}{\langle i|j\rangle}\langle j| \equiv |A_i\rangle\langle A_i|\mathbf{\Gamma}|A_j\rangle\langle A_j|$ we find

$$\begin{aligned} &= \mathrm{Tr}\Big[\Big(|1\rangle\overset{B,\mathbf{\Gamma}_R}{\langle 1|1\rangle}\langle 1| + |2\rangle\overset{B,\mathbf{\Gamma}_R}{\langle 2|2\rangle}\langle 2| + |1\rangle\overset{B,\mathbf{\Gamma}_R}{\langle 1|2\rangle}\langle 2| + |2\rangle\overset{B,\mathbf{\Gamma}_R}{\langle 2|1\rangle}\langle 1|\Big) \\ &\qquad \mathbf{G}|1\rangle\overset{A,\mathbf{\Gamma}_L}{\langle 1|1\rangle}\langle 1|\mathbf{G}^\dagger\Big]. \end{aligned} \tag{5.28}$$

All scalar quantities $\overset{A/B,\mathbf{\Gamma}}{\langle i|j\rangle}$ can be saved as an extra level of information. Hence $\overset{B,\mathbf{\Gamma}_R}{\langle 1|2\rangle}$ yields the right second molecular orbital projected onto the scattering states and the

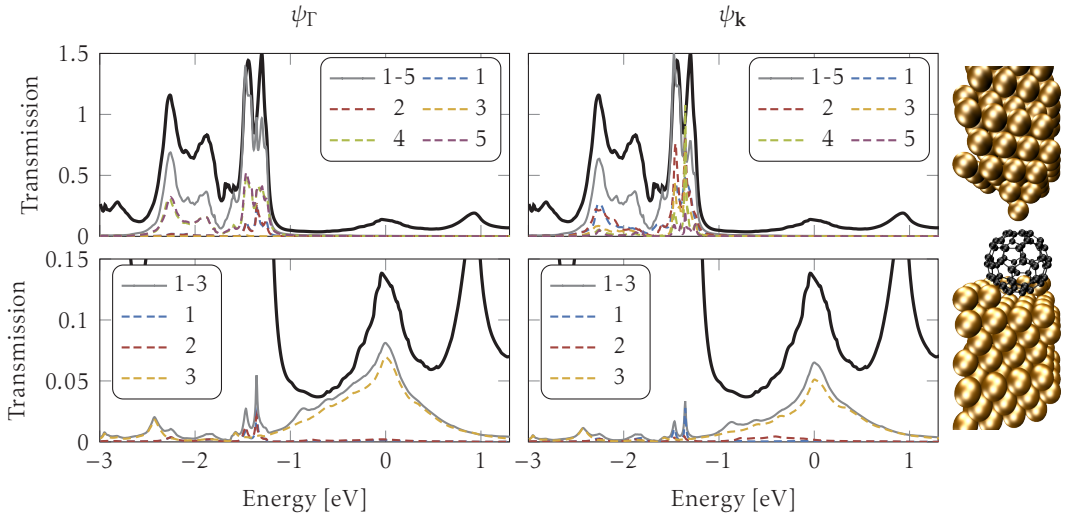


Figure 5.6 | Projected transmission on monolayer C_{60} molecule adsorbed on a Cu(111) surface. Projections of the 5 degenerate HOMO, $E < -1\,\text{eV}$, (top) levels as well as the 3 degenerate LUMO, $E = 0\,\text{eV}$, (bottom) levels are show with both non-dispersive (left) and dispersive (right) projectors. The black transmission curve is the non-projected transmission. Particularly for the LUMO projections in the dispersive projection one finds a better Lorentzian at $E = 0\,\text{eV}$ compared to the non-dispersive. This reveals the dispersive nature of the monolayer. Generally, projections describe well the HOMO/LUMO peaks while it is found that only a single LUMO eigenstate contributes to the transmission.

probability that this overlaps with the first molecular orbital. Essentially the projection removes the scattering states that are not the molecule eigenstates and retains only the scattering states which enters the projection molecule eigenstates. Consequently using the projected scattering states the spectral density matrix may also be used to track the density of states originating from a particular projected eigenstate. Importantly the projection states can either be real or complex eigenstates corresponding to Γ or $\mathbf{k}$ resolved projections. If the molecular eigenstates are non-dispersive in the Brillouin zone they yield the same result.

The projection method where used to estimate the molecular eigenstate transmissions through a C_{60} molecule [*134*]. In this setup the molecule is protruding the surface layer "digging" a hole in the cobber (111) surface. Due to the densely packed monolayer coverage of C_{60} there exists a slight dispersion in the Brillouin zone on the order of meV. The transmissions are calculated on a 13×13 Monkhorst-Pack grid. Figure 5.6 show a comparison of the full transmission (thick black) vs. the projected transmissions. The projections used are

$$\mathcal{T}_{L\{\}\to R\{\}} = \mathrm{Tr}\Big[\sum_{j,j'\in\{C_{60}\}}^{C_{60},\mathbf{\Gamma}_R} |j\rangle\langle j|j'\rangle\langle j'|\mathbf{G}\sum_{i,i'\in\{C_{60}\}}^{C_{60},\mathbf{\Gamma}_L} |i\rangle\langle i|i'\rangle\langle i'|\mathbf{G}^\dagger\Big], \tag{5.29}$$

i.e. a projection of both scattering states onto the molecular orbitals. The figures in the left column are the Γ-point projectors used in the entire Brillouin zone, while the right column figures are the $\mathbf{k}$ dependent projectors. All single integer projections are single molecular level projections, while $1\ldots X$ are the summed projector over the HOMO ($X = 5$) or LUMO ($X = 3$) levels. The HOMO levels cover a range between $-2.5\,\text{eV}$

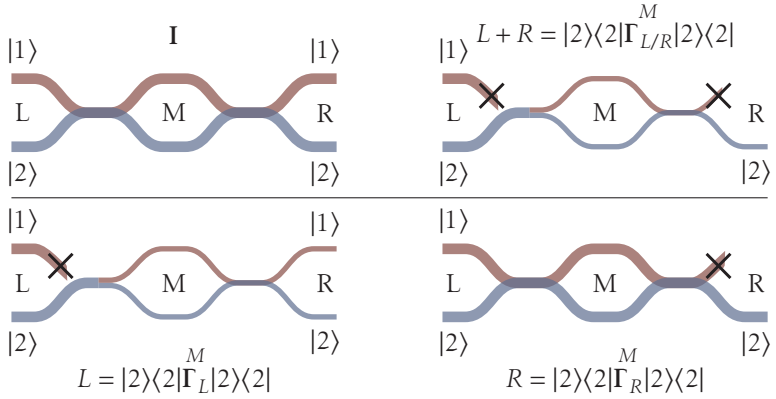


Figure 5.7 | Diagram for none(top-left), L(bottom-left), R(bottom-right) or $L+R$(top-right) projections. Region L/R, denote the incoming/outgoing scattering states while M is the traversing projected molecule. We use a 2 eigenstate molecule, $|1\rangle$ and $|2\rangle$, with projection onto $|2\rangle$. If both the left-right scattering states are projected the incoming and outgoing eigenstate $|1\rangle$ is filtered out. This is marked by ×. If only the left scattering state is projected the outgoing states may exit through either of the molecular states, while if only the right scattering state is projected the outgoing states may only exit through the projected state. Using $L+R$ both incoming and outgoing scattering states are projected.

to −1.25 eV while the LUMO peak is more localised at 0 eV. The hybridisation of the molecule with the surface is clear from the degeneracy lifting of the C_{60} molecule[7]. For the LUMO transmission projections labelled 2 and 1-3 we have

$$\mathcal{T}_{L^2\to R^2} = \mathrm{Tr}\Big[\sum_{j,j'\in\{2\}} |j\rangle\langle j|j'\rangle\langle j'|^{\mathrm{C}_{60},\mathbf{\Gamma}_R}\mathbf{G}\sum_{i,i'\in\{2\}} |i\rangle\langle i|i'\rangle\langle i'|^{\mathrm{C}_{60},\mathbf{\Gamma}_L}\mathbf{G}^\dagger\Big], \tag{5.30}$$

$$\mathcal{T}_{L^{\{1,2,3\}}\to R^{\{1,2,3\}}} = \mathrm{Tr}\Big[\sum_{j,j'\in\{1,2,3\}} |j\rangle\langle j|j'\rangle\langle j'|^{\mathrm{C}_{60},\mathbf{\Gamma}_R}\mathbf{G}\sum_{i,i'\in\{1,2,3\}} |i\rangle\langle i|i'\rangle\langle i'|^{\mathrm{C}_{60},\mathbf{\Gamma}_L}\mathbf{G}^\dagger\Big]. \tag{5.31}$$

As can be seen in the figure, a clear indication of which molecular orbitals contribute and which does not contribute to the transmission. Especially for the LUMO peak it is clear that only a single LUMO level is contributing. Further, the projected transmissions captures the separation of transmission through either HOMO or LUMO levels in the appropriate energy ranges. The dispersion of the eigenstates is also visible in the difference between the Γ and **k** transmissions. Clearly the **k** resolved transmission provides a sharper peak at the LUMO peak as would be expected for a Lorentzian broadened eigenstate.

An important detail of the projections is that the summed projections $1\ldots X$ does not sum up to the total transmission for the LUMO peak. This may be related to the method of projection. In the implemented projection method in TBtrans one may use either of three methods. These different methods are outlined in Figure 5.7 where a model of a molecule (M) connected to two leads L/R is shown. The molecule consists of 2 eigenstates $|1\rangle$ and $|2\rangle$. For no projection, top-left, the two eigenstates may scatter onto the molecule and may split into either of the eigenstates. The scattering state origin is lost as there may be mixing. Subsequently the same occurrence is found at the right electrode. In

[7] The lone C_{60} molecule has a 5-degenerate HOMO level and a 3-degenerate LUMO level.

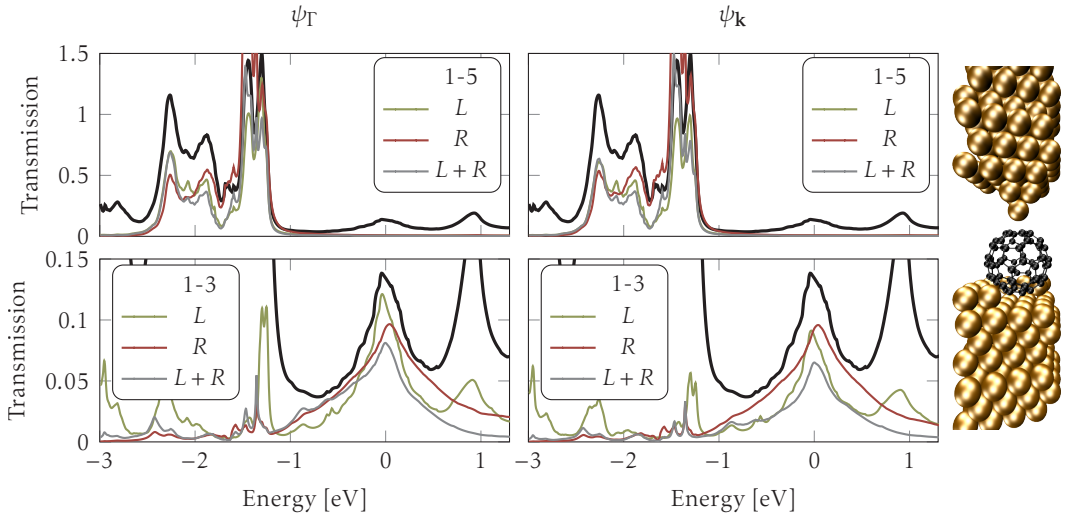


Figure 5.8 | See Figure 5.6 for details. Different methods of projections compared. Bulk/L only projects the incoming scattering states onto the molecular orbitals. STM/R only projects the outgoing scattering states onto the molecular orbitals. $L+R$ performs both projections. For the LUMO peak 1-3 the $L+R$ restricts the transmission which indicates a splitting of the projected scattering states. This is confirmed in the L projection where high peaks occur in the HOMO and LUMO+1 energy range. As the tip is weakly coupled to the C_{60} a small hybridisation occurs and the splitting is thus only present for L and not R.

bottom-left (method L) the incoming scattering states are projected *only* onto $|2\rangle$ which removes the $|1\rangle$ contribution. However, the incoming state may mix and leave in either of the two eigenstates. In the bottom-right figure (method R) the incoming scattering state is unchanged while the outgoing scattering state may only leave through $|2\rangle$. The full projection, top-right (method $L+R$), projects both the incoming and outgoing scattering states. The above three methods may indeed yield *very* different results. However, for geometry symmetric junctions at $V = 0$, or symmetric DOS, L and R methods are equal. Remark that transmission symmetry still applies.

In Figure 5.8 the three methods are compared for the same C_{60} system. The $L+R$ projections are the equivalent summed projections in Figure 5.6. The difference is mostly visible on the LUMO peak. The $L+R$ has the lowest transmission which indicates a mixing of the projected scattering states once they enter the molecule. This is confirmed in the L projection on LUMO 1-3 where high peaks occur in the HOMO and LUMO+1 energy range, suggesting a large hybridisation with the surface. Neither of the other two projections exhibit this mixing which we attribute to the weak coupling between the tip and the C_{60} molecule. Another feature on the R projections is that they exhibit little to no difference between Γ and $\mathbf{k}$ resolved projections. This is because the tip scattering states are non-dispersive, unlike the bulk electrode scattering states.

5.6 Summary

In this Chapter the basis for tight-binding calculations with Green functions have been presented. The focus has been on the transport code TBtrans and its implementation.

The basic quantities such as, density of states, spectral density of states, transmission spectra, transmission eigenchannel, electron thermal energy exchange, non-equilibrium current, bond-currents and molecular projected transmissions. All these quantities are calculated in a so-called "device" region which is the primary calculation unit that is the basis for all quantities.

Calculating the elastic current is performed by integrating the transmission function. To increase throughput with little loss of accuracy two interpolation schemes were presented. It was shown that spline interpolations was superior to the simplest linear interpolation scheme for an example system.

Bond-currents (or orbital currents) are describing the local current flow which is used for intuitive understanding of current flow in nanostructures. We have presented the importance of proper $\mathbf{k}$ point sampling where one *must* image both $\pm\mathbf{k}$ for interpretation. We introduced the *activity* current which an atom is responsible for is important when dealing with atoms having more than one orbital.

Lastly, we presented calculations of molecular orbital projections of the transmission function. This distinguishes between transmitting and non-transmitting molecular eigenstates. Detailed studies of molecular junctions can be performed via different projections, which may differ in terms of which scattering states are projected. It was shown that $\mathbf{k}$ resolved molecular projections are needed for dispersive molecular orbitals.

Chapter 6

Electrostatic gating 2D materials

A unique feature of graphene electrodes is that their electronic properties can easily be tuned by electrostatic gating. In fact, electrostatic gates can be used to increase the carrier density in graphene to above $10 \times 10^{13}\,\mathrm{cm}^{-2}$ [27]. For ion gating it has even been possible to reach carrier densities of $10 \times 10^{14}\,\mathrm{cm}^{-2}$ which correspond to a Fermi energy shift of about 1 eV [135]. For non-graphene electrode devices it has been shown that gating can tune resonances localised in the narrowest part of the junction, [26, 136] as the electronic states in the electrodes are affected only weakly by the gate-induced capacitive field [137]. However, for graphene electrodes, the low DOS and flat geometry, makes it comparable to the junction, and is thus likely to be perturbed similarly by gating. This peculiarity leads to a novel, yet largely unexplored, paradigm for graphene-based electronics. In effect graphene electrodes becomes a part of the device as they too are influenced by the gate. In electronic transport simulations, the effect of electrostatic gating and induced charge doping in the device has often been mimicked by rigidly shifting the position of the Fermi-level/chemical potential in calculations without explicitly including gate or dopants [18, 36, 138, 139]. However, despite accounting for some of the effects, these approaches neglect the self-consistent response of the device to the additional charge doping or the gate-induced electric field.

This chapter will start by introducing the empirical gate model. The model will be used on a graphene/graphite slab and also on a graphene nanoconstriction of which [*140*] is the resulting published paper.

6.1 Empirical gate model

A field effect setup consists of a gate, a dielectric and the system itself, in that stacking order. Applying a gate voltage charges the system like a capacitor setup, thus inducing an electrostatic potential gradient across the dielectric. The additional charge will redistribute to create a polarisation in the system along the electric field direction.

Such field effect setups can be realised in DFT software by employing a non-equilibrium Green function (NEGF) scheme [11], or by solving the Poisson equation with appropriate boundary conditions [47, 72]. The former is an expensive calculation while the latter method is relatively inexpensive.

The implementation can be schematised as shown in Figure 6.1 and it is instructive

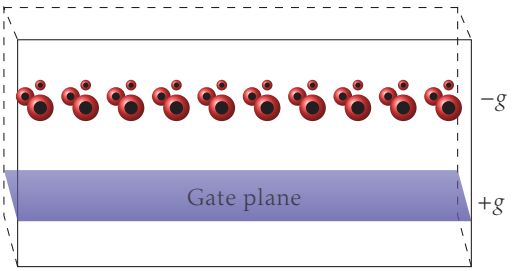


Figure 6.1 | Schematic view of the gate implementation by redistribution of charge. The gate electrode has a certain charge g (blue) and charges the system by an equal negative amount $-g$ (red), atoms shown as black dots. The charge redistribution for the system and the electric field are self-consistently calculated.

to have the equations governing a parallel plate capacitor setup in mind:

$$\mathbf{E} = \frac{V}{d} = \frac{g}{\epsilon_r \epsilon_0} \tag{6.1a}$$

$$V = \mathbf{E}d = \frac{gd}{\epsilon_r \epsilon_0} \tag{6.1b}$$

$$E = -Vq = \mathbf{E}dq = \frac{gqd}{\epsilon_r \epsilon_0}, \tag{6.1c}$$

where g is the charge on the gate(same and opposite charge in the system), d is the distance between the two plates, q a charge. V, E, and $\mathbf{E}$ are the potential difference, potential energy and electric field. ϵ_0 is the vacuum permittivity and ϵ_r is the relative permittivity. Analogous to a plate capacitor setup we assume that an applied gate voltage induces a charge $-\delta e^-$ in the system and a corresponding counter-charge $+\delta e^-$ in the gate plane. This situation is accounted for by charging the system with a given charge $g = -\delta e^-$, and by distributing homogeneously the corresponding counter-charge $+\delta e^-$ in a well defined region of the unit-cell, denoted *gate*, so that the overall system+*gate* remains charge neutral. Thus for $g > 0$ we have a p-doped system, similarly for $g < 0$ we have a n-doped system. Solving the Poisson equation inherently calculates the electric field between the gate and the system. As the calculation cell is periodic we apply the slab dipole correction [72] to terminate the periodic electric field induced by the charge redistribution. Another method is via changing the periodic boundary conditions [47]. In addition to the planar gates this method can readily be employed for arbitrary gate configurations. For instance a planar gate with an additional spherical charge which would correspond to a charge impurity in the gate medium. Such a gate model is referred to as *charge* gates. However, the gate model has a deficiency when the charge redistribution is not mirrored completely in the gate medium.

The gating method can also be adopted in transport calculations using NEGF if the gate is uniformly applied to the electrodes *and* the device[1]. Additionally, the gate at the electrodes must have a resulting electric field perpendicular to the applied bias to assert the correct boundary conditions. Our implementation resembles that of Brumme *et al.* [48, 49] except that we use an LCAO model, opposed to a plane-wave basis, which

[1] Strictly we only require the boundary conditions to be correctly described and hence *any* gate configuration which fulfils this may be used.

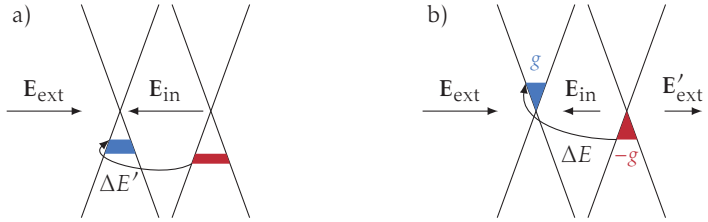


Figure 6.2 | Applied electric field on a slab of two non-interacting graphene sheets. The electric field will induce a charge-redistribution which will cost up to the potential energy that the electric field can supply $E_{\text{ext}} = gd\mathbf{E}_{\text{ext}}$, with d being the interlayer distance. a) Charge redistribution fully screens the electric field as $\Delta E < E_{\text{ext}}$. b) $\Delta E > E_{\text{ext}}$ and field penetrates, resulting in a non-vanishing $\mathbf{E}'_{\text{ext}}$.

means that the dielectric need not, necessarily, be simulated by a potential barrier to limit electronic penetration.

In conjunction with the charge gate model a Hartree gate model has also been implemented. This is an already widely used method and has already proven quite useful [16, 50, 88]. The Hartree gate model can be defined via this simple correction to the Hartree potential

$$V_H(\mathbf{r}) \leftarrow V_H(\mathbf{r}) + V_G(\mathbf{r}), \tag{6.2}$$

where $V_G(\mathbf{r})$ is the gate potential defined via planes, geometric spheres and/or boxes. Although exhibiting the electronic repulsion/attraction due to $+/-|V_G(\mathbf{r})|$ it does not add an explicit charging effect of the system. For metallic systems such a gate is sufficient while for low-DOS systems a charge gate is needed to account for charge redistribution.

Thus TranSIESTA is capable of calculating charging effects due to capacitor like setups while simultaneously adding repulsive/attractive potentials for a full control of the environment. In the following we will only concentrate on the charge gate model.

6.2 Gating N-layer graphite

The electrostatic dependence of graphene/graphite is an on-going interest for the field of graphene based nano-devices. Santos and Kaxiras have investigated [137] the field-dependence of pristine graphene and graphite with remarkably close results to experimental results [17, 141–146]. The theoretical studies were conducted using SIESTA however, using a constant electric field across the entire slab and it restricts it self to the Dirac point (E_D), i.e. no charge gating of the slab. Thus any redistribution of electrons occurs around the Dirac point where the DOS is low and the quantum capacitance is high [30, 147–149]. The quantum capacitance is an effect only arising in low DOS regimes and hence for small gating values, $|E_F - E_D| < 100\,\text{meV}$ [148]. In this section we will highlight the importance of taking into account the electric field *and* charging of a (few-layer) graphene system.

A simple picture (plate capacitor) describing the competing effects is shown in Figure 6.2. The quantum capacitance effect can be understood in terms of two non-interacting graphene sheets with a single gate. The two sheets tries to screen any applied electric field by inducing charge polarisation. However, the electric field can only be fully compensated if the energy cost of moving the electrons is lower than the potential energy in the electric field, Eq. (6.1). For gated graphene (a) the the energy cost of moving

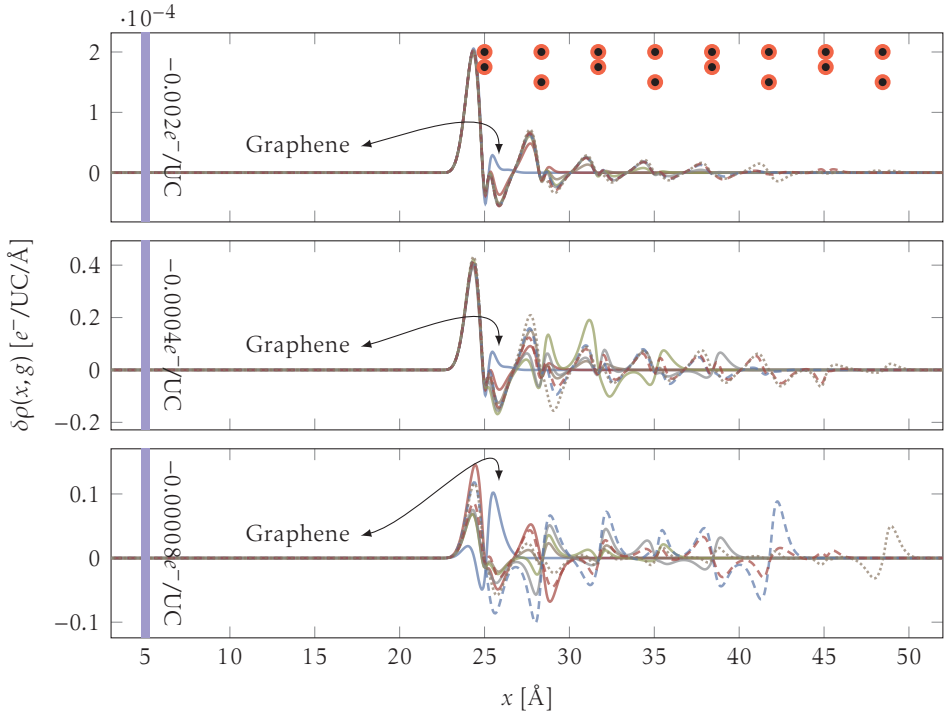


Figure 6.3 | Three different gating levels, each with 8 different number of layers in the graphite slab positioned as indicated in the top plot. The top figure has the highest gate level decreasing to the lowest gating level for the bottom plot. The top plot has a smoothly decaying screening length, adding additional layers have minimal influence on the preceding layers, which results in a short screening length. The middle configuration barely screens the field in the slab. The different layers have different screening properties which indicates a high barrier for interlayer charge redistribution. For the lowest gating level a penetration through all layers is present and the 8 different systems behave differently all together. This regime may in the limit of quantum capacitance.

charge from one layer to the other is relatively low. This is because the DOS is much higher at the Fermi level. For non-gated graphene (b), $E_F = E_D$, the DOS close to the Fermi level is very low and a large energy range is required to compensate the electric field. As such the screening capability is an effect of both the magnitude of the electric field *and* the gating level.

The following discussion is based on calculations made with SIESTA, a DZP basis set and a 44×44 Monkhorst-Pack grid, mesh cutoff of 276 Ry, exchange-correlation GGA-PBE [63].

Understanding the charge/gate effects of graphene/graphite layers become a non-trivial task as it is a non-linear effect. In Figure 6.3 we plot the charge redistribution due to the gate model described in Sec. 6.1. Here a symmetric gate/doping is used. A figure to remember for graphene is $\mathbf{E}_{g/\mathrm{UC}} = 33.4\,\mathrm{V/Å}$. This is the electric field for a plate capacitor charged with one electron on an area of the graphene unit-cell. The charge distribution is calculated via

$$\delta\rho(x,g) = \iint \mathrm{d}y\mathrm{d}z\rho_g(x,y,z) - \rho_0(x,y,z). \tag{6.3}$$

Every plot has 8 different lines corresponding to the number of layers in the graphite

slab. The top plot has graphene layers indicated by atoms (black dots), with indicated removal of charge (red). The far left blue region indicates the position of the gate and the associated charge added on the gate. The top figure has the highest electric field and the highest gating level, 0.002 e/UC ($E_F - E_D = 177$ meV for graphene), whereas the bottom figure has the lowest electric field and lowest gating level, 8×10^{-5} e/UC ($E_F - E_D = 84$ meV for graphene). Clearly there are two effects; 1) an intralayer redistribution due to polarisation across single layers; 2) an interlayer redistribution, long range polarisation. Comparing the top and bottom plots it is clear to see the electric field is fully screened in the top plot while the bottom plot seem only partially screened. The middle gate plot is also partially screening the electric field. The highly gated structure has a much shorter screening length while the lowest gating level does not screen nearly as good. However, the charge distribution in the first layer for the two top plots are equal for all different number of layers. This indicates a limit to the charge redistribution at that gate and electric field level for few layers of graphene. The high charge redistribution on the far side of the low gated system may be artefacts from the periodic calculation.

To conclude on this study; applying electric fields to <3D materials is a highly non-trivial task. The intrinsic quantum capacitance is creating non-linear effects. Indeed system properties are dependent on electric field strength and charge redistribution which is traced back to the low DOS region close to the Fermi level. Future studies should include gating effects such as those covered here. Additionally, a non-even gate/electric field complicates the analysis further. This is important as the electric field may be different than the equivalent capacitor setup. It is however, important to consider when additional non-linearity can be exposed via simple models which for instance may apply when the gating shifts the Fermi level far from the Dirac point.

6.3 Gating nanostructured graphene

In the following we extend the analysis of the gating method by introducing non-equilibrium effects using TranSIESTA NEGF capabilities. This combines the complexity of the gate method *and* the complexity of NEGF method.

Our systems are two geometrically similar, "left-right" symmetric graphene nanojunctions, formed by a nanoribbon (GNR) connected to pristine graphene electrodes, see Figure 6.4 and Figure 6.5. For zero gate/doping ($g = 0$) the former yields an electron-hole symmetric electronic structure (Hydrogen GNR), whereas the latter yields a e–h asymmetric electronic structure (Oxygen GNR) [150–152]. The gate is placed 20 Å beneath the graphene plane and we sweep the gating levels, g, according to $g \times 10^{13}\, e\,\mathrm{cm}^{-2}$.

In Figure 6.4, we plot the potential drop across the graphene constriction at 0.5 V for $g = 0$, a), and n-doped with $g = -2$ gating, b). The potential profile has been integrated in the perpendicular direction to the graphene surface for electronic densities above $\rho > \rho_\epsilon = 0.008\, e\,\mathrm{Å}^{-3}$ projected onto the x–y plane. The lower panels, c) and d), is a further projection onto the transport direction (x) as indicated in Figure 6.4b. At $g = 0$ we obtain an anti-symmetric potential drop in the transport direction ($\Delta V(x) = -\Delta V(-x)$) as expected for a fully e–h and left-right symmetric constriction. On the other hand, for $g = -2$ (n-doped), we see a clear pinning of the potential profile to the positive electrode, i.e. the potential drop at the negative electrode. Conversely, a calculation with $g = +2$ (p-doping) with 0.5 V display a pinning at the negative electrode, while for $g = +2$ and -0.5 V we regain the plot shown. This confirms the geometric symmetry.

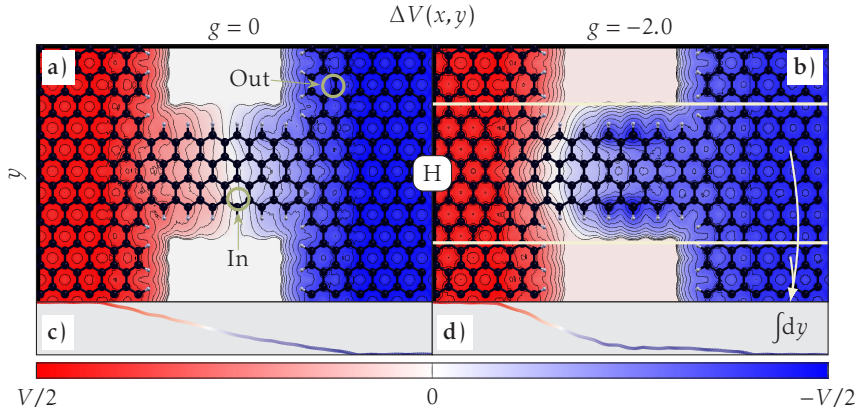


Figure 6.4 | Electronic Hartree potential drop integrated perpendicular to the plane and above a cutoff electron density $\rho_\epsilon = 0.008\, e\,\text{Å}^{-3}$ and projected to the graphene plane for the Hydrogen GNR, a), b). c) and d) are the contour plot further integrated in the box indicated in b). The non-gated system shows a linear gradient, whereas for $g < 0$ (n-doped) a pinning of the potential towards the right (positive) electrode.

In Figure 6.6 we show the transmission spectra for the hydrogen passivated constriction at 0 V, a), and 0.5 V, b), for different values of g each vertically shifted 1/2. As a measure of gating we track the position of two resonances, □ and ○ marks, corresponding to a resonance in the constriction located at the edge and in the center, respectively. The middle thick line is the transmission for $g = 0$, and is equivalent to earlier results where these resonances are discussed [36]. In addition, we plot the energy shift of the Dirac point for pristine graphene as vertical lines aligned at each of the two resonances at $g = 0$. These vertical lines match exactly the shift in chemical potential due to the doping in the electrodes. Discrepancies between the electrode gating (lines) and the resonance positions (dots) illustrate the difference in just rigidly shifting the resonances according to electrode doping, and a fully self-consistent calculation of the resonance positions. Importantly, at 0 V we find that the resonance peaks does *not* simply follow the gating. Moreover, the two peaks are shifting/gated independently of each other; the center resonance peak, ○, follows the pristine doping closer than the edge resonance peak, □, due to a difference in electrode coupling between the resonances. On the other hand, at 0.5 V we find that both peaks follow the pristine graphene electrode doping. As shown in Figure 6.4b, the junction behaves as an extension of the positive electrode and therefore the resonance position is pinned at the Fermi level of this particular electrode. The self-consistent calculation is needed to capture the correct transition with bias from semi-independent resonances to the pinned behavior.

Figure 6.5 are for the Oxygen terminated graphene nanoribbon. This nanoribbon has no e–h electronic DOS symmetry [150–152]. Similarly to the Hydrogen system we calculate for $g = 0$ and $g = -2$ at 0.5 V. a) shows that the Oxygen edges pins slightly to the negative electrode for zero gating, while gating, b), the entire ribbon is pinned to the positive electrode, equivalent to the Hydrogen case Figure 6.4b. This is also seen in the projected potential profiles Figure 6.5c and d). This confirms that the selectivity of the potential profile in the gated devices does not rely on the e–h symmetry of the junction, and conjectures the generality of this behaviour in systems with electrodes having low DOS around E_F, regardless of the electronic structure of the central part connecting the

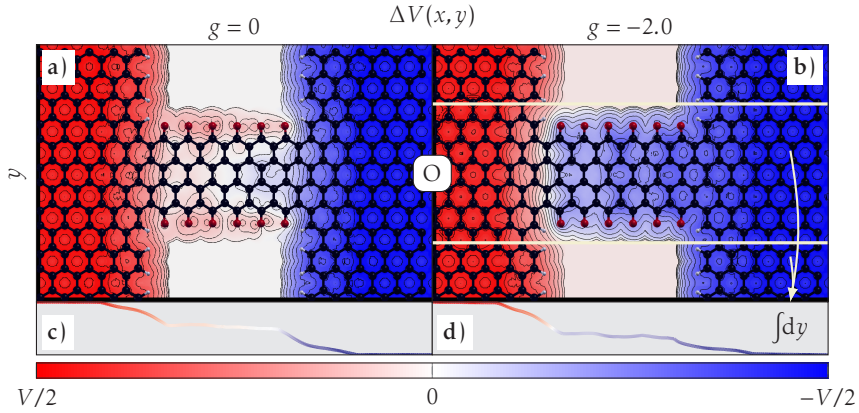


Figure 6.5 | Electronic Hartree potential drop integrated perpendicular to the plane and above a cutoff electron density $\rho_\epsilon = 0.008\, e\,\text{Å}^{-3}$ and projected to the graphene plane for the Oxygen GNR, a), b). c) and d) are the contour plot further integrated in the box indicated in b). The non-gated system shows a gradient at the GNR boundary, whereas for $g < 0$ (n-doped) a pinning of the potential towards the right (positive) electrode.

two electrodes.

The generic behavior of the potential drop just outlined is summarised in Figure 6.7, which shows the one-dimensional potential drop calculated for the hydrogen-terminated constriction for a number of different gates and positive bias voltages, similar to that of Figure 6.4c and d). Independently on the particular value of the bias voltage applied, gating the system always leads to a marked asymmetry of the potential drop across the constriction. For any value of n-doping, the potential drop pins always to the positive (right) electrode for positive bias. Similarly, for any value of p-doping, the system couples to the negative (left) electrode for positive bias. These results further demonstrate the general phenomena that does not depend on the particular values of applied gate and/or bias voltage. Furthermore, our calculations highlight the important fact that the charge neutrality point for the electrodes is a special case which does not extrapolate to the gated case. This becomes even more important if one considers the experimental difficulties in retaining a charge neutral sample [153, 154].

6.3.1 Voltage drop model

We will now consider a simple model which can explain the electrode selectivity of the voltage drop depending on the doping/electrostatic gating. Figure 6.8 is a guided reference for the following discussion. The position of the voltage drop can be obtained by considering the change in charge in the scattering region when applying a bias. If the scattering region becomes more positive, one can view it as the positive electrode extending into the scattering region and thus the voltage drop will occur closer to the negative electrode and *vice versa*. The change in charge in the scattering region is linked to the change in injected charge from left and right electrodes in the bias window, as noted in the Methods section. The linear dependence of the DOS in the graphene electrodes makes the coupling/broadening functions of the scattering region display the energy dependence,

$$\mathbf{\Gamma}_{L/R}(E) \propto |E - \mu_{L/R} + E_F|, \tag{6.4}$$

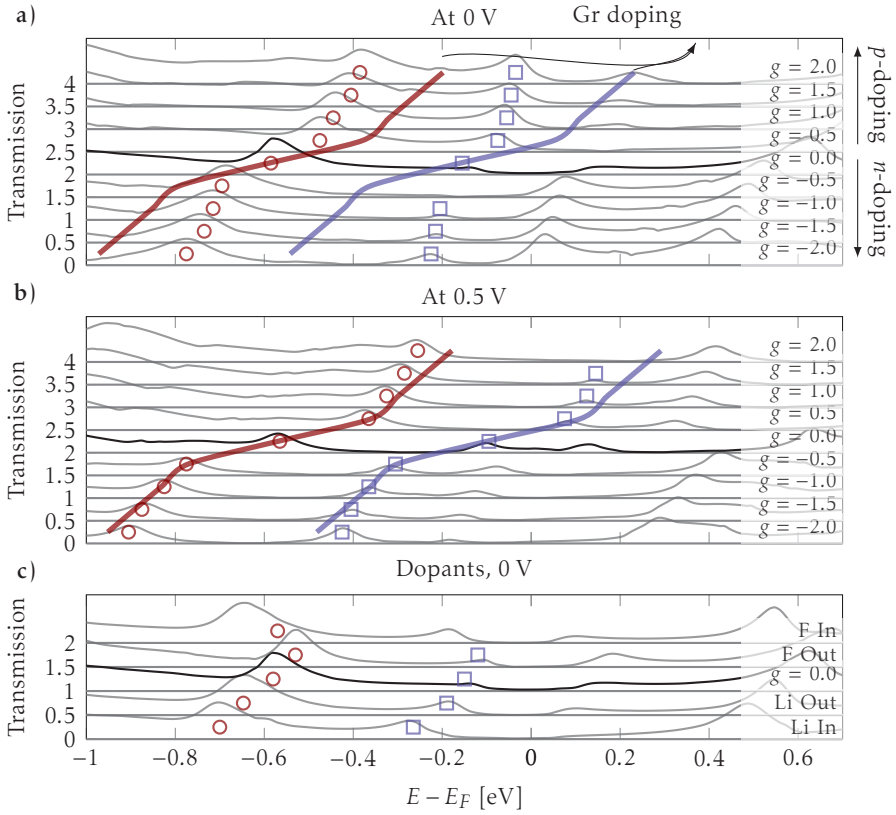


Figure 6.6 | Transmission spectra for the constriction at various doping levels for 0 V, a), 0.5 V, b), and for dopants, c). The middle line (black) at zero gating is a symmetric transmission function with two distinct resonances (marked, □ and ○). Gating the constriction shifts the resonances as indicated by the displacements of the marks. The full lines, crossing vertically the different doping levels, indicates the graphene electrode Fermi level shift due to the doping aligned at the $g = 0$ mark.

where $E = 0$ corresponds to the equilibrium Fermi level, E_F is the shift of Fermi level due to doping, $E_F \propto -g/|g|\sqrt{|g|}$, and $\mu_{L/R}$ is the change in the chemical potential of left/right electrodes with applied voltage bias (V). We will use $\mu_L = eV/2$ and $\mu_R = -eV/2$, and take $V > 0$. This definition means that the scattering region as a starting point will not preferentially select the left or right electrode for an electron-hole symmetric system, and the potential drop profile will be spatially anti-symmetric, $\Delta V(x) = -\Delta V(-x)$. We will now consider the voltage bias as a "perturbation" onto the system without bias, and calculate the change in charge in the scattering region. Thus we first neglect the change in potential set up by the change in charge, which again will impact the charge in the self-consistency. With this we have the density of scattering states from left and right, $\mathcal{A}_{L/R} \propto \mathbf{\Gamma}_{L/R} \propto |E - \mu_{L/R} + E_F|$, and the change in electrons(holes) injected from left(right) electrode can be written as,

$$\delta e = \int_0^{eV/2} \mathcal{A}_L\big(E\big)dE \propto \tfrac{eV}{2}\big(E_F - \tfrac{eV}{4}\big), \tag{6.5}$$

$$\delta h = \int_{-eV/2}^{0} \mathcal{A}_R\big(E\big)dE \propto \tfrac{eV}{2}\big(E_F + \tfrac{eV}{4}\big) \tag{6.6}$$

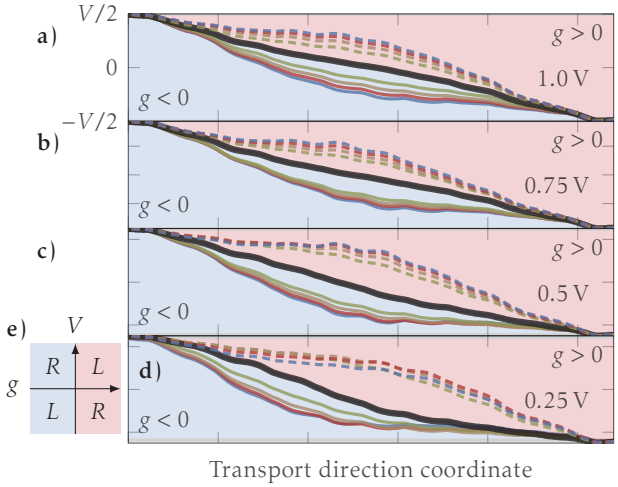


Figure 6.7 | Integrated Hartree potential profile in a region of width corresponding to the ribbon along the entire constriction. The thick middle line is the potential profile for $g = 0$. The blue regions correspond to n-doped graphene (full lines), while red are p-doped graphene (dashed lines). The non-gated calculations show a linear behavior whereas gated systems have a asymmetry between the left and right electrode DOS breaking the left-right anti-symmetry in the potential drop. e) summarises the trends where L/R means pinning to the left/right electrode.

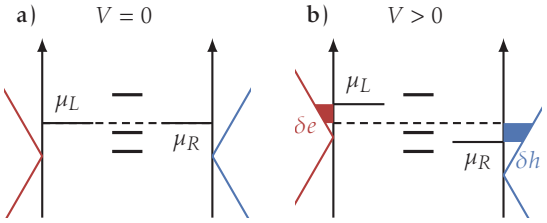


Figure 6.8 | Illustration of asymmetric coupling induced by doping out of symmetry. a) shows the zero bias configuration with broken e–h symmetry due to doping, $g < 0$. b) shows a difference among the electrode contributions in the bias window which *pins* the system to the right electrode.

where we assume $|V/2| < |E_F|$. The scenario is shown schematically in Figure 6.8b showing more injection of positive carriers $\delta h > \delta e$. Thus the scattering region will as the first response to the non-equilibrium filling become more positive and we conclude that for n-doping, $g < 0$ and $E_F > 0$, the positive electrode will "extend" into the constriction resulting in a voltage drop at the negative electrode, as seen in Figure 6.7. We stress that this behavior stems from the vanishing DOS of graphene at the Dirac point yielding a large relative difference between the electron/hole contributions. Contrary if we take E_F to be very large in Eqs. (6.5) and (6.6) we get $\delta e \approx \delta h$ and the constriction does not change its charge. Indeed, the pinning effect is smaller at 1 V compared to 0.5 V as seen in Figure 6.7a vs. c). This is due to the DOS of one lead being very close to zero at 0.5 V; $\mu_i - E_F \approx E_D$ with E_D being the Dirac point, and hence a much larger relative difference in DOS.

In order to substantiate that the voltage drop is controlled by the vanishing electrode DOS we smear the DOS energy dependence gradually into a flat function by introducing

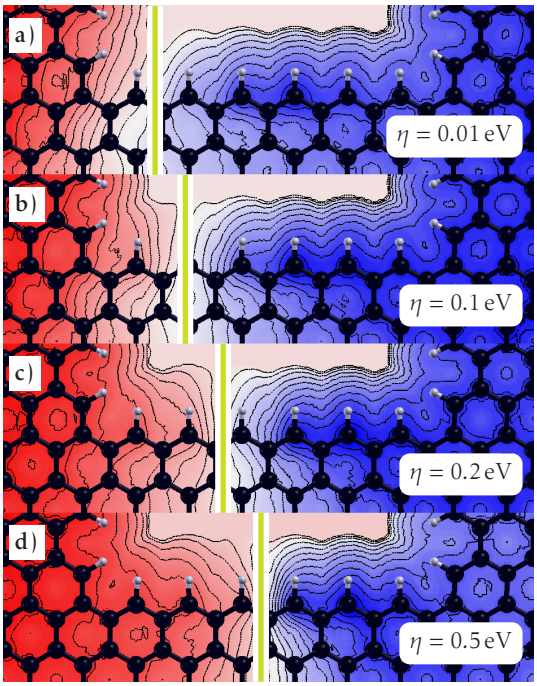


Figure 6.9 | Change of potential drop vs. level broadening parameter, $\eta_{L/R}$, for 0.5 V. Increasing values smear out the electrode DOS which evens out the electronic contribution from both electrodes in the bias window. The voltage drop becomes anti-symmetric at even charge injection rates from the two electrodes (large smearing).

an artificial increase in the broadening parameter, η, for the electrode self-energies in Eq. (3.16). Hence $\mathbf{\Gamma}_L(E) \approx \mathbf{\Gamma}_R(E)$ for $\eta \gg 0$ irrespective of the applied bias and gating. This forces $\delta e \approx \delta h$ and a resulting anti-symmetric voltage drop. Figure 6.9 shows the voltage drop in the middle part of the constriction for four η values. Clearly the anti-symmetric voltage drop is regained when $\eta_{L,R} \geq 0.5\,\text{eV}$. Note that since we have not made assumptions in the model about the nature of the constriction we anticipate that it can straightforwardly be applied to similar systems between graphene electrodes in the high-conductance regime.

Constriction, hydrogen terminated with dopants. Since Gr consists entirely of surface atoms it is also extraordinarily susceptible to external influences such as chemical modification or charged impurities. We will now discuss the influence of modifying the passivation or having adatoms [155–158] as a source of charge doping alternative to the electrostatic gating. We examine the effect of a donating lithium (Li) or an accepting flourine (F) adatom placed either inside or outside the constriction at the positions shown in Figure 6.4a. The Li or F atoms are positioned above the center of a hexagon, or ontop a carbon atom, respectively. In Figure 6.6c we show the transmission for the different adatom configurations. The transmission spectra indicate that very little scattering due to the dopants themselves takes place, especially when the adatoms are positioned outside the constriction. The doping effect is clearly seen from the shift in the two resonance peak positions. Li will n-dope the graphene constriction while F p-dope it. Surprisingly, we find that most of the charge transfer to the device resonances is maintained when the

dopants are moved outside the constriction. This suggests that nanostructured graphene devices will not necessarily be very sensitive to the actual position of the adatoms. In the case of F it is actually more efficient outside the constriction. Comparing the most significant peak with the field effect gating transmission curves we find that Li donates at least 0.2 electrons while F accepts at least 0.3 electrons from graphene. In addition, we find that a pinning of the potential to the positive/negative electrode occurs for Li(n-doping)/F(p-doping) for positive bias, consistent with the potential drops obtained from field effect gating (see Figure 6.4). Adatoms may therefore provide an alternative way to manipulate the voltage drop by pinning the potential to either of the two electrodes. This underlines the conclusion that the main effect is determined by the addition or removal of charge from the device, together with the uneven injection rates from the electrodes.

6.4 Summary

In this Chapter a new method for investigating the electronic structure of doped/gated structures has been presented. The method is based on simple capacitor arguments by moving charge from the system to a fixed user defined space where a gate is placed. This induces an intrinsic electrostatic contribution to the self-consistent calculations. Such a gate is denoted a *charge* gate. An explicit Hartree gate is also implemented and can be used in conjunction with the charge gate.

The charge gate model was used to understand the electronic structure and quantum capacitance effects of N-layer graphite under an external gate-potential. Importantly it is found that the screening length is dependent on the number of layers, gate strength and charging. For high gate levels the external field is fully screened and the top layer charge redistribution is constant for all N-layer graphite stacks. For low gates the external field penetrates the graphite and the charge redistribution is highly dependent on N due to the quantum capacitance.

Additionally the gate method was used in a non-equilibrium system with the NEGF method. We investigated a graphene nano-ribbon connected to two pristine graphene electrodes. In this system the charge gate induced a non-linearity in the induced electrostatic potential by a pinning effect. This is explained via a simple voltage model for graphene. Effectively the electrostatic potential was pinned to the electrode that had the highest coupling to the device region thus leaving the voltage drop at the opposite electrode. Additionally it was shown that adatoms with a doping effect such as flourine or lithium atoms also exhibit the pinning effect.

We expect that using gate models for graphene systems are generally important to describe the electronic structure in nanostructured graphene coupled to pristine graphene electrodes.

Chapter 7

N-electrode calculations

A large variety of experiments are based on multi-electrode setups. In particular for graphene, expectations of using it for nanoelectronics requires interconnects with very low contact resistances. Such low resistance electrodes have been seen in metals [43]. Another possibility is using other 2D materials as interconnects, or perhaps graphene itself. In this Chapter we investigate non-equilibrium calculations of 3 electrode graphene T-junctions based on the article [*159*]. Here we expand on the study in [*159*] by using more geometries. Furthermore we present NEGF calculations of the graphene T-junction. For further information on the introductory material we refer to the paper [*159*].

7.1 T-junction

Since the discovery of graphene [21, 22] it has been thought as a future replacement for several nanoelectronic components. However, a key problem that needs to be solved is how to efficiently connect electrodes and create circuitry of graphene, and 2D materials in general. Recently 1D contacts to graphene have been accomplished with 3D metal electrodes with surprisingly low contact resistance [43]. Metal contacts are needed when coupling the device together with other electronics via wires, however, interconnects in circuitry may be viably created by the same chemical species. In [*159*] we study an interconnect using a so-called T-junction. A T-junction is formed by the attachment of a semi-infinite sheet of graphene vertically onto a graphene sheet. The intrinsic sp^2 bonding of graphene will at the junction be altered to a sp^3 hybrid thus creating a tetrahedral-like bond on the connecting atom. The binding energies of these T-junctions with only carbon bonds are high with energies around 0.5 eV while such structures may be achieved by fusing or synthesis [*159*], additional introductory prospects and energetics can be found in the article.

Here we expand on the *Jacobsen et al.* [*159*] study and investigate the conductance properties of such T-junctions with different geometries. The calculations are performed using the GGA-PBE functional [63], a real-space grid cut-off of 254 Ry, filter cutoff 150 Ry [160], 17 transverse **k**-points per Ångstrom. Single-$\zeta^{\pm}$ (polarisation) is used as double-$\zeta^{\pm}$ showed no difference in geometries and transmissions. Bias calculations are performed with a non-equilibrium energy integral of $\Delta\epsilon = 0.01$ eV and $\eta = 0.002$ eV. The equilibrium contours start at -40 eV, connected to a line lifted 2.5 eV above the real energy axis with 16 points from -10 kT to ∞ by a Fermi quadrature. The electronic temperature is 25 meV. Transmission plots are created with 1700 transverse **k**-points per

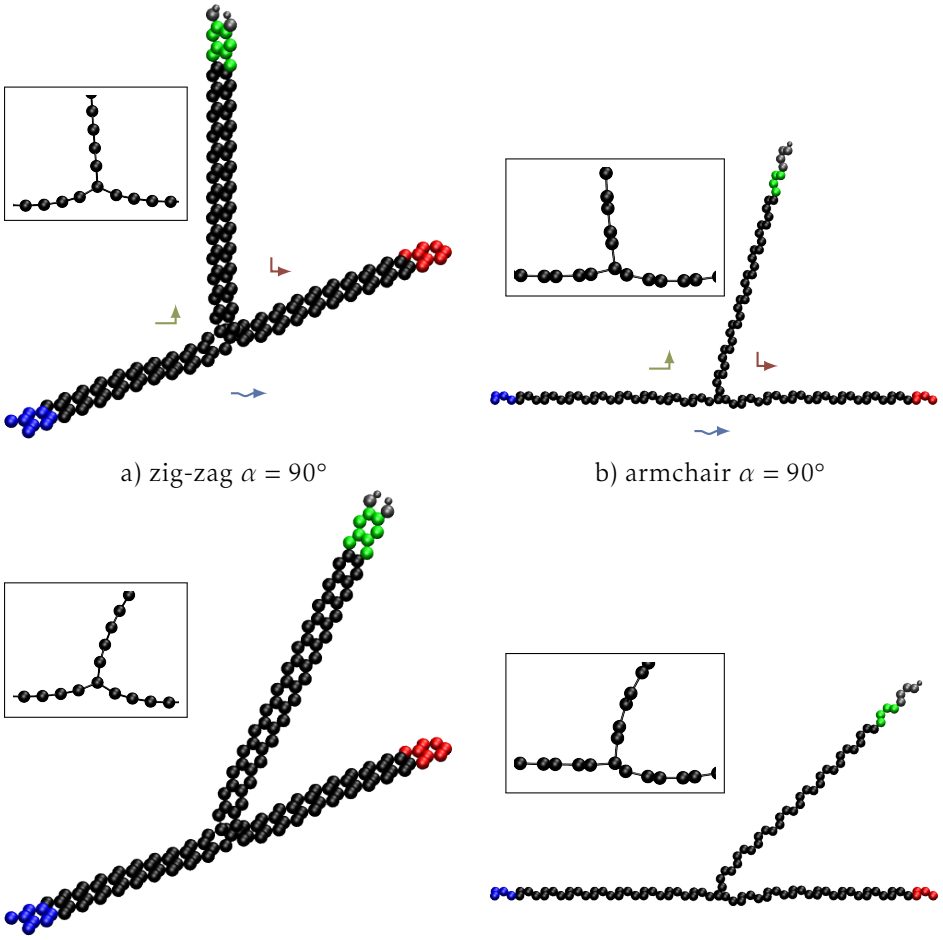


Figure 7.1 | Relaxed T-junction geometries. The transport directions are shown with ⇝(left to right), ⇂(left to top) and ↳(top to right). Small insets are a close-up side-view of the junction. a) and c) are zig-zag junctions with 90° and 55° degree angles with respect to the planar graphene. The 90° is symmetric for ⇂ and ↳. b) and d) are armchair junctions with 90° and 55° degree angles with respect to the planar graphene. The armchair junction is asymmetric. A clear transition from graphene sp^2 to tetrahedral-like sp^3 hybrid is seen for all junctions. The blue and red electrodes constitute the planar graphene sheet with an attached semi-infinite graphene terminated by an electrode (green).

Ångstrom with a subsequent interpolation [161]. We limit the study to encompass an armchair and a zig-zag connection. To compare the two directions we width-normalise the transmission and current. In Figure 7.1 the different relaxed geometries are shown. Both systems are investigated with $\alpha = 90°$ and $\alpha = 55°$. The angle α is the alignment of the third electrode with respect to the base graphene sheet, i.e. 90° corresponds to a perpendicular electrode. Remark that for both angles the relaxed geometry exhibits a vertical bonding to the graphene plane in accordance with a tetrahedral-like bonding. Hence an equilibrium bonding environment will preferentially uphold the tetrahedral-like bonding irrespective of the electrodes angle far from the junction. The intrinsic ripples in graphene will have negligible effect on the electronic properties [162], consequently it is expected that the long wave length ripples arising due to the junction

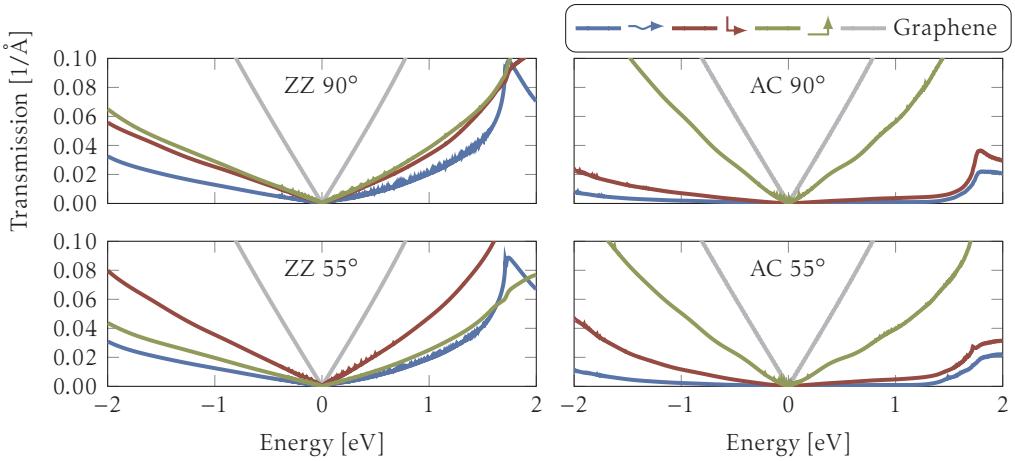
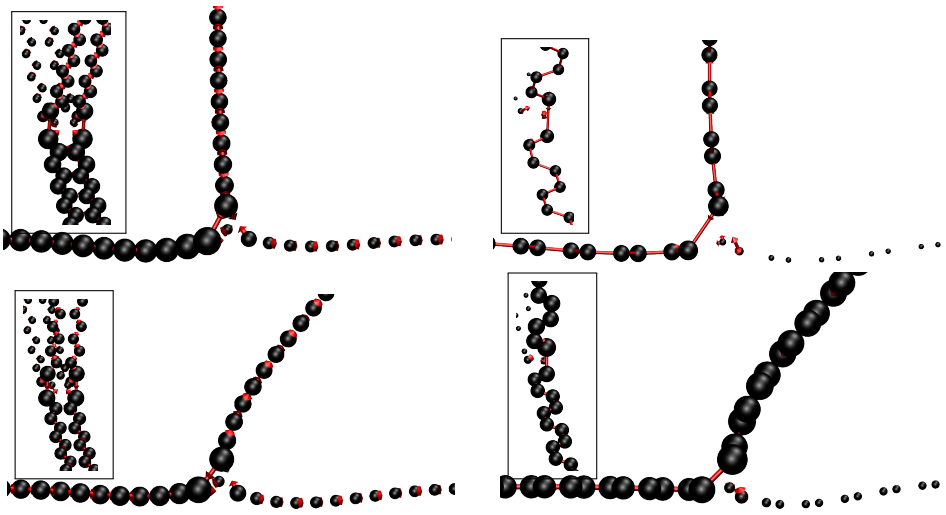


Figure 7.2 | Transmission (width normalised) for the different structures shown in Figure 7.1 at 0 V. Armchair (AC) and zig-zag (ZZ) for two different T-angles $\alpha = \{90°, 55°\}$. For comparison the pristine graphene curve is also shown. The lowest transmission for all geometries is the ⇝. ZZ 90° has by symmetry arguments equivalent ⇂ and ↳ transmissions. ZZ 55° has a slightly broken symmetry with ↳ having a larger transmission compared to ⇂. The ⇂ AC structure has a transmission remarkably close to pristine graphene, while ↳ and ⇝ have much lower transmissions.

also has negligible effect. In the following we denote each transmission path via three symbols; left to right(right to left) as ⇝, left to top(top to left) using ⇂ and top to right(right to top) via ↳. Note that transmission symmetries apply, see App. A.

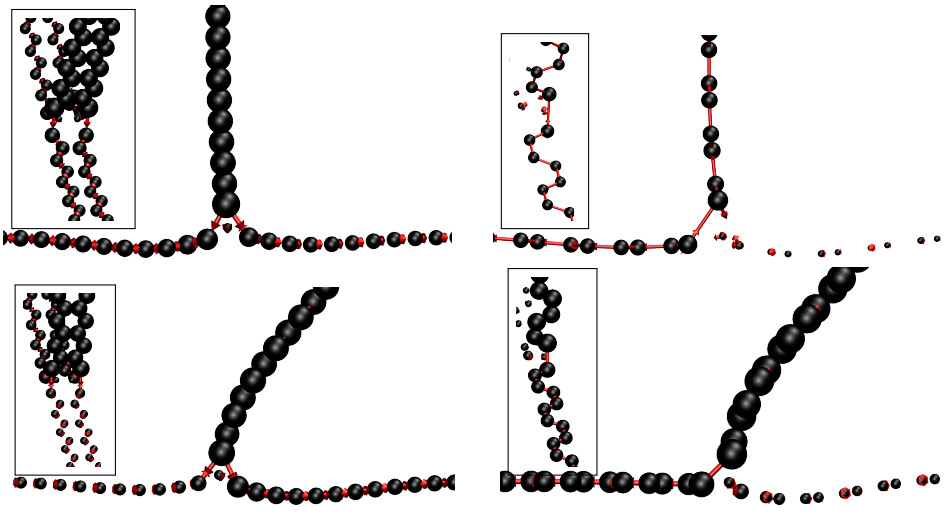
Figure 7.2 shows the 0 V transmission for the 4 structures (sorted in the same 2×2 diagram). For the ZZ 90° structure ⇂ and ↳ have almost equivalent transmissions which can be deduced by symmetry arguments from the junction. Interestingly ⇝ has a lower transmission compared to the other. The junction may be viewed in terms of *kinks* [163]. In path ⇂ a single kink disrupts the electron pathway, while ⇝ has a kink up, then kink down, i.e. two kinks. See inset Figure 7.1a. Effectively the left-right electrodes are decoupled which is counter intuitive given that the junction is formed from a flat graphene flake. In the ⇝ plane the tetrahedral-like bonding atom has 4 bonds completing the shell which effectively makes it inert. Hence the π cloud from the left electrode will always couple more to the top electrode as the overlap with the π orbitals in the top electrode are extensively larger than the direct coupling to the π orbitals on the right electrode. This also holds for ZZ 55° where ↳ transmission increases (larger π orbital overlap due to a lower angle towards the right electrode) and ⇂ decreases. ⇝ is largely unaffected which is expected as their coupling will only slightly depend on the vertical environment.

For the armchair junction we find ⇂ to almost retain a pristine graphene transmission while both ⇝ and ↳ have very low transmissions. Importantly, ⇂ and ↳ paths does not have a symmetric junction. Consequently if we shift the top electrode one atom towards either the left or right electrode the transmissions ⇂ and ↳ will swap. If one follows the alternating A/B site notation commonly used for graphene, the sequence for the ⇂ path is A-B-⊥-B-A while for ↳ it is A-B-⊥-A-B where ⊥ corresponds to a vertical bond. Hence the former bond sequence preserves the best conductance properties.



zig-zag left scattering states armchair left scattering states

Figure 7.3 | Bond-currents (non-orthogonal) originating from the left electrode ($E = -0.2\,\mathrm{eV}$). Ordered according to Figure 7.1. The large figures are sideviews of the system, while the insets are looking from above the left electrode, focusing on the junction. Red arrows are scaled according to the bond current. Atoms are scaled according to the activity current, Eq. (5.22). The tetrahedral-like bonded atom at the centre of the junction does not carry any bond-current in an energy range of −1 eV to 1 eV. Clearly the bond-currents are migrating towards the top electrode and the atoms close to the junction have an excess activity current.



zig-zag top scattering states armchair top scattering states

Figure 7.4 | Same as Figure 7.3 however from the top electrode scattering states. Similarly to the left scattering states there is no current passing the tetrahedral-like bonded atom. The top scattering states for the ZZ primarily propagate towards the closest electrode while for $\alpha = 90°$ it becomes evenly divided. For AC the division is dependent on the tetrahedral-like bonded atom which determines the high transmission path.

In Figs. 7.3 and 7.4 we plot the bond currents[1] as explained in Sec. 5.4 calculated using the non-orthogonal basis set. We note that bond-currents still maintain a physical interpretation. Atom sizes are scaled according to the activity current, Eq. (5.22), and normalised within each frame. Hence the largest atoms have a large excess activity current. The large figures are side views of the geometries, while the small insets are top skewed views from above the left electrode viewing down on the T-junction. For all junctions the bonding atom has no activity current, this has been checked in a range across −1 eV to 1 eV. For the ZZ geometries the lowest angle between two electrodes seem decisive for the highest transmission. Moreover, the left scattering states have an excess current compared to the top scattering states, corresponding to a larger circulating current in the junction around the bonding atom. This is apparent as all the top electrode atoms for the top scattering states have, roughly, the same size while the left electrode atoms for the left scattering states have increasing size with the largest being at the junction interface. For the AC geometries, the selectivity of the ⊥ transmission is also confirmed where ⊥ is dominant for $\alpha = 90°$ and ↳ increases for lower angles, albeit the increase is very low and thus the angle dependence on AC structures seem less pronounced. Clearly the asymmetric position of the top electrode has a high influence on the transport properties. In realistic devices this selectivity may be difficult to measure as it would require a pristine attachment of the T-junction with no bond deformations giving rise to discontinuities at the boundary. The bond currents aide in determining important geometrical effects for a graphene T-junction and shows that in fact the top electrode has the highest total transmission $\mathcal{T}_T$. Importantly, they highlight which atoms participate in carrying the transmission, and which does not. For the T-junction it is shown that the bonding atom is not participating in any current carrying transport.

7.1.1 Non-equilibrium

As shown in Sec. 6.3 non-equilibrium calculations of graphene junctions easily become elaborate as external gating may greatly influence the conductance and characteristics of the system [*140*]. Comparatively for the T-junction this complexity increases as we further add a third electrode. One could envision a device where the left-right and top electrodes were gated individually, additionally the three electrodes could have individual chemical potentials yielding 5 variables to tune for full device characteristics. 3 voltages and 2 gates.

Here we investigate the I-V characteristics of the T-junction. Current integration is performed using a fine energy grid of 0.001 eV with a spline interpolated Hamiltonian from full non-equilibrium calculations at −1.2 V to 1.2 V in steps of 0.3 V (9 calculations). The interpolation is in steps of 0.05 V, see Sec. 5.3 for details.

The applied bias is considered in two cases, 1) with $V/2 = \mu_L = \mu_R = -\mu_T$ and 2) with $V/2 = \mu_L = -\mu_T$ and $\mu_R = 0$. To work as a transistor-like system one would require case 2) with variation of the relative potential profiles. However, such investigations quickly become truly elaborate. The following arguments considering for the I-V characteristics can be translated to the transmissions shown in Figure 7.2.

Case 1 is seen in Figure 7.5 for all 4 systems. Note that there is no I-V curve for ⇝ as the Fermi levels are aligned. For the orthogonal ZZ system ⊥ and ↳ are symmetric and two equivalent I-V curves are found, as expected from the 0 V transmission. For ZZ,

[1] Note that the VMD view-ports are different between different images [164], whence the relative difference in atomic sizes can only be compared on single images.

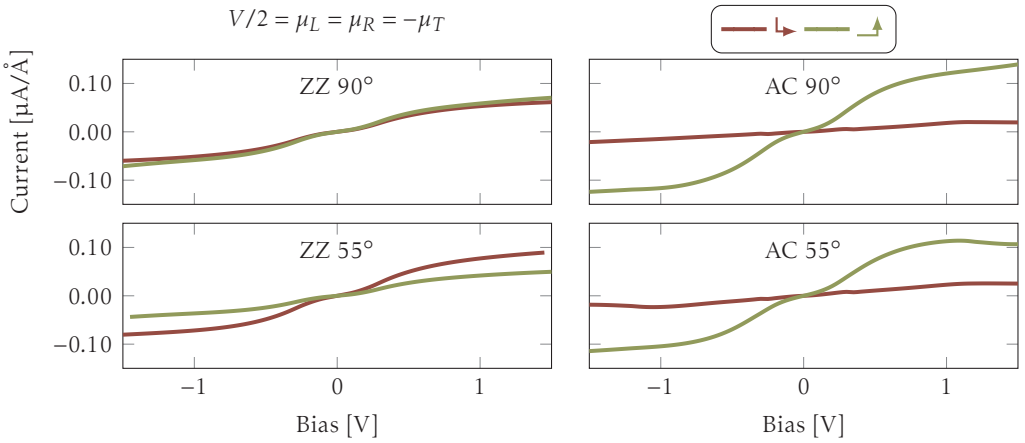


Figure 7.5 | I-V characteristics for ZZ and AC structures with different angles. The chemical potentials are in terms of the applied bias: $V/2 = \mu_L = \mu_R = -\mu_T$. For the ZZ structures the selectivity of the rotated top-electrode is seen as ↳ is roughly doubled with respect to ⮤ (for $\alpha = 55°$). For the AC structures the low angle dependency is seen as the current is roughly the same for $\alpha = 90°$ and $\alpha = 55°$. The relative difference is around 5.

$\alpha = 55°$, the increased transmission of ↳ is consistent with the I-V curves where ↳ is roughly twice as large as ⮤. For the AC structures the decoupling of the right electrode is seen as ⮤ is roughly 5 times larger than ↳ for both angles. Again, if the entire top electrode was shifted one atom, left or right, the I-V curves would be reversed. Here there is little difference between the angles as seen for the 0 V transmission.

In the second case the I-V characteristics is seen in Figure 7.6. Note that here all chemical potentials are different. The current scale is the same as that in Figure 7.5. Note that ⇝ and ↳ have half the bias width compared to ⮤. The high potential electrodes (left, top) will couple more to the device region as their DOS in the bias window is comparatively larger than the right electrode. See discussion in Sec. 6.3. Effectively the right electrode couples weakly and so the resulting current will be low. This effect is apparent across all structures, irrespective of the angles involved. The resulting current is carried in the ⮤ junction and it is roughly equivalent to the maximum current in Figure 7.5. This highlights the decoupling of the three electrodes. I.e. the angle has little influence on I-V characteristics as it is mainly dependent on the coupling of the electrodes to the device.

Comparing case 1 and 2 demonstrates a possible mechanism to control current carrying pathways by coupling/decoupling via a third electrode, thus "gating" the device using the third electrode[2]. Further studies are required to assert this for a wider range of potential differences. In particular calculations for $V/2 = \mu_L = -\mu_R$ and $\mu_T = 0$ might prove the coupling statements as the ⇝ is the smallest current for the investigated structures. Note that the conclusions made for zero bias are difficult to translate into a generic bias perspective as the relative chemical potentials and couplings are important for the I-V characteristics.

[2]We acknowledge the limitation of such "gating" as the off-state current will be high.

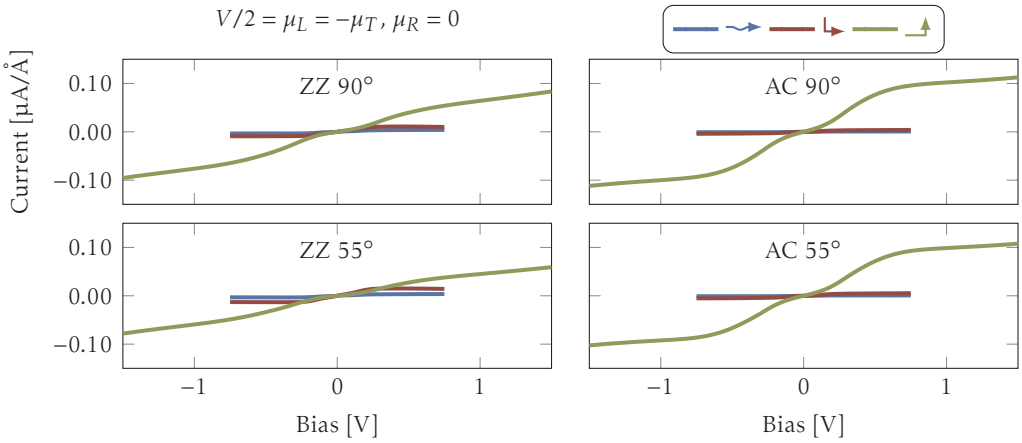


Figure 7.6 | I-V characteristics for ZZ and AC structures with different angles. The chemical potentials are in terms of the applied bias: $V/2 = \mu_L = -\mu_T$ and $\mu_R = 0$. For all structures ⤴ is the major current carrying pathway. In this bias configuration it also exhibits the largest difference between the electrode chemical potentials. There is nearly no angle dependence and in general the currents are similar to the maximum currents seen in Figure 7.5.

7.2 Summary

We have used the $N_\mathfrak{e}$ method in TranSIESTA and conducted transmission analysis on a T-junction system consisting of pristine graphene electrodes and a T junction. Using bond-currents we found that the bonding atom (tetrahedral-like bonded atom) had almost zero activity current in a range of 2 eV around the Fermi level. The current is thus redirected around the respective atom. The I-V characteristics shows that the electrode coupling is important, as shown previously in Figure 6.3. This makes the maximum current equivalent for the investigated systems. However, the current carrying pathway, ⤳, ⤴ or ↳ may be bias controlled by coupling/decoupling the last electrode.

Consequently, the T-junction may be a future prototype-device used to yield gate controlled junctions based on pure graphene electrodes. Such devices are interesting due to their low resistances and thus the need for contacting to a bulk material becomes superfluous if current-pathways can be controlled in a graphene *grid*.

T-junctions can have a great variation of the chemistry in the junction. As shown in [*159*] the bonding atom may be a substitutional boron/nitrogen/oxygen atom. Such bonding atoms are probably more likely to occur as the graphene edges may be exposed via chemical etchants composed of aforementioned atomic species. Such considerations are important for a full understanding of T-junctions. These investigations have been outside the scope of this thesis.

Chapter 8

Inelastic conductance in STM-graphene

Scanning Tunnelling Microscopy/Spectroscopy (STM/STS) is a commonly used method to investigate electronic structure of surfaces and 2D materials. This is partly due to, its high-resolution imaging but also its local spectroscopic capability. In particular STM/STS on graphene have shown interesting results with giant phonon induced conductances [165]. Furthermore are doping effects due to boron and nitrogen substitutions of increasing interest [166–*168*].

First an introduction to the concepts of inelastic conductance will be provided. The inelastic conductance is due to the electron-phonon interaction which creates steps at energies corresponding to phonon energies. The phonons can be estimated using the Hessian matrix as explained in Sec. 2.1.1. Then, we will turn to inelastic calculations in a 3-terminal system which proves extremely powerful in reproducing graphene's inelastic conductance in STS [165, 169, 170]. The results presented are based on the paper published in [*171*] and a subsequent collaborative work with *Lagoute et al.* [*168*].

8.1 Inelastic conductance

The following description is from the excellent implementation article[1] Frederiksen *et al.* [172]. For further explanations we refer to [139, 173].

The inelastic signal originates due to the interaction of the electronic structure with the phonons. The inelastic effect occurs at electron energies above a corresponding phonon energy. An excitation of a phonon can only occur when the bias is equal to or above the phonon frequency, $eV \geq \hbar\omega$. The electron-phonon coupling can be described using the extended Hamiltonian equation

$$\mathbf{H} = \mathbf{H}_e + \mathbf{H}_{\text{ph}} + \mathbf{H}_{e-\text{ph}} \tag{8.1}$$

$$\mathbf{H}_{\text{ph}} = \sum_{\lambda} \hbar\omega_\lambda \hat{b}^\dagger_\lambda \hat{b}_\lambda \tag{8.2}$$

$$\mathbf{H}_{e-\text{ph}} = \sum_{\lambda} \sum_{i,j} \mathbf{M}^\lambda_{i,j} \hat{c}^\dagger_\lambda \hat{c}_\lambda (\hat{b}^\dagger_\lambda + \hat{b}_\lambda), \tag{8.3}$$

[1] Note that [172] uses the term LOE. We adopt the terms LOE-WBA for [172] and LOE for [139].

with $\mathbf{H}_e$ being the electronic Hamiltonian under the Born-Oppenheimer approximation, Eq. (2.5). $\{\hat{b}, \hat{c}\}$ are the phonon and electron annihilation operators († creation). λ is the associated phonon index and $\{i, j\}$ are the electron indices. $\mathbf{M}^\lambda$ describes the coupling element between the electronic structure and the phonon mode λ. The electron phonon coupling is a function of the electronic Hamiltonian derivative with respect to atomic coordinates (I)

$$\mathbf{M}^\lambda = \sum_{I\nu} \left\langle i \middle| \frac{\partial \mathbf{H}_e}{\partial \mathbf{R}_{I\nu}} \middle| j \right\rangle \mathbf{v}^\lambda_{I\nu} \sqrt{\frac{\hbar}{2M_I\omega_\lambda}}, \tag{8.4}$$

where $\mathbf{v}^\lambda_{I\nu}$ is the normalised eigenvector for the phonon mode λ and frequency ω_λ. The following discussion will be based on the Lowest Order Expansion Wide Band Approximation (LOE-WBA) and the LOE expansion. For a full description of the self consistent Born approximation we refer to [172].

The phonon modes and electronic Hamiltonian derivatives are calculated via the frozen phonon method as described in Sec. 2.1.1. The LOE-WBA approximation describes the inelastic current (without elastic contribution) for a single phonon λ as

$$I^{\text{LOE–WBA}}_\lambda = \gamma_\lambda \mathcal{I}^{\text{sym}}_\lambda + \kappa_\lambda \mathcal{I}^{\text{asym}}_\lambda \tag{8.5}$$

$$\gamma_\lambda = \text{Tr}\left[\mathbf{G}^\dagger \mathbf{\Gamma}_L \mathbf{G}\left\{\mathbf{M}^\lambda \boldsymbol{\mathcal{A}}_R \mathbf{M}^\lambda + \tfrac{\mathrm{i}}{2}\left(\mathbf{\Gamma}_R \mathbf{G}^\dagger \mathbf{M}^\lambda \boldsymbol{\mathcal{A}} \mathbf{M}^\lambda - \text{H.c.}\right)\right\}\right] \tag{8.6}$$

$$\kappa_\lambda = \text{Tr}\left[\mathbf{G}^\dagger \mathbf{\Gamma}_L \mathbf{G}\left\{\mathbf{\Gamma}_R \mathbf{G}^\dagger \mathbf{M}^\lambda \left(\boldsymbol{\mathcal{A}}_R - \boldsymbol{\mathcal{A}}_L\right)\mathbf{M}^\lambda + \text{H.c.}\right\}\right], \tag{8.7}$$

where $\boldsymbol{\mathcal{A}} = \sum_\mathfrak{e} \boldsymbol{\mathcal{A}}_\mathfrak{e}$. The symmetric and anti-symmetric functions $\mathcal{I}$, are respectively

$$\mathcal{I}^{\text{sym}}_\lambda = \frac{e}{\pi\hbar}\left(2eV\langle n_\lambda\rangle + \frac{\hbar\omega_\lambda - eV}{e^{(\hbar\omega_\lambda - eV)/kT} - 1} - \frac{\hbar\omega_\lambda + eV}{e^{(\hbar\omega_\lambda + eV)/kT} - 1}\right), \tag{8.8}$$

$$\mathcal{I}^{\text{asym}}_\lambda = \frac{e}{2\pi\hbar}\int_{-\infty}^{\infty} \mathrm{d}\epsilon \Big[n_F(\epsilon) - n_F(\epsilon - eV)\Big]\mathcal{H}_{\epsilon'}\Big\{n_F(\epsilon' + \hbar\omega_\lambda) - n_F(\epsilon' - \hbar\omega_\lambda)\Big\}\{\epsilon\}, \tag{8.9}$$

where $\mathcal{H}$ is the Hilbert transform. We stress that the above equations governing the LOE-WBA requires the Green function, spectral function and scattering matrices at the Fermi level. In effect, the wide band approximation is that the electronic structure is slowly varying in the limited energy range of the phonon spectrum. Hence, for systems with significantly varying density of states around the Fermi level an inclusion of the changing electronic structure needs to be taken into account.

The LOE approximation improves the WBA by taking into account the change in electronic structure at the bias threshold of the phonon frequency [139]. Further, the LOE estimates the conductance above threshold as a constant contribution (linear in current). The main differences can be outlined schematically in Figure 8.1. The two electrodes, Left and Right, are shown with the associated applied bias eV. The DOS spectrum is marked in each of the two electrode regions by the thick black line. LOE-WBA and LOE are highlighted in different colours and the phonon excitation is shown as a wiggly line, $\hbar\omega$. An inelastic conductance contribution for phonon mode λ is constant for $\hbar\omega_\lambda \geq eV$ for both LOE-WBA and LOE. For LOE-WBA this contribution is for the spectral densities evaluated at $\epsilon = \epsilon_F$ (single line crossing both DOS's at same energy), while for LOE it is calculated using the spectral densities at $\epsilon = \pm eV/2$ (different lines in each electrode). These two methods yields different results as the the DOS may change drastically, even for low phonon energies. However, the wide band approximation has

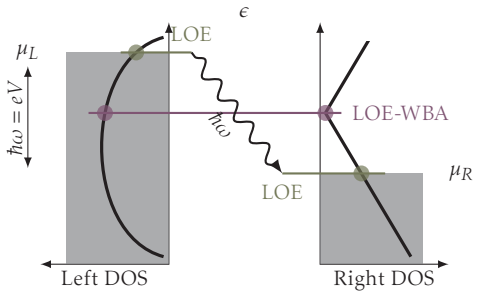


Figure 8.1 | Comparison of LOE-WBA and LOE contributions to the density of states from the two leads. The two methods have their own colour coding. The applied bias matches the phonon energy $\hbar\omega = eV$ which creates a constant contribution above $eV > \hbar\omega$ to the inelastic current in both approximations. The LOE-WBA only takes the Fermi level electronic structure into consideration. LOE takes into account the *at threshold* DOS from both the left (initial) state to the right (final) state.

proven sufficiently accurate in many cases. The LOE equations can be written without the elastic contribution as

$$\begin{aligned} I_\lambda^{\mathrm{LOE}} \approx \frac{e}{\pi\hbar}\int_{-\infty}^{\infty} \mathrm{d}\epsilon\, \Re\mathfrak{e}\,\mathrm{Tr}\Big[\mathbf{G}(\epsilon)\Sigma^\lambda(\epsilon)\mathcal{A}_L(\epsilon)\mathbf{\Gamma}_R(\epsilon)\Big]\Big(n_{F,L}(\epsilon)-n_{F,R}(\epsilon)\Big) + \\ \frac{e}{h}\sum_{\sigma=\pm}\Big(\coth\frac{\hbar\omega_\lambda}{2kT}-\coth\frac{\hbar\omega_\lambda+\sigma eV}{2kT}\Big)\times \\ \int_{-\infty}^{\infty}\mathrm{d}\epsilon\,\mathrm{Tr}\Big[\mathbf{M}^\lambda\widetilde{\mathcal{A}}_L(\epsilon)\mathbf{M}^\lambda\mathcal{A}_R(\epsilon_\sigma)\Big]\Big(n_{F,L}(\epsilon)-n_{F,R}(\epsilon)\Big) \end{aligned} \tag{8.10}$$

where the threshold dependent energies $\epsilon_\sigma \approx \mu_R = \mu_L + \sigma\hbar\omega_\lambda$ are used instead of the constant Fermi level as in LOE-WBA. The phonon self energy, Σ^λ, can be found elsewhere [139].

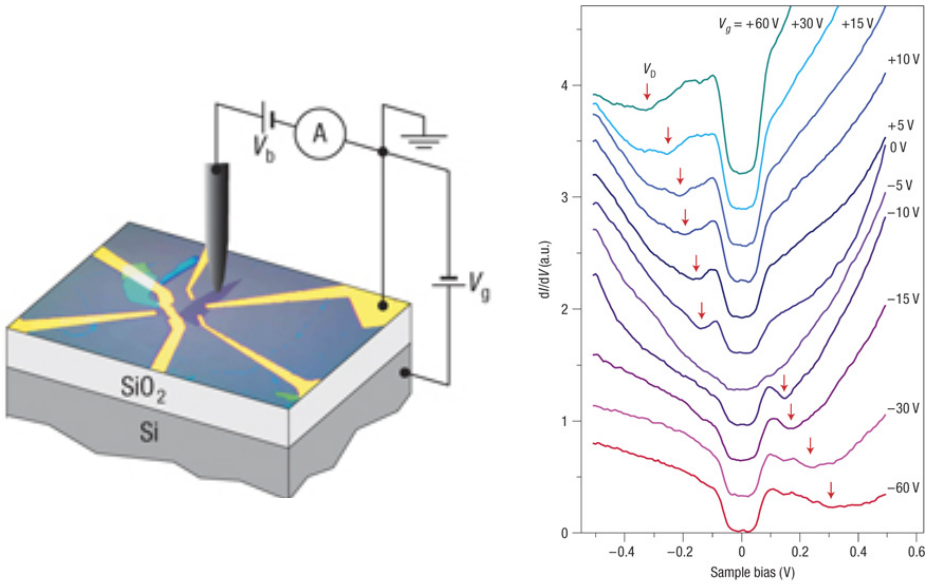
The inelastic currents are seen as discontinuities in the current which can be understood via a constant DOS model. Here the elastic current $I = \int G$ is linear plus jumps at $eV = \hbar\omega_\lambda$. Hence, $\mathrm{d}I/\mathrm{d}V$ yields a constant elastic conductance and a step for the inelastic conductance. A second differential $\mathrm{d}^2I/\mathrm{d}V^2$ only retains the inelastic contribution as a peak. The second derivative of the current is commonly known as the inelastic tunnelling spectroscopy signal (IETS) which may be divided by the conductance, $\mathrm{d}I/\mathrm{d}V$, to decouple strong vs. weak signals via rescaling. In effect, IETS is the *strength* of the signal. Both the non-scaled and the scaled are seen in literature [172].

This concludes the theoretical background for the calculations and we turn to an application of the above equations on graphene.

8.2 Inelastic tunnelling into graphene

Here we present the results from the published articles [*168*, *171*].

Experimental STS measurements on graphene have shown interesting phonon induced features [165, 174]. These features are found for both SiC and SiO_2 substrates. A "giant" phonon induced conductance with a gate dependency was found as seen in Figure 8.2 from [165]. a) shows the sample setup while b) shows the STS experiment for a range of gate voltages. Red arrows indicate the Dirac point in the graphene sample which



a) b)

Figure 8.2 | a) Experimental setup of STS experiment with graphene on SiO_2 substrate. b) STS spectra of gated graphene and inelastic conductance. Each arrow indicates the Dirac point of graphene and hence an exact doping profile can be mapped out by fitting the V_D dips vs. applied gate bias. Both from [165].

follows the gating via the well-known $\sqrt{}$ dependency of the doping. Several explanations exists which all are based on single phonon models [165, 169, 175]. The explanation can be thought of as an example of Fermi's Golden rule, which is an energy conservation rule of the rate; $R \propto \delta(\epsilon - \hbar\omega_\lambda)$. An electron from the tip has momentum $\mathbf{k}_\perp$ corresponding to an energy of eV. However, the only states available at energies close to the Fermi level lie at the K (Dirac) point in graphene thus having a high $\mathbf{k}_\parallel$ momentum. Energy conservation forces the incoming electron to have at least an energy corresponding to this momentum transfer. An excitation of a phonon at $q = \pm K$ with energy eV preserves energy with $q \pm K = 0$ and a momentum transfer is allowed. At $q \pm K$ the phonon spectrum has acoustic out-of-plane modes [176] at an energy around $\hbar\omega_\lambda = 65\,\text{meV}$. The half-width gap seen in Figure 8.2b corresponds well to this energy. Further the dip features, shown by arrows, indicate the graphene Dirac point which is probed via the STS spectra for various gating and bias levels.

We have studied the system using full first-principles calculations, thus calculating the full phonon spectrum, the electron-phonon couplings and the resulting LOE inelastic current as explained in the previous section [*168*, *171*]. Our model[2] setup can be seen in Figure 8.3a with the red atoms being the electrodes, left/right/tip with the tip being a gold chain with attached 4 atom gold tip. The black atoms indicates the region where the phonon modes are calculated, the green atoms are an extended electrode to allow the electronic structure to "relax" from the scattering region and is also used together with the black atoms to calculate the electron-phonon coupling matrix elements, see [172] for details.

[2]We use a split DZP basis set, a mesh cutoff of 200 Ry, a Monkhorst-Pack **k**-point mesh of $1 \times 2 \times 1$ and the LDA xc-functional [62] to calculate the electronic structure.

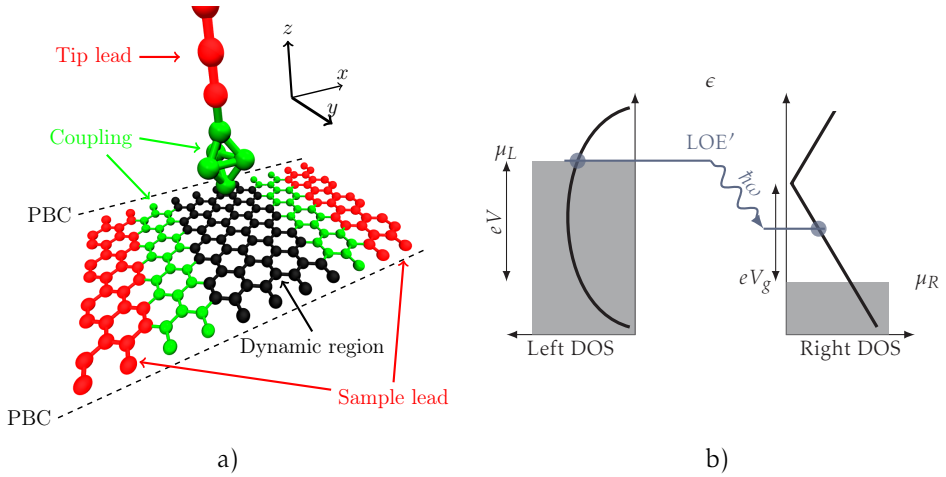


Figure 8.3 | a) Method setup for calculating the first-principle study of inelastic tunnelling into graphene. Phonon region (black atoms), electron-phonon coupling region (black+green), electrodes (red). From [171]. b) At and above threshold, $eV = \hbar\omega$, LOE adds a constant inelastic contribution for said phonon. While the LOE′ constantly follows the chemical potential of the left and estimates the contribution at threshold for all phonons thus probing the Right DOS.

Further characterisation of this setup is needed. In an STM setup the device is weakly coupled to the sample, i.e. the tunnelling regime. In this regime the graphene will retain a constant potential which is equal to the gating level. However, LOE expresses the spectral functions at $\pm eV/2$ as seen in Figure 8.1. To circumvent this restriction we adopt an altered scheme as shown in Figure 8.3b. Here the chemical potential in graphene is constant (Right DOS) and equal to $\epsilon_{F,G} = \epsilon_D + eV_g$ with ϵ_D being the Dirac point energy and eV_g the gating level. Note that the gating level is a shift for the entire electronic structure. The tip chemical potential follows the applied bias and equals $\epsilon_{F,T} = \epsilon_F + eV$ (Left DOS). Resulting in *threshold* contribution following the graphene DOS as indicated by LOE′. We remark that this approximation relies on the tip DOS to be roughly constant in the energy range investigated as the different bias contributions are from different relative energies. We have checked that the DOS on the tip atom is approximately constant in a range of 1 eV around the Fermi level.

In Figs. 8.4 we compare experimental measurements with calculated STS spectrum. The most prominent features are extremely well reproduced; the gap around ϵ_F and the dip features at the Dirac point in graphene. In b) top-right the phonon mode that dominates the inelastic contribution is shown with the *A-B* anti-symmetric out-of-plane displacements indicated by the arrows. For further understanding the different mode contributions the b) bottom-right plot is an equivalent calculation by taking into account the full phonon spectrum (full line) and only phonons with energies 60 meV to 70 meV (dashed line). A blue-shift and a magnitude increase is seen by including more phonons. These contributions originates from both lower phonon modes around 40 meV and higher optical (in-plane) phonon modes around 150 meV. This blue-shift can at $\epsilon_D = 100$ meV lead to an error of 32% when estimating the gating level from the dip feature. These findings may explain the difference seen in [177].

Being able to fully reproduce the inelastic signal for STS on graphene we proceed to investigate IETS signals on defected graphene. In this case defects are expected to

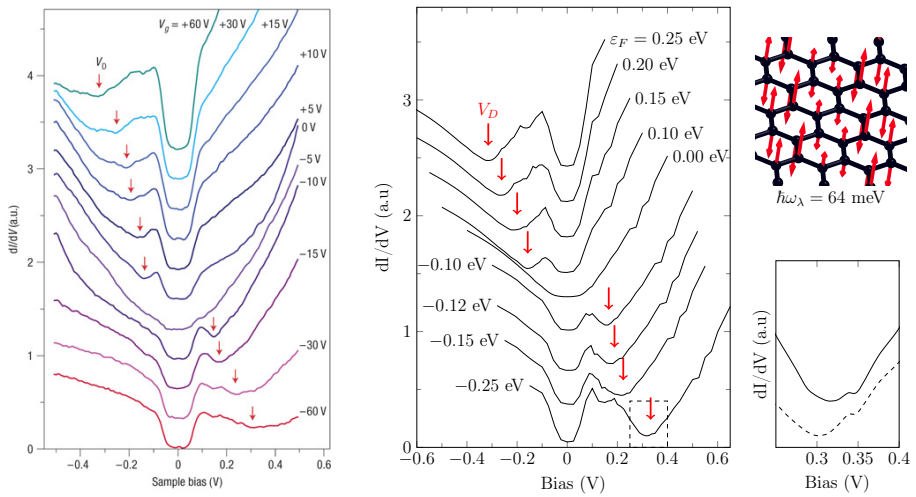


Figure 8.4 | a) Experimental findings from [165]. b) First principle calculations of the LOE' calculations which captures all notable features, the tunnelling gap around $V = 0$ and the gate dependent Dirac dip. The mode with the largest contribution can be seen in the top-right figure which is a phonon at K with anti-symmetric out-of-plane displacements. The bottom-right figure shows the boxed data (full line) together with only phonons in the range 60 meV to 70 meV (dashed line) to indicate that this is not only governed by single phonon excitations. From [*171*].

lift the $q = \pm K$ dependency and a corresponding high elastic signal. Hence the inelastic signals in the conductance are quenched and IETS is required to capture these effects. In Figure 8.5 we show the IETS signal for pristine graphene, a buckled Stone-Wales defect (SW) and a hydrogen terminated edge. For pristine graphene the same spectra is seen for hollow and bridge sites of the tip. By comparing all other spectra against the pristine one can find signature features due to the changing environment and, hence, phonon spectrum. The SW is buckled as a lower energy minimum is found for out-of-plane buckling [178]. The marked feature (b) of the bond-mode at the SW centre is higher than the corresponding graphene modes and we expect it to be visible in an experimental spectrum. The same calculation for the *flat* SW has a blue-shift of ~ 5 meV which also confirms the relaxation of the bond-mode with a buckled structure. A signature for the hydrogen terminated edge has a slight change as the only contributing modes are in-plane. The in-plane signature mode is fully dependent on the mass of the passivating specie via mass-scaled electron-phonon couplings. The out-of-plane hydrogen modes are hardly adding any inelastic signal. R_I percentages denote the fractional contribution from inelastic signals

$$R_I = \left.\frac{\mathrm{d}I}{\mathrm{d}V}\right|_{V=\max(\hbar\omega_\lambda)} \Big/ \left.\frac{\mathrm{d}I}{\mathrm{d}V}\right|_{V=0}. \tag{8.11}$$

Pristine graphene has $R_I = 95\%$ due to the $V = 0$ gap and all others have a significant elastic contribution lowering R_I.

In Figure 8.6 an equivalent analysis has been performed on graphene adsorbents. We investigate multiple different configurations and species. The hydrogen adsorbent has been investigated on-top, below, near-by and a graphane dimer which all but below have a significant high IETS peak corresponding to a hydrogen mode. All defects

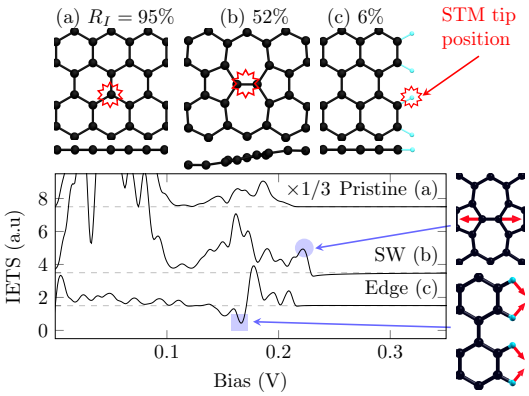


Figure 8.5 | IETS signals for 3 graphene structures. (a) Pristine graphene scaled by 1/3. (b) Energy minimum buckled Stone-Wales defect with finger-print signal at 225 meV higher than pristine graphene IETS signals. The corresponding mode is shown in the right inset. (c) Hydrogen terminated graphene edge, the hydrogen signal lies in the graphene phonon energy range but a significant hydrogen mode at 170 meV is seen. From [*171*].

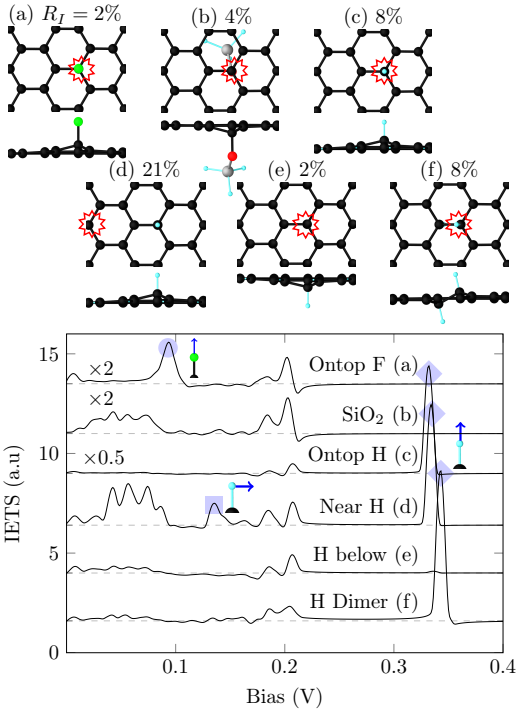


Figure 8.6 | IETS signals for 6 graphene adsorption defects as noted in the figure. Only "near H" and SiO_2 have a weak graphene background signal while all others show signatures related to the adsorbent. Remark that it is possible to decipher the above/below case of hydrogen. From [*171*].

are out-of-plane phonons while the near-by IETS is due to an in-plane phonon. The below case does not have a hydrogen phonon mode that couples to the tip. The IETS signal of the fluorine defect corresponds well to the mass-scaled hydrogen signal. This mass-scaling is achieved by dividing with $\sqrt{19\,\mathrm{au}}/\sqrt{1\,\mathrm{au}} \approx 4.3$. Our results indicate a measurable difference between the hydrogen above/below case which may determine whether hydrogen is intercalated or adsorbed [179, 180].

IETS measurements of adsorbents and defects exhibit signatures making the method suitable for determining the local environment. This is so far difficult with STM topography maps as it does not differ between species. Also, all defects induce mid-gap states which greatly increase the elastic conductance which makes IETS difficult. We expect more detailed IETS measurements to exhibit signatures for different specie environments and thus usable as finger prints.

In [*168*] we compared IETS measurements and calculations on nitrogen substituted graphene. The atomic substitution exhibited an increase in the elastic contribution due to the broken symmetry, much like the defects shown in Figure 8.5. The calculations reproduce the gap features which remains a consequence of the Dirac point in graphene.

8.3 Summary

In this Chapter we have presented the LOE-WBA and LOE methods used to calculate the inelastic transmission due to electron-phonon coupling. We use the revised LOE method to reproduce inelastic signals for pristine graphene. To correctly capture the low DOS regime around the Dirac point one needs to *follow* the graphene DOS to account for the drastic change under non-equilibrium. This slight change captures all seen effects in experimental work and our calculations assert the out-of-plane mode to be dominant in the inelastic contribution. However, we emphasise the importance of a wider range of phonons which exhibit a blue-shift and magnitude shift of the IETS signal. We also presented several inelastic signals for various defects in graphene using the same approach. Structural defects such as Stone-Wales and hydrogen terminated graphene. Additionally adatoms, hydrogen, fluorine and SiO_2 was investigated. A generic feature for all defects was that the elastic transmission lifted the momentum rule, thus drastically increasing the elastic signal. This unfortunately makes IETS measurements more difficult.

Chapter 9

Conclusion & Outlook

9.1 Conclusion

In this thesis we have considered and expanded the non-equilibrium Green function (NEGF) theory in TranSIESTA. TranSIESTA is a NEGF code based on density functional theory within SIESTA.

The equations for the NEGF approach were expanded and generalised for generic $N_{\mathfrak{C}} \geq 1$ calculations. We emphasised that the complexity of solving $N_{\mathfrak{C}} > 2$ NEGF systems is primarily governed by the number of different electrode electronic distributions and secondly by the number of electrodes. A major workload of the NEGF method is the calculation of the density matrix via two distinct energy integrals, one called the equilibrium contribution and the other the non-equilibrium contribution. The equilibrium contribution can be calculated in the complex plane using different methods. We compared several different methods. In particular the continued fraction method and the residue theorem with variations. For some cases, the residue theorem converge faster than the continued fraction while for others it may be reversed. However, the continued fraction method has the benefit of using a single convergence parameter.

We highlighted the importance of the boundary conditions for NEGF calculations. The $N_{\mathfrak{C}} > 2$ case was covered as it is a non-trivial task to fully describe the electrostatics. A simple scheme only forcing the boundary conditions on the electrodes was presented. This is the simplest form estimating the *true* electrostatic boundary potential. It was shown that convergence rate and charge conservation was much improved by using a more complete electrostatic guess.

Most importantly, the NEGF method is based on inverting a matrix with dimension equal to the number of basis orbitals in the system. We showed that any $N_{\mathfrak{C}} \geq 1$ system may be described in a quasi 1D matrix such that an efficient block tridiagonal matrix inversion algorithm could be used. To reduce the bandwidth of the quasi 1D matrix several pivoting algorithms were tested and implemented. It was also shown that the implemented algorithms scaled very efficiently using a hybrid MPI+OpenMP parallelisation. The parallel fraction using the threading of TranSIESTA is an impressive 95%.

In conjunction with TranSIESTA a post-processing tool, TBtrans has been fully re-implemented. Several new features have emerged including; very efficient transmission calculation for extreme scale systems, bias interpolation schemes by interpolating the Hamiltonian and bond-currents to quantify local currents in nanostructures. For bond-

currents it is important to consider the entire Brillouin zone when visualising a periodic system. We also introduced a highly advanced eigenstate transmission projection method. This can distinguish between conducting and non-conducting localised eigenstates. This provide a general analysis tool to understand the complexity of transmission pathways in the view of localised eigenstates.

We implemented and used a gating method to perform investigations on graphene nanostructures. This gate-model takes into account the charging effects in a capacitor-like setup which is important when considering low-DOS devices such as graphene around the Dirac point. The gating model was used on a N-layer graphene (graphite) and it was shown that the gating effects are non-linear in terms of the screening length for different number of layers and gating levels. Additionally, a graphene nanoribbon was studied with NEGF and the gate model. This revealed a peculiar pinning effect which created an unsymmetric potential drop across a symmetric junction. This was attributed to a difference in coupling strength due to the difference in DOS arising from gating *and* NEGF. This simple understanding was confirmed by both adatom dopants and via smearing the non-equilibrium DOS that accounts for the difference in coupling strength. Such an unsymmetric potential drop may be experimentally found using Kelvin Probe Force Microscopy.

Subsequently, graphene interconnects were studied. We investigated pure graphene interconnects using a T-junction setup. This setup consisted of 3 electrodes in four different geometric configurations. An I-V characteristic was carried out in two different linear bias configurations. The pinning effect seen for the graphene nanoconstriction was recurring for the T-junction exhibiting a transistor-like effect. However, the relative difference was only ~ 5.

Lastly, we reproduced inelastic tunnelling on graphene in an STM setup using first principles DFT+NEGF theory. This setup included a 3-electrode geometry using the extended lowest order expansion of the inelastic current. The confident results spurred a set of predictive IETS signals for various defects in graphene. Both structural defects and adsorbant defects were considered.

9.2 Outlook

This thesis has contributed to a more elaborate NEGF method enabling $N_{\mathfrak{e}} \geq 1$ electrode calculations and gated systems. The importance of the boundary conditions for the Hartree potential was shown. Future studies may improve on Poisson solvers enabling other boundary conditions such as Dirichlet or Neumann. This would improve precision and convergence rate.

A fully N-branched inversion scheme may be implemented to enhance the inversion algorithm. This algorithm is much like the currently used algorithm in TBtrans. The block-tri-diagonal matrix may be divided into several quasi-1D blocks orthogonal to each other and sub-partitions of the matrix inversion may be performed. This abstraction will reduce memory requirements as well as increase throughput for very complex systems.

By using other Poisson solvers one may also improve the gate method by ensuring better boundary conditions. This is related to a difficulty in expressing the correct dipole for multi-layer structures with a gate strictly confined in a region. However, the models simplicity and usefulness is expected to be of relevance for the community of 2D materials.

Appendix A

Symmetries

Symmetries can be one of the most powerful tools to reduce problems to manageable sizes. Whenever you find symmetry, *use it*. In the following no time-dependency is shown, and it should not be regarded as a dependent variable. The system is assumed to conserve time reversal symmetry. This means that the solution to the Schrödinger equation is symmetric in t

$$\Psi(\mathbf{r}, t) \text{ and } \Psi^*(\mathbf{r}, -t), \tag{A.1}$$

are both solutions to the Schrödinger equation. Hence $\mathbf{B} = 0$ is assumed.

In this section the $\mathbf{k}$ dependence will be explicit and these operators are defined

$$\{\cdot\}^{\mathrm{T}} \quad \text{, transpose} \tag{A.2}$$

$$\{\cdot\}^{*} \quad \text{, element wise complex conjugate} \tag{A.3}$$

$$\{\cdot\}^{\dagger} \equiv (\{\cdot\}^{*})^{\mathrm{T}} \quad \text{, transpose and element wise complex conjugate.} \tag{A.4}$$

The Hamiltonian and overlap matrices are Hermitian matrices and obey the usual $\mathbf{k}$ symmetry

$$\mathbf{H}_{\mathbf{k}}^{\mathrm{T}} = \mathbf{H}_{-\mathbf{k}} \tag{A.5a}$$

$$\mathbf{H}_{\mathbf{k}}^{*} = \mathbf{H}_{-\mathbf{k}} \tag{A.5b}$$

$$\mathbf{H}_{\mathbf{k}}^{\dagger} = \mathbf{H}_{\mathbf{k}} \tag{A.5c}$$

the same thing applies for the overlap matrix:

$$\mathbf{S}_{\mathbf{k}}^{\mathrm{T}} = \mathbf{S}_{-\mathbf{k}} \tag{A.6a}$$

$$\mathbf{S}_{\mathbf{k}}^{*} = \mathbf{S}_{-\mathbf{k}} \tag{A.6b}$$

$$\mathbf{S}_{\mathbf{k}}^{\dagger} = \mathbf{S}_{\mathbf{k}} \tag{A.6c}$$

From the above symmetries the sub-matrix symmetries in a Hermitian matrix are

$$\mathbf{V}_{\mathbf{k}\{i,j\}}^{\mathrm{T}} = \mathbf{V}_{-\mathbf{k}\{j,i\}} \tag{A.7a}$$

$$\mathbf{V}_{\mathbf{k}\{i,j\}}^{*} = \mathbf{V}_{-\mathbf{k}\{i,j\}} \tag{A.7b}$$

$$\mathbf{V}_{\mathbf{k}\{i,j\}}^{\dagger} = \mathbf{V}_{\mathbf{k}\{j,i\}} \tag{A.7c}$$

The Green function is defined in the usual way (the retarded Green function)

$$\mathbf{G}_{\mathbf{k}} = \left[(E + \mathrm{i}\eta)\mathbf{S}_{\mathbf{k}} - \mathbf{H}_{\mathbf{k}}\right]^{-1}, \tag{A.8}$$

and simultaneously we define the advanced Green function

$$\mathbf{G}^A_\mathbf{k} = \left[(E - \mathrm{i}\eta)\mathbf{S}_\mathbf{k} - \mathbf{H}_\mathbf{k}\right]^{-1}. \tag{A.9}$$

In the following we set $\mathbf{S} = \mathbf{I}$ and assume it implicit without loss of generality. The following symmetries for the Green function applies

$$\mathbf{G}^\mathrm{T}_\mathbf{k} = \left[E + \mathrm{i}\eta - \mathbf{H}^\mathrm{T}_\mathbf{k}\right]^{-1} = \left[E + \mathrm{i}\eta - \mathbf{H}_{-\mathbf{k}}\right]^{-1} = \mathbf{G}_{-\mathbf{k}} \tag{A.10a}$$

$$\mathbf{G}^*_\mathbf{k} = \left[E - \mathrm{i}\eta - \mathbf{H}^*_\mathbf{k}\right]^{-1} = \left[E - \mathrm{i}\eta - \mathbf{H}_{-\mathbf{k}}\right]^{-1} = \mathbf{G}^A_{-\mathbf{k}} \tag{A.10b}$$

$$\mathbf{G}^\dagger_\mathbf{k} = \mathbf{G}^A_\mathbf{k}. \tag{A.10c}$$

The self energy term can be written in this form

$$\Sigma_\mathbf{k} = \mathbf{V}_{\mathbf{k}\{1,2\}}\mathbf{g}_\mathbf{k}\mathbf{V}_{\mathbf{k}\{2,1\}} = \mathbf{V}_{\mathbf{k}\{1,2\}}\mathbf{g}_\mathbf{k}\mathbf{V}^\dagger_{\mathbf{k}\{1,2\}}. \tag{A.11}$$

Symmetries governing the self energies are

$$\Sigma^\mathrm{T}_\mathbf{k} = \mathbf{V}^\mathrm{T}_{\mathbf{k}\{2,1\}}\mathbf{g}^\mathrm{T}_\mathbf{k}\mathbf{V}^\mathrm{T}_{\mathbf{k}\{1,2\}} = \Sigma_{-\mathbf{k}} \tag{A.12a}$$

$$\Sigma^*_\mathbf{k} = \mathbf{V}^*_{\mathbf{k}\{1,2\}}\mathbf{g}^*_\mathbf{k}\mathbf{V}^*_{\mathbf{k}\{2,1\}} = \Sigma^A_{-\mathbf{k}} \tag{A.12b}$$

$$\Sigma^\dagger_\mathbf{k} = \mathbf{V}_{\mathbf{k}\{1,2\}}\mathbf{g}^\dagger_\mathbf{k}\mathbf{V}_{\mathbf{k}\{2,1\}} = \Sigma^A_\mathbf{k} \tag{A.12c}$$

As seen from Eqs. (A.10) the same symmetries applies to the self energy as the Green function.

An often used quantity is the scattering matrix $\mathbf{\Gamma}_\mathbf{k}$ which we define as

$$\mathbf{\Gamma}_\mathbf{k} = \mathrm{i}[\Sigma_\mathbf{k} - \Sigma^\dagger_\mathbf{k}], \tag{A.13}$$

with the following symmetries

$$\mathbf{\Gamma}^\mathrm{T}_\mathbf{k} = \mathrm{i}\Sigma^\mathrm{T}_\mathbf{k} - \mathrm{i}\left(\Sigma^\dagger_\mathbf{k}\right)^\mathrm{T} = \mathbf{\Gamma}_{-\mathbf{k}} \tag{A.14a}$$

$$\mathbf{\Gamma}^*_\mathbf{k} = -\mathrm{i}\Sigma^*_\mathbf{k} + \mathrm{i}\left(\Sigma^\dagger_\mathbf{k}\right)^* = \mathbf{\Gamma}_{-\mathbf{k}} \tag{A.14b}$$

$$\mathbf{\Gamma}^\dagger_\mathbf{k} = -\mathrm{i}\Sigma^\dagger_\mathbf{k} + \mathrm{i}\left(\Sigma^\dagger_\mathbf{k}\right)^\dagger = \mathbf{\Gamma}_\mathbf{k} \tag{A.14c}$$

which is seen to retain the symmetries of a regular Hermitian matrix.

A.1 Transmission symmetries

Here we apply the symmetries derived above for the Green function transport formalism. We need the scattering matrices for two electrodes, $\mathbf{\Gamma}_{\mathfrak{e},\mathbf{k}}$ and $\mathbf{\Gamma}_{\mathfrak{e}',\mathbf{k}}$.

The transmission for a certain $\mathbf{k}$-point can be written as

$$T_{\mathbf{k},\mathfrak{e}'\to\mathfrak{e}} = \mathrm{Tr}\left[\mathbf{\Gamma}_{\mathfrak{e},\mathbf{k}}\mathbf{G}_\mathbf{k}\mathbf{\Gamma}_{\mathfrak{e}',\mathbf{k}}\mathbf{G}^\dagger_\mathbf{k}\right] \tag{A.15}$$

which is the transmission from $\mathfrak{e}'$ to $\mathfrak{e}$. The last triple product is also known as the *spectral function* $\mathcal{A}_{\mathfrak{e},\mathbf{k}} \equiv \mathbf{G}_\mathbf{k}\mathbf{\Gamma}_{\mathfrak{e},\mathbf{k}}\mathbf{G}^\dagger_\mathbf{k}$. A rule of thumb is that the spectral function in the

transmission function is the *origin* of the electron and the remaining scattering matrix is the *absorbing* electrode, whence $\mathfrak{e}' \to \mathfrak{e}$. The symmetries of $\mathcal{A}_{\mathfrak{e},\mathbf{k}}$ are the following

$$\mathcal{A}^{\mathrm{T}}_{\mathfrak{e},\mathbf{k}} = \mathbf{G}^{\dagger}_{-\mathbf{k}}\mathbf{\Gamma}_{\mathfrak{e},-\mathbf{k}}\mathbf{G}_{-\mathbf{k}} \equiv \widetilde{\mathcal{A}}_{\mathfrak{e},-\mathbf{k}}, \tag{A.16a}$$

$$\mathcal{A}^{*}_{\mathfrak{e},\mathbf{k}} = \mathbf{G}^{\dagger}_{-\mathbf{k}}\mathbf{\Gamma}_{\mathfrak{e},-\mathbf{k}}\mathbf{G}_{-\mathbf{k}} = \widetilde{\mathcal{A}}_{\mathfrak{e},-\mathbf{k}}, \tag{A.16b}$$

$$\mathcal{A}^{\dagger}_{\mathfrak{e},\mathbf{k}} = \mathcal{A}_{\mathfrak{e},\mathbf{k}}, \tag{A.16c}$$

with $\widetilde{\mathcal{A}}_{\mathfrak{e},\mathbf{k}}$ being the time reversed version of $\mathcal{A}_{\mathfrak{e},\mathbf{k}}$.

There are general rules for the trace of a function, i.e. it is invariant under cyclic permutations and is also invariant if taking the transpose. This means we can do

$$T_{\mathbf{k},\mathfrak{e}'\to\mathfrak{e}} = \mathrm{Tr}\left[\mathbf{\Gamma}_{\mathfrak{e},\mathbf{k}}\mathbf{G}_{\mathbf{k}}\mathbf{\Gamma}_{\mathfrak{e}',\mathbf{k}}\mathbf{G}^{\dagger}_{\mathbf{k}}\right] = \mathrm{Tr}\left[\mathbf{\Gamma}_{\mathfrak{e},\mathbf{k}}\mathcal{A}_{\mathfrak{e}',\mathbf{k}}\right] = \mathrm{Tr}\left[\mathbf{\Gamma}_{\mathfrak{e}',\mathbf{k}}\widetilde{\mathcal{A}}_{\mathfrak{e},\mathbf{k}}\right], \tag{A.17}$$

$$T_{\mathbf{k},\mathfrak{e}'\to\mathfrak{e}} = \mathrm{Tr}\left[\mathbf{\Gamma}_{\mathfrak{e},\mathbf{k}}\mathcal{A}_{\mathfrak{e}',\mathbf{k}}\right]^{\mathrm{T}} = \mathrm{Tr}\left[\widetilde{\mathcal{A}}_{\mathfrak{e}',-\mathbf{k}}\mathbf{\Gamma}_{\mathfrak{e},-\mathbf{k}}\right] \tag{A.18}$$

$$= \mathrm{Tr}\left[\mathbf{\Gamma}_{\mathfrak{e},-\mathbf{k}}\widetilde{\mathcal{A}}_{\mathfrak{e}',-\mathbf{k}}\right] = \mathrm{Tr}\left[\mathbf{\Gamma}_{\mathfrak{e}',-\mathbf{k}}\mathcal{A}_{\mathfrak{e},-\mathbf{k}}\right] = T_{-\mathbf{k},\mathfrak{e}\to\mathfrak{e}'}, \tag{A.19}$$

which is the time reversal symmetry argument. We can also take the complex conjugate

$$T_{\mathbf{k},\mathfrak{e}'\to\mathfrak{e}} = \mathrm{Tr}\left[\mathbf{\Gamma}_{\mathfrak{e},\mathbf{k}}\mathcal{A}_{\mathfrak{e}',\mathbf{k}}\right]^{*} = \mathrm{Tr}\left[\mathbf{\Gamma}_{\mathfrak{e},-\mathbf{k}}\widetilde{\mathcal{A}}_{\mathfrak{e}',-\mathbf{k}}\right] = T_{-\mathbf{k},\mathfrak{e}\to\mathfrak{e}'}, \tag{A.20}$$

which proves that the transmission is a real quantity.

Further we can calculate the sum of all spectral functions

$$\begin{aligned}\mathcal{A}_{\mathbf{k}} \equiv \sum_{\mathfrak{e}}\mathcal{A}_{\mathbf{k},\mathfrak{e}} &= \mathrm{i}\sum_{\mathfrak{e}}\mathbf{G}_{\mathbf{k}}\left[\Sigma_{\mathbf{k},\mathfrak{e}} - \Sigma^{\dagger}_{\mathbf{k},\mathfrak{e}}\right]\mathbf{G}^{\dagger}_{\mathbf{k}} = \mathrm{i}\left[\mathbf{G}_{\mathbf{k}} - \mathbf{G}^{\dagger}_{\mathbf{k}} + 2\mathrm{i}\eta\mathbf{G}_{\mathbf{k}}\mathbf{I}\mathbf{G}^{\dagger}_{\mathbf{k}}\right] \\ &\cong \mathrm{i}[\mathbf{G}_{\mathbf{k}} - \mathbf{G}^{\dagger}_{\mathbf{k}}],\end{aligned} \tag{A.21}$$

where $\mathbf{I}$ may be replaced by the overlap matrix if needed. The symmetries for $\mathcal{A}_{\mathbf{k}}$ are the general Hermitian symmetries

$$\mathcal{A}^{\mathrm{T}}_{\mathbf{k}} = \mathrm{i}(\mathbf{G}_{-\mathbf{k}} - \mathbf{G}^{\dagger}_{-\mathbf{k}}) = \mathcal{A}_{-\mathbf{k}} \tag{A.22a}$$

$$\mathcal{A}^{*}_{\mathbf{k}} = -\mathrm{i}(\mathbf{G}^{\dagger}_{-\mathbf{k}} - \mathbf{G}_{-\mathbf{k}}) = \mathcal{A}_{-\mathbf{k}} \tag{A.22b}$$

$$\mathcal{A}^{\dagger}_{\mathbf{k}} = \mathcal{A}_{\mathbf{k}}. \tag{A.22c}$$

Using Eqs. (A.22) we can easily prove that the transmission is symmetric of $\mathbf{k}$. We will in the following limit ourselves to two electrodes without loss of ambiguity

$$T_{\mathbf{k},\mathfrak{e}'\to\mathfrak{e}} = \mathrm{Tr}[\mathbf{\Gamma}_{\mathfrak{e},\mathbf{k}}\mathbf{G}_{\mathbf{k}}\mathbf{\Gamma}_{\mathfrak{e}',\mathbf{k}}\mathbf{G}^{\dagger}_{\mathbf{k}}] = \mathrm{Tr}[\mathbf{\Gamma}_{\mathfrak{e},\mathbf{k}}\mathcal{A}_{\mathbf{k}}] - \mathrm{Tr}[\mathbf{\Gamma}_{\mathfrak{e},\mathbf{k}}\mathcal{A}_{\mathfrak{e},\mathbf{k}}] \tag{A.23}$$

$$= \mathrm{i}\,\mathrm{Tr}\left[\mathbf{\Gamma}_{\mathfrak{e},\mathbf{k}}(\mathbf{G}_{\mathbf{k}} - \mathbf{G}^{\dagger}_{\mathbf{k}})\right] - \mathrm{Tr}[\mathbf{\Gamma}_{\mathfrak{e},\mathbf{k}}\mathcal{A}_{\mathfrak{e},\mathbf{k}}] = \mathrm{i}\,\mathrm{Tr}\left[\mathbf{\Gamma}_{\mathfrak{e},\mathbf{k}}(\mathbf{G}_{\mathbf{k}} - \mathbf{G}^{\dagger}_{\mathbf{k}})\right]^{\mathrm{T}} - \mathrm{Tr}[\mathbf{\Gamma}_{\mathfrak{e},\mathbf{k}}\mathcal{A}_{\mathfrak{e},\mathbf{k}}]^{\mathrm{T}} \tag{A.24}$$

$$= \mathrm{i}\,\mathrm{Tr}\left[\mathbf{\Gamma}_{\mathfrak{e},-\mathbf{k}}(\mathbf{G}_{-\mathbf{k}} - \mathbf{G}^{\dagger}_{-\mathbf{k}})\right] - \mathrm{Tr}[\mathbf{\Gamma}_{\mathfrak{e},-\mathbf{k}}\mathcal{A}_{\mathfrak{e},-\mathbf{k}}] = T_{-\mathbf{k},\mathfrak{e}'\to\mathfrak{e}}. \tag{A.25}$$

In effect one only needs to calculate the transmission for one $\mathbf{k}$ point and consequently the transmission for $-\mathbf{k}$ is known.

Appendix B

BTD Pivoting

This chapter further expands on the different pivoting algorithms in TranSIESTA. The structure we study is shown in Figure B.1 and is equivalent to the system in Figure 4.8. Note that the initial orbital arrangement can also be seen in Figure 4.8.

The system consists of 2,400 atoms with 9 orbitals per atom, totalling at 21,600 orbitals. The used pivoting algorithms are

CM/rev-CM Cuthill-Mckee [108], "rev" denotes the reverse pivoting which for Cuthill-Mckee has proven a much better candidate for reducing bandwiths [107]

GPS/rev-GPS Gibbs-Poole-Stockmeyer [109] algorithm for reducing bandwidth. This is a much more advanced method in that it first finds the periphery of the graph and creates a permutation matrix to reduce the bandwidth with the periphery points as end-points. As this has *implicit* reverse capability very rarely (if ever) should the reverse algorithm outperform the standard algorithm.

PCG/rev-PCG Peripheral connectivity graph. Based on the periphery end points from the GPS algorithm the subsequent bandwidth reduction is performed by a simple connectivity graph starting from the end-point. In this case can the reverse algorithm outperform the standard PCG.

GGPS/rev-GGPS Generalised Gibbs-Poole-Stockmeyer [110] algorithm for reducing bandwidth. This is a more complex variant of GPS and does better periphery analysis, as well as better intermediate analysis on the optimum path chosen for minimising the bandwidth. This also *implicitly* has reverse capability and very rarely (if ever) should the reverse algorithm outperform the standard algorithm.

In Figure B.2 the pivoting algorithms outlined above are shown for the 3-electrode system. The percentage of the total matrix elements is shown in the top-right while the lower-left denotes the used pivoting algorithm. The number of elements is reduced to approximately 16%-17%.

In Figure B.2 the pivoting algorithms outlined above are shown for the 4-electrode system. Figures are outlined similarly to Figure B.2. The number of elements is reduced to approximately 13%. The difference between the 3 and 4 electrode system indicates the difficulty in reducing the bandwidth as the 3-electrode system should have a smaller bandwidth compared to the 4-electrode because there is a dense block corresponding to the self-energy of the additional electrode.

TranSIESTA is implemented such that any custom atomic pivoting scheme may be used. Hence, one may test other pivoting algorithms on equal footing with the already implemented methods.

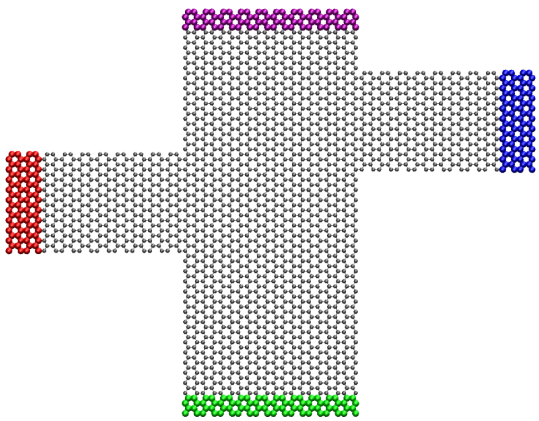


Figure B.1 | System investigated for pivoting algorithms. 3/4 electrodes coloured heavily and randomised atoms. 2,400 atoms with 9 orbitals per atom, yielding 21,600 orbitals.

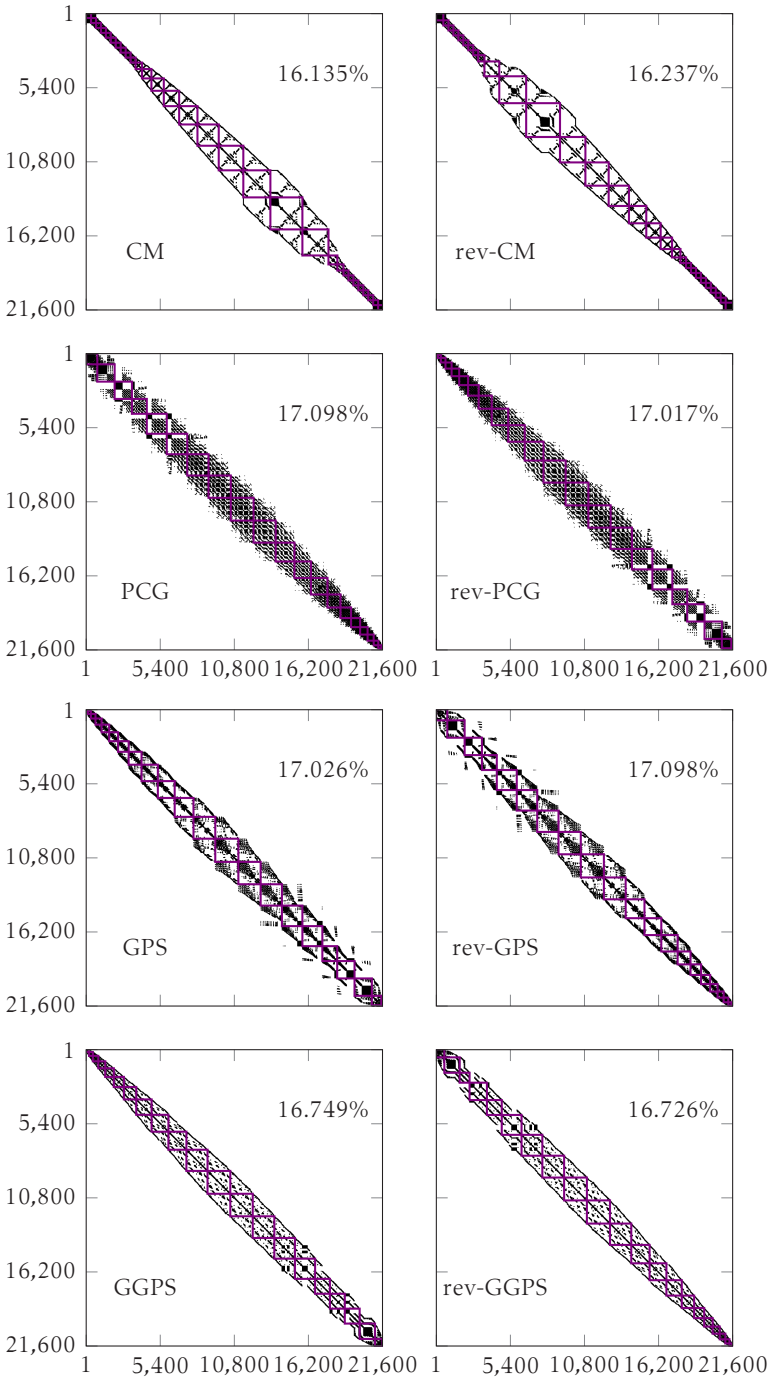


Figure B.2 | Cuthill-Mckee, peripheral connectivity graph, Gibbs-Poole-Stockmeyer and generalised Gibbs-Poole-Stockmeyer algorithms (top), and the equivalent reverse algorithms (bottom) for the 3-electrode system.

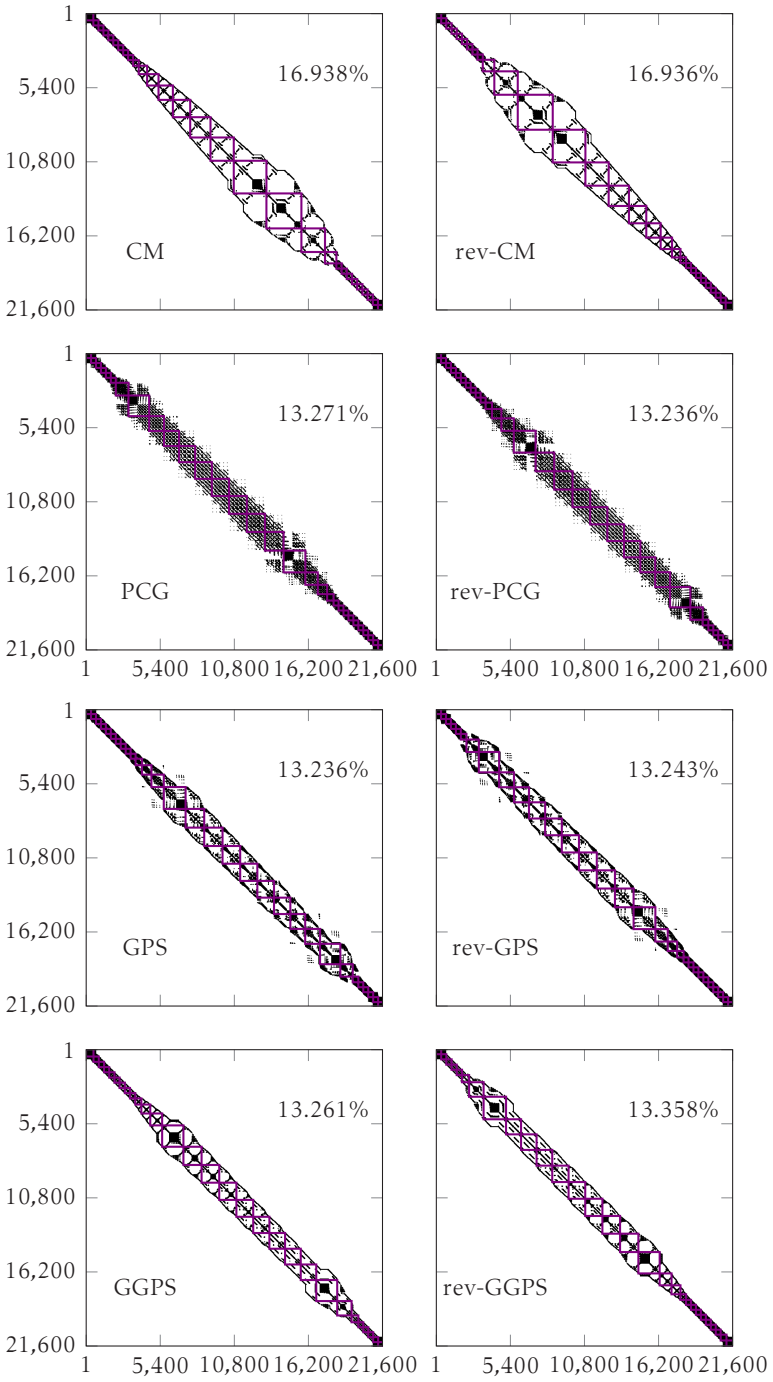


Figure B.3 | Cuthill-Mckee, peripheral connectivity graph, Gibbs-Poole-Stockmeyer and generalised Gibbs-Poole-Stockmeyer algorithms (top), and the equivalent reverse algorithms (bottom) for the 4-electrode system.

Appendix C

Tight-binding calculations

An important feature in TBtrans is the capability of performing large scale tight-binding transport calculations. To create parameterisations in native TBtrans data format we have developed a Python[1] package called sids[2]. The full usage of sids is beyond the scope of this section.

An example of a transport tight-binding calculation can be seen in the following code:

```
#!/usr/bin/env python
import numpy as np
import sids as si

alat = 1.42

# Setup tight-binding parameters
on, r0 = [-0.6 , 1.], 0. # on-site
n1, r1 = [-2.7 , 0.], alat # nearest neighbor
n2, r2 = [-0.2 , 0.], alat*3**.5 # next-nearest
n3, r3 = [-0.18, 0.], alat*2 # next-next-nearest

# Add small distance to remove numerical noise
dR = np.array([r0,r1,r2,r3]) + 0.1

# Define maximum orbital range
C = si.Atom(6, R = dR[-1])
# Create periodic structure
el = si.geom.graphene(alat, C).rotatec(-30).translate([alat/2,alat/2,0])
el.cell[1,1] *= -1 # only for VMD

# Create tight-binding parameter set and save it
si.TightBinding(el).construct(dR, [on,n1,n2,n3]).write('electrode.nc')

# Create full electrode and device
el = el.repeat(30, 0)
dev = el.tile(30, 1)
# Remove 1% atoms as defects (set same seed for reproducability)
np.random.seed(13240508)
rem = np.random.randint(len(dev), size=int(len(dev)*0.01/100))

# Reduce and add electrodes
dev = el.append(dev.remove(rem), 1).append(el, 1)

si.TightBinding(dev).construct(dR, [on,n1,n2,n3]).write('device.nc')
```

[1] Both Python2 and Python3 compliant.

[2] http://github.com/zerothi/sids

The script perform these operations:

1. Initialise needed packages, sids and numpy
2. Create atom with correct orbital range of 3rd nearest neighbour
3. Create electrode of the graphene unit-cell
4. Create tight-binding parameters in a 3rd nearest neighbour model [39, set D]
5. Expand minimal electrode to full device size and remove 1% of the atoms for a defected system
6. Create tight-binding parameters for full device

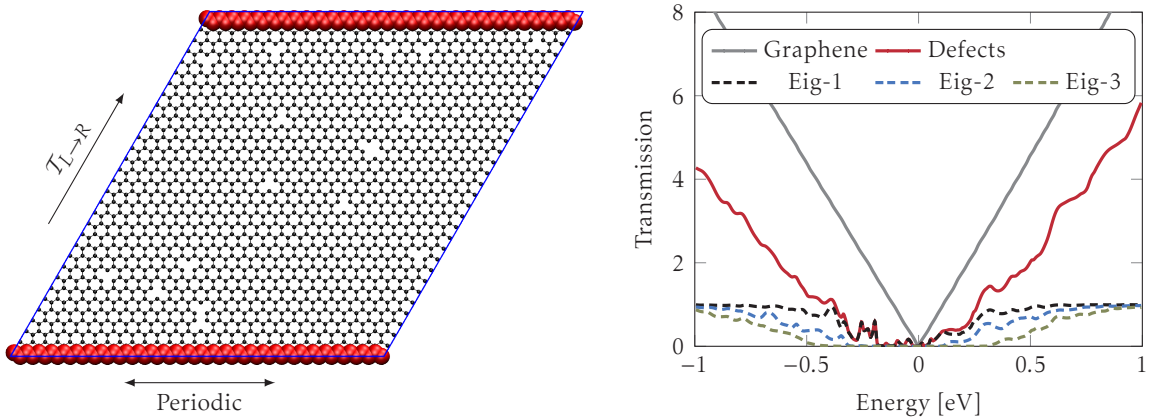
To create the following plots one needs the following FDF input file:

```
# Set device parameters
TBT.TSHS device.nc

# Calculating DOS of electrodes, also calculates bulk transmission of the electrodes
TBT.DOS.Elecs

# Calculate transmission eigenvalues (highest 4)
TBT.T.Eig 4

# Define k-grid [a, b, c] cell directions
%block TBT.k
 diag 51 1 1
%endblock

# Electrode setup
TBT.Elecs.Eta 0.001 eV
%block TBT.Elec.Left
  TSHS electrode.nc
  chemical-potential Left
  semi-inf-direction -a2
  electrode-position 1
  rep-a1 30 # use repetition
%endblock
%block TBT.Elec.Right
  TSHS electrode.nc
  chemical-potential Right
  semi-inf-direction +a2
  electrode-position end -1
  rep-a1 30 # use repetition
%endblock
```

To greatly speed up execution time we use the repetition scheme as defined in Eqs. (2.24) and (3.28).

The system, transmission and first 3 transmission eigenchannels are shown in Figure C.1. The system examined is shown in a). The electrodes are marked by big red atoms and $\mathbf{k}$-points are used in the transverse direction. Electrons originate from the bottom electrode. In b) the transmission for the defected structure is compared against pristine graphene. Furthermore the eigenchannel transmission is shown. With the eigenchannel transmission one can track the opening of new channels. New channels open approximately when the latest eigenchannel reaches 1/2.

This approach shows a generic implementation which can generate custom tight-binding parameterisations for a highly scalable transport code.



a) b)

Figure C.1 | a) Graphene with 1% defects. The electrodes are marked by the large red atoms. 51 transverse **k**-points is used in the transport calculation. Note the skewed transport direction aligned with the graphene unit-cell. b) The corresponding transmission of a 3rd nearest neighbour model for the geometry shown compared to pristine graphene (gray). There are no corrections neighbours to the defects. We also plot the first 3 eigenchannel transmissions for tracking their increase. They all 3 converge to 1 for high transmissions.

Appendix D

Publications

OPEN ACCESS
IOP Publishing
Journal of Physics: Condensed Matter
J. Phys.: Condens. Matter **27** (2015) 015001 (6pp)
doi:10.1088/0953-8984/27/1/015001

Spectroscopy of transmission resonances through a C_{60} junction

N L Schneider[1], N Néel[1], N P Andersen[2], J T Lü[2,3], M Brandbyge[2], J Kröger[4] and R Berndt[1]

[1] Institut für Experimentelle und Angewandte Physik, Christian-Albrechts-Universität zu Kiel, D-24098 Kiel, Germany
[2] DTU Nanotech, Technical University of Denmark, Anker Engelunds Vej 1 Bygning 101A, DK-2800 Kongens Lyngby, Denmark
[3] School of Physics, Huazhong University of Science and Technology, Wuhan, China
[4] Institut für Physik, Technische Universität Ilmenau, D-98693 Ilmenau, Germany

E-mail: joerg.kroeger@tu-ilmenau.de

Received 12 August 2014, revised 11 October 2014
Accepted for publication 24 October 2014
Published 19 November 2014

Abstract

Electron transport through a single C_{60} molecule on Cu(1 1 1) has been investigated with a scanning tunnelling microscope in tunnelling and contact ranges. Single-C_{60} junctions have been fabricated by establishing a contact between the molecule and the tip, which is reflected by a down-shift in the lowest unoccupied molecular orbital resonance. These junctions are stable even at elevated bias voltages enabling conductance measurements at high voltages and nonlinear conductance spectroscopy in tunnelling and contact ranges. Spectroscopy and first principles transport calculations clarify the relation between molecular orbital resonances and the junction conductance. Due to the strong molecule–electrode coupling the simple picture of electron transport through individual orbitals does not hold.

Keywords: scanning tunnelling microscopy, scanning tunnelling spectroscopy, single-molecule junction, transport calculation

(Some figures may appear in colour only in the online journal)

1. Introduction

The transport of electrons through atomic-scale contacts between two electrodes may be interpreted in terms of transport channels—quantum states extending between the electrodes—and their transmission probabilities τ_n [1]. In calculations, these probabilities vary drastically as a function of the electron energy [2–5]. Experiments have addressed the low-bias conductance [2, 6–14] and therefore typically provided little information on the transport channels and their τ_n. Metallic contacts between superconducting leads are a notable exception. In this case, Andreev reflections have been used to experimentally determine the τ_n at low bias [2, 6, 7]. However, experimental data on the variation of τ_n with the electron energy E were not reported.

Molecular junctions are expected to exhibit more and sharper structure of $\tau_n(E)$ [15]. The energy gap between occupied and unoccupied states of many molecules used in contact experiments is of the order of electron volts. As a result, probing their contributions to the conductance at contact is difficult. At the required elevated voltages and correspondingly large currents heating of the junctions occurs [16–19] and may lead to their destruction [17]. In a previous break-junction experiment the number of transport channels in benzene junctions has been determined using shot noise measurements [20]. To date, hardly any experimental data are available for highly conductive molecular junctions [20, 21].

Here, we report results from single-molecule contacts to C_{60} on Cu(1 1 1). Owing to a C_{60}-induced reconstruction the contacts are stable enough for conductance spectroscopy $[G(V)]$ at elevated bias voltages. Conductance resonances are observed and quantitatively analyzed using first-principles calculations. As expected, the molecular orbitals leave their

footprint on $G(V)$. However, it turns out that a picture of parallel transport through individual orbitals is too simple and only accounts for a fraction of the total conductance.

2. Experiment

Experiments were performed with a scanning tunnelling microscope (STM) operated at 8 K and in ultrahigh vacuum with a base pressure of 10^{-9} Pa. Chemically etched W tips and Cu(1 1 1) surfaces were cleaned by Ar^+ bombardment and annealing. C_{60} molecules were sublimated from a Ta crucible and adsorbed to clean Cu(1 1 1) at room temperature. After C_{60} deposition the surface was annealed at 500 K for 10 min. This preparation leads to the formation of well ordered C_{60} islands and a reconstruction of the Cu(1 1 1) surface [22]. To form a single-molecule contact the STM tip was brought closer towards the center of a molecule and the current was simultaneously recorded. Before and after contact experiments STM images and spectra of the differential conductance ($\mathrm{d}I/\mathrm{d}V$) were recorded to detect tip or molecule modifications. It turned out that the junctions are stable up to currents of ≈20 μA at elevated voltages of ≈1 V. Spectroscopy of $\mathrm{d}I/\mathrm{d}V$ was performed by modulating the sample voltage (10 $\mathrm{mV_{rms}}$, 8 kHz) and measuring the current response with a lock-in amplifier.

3. Theory

To simulate experimental data, tunnelling and contact junctions were modelled by a tetrahedral Cu tip attached to a 4 × 4 surface unit cell in a 7-layer Cu slab (substrate) and a C_{60} molecule adsorbed with a C hexagon to an on-top Cu(1 1 1) site (inset to figure 3(*b*)). The electronic structure, contact formation and conductance of the C_{60} junction were calculated within density functional theory (DFT) with the generalized gradient approximation (GGA-PBE) [23] to the exchange-correlation functional and a 2 × 2 surface k-point sampling. A localized atomic orbital basis set (SIESTA) [24] as well as a plane-wave basis set (VASP) [25] were used in order to access and avoid basis set superposition errors, which are present in calculations based on the linear combination of atomic orbitals. A series of calculations were performed in which the tip and the surface were approached towards each other in steps of 0.1 Å by decreasing the unit cell dimension in the approach direction. Relaxations of the tip tetraeder, the C_{60} molecule, and the two outermost surface layers were considered in these calculations. The TRANSIESTA [26] method was then applied to perform transport calculations of the linear conductance as well as non-equilibrium calculations of the current–voltage characteristics for two selected junction configurations in the tunnelling and contact range.

4. Results and discussion

The inset to figure 1 shows a constant-current STM image from the interior of a C_{60} island. Similar to previous observations from other surfaces [27–30], the molecules exhibit three protrusions, which are due to the next-to-lowest unoccupied

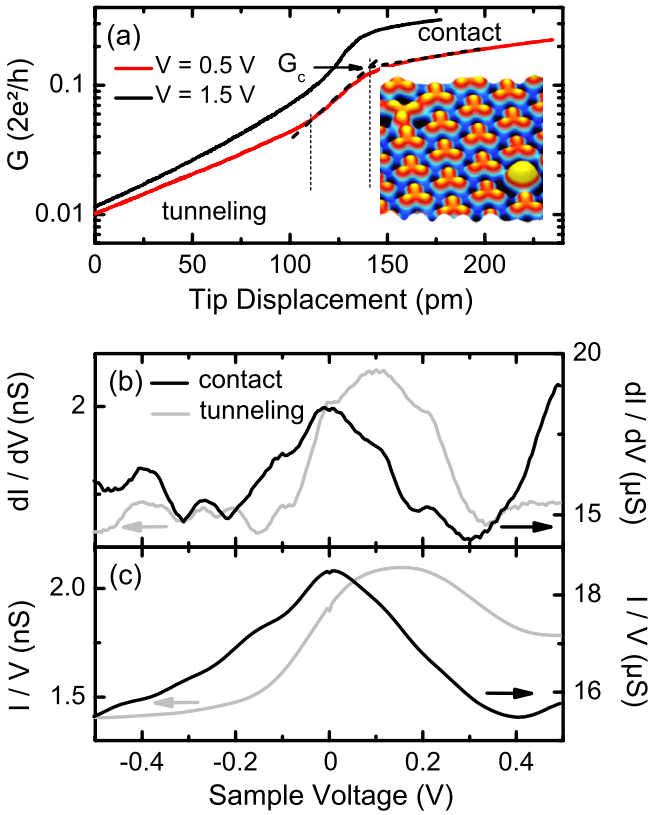


Figure 1. (*a*) Conductance of C_{60} on Cu(1 1 1) in a STM versus displacement of the tip towards the molecule. Red and black lines indicate data recorded at sample voltages $V = 0.5$ V and 1.5 V, respectively. Dashed lines illustrate the definition of the point of contact formation and the corresponding conductance G_c. Vertical lines separate different conductance ranges, namely tunnelling, transition, and contact. Inset: pseudo-three-dimensional representation of a constant-current STM image of a C_{60} monolayer on Cu(1 1 1) (1.5 V, 100 pA, 6 × 6 $\mathrm{nm^2}$). (*b*) $\mathrm{d}I/\mathrm{d}V$ and (*c*) I/V curves as a function of V acquired at fixed tip heights in the tunnelling range (grey) and at contact (black). Tip heights were set by disabling the STM feedback loop at 0.6 V and, respectively, 0.7 nA and 11.6 μA in tunnelling and contact ranges. The small sharp feature at $V = 0$ in (*c*) is a numerical artefact.

molecular orbital (LUMO+1). This orbital is centred at the three C pentagons that surround a hexagon in the observed trifoliate way. When the tip is brought closer to a molecule the conductance varies as displayed in figure 1(*a*). Tunnelling, transition and contact ranges are defined using the intersections of exponential fits to the conductance data (indicated in the lower curve of figure 1(*a*)) [31]. The two data sets shown were recorded at sample voltages $V = 0.5$ V and 1.5 V and exhibit different conductances G_c at the transition to contact, namely ≈0.15 G_0 and ≈0.25 G_0, respectively.

To relate the bias dependence of the conductance to the electronic structure of the adsorbed molecule, $\mathrm{d}I/\mathrm{d}V$ spectra were acquired at constant tip–sample separations. The grey and black lines in figure 1(*b*) show data sets from the tunnelling and contact ranges, respectively. In the tunnelling range constant-height $\mathrm{d}I/\mathrm{d}V$ spectra of C_{60} can be routinely recorded over a fairly wide range of bias voltages. At contact, however, currents on the order of 10 μA flow and the junction usually becomes unstable at much lower voltages. Owing to

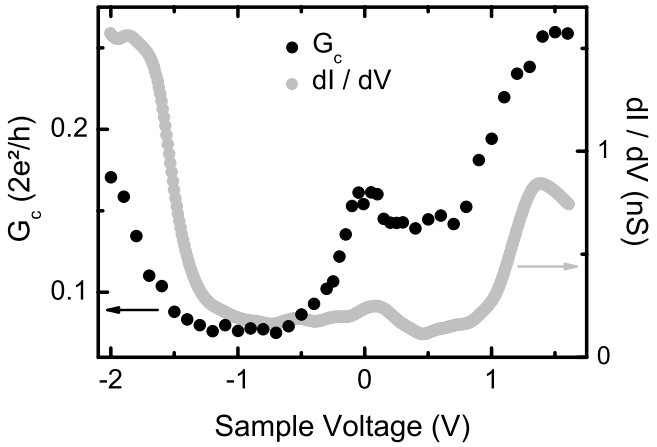


Figure 2. Contact conductance G_c (black) and tunnelling $\mathrm{d}I/\mathrm{d}V$ data (grey) versus sample voltage. G_c has been extracted from individual conductance-versus-displacement curves (see figure 1(*a*)) acquired at voltages between −2 V and 1.6 V.

the particular stability of the structures used here, the range from −0.5 to 0.5 V can be probed. The tunnelling data of figure 1(*b*) show a peak centred at ≈100 mV. Previous reports have shown that it is due to the C_{60} LUMO [22, 32]. At contact, a similar peak is observed, albeit broadened and shifted to ≈0 mV. Our calculations (*vide infra*) reveal a hybridization of C_{60} with the tip. It is not clear that the contact data may simply be interpreted in terms of a density of states of the junction. Ignoring this issue for the moment, we find that the differences of the contact spectrum are consistent with the calculated electronic structure of the junction. Figure 1(*c*) shows the conductances $G = I/V$, which were recorded along with the $\mathrm{d}I/\mathrm{d}V$ spectra.

To extend the accessible range of voltages at contact, a different method was used. Rather than sweeping the voltage at a fixed tip–molecule distance, the current was recorded as a function of the tip height while keeping V fixed. The conductance G_c at contact formation was then extracted as described above (figure 1). The results are depicted in figure 2 (black) together with a constant-height tunnelling $\mathrm{d}I/\mathrm{d}V$ spectrum (grey). Over the investigated voltage range $-2\,\mathrm{V} \leqslant V \leqslant 1.6\,\mathrm{V}$, the contact conductance G_c varies significantly between $0.07\,G_0$ and $0.26\,G_0$. Moreover, the maxima at $V \approx 0$ V and ≈1.5 V) are close to maxima of the tunnelling $\mathrm{d}I/\mathrm{d}V$ data at ≈100 mV (LUMO) and ≈1.3 V (LUMO+1).

DFT calculations based on the structure shown in the inset to figure 3(*b*) were performed to rationalize the experimentally observed shift of the LUMO resonance to lower energies upon the tunnelling-to-contact transition. Figure 3 shows the zero-bias transmission functions calculated with a 9 × 9 surface k-point sampling. The peak-like structure close to the Fermi energy (E_F) is due to the LUMO resonance, which clearly shifts towards E_F upon decreasing tip–C_{60} distances and thus increasing hybridization. According to a Bader charge analysis based on the VASP calculations [33] a charge of ≈0.4 e is transferred from the tip to the molecule. For large tip–molecule distances in the tunnelling range the transmission function exhibits an approximate exponential variation with

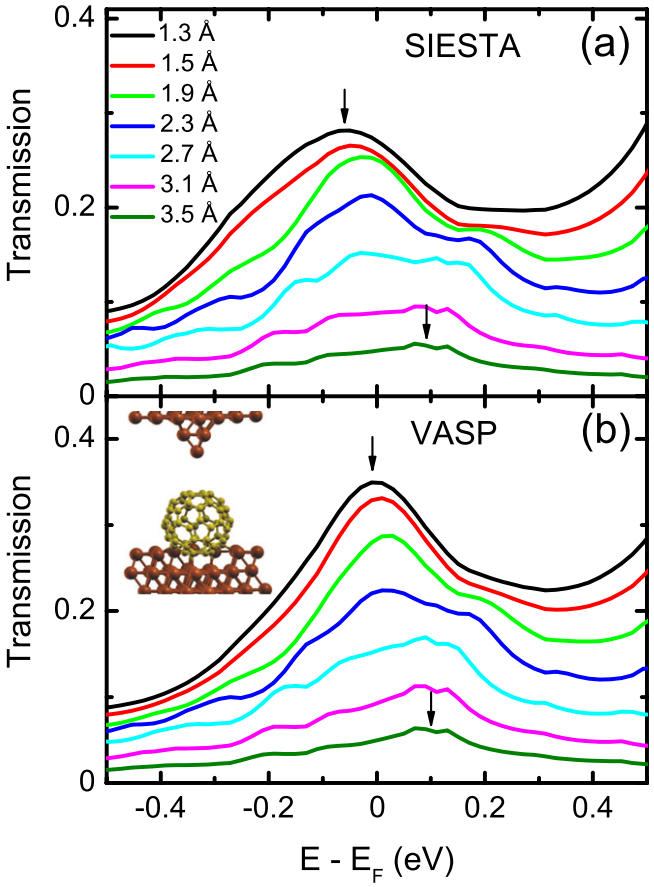


Figure 3. Transmission functions calculated using relaxed geometries obtained from SIESTA (*a*) and VASP (*b*) with PBE-GGA labelled by the unrelaxed tip–C_{60} distances measured from the tip apex atom to the C hexagon plane. Beyond contact formation at 1.7 Å (unrelaxed distance) the relaxed tip–C_{60} distance hardly changes. Instead the tip is progressively compressed in both SIESTA and VASP calculations. The transmission functions show the change of the C_{60} LUMO resonance close to the Fermi energy (E_F) at zero bias during contact formation. A clear lowering of the resonance energy towards E_F is observed (vertical arrows). The functional form of the transmission versus energy roughly shows an exponential decrease in magnitude with increasing tip–C_{60} distance in the tunnelling range starting from ≈2.7 Å. The inset to (*b*) shows the structural model of the junction used in the calculations.

the distance. The contact formation may be observed as a deviation from this scaling behaviour due to the onset of chemical interactions leading to a resonance shift and broadening. Such deviations are indeed present for a tip–hexagon distance between 2.3 and 2.7 Å with a corresponding conductance of $\approx 0.2\,G_0$. According to the experiments the contact conductance close to zero bias voltage is $\approx 0.16\,G_0$ (figure 2). For tip–molecule distances at which repulsive interactions start to deform the tip apex into a flat geometry the conductance is close to $0.3\,G_0$. At positive energies, starting from ≈0.4 eV the tail of a second transmission resonance has been observed in the calculations (figure 3).

Figure 3 compares results using geometries obtained from SIESTA (figure 3(*a*)) and VASP PBE-GGA (figure 3(*b*)) calculations. Both methods lead to virtually identical evolutions of the energy-dependent transmission functions. In the SIESTA calculations the tip apex atom is slightly

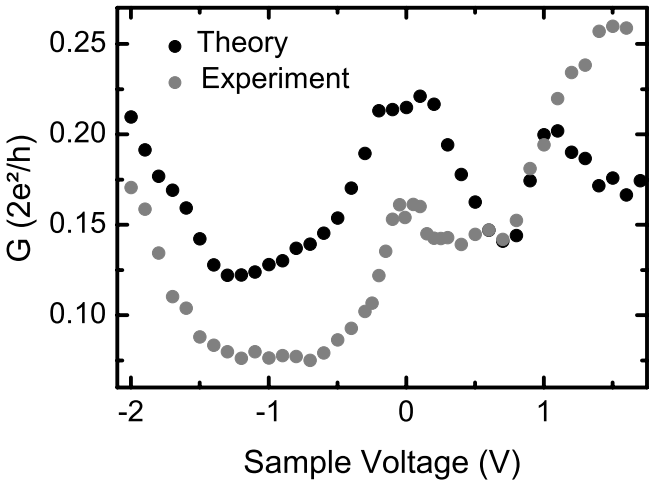


Figure 4. Comparison of experimental G_c (grey) and calculated conductances (black). In the calculations the tip apex atom was separated by 2 Å from the closest C_{60} hexagon.

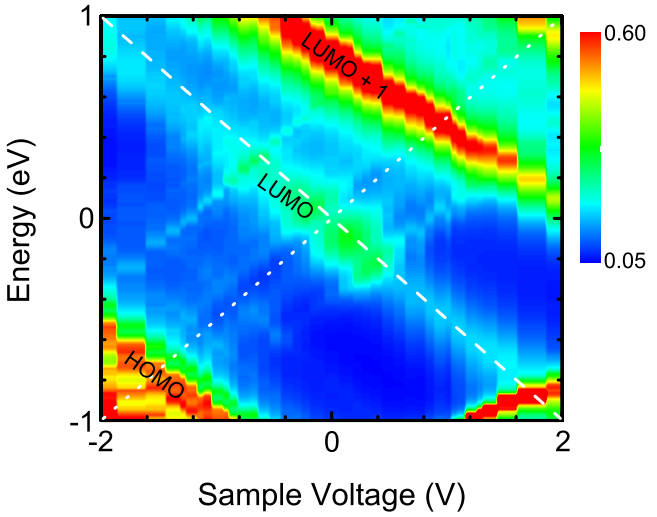


Figure 5. Density plot of the transmission function, $T(E, V) = \sum_n \tau_n(E, V)$, for different energies E and voltages V in the contact range. The energies of the transmission maxima labelled HOMO, LUMO, and LUMO+1 approximately follow the chemical potential of the sample $-eV/2$ (dashed line). Weaker transmission features are related to tip states and follow the chemical potential of the tip ($+eV/2$, dotted line).

stretched towards the C_{60} molecule upon approaching the tip to the surface. Further approach of the tip leads to a repulsive tip–molecule interaction and deforms the tip apex towards a flat geometry. Including van der Waals forces in the VASP calculations [34] (not shown) did not lead to markedly different conductance behaviour with tip displacement. While the calculated conductances in the contact range are in good agreement with the experimentally observed values, the calculated exponential variation of the conductance with the tip–molecule distance in the tunnelling range is somewhat larger than in the experiments. This observation is probably due to the use of an atomic basis set description of the tip [35]. In addition, as is evident from figure 1(*a*) the slope of conductance-displacement characteristics depends on the bias voltage both in tunnelling and contact ranges. Thus, the bias voltage plays a significant role in the effective tunnelling barrier as well as in the contact formation. This effect could likewise involve bias voltage-induced atomic relaxations. In the calculations we neglect the computationally very demanding bias-induced relaxations but note that these can lead to significant forces in the contact range [19].

Next, full non-equilibrium calculations based on the junction geometries obtained from SIESTA were performed. The resulting bias voltage-dependent conductances at contact in figure 4 (black) displays much similarity with the experimental data (grey). A resonance with ≈1 V full width at half maximum is centred around 0 V. In addition, resonances are observed in the calculations for negative and positive bias voltages, which are similar to the experimental results.

Figure 5 shows how these transmission maxima change with the sample voltage. As expected from the strong C_{60}–Cu(1 1 1) coupling, the maxima essentially follow the sample chemical potential, which is defined as $-eV/2$ with V the sample voltage. Slight deviations from the evolution of the sample chemical potential are due to a small coupling between the molecule and the tip. As exposed in detail next, the notion of individual molecular states and their identification with transmission maxima requires some caution due to the strong molecule–electrode coupling. Allowing only transport via the highest occupied molecular orbital (HOMO), LUMO, or LUMO+1 states of the molecular region in the calculation does indeed warrant this designation. However, the simple picture of parallel transport via each of these orbitals only accounts for a fraction of the total transmission, as demonstrated below.

The standard Green function expression [26] for the elastic transmission, T, reads

$$T(E) = \mathrm{Tr}\left[G(E)\Gamma_{\mathrm{L}}(E)G^{\dagger}(E)\Gamma_{\mathrm{R}}(E)\right], \qquad (1)$$

where G is the retarded Green function matrix and $\Gamma_{\mathrm{L/R}}$ the electrode coupling matrices in the full basis set describing the scattering region. The transmission eigenchannels τ_n, with $T(E) = \sum_n \tau_n(E)$ (figure 6(*a*)) [15], provide an exact decomposition of the total transmission, but do typically not show a separation into molecular orbitals. The dominant transmission eigenchannel (black line in figure 6(*a*), transmission probability $\tau_1(E)$) closely follows the C_{60} HOMO, LUMO, and LUMO+1. The single channel giving rise to the LUMO transmission is due to the coupling of the rotational symmetric *s* orbital on the tip around the Fermi energy. Close to the LUMO+1 energy (≈0.9 eV) three channels contribute to the conductance.

To test whether transport takes place in parallel via molecular orbitals of C_{60} the eigenstates of the molecule-projected self-consistent Hamiltonian (MPSH) [36] were calculated, which correspond to the HOMO, LUMO, and LUMO+1. The coupling matrices in equation (1) were then projected onto each of these orbitals in order to evaluate the transmission probability of electrons that enter and exit the C_{60} junction via one of these orbitals. The projected electrode couplings read $\Gamma^{\alpha}_{\mathrm{L/R}} = P_{\alpha}\Gamma_{\mathrm{L/R}}P_{\alpha}$, where α is one of the molecular orbitals and P_{α} a corresponding projector

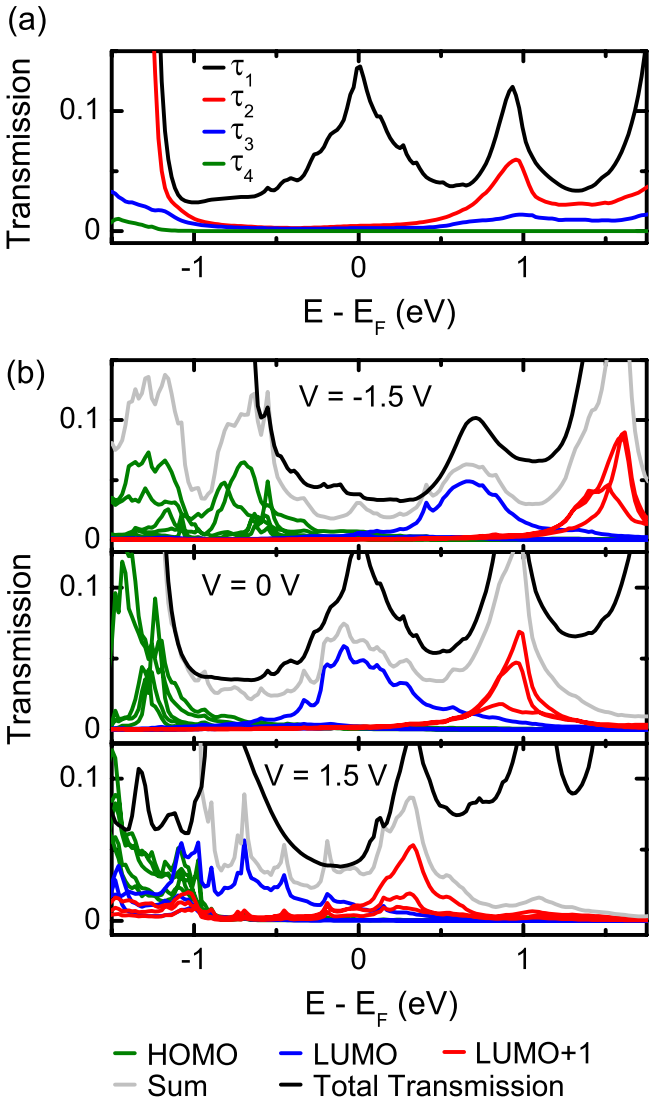


Figure 6. (*a*) Eigenchannel transmissions, $\tau_n(E)$ ($n = 1, \ldots, 4$), at zero bias for the contact range (tip–C_{60} distance: 1.9 Å) averaged over all k points. The dominant transmission channel (black, transmission probability $\tau_1(E)$) closely follows the C_{60} HOMO, LUMO, and LUMO+1. Around the energy of the LUMO+1 (≈0.9 eV) three channels contribute to the conductance. (*b*) Orbital-projected transmission functions $T_\alpha(E)$ for the C_{60} HOMO (green), LUMO (blue) and LUMO+1 (red) resonances evaluated at the indicated voltages (tip–C_{60} distance: 1.9 Å, i.e. in the contact range). Several contributions are observed from each orbital due to their partial degeneracy. The grey line depicts the sum of all transmissions, $\sum_\alpha T_\alpha(E)$, while the solid black line is the total transmission calculated according to equation (1).

for the same subset region. Using $\Gamma^\alpha_{L/R}$ in equation (1) the resulting T_α,

$$T_\alpha(E) = \mathrm{Tr}\left[G(E)\Gamma^\alpha_L(E)G^\dagger(E)\Gamma^\alpha_R(E)\right], \qquad (2)$$

is interpreted as the electron transmission via orbital α. Therefore, $T_\alpha(E)$ may be used to judge the extent to which the assignment of individual orbitals to a specific transmission feature is valid. Figure 6(*b*) shows the k-averaged transmission functions $T_\alpha(E)$ for the HOMO (green), LUMO (blue) and the LUMO+1 (red) at bias voltages of −1.5 V (top), 0 V (middle),

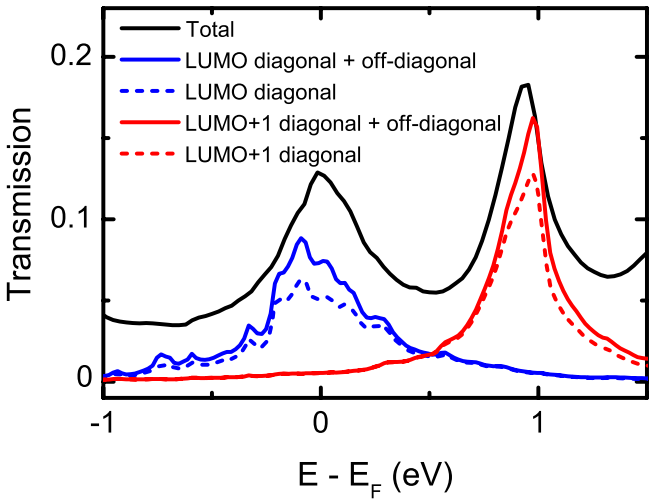


Figure 7. Contributions to LUMO and LUMO+1 transmissions from diagonal (dashed lines) terms (α = LUMO, LUMO + 1 in equation (2)), and off-diagonal/mixing (full lines) contributions to the LUMO, LUMO+1 transmission from nearby orbitals (HOMO, LUMO, LUMO+1) (equation (3)).

+1.5 V (bottom). The fivefold (threefold) degeneracy of the HOMO (LUMO, LUMO+1) of the free C_{60} molecule is partly lifted for C_{60} attached to the electrodes. Each degenerate orbital contributes to the transmission. The sum of all transmission curves is plotted as a grey line.

It is clear from figure 6(*b*) that the projected transmissions of the HOMO, LUMO, and LUMO+1 do indeed follow the main peaks in the total transmission. The shift of individual transmission peaks to lower energies with increasing bias voltage (figure 5) is also visible. However, the projected transmissions of HOMO, LUMO, LUMO+1 do not add up to the total transmission (black line in figure 6(*b*)) in the energy range where we expect the conductance to take place in these. Mostly the sum of projected transmissions (grey line) is lower than the total transmission (black line)[4]. In figure 6(*b*) we have restricted the calculations to the HOMO, LUMO and LUMO+1. It is necessary to include more molecular orbitals in the sum to obtain peaks farther from E_F. For instance, in figure 6(*b*) the total transmission peaks for energies exceeding 1 eV cannot be accounted for by the contributions from the chosen projected orbitals (grey line) and would require inclusion of the LUMO+2. For obtaining the full picture the off-diagonal contributions from different α in equation (2) must be considered. Electrons enter and exit the molecule via different orbitals α, α' and may play a significant role for the transmission. Due to the strong molecule–electrode coupling the resonances originating from the molecular orbitals have weight inside the metal and mix with each other. The terms that describe the mixing are mainly positive leading to a lower sum of projected transmission. Occasionally they are negative which leads to a higher projected transmission at certain energies. The mixing can be quantified by the corresponding

[4] At some energies the sum of projected transmissions occasionally exceeds the total transmission, which is due to interference phenomena.

off-diagonal transmission,

$$T_{\alpha,\alpha'}(E) = \mathrm{Tr}\left[G(E)\Gamma_{\mathrm{L}}^{\alpha}(E)G^{\dagger}(E)\Gamma_{\mathrm{R}}^{\alpha'}(E)\right], \qquad (3)$$

where α and α' are different. In figure 7 we show the transmission for the LUMO and the LUMO+1 with (full lines) and without (dashed lines) the mixing with nearby orbitals including HOMO, LUMO, LUMO+1. Including the mixing yields a significant contribution and thus mixing between orbitals due to the strong coupling in the junction plays a significant role. Thus we conclude that the approximation of parallel transport through individual C_{60}-states neglects significant contributions to the conductance.

5. Conclusion

Conductance spectroscopy and first-principles transport calculations clarify the role molecular orbital resonances play in determining the conductance of a molecular junction at contact. A picture of electron transport through individual orbitals [2, 37] does not hold. Rather, for strong molecule–electrode couplings mixing of orbitals must be considered for the correct description of the junction conduction.

Acknowledgments

Financial support by the Deutsche Forschungsgemeinschaft through SFB 677 is acknowledged. We thank T Frederiksen (San Sebastian) for useful comments and the DCSC for computational resources.

References

[1] Landauer R 1970 *Phil. Mag.* **21** 863
[2] Scheer E, Agraït N, Cuevas J C, Yeyati A L, Ludoph B, Martín-Rodero A, Bollinger G R, van Ruitenbeek J M and Urbina C 1998 *Nature* **394** 154
[3] Brandbyge M, Kobayashi N and Tsukada M 1999 *Phys. Rev.* B **60** 17064
[4] Pauly F, Viljas J K, Huniar U, Häfner M, Wohlthat S, Bürkle M, Cuevas J C and Schön G 2008 *New J. Phys.* **10** 125019
[5] Bilan S, Zotti L A, Pauly F and Cuevas J C 2012 *Phys. Rev.* B **85** 205403
[6] Scheer E, Joyez P, Esteve D, Urbina C and Devoret M H 1997 *Phys. Rev. Lett.* **78** 3535
[7] Naaman O and Dynes R C 2004 *Solid State Commun.* **129** 299
[8] Agraït N, Untiedt C, Rubio-Bollinger G and Vieira S 2002 *Phys. Rev. Lett.* **88** 216803
[9] Néel N, Kröger J, Berndt R and Pehlke E 2008 *Phys. Rev.* B **78** 233402
[10] Vitali L, Ohmann R, Stepanow S, Gambardella P, Tao K, Huang R, Stepanyuk V S, Bruno P and Kern K 2008 *Phys. Rev. Lett.* **101** 216802
[11] Calvo M R, Fernandez-Rossier J, Palacios J J, Jacob D, Natelson D and Untiedt C 2009 *Nature* **458** 1150
[12] Néel N, Kröger J and Berndt R 2010 *Phys. Rev.* B **82** 233401
[13] Bork J, Zhang Y H, Diekhöner L, Borda L, Simon P, Kroha J, Wahl P and Kern K 2011 *Nat. Phys.* **7** 901
[14] Choi D J, Rastei M V, Simon P and Limot L 2012 *Phys. Rev. Lett.* **108** 266803
[15] Paulsson M and Brandbyge M 2007 *Phys. Rev.* B **76** 115117
[16] Néel N, Kröger J, Limot L, Frederiksen T, Brandbyge M and Berndt R 2007 *Phys. Rev. Lett.* **98** 065502
[17] Schulze G *et al* 2008 *Phys. Rev. Lett.* **100** 136801
[18] Néel N, Kröger J and Berndt R 2011 *Nano Lett.* **11** 3593
[19] Ulstrup S, Frederiksen T and Brandbyge M 2012 *Phys. Rev.* B **86** 245417
[20] Kiguchi M, Tal O, Wohlthat S, Pauly F, Krieger M, Djukic D, Cuevas J C and van Ruitenbeek J M 2008 *Phys. Rev. Lett.* **101** 046801
[21] Chen W, Widawsky J R, Vázquez H, Schneebeli S T, Hybertsen M S, Breslow R and Venkataraman L 2011 *J. Am. Chem. Soc.* **133** 17160–3
[22] Pai W W *et al* 2010 *Phys. Rev. Lett.* **104** 036103
[23] Perdew J P, Burke K and Ernzerhof M 1996 *Phys. Rev. Lett.* **77** 3865
[24] Soler J M, Artacho E, Gale J D, Garcia A, Junquera J, Ordejón P and Sánchez-Portal D 2002 *J. Phys.: Condens. Matter* **14** 2745
[25] Kresse G and Furthmüller J 1996 *Phys. Rev.* B **54** 11169
[26] Brandbyge M, Mozos J L, Ordejón P, Taylor J and Stokbro K 2002 *Phys. Rev.* B **65** 165401
[27] Abel M, Dimitriev A, Fasel R, Lin N, Barth J V and Kern K 2003 *Phys. Rev.* B **67** 245407
[28] Néel N, Kröger J and Berndt R 2008 *Nano Lett.* **8** 1219
[29] Néel N, Limot L, Kröger J and Berndt R 2008 *Phys. Rev.* B **77** 125431
[30] Schull G, Néel N, Becker M, Kröger J and Berndt R 2008 *New J. Phys.* **10** 065012
[31] Kröger J, Néel N and Limot L 2008 *J. Phys.: Condens. Matter* **20** 223001
[32] Wang L L and Cheng H P 2004 *Phys. Rev.* B **69** 045404
[33] Sanville E, Tang W and Henkelman G 2009 *J. Phys.: Condens. Matter* **21** 084204
[34] Grimme S 2006 *J. Comput. Chem.* **27** 1787
[35] Garcia-Lekue A and Wang L W 2010 *Phys. Rev.* B **82** 035410
[36] Stokbro K, Taylor J, Brandbyge M, Mozos J L and Ordejón P 2003 *Comput. Mater. Sci.* **27** 151–60
[37] Cuevas J C, Yeyati A L and Martín-Rodero A 1998 *Phys. Rev. Lett.* **80** 1066

PHYSICAL REVIEW B **91**, 121403(R) (2015)

Unravelling the role of inelastic tunneling into pristine and defected graphene

Mattias L. N. Palsgaard, Nick P. Andersen, and Mads Brandbyge*

Center for Nanostructured Graphene, Department of Micro- and Nanotechnology, Technical University of Denmark, Ørsteds Plads, Building 345E, DK-2800 Kongens Lyngby, Denmark

(Received 11 October 2014; revised manuscript received 17 February 2015; published 5 March 2015)

We present a first principles method for calculating the inelastic electron tunneling spectroscopy (IETS) on gated graphene. We reproduce experiments on pristine graphene and point out the importance of including several phonon modes to correctly estimate the local doping from IETS. We demonstrate how the IETS of typical imperfections in graphene can yield characteristic fingerprints revealing, e.g., adsorbate species or local buckling. Our results show how care is needed when interpreting scanning tunneling microscopy images of defects due to suppression of the elastic tunneling on graphene.

DOI: 10.1103/PhysRevB.91.121403 PACS number(s): 63.22.Rc, 63.20.dk, 68.37.Ef, 72.10.Di

Imperfections such as lattice defects, edges, and impurity/dopant atoms can degrade the superb transport properties of graphene [1–4], or may, if controlled, lead to new functionality [5]. Scanning tunneling microscopy and spectroscopy (STM/STS) have been used extensively to obtain insights into the local electronic structure of graphene with atomic resolution [6–10]. However, contrary to most STM/STS experiments where elastic tunneling plays the dominant role, for graphene the inelastic tunneling prevails. This was clearly demonstrated experimentally as a "giant" signal in the second derivative of the current with regard to voltage obtained in inelastic electron tunneling spectroscopy (IETS) performed on gated, pristine graphene with STM [6–8]. The pronounced inelastic features are rooted in the electronic structure of graphene. The electrons have to enter the Dirac points corresponding to a finite in-plane momentum leading to weak elastic tunneling. The IETS signal of pristine graphene has been reproduced qualitatively by Wehling et al. by considering the change in the wave-function decay when displacing the carbon atoms along a selected frozen zone-boundary out-of-plane phonon [11]. In general, the important role of the inelastic process complicates the interpretation of STM results on graphene. Ideally, STM images on graphene structures should be accompanied by local STS/IETS measurements, in order to distinguish between contributions from the inelastic and elastic channel. On the other hand, first principles calculations based on density functional theory (DFT) often provide essential unbiased insights into STM/STS/IETS experiments to help the interpretation.

In this Rapid Communication we present a method for DFT calculations of the STS/IETS on gated graphene. We demonstrate its predictive power by reproducing from first principles the features of the experimental results for the giant inelastic conductance of gated pristine graphene [6–8]. We then provide results for IETS signals of defected graphene systems by determining the relative impact on the current of the various phonon modes. In particular, we identify inelastic fingerprints of selected defects, suggesting that IETS measurements can be a powerful tool in the characterization of imperfect graphene. Our analysis also illustrates how one should keep in mind the in-plane momentum conservation when performing STM on graphene. In particular, we demonstrate how defects can locally lift the suppression of elastic tunneling. The resulting increased local conductance may be misinterpreted as a high local density of states (LDOS).

Method. The calculations are performed with DFT using the SIESTA/TRANSIESTA [12,13] code and the INELASTICA package for inelastic transport [14]. Our system, shown in Fig. 1, is divided into a top lead (source), device and sample lead (drain) following the standard transport setup [13–15]. We consider a suspended graphene sheet located 5 Å below the tip of a gold STM probe model and a voltage bias between the tip and sample leads. The electron-phonon coupling ($\mathbf{M}^{\lambda}$) is calculated in the coupling region (green+black atoms in Fig. 1) of phonon modes (index λ) calculated in a dynamical region (black atoms) as described in Ref. [14]. Floating orbitals are included between the STM tip and the graphene sample, to give a better description of the vacuum [16].

Following the lowest order expansion (LOE) [17], simplified and efficient expressions for the IETS signals can be derived under the assumption of weak electron-phonon coupling. The LOE expressions involve just the evaluation of the spectral density matrices for states coming from the tip/sample, $\mathbf{A}_{T/S}(\varepsilon)$, at a discrete number of chemical potentials, $\varepsilon = \mu_T, \mu_S$, corresponding to the threshold voltage bias (V_b) for excitation of a given phonon (λ), $|\mu_T - \mu_S| = \hbar\omega_\lambda$. Thus the LOE expression does not *per se* reflect changes in the DOS above the phonon excitation threshold. However, in the context of STS on gated graphene, this is highly relevant since the DOS behavior leads to a distinct dip in the differential conductance at a specific applied voltage, $V_b = V_D$ [6], enabling a determination of the local chemical potential of graphene.

In order to encompass this important variation in the DOS above threshold we make the following observations (see also Fig. 2). The expressions for the current which gives rise to inelastic signals have a Fermi's golden-rule-like form at low temperature,

$$I_i \approx \frac{e}{\hbar}\Theta\left(e|V_b| - \hbar\omega_\lambda\right) \times \int_{\mu_S \pm \hbar\omega_\lambda}^{\mu_T} d\varepsilon \,\mathrm{Tr}\left[\mathbf{M}_\lambda \tilde{\mathbf{A}}_T(\varepsilon)\mathbf{M}_\lambda \mathbf{A}_S(\varepsilon \mp \hbar\omega_\lambda)\right], \quad (1)$$

*mads.brandbyge@nanotech.dtu.dk

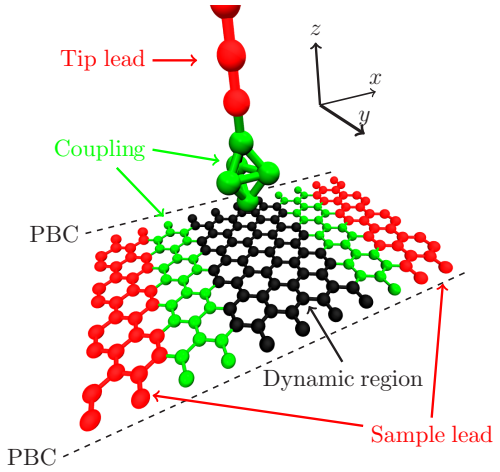


FIG. 1. (Color online) The system setup with semi-infinite leads (red), device region (green), dynamic region (black), and periodic boundary conditions along the dashed lines.

for $V_b \gtrless 0$, where Θ is the Heaviside step function and $\tilde{\mathbf{A}}_T$ is the time reversed $\mathbf{A}_T$. Above threshold ($|V_b| > \hbar\omega_\lambda$) the step behavior is unimportant and we are left with the bias behavior of the integral. For finite bias the states in the device, that is, the spectral functions, change with V_b. However, in this STM setup, the device is strongly coupled to the sample lead and very weakly coupled to the tip lead. Consequently, the potential in the device is pinned to that of the sample lead, which is the Fermi level ε_F of graphene. The DOS of the gold STM probe varies slowly with regard to energy. Thus the *only* important voltage dependent term in Eq. (1) is the tip chemical potential defined by $\mu_T = \varepsilon_F + eV_b$, with $\mu_S = \varepsilon_F$, yielding the differential conductance expression

$$\partial_{V_b} I_i \approx \gamma_\lambda \partial_{V_b} \mathcal{I}^{\mathrm{sym}}, \tag{2}$$

where

$$\gamma_\lambda = \mathrm{Tr}\left[\mathbf{M}_\lambda \tilde{\mathbf{A}}_T(eV_b + \varepsilon_F)\mathbf{M}_\lambda \mathbf{A}_S(eV_b + \varepsilon_F \mp \hbar\omega_\lambda)\right], \tag{3}$$

for $V_b \gtrless 0$, and $\mathcal{I}^{\mathrm{sym}}$ is a temperature broadened version of the step function in Eq. (1) [14]. Equation (2) is equivalent to the usual LOE expression but valid above threshold due to the constant tip DOS. The same argument can be applied to the

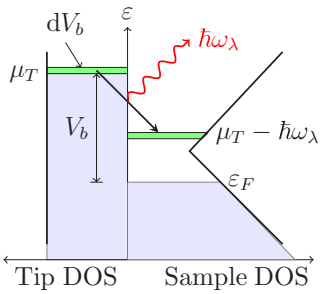


FIG. 2. (Color online) Energy diagram showing the important (green) contributions of the tip/sample DOS when calculating the differential conductance for $V_b > \hbar\omega_\lambda$.

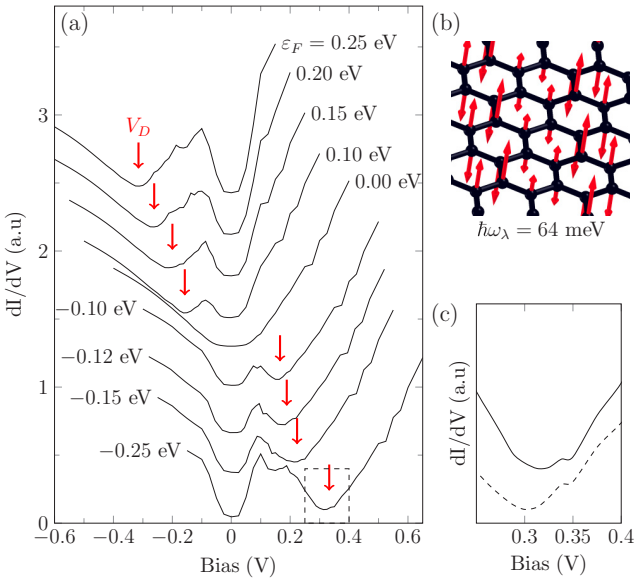


FIG. 3. (Color online) (a) Calculated STS spectra of pristine graphene at different Fermi levels. (b) Out-of-plane acoustic graphene phonon. (c) Closeup of the dip at V_D for $\varepsilon_F = -0.25$ eV, when including only phonons in the 0–400 meV (solid) and 60–70 meV (dashed) energy range.

other terms in LOE. See supplemental material [18] for details of the calculation used, to obtain STS spectra.

Results—pristine graphene. Calculated STS spectra on pristine graphene for a number of different applied gate voltages (Fermi levels) are shown in Fig. 3(a). The gap feature around $V_b = 0$ of width 0.13 V is reproduced in detail and the dip at V_D, caused by inelastic tunneling into the charge neutrality point of graphene, appears outside the gap, as seen in experiments [6–8]. As the gate is applied, V_D moves across the spectrum, changing polarity, while the position and width of the gap feature is stable.

Most major steps in differential conductance come from acoustic out-of-plane phonons at energies just below 67 meV [19]. In particular, the mode shown in Fig. 3(b) gives a large contribution. However, we find that acoustic out-of-plane graphene phonons with energies as low as 42 meV give considerable contributions as well. We also find important inelastic signals from optical graphene phonons at energies above 67 meV. The additional features away from 67 mV make up about half the signal, and have not been included in previous studies [11]. If we restrict our calculations to phonons in the 60–70 meV range, we obtain a 15 mV change in V_D [see Fig. 3(c)] and changes in both the width and height of the inelastic gap. The change in V_D is caused mostly by the experimentally observed [8] inelastic signal near 150 mV, coming from the optical in-plane modes, and occurs for $|V_D| > 150$ mV. In STS experiments V_D is used to extract the energy position of the charge neutrality point from $E_D = e|V_D| - \hbar\omega_0$, where $\hbar\omega_0 = 63$ meV is half the width of the gap feature which corresponds to the energy of an acoustic out-of-plane graphene phonon [6,7]. The change in V_D could explain why all points with $|E_D| < 100$ meV in the E_D versus gate voltage plot of Ref. [7] fall below the fitted line. The local

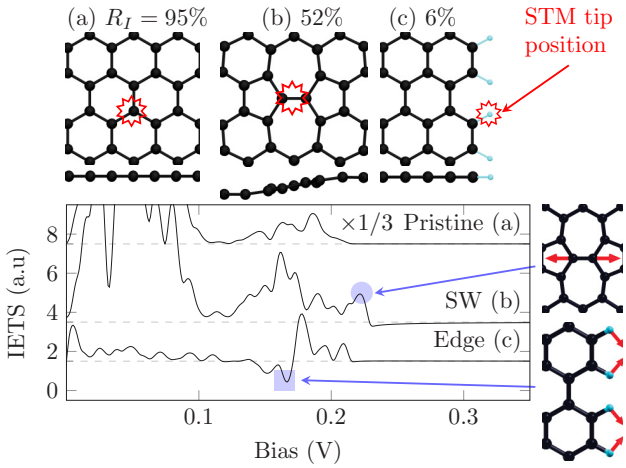


FIG. 4. (Color online) IETS as a function of bias for pristine graphene, a Stone-Wales defect, and a hydrogen passivated armchair edge (geometries shown above the plot). The blue marker indicates characteristic signals. The fraction of differential conductance coming from the inelastic channel (R_I) is shown above the geometries.

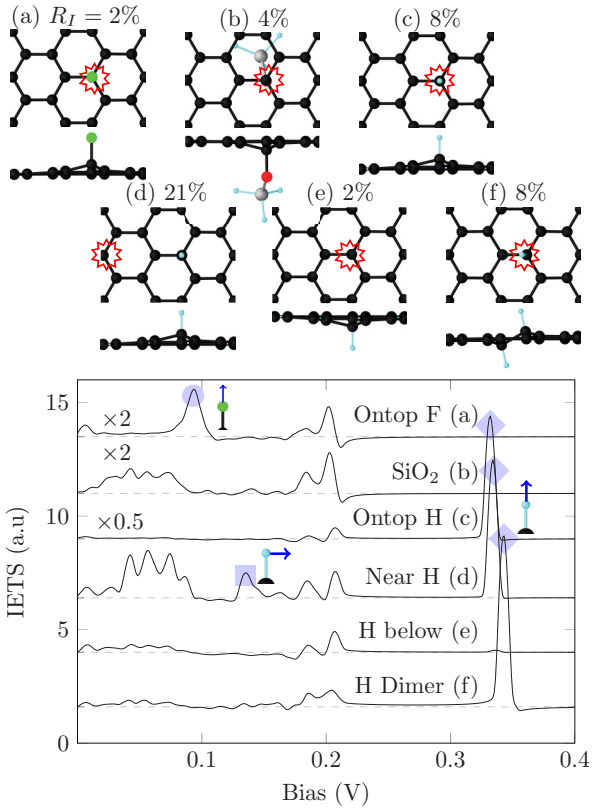


FIG. 5. (Color online) IETS as a function of bias voltage, for various adsorbates on graphene (geometries shown above the plot). Fingerprints for each adsorbate are marked and the phonon shown. The fraction of differential conductance coming from the inelastic channel (R_I) is shown above the geometries.

charge-carrier density (n) of graphene is also extracted from V_D in STS experiments [7,9,10]. Mistaking $E_D = 100$ meV for $E_D = 115$ meV results in a 32% error in n. To capture these experimental details one must include several phonons, and account for their impact in an *ab initio* manner.

Encouraged by the agreement for pristine graphene, we next predict the inelastic signals from various defects to shed light on what information can be obtained from STM-IETS.

Results—structurally defected graphene. In Fig. 4 we show the calculated IETS spectra from an on-top position in pristine graphene [Fig. 4(a)], directly above a Stone-Wales defect (SW) [Fig. 4(b)], and above a passivated armchair edge [Fig. 4(c)]. The result shown for pristine graphene is the same at hollow sites and bridge sites. The gap feature shows up as giant peaks in the IETS below 100 mV for pristine graphene. The signals from in-plane graphene phonons are also seen at 150 mV and above. We find that the low voltage IETS above a SW are very similar to that of pristine graphene. Gap like features have also been observed experimentally for regions with heptagon-pentagon defects [20]. However, a characteristic signal can be seen at $V_b = 223$ mV bias, above any of the pristine graphene phonon bands which can be traced to the high-frequency stretch mode localized at the twisted C-C bond shown in Fig. 4.

Ignoring the out-of-plane buckling introduced to the graphene sheet near a SW, and calculating the IETS for a flat SW system, leads to a 5 mV blueshift of the signal from the twisted C-C bond, as previously proposed [21]. Furthermore, signals from in-plane modes at 150 mV in Fig. 4(b) generally become weaker in the flat system, indicating a weaker coupling of these modes to the out-of-plane current. We also see strong signals at low bias. These signals are caused by low-frequency sinelike out-of-plane modes. These modes couple strongly to the current because they break the mirror symmetry across the twisted C-C bond. In the buckled system, this symmetry is inherently broken, leading to an increase in elastic tunneling. Measuring strong low bias inelastic signals and a 228 mV signal above a SW therefore indicates that it is in a metastable flat configuration, whereas increased elastic transmission and a 223 mV signal is a sign of local buckling. Above a passivated armchair edge, a dip in IETS is seen at $V_b = 168$ mV, which is caused by a collective transverse mode of the hydrogen atoms shown in Fig. 4. Changing the mass of the passivating agent to that of fluor, we observe a corresponding change in the position of the inelastic signals. This indicates that IETS can be used to obtain knowledge of graphene edge passivation.

Results—adsorbates on graphene. In Fig. 5 we show IETS spectra from a range of different covalently bonded impurities. For all systems, delocalized in-plane graphene modes cause signals around 200 mV. In Fig. 5(a) a clear inelastic signal from the longitudinal mode of a fluor adsorbate is seen at 95 mV. Above a hydrogen adsorbate, we see a strong inelastic peak at 332 mV caused by the stretch mode of the C-H bond [see Fig. 5(c)]. This signal serves as a fingerprint for a hydrogen impurity above the graphene sheet as opposed to below where the signal disappears, as can be seen in Fig. 5(e). The corresponding STS spectra show a strong zero-energy peak [22]; this behavior is, however, expected for all covalently bonded impurities [23], above or below the sheet, and therefore cannot be used as a fingerprint.

The STS spectra on the hydrogenated system with the probe above a carbon atom 4.25 Å laterally away from the

impurity in Fig. 5(d) show additional signals. The graphene out-of-plane phonon signals reappear and a signal is also seen at 134 mV, caused by a transverse mode of the C-H bond. Above a graphanelike hydrogen dimer [Fig. 5(f)] the signal caused by the C-H bond stretch mode is seen, however, here it is caused by two degenerate modes and blueshifted by 11–16 mV, indicating a lower energy configuration.

Common for all the imperfect systems is that the gap seen in pristine graphene is quenched and considerable elastic tunneling is seen, as indicated by the severe reduction of the inelastic conductance ratio (R_I) in Figs. 4 and 5. Out-of-plane corrugations in the graphene sheet can lift the suppression of elastic tunneling if they are on the same length scale as the graphene lattice constant [11]. Our results indicate that defects can also lift the suppression locally. This is because the selection rules causing the suppression in pristine graphene are a result of the translational symmetry of the crystal lattice. When this symmetry is broken, the suppression is lifted and the elastic tunneling dominates. The expected order of magnitude change in tunneling conductance should lead to bright spots in STM topographies. In the case of granular chemical vapor deposited (CVD) graphene, protruding grain boundaries are often attributed to localized electronic states [20]. Our results here point out that one may expect increased tunneling near disordered areas of graphene, even if no localized electronic states are present and the area is completely flat. As seen in Figs. 5(b) and 5(e), this is also the case for strong interactions with a SiO_2 substrate or hydrogen sitting below the graphene sheet, which should therefore be visible as protruding from the graphene sheet.

In summary, we have presented a first principles method and used it for calculations of IETS and STS spectra of pristine and defected graphene. We showed how measured STS spectra on pristine gated graphene can be reproduced in detail as a function of gating. The inclusion of several phonons had a strong impact on all aspects of the STS spectrum of pristine graphene. In particular, we found that including optical in-plane phonons changed the V_D value for certain gate voltages. This is of importance for studies where IETS is used to probe the local doping of graphene [7,9,10] where it may lead to a significant overestimation of the local charge inhomogeneity. We predicted the IETS of typical imperfections in graphene, and demonstrated how these can yield characteristic fingerprints revealing, e.g., adsorbate species or local buckling. Additional elastic contributions above defects should make them protrude in STM regardless of the actual geometric or electronic structure, and care is needed when interpreting STM images.

We gratefully acknowledge discussions with Thomas Frederiksen, Aran Garcia-Lekue, and Rasmus Bjerregaard Christensen. The Center for Nanostructured Graphene (CNG) is sponsored by the Danish Research Foundation, Project DNRF58.

[1] A. Geim and K. S. Novoselov, Nat. Mater. **6**, 183 (2007).
[2] A. H. C. Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, Rev. Mod. Phys. **81**, 109 (2009).
[3] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, Science **306**, 666 (2004).
[4] A. C. Ferrari, J. C. Meyer, V. Scardaci, C. Casiraghi, M. Lazzeri, F. Mauri, S. Piscanec, D. Jiang, K. S. Novoselov, S. Roth, and A. K. Geim, Phys. Rev. Lett. **97**, 187401 (2006).
[5] J. G. Pedersen, T. Gunst, T. Markussen, and T. G. Pedersen, Phys. Rev. B **86**, 245410 (2012).
[6] Y. Zhang, V. W. Brar, F. Wang, C. Girit, Y. Yayon, M. Panlasigui, A. Zettl, and M. Crommie, Nat. Phys. **4**, 627 (2008).
[7] R. Decker, Y. Wang, V. W. Brar, W. Regan, H. Z. Tsai, Q. Wu, W. Gannett, A. Zettl, and M. F. Crommie, Nano Lett. **11** (2011).
[8] V. W. Brar, S. Wickenburg, M. Panlasigui, C.-H. Park, T. O. Wehling, Y. Zhang, R. Decker, C. Girit, A. V. Balatsky, S. G. Louie, A. Zettl, and M. F. Crommie, Phys. Rev. Lett. **104**, 036805 (2010).
[9] P. Cao, J. O. Varghese, K. Xu, and J. R. Heath, Nano Lett. **12**, 1459 (2012).
[10] S. J. Goncher, L. Zhao, A. N. Pasupathy, and G. W. Flynn, Nano Lett. **13**, 1386 (2013).
[11] T. O. Wehling, I. Grigorenko, A. I. Lichtenstein, and A. V. Balatsky, Phys. Rev. Lett. **101**, 216803 (2008).
[12] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, J. Phys.: Condens. Matter **14**, 2745 (2002).
[13] M. Brandbyge, J. L. Mozos, P. Ordejón, J. Taylor, and K. Stokbro, Phys. Rev. B **65**, 165401 (2002).
[14] T. Frederiksen, M. Paulsson, M. Brandbyge, and A. P. Jauho, Phys. Rev. B **75**, 205413 (2007).
[15] We use a split polarized double zeta (DZP) basis set, a mesh cutoff of 200 Ry, a Monkhorst-Pack k-point mesh of $1 \times 2 \times 1$, and the local density approximation (LDA) exchange-correlation (xc) functional [24] to calculate the electronic structure. A supercell dimension (k_y points) of 27 Å × 12.8 Å(101)/34.5 Å × 25.5 Å(Γ)/49.3 Å × 12.8 Å(Γ)/27 Å × 17 Å(Γ) is used for inelastic transport for the pristine/SW/edge/adsorbate configuration.
[16] A. Garcia-Lekue and L. W. Wang, Phys. Rev. B **82**, 035410 (2010).
[17] J. T. Lü, R. B. Christensen, G. Foti, T. Frederiksen, T. Gunst, and M. Brandbyge, Phys. Rev. B **89**, 081405 (2014).
[18] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.91.121403 for convergence tests and details of the STS calculation including above excitation threshold DOS effects.
[19] M. Mohr, J. Maultzsch, E. Dobardzic, S. Reich, I. Milosevic, M. Damnjanovic, A. Bosak, M. Krisch, and C. Thomsen, Phys. Rev. B **76**, 035439 (2007).

[20] L. Zhao, M. Levendorf, S. Goncher, T. Schiros, L. Paova, A. Zabet-Khosousi, K. T. Rim, C. Gutierez, D. Nordlund, C. Jaye, M. Hybertsen, D. Reichman, G. W. Flynn, J. Park, and A. N. Pasupathy, Nano Lett. **13**, 4659 (2013).

[21] J. Ma, D. Alfe, A. Michaelides, and E. Wang, Phys. Rev. B **80**, 033407 (2009).

[22] M. Scheffler, D. Haberer, L. Petaccia, M. Farjam, R. Schlegel, D. Baumann, T. Hänke, A. Gruneis, M. Knupfer, C. Hess, and B. Buchner, ACS Nano **6**, 10590 (2012).

[23] T. O. Wehling, M. I. Katsnelson, and A. I. Lichtenstein, Phys. Rev. B **80**, 085428 (2009).

[24] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

Giant tunnel-electron injection in nitrogen-doped graphene

Jérôme Lagoute,[1,*] Frédéric Joucken,[2] Vincent Repain,[1] Yann Tison,[1,†] Cyril Chacon,[1] Amandine Bellec,[1] Yann Girard,[1] Robert Sporken,[2] Edward H. Conrad,[3] François Ducastelle,[4] Mattias Palsgaard,[5] Nick Papior Andersen,[5] Mads Brandbyge,[5] and Sylvie Rousset[1]

[1]*Laboratoire Matériaux et Phénomènes Quantiques, Université Paris Diderot-Paris 7, Sorbonne Paris Cité, CNRS, UMR 7162, 10, Rue A. Domon et L. Duquet, 75205 Paris 13, France*

[2]*Research Center in Physics of Matter and Radiation (PMR), Université de Namur, 61 Rue de Bruxelles, 5000 Namur, Belgium*

[3]*The Georgia Institute of Technology, Atlanta, Georgia 30332-0430, USA*

[4]*Laboratoire d'Etude des Microstructures, ONERA-CNRS, BP 72, 92322 Châtillon Cedex, France*

[5]*Center for Nanostructured Graphene, Department of Micro- and Nanotechnology, Technical University of Denmark, Ørsteds Plads, Bldg. 345E, DK-2800 Kongens Lyngby, Denmark*

(Received 12 November 2014; revised manuscript received 3 March 2015; published 31 March 2015)

Scanning tunneling microscopy experiments have been performed to measure the local electron injection in nitrogen-doped graphene on SiC(000$\bar{1}$) and were successfully compared to *ab initio* calculations. In graphene, a gaplike feature is measured around the Fermi level due to a phonon-mediated tunneling channel. At nitrogen sites, this feature vanishes due to an increase of the elastic channel that is allowed because of symmetry breaking induced by the nitrogen atoms. A large conductance enhancement by a factor of up to 500 was measured at the Fermi level by comparing local spectroscopy at nitrogen sites and at carbon sites. Nitrogen doping can therefore be proposed as a way to improve tunnel-electron injection in graphene.

DOI: 10.1103/PhysRevB.91.125442 PACS number(s): 73.22.Pr, 63.22.Rc, 68.37.Ef

The exploitation of the electronic properties of graphene allows us to envision the development of new electronics based on carbon materials [1–3]. Due to the particular band structure of graphene, low-energy electronic states, driving the current in transport devices, are only available at the large parallel momentum ($k_{\parallel}$) K and K' points in graphene. As a consequence, perpendicular tunneling of electrons into a graphene sheet is quenched at low bias voltage due to the competing large $k_{\parallel}$ momentum conservation and the exponentially decaying tunneling probability with increasing $k_{\parallel}$. Scanning tunneling microscopy (STM) experiments on graphene on SiO_2 have revealed that above 63 mV an inelastic channel corresponding to the excitation of an out-of-plane phonon is opened, which enhances substantially the tunneling current [4]. A gaplike feature in the dI/dV spectroscopy is therefore observed around the Fermi level E_F [4–8]. Interestingly, the same feature has been measured in spin devices using magnetic tunnel junctions [9] where magnetoresistance measurements were performed at bias voltages larger than the gap feature. Using a large bias in such devices can be a limitation as the magnetoresistance signal decreases with the bias voltage. Therefore, enhancing the electron injection at low energy in graphene turns out to be a cornerstone for improving the performances of graphene-based electronic and spintronic devices. As the quenching of electron injection is due to the momentum conservation of electrons imposed by the symmetry of the graphene sheet, breaking the symmetry of the system can lead to an enhancement of electron injection. In that respect, introducing atomic defects is a promising strategy to restore an elastic channel for vertical injection in graphene. The nitrogen doping of graphene obtained by substitution of nitrogen atoms for carbon atoms is a suitable route to achieve well-controlled and well-characterized point defects with limited atomic relaxation [6–8] while preserving the band structure of graphene. At the atomic level, nitrogen doping induces a redistribution of the electron density on one sublattice and a localized resonance at nitrogen sites [8] that have been widely studied theoretically [10]. In this paper, we perform scanning tunneling spectroscopy (STS) experiments to study the local electron injection in nitrogen-doped graphene on SiC(000$\bar{1}$). dI/dV and tunneling current decay length spectra reveal that below the out-of-plane phonon energy, elastic tunneling is allowed at nitrogen sites leading to a large enhancement of the electron injection by a factor of up to 500 close to zero bias. This interpretation is supported by first-principle calculations and well explained by a simple analytical model. Our findings allow us to propose nitrogen doping of graphene as an efficient way to improve electron injection in graphene without altering its band structure.

The graphene sample was grown by the confinement control sublimation (CCS) method on SiC(000$\bar{1}$) [11]. It consists of about five non-Bernal stacked layers on top of SiC. The postsynthesis doping was achieved in an ultrahigh-vacuum (UHV) chamber by exposing the sample during 30 mn to a nitrogen radical flux produced by a remote (∼30 cm) RF plasma source (MPD21 from Oxford Applied Research [12]) fed with N_2 (purity 99.999%) [8]. The scanning tunneling experiments were performed with a UHV Low-Temperature STM (4.2 K) from Omicron GmbH using electrochemically etched tungsten tips. dI/dV spectra were acquired with a lock-in detector at 710 Hz and a modulation amplitude of 24 mV. Prior to spectroscopic measurement on graphene the STM tip was calibrated on a Au(111) substrate by applying voltage pulses until the dI/dV spectrum shows the onset of the Shockley surface state. This procedure is necessary to obtain reliable spectra as it was mentioned by Zhang *et al.* [4].

*jerome.lagoute@univ-paris-diderot.fr

†Present address: Université de Pau et des Pays de l'Adour, IPREM - ECP CNRS UMR 5254, technopole hélioparc, 2 av. du Président Angot, 64053 Pau Cedex 09, France.

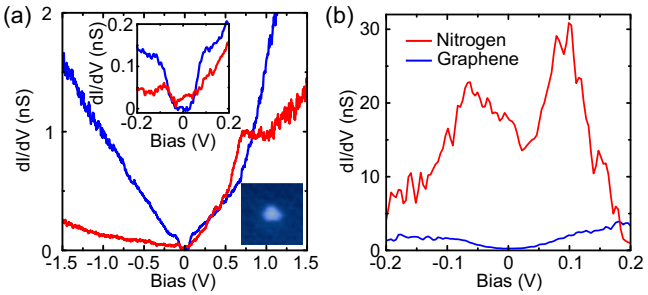


FIG. 1. (Color online) (a) dI/dV spectra measured above graphene far from the nitrogen atoms (blue) and above a nitrogen atom (red). The spectrum in inset is a zoom at low bias voltages of the spectra. The image in inset is a 2×2 nm^2 image of the nitrogen atom where the measurement was performed. (b) dI/dV spectra around E_F on carbon (blue) and nitrogen (red) sites measured in constant height mode conditions prior to the spectroscopy measurement (110 pA at 100 mV above graphene carbon area).

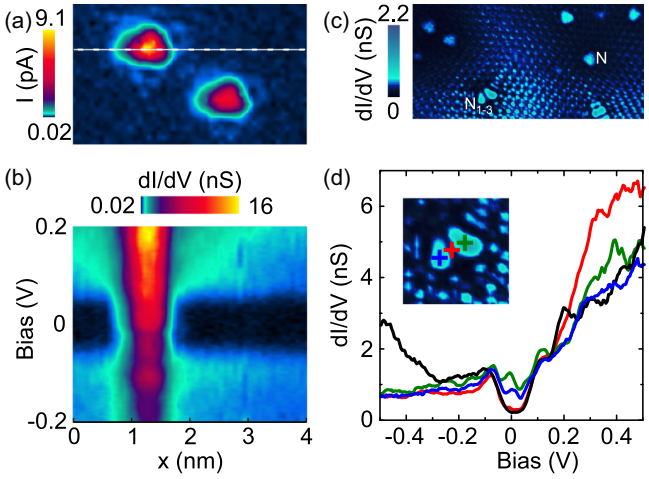


FIG. 2. (Color online) (a) Current image measured at 1.3 mV in constant height mode (4×2.4 nm^2, 12 pA at 100 mV above graphene carbon area) on which STS measurements have been performed at each point. (b) Two-dimensional (2D) map representing the spectra measured along the line shown in (a). (c) Conductance map measured in constant current mode at 2.5 mV (12×6 nm^2). The marks N and N_{1-3} indicate one single substitutional nitrogen atom, and a nitrogen pair with two nitrogens in second neighbor position. (d) Spectra measured on a pair of nitrogen atoms in second neighbor positions. The black curve was measured on the graphene area far from nitrogen atoms, the color curves were measured at the positions marked by the cross in the inset that is a zoom of (c) corresponding to the nitrogen pair.

After synthesis and doping, the sample was transported in the atmosphere and outgassed in UHV at ~800 °C before the STM measurements.

Nitrogen atoms substituted to carbon atoms (graphitic nitrogen) in graphene on SiC(000$\bar{1}$) appear in the STM images as bright spots with a triangular shape [bottom right inset in Fig. 1(a)]. This can be attributed to a combination of charge transfer between the nitrogen and the three neighboring carbon atoms [8] and the opening of the elastic channel as shown in the following. The dI/dV spectrum in Fig. 1(a) measured on graphene far from the nitrogen atoms reveals a gaplike feature of 130 mV around E_F. This feature has been attributed to the inelastic excitation of an out-of-plane phonon [4]. Indeed, momentum conservation in the tunneling process imposes a large in-plane momentum $k_\parallel$ for electrons tunneling to graphene where the only available states are at the K points. However, above a threshold voltage corresponding to the energy of an out-of-plane graphene phonon, an inelastic tunneling channel is opened, which leads to a large increase of the tunneling current. In the dI/dV spectrum this inelastic excitation appears as two onsets symmetric with respect to E_F at the energy of the phonon. The spectrum measured above a nitrogen atom is totally different [Fig. 1(a)]. The inelastic signal almost disappears, only fainted onsets can be seen [top inset in Fig. 1(a)]. This is indicative of a strong decrease of the inelastic/elastic transmission ratio. A quantitative comparison of the conductance at nitrogen and carbon sites cannot be obtained from the curve in Fig. 1(a) as these spectra were measured with a constant current condition before measurement ($U = 1V, I = 500$ pA). Indeed the nitrogen atoms lie in the graphene plane as it was suggested by voltage-dependent STM measurements [8] and confirmed by *ab initio* calculations on monolayer [13] and bilayer [14,15] graphene. Electronic effects lead to an apparent height of about 1 Å, the tip-sample distance is larger above the nitrogen atom than above graphene. In order to quantitatively compare the conductance at carbon and nitrogen sites, spectra have to be measured with the same tip-sample distance (constant height mode). Such data are displayed in Fig. 1(b). The measurements obtained with the same tip-sample separation reveal a huge increase of dI/dV above nitrogen. At E_F the conductance at the Fermi level above a nitrogen atom is 70 times larger than on the graphene sheet revealing a very large enhancement of electron injection at nitrogen sites.

The increase of conductance above nitrogen atoms was systematically observed as shown in Fig. 2. Figure 2(a) displays the current image measured in constant height mode on an area with two nitrogen atoms. Figure 2(b) shows the evolution of the spectra along a line crossing a nitrogen atom [cf. dotted line of Fig. 2(a)] revealing the increase of the dI/dV signal above nitrogen over a distance characterized by a full width at half maximum of 5 Å. This extent corresponds to the typical size of the triangular pattern associated with a nitrogen atom in the STM images. Such a spatial extension also corresponds to the variation of the potential around a nitrogen atom [10]. The ratio between the conductance at the Fermi level above graphene (0.017 nS) and above nitrogen (9.15 nS) in Fig. 2 is about 500, which is larger than the above-mentioned measure (ratio of 70). As will be discussed below, we attribute this difference to the tip-sample distance that is larger in the second case, which increases the ratio of conductance between nitrogen and graphene. In Fig. 2(c) we show a conductance map measured in constant current mode at 2.5 mV. Although this mode reduces the contrast, it allows us to obtain more details along the whole image. This image shows again that the conductance is stronger above the nitrogen atoms confirming that the phonon gap feature vanishes above nitrogen as previously observed in the spectrum of Fig. 1(a). Interestingly, on a nitrogen pair identified as two nitrogen

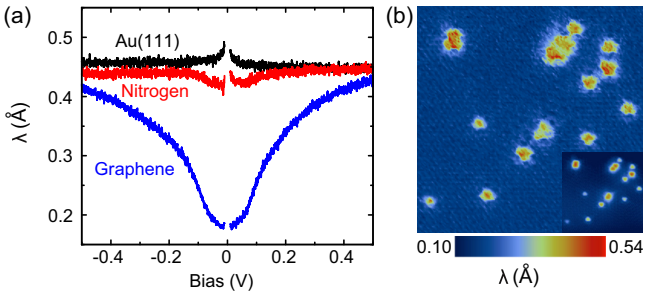


FIG. 3. (Color online) (a) Voltage dependent current decay length λ measured on a reference Au(111) sample (black), N-doped graphene carbon area (blue) and above a graphitic nitrogen atom in graphene (red). (b) λ mapping at 20 mV extracted from spectra measured at each point of the topographic image shown in the inset (10×10 nm^2, $U = 0.1$ V, $I = 500$ pA).

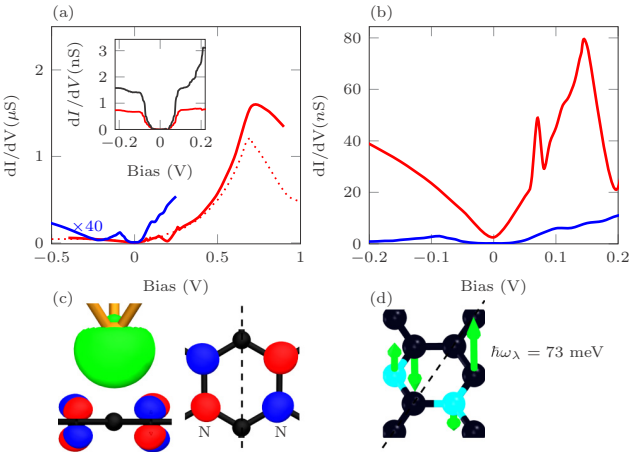


FIG. 4. (Color online) (a) Calculated dI/dV spectra above pristine graphene (blue) and a nitrogen (red) with (solid) and without (dotted) electron-phonon coupling. The spectrum in inset is a close-up of the inelastic conductance steps near E_F on a single nitrogen (red) and at the center of a nitrogen dimer (black) in graphene. (b) Calculated constant height dI/dV spectra around E_F on pristine graphene (blue) and a nitrogen (red) sites. (c) Resonant scattering state of the nitrogen dimer blue/red indicates positive/negative isovalues. (d) Out-of-plane phonon that mediates tunneling into the scattering state of (c).

atoms in second neighbor positions, a strong contrast with atomic resolution is clearly observed in the dI/dV map. The spectra in Fig. 2(d) measured on the nitrogen atoms and between the nitrogen atoms of this pair clearly reveal that although the inelastic feature is almost suppressed above the nitrogen atoms, it is unaffected at the central position between the N atoms. This is in line with previously reported data showing that the inelastic feature is observed at the center of a nitrogen pair [7].

Details on the tunneling mechanism at nitrogen sites can be obtained by measuring the voltage-dependent current decay length λ as discussed for undoped graphene on SiO_2 [4]. In that case, the decay length inside and outside the phonon gap has been measured to be 0.25 Å and 0.5 Å respectively [4] by fitting, at different bias voltages, the $I(z)$ curves with the function $I = \exp(-z/\lambda)$. Here we performed a continuous voltage-dependent measure of λ using a lock-in technique. A sinusoidal modulation at 715 Hz with an amplitude of 0.06 Å was applied to the piezo z signal. A lock-in detection was used to record the corresponding current response allowing to measure the dI/dz signal. The measures were performed with a closed feedback loop in order to keep the current constant during the spectrum and obtain a signal directly proportional to $1/\lambda$. After dividing by the mean value of the current we obtain the $\lambda(V)$ curve. The resulting spectra are shown in Fig. 3(a). The $\lambda(V)$ curve measured above graphene looks similar to previously reported $\lambda(V)$ curves obtained by fitting I versus z measurements [4]. We obtain λ values ∼0.18 Å inside the phonon gap and 0.42 Å outside the phonon gap. As a reference we have measured the $\lambda(V)$ curve on a Au(111) with the same tip just before measuring on the graphene sample. The corresponding curve shows the expected value of 0.45 Å for λ independent of the voltage. Above a nitrogen atom, the $\lambda(V)$ curve obtained is in strong contrast with the one measured on graphene. The $\lambda(V)$ curve above nitrogen is similar to the one measured on Au(111) suggesting that the transport is dominated by elastic tunneling. A slight phonon feature is nevertheless observable with a much weaker amplitude (0.42 Å in the gap, 0.44 Å outside the gap). The use of the lock-in technique allows us to perform a systematic measure of local $\lambda(V)$ spectra at each point of an image and to extract a λ mapping at any bias voltage. An example of such an image is displayed in Fig. 3(b) showing a λ mapping at 20 mV measured simultaneously with the image shown in the inset. This mapping shows a uniform value of λ on the carbon areas of around ∼0.2 Å and a clear larger value on nitrogen atoms of ∼0.4 Å. A consequence of the different decreasing lengths at nitrogen and carbon sites is that the conductance ratio between these sites is expected to be distance dependent. Indeed this ratio is expected to increase with the tip-sample distance z according to the quantity $\exp[z(1/\lambda_C - 1/\lambda_N)]$ where λ_C and λ_N are the decreasing lengths at E_F above carbon and nitrogen sites respectively. This allows us to understand the different dI/dV ratios measured to be 70 in Fig. 1(b) and 500 in Fig. 2(b). In these measurements, the current above the graphene area at a reference bias of 100 mV is $I_1 = 110$ pA in the Fig. 1(b) and $I_2 = 12$ pA in the Fig. 2(b). This means that the tip-sample distance in the two measurements varies by $\Delta z(V) = \lambda_C(V)\ln(\frac{I_1}{I_2}) = 0.55$ Å [using $\lambda_C(V) = 0.25$ Å for $V = 100$ mV]. The ratios of conductance are therefore expected to vary by a factor of 5.7 (with $\lambda_C = 0.18$ Å and $\lambda_N = 0.42$ Å at the Fermi level), which is close to the ratio 500/70.

Using the DFT-NEGF method and setup described in Ref. [16],[1] dI/dV spectra including vibrational effects were calculated directly above a nitrogen dopant in graphene and above pristine graphene. For all calculations shown we use a tip-to-sample distance of $d = 5$ Å since tunneling at larger

[1]We use a split DZP basis set, a mesh cutoff of 200 Ry, a Monkhorst-Pack k-point mesh of $1 \times 2 \times 1$ and the Ceperley-Alder LDA functional [22] to calculate the electronic structure. Supercell dimension and k_y of 27×12.8 Å(101)/27×17 Å(81) is used to calculate inelastic transport for pristine/nitrogen doped graphene.

distances is not described well by the LCAO basis set used [17]. To simulate single nitrogen sites, the nitrogen atoms were separated by 1.7 nm. To test that this separation was sufficient the STS spectrum was calculated exactly between adjacent nitrogen pairs (8.5 Å away from either site) and seen to coincide with that of pristine graphene. The dI/dV spectra on a nitrogen and pristine graphene are shown in Fig. 4(a). The dotted line shows the equivalent spectrum above nitrogen without electron-phonon coupling. Note that the spectrum on graphene has been enhanced to allow a better comparison with the experimental data of Fig. 1(a) where the constant current imaging condition brings the intensity of the signal above graphene close to the intensity of the signal above nitrogen. For pristine graphene, the spectrum is dominated by the inelastic features. As discussed previously [16] the 130 meV gap is reproduced, and the dip caused by inelastic tunneling into the Dirac point is seen at the expected bias value. As in the measured dI/dV spectra of Fig. 1(a) a broad resonance emerges at $V_b = 700$ meV above nitrogen, and the inelastic features near the Fermi level are not seen. Comparing the dI/dV spectra above nitrogen with and without electron-phonon coupling, it is clear that the elastic channel dominates, and the inclusion of vibrations only leads to a stronger resonance slightly shifted in bias. In Fig. 4(b) the calculated dI/dV spectra around E_F are shown on the same scale. Comparing these spectra with the constant height measurement of Fig. 1(b), both the conductance magnitudes and their ratio suggest that $d > 5$ Å in the experiment as expected.

The inelastic steps in dI/dV excluding elastic contributions on nitrogen and at the center of a nitrogen dimer are investigated using the lowest-order expansion [18]. This method was previously found to give a good description of the gap feature of pristine graphene.[16] The results are shown in the inset of Fig. 4(a). The inelastic steps above a single nitrogen are of similar magnitude to those of pristine graphene, but the elastic contribution is an order of magnitude larger above nitrogen making them nearly invisible in measurements. At the center of the nitrogen dimer the inelastic feature is about three times larger. This additional inelastic feature is a result of the mirror symmetry along armchair direction between the two nitrogen sites. The resonant scattering state of the dimer system shown in Fig. 4(c) has odd symmetry at the dimer center and consequently does not couple elastically to the *s*-like state of the STM-tip. In the presence of the 73 meV out-of-plane phonon shown in Fig. 4(d) the mirror symmetry is broken and a large inelastic step in conductance is seen in very good agreement with the experimental data of Fig. 2(d). Therefore, the DFT-NEGF results support the conclusion based on the measurements that the opening of an elastic channel above nitrogen substitutions leads to a giant increase in tunneling conductance.

The enhancement of the elastic tunnel electron injection at nitrogen sites can be rationalized further in terms of a simple model. In the presence of disorder, k conservation is broken and an elastic tunneling channel is opened. For the case of N doping the Hamiltonian of the system now includes a potential V localized on the impurity sites that produces a resonance about 0.5 eV above the Dirac point in the bulk π states [10]. For simplicity we will assume that the free-electron states can be described locally in terms of atomiclike functions. One of them has a π-like character (state of symmetry A_{2u} [19] also denoted 1^- in Ref. [20]) and can be coupled to the genuine π states of pristine graphene. Assuming the potential to be localized on the nitrogen atoms, its relevant matrix elements for a single impurity at site 0 only involve the π state at this site $|\pi,0\rangle$ as well as the localized function associated with the A_{2u} state $|A_{2u},0\rangle$. In k space the potential obviously couples all states. In this very localized limit, the main matrix element $\langle\pi,0|V|\pi,0\rangle$ has been estimated to be about 10 eV [10]. Because of the different extensions of the orbitals the other matrix element should be much weaker, allowing us to keep just one off-diagonal term $v = \langle\pi,0|V|A_{2u},0\rangle$. We can now directly adapt the calculation by Wehling *et al.* [21] and calculate the self-energy and the local Green's function corresponding to the new Γ elastic channel. Finally, in a second-order perturbation theory, the local density of states on the nitrogen site is $n_\Gamma(E) \simeq v^2 n_\pi(E)/(E - E_\sigma)^2$ where $n_\pi(E)$ is the corresponding density of states of the π states in the presence of the nitrogen impurities [10] and E_σ is the energy of the free-electron -like state [21].

Close to the Dirac point $E \simeq 0$ and the intensity of the new channel is a fraction $(v/\lambda_{el-ph})^2$ of the inelastic channel, where λ_{el-ph} is the electron phonon coupling strength. Assuming v to be at least equal to λ_{el-ph} (about 0.5 eV [21]) we see that the elastic channel can be at least as efficient as the inelastic channel in agreement with the experiments and the above *ab initio* calculations.

In conclusion, using STM/STS we measured an increase of differential conductance leading to a vanishing of the phonon gaplike feature at nitrogen sites. This leads to a two orders of magnitude increase in electron injection into graphene. First-principles calculations reveal that although the inelastic excitation of graphene phonon still occurs at nitrogen sites, an elastic channel opens that substantially increase the conductance. Therefore we expect that nitrogen doping of graphene can be used to improve tunnel electron injection in graphene-based devices.

This work has been supported by the Labex SEAM Program No. ANR-11-LABX-86 in the framework of the Program No. ANR-11-IDEX-0005-02. The research leading to these results has received funding from the European Union Seventh Framework Programme under Grant Agreement No. 604391 Graphene Flagship. V.R. thanks the Institut Universitaire de France for support. E.H.C. acknowledges support by the National Science Foundation under Grant No. DMR-1005880. We thank Luc Henrard for fruitfull discussions.

[1] C. Berger, Z. Song, T. Li, X. Li, A. Y. Ogbazghi, R. Feng, Z. Dai, A. N. Marchenkov, E. H. Conrad, P. N. First, and W. A. de Heer, J. Phys. Chem. B **108**, 19912 (2004).

[2] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, Science **306**, 666 (2004).

[3] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, Rev. Mod. Phys. **81**, 109 (2009).
[4] Y. Zhang, V. W. Brar, F. Wang, C. Girit, Y. Yayon, M. Panlasigui, A. Zettl, and M. F. Crommie, Nature Phys. **4**, 627 (2008).
[5] R. Decker, Y. Wang, V. W. Brar, W. Regan, H.-Z. Tsai, Q. Wu, W. Gannett, A. Zettl, and M. F. Crommie, Nano Lett. **11**, 2291 (2011).
[6] L. Zhao, R. He, K. T. Rim, T. Schiros, K. S. Kim, H. Zhou, C. Gutiérrez, S. P. Chockalingam, C. J. Arguello, L. Pálová, D. Nordlund, M. S. Hybertsen, D. R. Reichman, T. F. Heinz, P. Kim, A. Pinczuk, G. W. Flynn, and A. N. Pasupathy, Science **333**, 999 (2011).
[7] R. Lv, Q. Li, A. R. Botello-Méndez, T. Hayashi, B. Wang, A. Berkdemir, Q. Hao, A. L. Eléas, R. Cruz-Silva, H. R. Gutiérrez, Y. A. Kim, H. Muramatsu, J. Zhu, M. Endo, H. Terrones, J.-C. Charlier, M. Pan, and M. Terrones, Sci. Rep. **2**, 586 (2012).
[8] F. Joucken, Y. Tison, J. Lagoute, J. Dumont, D. Cabosart, B. Zheng, V. Repain, C. Chacon, Y. Girard, A. R. Botello-Méndez, S. Rousset, R. Sporken, J.-C. Charlier, and L. Henrard, Phys. Rev. B **85**, 161408 (2012).
[9] B. Dlubak, M.-B. Martin, R. S. Weatherup, H. Yang, C. Deranlot, R. Blume, R. Schloegl, A. Fert, A. Anane, S. Hofmann, P. Seneor, and J. Robertson, ACS Nano **6**, 10930 (2012).
[10] P. Lambin, H. Amara, F. Ducastelle, and L. Henrard, Phys. Rev. B **86**, 045448 (2012).
[11] W. A. de Heer, C. Berger, M. Ruan, M. Sprinkle, X. Li, Y. Hu, B. Zhang, J. Hankinson, and E. Conrad, Proc. Natl. Acad. Sci. USA **108**, 16900 (2011).
[12] R. P. Vaudo, Z. Yu, J. W. Cook, and J. F. Schetzina, Opt. Lett. **18**, 1843 (1993).
[13] B. Zheng, P. Hermet, and L. Henrard, ACS Nano **4**, 4165 (2010).
[14] S.-O. Guillaume, B. Zheng, J.-C. Charlier, and L. Henrard, Phys. Rev. B **85**, 035444 (2012).
[15] Luc Henrard (private communication).
[16] M. L. N. Palsgaard, N. P. Andersen, and M. Brandbyge, Phys. Rev. B **91**, 121403(R) (2015).
[17] A. Garcia-Lekue and L. W. Wang, Phys. Rev. B **82**, 035410 (2010).
[18] J.-T. Lü, R. B. Christensen, G. Foti, T. Frederiksen, T. Gunst, and M. Brandbyge, Phys. Rev. B **89**, 081405 (2014).
[19] E. Kogan, V. U. Nazarov, V. M. Silkin, and M. Kaveh, Phys. Rev. B **89**, 165430 (2014).
[20] V. M. Silkin, J. Zhao, F. Guinea, E. V. Chulkov, P. M. Echenique, and H. Petek, Phys. Rev. B **80**, 121408 (2009).
[21] T. O. Wehling, I. Grigorenko, A. I. Lichtenstein, and A. V. Balatsky, Phys. Rev. Lett. **101**, 216803 (2008).
[22] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2016, **18**, 1025

Manipulating the voltage drop in graphene nanojunctions using a gate potential†

Nick Papior,*[ab] Tue Gunst,[ab] Daniele Stradi[ab] and Mads Brandbyge[ab]

Received 4th August 2015,
Accepted 30th November 2015

DOI: 10.1039/c5cp04613k

www.rsc.org/pccp

Graphene is an attractive electrode material to contact nanostructures down to the molecular scale since it can be gated electrostatically. Gating can be used to control the doping and the energy level alignment in the nanojunction, thereby influencing its conductance. Here we investigate the impact of electrostatic gating in nanojunctions between graphene electrodes operating at finite bias. Using quantum transport simulations based on density functional theory, we show that the voltage drop across symmetric junctions changes dramatically and controllably in gated systems compared to non-gated junctions. In particular, for p-type(n-type) carriers the voltage drop is located close to the electrode with positive(negative) polarity, the potential of the junction is pinned to the negative(positive) electrode. We trace this behaviour back to the vanishing density of states of graphene in the proximity of the Dirac point. Due to the electrostatic gating, each electrode exposes different density of states in the bias window between the two different electrode Fermi energies, thereby leading to a non-symmetry in the voltage drop across the device. This selective pinning is found to be independent of device length when carriers are induced either by the gate or dopant atoms, indicating a general effect for electronic circuitry based on graphene electrodes. We envision this could be used to control the spatial distribution of Joule heating in graphene nanostructures, and possibly the chemical reaction rate around high potential gradients.

Introduction

Graphene (Gr) shows great promise as a central material for future two-dimensional (2D) nanoelectronic applications.[1,2] In particular, its semi-metallic character and its record high mean-free path[3] makes it a top candidate for ultra-fast and flexible electronic components.[4,5] Fuelled by these perspectives, nanostructured devices down to the molecular scale using electrodes based on Gr have recently been put forward.[6–9] In their most generic form, these devices are composed by a Gr constriction where the narrowest junction consists of a Gr nanoribbon (GNR)[10,11] or an organic molecule.[6–9] More complex structures such as Gr antidot lattices[12,13] can also be viewed as consisting of a network of constrictions.

A unique feature of Gr electrodes is that their electronic properties can easily be tuned by electrostatic gating. In fact, electrostatic gates can be used to increase the carrier density in Gr up to above 10^{13} cm^{-2}.[14] For ion gating it has even been possible to reach carrier densities of 10^{14} cm^{-2} which correspond to a Fermi energy shift of about 1 eV.[15] It has been shown that gating can be used to tune the resonances localized in the narrowest part of the junction,[16,17] as the electronic states of the electrodes are usually affected only weakly by the gate-induced capacitive field due to effective screening by the high density of states (DOS).[18] However, for Gr electrodes, the lower DOS and its flat geometry, makes it comparable to the junction itself, and is thus likely to be perturbed similarly by gating. This peculiarity leads to a novel, yet largely unexplored, paradigm for graphene-based electronics, as the transport characteristics of the device ultimately depend on the response of the entire system to the gate. In electronic transport simulations, the effect of electrostatic gating and induced doping charge in the device has often been mimicked by rigidly shifting the position the Fermi-level/chemical potential in calculations without explicitly including the gate or dopants.[19–22] However, despite accounting for some of the effects, these approaches completely neglect the self-consistent response of the device to the additional charge doping or the gate-induced electric field.

Here, we investigate these issues by extending the TranSIESTA electronic transport package, based on density functional theory and nonequilibrium Green function (DFT–NEGF),[23,24] with the inclusion of a physically motivated gate model, see Fig. 1a. This improvement allows us to consider on an equal footing the effect of charge doping, capacitive gate field, and of the finite bias voltage in our DFT + NEGF simulations (see the Methods section

[a] *Department of Nanotech, Technical University of Denmark, H.C. Ørsteds plads 345b, Kgs. Lyngby, Denmark. E-mail: nickpapior@gmail.com*
[b] *Center for Nanostructured Graphene (CNG), Denmark*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c5cp04613k

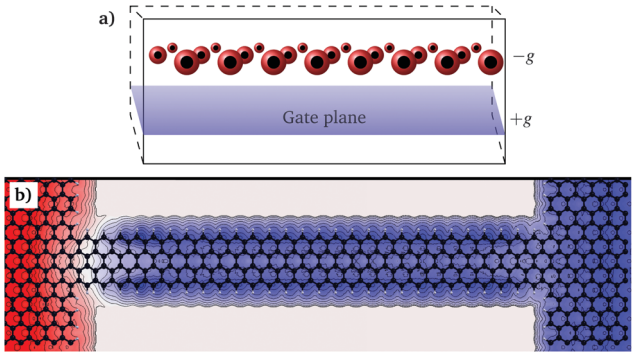


Fig. 1 (a) implementation of the field effect gate model. Redistribution of charge from atoms to gate plane. (b) resulting voltage drop for a 8.3 nm long constriction including a field effect gate of $n = 2 \times 10^{13}$ e$^-$ cm^{-2} and a bias of 0.5 V. Contour lines are separated by 0.022 V.

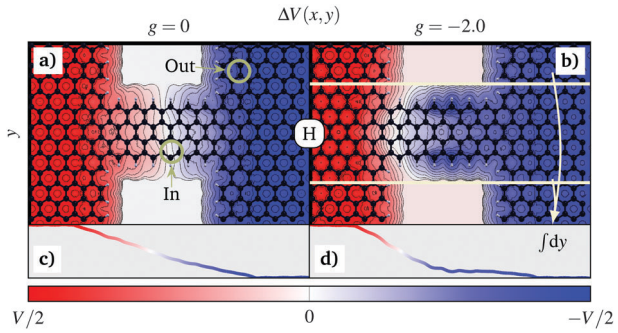


Fig. 2 Electronic Hartree potential drop integrated perpendicular to the plane and above a cutoff electron density $\rho_\varepsilon = 0.008$ e Å^{-3} and projected to the graphene plane for the Hydrogen GNR, (a) and (b). (c) and (d) are the contour plot further integrated in the box indicated in (b). The non-gated system shows a linear gradient, whereas for $g < 0$ (n-doped) a pinning of the potential towards the right (positive) electrode.

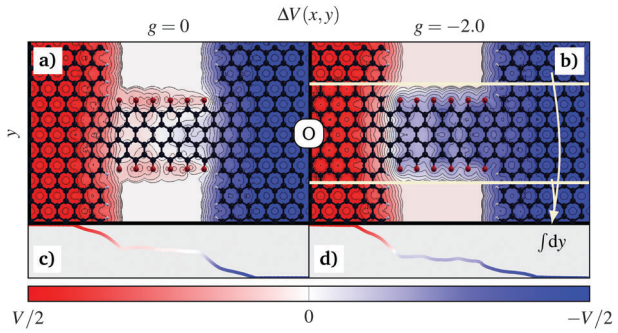


Fig. 3 Electronic Hartree potential drop integrated perpendicular to the plane and above a cutoff electron density $\rho_\varepsilon = 0.008$ e Å^{-3} and projected to the graphene plane for the Oxygen GNR, (a) and (b). (c) and (d) are the contour plot further integrated in the box indicated in (b). The non-gated system shows a gradient at the GNR boundary, whereas for $g < 0$ (n-doped) a pinning of the potential towards the right (positive) electrode.

for additional details of the implementation). We apply this methodology to Gr constrictions consisting of nanoribbon junctions between Gr electrodes. For these we demonstrate how the transport characteristics depend in a non-trivial way on the applied source/drain and gate voltages. As seen in Fig. 1b, upon gating and bias, the voltage drop is pinned to the electrode of a given polarity depending on the doping type and the bias, even for a constriction of 8.3 nm. We can relate the phenomenon to the gate-dependent behavior of the voltage drop in the system which, in turn, can be traced back to the energy dependence of the DOS in the Gr electrodes. The electronic structure of the semi-metallic Gr electrodes displays zero DOS at its charge neutrality point, and a linear increase of the DOS away from it (V-shape). Our analysis demonstrates how the V-shaped DOS in the electrodes control the voltage drop in the Gr junctions indicating a quite generic scenario.

The control of the position of the voltage drop on the nanoscale with gate could be useful in practical applications. We envision this feature, f.ex., could be used to tune the spatial distribution of Joule heating in the device and influence its breaking at the nanoscale.[21,25] Our results highlight the importance of using fully self-consistent electronic transport simulations to predict and design the gating behavior under operating conditions of the emerging class of devices with electrodes having a vanishing DOS.[18]

Results and discussion

We have applied our method to two geometrically similar, "left-right" symmetric Gr nanojunctions, formed by a Gr nanoribbon connected to pristine Gr electrodes, see Fig. 2 and 3. For zero gate/doping ($g = 0$) the former yield an electron–hole symmetric electronic structure (hydrogen GNR), whereas the latter yield a e–h non-symmetric electronic structure (Oxygen GNR).[26–28] The hydrogen-terminated system is also investigated using dopant atoms instead of the electrostatic gating.[11] The simulation unit cell has an area of ∼200 Gr unit cells. The gate is placed 20 beneath the planar Gr structure and we sweep the gating levels (g) according to $g \times 10^{13}$ e$^-$ cm^{-2}.

In Fig. 2a and b, we plot the potential drop across the Gr constriction at 0.5 V for $g = 0$, (a), and n-doped with $g = -2$ gating, (b). The potential profile has been integrated in the perpendicular direction to the Gr surface for electronic densities above $\rho_\varepsilon = 0.008$ e Å^{-3} projected onto the x–y plane. The lower panels, (c) and (d), is a further projection onto the transport direction (x) as indicated in Fig. 2b. At $g = 0$ we obtain an anti-symmetric potential drop in the transport direction ($\Delta V(x) = -\Delta V(-x)$) as expected for a fully e–h and left-right symmetric constriction. On the other hand, in the $g = -2$ (n-doped Gr), we see a clear pinning of the potential profile to the positive electrode, the potential drop at the negative electrode. Conversely, calculations with $g = +2$ (p-doping) with 0.5 V display a pinning at the negative electrode, while for $g = +2$ and −0.5 V we regain the plot shown. This confirms the geometric symmetry.

In Fig. 4 we show the electron transmission spectra for the hydrogen passivated constriction at 0 V, (a), and 0.5 V, (b), for different values of g each vertically shifted 1/2. As a measure of gating we track the position of two resonances, □ and ○ dots, corresponding to a resonance in the constriction located at the

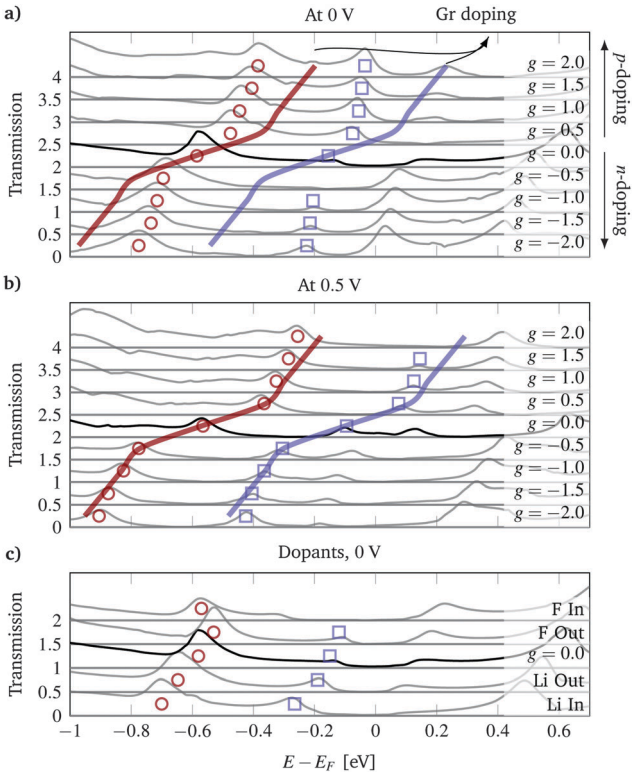


Fig. 4 Transmission spectra for the constriction at various doping levels for 0 V, (a), 0.5 V, (b), and for dopants, (c). The middle line (black) at zero gating is a symmetric transmission function with two distinct resonances (marked □, and ○). Gating the constriction shifts the resonances as indicated by the displacements of the marks. The full lines, crossing vertically the different doping levels, indicates the graphene electrode Fermi level shift due to the doping aligned at the g = 0 mark.

edge and in the center, respectively. The middle thick line is the transmission for $g = 0$, and is equivalent to earlier results where these resonances are discussed.[21] In addition, we plot the energy shift of the Dirac point for pristine graphene as vertical lines aligned at each of the two resonances at $g = 0$. These vertical lines match exactly the shift in chemical potential due to the doping in the electrodes. Discrepancies between the electrode gating (lines) and the resonance positions (dots) illustrate the difference in just rigidly shifting the resonances according to electrode doping, and a fully self-consistent calculation of the resonance positions. Importantly, at 0 V we find that the resonance peaks does not simply follow the gating. Moreover, the two peaks are shifting/gated independently of each other; the center resonance peak, ○, follows the pristine doping closer than the edge resonance peak, □, due to a difference in electrode coupling between the resonances. On the other hand, at 0.5 V we find that both peaks follow the pristine graphene electrode doping. As shown in Fig. 2b, the junction behaves as an extension of the positive electrode and therefore the resonance position is pinned at the Fermi level of this particular electrode. The self-consistent calculation is needed to capture the correct transition with bias from semi-independent resonances to the pinned behavior. The same calculation was performed on a 8.3 nm long ribbon Fig. 1b exhibiting the same pinning feature.

Fig. 3 are for the Oxygen terminated graphene nanoribbon. This nanoribbon has no e–h electronic DOS symmetry.[26–28] Similarly to the hydrogen system we calculate for $g = 0$ and $g = -2$ at 0.5 V. a shows that the Oxygen edges pins slightly to the negative electrode for zero gating, while gating, (b), the entire ribbon is pinned to the positive electrode, equivalent to the hydrogen case Fig. 2b. This is also seen in the projected potential profiles Fig. 3c and d. This confirms that the selectivity of the potential profile in the gated devices does not rely on the e–h symmetry of the junction, and conjectures the generality of this behavior in systems with electrodes having V-shaped DOS around E_F, regardless of the electronic structure of the central part connecting the two electrodes.

The generic behavior of the potential drop just outlined is summarized in Fig. 5, which shows the one-dimensional potential drop calculated for the hydrogen-terminated constriction for a number of different gates and positive bias voltages, similar to that of Fig. 2c and d. Independently on the particular value of the bias voltage applied, gating the system always leads to a marked asymmetry of the potential drop across the constriction. For any value of n-doping, the potential drop pins always to the positive (right) electrode for positive bias. Similarly, for any value of p-doping, the system couples to the negative (left) electrode for positive bias. These results further demonstrate the general phenomenon that does not depend on the particular values of applied gate or bias voltage. Furthermore, our calculations highlight the important fact that the charge neutrality point for the electrodes is a special case which does not extrapolate to the gated case. This becomes even more important if one considers the experimental difficulties in retaining a charge neutral sample.[29,30]

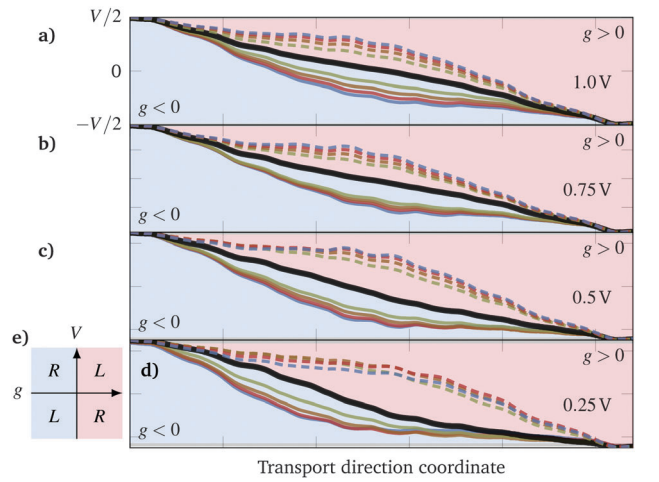


Fig. 5 (a)–(d) Integrated Hartree potential profile in a region of width corresponding to the ribbon along the entire constriction for varying gate levels and applied bias. The thick middle line is the potential profile for $g = 0$. The blue regions correspond to n-doped graphene (full lines), while red are p-doped graphene (dashed lines). The non-gated calculations show a linear behavior whereas gated systems have a non-symmetry between the left and right electrode DOS breaking the left-right anti-symmetry in the potential drop. (e) summarizes the trends where L/R means pinning to the left/right electrode.

Voltage drop model

We will now consider a simple model which can explain the electrode selectivity of the voltage drop depending on the doping/electrostatic gating. Fig. 6 is a guided reference for the following discussion. The position of the voltage drop can be obtained by considering the change in charge in the scattering region when applying a bias. If the scattering region becomes more positive, one can view it as the positive electrode extending into the scattering region and thus the voltage drop will occur closer to the negative electrode and *vice versa*. The change in charge in the scattering region is linked to the change in injected charge from left and right electrodes in the bias window, as noted in the Methods section. The linear dependence of the DOS in the graphene electrodes makes the coupling/broadening functions of the scattering region display the energy dependence,

$$\Gamma_{\mathrm{L/R}}(E) \propto |E - \mu_{\mathrm{L/R}} + E_{\mathrm{F}}|, \tag{1}$$

where $E = 0$ corresponds to the equilibrium Fermi level, E_{F} is the shift of Fermi level due to doping, $E_{\mathrm{F}} \propto \sqrt{g}$, and $\mu_{\mathrm{L/R}}$ is the change in the chemical potential of left/right electrodes with applied voltage bias (V). We will use $\mu_{\mathrm{L}} = eV/2$ and $\mu_{\mathrm{R}} = -eV/2$, and take $V > 0$. This definition means that the scattering region as a starting point will not preferentially select the left or right electrode for an electron–hole symmetric system, and the potential drop profile will be spatially anti-symmetric, $\Delta V(x) = -\Delta V(-x)$. We will now consider the voltage bias as a "perturbation" onto the system without bias, and calculate the change in charge in the scattering region. Thus we first neglect the change in potential set up by the change in charge, which again will impact the charge in the self-consistency. With this we have the density of scattering states from left and right, $A_{\mathrm{L/R}} \propto \Gamma_{\mathrm{L/R}} \propto |E - \mu_{\mathrm{L/R}} + E_{\mathrm{F}}|$, and the change in electrons(holes) injected from left(right) electrode can be written as,

$$\delta\mathrm{e} = \int_0^{eV/2} A_{\mathrm{L}}(E)\mathrm{d}E \propto \frac{eV}{2}\left(E_{\mathrm{F}} - \frac{eV}{4}\right), \tag{2}$$

$$\delta\mathrm{h} = \int_{-eV/2}^{0} A_{\mathrm{R}}(E)\mathrm{d}E \propto \frac{eV}{2}\left(E_{\mathrm{F}} + \frac{eV}{4}\right) \tag{3}$$

where we assume $|V/2| < |E_{\mathrm{F}}|$. The scenario is shown schematically in Fig. 6b showing more injection of positive carriers $\delta\mathrm{h} > \delta\mathrm{e}$. Thus the scattering region will as the first response to the nonequilibrium filling become more positive and we conclude that for n-doping, $g < 0$ and $E_{\mathrm{F}} > 0$, the positive electrode will "extend" into the constriction resulting in a voltage drop at the negative electrode, as seen in Fig. 5. We stress that this behavior stems from the vanishing DOS of graphene at the Dirac point yielding a large relative difference between the electron/hole contributions. Contrary if we take E_{F} to be very large in eqn (2) and (3) we get $\delta\mathrm{e} \approx \delta\mathrm{h}$ and the constriction does not change its charge. Indeed, the pinning effect is smaller at 1 V compared to 0.5 V as seen in Fig. 5a *vs.* 5c. This is due to the DOS of one lead being very close to zero at 0.5 V; $\mu_i - E_{\mathrm{F}} \approx E_{\mathrm{D}}$ with E_{D} being the Dirac point, and hence a much larger relative difference in DOS.

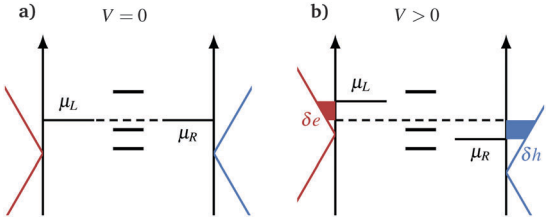


Fig. 6 Illustration of non-symmetric coupling induced by doping out of symmetry. (a) Shows the zero bias configuration with broken e–h symmetry due to doping, $g < 0$. (b) Shows a difference among the electrode contributions in the bias window which pins the system to the right electrode.

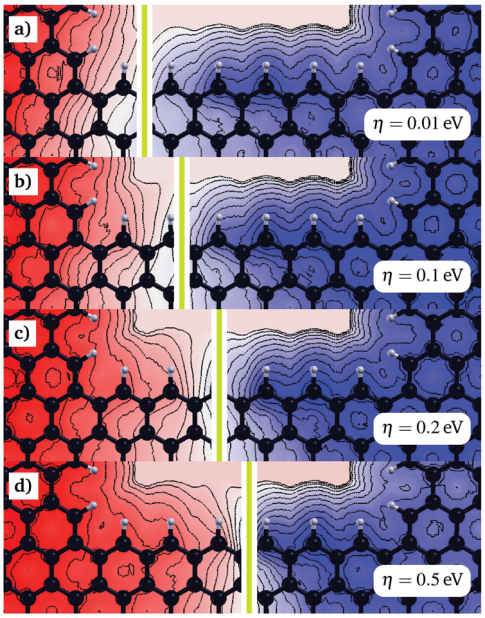


Fig. 7 Change of potential drop *vs.* level broadening parameter, $\eta_{\mathrm{L/R}}$, for 0.5 V. Increasing values smear out the electrode DOS which evens out the electronic contribution from both electrodes in the bias window. The voltage drop becomes anti-symmetric at even charge injection rates from the two electrodes (large smearing).

In order to substantiate that the voltage drop is controlled by the vanishing electrode DOS we smear the DOS energy dependence gradually into a flat function by introducing an artificial increase in the broadening parameter, η, for the electrode self-energies in eqn (4). Hence $\Gamma_{\mathrm{L}}(E) \approx \Gamma_{\mathrm{R}}(E)$ for $\eta \gg 0$ irrespective of the applied bias and gating. This forces $\delta\mathrm{e} \approx \delta\mathrm{h}$ and a resulting anti-symmetric voltage drop. Fig. 7 shows the voltage drop in the middle part of the constriction for four η values. Clearly the anti-symmetric voltage drop is regained when $\eta_{\mathrm{L,R}} \geq 0.5$ eV. Note that since we have not made assumptions in the model about the nature of the constriction we anticipate that it can straightforwardly be applied to similar systems between graphene electrodes in the high-conductance regime.

Constriction, hydrogen terminated with dopants

Since Gr consists entirely of surface atoms it is also extraordinarily susceptible to external influences such as chemical modification or charged impurities. We will now discuss the influence of

modifying the passivation or having adatoms[31–34] as a source of charge doping alternative to the electrostatic gating. We examine the effect of a donating lithium (Li) or an accepting flourine (F) adatom placed either inside or outside the constriction at the positions shown in Fig. 2a. The Li or F atoms are positioned above the center of a hexagon, or ontop a Carbon atom, respectively. In Fig. 4c we show the transmission for the different adatom configurations. The transmission spectra indicate that very little scattering due to the dopants themselves takes place, especially when the adatoms are positioned outside the constriction. The doping effect is clearly seen from the shift in the two resonance peak positions. Li will n-dope the graphene constriction while F p-dope it. Surprisingly, we find that most of the charge transfer to the device resonances is maintained when the dopants are moved outside the constriction. This suggests that nanostructured graphene devices will not necessarily be very sensitive to the actual position of the adatoms. In the case of F it is actually more efficient outside the constriction. Comparing the most significant peak with the field effect gating transmission curves we find that Li donates at least 0.2 electrons while F accepts at least 0.3 electrons from graphene. In addition, we find that a pinning of the potential to the positive/negative electrode occurs for Li(n-doping)/F(p-doping) for positive bias, consistent with the potential drops obtained from field effect gating (see Fig. 2). Adatoms may therefore provide an alternative way to manipulate the voltage drop by pinning the potential to either of the two electrodes. This underlines the conclusion that the main effect is determined by the addition or removal of charge from the device, together with the uneven injection rates from the electrodes.

Conclusion

We have implemented an electrostatic gate method which introduce charge carriers and the corresponding electric field in a capacitor-like setup in self-consistent DFT–NEGF calculations with open boundary conditions to semi-infinite electrodes. The gate method has been applied to several graphene constrictions where the narrowest junction corresponds to a graphene nanoribbon with either hydrogen or oxygen passivation. For positive voltage bias and with electrostatic gating the junction potential gets preferentially pinned to the positive(negative) electrode for n(p)-type doping charge, and *vice versa* for polarity changes of gating and/or bias. Thus the position of the voltage drop can be manipulated by the gate potential or correspondingly from charge doping from adatoms. The constrictions was found to couple selectively to the electrode with the highest DOS contribution in the bias window. The behavior was traced back to the vanishing DOS of graphene close to the Dirac point. A simple perturbation model showed how the selectivity is due to the low DOS of graphene around the Fermi level, irrespective of the details of the junction electronic structure. The V-shaped DOS is also true for the local DOS at armchair edges.[19] Thus we anticipate that our results also apply to molecular junctions more weakly coupled *via* a barrier to armchair edges of graphene. We suggest that this selectivity and high potential gradient can be utilized in experiments on nanostructured graphene or similar 2D materials to control regions of reactivity, manipulate polar adsorbates, or providing control of and insights into the local Joule heating.[25,35] We expect that Kelvin Atomic Force Microscopy,[36] Scanning Tunnelling Potentiometry[37] or Low-Energy Electron Potentiometry[38] to be suitable experimental techniques to examine the effect pointed out here in nanostructured graphene.

Methods

The simulations have been performed using the SIESTA/TranSIESTA code with the PBE-GGA functional for exchange-correlation[39] and a SZP basis-set. A confinement radii determined from an energy shift of 230 meV. The real-space grid cutoff was 230 Ry. The electronic temperature has been set to 25 meV (50 meV for the O-terminated constriction). Unless stated otherwise, the smearing parameter η was set to 10^{-2} eV. The geometries were relaxed until all forces were smaller than 5×10^{-2} eV Å^{-1}. Five transverse k-points were used in the electronic structure calculation. This was increased to between 25 and 50 k-points in the transport calculations. The transmission data have subsequently been interpolated.[40] A vacuum gap of 120 was used in the direction normal to the constriction plane.

Our field effect setup consists of a gate electrode, a dielectric, and the system, here being the graphene nanojunctions. Applying a gate voltage charges the system and electrodes like in a capacitor setup, thus inducing an electrostatic potential gradient across the dielectric, which in this implementation is vacuum. The additional charge will redistribute to create a polarization in the system along the electric field direction. Such field effect setups can be realized in open-boundary DFT calculations by employing a nonequilibrium Green function (NEGF) scheme,[24,41] or by solving the Poisson equation with appropriate boundary conditions.[42,43] The former is a computationally expensive calculation compared to the latter.

Analogous to a plate capacitor setup we assume that an applied gate voltage induces an electron charge $-\delta e^-$ in the system and a corresponding counter-charge $+\delta e^-$ in the gate plane. This situation is accounted for by charging the system with a given electron charge $g = -\delta e^-$, and by distributing homogeneously the corresponding counter-charge $+\delta e^-$ in a well defined region of the unit-cell, denoted gate, so that the overall system + gate remains charge neutral. The setup is shown schematically in Fig. 1a. Thus for $g > 0$ we have a p-doped system, similarly for $g < 0$ we have a n-doped system. Solving the Poisson equation inherently calculates the electric field between the gate and the system. As the calculation cell is periodic we apply the slab dipole correction[42] to terminate the periodic electric field induced by the charge redistribution. The gating method can readily be adopted to transport calculations using NEGF if the gate is uniformly applied to the electrodes and the device. Additionally, the gate at the electrodes must have a resulting electric field perpendicular to the applied bias to assert the correct boundary conditions. Our implementation resembles that of Brumme[44,45] except that we use a linear combination of atomic orbitals method, which means that

the dielectric need not be simulated by a potential barrier to limit electronic penetration.

We note that the DFT–NEGF[24] calculation relies on calculating the density by occupying the left and right scattering states to the different respective chemical potentials. This is done by integrating the left/right spectral density matrices, $\mathbf{A}_{L/R}$, given in terms of the retarded Greens function, $\mathbf{G}$,

$$\mathbf{A}_{L/R}(E) = \mathbf{G}(E)\Gamma_{L/R}(E)\mathbf{G}^{\dagger}(E), \quad (4)$$

$$\mathbf{G}(E) = [(E + i\eta)\mathbf{S} - \mathbf{H} - \Sigma_L(E) - \Sigma_R(E)]^{-1}. \quad (5)$$

Here $\mathbf{H}$, $\mathbf{S}$, $\Gamma_{L/R}(E) = i[\Sigma_{L/R}(E) - \Sigma^{\dagger}_{L/R}(E)]$ are the Hamiltonian, the overlap and the electrode broadening matrices. The parameter $\eta \rightarrow 0^{+}$ introduce a vanishingly small broadening of DOS. However, a finite η broadens the electrode DOS.

The simple Voltage drop model is developed based on the following more detailed description. We consider a left-right symmetric conductor. In nonequilibrium the density (matrix) can formally be written at as an ''equilibrium'' contribution corresponding to the equilibrium Fermi energy, E_F, plus two ''nonequilibrium'' contributions originating from the change in filling of left and right originating scattering states, say, $\mu_L > E_F > \mu_R$. The ''nonequilibrium'' terms corresponding to negative charge injection from the negative electrode, and positive charge injection from the positive electrode,

$$\rho = -\frac{1}{\pi}\int \mathrm{d}E\,\mathrm{Im}\,(E)n_{F,E_F} + \delta e - \delta h, \quad (6)$$

where δe and δh are defined in eqn (2) and (3). We choose $E_F = (\mu_L + \mu_R)/2$ and consider the different fillings as a perturbation. If we neglect the resulting Landauer dipole field in $\mathbf{H}$, which appear in the response to this perturbation in the self-consistent DFT–NEGF calculation, then the first ''equilibrium'' term can not break left-right symmetry and result in a left-right symmetric density. It is then clear that the symmetry breaking and charge in the device is determined by the competition between the latter two contributions which are of opposite sign.

The systems studied here belong to the class highly conducting carbon junctions for which the DFT–NEGF method has been compared favorably to detailed experiments both in the linear[46,47] and non-linear conductance regime.[48] In any case, we are here mainly interested in the qualitative aspects of the behavior of the voltage drop.

Acknowledgements

We thank Osamu Sugino and Pablo Ordejón for useful comments on the method employed, and the Danish e-Infrastructure Cooperation (DeIC) for providing computer resources. The Center for Nanostructured Graphene (CNG) is sponsored by the Danish Research Foundation, Project DNRF58. DS acknowledges the support from the H. C. Ørsted COFUND postdoctoral program at DTU. The contour plots have been prepared with XCrySDen.[49]

References

1 A. Geim and K. S. Novoselov, *Nat. Mater.*, 2007, **6**(3), 183–191.
2 H. Raza, *Graphene nanoelectronics: metrology, synthesis, properties and applications*, Springer, 2011.
3 K. I. Bolotin, K. J. Sikes, Z. Jiang, M. Klima, G. Fudenberg, J. Hone, P. Kim and H. L. Stormer, *Solid State Commun.*, 2008, **146**, 351–355.
4 K. S. Kim, Y. Zhao, H. Jang, S. Y. Lee, J. M. Kim, K. S. Kim, J.-H. Ahn, P. Kim, J.-Y. Choi and B. H. Hong, *Nature*, 2009, **457**, 706–710.
5 T. Georgiou, R. Jalil, B. D. Belle, L. Britnell, R. V. Gorbachev, S. V. Morozov, Y.-J. Kim, A. Gholinia, S. J. Haigh, O. Makarovsky, L. Eaves, L. A. Ponomarenko, A. K. Geim, K. S. Novoselov and A. Mishchenko, *Nat. Nanotechnol.*, 2013, **8**, 100–103.
6 Y. Cao, S. Dong, S. Liu, Z. Liu and X. Guo, *Angew. Chem., Int. Ed.*, 2013, **52**, 3906.
7 C. Jia, J. Wang, C. Yao, Y. Cao, Y. Zhong, Z. Liu, Z. Liu and X. Guo, *Angew. Chem., Int. Ed.*, 2013, **52**, 8666.
8 K. Ullmann, P. B. Coto, S. Leitherer, A. Molina-Ontoria, N. Martn, M. Thoss and H. B. Weber, *Nano Lett.*, 2015, **15**, 3512–3518.
9 H. Sadeghi, J. A. Mol, C. S. Lau, G. A. D. Briggs, J. Warner and C. J. Lambert, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 2658–2663.
10 S. Dutta and S. K. Pati, *J. Mater. Chem.*, 2010, **20**, 8207.
11 Q. Tang, Z. Zhou and Z. Chen, *Nanoscale*, 2013, **5**, 4541.
12 T. G. Pedersen, C. Flindt, J. Pedersen, N. A. Mortensen, A. Jauho and K. Pedersen, *Phys. Rev. Lett.*, 2008, **100**, 136804.
13 J. Bai, X. Zhong, S. Jiang, Y. Huang and X. Duan, *Nat. Nanotechnol.*, 2010, **5**, 190–194.
14 J. Ye, M. F. Craciun, M. Koshino, S. Russo, S. Inoue, H. Yuan, H. Shimotani, A. F. Morpurgo and Y. Iwasa, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 13002–13006.
15 Y. Yin, Z. Cheng, L. Wang, K. Jin and W. Wang, *Sci. Rep.*, 2014, **4**, 5758.
16 H. Song, Y. Kim, Y. H. Jang, H. Jeong, M. A. Reed and T. Lee, *Nature*, 2009, **462**, 1039–1043.
17 F. Prins, A. Barreiro, J. W. Ruitenberg, J. S. Seldenthuis, N. Aliaga-Alcalde, L. M. K. Vandersypen and H. S. J. van der Zant, *Nano Lett.*, 2011, **11**, 4607–4611.
18 E. J. G. Santos and E. Kaxiras, *Nano Lett.*, 2013, **13**, 898–902.
19 D. A. Ryndyk, J. Bundesmann, M.-H. Liu and K. Richter, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **86**, 195425.
20 B. K. Nikolić, K. K. Saha, T. Markussen and K. S. Thygesen, *J. Comput. Electron.*, 2012, **11**, 78–92.
21 T. Gunst, J.-T. Lü, P. Hedegård and M. Brandbyge, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **88**, 161401.
22 J.-T. Lü, R. B. Christensen, G. Foti, T. Frederiksen, T. Gunst and M. Brandbyge, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2014, **89**, 081405.
23 J. M. Soler, E. Artacho, J. D. Gale, A. Garca, J. Junquera, P. Ordejón and D. Sánchez-Portal, *J. Phys.: Condens. Matter*, 2002, **14**, 2745–2779.
24 M. Brandbyge, J.-L. Mozos, P. Ordejón, J. Taylor and K. Stokbro, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2002, **65**, 1–17.
25 R. Murali, Y. Yang, K. Brenner, T. Beck and J. D. Meindl, *Appl. Phys. Lett.*, 2009, **94**, 243114.

26 F. Cervantes-Sodi, G. Csányi, S. Piscanec and A. C. Ferrari, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2008, **77**, 165427.
27 G. Cantele, Y. S. Lee, D. Ninno and N. Marzari, *Nano Lett.*, 2009, **9**, 3425–3429.
28 D. Selli, M. Baldoni, A. Sgamellotti and F. Mercuri, *Nanoscale*, 2012, **4**, 1350.
29 J. Martin, N. Akerman, G. Ulbricht, T. Lohmann, J. H. Smet, K. von Klitzing and A. Yacoby, *Nat. Phys.*, 2008, **4**, 144–148.
30 A. K. M. Newaz, Y. S. Puzyrev, B. Wang, S. T. Pantelides and K. I. Bolotin, *Nat. Commun.*, 2012, **3**, 734.
31 K. Nakada and A. Ishii, *Solid State Commun.*, 2011, **151**, 13–16.
32 P. Marconcini, A. Cresti, F. Triozon, G. Fiori, B. Biel, Y.-M. Niquet, M. Macucci and S. Roche, *ACS Nano*, 2012, **6**, 7942–7947.
33 A. López-Bezanilla, F. Triozon and S. Roche, *Nano Lett.*, 2009, **9**, 2537–2541.
34 B. Biel, X. Blase, F. Triozon and S. Roche, *Phys. Rev. Lett.*, 2009, **102**, 096803.
35 L. Dong, S. Youkey, J. Bush, J. Jiao, V. M. Dubin and R. V. Chebiam, *J. Appl. Phys.*, 2007, **101**, 024320.
36 L. Yan, C. Punckt, I. A. Aksay, W. Mertin and G. Bacher, *Nano Lett.*, 2011, **11**, 3543–3549.
37 P. Willke, T. Druga, R. G. Ulbrich, M. A. Schneider and M. Wenderoth, *Nat. Commun.*, 2015, **6**, 6399.
38 J. Kautz, J. Jobst, C. Sorger, R. M. Tromp, H. B. Weber and S. J. van der Molen, *Sci. Rep.*, 2015, **5**, 13604.
39 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
40 J. T. Falkenberg and M. Brandbyge, *Beilstein J. Nanotechnol.*, 2015, **6**, 1603–1608.
41 D. A. Areshkin and B. K. Nikolić, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2010, **81**, 155450.
42 L. Bengtsson, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1999, **59**, 12301–12304.
43 M. Otani and O. Sugino, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2006, **73**, 115407.
44 T. Brumme, M. Calandra and F. Mauri, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2014, **89**, 1–11.
45 T. Brumme, M. Calandra and F. Mauri, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2015, **91**, 155436.
46 G. Schull, T. Frederiksen, M. Brandbyge and R. Berndt, *Phys. Rev. Lett.*, 2009, **103**, 206803.
47 T. Frederiksen, G. Foti, F. Scheurer, V. Speisser and G. Schull, *Nat. Commun.*, 2014, **5**, 3659.
48 N. L. Schneider, N. Néel, N. P. Andersen, J. T. Lü, M. Brandbyge, J. Kröger and R. Berndt, *J. Phys.: Condens. Matter*, 2015, **27**, 015001.
49 A. Kokalj, *J. Mol. Graphics Modell.*, 1999, **17**, 176–179.

All-graphene edge contacts: Electrical resistance of graphene T-junctions

Kåre Wedel Jacobsen[a,b], Jesper Toft Falkenberg[a,c], Nick Papior[a,c], Peter Bøggild[a,c], Antti-Pekka Jauho[a,c], Mads Brandbyge[a,c,*]

[a] *Center for Nanostructured Graphene (CNG)*
[b] *Dept. of Photonics Engineering, Technical University of Denmark, Ørsteds Plads, Bldg. 343, DK-2800 Kongens Lyngby, Denmark*
[c] *Dept. of Micro- and Nanotechnology, Technical University of Denmark, Ørsteds Plads, Bldg. 345E, DK-2800 Kongens Lyngby, Denmark*

Abstract

Using *ab-initio* methods we investigate the possibility of three-terminal graphene "T-junction" devices and show that these all-graphene edge contacts are energetically feasible when the 1D interface itself is free from foreign atoms. We examine the energetics of various junction structures as a function of the atomic scale geometry. Three-terminal equilibrium Green's functions are used to determine the transmission spectrum and contact resistance of the system. We find that the most symmetric structures have a significant binding energy, and we determine the contact resistances in the junction to be in the range of $1-10\ \mathrm{k\Omega}\,\mu\mathrm{m}$ which is comparable to the best contact resistance reported for edge-contacted graphene-metal contacts[1, 2]. We conclude that conducting all-carbon T-junctions should be feasible.

1. Introduction

Two-dimensional (2D) materials are being vigorously investigated as a platform for nano-scale electronics due to their potential use *e.g.* in flexible electrodes[1], high performance electronics, photovoltaics and spintronics[3]. Graphene plays a key role, not only because it was the first 2D material to be isolated and experimentally characterized[4], but also because of its

*Corresponding author. Tel: +45 45 25 63 28. E-mail: mads.brandbyge@nanotech.dtu.dk (Mads Brandbyge)

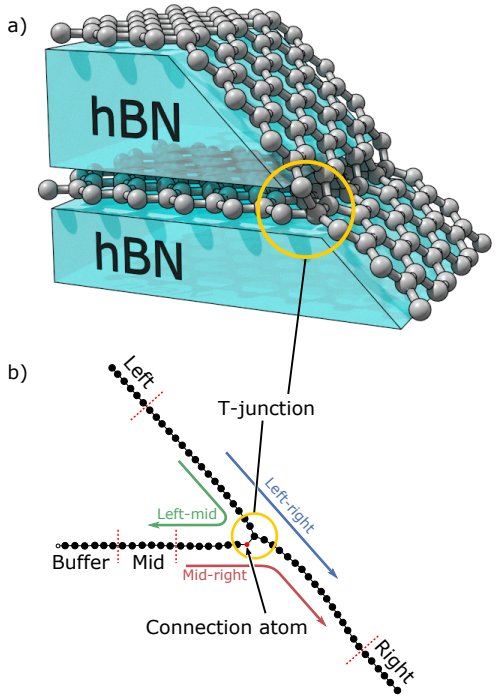


Figure 1: **a)** Sketch of a possible experimental realization where the GTJ is partly encapsulated in a stack of hexagonal boron-nitride. **b)** Schematic of the graphene T-junction system (side view) consisting of 3 semi-infinite graphene electrodes ("Mid"/"Buffer", "Left", "Right") used in the DFT-EGF calculation. In the calculations that include passivation, the connection atom was substituted with other elements. The structure shown here is the result of a nitrogen substitution in the ZZ geometry.

extraordinary electronic properties, which can be harnessed by various types of nanostructuring and chemical functionalization[5, 6, 7, 8, 9, 10, 11, 12].

Recently graphene has played a role as contact electrode to semi-conducting transition metal di-chalcogenides (TMDC) encapsulated in insulating hexagonal boron-nitride (hBN) layers[2]. In these devices graphene is connected to the external circuit via one-dimensional edge-contacts to 3D metal electrodes[2, 13]. In the fabrication process the hBN-G-hBN stack is etched with a slope of approximately 45° resulting in a graphene edge being exposed to subsequent metal electrode deposition.

While these studies have so far focused on planar devices it is relevant to investigate various ways of extending the 2D structures into 3D circuitry.

To this end, and inspired by the experimentally realized 1D edge contacts[2], we here use first principles calculations to investigate graphene T-junctions (GTJ), where the bulk metal electrode is replaced by graphene as illustrated in Fig. 1 and 2. The out-of-plane bonding is possible due to the sp^2 nature of graphene, which can hybridize further to sp^3 and thus allows the formation of a "standing" sheet (or ribbon). The electronic properties of the graphene T-junctions are only limited by the junction itself as the long range ripples are an intrinsic detail in graphene[14]. Formation of such a T-junction requires the edge atoms of one layer to form covalent bonds to the plane of another layer, which can either be done by fusing or by synthesis. Coalescing or fusing of separate carbon nanostructures can either be achieved through Joule heating[15, 16], ion[17, 18] or electron[19] irradiation, where the extraordinary ability of sp^2 carbon nanostructures to self-repair[20] can be exploited to reach well-defined, stable, low-energy configurations. In principle, the alternative approach of bottom-up synthesis can can also lead to creation of complex hybrid all-carbon architectures with interconnections such as graphene carbon-nanotubes[1, 21], 3D interconnected graphene "foam"[22] and vertical T-junction-like "nanowalls"[23, 24]. While graphene-nanotube two-terminal systems have been investigated by first principles calculations[25], the electronic transport properties of T-junctions involving three semi-infinite graphene or graphene nano-ribbon (GNR) electrodes have not, to the best of our knowledge, been investigated. The paper is organized as follows: In Sec. 2 we describe the systems and computational method. In Sec. 3 we present the results for the energetics and structure of both infinitely wide junctions and narrow ribbon junctions to the graphene plane, and discuss the electronic transmission as a function of electronic energy (doping level or gate voltage) and electrical contact resistance. Finally, we summarize the results and present an outlook of future work and experimental realization of graphene T-junctions in Sec. 4.

2. Systems and methods

In order to predict the atomic structures and their binding energy we employ density functional theory calculations (DFT). Electronic transport is calculated by the Green's function method[26] extended to the multi-terminal case, which in the present work involve three electrodes denoted "Left", "Right" and "Mid", see Fig. 1. The middle electrode corresponds to a graphene layer which for instance could be encapsulated in a hBN stack

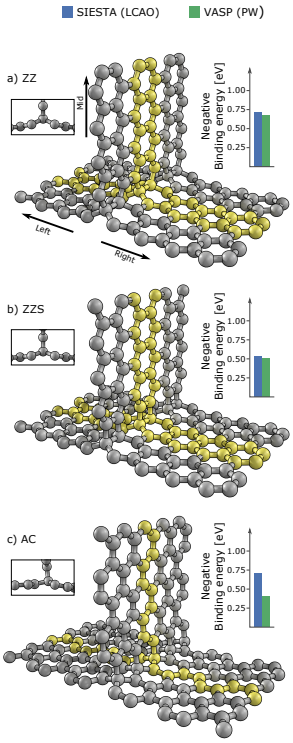


Figure 2: Sketch showing the three main principal structures of interest. The structures are periodic in the transverse direction and yellow indicates atoms belonging to the unit cell. The insets shows the binding energy of the middle part as calculated with the SIESTA (blue bar) and VASP (green bar) codes which are in reasonable agreement. The insets show sideviews of the structures, clearly illustrating the sp^3 geometry at the junction.

as shown in Fig. 1. Here only non-gated graphene electrodes are considered, while in general care must be taken to account properly for gating effects of graphene electrodes[27]. The geometrical structure of the junction is, as we shall see, a main factor determining the resistance.

We first focus on the two main symmetry directions of current flow in graphene (zig-zag and armchair) and consider three different principal structures corresponding to perfect match in the junction as shown in Fig. 2. We note that the structures have been rotated compared to Fig. 1 so that the "mid"-section is pointing out-of-plane of the graphene sheet going from "left" to "right". The simplest possible connection is denoted "ZZ" according to the zig-zag edge of the attached "ribbon" as shown in Fig. 2a, and is mirror

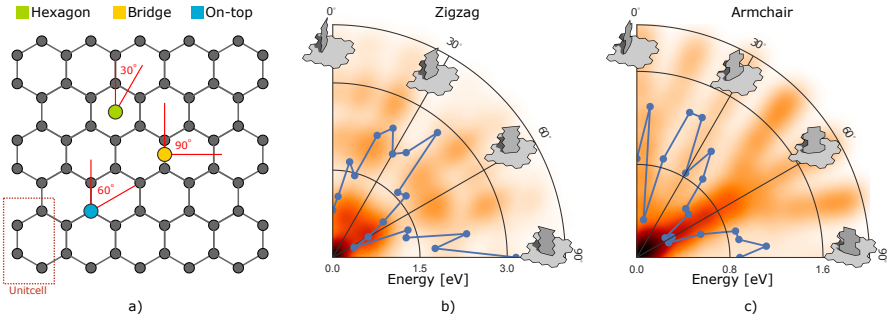


Figure 3: **a)** The setup for the starting configurations of the calculations of ribbon T-junctions. Each calculation was initiated with the ribbons placed one bondlength above the base sheet and with the bottom corner atom above one of the three sites, hexagon (green), bridge (yellow) or on-top (blue). The ribbons were rotated around the anchor points in steps of 5° between the directions indicated with red. **b)**, **c)** The blue curve shows the mean energy of the three ribbons started at each angle (one for each site). It is shown atop a gaussian kernel density plot (red colors) of the two bottom rows of atoms in each of the relaxed ribbons translated to have the outermost atom in the same point and projected onto the plane. The orientation of the underlying graphene sheet is as in a). Both types of ribbons energetically favor the 0° and 60° directions, making the zig-zag ribbon similar to the ZZ(S) structure in Fig. 2, but the armchair ribbon different from AC in the same figure.

symmetric around the middle part and involves only hexagonal carbon rings. The second principal structure is a shifted zig-zag (ZZS) and is also mirror symmetric, but the middle part has been shifted one half of a unit cell resulting in 4- or 8-ring transition in the T-junction. The armchair junction (AC) in Fig. 2c has no mirror-symmetry between the electrodes. However, it is noted that the left-mid and the mid-right transitions have similar grain boundary types in the junctions. In Sec. 3.2 we will see how this similarity is reflected in the transmission through the junction. For all these structures we employ periodic boundary conditions (PBC) along the one-dimensional junction. We have also investigated junctions with an initial angle different from 90° between the "mid" section and the base sheet. When allowed to move freely, however, the "mid" section relaxes towards the symmetric configuration in all cases.

In order to examine the role of a less symmetric match in the T-junction we also investigate *rotated* junctions by attaching the nano-ribbons in the "Mid" position on top of an infinite graphene sheet (still containing "Left"/"Right" electrodes). Both zig-zag and armchair nano-ribbons are investigated by plac-

ing the bottom right carbon atom one bond-length above the graphene sheet consisting of 10×5 4-atom unit cells (indicated in Fig. 3). The attachment site on the graphene sheet is chosen among the three high-symmetry sites, *on-top*, *bridge*, and *hexagon*, as illustrated in Fig. 3a. For each attachment site, the ribbons are rotated in steps of five degrees within the angles as indicated in the figure.

In summary, with the use of periodic boundary conditions, we examine T-junctions formed by infinte graphene sheets as shown in Fig. 2 and finite-width ribbons intersecting with an infinite graphene layer as depicted in the small illustrations in Fig. 3.

2.1. Computational methods

The DFT calculations were performed using the software packages SIESTA[28] and VASP[29]. The former utilizes a localized basis set (LCAO), which allows much faster calculations than the the more accurate plane-wave basis in VASP, which provides a quality check on the total energies of the LCAO method. A SZP basis set was chosen for SIESTA after noting only negligible differences in the resulting relaxed geometries and transmission spectra when comparing to a DZP basis set. The plane-wave calculations used a cut-off of 400 eV. The PBE-GGA functional for exchange-correlation[30] was used for both methods, as well as an atomic force tolerance of 0.04 eV/Å. Additionally, the SIESTA calculations used confinement radii determined from an energy shift of 275 meV with a real-space grid cutoff energy of 300 Ry. The PBC of the principal structures along the 1D junction was utilized in k-point sampling by a Monkhorst-Pack grid of $15 \times 1 \times 1$ ensuring relative energy convergence. In the subsequent transport calculations, the three electrodes ("left", "right", and "mid") were all treated as semi-infinite while the system was modelled as periodic in the transverse direction. The transport calculations were performed using the TranSIESTA[26, 31] method extended with a recently implemented N-electrode capability following Saha et al. [32]. This allows the description of proper boundaries for the three semi-infinite graphene leads. All electrodes are described using surface self-energies from separate bulk calculations. Here we focus on low bias properties and only present equilibrium transport calculations while full non-equilibrium calculations are presented elsewhere[31]. We extended the "Mid" electrode using "buffer" atoms[26] in order to obtain a bulk electrode potential profile for this. Transmission calculations are performed using 100 k-points and post-processed using the interpolation technique described by Falkenberg

and Brandbyge [33] in order to obtain well converged smooth transmission functions.

3. Results

3.1. Energetics

All three structures shown in Fig. 2 have a negative binding energy, and could thus be experimentally feasible. The total energy calculations were consistent when comparing energies from SIESTA with VASP after relaxing the atoms, as is shown as bars to the right of the structures in Fig. 2.

We investigate the rotation of ribbons perpendicularly attached to a graphene sheet, but attached to the sites shown in Fig. 3a and allowing the atoms to relax. In the calculation, one row of atoms in the base sheet was fixed in space while all other atoms – including the entire ribbon – was free to relax. The zig-zag and armchair ribbons were both four rows of atoms wide. Only negligible rotation is observed of the carbon atoms in the GNRs the furthest away from the base sheet, thus allowing us to define a starting angle. The averaged energy of the three ribbon configurations is shown in Fig. 3b,c. The blue line is the mean energy for the three relaxed structures at each starting angle subtracted the minimum energy configuration of the entire set of structures. The angles in Fig 3b,c indicate the initial rather than the final angle of the relaxed structure, which may be slightly different. In order to examine the relaxed direction a density map of atoms is shown as the background of the energy plot. All relaxed configurations have been translated and projected to the graphene plane.

We see that the zig-zag ribbons preferentially will be oriented in the directions similar to the principal periodic structure in Fig. 2a, and are thus corresponding to a geometric transition similar to that of pristine graphene. Note that the armchair ribbons also have the lowest energies when oriented along the armchair direction of the base sheet and are thus *not* similar to the periodic armchair structure described earlier, cf. Fig. 2c. In this configuration, the ribbons are situated symmetrically, but in a way that cannot be periodic without inducing a substantial strain in the graphene layers. Fig. 3c seem to break the 60°-rotation symmetry as the energies at 0° and 5° are much larger than the ones around the other armchair direction. This is caused by one of the three calculated structures in each case that quickly relaxed in to a local minimum with very high energy, thus raising the average energy considerably.

3.2. Transmission

The transmission spectra shown in Fig. 4 show conductance per junction length for each of the three principal systems and between each of the three electrodes defined in Fig. 2. The gray transmission curve is that of pristine graphene. The geometric symmetries of the periodic zig-zag structures (ZZ, ZZS) are reflected in the transmission spectra as the left-mid (blue) and the mid-right (green) curves are identical. Interestingly, it is found that the transmission T through the base sheet (*i.e.* left-to-right) is lower than that from the base and into the mid-terminal for all principal structures. In contrast to the ZZ and ZZS structures the armchair (AC) transmission spectrum display a bandgap-like feature for low energies, but yield the highest transmission into the mid-electrode from left (green curve in Fig. 4 AC) of all the structures.

We can compare the left-right transmission $T_{\text{left-right}}$ through the ZZ structure to that found for hydrogenated kinked graphene[12]. This can be done by realizing that the row of atoms in the mid-part closest to the junction can be exchanged by hydrogen atoms while the rest of the mid-atoms can be removed. Since the junction atoms are allowed to relax we locally have an sp^3-configuration directly comparable to the hydrogenated kinks acting as transmission barriers. There are minor numerical differences in normalized transmission but the trends remain the same: hole transport slowly grows to around 0.03 $G_0/$Å at -2 eV, while the electron transmission grows more rapidly to 0.1 $G_0/$Å at 2 eV.

As previously stated, there are similarities in the atomic transitions from left-to-right and from mid-to right in the AC structure. One could thus expect $T_{\text{left-right}}$ and $T_{\text{mid-right}}$ to differ from $T_{\text{left-mid}}$ and this is also what is found in Fig. 4 AC. The $T_{\text{left-right}}$ and $T_{\text{mid-right}}$ transmission spectra are rather similar even though the angle the electrons have to pass through differs greatly.

For the sake of completeness, we have also briefly investigated the effect of forcing the "mid" part into an non-orthogonal angle with the base graphene. As previously mentioned, the ZZ(S) T-junctions relax towards the symmetric structures shown in Fig. 2 and to avoid this, we have in one case fixed the "mid" terminal part to move doing these relaxations. With an imposed angle of 55° between the "mid" and "right" terminals in an ZZ configuration, only minor changes in the transmission spectrum was observed. The left-to-right and left-to-mid transmissions remained almost unaffected while the mid-to-right lowered a bit to have roughly the same behavior as left-to-right.

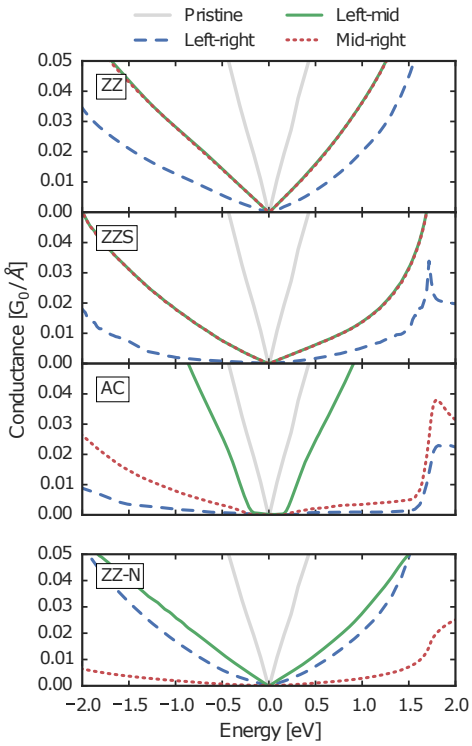


Figure 4: Transmission through the T-junctions for the structures in Fig. 2 (Fermi energy, $E_F = 0$). The bottom frame shows the transmission spectrum for ZZ-N where nitrogen replaces carbon as connecting atoms in the arm of the ZZ junction (cf. red atom in Fig. 1b).

3.3. Passivation

Wang et al. suggest[2] that graphene edges exposed after etching might be passivated with H, F or O. They further treat the edges with an O_2 plasma to change chemical composition[34, 35]. After such processing the edges may be oxygen terminated while other contaminants remaining from the edges should have been removed. To investigate the effects of this edge termination on the transport through the junction we substituted the edge atoms (see connecting atom in Fig. 1b) of the principal T-junction structures with H, F, O or N. We found, as expected, that H and F passivated edges did not bind to the graphene base sheet. On the other hand we found that while O only bonds in one case (ZZ), N binds to both in the ZZ and ZZS structures. A negative formation energy has been obtained for N on a clean zig-zag edge, originating from a N_2 gas[36] making incorporation during hBN etching a possibility. Most importantly, both O and N in the connection introduce an angle between the mid part and the base sheet around 45°, see Fig. 1b which shows the relaxed ZZ-N structure. The same size as the angle in the hypothetical structure based on the hBN-G-hBN setup due to the etch[2] as illustrated in Fig. 1. As seen from Fig. 4 (ZZ-N) the introduction of N in the junction and the resulting angle breaks the left-mid and right-mid symmetry in the transmission spectrum. Now left-mid transmission is significantly higher than that of right-mid while the left-right transmission is nearly unchanged compared to the C-only junction (ZZ). Even though the symmetry breaking due to the nitrogen completely changes the transmission, the magnitude of the transmission does not to vary much between the various configurations. We note that similar high left-mid/left-right transmissions were obtained for O passivation.

For all the relaxed structures with a negative binding energy, similar calculations, *i.e.* the same number of atoms, were made, but without a connection between the mid part and the base sheet. That is, the atoms were lifted off the graphene sheet such that no contact was possible between the two. This setup was then relaxed, and the energy was used as reference for the binding energy. None of the edge-passivated structures had lower energies when forming a T-junction and are thus unlikely to occur.

3.4. Junction resistance

In order to determine the resistance of the contact between the graphene sheet and the attached "Mid" graphene terminal, we calculated the mean junction resistance for electrons with energies on each side of the equilibrium

Fermi energy ($E_F = 0$) in the $0\,\mathrm{eV} - 0.5\,\mathrm{eV}$ range, and likewise for holes, in the $-0.5\,\mathrm{eV} - 0\,\mathrm{eV}$ range. This can be compared to the pristine graphene result for both principal structures and their ribbon counterparts. The result is shown in Fig. 5. Note that since the armchair ribbons have a bandgap we average from the valence/conduction band edges for holes/electrons in this specific case. This explains why some of the AC ribbon resistivities are smaller than their corresponding values for the infinite structures. In general, it is seen that the infinite structures are much more electron-hole-symmetric than the ribbons and have lower resistances. This can clearly be seen in the combined mid-right/left-mid resistance,

$$R_{\mathrm{combined}} = \left(\frac{1}{R_{\mathrm{MR}}} + \frac{1}{R_{\mathrm{LM}}} \right)^{-1}$$

which takes into account both sides of the base layer through the junction.

In an experimental realization of the T-junction it is likely that the connection between the two parts consists of many different bonding configurations due to disorder, defects and the specific orientation of the sheets. In such a case, the transmission would be dominated by the most transparent interface and would probably average to some effective resistance over the junction length.

4. Discussion and Conclusions

Using *ab initio* calculation methods, we find that three-dimensional graphene T-junctions are energetically feasible when no passivation is present in the connection. For N and O passivation some of the junctions are stable, but unlikely to occur because of their relatively high formation energy. The analysis of junctions created between graphene ribbons and an infinite base sheet using a large ensemble of possible starting configurations, revealed a preferential junction orientation along the armchair direction. Interestingly, this was regardless of whether the ribbons were of zigzag or armchair type.

Utilizing the three-terminal Green's functions transport capabilities of TranSIESTA, we find that the transmissive properties of the junctions depend heavily on the specific connection geometry, but that the transmission from the base graphene sheet to the middle terminal is larger than the one through the base sheet. The contact resistance of various metals end-contacted to graphene encapsulated in hBN have been measured[2]. The

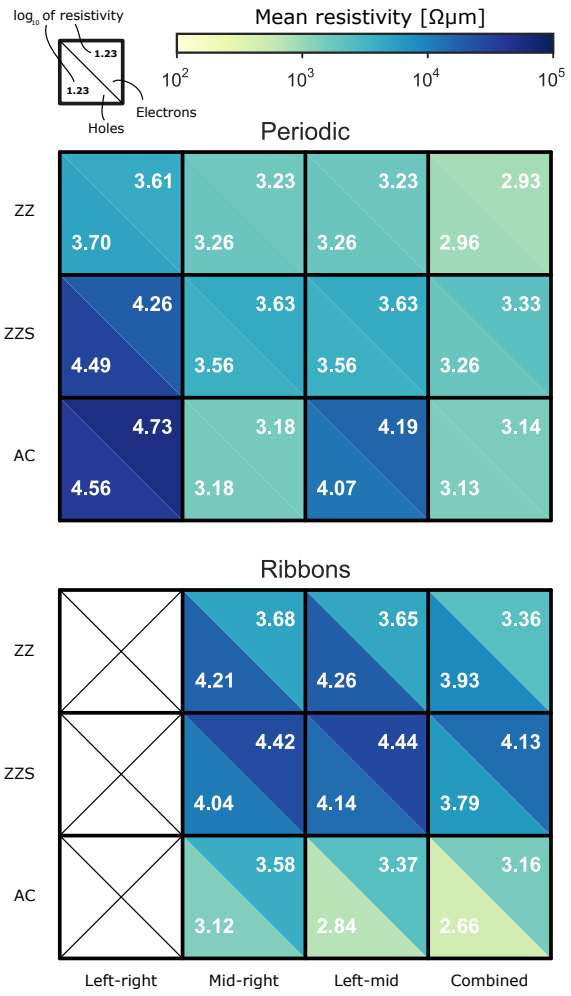


Figure 5: Resistance times length in the T-junctions compared to similar ribbons. The resistivity is taken as the mean value of electrons with energies between 0 eV and 0.5 eV on both the hole and the electron side.

lowest are obtained using Cr (~ 0.1 k$\Omega\,\mu$m), followed by Ti (~ 10 k$\Omega\,\mu$m) and Al, Pd (~ 100 k$\Omega\,\mu$m).

Our results show that not only are such T-junctions feasible, but also in some configurations exhibit superior contact properties. More importantly, the emerging van der Waals technique[37, 2] provides an ideal platform for creating such edge-plane contact architectures in a controlled and possibly scalable manner. Since a number of techniques to form seamless junctions between carbon nano-structures have been demonstrated experimentally, the T-junction architecture opens for a number of exciting possibilities. While we show that the behavior of the all-graphene junctions is already rich, structurally similar materials that should immediately be possible to join and combine in T-junctions include hexagonal boron nitride and boron-carbon-nitrogen alloys[38], where a mix of boron, carbon and nitrogen atoms allows for tunability of the electrical properties, paving the way for bandgap engineered T-junctions. A T-junction with a single sheet of graphene terminated with a few atomic rows of hBN grown by in-plane heterostructures[39] would be a compelling 1D equivalent of a field effect transistor, with the shortest channel imaginable. In fact, the topological similarity with several transistor geometries combined with in-plane heterosynthesis suggests that other 1D analogues of conventional semiconductor components should be possible to realize. Finally, control of the edge chemistry[40] opens for a wide range of possibilities in terms of tailoring the structural and electrical properties of such junctions.

Acknowledgement

The authors thank Bjarke Sørensen Jessen, Lene Gammelgaard, and Filippo Pizzocchero for sharing their experimental insight and ideas, and Dr. Gemma Solomon for valuable comments. Furthermore, the authors thank the Danish e-Infrastructure Cooperation (DeIC) for providing computer resources, EC Graphene FET Flagship (contract number 604391), and the Lundbeck foundation for support (R95-A10510). The Center for Nanostructured Graphene (CNG) is sponsored by the Danish National Research Foundation (Project DNRF103).

References

[1] D. Akinwande, N. Petrone, J. Hone, Two-dimensional flexible nanoelectronics, Nat. Comm. 5 (2014) 5678.

[2] L. Wang, I. Meric, P. Huang, Q. Gao, Y. Gao, H. Tran, et al., One-Dimensional Electrical Contact to a Two-Dimensional Material, Science 342 (6158) (2013) 614–617.

[3] K. Novoselov, V. Fal'ko, L. Colombo, P. Gellert, M. Schwab, K. Kim, A roadmap for graphene, Nature 490 (7419).

[4] A. Geim, K. Novoselov, The rise of graphene, Nat. Mater. 6 (2007) 183–191.

[5] M. Han, B. Özyilmaz, Y. Zhang, P. Kim, Energy Band-Gap Engineering of Graphene Nanoribbons, Physical Review Letters 98 (20) (2007) 206805.

[6] T. Pedersen, C. Flindt, J. Pedersen, N. Mortensen, A.-P. Jauho, K. Pedersen, Graphene Antidot Lattices: Designed Defects and Spin Qubits, Physical Review Letters 100 (13) (2008) 136804.

[7] D. Elias, R. Nair, T. Mohiuddin, Control of graphene's properties by reversible hydrogenation: evidence for graphane, Science 323 (January) (2009) 610–613.

[8] K.-I. Ho, C.-H. Huang, J.-H. Liao, W. Zhang, L.-J. Li, C.-S. Lai, et al., Fluorinated graphene as high performance dielectric materials and the applications for graphene nanoelectronics, Scientific reports 4 (2014) 5893.

[9] J. Sofo, A. Chaudhari, G. Barber, Graphane: A two-dimensional hydrocarbon, Physical Review B 75 (15) (2007) 153401.

[10] J. Oostinga, H. Heersche, X. Liu, A. Morpurgo, L. Vandersypen, Gate-induced insulating state in bilayer graphene devices, Nat. Mater. 7 (2) (2008) 151–157.

[11] T. Ohta, A. Bostwick, T. Seyller, K. Horn, E. Rotenberg, Controlling the electronic structure of bilayer graphene, Science 313 (August) (2006) 951–955.

[12] J. Rasmussen, T. Gunst, P. Bøggild, A.-P. Jauho, M. Brandbyge, Electronic and transport properties of kinked graphene, Beilstein J. Nanotechnol. 4 (1) (2013) 103–110.

[13] X. Cui, G.-H. Lee, Y. Kim, G. Arefe, P. Huang, C.-H. Lee, et al., Multi-terminal transport measurements of MoS2 using a van der Waals heterostructure device platform, Nat. Nanotechnol. 10 (6) (2015) 534–540.

[14] A. Fasolino, J. Los, M. Katsnelson, Intrinsic ripples in graphene, Nat. Mater. 6 (11) (2007) 858–861.

[15] Z. Qi, C. Daniels, S. Hong, Y. Park, V. Meunier, M. Drndic, et al., Electronic Transport of Recrystallized Freestanding Graphene Nanoribbons, ACS Nano 9 (4) (2015) 3510–3520.

[16] R. Zou, Z. Zhang, K. Xu, L. Jiang, Q. Tian, Y. Sun, et al., A method for joining individual graphene sheets, Carbon 50 (13) (2012) 4965–4972.

[17] X. Wu, H. Zhao, M. Zhong, H. Murakawa, M. Tsukamoto, Molecular dynamics simulation of graphene sheets joining under ion beam irradiation, Carbon 66 (2014) 31–38.

[18] X. Wu, H. Zhao, M. Zhong, H. Murakawa, M. Tsukamoto, The Formation of Molecular Junctions between Graphene Sheets, Materials Transactions 54 (6) (2013) 940–946.

[19] M. Terrones, H. Terrones, F. Banhart, J.-C. Charlier, P. Ajayan, Coalescence of single-walled carbon nanotubes, Science 288 (5469) (2000) 1226–1229.

[20] R. Zan, Q. Ramasse, U. Bangert, K. Novoselov, Graphene reknits its holes, Nano Lett. 12 (8) (2012) 3936–3940.

[21] Z. Yan, Z. Peng, G. Casillas, J. Lin, C. Xiang, H. Zhou, et al., Rebar graphene, ACS nano 8 (5) (2014) 5061–5068.

[22] Z. Chen, W. Ren, L. Gao, B. Liu, S. Pei, H.-M. Cheng, Three-dimensional flexible and conductive interconnected graphene networks grown by chemical vapour deposition, Nat. Mater. 10 (6) (2011) 424–428.

[23] N. Lisi, R. Giorgi, M. Re, T. Dikonimos, L. Giorgi, E. Salernitano, et al., Carbon nanowall growth on carbon paper by hot filament chemical vapour deposition and its microstructure, Carbon N.Y. 49 (6) (2011) 2134–2140.

[24] S. Kumar, T. van der Laan, A. E. Rider, L. Randeniya, K. K. Ostrikov, Multifunctional Three-Dimensional T-Junction Graphene Micro-Wells: Energy-Efficient, Plasma-Enabled Growth and Instant Water-Based Transfer for Flexible Device Applications, Advanced Functional Materials 24 (39) (2014) 6114–6122.

[25] F. Novaes, R. Rurali, P. Ordejón, Electronic transport between graphene layers covalently connected by carbon nanotubes, ACS Nano 4 (12) (2010) 7596–7602.

[26] M. Brandbyge, J. Mozos, P. Ordejón, J. Taylor, K. Stokbro, Density-functional method for nonequilibrium electron transport, Phys. Rev. B 65 (2002) 165401.

[27] N. Papior, T. Gunst, D. Stradi, M. Brandbyge, Manipulating the voltage drop in graphene nanojunctions using a gate potential, Phys. Chem. Chem. Phys. 18 (2016) 1025–1031, doi:10.1039/C5CP04613K.

[28] J. Solér, E. Artacho, J. Gale, A. García, J. Junquera, P. Ordejón, et al., The SIESTA method for ab initio order-N materials simulation, Journal of Physics: Condensed Matter 14 (11) (2002) 2745–2779.

[29] G. Kresse, J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Physical Review B 54 (16) (1996) 11169.

[30] J. Perdew, K. Burke, M. Ernzerhof, Generalized Gradient Approximation Made Simple, Physical Review Letters 77 (18) (1996) 3865–3868.

[31] N. Papior, et al., Unpublished .

[32] K. Saha, W. Lu, J. Bernholc, V. Meunier, First-principles methodology for quantum transport in multiterminal junctions, The Journal of Chemical Physics 131 (16) (2009) 164105.

[33] J. Falkenberg, M. Brandbyge, Simple and efficient way of speeding up transmission calculations with k-point sampling, Beilstein J. Nanotechnol. 6 (2015) 1603–1608.

[34] A. Felten, C. Bittencourt, J. Pireaux, G. Van Lier, J. Charlier, Radio-frequency plasma functionalization of carbon nanotubes surface O_2,

NH_3, and CF_4 treatments, Journal of Applied Physics 98 (2005) (2005) 074308.

[35] C. Bittencourt, C. Navio, A. Nicolay, B. Ruelle, T. Godfroid, R. Snyders, et al., Atomic Oxygen Functionalization of Vertically Aligned Carbon Nanotubes, The Journal of Physical Chemistry C 115 (2011) 20412–20418.

[36] J. Jiang, J. Turnbull, W. Lu, P. Boguslawski, J. Bernholc, Theory of nitrogen doping of carbon nanoribbons: Edge effects, The Journal of Chemical Physics 136 (1) (2012) 014702.

[37] A. Geim, I. Grigorieva, Van der Waals heterostructures, Nature 499 (7459) (2013) 419–425.

[38] B. Muchharla, A. Pathak, Z. Liu, L. Song, T. Jayasekera, S. Kar, et al., Tunable electronics in large-area atomic layers of boron–nitrogen–carbon, Nano Lett. 13 (8) (2013) 3476–3481.

[39] Z. Liu, L. Ma, G. Shi, W. Zhou, Y. Gong, S. Lei, et al., In-plane heterostructures of graphene and hexagonal boron nitride with controlled domain sizes, Nat. Nanotechnol. 8 (2) (2013) 119–124.

[40] X. Zhang, O. Yazyev, J. Feng, L. Xie, C. Tao, Y.-C. Chen, et al., Experimentally engineering the edge termination of graphene nanoribbons, ACS Nano 7 (1) (2012) 198–202.

Bibliography

[1] D. A. Reed and J. Dongarra. "Exascale computing and big data". In: *Communications of the ACM* 58.7 (June 2015), pp. 56–68. DOI: 10.1145/2699414 (cit. on p. 1).

[2] Nobelprize.org. *The Nobel Prize in Physics 1956*. 2014 (cit. on p. 1).

[3] R. Chau, B. Doyle, S. Datta, J. Kavalieros, and K. Zhang. "Integrated nanoelectronics for the future". In: *Nature Materials* 6.11 (Nov. 2007), pp. 810–812. DOI: 10.1038/nmat2014 (cit. on pp. 1, 2).

[4] P. Hohenberg and W. Kohn. "Inhomogeneous Electron Gas". In: *Physical Review* 136.3B (Nov. 1964), B864–B871. DOI: 10.1103/PhysRev.136.B864 (cit. on pp. 2, 6).

[5] W. Kohn and L. J. Sham. "Self-Consistent Equations Including Exchange and Correlation Effects". In: *Physical Review* 140.4A (Nov. 1965), A1133–A1138. DOI: 10.1103/PhysRev.140.A1133 (cit. on pp. 2, 7).

[6] J. M. Soler, E. Artacho, J. D. Gale, A. Garca, J. Junquera, P. Ordejn, and D. Snchez-Portal. "The SIESTA method for ab initio order-N materials simulation". In: *Journal of Physics: Condensed Matter* 14.11 (Mar. 2002), pp. 2745–2779. DOI: 10.1088/0953-8984/14/11/302 (cit. on pp. 2, 5, 12, 14).

[7] G. Kresse and J. Furthmller. "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set". In: *Physical Review B* 54.16 (Oct. 1996), pp. 11169–11186. DOI: 10.1103/PhysRevB.54.11169 (cit. on pp. 2, 12).

[8] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, and M. C. Payne. "First principles methods using CASTEP". In: *Zeitschrift fr Kristallographie - Crystalline Materials* 220.5/6 (Jan. 2005). DOI: 10.1524/zkri.220.5.567.65075 (cit. on pp. 2, 12).

[9] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen. "Real-space grid implementation of the projector augmented wave method". In: *Physical Review B* 71.3 (Jan. 2005), p. 035109. DOI: 10.1103/PhysRevB.71.035109 (cit. on pp. 2, 12).

[10] S. Sanvito, C. J. Lambert, J. H. Jefferson, and A. M. Bratkovsky. "General Green's-function formalism for transport calculations with *spd* Hamiltonians and giant magnetoresistance in Co- and Ni-based magnetic multilayers". In: *Physical Review B* 59.18 (May 1999), pp. 11936–11948. DOI: 10.1103/PhysRevB.59.11936 (cit. on pp. 2, 21, 23, 27, 32, 34, 48).

[11] M. Brandbyge, J.-L. Mozos, P. Ordejn, J. Taylor, and K. Stokbro. "Density-functional method for nonequilibrium electron transport". In: *Physical Review B* 65.16 (Mar. 2002), pp. 1–17. DOI: 10.1103/PhysRevB.65.165401 (cit. on pp. 2, 23, 27, 30–32, 41, 51, 61).

[12] A. R. Rocha, V. M. Garca-Surez, S. Bailey, C. Lambert, J. Ferrer, and S. Sanvito. "The Smeagol method for spin- and molecular-electronics". In: (Oct. 2005). arXiv: 0510083 [cond-mat] (cit. on pp. 2, 23, 27, 34, 41).

[13] R. Li, J. Zhang, S. Hou, Z. Qian, Z. Shen, X. Zhao, and Z. Xue. "A corrected NEGF+DFT approach for calculating electronic transport through molecular devices: Filling bound states and patching the non-equilibrium integration". In: *Chemical Physics* 336.2-3 (July 2007), pp. 127–135. DOI: 10.1016/j.chemphys.2007.06.011 (cit. on pp. 2, 27, 30).

[14] A. Pecchia, G. Penazzi, L. Salvucci, and A. Di Carlo. "Non-equilibrium Green's functions in density functional tight binding: method and applications". In: *New Journal of Physics* 10.6 (June 2008), p. 065022. DOI: 10.1088/1367-2630/10/6/065022 (cit. on pp. 2, 4, 23).

[15] K. K. Saha, W. Lu, J. Bernholc, and V. Meunier. "First-principles methodology for quantum transport in multiterminal junctions." In: *The Journal of chemical physics* 131.16 (Oct. 2009), p. 164105. DOI: 10.1063/1.3247880 (cit. on pp. 2, 23, 26, 30, 34, 47).

[16] T. Ozaki, K. Nishio, and H. Kino. "Efficient implementation of the nonequilibrium Green function method for electronic transport calculations". In: *Physical Review B* 81.3 (Jan. 2010), p. 035116. DOI: 10.1103/PhysRevB.81.035116 (cit. on pp. 2, 12, 23, 24, 27, 32, 34, 41, 52, 63).

[17] J. Chen, M. Badioli, P. Alonso-Gonzlez, S. Thongrattanasiri, F. Huth, J. Osmond, M. Spasenovi, A. Centeno, A. Pesquera, P. Godignon, A. Zurutuza Elorza, N. Camara, F. J. G. de Abajo, R. Hillenbrand, and F. H. L. Koppens. "Optical nano-imaging of gate-tunable graphene plasmons". In: *Nature* (June 2012). DOI: 10.1038/nature11254 (cit. on pp. 2, 63).

[18] B. K. Nikoli, K. K. Saha, T. Markussen, and K. S. Thygesen. "First-principles quantum transport modeling of thermoelectricity in single-molecule nanojunctions with graphene nanoribbon electrodes". In: *Journal of Computational Electronics* 11.1 (2012), pp. 78–92. DOI: 10.1007/s10825-012-0386-y. arXiv: 1111.0106 (cit. on pp. 2, 61).

[19] S. J. Tans, A. R. M. Verschueren, and C. Dekker. "Room-temperature transistor based on a single carbon nanotube". In: *Nature* 393.6680 (May 1998), pp. 49–52. DOI: 10.1038/29954 (cit. on p. 2).

[20] A. Javey, J. Guo, Q. Wang, M. Lundstrom, and H. Dai. "Ballistic carbon nanotube field-effect transistors". In: *Nature* 424.6949 (Aug. 2003), pp. 654–657. DOI: 10.1038/nature01797 (cit. on p. 2).

[21] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov. "Electric field effect in atomically thin carbon films." In: *Science (New York, N.Y.)* 306.5696 (Oct. 2004), pp. 666–9. DOI: 10.1126/science.1102896 (cit. on pp. 2, 73).

[22] A. K. Geim and K. S. Novoselov. "The rise of graphene". In: *Nature Materials* 6.3 (Mar. 2007), pp. 183–191. DOI: 10.1038/nmat1849 (cit. on pp. 2, 3, 73).

[23] F. Schwierz. "Graphene transistors". In: *Nature Nanotechnology* 5.7 (July 2010), pp. 487–496. DOI: 10.1038/nnano.2010.89 (cit. on p. 2).

[24] M. J. Allen, V. C. Tung, and R. B. Kaner. "Honeycomb Carbon: A Review of Graphene". In: *Chemical Reviews* 110.1 (Jan. 2010), pp. 132–145. DOI: 10.1021/cr900070d (cit. on p. 2).

[25] K. S. Novoselov, V. I. Fal'ko, L. Colombo, P. R. Gellert, M. G. Schwab, and K. Kim. "A roadmap for graphene". In: *Nature* 490.7419 (Oct. 2012), pp. 192–200. DOI: 10.1038/nature11458 (cit. on p. 2).

[26] F. Prins, A. Barreiro, J. W. Ruitenberg, J. S. Seldenthuis, N. Aliaga-Alcalde, L. M. K. Vandersypen, and H. S. J. van der Zant. "Room-temperature gating of molecular junctions using few-layer graphene nanogap electrodes." In: *Nano Lett.* 11.11 (Nov. 2011), pp. 4607–11. DOI: 10.1021/nl202065x (cit. on pp. 2, 3, 61).

[27] J. Ye, M. F. Craciun, M. Koshino, S. Russo, S. Inoue, H. Yuan, H. Shimotani, A. F. Morpurgo, and Y. Iwasa. "Accessing the transport properties of graphene and its multilayers at high carrier density". In: *Proceedings of the National Academy of Sciences* 108.32 (Aug. 2011), pp. 13002–13006. DOI: 10.1073/pnas.1018388108 (cit. on pp. 2, 61).

[28] J. R. Williams, L. DiCarlo, and C. M. Marcus. "Quantum Hall Effect in a Gate-Controlled p-n Junction of Graphene". In: *Science* 317.5838 (Aug. 2007), pp. 638–641. DOI: 10.1126/science.1144657 (cit. on p. 2).

[29] S. S. Datta, D. R. Strachan, E. J. Mele, and a. T. C. Johnson. "Surface potentials and layer charge distributions in few-layer graphene films." In: *Nano letters* 9.1 (Jan. 2009), pp. 7–11. DOI: 10.1021/nl8009044 (cit. on pp. 2, 4).

[30] F. Giannazzo, S. Sonde, V. Raineri, and E. Rimini. "Screening length and quantum capacitance in graphene by scanning probe microscopy." In: *Nano letters* 9.1 (Jan. 2009), pp. 23–9. DOI: 10.1021/nl801823n (cit. on pp. 2, 4, 63).

[31] S. K. Choudhary and A. K. Gupta. "Scanning Tunneling Microscopy and Spectroscopy study of charge inhomogeneities in bilayer Graphene". In: (July 2010), p. 5. DOI: 10.1016/j.ssc.2010.12.010. arXiv: 1007.4417 (cit. on pp. 2, 4).

[32] Q. H. Wang, Z. Jin, K. K. Kim, A. J. Hilmer, G. L. C. Paulus, C.-J. Shih, M.-H. Ham, J. D. Sanchez-Yamagishi, K. Watanabe, T. Taniguchi, J. Kong, P. Jarillo-Herrero, and M. S. Strano. "Understanding and controlling the substrate effect on graphene electron-transfer chemistry via reactivity imprint lithography." In: *Nature chemistry* 4.9 (Sept. 2012), pp. 724–32. DOI: 10.1038/nchem.1421 (cit. on p. 2).

[33] A. Bandyopadhyay, S. S. R. K. C. Yamijala, and S. K. Pati. "Tuning the electronic and optical properties of graphene and boron-nitride quantum dots by molecular charge-transfer interactions: a theoretical study." In: *Physical chemistry chemical physics : PCCP* 15.33 (July 2013), pp. 13881–7. DOI: 10.1039/c3cp51510a (cit. on p. 2).

[34] M. Bruna, A. K. Ott, M. Ijas, D. Yoon, U. Sassi, and A. C. Ferrari. "Doping dependence of the Raman spectrum of defected graphene". In: (May 2014). arXiv: 1405.4264 (cit. on p. 2).

[35] T. G. Pedersen, C. Flindt, J. Pedersen, N. A. Mortensen, A. Jauho, and K. Pedersen. "Graphene Antidot Lattices: Designed Defects and Spin Qubits". In: *Physical Review Letters* 100.13 (Apr. 2008), p. 136804. DOI: 10.1103/PhysRevLett.100.136804 (cit. on p. 3).

[36] T. Gunst, J.-T. L, P. Hedegrd, and M. Brandbyge. "Phonon excitation and instabilities in biased graphene nanoconstrictions". In: *Physical Review B* 88.16 (Oct. 2013), p. 161401. DOI: 10.1103/PhysRevB.88.161401 (cit. on pp. 3, 61, 66).

[37] M. Y. Han, B. zyilmaz, Y. Zhang, and P. Kim. "Energy Band-Gap Engineering of Graphene Nanoribbons". In: *Physical Review Letters* 98.20 (May 2007), p. 206805. DOI: 10.1103/PhysRevLett.98.206805 (cit. on p. 3).

[38] L. Yan, C. Punckt, I. A. Aksay, W. Mertin, and G. Bacher. "Local Voltage Drop in a Single Functionalized Graphene Sheet Characterized by Kelvin Probe Force Microscopy". In: *Nano Letters* 11.9 (Sept. 2011), pp. 3543–3549. DOI: 10.1021/nl201070c (cit. on p. 3).

[39] Y. Hancock, A. Uppstu, K. Saloriutta, A. Harju, and M. J. Puska. "Generalized tight-binding transport model for graphene nanoribbon-based systems". In: *Physical Review B* 81.24 (June 2010), p. 245402. DOI: 10.1103/PhysRevB.81.245402 (cit. on pp. 3, 23, 54, 55, 100).

[40] D. Abergel, V. Apalkov, J. Berashevich, K. Ziegler, and T. Chakraborty. "Properties of graphene: a theoretical perspective". In: *Advances in Physics* 59.4 (July 2010), pp. 261–482. DOI: 10.1080/00018732.2010.487978 (cit. on p. 3).

[41] B. J. LeRoy, J. Kong, V. K. Pahilwani, C. Dekker, and S. G. Lemay. "Three-terminal scanning tunneling spectroscopy of suspended carbon nanotubes". In: *Physical Review B* 72.7 (Aug. 2005), p. 075413. DOI: 10.1103/PhysRevB.72.075413 (cit. on p. 4).

[42] Y. Wu, D. B. Farmer, W. Zhu, S.-J. Han, C. D. Dimitrakopoulos, A. A. Bol, P. Avouris, and Y.-M. Lin. "Three-Terminal Graphene Negative Differential Resistance Devices". In: *ACS Nano* 6.3 (Mar. 2012), pp. 2610–2616. DOI: 10.1021/nn205106z (cit. on p. 4).

[43] L. Wang, I. Meric, P. Y. Huang, Q. Gao, Y. Gao, H. Tran, T. Taniguchi, K. Watanabe, L. M. Campos, D. A. Muller, J. Guo, P. Kim, J. Hone, K. L. Shepard, and C. R. Dean. "One-Dimensional Electrical Contact to a Two-Dimensional Material". In: *Science* 342.6158 (Nov. 2013), pp. 614–617. DOI: 10.1126/science.1244358 (cit. on pp. 4, 73).

[44] Z. Chen, W. Ren, L. Gao, B. Liu, S. Pei, and H.-M. Cheng. "Three-dimensional flexible and conductive interconnected graphene networks grown by chemical vapour deposition". In: *Nature Materials* 10.6 (June 2011), pp. 424–428. DOI: 10.1038/nmat3001 (cit. on p. 4).

[45] L. Yu, Y.-H. Lee, X. Ling, E. J. G. Santos, Y. C. Shin, Y. Lin, M. Dubey, E. Kaxiras, J. Kong, H. Wang, and T. Palacios. "Graphene/MoS 2 Hybrid Technology for Large-Scale Two-Dimensional Electronics". In: *Nano Letters* 14.6 (June 2014), pp. 3055–3063. DOI: 10.1021/nl404795z (cit. on p. 4).

[46] P. Puczkarski, P. Gehring, C. S. Lau, J. Liu, A. Ardavan, J. H. Warner, G. A. D. Briggs, and J. A. Mol. "Three-terminal graphene single-electron transistor fabricated using feedback-controlled electroburning". In: *Applied Physics Letters* 107.13 (Sept. 2015), p. 133105. DOI: 10.1063/1.4932133 (cit. on p. 4).

[47] M. Otani and O. Sugino. "First-principles calculations of charged surfaces and interfaces: A plane-wave nonrepeated slab approach". In: *Physical Review B* 73.11 (Mar. 2006), p. 115407. DOI: 10.1103/PhysRevB.73.115407 (cit. on pp. 4, 61, 62).

[48] T. Brumme, M. Calandra, and F. Mauri. "Electrochemical doping of few-layer ZrNCl from first principles: Electronic and structural properties in field-effect configuration". In: *Physical Review B - Condensed Matter and Materials Physics* 89 (2014), pp. 1–11. DOI: 10.1103/PhysRevB.89.245406 (cit. on pp. 4, 62).

[49] T. Brumme, M. Calandra, and F. Mauri. "First-principles theory of field-effect doping in transition-metal dichalcogenides: Structural properties, electronic structure, Hall coefficient, and electrical conductivity". In: *Physical Review B* 91 (2015), p. 155436. DOI: 10.1103/PhysRevB.91.155436. arXiv: 1501.07223 (cit. on pp. 4, 62).

[50] J. Chen, K. S. Thygesen, and K. W. Jacobsen. "Ab initio nonequilibrium quantum transport and forces with the real-space projector augmented wave method". In: *Physical Review B* 85.15 (Apr. 2012), p. 155140. DOI: 10.1103/PhysRevB.85.155140. arXiv: 1204.4175 (cit. on pp. 4, 23, 27, 34, 63).

[51] Y. Okuno and T. Ozaki. "First-principles study of multiterminal quantum interference controlled molecular devices". In: *Journal of Physical Chemistry C* 117.1 (2013), pp. 100–109. DOI: 10.1021/jp309455n (cit. on pp. 4, 23, 34).

[52] C. Kittel. *Introduction to Solid State Physics*. 8th ed. John Wiley & Sons, 2005 (cit. on pp. 5, 7, 9, 11).

[53] R. M. Martin. *Electronic Structure: Basic Theory and Practical Methods*. 1st ed. Cambridge University Press, 2004 (cit. on pp. 5, 7).

[54] J. Kohanoff. *Electronic Structure Calculations for Solids and Molecules*. 1st ed. Cambridge University Press, 2006 (cit. on pp. 5, 7, 9).

[55] J. C. Cuevas and E. Scheer. *Molecular Electronics: An Introduction to Theory and Experiment*. 1st ed. World Scientific, 2010 (cit. on pp. 5, 7, 9, 15, 47).

[56] M. Born and R. Oppenheimer. "Zur Quantentheorie der Molekeln". In: *Annalen der Physik* 389.20 (1927), pp. 457–484. DOI: 10.1002/andp.19273892002 (cit. on p. 6).

[57] R. O. Jones and O. Gunnarsson. "The density functional formalism, its applications and prospects". In: *Reviews of Modern Physics* 61.3 (July 1989), pp. 689–746. DOI: 10.1103/RevModPhys.61.689 (cit. on p. 6).

[58] A. Zangwill. "A half century of density functional theory". In: *Physics Today* 68.7 (2015), pp. 34–39. DOI: 10.1063/PT.3.2846 (cit. on p. 6).

[59] R. Stowasser and R. Hoffmann. "What Do the Kohn-Sham Orbitals and Eigenvalues Mean?" In: *Journal of the American Chemical Society* 121.14 (Apr. 1999), pp. 3414–3420. DOI: 10.1021/ja9826892 (cit. on p. 7).

[60] R. Peverati and D. G. Truhlar. "Quest for a universal density functional: the accuracy of density functionals across a broad spectrum of databases in chemistry and physics". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 372.2011 (Feb. 2014), pp. 20120476–20120476. DOI: 10.1098/rsta.2012.0476 (cit. on p. 8).

[61] M. A. Marques, M. J. Oliveira, and T. Burnus. "Libxc: A library of exchange and correlation functionals for density functional theory". In: *Computer Physics Communications* 183.10 (Oct. 2012), pp. 2272–2281. DOI: 10.1016/j.cpc.2012.05.007 (cit. on p. 8).

[62] J. P. Perdew and A. Zunger. "Self-interaction correction to density-functional approximations for many-electron systems". In: *Physical Review B* 23.10 (May 1981), pp. 5048–5079. DOI: 10.1103/PhysRevB.23.5048 (cit. on pp. 8, 84).

[63] J. P. Perdew, K. Burke, and M. Ernzerhof. "Generalized Gradient Approximation Made Simple". In: *Physical Review Letters* 77.18 (Oct. 1996), pp. 3865–3868. DOI: 10.1103/PhysRevLett.77.3865 (cit. on pp. 8, 64, 73).

[64] R. P. Feynman. "Forces in Molecules". In: *Physical Review* 56.4 (Aug. 1939), pp. 340–343. DOI: 10.1103/PhysRev.56.340 (cit. on p. 8).

[65] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi. "Phonons and related crystal properties from density-functional perturbation theory". In: *Reviews of Modern Physics* 73.2 (July 2001), pp. 515–562. DOI: 10.1103/RevModPhys.73.515 (cit. on p. 8).

[66] J. M. Pruneda, S. K. Estreicher, J. Junquera, J. Ferrer, and P. Ordejn. "Ab initio local vibrational modes of light impurities in silicon". In: *Physical Review B* 65.7 (Feb. 2002), p. 075210. DOI: 10.1103/PhysRevB.65.075210 (cit. on p. 9).

[67] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch. "QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials". In: *Journal of Physics: Condensed Matter* 21.39 (Sept. 2009), p. 395502. DOI: 10.1088/0953-8984/21/39/395502 (cit. on p. 12).

[68] X. Gonze, J.-M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G.-M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, P. Ghosez, J.-Y. Raty, and D. Allan. "First-principles computation of material properties: the ABINIT software project". In: *Computational Materials Science* 25.3 (Nov. 2002), pp. 478–492. DOI: 10.1016/S0927-0256(02)00325-7 (cit. on p. 12).

[69] R. S. Mulliken. "Electronic Population Analysis on LCAO[Single Bond]MO Molecular Wave Functions. I". In: *The Journal of Chemical Physics* 23.10 (1955), p. 1833. DOI: 10.1063/1.1740588 (cit. on pp. 13, 47, 53).

[70] C. Fonseca Guerra, J.-W. Handgraaf, E. J. Baerends, and F. M. Bickelhaupt. "Voronoi deformation density (VDD) charges: Assessment of the Mulliken, Bader, Hirshfeld, Weinhold, and VDD methods for charge analysis". In: *Journal of Computational Chemistry* 25.2 (Jan. 2004), pp. 189–210. DOI: 10.1002/jcc.10351 (cit. on pp. 13, 53).

[71] M. Soriano and J. J. Palacios. "Theory of projections with non-orthogonal basis sets: Partitioning techniques and effective Hamiltonians". In: *Physical Review B* 90.7 (Apr. 2014), p. 12. arXiv: 1404.2043 (cit. on p. 13).

[72] L. Bengtsson. "Dipole correction for surface supercell calculations". In: *Physical Review B* 59.19 (May 1999), pp. 12301–12304. DOI: 10.1103/PhysRevB.59.12301 (cit. on pp. 14, 61, 62).

[73] M. C. M. Wright. "Green function or Green's function?" In: *Nature Physics* 2.10 (Oct. 2006), pp. 646–646. DOI: 10.1038/nphys411 (cit. on p. 15).

[74] J. J. Sakurai. *Modern Quantum Mechanics*. rev. 1994 (cit. on pp. 15, 47).

[75] S. Datta. *Electronic Transport in Mesoscopic Systems*. Cambridge: Cambridge University Press, 1995. DOI: 10.1017/CBO9780511805776 (cit. on pp. 15, 47).

[76] M. P. L. Sancho, J. M. L. Sancho, and J. Rubio. "Highly convergent schemes for the calculation of bulk and surface Green functions". In: *Journal of Physics F: Metal Physics* 15.4 (Apr. 1985), pp. 851–858. DOI: 10.1088/0305-4608/15/4/009 (cit. on pp. 19–21, 34).

[77] D. H. Lee and J. D. Joannopoulos. "Simple scheme for surface-band calculations. II. The Green's function". In: *Physical Review B* 23.10 (May 1981), pp. 4997–5004. DOI: 10.1103/PhysRevB.23.4997 (cit. on p. 21).

[78] D. H. Lee and J. D. Joannopoulos. "Simple scheme for surface-band calculations. I". In: *Physical Review B* 23.10 (May 1981), pp. 4988–4996. DOI: 10.1103/PhysRevB.23.4988 (cit. on p. 21).

[79] A. Umerski. "Closed-form solutions to surface Green's functions". In: *Physical Review B* 55.8 (Feb. 1997), pp. 5266–5275. DOI: 10.1103/PhysRevB.55.5266 (cit. on p. 21).

[80] J.-S. Wang, J. Wang, and J. T. L. "Quantum thermal transport in nanostructures". In: *The European Physical Journal B* 62.4 (May 2008), pp. 381–404. DOI: 10.1140/epjb/e2008-00195-8 (cit. on p. 21).

[81] I. Rungger and S. Sanvito. "Algorithm for the construction of self-energies for electronic transport calculations based on singularity elimination and singular value decomposition". In: *Physical Review B* 78.3 (July 2008), p. 035407. DOI: 10.1103/PhysRevB.78.035407 (cit. on pp. 21, 23).

[82] Z. Qian, R. Li, S. Hou, Z. Xue, and S. Sanvito. "An efficient nonequilibrium Green's function formalism combined with density functional theory approach for calculating electron transport properties of molecular devices with quasi-one-dimensional electrodes." In: *The Journal of chemical physics* 127.19 (Nov. 2007), p. 194710. DOI: 10.1063/1.2804876 (cit. on pp. 23, 34).

[83] T. N. Todorov. "Tight-binding simulation of current-carrying nanostructures". In: *Journal of Physics: Condensed Matter* 14.11 (Mar. 2002), pp. 3049–3084. DOI: 10.1088/0953-8984/14/11/314 (cit. on pp. 23, 53).

[84] S. M.-M. M.-M. Dubois, A. Lopez-Bezanilla, A. Cresti, F. Triozon, B. Biel, J.-C. Charlier, and S. Roche. "Quantum Transport in Graphene Nanoribbons: Effects of Edge Reconstruction and Chemical Reactivity". In: *ACS Nano* 4.4 (Apr. 2010), pp. 1971–1976. DOI: 10.1021/nn100028q (cit. on p. 23).

[85] A. Pecchia and A. D. Carlo. "Atomistic theory of transport in organic and inorganic nanostructures". In: *Reports on Progress in Physics* 67.8 (Aug. 2004), pp. 1497–1561. DOI: 10.1088/0034-4885/67/8/R04 (cit. on pp. 23, 47, 53).

[86] R. Zhang, I. Rungger, S. Sanvito, and S. Hou. "Current-induced energy barrier suppression for electromigration from first principles". In: *Physical Review B* 84.8 (Aug. 2011), p. 085445. DOI: 10.1103/PhysRevB.84.085445 (cit. on pp. 24, 53).

[87] M. Di Ventra and S. T. Pantelides. "Hellmann-Feynman theorem and the definition of forces in quantum time-dependent and transport problems". In: *Physical Review B* 61.23 (June 2000), pp. 16207–16212. DOI: 10.1103/PhysRevB.61.16207 (cit. on p. 24).

[88] *Atomistix ToolKit version 2015.0*. 2015 (cit. on pp. 27, 32, 63).

[89] H. Takahasi and M. Mori. "Double exponential formulas for numerical integration". In: *Publications of the Research Institute for Mathematical Sciences* 9.3 (1974), pp. 721–741. DOI: 10.2977/prims/1195192451 (cit. on p. 27).

[90] T. Ozaki. "Continued fraction representation of the Fermi-Dirac function for large-scale electronic structure calculations". In: *Physical Review B* 75.3 (Jan. 2007), p. 035123. DOI: 10.1103/PhysRevB.75.035123 (cit. on pp. 27, 28).

[91] K. S. Thygesen and K. W. Jacobsen. "Interference and k-point sampling in the supercell approach to phase-coherent transport". In: *Physical Review B* 72.3 (July 2005), p. 033401. DOI: 10.1103/PhysRevB.72.033401 (cit. on p. 32).

[92] S. Ulstrup, T. Frederiksen, and M. Brandbyge. "Nonequilibrium electron-vibration coupling and conductance fluctuations in a C60 junction". In: *Physical Review B* 86.24 (Dec. 2012), p. 245417. DOI: 10.1103/PhysRevB.86.245417 (cit. on p. 32).

[93] E. M. Godfrin. "A method to compute the inverse of an n-block tridiagonal quasi-Hermitian matrix". In: *Journal of Physics: Condensed Matter* 3.40 (Oct. 1991), pp. 7843–7848. DOI: 10.1088/0953-8984/3/40/005 (cit. on pp. 34, 35).

[94] Y. Xue, S. Datta, and M. A. Ratner. "First-principles based matrix Green's function approach to molecular electronic devices: general formalism". In: *Chemical Physics* 281.2-3 (Aug. 2002), pp. 151–170. DOI: 10.1016/S0301-0104(02)00446-9 (cit. on p. 34).

[95] B. Feldman, T. Seideman, O. Hod, and L. Kronik. "A real-space method for highly parallelizable electronic transport calculations". In: (Jan. 2014). arXiv: 1401.0782 (cit. on p. 34).

[96] G. Thorgilsson, G. Viktorsson, and S. Erlingsson. "Recursive Greens function method for multi-terminal nanostructures". In: *Journal of Computational Physics* 261 (Mar. 2014), pp. 256–266. DOI: 10.1016/j.jcp.2013.12.054 (cit. on p. 34).

[97] D. E. Petersen, H. H. B. Srensen, P. C. Hansen, S. Skelboe, and K. Stokbro. "Block tridiagonal matrix inversion and fast transmission calculations". In: *Journal of Computational Physics* 227.6 (Mar. 2008), pp. 3174–3190. DOI: 10.1016/j.jcp.2007.11.035 (cit. on pp. 34, 49).

[98] K. Takahashi. "Formation of Sparse Bus Impedance Matrix and Its Application to Short Circuit Study". In: *Proc. PICA Conference* June (1973) (cit. on p. 35).

[99] P. Amestoy, I. Duff, and J.-Y. L'Excellent. "Multifrontal parallel distributed symmetric and unsymmetric solvers". In: *Computer Methods in Applied Mechanics and Engineering* 184.2-4 (Apr. 2000), pp. 501–520. DOI: 10.1016/S0045-7825(99)00242-X (cit. on pp. 35, 42, 45).

[100] P. R. Amestoy, I. S. Duff, J.-Y. L'Excellent, and J. Koster. "A Fully Asynchronous Multifrontal Solver Using Distributed Dynamic Scheduling". In: *SIAM Journal on Matrix Analysis and Applications* 23.1 (Jan. 2001), pp. 15–41. DOI: 10.1137/S0895479899358194 (cit. on pp. 35, 42, 45).

[101] S. Li, S. Ahmed, G. Klimeck, and E. Darve. "Computing entries of the inverse of a sparse matrix using the FIND algorithm". In: *Journal of Computational Physics* 227.22 (Nov. 2008), pp. 9408–9427. DOI: 10.1016/j.jcp.2008.06.033 (cit. on p. 35).

[102] L. Lin, C. Yang, J. Lu, L. Ying, and W. E. "A Fast Parallel Algorithm for Selected Inversion of Structured Sparse Matrices with Application to 2D Electronic Structure Calculations". In: *SIAM Journal on Scientific Computing* 33.3 (Jan. 2011), pp. 1329–1351. DOI: 10.1137/09077432X (cit. on pp. 35, 42, 45).

[103] L. Lin, C. Yang, J. C. Meza, J. Lu, L. Ying, and W. E. "SelInv—An Algorithm for Selected Inversion of a Sparse Symmetric Matrix". In: *ACM Transactions on Mathematical Software* 37.4 (Feb. 2011), pp. 1–19. DOI: 10.1145/1916461.1916464 (cit. on p. 35).

[104] U. Hetmaniuk, Y. Zhao, and M. Anantram. "A nested dissection approach to modeling transport in nanodevices: Algorithms and applications". In: *International Journal for Numerical Methods in Engineering* 95.7 (Aug. 2013), pp. 587–607. DOI: 10.1002/nme.4518 (cit. on p. 35).

[105] M. G. Reuter and J. C. Hill. "An efficient, block-by-block algorithm for inverting a block tridiagonal, nearly block Toeplitz matrix". In: *Computational Science & Discovery* 5.1 (July 2012), p. 014009. DOI: 10.1088/1749-4699/5/1/014009 (cit. on pp. 35, 36).

[106] D. Mamaluy, D. Vasileska, M. Sabathil, T. Zibold, and P. Vogl. "Contact block reduction method for ballistic transport and carrier densities of open nanostructures". In: *Physical Review B* 71.24 (June 2005), p. 245321. DOI: 10.1103/PhysRevB.71.245321 (cit. on p. 38).

[107] D. J. Mason, D. Prendergast, J. B. Neaton, and E. J. Heller. "Algorithm for efficient elastic transport calculations for arbitrary device geometries". In: *Physical Review B* 84.15 (Oct. 2011), p. 155401. DOI: 10.1103/PhysRevB.84.155401 (cit. on pp. 38, 95).

[108] E. Cuthill and J. McKee. "Reducing the bandwidth of sparse symmetric matrices". In: *ACM Proceedings of the 1969 24th national conference*. New York, New York, USA: ACM Press, 1969, pp. 157–172. DOI: 10.1145/800195.805928 (cit. on pp. 38, 95).

[109] N. Gibbs, J. Poole, W.G, and P. Stockmeyer. "An algorithm for reducing the bandwidth and profile of a sparse matrix". In: *SIAM Numerical Analysis* 2 (1976), pp. 236–250 (cit. on pp. 38, 39, 95).

[110] Q. Wang, Y.-C. Guo, and X.-W. Shi. "A generalized GPS algorithm for reducing the bandwidth and profile of a sparse matrix". In: *Progress In Electromagnetics Research* 90 (2009), pp. 121–139 (cit. on pp. 38, 95).

[111] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide*. Third. Society for Industrial and Applied Mathematics, 1999 (cit. on p. 41).

[112] J. W. Demmel, S. C. Eisenstat, J. R. Gilbert, X. S. Li, and J. W. H. Liu. "A Supernodal Approach to Sparse Partial Pivoting". In: *SIAM Journal on Matrix Analysis and Applications* 20.3 (Jan. 1999), pp. 720–755. DOI: 10.1137/S0895479895291765 (cit. on p. 42).

[113] P. R. Amestoy, A. Guermouche, J.-Y. Lxcellent, and S. Pralet. "Hybrid scheduling for the parallel solution of linear systems". In: *Parallel Computing* 32.2 (Feb. 2006), pp. 136–156. DOI: 10.1016/j.parco.2005.07.004 (cit. on p. 42).

[114] B. A. Lippmann and J. Schwinger. "Variational principles for scattering processes. I". In: *Physical Review* 79.3 (1950), pp. 469–480. DOI: 10.1103/PhysRev.79.469 (cit. on p. 47).

[115] M. Bttiker. "Four-Terminal Phase-Coherent Conductance". In: *Physical Review Letters* 57.14 (Oct. 1986), pp. 1761–1764. DOI: 10.1103/PhysRevLett.57.1761 (cit. on p. 47).

[116] M. Buttiker. "Symmetry of electrical conduction". In: *IBM Journal of Research and Development* 32.3 (May 1988), pp. 317–334. DOI: 10.1147/rd.323.0317 (cit. on p. 47).

[117] M. Buttiker. "Coherent and sequential tunneling in series barriers". In: *IBM Journal of Research and Development* 32.1 (Jan. 1988), pp. 63–75. DOI: 10.1147/rd.321.0063 (cit. on p. 47).

[118] A. Levy Yeyati and M. Bttiker. "Scattering phases in quantum dots: An analysis based on lattice models". In: *Physical Review B* 62.11 (Sept. 2000), pp. 7307–7315. DOI: 10.1103/PhysRevB.62.7307 (cit. on p. 47).

[119] J. Ferrer, C. J. Lambert, V. M. Garca-Surez, D. Z. Manrique, D. Visontai, L. Oroszlany, R. Rodrguez-Ferrads, I. Grace, S. W. D. Bailey, K. Gillemot, H. Sadeghi, and L. a. Algharagholy. "GOLLUM: a next-generation simulation tool for electron, thermal and spin transport". In: *New Journal of Physics* 16.9 (2014), p. 093029. DOI: 10.1088/1367-2630/16/9/093029 (cit. on p. 47).

[120] B. G. Cook, P. Dignard, and K. Varga. "Calculation of electron transport in multiterminal systems using complex absorbing potentials". In: *Physical Review B* 83.20 (May 2011), p. 205105. DOI: 10.1103/PhysRevB.83.205105 (cit. on p. 47).

[121] M. Brandbyge, N. Kobayashi, and M. Tsukada. "Conduction channels at finite bias in single-atom gold contacts". In: *Phys. Rev. B* 60.24 (Dec. 1999), pp. 17064–17070. DOI: 10.1103/PhysRevB.60.17064 (cit. on p. 51).

[122] M. Paulsson and M. Brandbyge. "Transmission eigenchannels from nonequilibrium Green's functions". In: *Physical Review B* 76.11 (Sept. 2007), pp. 1–7. DOI: 10.1103/PhysRevB.76.115117 (cit. on p. 51).

[123] H. Nakamura. "First-Principles Study for Detection of Inelastic Electron Transport in Molecular Junctions by Internal Substitution". In: *The Journal of Physical Chemistry C* 114.28 (July 2010), pp. 12280–12289. DOI: 10.1021/jp9114223 (cit. on p. 53).

[124] A. A. Stuchebrukhov. "Tunneling currents in proteins: Nonorthogonal atomic basis sets and Mulliken population analysis". In: *The Journal of Chemical Physics* 107.16 (Oct. 1997), pp. 6495–6498. DOI: 10.1063/1.474308 (cit. on p. 53).

[125] P.-O. Lwdin. "On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals". In: *The Journal of Chemical Physics* 18.3 (1950), pp. 365–375. DOI: 10.1063/1.1747632 (cit. on pp. 53, 55).

[126] A. A. Stuchebrukhov. "Tunneling currents in electron transfer reaction in proteins. II. Calculation of electronic superexchange matrix element and tunneling currents using nonorthogonal basis sets". In: *The Journal of Chemical Physics* 105.24 (1996), p. 10819. DOI: 10.1063/1.472890 (cit. on p. 53).

[127] K. Stokbro, J. Taylor, M. Brandbyge, J.-L. Mozos, and P. Ordejn. "Theoretical study of the nonlinear conductance of Di-thiol benzene coupled to Au(111) surfaces via thiol and thiolate bonds". In: *Computational Materials Science* 27.1-2 (Mar. 2003), pp. 151–160. DOI: 10.1016/S0927-0256(02)00439-1 (cit. on p. 55).

[128] R. H. M. Smit, Y. Noat, C. Untiedt, N. D. Lang, M. C. van Hemert, and J. M. van Ruitenbeek. "Measurement of the conductance of a hydrogen molecule." In: *Nature* 419.6910 (Oct. 2002), pp. 906–9. DOI: 10.1038/nature01103 (cit. on p. 55).

[129] J. Tomfohr and O. F. Sankey. "Theoretical analysis of electron transport through organic molecules." In: *The Journal of chemical physics* 120.3 (Jan. 2004), pp. 1542–54. DOI: 10.1063/1.1625911 (cit. on p. 55).

[130] L. Venkataraman, J. E. Klare, C. Nuckolls, M. S. Hybertsen, and M. L. Steigerwald. "Dependence of single-molecule junction conductance on molecular conformation." In: *Nature* 442.7105 (2006), pp. 904–907. DOI: 10.1038/nature05037 (cit. on p. 55).

[131] S. Sen and S. Chakrabarti. "How Important Is the Position of the Molecule between the Electrodes in Tuning Negative Differential Resistance Behavior?" In: *Journal of Physical Chemistry C* 112.39 (Oct. 2008), pp. 15537–15542. DOI: 10.1021/jp804754h (cit. on p. 55).

[132] H. Song, M. a. Reed, and T. Lee. "Single molecule electronic devices." In: *Advanced materials (Deerfield Beach, Fla.)* 23.14 (Apr. 2011), pp. 1583–1608. DOI: 10.1002/adma.201004291 (cit. on p. 55).

[133] T. Rangel, G.-M. Rignanese, and V. Olevano. "Can molecular projected density of states (PDOS) be systematically used in electronic conductance analysis?" In: *Beilstein Journal of Nanotechnology* 6 (June 2015), pp. 1247–1259. DOI: 10.3762/bjnano.6.128 (cit. on p. 56).

[134] **N. L. Schneider, N. Nel, N. P. Andersen, J. T. L, M. Brandbyge, J. Krger, and R. Berndt. "Spectroscopy of transmission resonances through a C60 junction". In: *Journal of Physics: Condensed Matter* 27.1 (2015), p. 015001. DOI: 10.1088/0953-8984/27/1/015001 (cit. on p. 57).**

[135] Y. Yin, Z. Cheng, L. Wang, K. Jin, and W. Wang. "Graphene, a material for high temperature devices–intrinsic carrier density, carrier drift velocity, and lattice energy." en. In: *Sci. Rep.* 4 (Jan. 2014), p. 5758. DOI: 10.1038/srep05758 (cit. on p. 61).

[136] H. Song, Y. Kim, Y. H. Jang, H. Jeong, M. A. Reed, and T. Lee. "Observation of molecular orbital gating". In: *Nature* 462.7276 (Dec. 2009), pp. 1039–1043. DOI: 10.1038/nature08639 (cit. on p. 61).

[137] E. J. G. Santos and E. Kaxiras. "Electric-field dependence of the effective dielectric constant in graphene". In: *Nano Letters* 13.3 (2013), pp. 898–902. DOI: 10.1021/n1303611v (cit. on pp. 61, 63).

[138] D. A. Ryndyk, J. Bundesmann, M.-H. Liu, and K. Richter. "Edge state effects in junctions with graphene electrodes". In: *Phys. Rev. B* 86.19 (Nov. 2012), p. 195425. DOI: 10.1103/PhysRevB.86.195425 (cit. on p. 61).

[139] J.-T. L, R. B. Christensen, G. Foti, T. Frederiksen, T. Gunst, and M. Brandbyge. "Efficient calculation of inelastic vibration signals in electron transport: Beyond the wide-band approximation". In: *Physical Review B* 89.8 (Feb. 2014), p. 081405. DOI: 10.1103/PhysRevB.89.081405 (cit. on pp. 61, 81–83).

[140] **N. Papior, T. Gunst, D. Stradi, and M. Brandbyge. "Manipulating the voltage drop in graphene nanojunctions using a gate potential". In: *Phys. Chem. Chem. Phys.* 18 (2016), pp. 1025–1031. DOI: 10.1039/C5CP04613K (cit. on pp. 61, 77).**

[141] G. E. Jellison, J. D. Hunn, and H. N. Lee. "Measurement of optical functions of highly oriented pyrolytic graphite in the visible". In: *Physical Review B* 76.8 (Aug. 2007), p. 085125. DOI: 10.1103/PhysRevB.76.085125 (cit. on p. 63).

[142] H. Min, B. Sahu, S. K. Banerjee, and A. H. MacDonald. "Ab initio theory of gate induced gaps in graphene bilayers". In: *Physical Review B* 75.15 (Apr. 2007), p. 155115. DOI: 10.1103/PhysRevB.75.155115 (cit. on p. 63).

[143] D. C. Elias, R. V. Gorbachev, A. S. Mayorov, S. V. Morozov, A. A. Zhukov, P. Blake, L. A. Ponomarenko, I. V. Grigorieva, K. S. Novoselov, F. Guinea, and A. K. Geim. "Dirac cones reshaped by interaction effects in suspended graphene". In: *Nature Physics* 7.9 (July 2011), pp. 701–704. DOI: 10.1038/nphys2049 (cit. on p. 63).

[144] Z. Fei, A. S. Rodin, G. O. Andreev, W. Bao, A. S. McLeod, M. Wagner, L. M. Zhang, Z. Zhao, M. Thiemens, G. Dominguez, M. M. Fogler, A. H. C. Neto, C. N. Lau, F. Keilmann, and D. N. Basov. "Gate-tuning of graphene plasmons revealed by infrared nano-imaging". In: *Nature* (June 2012). DOI: 10.1038/nature11253 (cit. on p. 63).

[145] Y. Wang, V. W. Brar, A. V. Shytov, Q. Wu, W. Regan, H.-Z. Tsai, A. Zettl, L. S. Levitov, and M. F. Crommie. "Mapping Dirac quasiparticles near a single Coulomb impurity on graphene". In: *Nature Physics* 8.9 (July 2012), pp. 653–657. DOI: 10.1038/nphys2379 (cit. on p. 63).

[146] B. Fallahazad, Y. Hao, K. Lee, S. Kim, R. S. Ruoff, and E. Tutuc. "Quantum Hall effect in Bernal stacked and twisted bilayer graphene grown on Cu by chemical vapor deposition". In: *Physical Review B* 85.20 (May 2012), p. 201408. DOI: 10.1103/PhysRevB.85.201408 (cit. on p. 63).

[147] S. Luryi. "Quantum capacitance devices". In: *Applied Physics Letters* 52 (1988), pp. 501–503. DOI: 10.1063/1.99649 (cit. on p. 63).

[148] S. Droscher, P. Roulleau, F. Molitor, P. Studerus, C. Stampfer, K. Ensslin, and T. Ihn. "Quantum capacitance and density of states of graphene". In: *Applied Physics Letters* 96.15 (2010), p. 152104. DOI: 10.1063/1.3391670. arXiv: 1001.4690v1 (cit. on p. 63).

[149] E. J. G. Santos and E. Kaxiras. "Electrically driven tuning of the dielectric constant in MoS2 layers." In: *ACS nano* 7.12 (Dec. 2013), pp. 10741–6. DOI: 10.1021/nn403738b (cit. on p. 63).

[150] F. Cervantes-Sodi, G. Csnyi, S. Piscanec, and A. C. Ferrari. "Edge-functionalized and substitutionally doped graphene nanoribbons: Electronic and spin properties". In: *Physical Review B* 77.16 (Apr. 2008), p. 165427. DOI: 10.1103/PhysRevB.77.165427 (cit. on pp. 65, 66).

[151] G. Cantele, Y. S. Lee, D. Ninno, and N. Marzari. "Spin channels in functionalized graphene nanoribbons". In: *Nano Letters* 9.10 (2009), pp. 3425–3429. DOI: 10.1021/n1901557x. arXiv: 0909.1253 (cit. on pp. 65, 66).

[152] D. Selli, M. Baldoni, A. Sgamellotti, and F. Mercuri. "Redox-switchable devices based on functionalized graphene nanoribbons". In: *Nanoscale* 4.4 (2012), p. 1350. DOI: 10.1039/c2nr11743f (cit. on pp. 65, 66).

[153] J. Martin, N. Akerman, G. Ulbricht, T. Lohmann, J. H. Smet, K. von Klitzing, and A. Yacoby. "Observation of electron-hole puddles in graphene using a scanning single-electron transistor". In: *Nature Physics* 4.2 (Feb. 2008), pp. 144–148. DOI: 10.1038/nphys781. arXiv: 0705.2180 (cit. on p. 67).

[154] A. K. M. Newaz, Y. S. Puzyrev, B. Wang, S. T. Pantelides, and K. I. Bolotin. "Probing charge scattering mechanisms in suspended graphene by varying its dielectric environment". en. In: *Nat Commun* 3 (Mar. 2012), p. 734. DOI: 10.1038/ncomms1740 (cit. on p. 67).

[155] K. Nakada and A. Ishii. "Migration of adatom adsorption on graphene using DFT calculation". In: *Solid State Communications* 151.1 (Jan. 2011), pp. 13–16. DOI: 10.1016/j.ssc.2010.10.036 (cit. on p. 70).

[156] P. Marconcini, A. Cresti, F. Triozon, G. Fiori, B. Biel, Y.-M. Niquet, M. Macucci, and S. Roche. "Atomistic Boron-Doped Graphene Field-Effect Transistors: A Route toward Unipolar Characteristics". In: *ACS Nano* 6.9 (Sept. 2012), pp. 7942–7947. DOI: 10.1021/nn3024046 (cit. on p. 70).

[157] A. Lpez-Bezanilla, F. Triozon, and S. Roche. "Chemical Functionalization Effects on Armchair Graphene Nanoribbon Transport". In: *Nano Letters* 9.7 (July 2009), pp. 2537–2541. DOI: 10.1021/n1900561x (cit. on p. 70).

[158] B. Biel, X. Blase, F. Triozon, and S. Roche. "Anomalous Doping Effects on Charge Transport in Graphene Nanoribbons". In: *Phys. Rev. Lett.* 102.9 (Mar. 2009), p. 096803. DOI: 10.1103/PhysRevLett.102.096803 (cit. on p. 70).

[159] K. W. Jacobsen, J. T. Falkenberg, N. Papior, P. Bggild, A.-P. Jauho, and M. Brandbyge. "All-graphene edge contacts: Electrical resistance of graphene T-junctions". In: *Carbon* (2016). DOI: 10.1016/j.carbon.2016.01.084 (cit. on pp. 73, 79).

[160] J. M. Soler and E. Anglada. "Optimal Fourier filtering of a function that is strictly confined within a sphere". In: *Computer Physics Communications* 180.7 (July 2009), pp. 1134–1136. DOI: 10.1016/j.cpc.2009.01.017 (cit. on p. 73).

[161] J. T. Falkenberg and M. Brandbyge. "Simple and efficient way of speeding up transmission calculations with k -point sampling". In: *Beilstein Journal of Nanotechnology* 6 (July 2015), pp. 1603–1608. DOI: 10.3762/bjnano.6.164. arXiv: 1505.03267 (cit. on p. 74).

[162] A. Fasolino, J. H. Los, and M. I. Katsnelson. "Intrinsic ripples in graphene". In: (Apr. 2007), p. 14. DOI: 10.1038/nmat2011. arXiv: 0704.1793 (cit. on p. 74).

[163] J. T. Rasmussen, T. Gunst, P. Bggild, A.-P. Jauho, and M. Brandbyge. "Electronic and transport properties of kinked graphene". In: *Beilstein Journal of Nanotechnology* 4 (Feb. 2013), pp. 103–110. DOI: 10.3762/bjnano.4.12 (cit. on p. 75).

[164] W. Humphrey, A. Dalke, and K. Schulten. "VMD: Visual molecular dynamics". In: *Journal of Molecular Graphics* 14.1 (1996), pp. 33–38. DOI: 10.1016/0263-7855(96)00018-5 (cit. on p. 77).

[165] Y. Zhang, V. W. Brar, F. Wang, C. Girit, Y. Yayon, M. Panlasigui, A. Zettl, and M. F. Crommie. "Giant phonon-induced conductance in scanning tunnelling spectroscopy of gate-tunable graphene". In: *Nature Physics* 4.8 (July 2008), pp. 627–630. DOI: 10.1038/nphys1022 (cit. on pp. 81, 83, 84, 86).

[166] L. S. Panchakarla, K. S. Subrahmanyam, S. K. Saha, A. Govindaraj, H. R. Krishnamurthy, U. V. Waghmare, and C. N. R. Rao. "Synthesis, Structure, and Properties of Boron- and Nitrogen-Doped Graphene". In: *Advanced Materials* (Aug. 2009), NA–NA. DOI: 10.1002/adma.200901285 (cit. on p. 81).

[167] H. Wang, T. Maiyalagan, and X. Wang. "Review on Recent Progress in Nitrogen-Doped Graphene: Synthesis, Characterization, and Its Potential Applications". In: *ACS Catalysis* 2.5 (May 2012), pp. 781–794. DOI: 10.1021/cs200652y (cit. on p. 81).

[168] **J. Lagoute, F. Joucken, V. Repain, Y. Tison, C. Chacon, A. Bellec, Y. Girard, R. Sporken, E. H. Conrad, F. Ducastelle, M. Palsgaard, N. P. Andersen, M. Brandbyge, and S. Rousset. "Giant tunnel-electron injection in nitrogen-doped graphene". In: *Physical Review B* 91.12 (2015), pp. 1–5. DOI: 10.1103/PhysRevB.91.125442 (cit. on pp. 81, 83, 84, 87).**

[169] T. O. Wehling, I. Grigorenko, A. I. Lichtenstein, and A. V. Balatsky. "Phonon-Mediated Tunneling into Graphene". In: *Physical Review Letters* 101.21 (Nov. 2008), p. 216803. DOI: 10.1103/PhysRevLett.101.216803 (cit. on pp. 81, 84).

[170] N. M. R. Peres, L. Yang, and S.-W. Tsai. "Local density of states and scanning tunneling currents in graphene". In: *New Journal of Physics* 11.9 (Sept. 2009), p. 095007. DOI: 10.1088/1367-2630/11/9/095007 (cit. on p. 81).

[171] **M. L. N. Palsgaard, N. P. Andersen, and M. Brandbyge. "Unravelling the role of inelastic tunneling into pristine and defected graphene". In: *Physical Review B* 91.12 (2015). DOI: 10.1103/PhysRevB.91.121403 (cit. on pp. 81, 83–87).**

[172] T. Frederiksen, M. Paulsson, M. Brandbyge, and A.-P. Jauho. "Inelastic transport theory from first principles: Methodology and application to nanoscale devices". In: *Physical Review B* 75.20 (May 2007), pp. 1–22. DOI: 10.1103/PhysRevB.75.205413 (cit. on pp. 81–84).

[173] M. Paulsson, T. Frederiksen, and M. Brandbyge. "Modeling inelastic phonon scattering in atomic- and molecular-wire junctions". In: *Physical Review B* 72.20 (Nov. 2005), pp. 1–4. DOI: 10.1103/PhysRevB.72.201101 (cit. on p. 81).

[174] V. W. Brar, Y. Zhang, Y. Yayon, T. Ohta, J. L. McChesney, A. Bostwick, E. Rotenberg, K. Horn, and M. F. Crommie. "Scanning tunneling spectroscopy of inhomogeneous electronic structure in monolayer and bilayer graphene on SiC". In: *Applied Physics Letters* 91.12 (2007), p. 122102. DOI: 10.1063/1.2771084 (cit. on p. 83).

[175] J. Cervenka, K. van de Ruit, and C. F. J. Flipse. "Giant inelastic tunnelling in epitaxial graphene mediated by localized states". In: *Physical Review B* 247.205403 (2010), pp. 2992–2996 (cit. on p. 84).

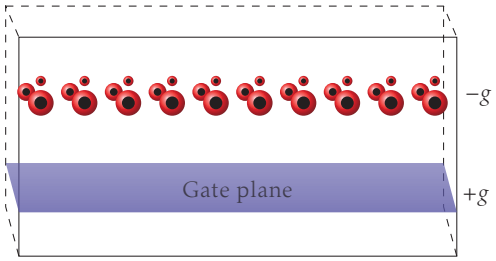
[176] M. Mohr, J. Maultzsch, E. Dobardi, S. Reich, I. Miloevi, M. Damnjanovi, A. Bosak, M. Krisch, and C. Thomsen. "Phonon dispersion of graphite by inelastic x-ray scattering". In: *Physical Review B* 76.3 (July 2007), p. 035439. DOI: 10.1103/PhysRevB.76.035439 (cit. on p. 84).

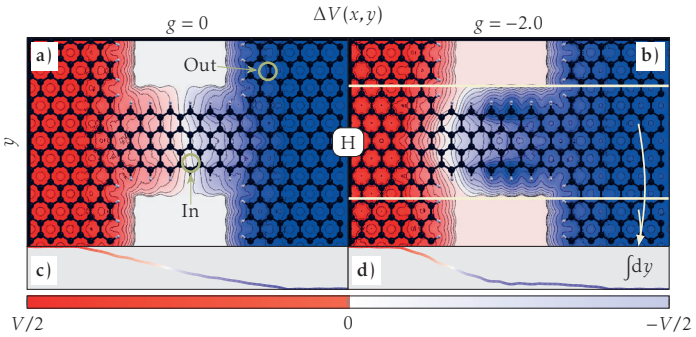
[177] R. Decker, Y. Wang, V. W. Brar, W. Regan, H. Z. Tsai, Q. Wu, W. Gannett, A. Zettl, and M. F. Crommie. "Local Electronic Properties of Graphene on a BN Substrate via Scanning Tunnelling Microscopy". In: *Nano Letters* 11 (2011), pp. 2291–2295 (cit. on p. 85).

[178] J. Ma, D. Alf, A. Michaelides, and E. Wang. "Stone-Wales defects in graphene and other planar sp2-bonded materials". In: *Physical Review B* 80.3 (July 2009), p. 033407. DOI: 10.1103/PhysRevB.80.033407 (cit. on p. 86).

[179] T. O. Wehling, M. I. Katsnelson, and A. I. Lichtenstein. "Impurities on graphene: Midgap states and migration barriers". In: *Physical Review B* 80.085428 (2009) (cit. on p. 87).

[180] M. Scheffler, D. Haberer, L. Petaccia, M. Farjam, R. Schlegel, D. Baumann, T. Hnke, A. Grneis, M. Knupfer, C. Hess, and B. Bchner. "Probing local hydrogen impurities in quasi-free-standing graphene". In: *ACS nano* 6.12 (Dec. 2012), pp. 10590–7. DOI: 10.1021/nn303485c (cit. on p. 87).





Published by:
DTU Nanotech
Department of Micro- and Nanotechnology
Technical University of Denmark
Ørsteds Plads, building 345C
DK-2800 Kgs. Lyngby