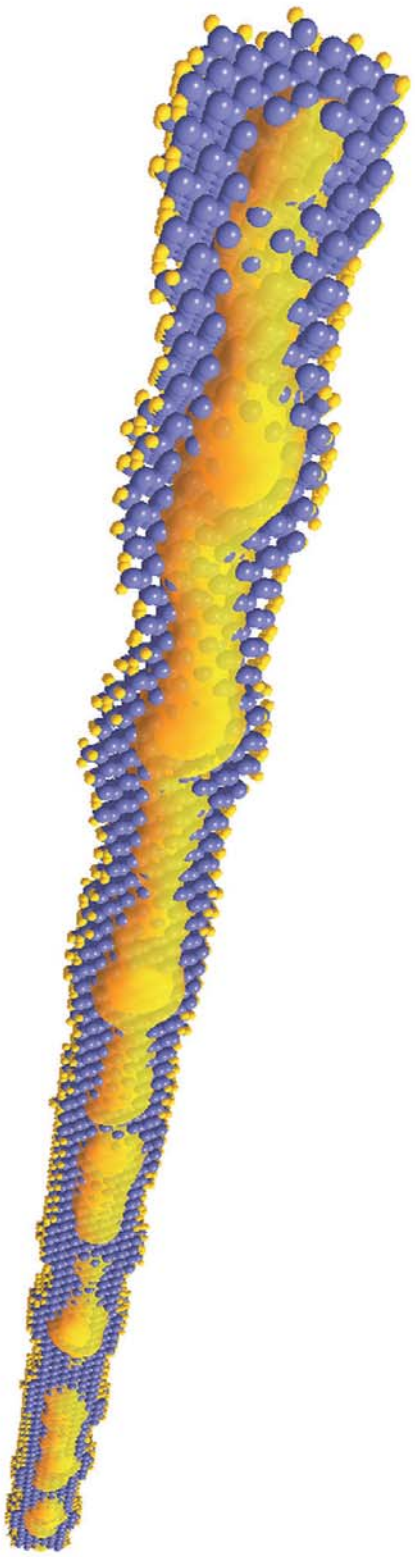
Report on multiscale approaches to modelling for nanotechnology



PHANTOMS foundation

cea

UNIVERSITÀ DI PISA

Report on multiscale approaches to modelling for nanotechnology

Editors:

M. Macucci, S. Roche, A. Correia

Contents

Introduction

M. Macucci
Dipartimento di Ingegneria dell'Informazione,
Università di Pisa,
Via Caruso 16, I-56122 Pisa, Italy

S. Roche
CEA/DSM/DRFMC/SPSMS,
17 avenue des Martyrs, 38054 Grenoble, France

A. Correia
Phantoms Foundation,
Pabellón C, Universidad Autónoma de Madrid,
Cantoblanco, 28049 Madrid, Spain

Nanoscience and Nanotechnology funding worldwide

Nanotechnology is now taking its first steps outside the laboratory, and it is demonstrating an immense potential for the manufacture of consumer goods, thus offering tangible and important perspectives for the economy [1] in the next few decades. Because of this expected economic impact, nanotechnology has aroused great interest among funding agencies, scientific policymakers, organizations, institutions and companies of the world's most developed countries.

Nanotechnology represents one of the fastest growing areas of R&D. In the period 1997–2005 worldwide investment in Nanotechnology research and development has exhibited an approximately ninefold increase, from $432 million to $4.2 billion [2]. This represents an average annual growth rate of 32%. The first set of coordinated efforts oriented to promote Nanotechnology at the national scale took place in the US in 1996, when several federal agencies launched the National Nanotechnology Initiative (NNI) [3] that has invested more than $2700 million in the 1997–2003 period to support long-term nanoscale R&D. In 2003 the US Government allocated $43.7 billion over a four year period to Nanotechnology R&D. Current estimates indicate that the global US Government spending for 2006 has been $6 billion. In addition to the federal initiatives, an important effort has been carried out at the level of the various US state governments, as well as within some major companies (Motorola, Intel, Hewlett-Packard, IBM, etc).

Industrialized Asian countries have promoted the development of Nanotechnology in the industrial and governmental sectors, with investment levels similar to those of the USA. Countries such as Taiwan and Korea have made a great effort to keep their current privileged positions in the control of Nanotechnology know-how. As an example, during the period 1997–2003 Japan has invested $2850 million on Nanotechnology promotion, an investment even higher than that of the USA. Korea launched the 10 Year Plan for the promotion of Nanotechnology in 2001 with about $1485 million committed — from the government ($983.5 million) and industry ($501.5 million) — aiming to enter the group of the top 5 nanotechnology countries in the world. Its ambitious program focuses on building infrastructure/facilities, strategic R&D and education. Taiwan launched its Nanotechnology program in 2002 and the funding is about $630 million over 6 years. China has also recently joined this race, making millionaire investments for the creation of several university-industry poles dedicated to Nanotechnology, emphasizing its leadership in nanomaterials.

Europe has intensively promoted Nanotechnology [4] within the VI [5, 6] and VII [7, 8, 9] Research Framework Programs (denoted as FP6 and FP7, respectively), through the thematic Area denominated "Nanotechnologies and nano-sciences, knowledge-based multifunctional materials and new production processes and devices" (NMP) that operated with a budget of 1300 million euros for the period 2003–2006 [5] and with the creation of the NID proactive initiative within the ICT (Information Communication Technologies) program [6]. In addition, other FP7 themes such as Health, Transport or Energy also launched funding calls directly related with Nanoscience & Nanotechnology [7]. There has been a boom of European initiatives dedicated to develop and popularize Nanotechnology, to the point of having today almost 200 national or regional networks. In the FP7, Nanotechnology maintains its outstanding role, as proved by the creation of the NanoICT proactive initiative (ICT) [6] or of the Technological Platforms, characterized by a strong industrial component, to drive the technological development in areas such as Nanoelectronics (ENIAC/EANEAS) [9]. The following table illustrates the estimated economic investment that nanotechnology has attracted around the world during the last decade.

Table 1: Estimated worldwide Nanotechnology funding (M$/year) in the 1997–2007 period.

	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Europe	126	151	179	200	225	400	650	950[4]	1050[4]		
Japan	120	135	157	245	465	750	810[3]	875[4]	950[4]		
USA[2]	116	190	255	270	422	604	862	989[4]	1200[4]	1351[5]	1392[5]
Others[6]	70	83	96	110	380	520	511[3]	900[4]	1000[4]		
TOTAL	432	559	687	825	1502	2274	2833	3714	4200		

[1] Source: Ref. [10].

[2] Do not include regional initiatives.

[3] Source: EU [4].

[4] Source: National Science Foundation (USA).

[5] Source: Ref. [11].

[6] Others: Australia, Korea, Canada, Taiwan, China, Russia, Singapore, Eastern Europe.

Nanomodelling Initiatives

Simulation of nanostructures and modelling of physical phenomena occurring at the nanoscale is crucial for the development of Nanotechnology. Emerging research areas such as Molecular Electronics, Biotechnology, Nanophotonics, Nanofluidics, Spintronics, and Quantum Computing could lead, in the mid-term future, to essential breakthroughs in information processing, in medicine, in biology and in manufacturing techniques. Modelling the characteristics of the nanodevices that would be at the heart of such innovations is therefore becoming more and more important and should make possible to: (i) provide a visual description of what happens inside a device, (ii) optimise the devices under study, (iii) improve the understanding of nanoscale properties (physical, chemical, biological). Awareness of the relevance of nano modelling has prompted the scientific community and funding agencies to promote several initiatives worldwide aimed at strengthening research in this area [12, 13].

In France, CEA has recently launched a new platform focused on "simulation tool developments for Nanosciences and Nanotechnologies". This programme **currently** gathers more than 40 permanent research staff with expertise in computer simulation for nanoelectronics, spintronics, molecular and organic electronics, or new energy sources. The platform is made up of three different centers of expertise, namely "AB INITIO HANDBOOK for Nanosciences", "NANOSCOPE" and "NANOSPECTROMETER" that respectively target various aspects of simulation including the development of state of the art *ab initio* calculations, molecular dynamics, and the study of physical properties (optics, transport, etc.). The first center focuses on setting up a database of material science information derived from first principles calculations (activation energies of dopants in semiconducting materials, thermodynamics of nanoscale structures, catalytic ability of small sized clusters). The "NANOSCOPE" center investigates

the dynamical phenomena at the nanoscale (dopant diffusion, interface and defect formation), whereas the "NANOSPECTROMETER" center tackles the development of efficient tools to simulate physical properties of novel materials (nanotubes, nanowires, graphene, organics, molecular systems) and related devices (field effect transistors, interconnects).

Other actions, such as the WEB-based initiative "Modelling for Nanotechnology" (M4nano) [12], led by four Spanish Institutions: Phantoms Foundation, Parque Científico de Madrid (PCM), Universidad Autónoma de Madrid (UAM) and Universidad Complutense, aims to maintain a continuous and structured flow of information among research groups, thereby preventing fragmentation of research efforts in Nanomodelling.

In the US, the nanoHUB [13] provides online simulation tools and was created by the NSF-funded Network for Computational Nanotechnology (NCN). This network has received in 2007 a five-year, $18.25 million grant from the National Science Foundation (NSF) to support the US National Nanotechnology Initiative with expanded capabilities and services for computer simulations.

Nanomodelling research overview

During the second half of the last century, the usage of computers for investigations of basic and technological issues in the physical sciences and engineering (as well as in various other fields, including economy, finance, weather forecasting, and the health sciences) evolved and transformed from being aimed primarily at enabling and accelerating certain time-consuming numerical manipulations to their present status, in which computer-based simulations serve as an indispensable and powerful tool of discovery, supplementing and complementing (more traditional) laboratory experiments and analytical theory as the pillars of scientific exploration [14]. New tools, algorithms and computational capabilities resulting from the increasing performance of supercomputers are envisioned to radically change the development of innovation in Information and Communication Technologies.

In the near future, the controlled and structured development of advanced numerical simulation tools and platforms could become a cornerstone of European excellence and leadership in emerging ICT innovation.

In addition, the engineering and integration of new low-dimensional materials (carbon nanotubes [15], semiconducting nanowires [16], quasi-1D organic compounds, two-dimensional graphene [17], molecular crystals [18], or biomolecules (DNA) and other molecular systems [19]), along with the mastering of quantum phenomena emerging at the nanoscale, increasingly demand for more realistic simulation of atomic-scale features of device components (material interfaces, chemical heterogeneity, conductance properties) as well as for a more sophisticated treatment of quantum physics (interactions, many-body effects, out-of-equilibrium phenomena), which will ultimately dominate any underlying device characteristics, the design of novel functionalities, and circuitry performance.

In particular, despite their successful contribution to ultimate MOSFET simulation, Monte Carlo approaches remain tremendously difficult to bridge to state of the art Density-Functional Theory (DFT) —based methods, whose ultimate developments allow higher accuracy in the description of atomistic scale

features and complex phenomena [20] and, in general, in full quantum approaches. Bringing *ab initio* capability to the device simulation framework will be essential in the quest for understanding the role of quantum phenomena and further achieving optimal control of such phenomena emerging at the nanoscale. For instance, the severe device characteristic fluctuations due to uncontrolled/unwanted distribution of dopants in the Si-channel and the effect of interfaces (dielectric layer) on charge carrier mobilities jeopardize the establishment of predicting computational tools (and compact SPICE-like models) essential for engineers for the proper design of circuit architectures for future nodes of advanced microelectronics [21]. Many novel materials such as carbon nanotubes, semiconducting nanowires, biomolecules (DNA), and two-dimensional graphene are full of promises, and seem to open unprecedented opportunities for applications and innovation in ICT (as envisioned in the BEYOND CMOS ITRS roadmap), but sophisticated computational frameworks are scarce and not versatile enough. Often, the fundamental and spectacular properties revealed by state of the art *ab initio* methods are not transferable into realistic device simulation, and the lack of well calibrated simplified models prohibits the study of essential statistical features such as device variability.

Innovation in ICT will also crucially need a deeper understanding of quantum phenomena involving spin (spintronics), phonons (thermal management, dissipation) and electromechanical degrees of freedom. Recent achievements in spintronics with carbon nanotubes [22] demonstrate the usefulness of sophisticated DFT-based computational approaches to achieve some predictability of novel material and device performances. More can be done if state of the art *ab initio* approaches are combined with advanced quantum transport methods and device simulation multiscale strategies.

The aim of the present report is to provide the reader with an overview of topics and techniques that are currently relevant in the field of nanodevices and nanocircuit modelling, with a specific emphasis on an aspect which is rapidly gaining importance, i.e. multi-scale modelling. It is our intention to highlight the most important open problems, the promising new routes for computational developments and underlying modelling needs of the nanoelectronics community.

Although new device concepts suitable to replace, as the mainstream technology, the CMOS workhorse, once its scaling limits will be reached, are yet to come, significant results have been achieved in terms of understanding of transport, noise, electronic structure and general physical properties of building blocks for innovative circuits. As a result of the development of optimized techniques and of the steep increase of the available computational power, atomistic simulations of relatively extended objects have become possible, and have extended the reach of first-principle simulation to functional structures. However, it is not conceivable to model complete circuits or even complete devices with a fully atomistic approach, due to the huge computational size of the resulting problem. It is therefore necessary to pursue an approach with multiple levels of approximation, adapted to the treatment of the different building blocks: from the most advanced and detailed atomistic treatment applied to a contact or small active region to semiclassical approximations for the description of parts of circuits made up of several interacting devices. This corresponds to a *multi-scale modelling* approach that defines a hierarchy of simulation tools analogous, from some points of view, to what is established in traditional microelectronics, with less refined models that are applied to higher-level structures exploiting parameters and functional relationships that are

derived with more refined approaches applied to simpler blocks.

The big challenge for this multi-scale hierarchy for nanoelectronics, compared to the existing tools for microelectronics, is represented by the much wider scope involved and by its intrinsically interdisciplinary nature, which includes not only many different branches of physics and engineering, but also chemistry and biology. Quite different approaches must be bridged in order to build a seamless hierarchy of simulation tools that can be applied effectively also by nonspecialists and that can actually support product development as nanoelectronic technologies are progressively transferred to industry.

Nanomodelling report overview

The present report intends to provide an overview, albeit limited and certainly not exhaustive, of relevant aspects of modelling at the nanoscale, pointing out some important issues that are still open and affording the reader that is not yet active in the field with an introduction to several widely used techniques and with a large body of references.

This review has been written by experts in the fields of computational modelling, most of them have strongly contributed to the development of European excellence in recent years, and have been leading EU-initiative over FP5, FP6 and FP7. Although more efforts will be needed to bridge different communities from *ab initio* development to device simulation, contributors of this report are overviewing promising methodologies to fill the gap between scientific communities, establishing some framework for further promoting European-based networking activities and coordination.

In Chapter 1 an approach for molecular-scale simulations based on first principles is presented, with a focus on the combination of the Density Functional Technique (DFT) with the Nonequilibrium Green's Function (NEGF) formalism. Such an approach has allowed a very successful treatment of conduction through single molecules, and the obtained results generally agree well at a qualitative level with experimental data.

Chapter 2 presents novel order N real space methods to investigate quantum charge transport within the frameworks of both Kubo-Greenwood and Landauer-Büttiker approaches. These order N methodologies allow for realistic modelling of carbon nanotubes, semiconducting nanowires and graphene-based material and devices. These computational schemes combine state of the art *ab initio* calculations with sophisticated semi-empirical models, and give quantitative access to the fundamental transport length scales (mean free path, charge mobilities, localization lengths) in realistic models of chemically modified (doped, functionalized) low-dimensional systems.

Chapter 3 deals with the integration of the Non Equilibrium Green's Function formalism with DFT (Density Functional Theory) for the investigation of charge and spin transport in carbon based nanodevices.

A relevant aspect of atomistic simulation is treated in Chapter 4, where analytical approximations to inter-atomic potentials are discussed, which are a simplification needed to handle structures containing a large number of atoms with semiclassical techniques. These new atomistic potential are of relevance for further exploration of fundamental and intrinsic properties of microelectronics materials.

Chapter 5 focuses on an introduction to the $k \cdot p$ method and its applications to the investigation of heterostructures. This technique represents an intermediate stage between effective mass approaches and fully atomistic descriptions, and is particularly useful for the treatment of nanostructures with a relatively large number of atoms, such as carbon nanotubes or the active area of semiconductor nanodevices.

An in-depth discussion of the application of Green's function techniques to interacting systems is presented in Chapter 6, with specific emphasis on systems operating in the Coulomb Blockade regime. This chapter presents state of the art advances in extending the conventional computational methods to include non equilibrium phenomena, and tackle with electron-phonon coupling in molecular systems or DNA.

Chapter 7 introduces the Büttiker formalism for the investigation of shot noise in nanostructures and details its application to a few examples, such as diffusive wires or chaotic cavities, starting from an effective mass approximation.

The development of efficient analytical approximations is instead the subject of Chapter 8, where analytical models suitable for fast numerical processing within computer algorithms are presented for the specific case of gated circular quantum dots.

Chapter 9 offers an outline of Quantum Monte Carlo Methods, which are essential for the determination of parameters typical of strongly interacting systems that can be transferred to levels of higher approximation, such as DFT.

Chapter 10 summarizes molecular dynamics techniques and their relevance in the framework of multiscale modelling.

Finally Chapter 11 overviews the potential of a generalized image charge method for computing electric field effects in scanning probe microscopies. The versatility of the method allows for an analysis of the electrostatic problem as a function of the tip apex geometry, and can be easily extended to analyze multilayered and anisotropic samples.

Bibliography

[1] A. Correia, M. Pérez, J. J. Sáenz, and P. A. Serena, "Nanotechnology applications: a driving force for R&D investment", Physica Status Solidi (a) **204**, 1611 (2007).

[2] Marks & Clerk, "Nanotechnology", Report (2006).

[3] US National Nanotechnology Initiative (NNI): `http://www.nano.gov`

[4] "Some Figures about Nanotechnology R&D in Europe and Beyond", European Commission, Research DG (`http://cordis.europa.eu/nanotechnology`).

[5] UE FP6 NMP Program: `http://www.cordis.lu/nmp/home.html`

[6] UE FP6 "NID" Proactive Initiative: `http://cordis.europa.eu/ist/fet/nid.htm`
UE FP7 "Nanoscale ICT Devices and Systems" Proactive Initiative:
`http://cordis.europa.eu/fp7/ict/fet-proactive/nanoict_en.html`

[7] UE FP7: http://cordis.europa.eu/fp7/cooperation/nanotechnology_en.html

[8] EU FP7 Nanotechnology funding opportunities:
http://cordis.europa.eu/nanotechnology/src/eu_funding.htm

[9] EU FP7 Technological Platforms:
http://cordis.europa.eu/technology-platforms/home_en.html
European Nanoelectronics Initiative Advisory Council (ENIAC): http://www.eniac.eu/

[10] M. C. Roco, "International Perspective on Government Nanotechnology Funding in 2005", Journal of Nanoparticle Research **7**(6), 707 (2005).

[11] US National Nanotechnology Initiative (NNI) budget overview:
http://www.nano.gov/html/about/funding.html
See also NNI FY 2008 budget & highlights:
http://www.nano.gov/pdf/NNI_FY08_budget_summary-highlights.pdf

[12] M4nano Initiative: http://www.m4nano.

[13] nanoHUB: http://www.nanohub.org

[14] U. Landman, "Materials by numbers: Computations as tools of discovery", PNAS **102**, 6671 (2005).

[15] J. C. Charlier, X. Blase, and S. Roche, "Electronic and transport properties of nanotubes", Review of Modern Physics **79**, 677 (2007).

[16] R. Beckman, E. Johnston-Halperin, Yi Luo,J. E. Green, and J. R. Heath, "Bridging Dimensions:Demultiplexing Ultrahigh-Density Nanowire Circuits", Science **310**, 465 (2005);
R. S. Friedman, M. C. McAlpine, D. S. Ricketts, D. Ham, and C. M. Lieber, "High-speed integrated nanowire circuits", Nature **434**, 1085 (2005);
L. Samuelson, "Self-forming nanoscale devices", Materials Today **6**, 22 (2003).

[17] K. 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", Science **306**, 666 (2004);
C. Berger, Z. Song, Tianbo Li, Xuebin Li, A. Y. Ogbazghi, Rui Feng, Zhenting Dai, A. N. Marchenkov, E. H. Conrad, P. N. First, and W. A. de Heer, "Ultrathin epitaxial graphite: 2D electron gas properties and a route toward graphene-based nanoelectronics", J. Phys. Chem. B **108**, 19912 (2004);
M. C. Lemme, T. J. Echtermeyer, M. Baus, and H. Kurz, "A Graphene Field Effect Device", IEEE Electron Device Letters **28**(4), 282 (2007).

[18] J. L. Bredas, J. P. Calbert, D. A. da Silva Filho, and J. Cornil, "Organic semiconductors: A theoretical haracterization of the basic parameters governing charge transport", PNAS **99**, 5804 (2002).

[19] G. Cuniberti, G. Fagas, and K. Richter, "Introducing Molecular Electronics", Lectures Notes in Physics **680** (Springer Verlag, Berlin/Heidelberg, 2005).

[20] M. A. L. Marques, C. A. Ulrich, F. Nogueira, A. Rubio, K. Burke, and E. K. U. Gross, "Time-Dependent Density Functional Theory", Lectures Notes in Physics **706** (Springer Verlag, Berlin/Heidelberg, 2006)

[21] A. Asenov, G. Roy, F. Adamu-Lema, A. R. Brown, S. Roy, and A. Asenov, "Intrinsic Parameter Fluctuations in Conventional MOSFETs until the End of the ITRS: A Statistical Simulation Study", Journal of Physics: Conference Series, Vol. **38**, pp.188 (2006).

[22] L. E. Hueso, J. M. Pruneda, V. Ferrari, G. Burnell, J. P. Valdés-Herrera, B. D. Simons, P. B. Littlewood, E. Artacho, A. Fert, and N. D. Mathur, Transformation of spin information into large electrical signals using carbon nanotubes, Nature **445**, 410 (2007).

Chapter 1

First-principles approach for the calculation of electronic transport at the molecular scale

M. Brandbyge, T. Frederiksen, M. Paulsson
Department of Micro and Nanotechnology, NanoDTU,
Technical University of Denmark,
Ørsteds Plads, Bldg. 345E, DK-2800 Lyngby, Denmark

Abstract. A brief introduction to atomistic modeling of electronic transport properties of nano-devices is given. We highlight a popular approach which combine density functional theory and non-equilibrium Green's functions, and show applications of it to elastic transport, as well as inelastic transport due to phonon scattering. An approximate scheme which enables calculations of the inelastic transport for relatively large systems is described.

1.1 Introduction

Reliable modeling of electron transport at the nanoscale is a field of increasing importance, both due to the downscaling of the well-known semiconductor electronic technology, but also due to the great progress and interest in molecular based electronics. The density functional theory (DFT) approach offers an atomistic description of total energy properties of nanosystems without system specific adjustable parameters. Further, in combination with the non-equilibrium Green's function (NEGF) method it has recently become a popular approach to atomistic-based quantum transport in nanosystems [1, 2, 3, 4]. In fact, recently a Danish based software company has put a commercial implementation of this approach

on the market (`www.atomistix.com`).

The aim of this contribution is to give a brief introduction to the combined DFT-NEGF transport method applied to nanostructures based on an atomistic description. DFT offers a overall good description of chemical bonding and atomic structure, while at the same time it is possible to do calculations on systems involving hundreds to thousands of atoms. Typical transport systems are, due to the involvement of two or more electrodes connecting the actual device, often at least in this range of size. There is always a trade off when it comes to accuracy in the method describing the electronic structure and the size of the system. It is widely believed that DFT is a good starting point when it comes to balancing accuracy of method and system size for describing the atomic structure. On the other hand when it comes to transport properties it is clear that DFT is by construction targeting the total energy of the system and not transport. During the last couple of decades DFT has been used with increasing popularity to calculate transport and to establish for what systems and what transport properties DFT may yield reasonable results. It is furthermore possible to use DFT as a starting point for more involved electronic structure methods such as, time-dependent DFT [5], the GW approximation [6], or self-interaction corrected DFT [7]. In weakly coupled molecular conductors electron-electron interaction effects play a significant role. While some Coulomb blockade effects have been described using spin-density functional theory [8], the correlation effects are more complicated to treat. In this direction the addition of a Hubbard-like term on top of the DFT Hamiltonian have been used [9]. These more advanced developments may, on the other hand, lead to compromises in the system-size which can be undertaken. It is therefore interesting to investigate to what extent we may obtain good agreement and predictive power with the DFT-NEGF approach by comparison with available experimental data for various transport properties in and system types.

In the following we will briefly present the DFT-NEGF transport method, and application of this to both the elastic and inelastic transport due to vibrational interactions/phonons.

1.2 Elastic transport

More than two decades ago atomic contacts between free-electron (jellium) electrodes were investigated theoretically by Lang [10] using DFT combined with scattering theory, and Ferrer *et al.* [11] used the empirical one-orbital tight-binding method combined with NEGF. The use of the *combined* DFT and NEGF methods has in recent years gained popularity in the study of molecular-scale contacts. The strong coupling limit poses some challenges for the theoretical method on several points. Firstly, the coupling of the contact region to the electrodes can not be treated using perturbation theory since there are strong chemical bonds between all regions rather than tunnel barriers [12]. Secondly, there is from the beginning no obvious way to divide the system into two electrode parts like what is done in the tunnel perturbation approach, e.g., used for STM theory. One way to deal with the latter problem is to use an

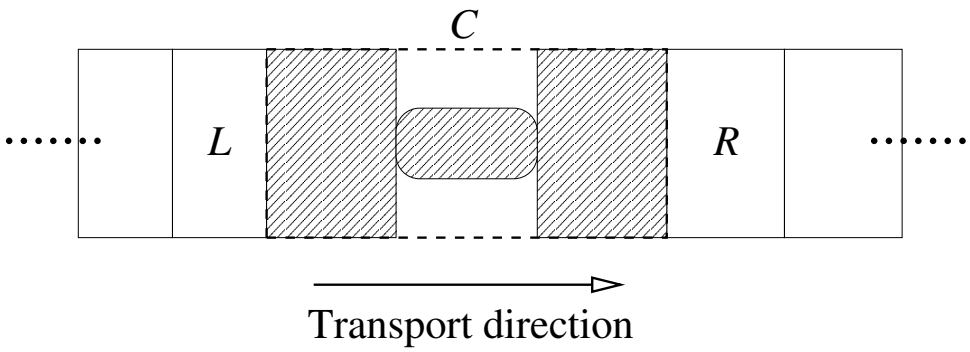


Figure 1.1: Schematic picture of the division of the system into a central contact (C), and ideal left (L), right (R) electrode regions.

atomic orbital basis-set[1] to expand the wave functions,

$$\psi_\mu(\mathbf{r}) = \sum_\alpha \phi_\alpha(\mathbf{r})\, c_\alpha^\mu. \tag{1.1}$$

When the basis functions $\phi_\alpha(\mathbf{r})$ has finite range in space this enables one to divide space (Hilbert space) into separate pieces chosen for computational convenience and calculate the electronic current between these pieces. Since this system is neither isolated nor periodic one can not use the normal periodic boundary conditions which are a standard ingredient in electronic structure calculations. This situation—combined with the non-perturbative nature—can be tackled within the DFT by calculating the exact scattering states. An alternative is to use NEGF [13, 14]. As with the scattering approach one can with the NEGF method obtain the electronic current, density and potential at a finite applied voltage. Furthermore the NEGF method facilitates the inclusion of interactions such as the electron-phonon coupling, as described later Sec. 1.3.

Now we will first briefly sketch the combined DFT-NEGF approach [1], and then go though some illustrative applications. Detailed reviews covering several related methods can be found in recent reviews, e.g., Refs. [15, 16, 17]. We consider the following system setup: Left semi-infinite electrode (L), contact region (C), and right semi-infinite electrode (R). We use a atomic orbital basis-set with a finite range as implemented in the SIESTA DFT code [18]. This basis enables us to split space into these regions.[2] The electrode regions L and R are chosen to have a perfect layer structure with a potential converged to the bulk values, that is, all disturbances in the C region is assumed to be screened out here. There is no orbital overlap or interaction *directly* between L and R regions. The set-up is illustrated in Fig. 1.1. In

[1]The orbital index α runs over all orbitals on all atoms in the system.

[2]The fact that these basis functions have an overlap $\langle\phi_\alpha|\phi_\beta\rangle = \mathbf{S}_{\alpha,\beta}$ has to be taken into account but does not alter the fundamental ideas. See, e.g., Ref. [19, 1, 20] for further details.

this case the Hamiltonian matrix takes the form,

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_L & \mathbf{V}_L & 0 \\ \mathbf{V}_L^\dagger & \mathbf{H}_C & \mathbf{V}_R \\ 0 & \mathbf{V}_R^\dagger & \mathbf{H}_R \end{pmatrix}, \tag{1.2}$$

where the $\mathbf{H}_R$ is a semi-infinite tri-diagonal matrix,

$$\mathbf{H}_R = \begin{pmatrix} \mathbf{h}_R & \mathbf{v}_R & 0 & \cdots & \\ \mathbf{v}_R^\dagger & \mathbf{h}_R & \mathbf{v}_R & 0 & \cdots \\ 0 & \mathbf{v}_R^\dagger & \mathbf{h}_R & \mathbf{v}_R & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}, \tag{1.3}$$

and like-wise for $\mathbf{H}_L$. The intra-layer ($\mathbf{h}$) and inter-layer ($\mathbf{v}$) Hamiltonians are identical to the corresponding Hamiltonians for the semi-infinitely repeated layer structure for the L or R electrodes and can be calculated once and for all using periodic boundary conditions. The electron density, $n(\mathbf{r})$, is obtained via the density-matrix, $\mathbf{D}_{\alpha\beta}$,

$$n(\mathbf{r}) = \sum_{\alpha,\beta} \phi_\alpha(\mathbf{r})\, \mathbf{D}_{\alpha\beta}\, \phi_\beta(\mathbf{r})\,, \tag{1.4}$$

which again is related to the retarded Green's function matrix, $\mathbf{G}$,

$$\mathbf{G}(E) = (E + i\delta - \mathbf{H})^{-1}\,, \tag{1.5}$$

as

$$\mathbf{D} = \int \frac{dE}{2\pi}\; n_F(E - E_F)\; i\left[\mathbf{G}(E) - \mathbf{G}^\dagger(E)\right]\,. \tag{1.6}$$

This latter equation is *only* true in *equilibrium* when no voltage is applied between left and right electrodes. We return to the *non-equilibrium* situation shortly.

We note that obtaining $\mathbf{G}$ involves the inversion of an *infinite* matrix which is not practical! On the other hand all interesting properties take place within the C region since here the electron density and potential differs from the bulk values in the electrodes (*i.e.*, matrix elements differs from the values in $\mathbf{h}_{L,R}$). So we basically want to consider finite matrices involving this region. For orbitals *inside* the C region we can in fact write this part of $\mathbf{G}$ exactly as an inversion of a *finite* matrix (which we just denote by $\mathbf{G}$ from now on),

$$\mathbf{G}(E) = \left[E + i\delta - \mathbf{H}_C - \mathbf{\Sigma}_L(E) - \mathbf{\Sigma}_R(E)\right]^{-1}\,, \tag{1.7}$$

where the so-called (one-electron) self-energies, $\mathbf{\Sigma}_{L,R}$, fully take into account the coupling of the C region to the L and R. The self-energies can be calculated exactly due to the perfect semi-infinite layer structure

of the electrodes, which translates into a tri-diagonal semi-infinite matrix we formally can write

$$\mathbf{H}_R = \begin{pmatrix} \mathbf{h}_R & \mathbf{v}_R \\ \mathbf{v}_R^\dagger & \mathbf{H}_R \end{pmatrix}, \tag{1.8}$$

We can obtain its inverse restricted to the first R unit-cell, $\mathbf{g}_R$, by an efficient iterative procedure [21]

$$(E + i\delta - \mathbf{H}_R)^{-1} = \begin{pmatrix} \mathbf{g}_R(E) & \cdots \\ \vdots & \ddots \end{pmatrix}, \tag{1.9}$$

where we do not specify the other elements (marked by "$\cdots$"). With this at hand, and similarly for L, we can solve the matrix equations (1.5) and (1.7) to find

$$\mathbf{\Sigma}_L(E) = \mathbf{V}_L^\dagger \mathbf{g}_L \mathbf{V}_L\,, \;\; \mathbf{\Sigma}_R(E) = \mathbf{V}_R^\dagger \mathbf{g}_R \mathbf{V}_R\,. \tag{1.10}$$

Roughly speaking, the real part of $\mathbf{\Sigma}_{L,R}$ describes the change in energy levels in region C due to the bond-formation with the L, R electrodes whereas the imaginary part is a current-operator describing the decay (inverse life-time) of electronic states located inside region C. This latter "escape-rate" is directly related to the electron transport and is denoted by,

$$\mathbf{\Gamma}_L(E) = i[\mathbf{\Sigma}_L(E) - \mathbf{\Sigma}_L^\dagger(E)]\,, \tag{1.11}$$

and likewise for R. Using these and Eq. (1.7) we can also write the density matrix which specify the density inside region C,

$$\mathbf{D} = \int \frac{dE}{2\pi}\; n_F(E - E_F)[(\mathbf{G}\mathbf{\Gamma}_L\mathbf{G}^\dagger)(E) + (\mathbf{G}\mathbf{\Gamma}_R\mathbf{G}^\dagger)(E)]\,, \tag{1.12}$$

where we have assumed all matrices (orbital indices) now due to the "self-energy trick" are restricted to the C region only and discarded the "CC" labels. One can show [1] that the first/second term in Eq. (1.12) correspond to the electron density in region C due to the filling of scattering states originating in the left/right electrode, that is,

$$[\mathbf{A}_L(E)]_{\alpha\beta} \equiv \left[(\mathbf{G}\mathbf{\Gamma}_L\mathbf{G}^\dagger)(E)\right]_{\alpha\beta} = \sum_l (c_\alpha^l)^* c_\beta^l\,, \tag{1.13}$$

where c_α^l are the orbital coefficients of the scattering states[3] at energy E originating from electrode L labeled by l, and similarly for the right.

In equilibrium these scattering states are filled to the common Fermi level $E_F^L = E_F^R = E_F$. Out of equilibrium this is not the case: In this case there is a voltage drop and a difference in the filling of the

[3]The scattering states are flux-normalized and divided by $\hbar$.

scattering states and the two terms in Eq. (1.12) acquire a different Fermi function,

$$\mathbf{D} = \int \frac{dE}{2\pi}\, n_F(E - E_F^L)\mathbf{A}_L(E) \,+\, n_F(E - E_F^R)\mathbf{A}_R(E)\,. \tag{1.14}$$

In the language of NEGF this is written in terms of the so-called "lesser" or electron distribution Green's function [13, 14],

$$\mathbf{D} = -i \int \frac{dE}{2\pi}\, \mathbf{G}^<(E)\ ,\ \ \mathbf{G}^<_{\alpha\beta}(t) = i\langle \hat{c}^\dagger_\alpha(t)\hat{c}_\beta\rangle. \tag{1.15}$$

The two electrodes are assumed still to be described by their bulk quantities as in Eq. (1.3) except for a constant shift in the potential-zero and Fermi energy. Thus it is assumed that the current has spread out and the change in density due to the non-equilibrium is screened in the L and R regions. In the DFT method we calculate the electronic density and potential in a self-consistent cycle but now using the expression in Eq. (1.14) for the electron density inside C. This enables us to calculate the voltage-drop in a current-carrying device, which in the case of a strong contact is a non-trivial quantity which does not only include the response of the electrons due to the electric potential but also the effect of the non-equilibrium filling (the current).

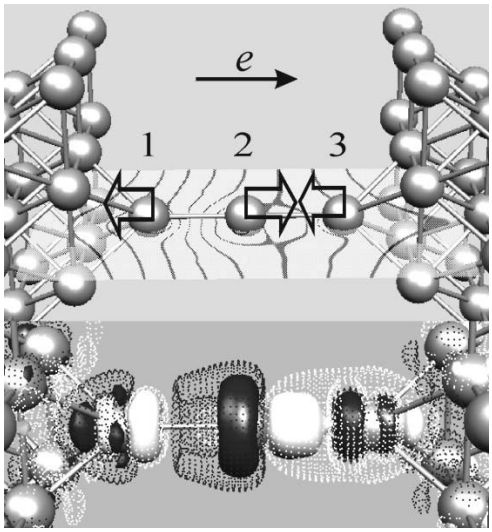


Figure 1.2: Atomic gold wire connecting (100) electrodes. (a) The contours indicate the voltage drop *i.e.*change in one-electron potential, $\mathcal{V}$ from 0V to 1V ($\mathcal{V}_{1V}(\mathbf{r}) - \mathcal{V}_{0V}(\mathbf{r})$). The arrows indicate the direction of the forces on the atoms due to the non-equilibrium. (b) Iso-density surfaces for the *change* in density from 0V to 1V. Dark is deficit and white is extra electron density. The solid (dotted) surface correspond to $\pm 5{\cdot}10^{-4} e/\text{Å}^3$ ($\pm 2{\cdot}10^{-4} e/\text{Å}^3$). From Ref. [22].

An illustration of this is shown in Fig. 1.2 where an 3-atom long atomic gold wire is considered connecting two (100) electrodes with $E_F^L - E_F^R = 1$ eV. The voltage is not dropping linearly across the wire but mainly in the bond between atom 1 and 2. The reason for the asymmetry can be traced back to the fact that the wire loose electronic charge with the applied voltage and become slightly more positive with bias. The non-equilibrium situation changes the electron density in the atomic bonds and will lead to forces and structural changes. The voltage drop has been analyzed in calculations for molecular

conductors in e.g. [23, 24].

In the case of no interactions beyond the mean-field potential (*e.g.*electron-phonon cf. next section) the current can be cast in the Landauer-Büttiker [25] form,

$$I = \mathrm{G}_0 \int dE \, (n_F(E - E_F^L) - n_F(E - E_F^R)) \, \mathrm{Tr}[\mathbf{t}^\dagger \mathbf{t}](E) \tag{1.16}$$

$$= \mathrm{G}_0 \int dE \, (n_F(E - E_F^L) - n_F(E - E_F^R)) \, T_{\mathrm{tot}}(E) \, , \tag{1.17}$$

where $\mathrm{G}_0 = 2e^2/\mathrm{h}$ is the conductance quantum. The transmission amplitude matrix involves the current-operators to the left and right electrodes and the matrix elements of the retarded Green's function between orbitals connecting L and R in a Fisher-Lee type relation [26, 27],

$$\mathbf{t} = (\mathbf{\Gamma}_R)^{1/2} \, \mathbf{G} \, (\mathbf{\Gamma}_L)^{1/2} \, . \tag{1.18}$$

The transmission itself depends on the applied voltage through the change in the self-consistent potential landscape in the C region (*i.e.*the change in $\mathbf{H}_C$ in Eq. (1.2) with applied voltage) and the rigid potential shifts of the electrodes relative to each other. An equivalent and popular way to write the current is,

$$I = \mathrm{G}_0 \int dE \, (n_F(E - E_F^L) - n_F(E - E_F^R)) \, \mathrm{Tr}[\mathbf{\Gamma}_R \mathbf{A}_L] \, , \tag{1.19}$$

corresponding to finding the current into the right electrode using $\mathbf{\Gamma}_R$ of the scattering states originating from the left electrode described by $\mathbf{A}_L$.

1.2.0.1 Example: Atomic metal contacts

For atomic metal contacts the total transmission, $T_{\mathrm{tot}}(E)$ can take values greater than 1 corresponding to several transmitting channels. This can be investigated using the so-called transmission eigen-channels [28, 29],

$$T_{\mathrm{tot}}(E) = \sum_n \tau_n(E), \tag{1.20}$$

which correspond to a basis change of scattering states so that $\mathbf{t}$ becomes diagonal with values $0 \leq \tau_n \leq 1$ in the diagonal. By plotting the corresponding scattering states at a particular energy one can get information on what orbitals contribute to the conduction. As an example, we consider a platinum atomic wire in Fig. 1.3 where two highly transmitting channels exist for that particular atomic configuration. From the plots of these two channels it is qualitatively seen how the d-orbitals enter these: The first channel has an angular momentum $m = 0$ character and involves d_{z^2} orbitals on the wire atoms while the second channel involves higher angular momentum d orbitals.[4]

In the case of atomic gold wires the d-electrons do not participate significantly in the conduction which is mainly carried by a single channel consisting of zero angular momentum states around the wire axis (mainly $6s$). The $6s$ orbitals have a greater range and larger ss matrix elements compared to the d

[4]Note that the left-right symmetry is broken since we consider scattering states originating in the right electrode.

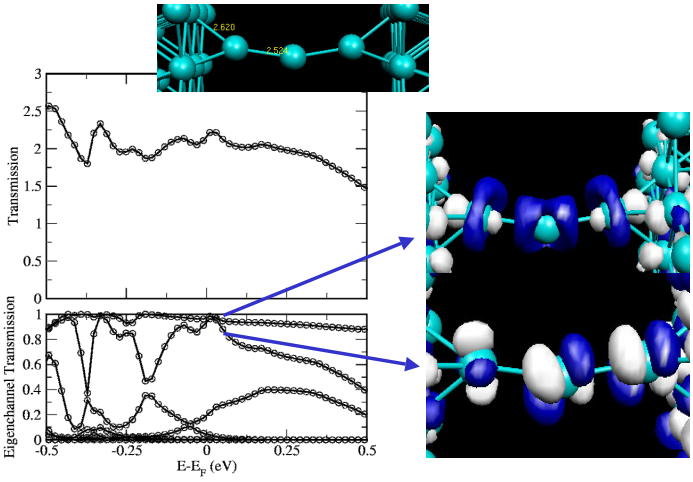


Figure 1.3: The total transmission and eigenchannel transmissions for a platinum wire (at zero voltage). The scattering states corresponding to the two highest transmitting channels at the Fermi energy are shown. These clearly involve d orbitals of different character (the color indicates the phase of the wavefunction). The channel involving d_{z^2} orbitals(upper) has least variation in eigenchannel transmission with energy.

orbitals. The effect of this can be seen in the big variation of the conductance with interatomic distances in Pt compared to Au [30], but also in the $I-V$ characteristics.

To illustrate this latter effect we consider in Fig. 1.4 a comparison between simplified gold and platinum single-atomic contacts. For Au the transmission at zero bias is dominated by a single, broad channel of mainly $6s$ character resulting from the strong coupling of these orbitals. For almost constant channel transmissions within the voltage window, and without a change of the transmission behavior with voltage, we anticipate from Eq. (1.20) and (1.17) a quite linear $I-V$. This turns out to be the case since we find that the change in channel behavior with bias below a couple of volts is very small for Au. This is in contrast to Pt. Here we find 4 channels with significant contributions and a much stronger variation with energy for zero bias. Note that the d_{yz}, d_{zx} channels are degenerate while the d_{xy} and $d_{x^2-y^2}$ channels are split due to the symmetry of the (100) electrodes. Additionally the s and d_{z^2} channel at certain energies split into two contributing channels. From the strong variation with energy and rich structure we do not anticipate a linear $I-V$ and furthermore we cannot expect the channel structure to be independent of bias: Indeed for finite bias we find that a significant change in the $\{d_{yz}, d_{zx}, d_{xy}, d_{x^2-y^2}\}$ derived channels which become less transmitting and downshifted in energy, whereas the broader s derived channel is not prone to the shift in potential.

Thus we find that the reason for the decrease in conductance with bias for Pt is due to the significant participation of the d electrons in the transport: The d electrons are more easily scattered by the voltage induced potential which in this case is of the same order of magnitude as the strength of the coupling to the electrodes. The same argument goes for the variation with atom-electrode distance. Here it is essentially the d contributions which decrease as the distance increases.

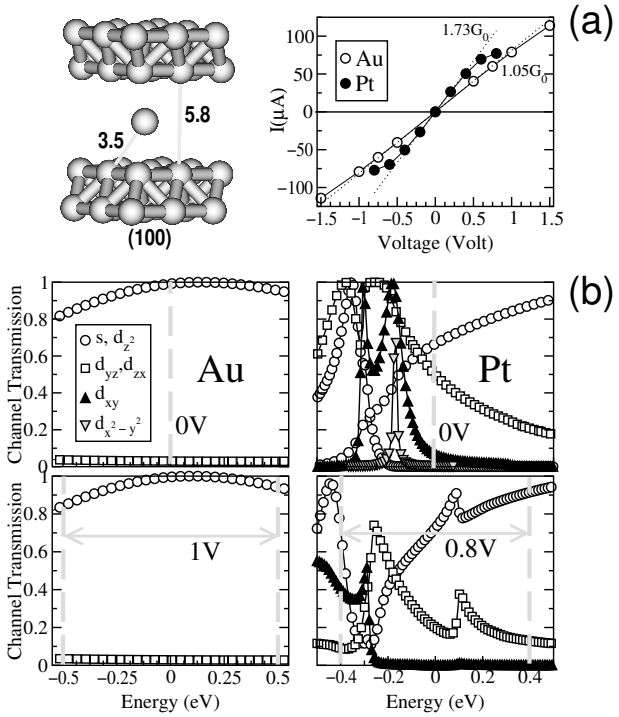


Figure 1.4: Comparison of the $I-V$ characteristics of a simplified Au and Pt single-atom contact. The z direction is defined to be perpendicular to the electrode surfaces. (a) Atomic structure (atom-surface distance shown in Å) of a single atom contact of Au and Pt between (100) electrodes and corresponding $I-V$. (b) Eigenchannel decomposition of the total transmission through the atoms for Au for 0 V and 1 V (left panels) and Pt for 0 V and 0.8 V (right panels). The channels are labeled by their main orbital components (z is the direction perpendicular to the electrode surfaces). The gray dashed lines indicate the voltage window. From Ref. [30].

1.3 Inelastic transport

Steered partly by the experimental findings there is now an increasing interest in the theoretical modeling of the electron-vibration[5] processes in electron transport. While there has already been many studies devoted to transport with phonon interaction based on model-type Hamiltonians emphasizing various aspects of the transport [31, 32, 33, 34, 35, 36], there has only been a hand-full of studies based on a first-principles description including all aspects of the e-ph transport problem within the same approach, such as atomic structure, vibrational modes and frequencies, and transport. The inelastic transport problem is conveniently formulated using the tight-binding description of the electronic structure discussed in the previous section. Within this formulation the transport can be based on electronic structure ranging from approximate parameterizations, such as the Hückel or extended Hückel [37, 38], to DFT and beyond. In the following we will give a few examples of work based on first-principles methods, which so far almost

[5]We will here use the term "phonon" also for localized quantum vibrations and use the term phonon and vibration on equal footing.

entirely has been DFT-based.

In the tunneling regime inelastic effects has been investigated by Scanning Tunneling Microscopy (STM) yielding atomic resolution of the inelastic tunneling process. The STM inelastic images been simulated theoretically by Lorente and Persson based on DFT and the Tersoff-Hamann approach [39, 40]. Also the controlled conformational changes, motion and chemistry induced by the inelastic tunnel current in STM has been addressed [41, 42, 43]. More recently the regime where a molecular/atomic-scale conductor is strongly coupled to both electrodes has been investigated. Chen and co-workers have studied inelastic scattering and local heating in an atomic gold contact, a thiol-bonded benzene [44, 45], and alkane-thiols [46]. The inelastic signals were calculated using a golden-rule-type of expression and the DFT scattering states where the systems are adsorbed on electrodes represented by a jellium [47]. Based on the self-consistent tight-binding procedure with parameters obtained from DFT Pecchia *et al.* [2] has considered on vibrational effects in octane-thiols bonded to gold electrodes [48] using NEGF and the Born approximation for the self-energy of the phonon interaction. Recently Solomon *et al.* used this method to simulate the experimental IETS spectra of Wang [49] for this system. Sergueev *et al.* studied a 1,4-benzenedithiolate molecule contacted by two aluminum leads [50]. This study addressed the bias dependence of the vibrational modes and e-ph coupling constants, which were calculated using NEGF-DFT method at finite bias. While the vibrational spectrum is almost unchanged, a significant increase was found for high biases (> 0.5 Volts). Finally, Troisi and Ratner has suggested an simplified approach from which IETS signals can be calculated approximately based on a calculation on an isolated molecule and neglecting the electrodes [51]. The hybrid DFT B3LYP method was used to describe vibrations and electronic structure. A related method has also been used by Jiang *et al.* [52].

We have also looked at atomic gold wires and hydrocarbon molecules between gold electrodes using DFT and we will give a brief introduction and a few examples in the next section.

1.3.1 Phonon interaction: Self-consistent Born Approximation (SCBA)

Again, the starting point is the system set-up and electronic Hamiltonian in Eq. (1.2). We employ the standard adiabatic approximation and consider the atomic positions as a parameter in the electronic Hamiltonian. We restrict the atomic motion to the central region C, use the harmonic approximation, and determine the normal vibrational modes (λ) and their frequencies (ω_λ) for selected atoms embedded in region C, denoted by the small displacements Q_λ. We can choose the C region large enough so only $\mathbf{H}_C$ will depend on the atomic motion around the equilibrium positions. To lowest order we have

$$\mathbf{H}_C(Q) \approx \mathbf{H}_C(0) + \sum_\lambda \frac{\partial \mathbf{H}_C}{\partial \mathbf{Q}^\lambda} \cdot \mathbf{Q}^\lambda, \tag{1.21}$$

or explicitly quantizing the harmonic motion,

$$\mathbf{H}_C(Q) \approx \mathbf{H}_C(0) + \sum_\lambda \mathbf{M}_\lambda (\hat{b}_\lambda^\dagger + \hat{b}_\lambda)\,, \tag{1.22}$$

where $\hat{b}_\lambda^\dagger$ is the phonon creation operator. The coupling matrices for each mode, $\mathbf{M}_\lambda$, are calculated in the atomic orbital basis set using finite differences [53]. Thus all modes and parameters $(\omega_\lambda, Q_\lambda, \mathbf{M}_\lambda)$, can be determined using DFT without using fitting parameters. The NEGF method is especially ideal for a systematic treatment of the non-equilibrium situation in combination with interactions beyond the mean-field approximation, *e.g.*as employed by DFT. It is possible to go to infinite order in the perturbing $\mathbf{M}_\lambda$ using the self-consistent Born approximation (SCBA) [14]. This amounts to solving the not only the retarded Green's function, $\mathbf{G}$, including the phonon interaction, but also solve for the Green's function related to the occupied states ($\mathbf{G}^<$) introduced in the previous section, and equivalently for the empty states ($\mathbf{G}^>$). Note that these are not simply related via a Fermi function due to the non-equilibrium situation. Since the emission or absorbtion of phonons depend on the filling of states, the phonon self-energies the scattering used to obtain the Green's functions will depend on the Green's functions themselves. Thus a self-consistent calculation is called for. For example, the filled states at energy E ($\mathbf{G}^<(E)$) depend on the scattering rate for scattering *into* these states due to the phonon interaction, as described by the "lesser" phonon self-energy:

$$\mathbf{\Sigma}_\lambda^<(E) = \mathbf{M}_\lambda \left[(N_\lambda + 1)\mathbf{G}^<(E + \hbar\omega_\lambda) + N_\lambda \mathbf{G}^<(E - \hbar\omega_\lambda)\right] \mathbf{M}_\lambda \,. \tag{1.23}$$

Here the first term correspond to phonon emission and involves thus the occupied states at $E + \hbar\omega$, while the second to phonon absorption, N_λ being the phonon occupation of mode λ. We will not go further into the formalism here, but just conclude that even the lowest order self-energy for the inelastic scattering yield much more demanding computations involving a grid of different energies which mix in the equations, and a self-consistency, as compared to the elastic case. More details on the theory, equations, and computation can be found in Ref. [54]. In the next section we discuss a simpler approximate approach to the inelastic transport valid in the weak e-ph coupling limit, which involve about the same amount of computation as the elastic conductance, once the phonons and phonon-couplings are determined.

1.3.1.1 Example: Atomic gold wire

As an example of the method we consider the application of it to the inelastic transport in 4-atom long gold wires just before rupture [55]. This is probably the most well-studied atomic-sized conductor, and a great deal of detailed information is available from experiments for comparison. For example it is known that the typical force mediated by the wire at fracture is 1.5 nN [56]. We consider in Fig. 1.5 two stages of strain corresponding to electrode separations of $L = 12.22$ Å and $L = 12.68$ Å corresponding to a force of about 0.5 nN and 1.5 nN. In the results displayed in Fig. 1.6 the phonon occupation is kept constant (almost zero) corresponding to the experimental temperature $T = 4.2$ K. This corresponds to a situation where the vibration is loosing all its energy obtained from the electronic current to an external source such as phonons in the electrodes.[6] The results of the calculation reproduce main features of the experiments: (i) A single main conductance drop is observed, (ii) the order of magnitude of the conductance drop, (iii)

[6]The damping is not included in the dynamics of the oscillator which is valid when the damping rate is much smaller than the vibrational frequency.

the mode softening, and (iv) the increased phonon signal with strain. A frequency shift with elongation ($\Delta\omega/\Delta L = -7$ meV/Å) corresponding to a softening of the bonds in the wire is also in accordance with the experiments.

The fact that only a single main drop is observed can be traced back to the symmetry of the electronic states at the Fermi level. These states are mainly composed of s orbitals (see Fig. 1.6) as discussed in the previous section. The symmetry dictates that this state mainly couple to longitudinal vibrations where the bond-length is alternating inside the wire.

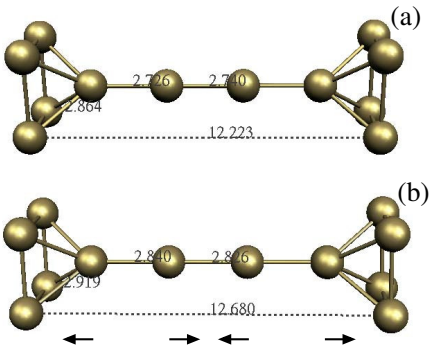


Figure 1.5: Geometry of a four atom gold wire under two different states of stress corresponding to an electrode separation of (a) $L = 12.22$Å and (b) $L = 12.68$Å. The gold electrodes are modelled by perfect (100) surfaces. The alternating bond length (ABL) modes, which cause the inelastic scattering, are shown schematically below each structure, together with mode energies ω_λ and reduced conductance drop $\Delta G/G(0\text{V})$. The displayed interatomic distances are measured in units Å.

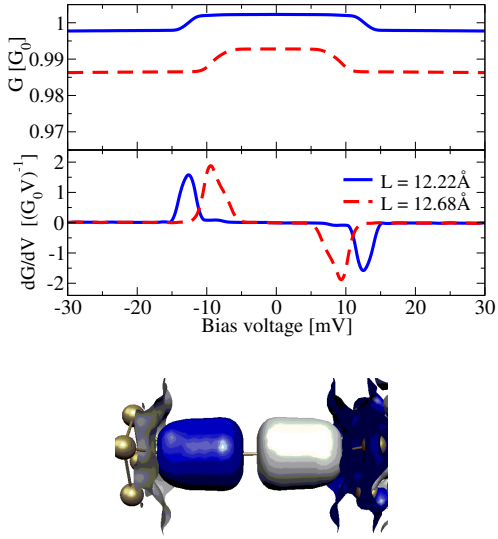


Figure 1.6: Differential conductance and its derivative for the four atom gold wire at two different tensions in the case where the oscillators are externally damped ($N_\lambda \approx 0$). All modes are included in this calculation. Below, the state at E_F responsible for the transport is shown (similar for both chain-lengths).

A closer look at the experimental results show a slight slope of the conductance versus voltage after the onset of the vibrational excitation [57]. This can be explained to be due to heating of the vibrations caused by the electronic excitation. We can take the opposite limit assuming that there is no external damping mechanism (e.g. phonons in the electrodes) keeping the occupation fixed. In this case the vibration can only loose energy to electronic excitations and the number of phonons is determined by the excitation and de-excitation due to the coupling to the electronic degrees of freedom. For a given bias voltage we can use the fact that the system is in a steady state and we require that the net power into the vibration must be zero. The power can be obtained using the NEGF as well [55]. This in turn puts a restriction on N_λ. For simplicity we include only the most important mode. The idea is illustrated in Fig. 1.7 where the power balance for different fixed mode occupations is shown. For low bias and finite phonon excitation we will have energy transfer from the phonon subsystem to the electrons - that is vibrational damping due to the electron-hole pair excitation. For increasing voltage each power curve crosses the abscissa at

one particular point (marked with a dark circle) corresponding to power balance between the electronic and phonon subsystems: Hence the crossing point sets the occupation at the corresponding bias voltage when we assume a steady state. At a voltage $V = 55$mV the occupation is found to be the same as *if* the mode was occupied according to a Bose-Einstein distribution with temperature $T = 300$K. The conductance calculation is shown in Fig. 1.8a. Compared with the externally damped results Fig. 1.6, the notable differences are a slightly larger drop as well as a finite slope in the conductance beyond the onset of inelastic scattering. This increase in backscattering with voltage beyond the threshold is simply due to the fact that the probability for emitting a phonon and thus backscattering increase with increasing number of phonons present.

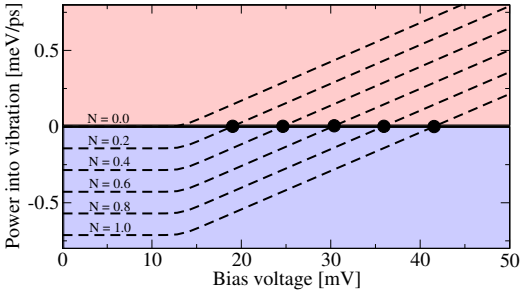


Figure 1.7: Net power transferred from the electrons to the phonon mode vs. bias voltage for different (fixed) occupations N. The data shown here for $L = 12.22$Å are representative for both geometries.

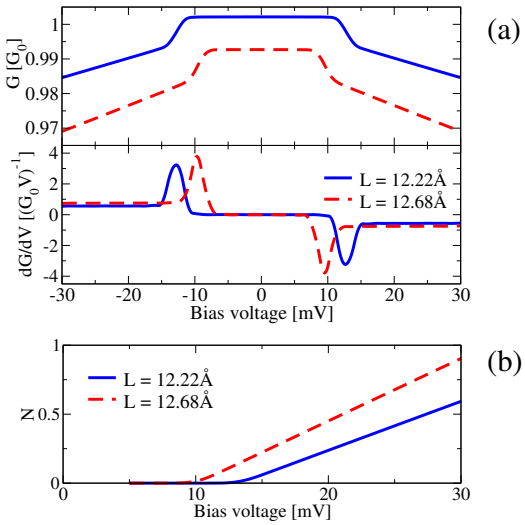


Figure 1.8: (a) Differential conductance and its derivative for the four atom gold wire at two different tensions in the externally undamped limit. Only the most important mode is included in this calculation. (b) Mode occupation N vs. bias voltage.

Quantitatively we find a slope (dG/dV) which is only slightly larger than detected for relatively long atomic gold wires. Including the damping due to the coupling to electrode phonons one can expect that the typical damping conditions lead to conductance curves in between Fig. 1.6 and 1.8a. However, measurements are the closest to the externally undamped limit, which suggests that such mechanisms are weak and that the mode "heating" a significant element.

1.3.2 Phonon interaction: Lowest order expansion (LOE)

As mentioned above, the solution of the SCBA equations is a daunting numerical task for systems consisting of more than a handful of atoms. However, for systems where the e-ph coupling is weak and the density of states (DOS) varies slowly with energy, we have previously derived the LOE approximation [58].

The LOE approximation assumes that the retarded and advanced single-particle Green's functions and electrode self-energies are energy *independent*. We can then expand the current and power expressions to the lowest order (second) in e-ph couplings $\mathbf{M}_\lambda$ and perform the energy integrations analytically. These integrals consist of products of Fermi-Dirac functions and their Hilbert transforms. The LOE thus retains the Pauli exclusion principle for fermionic particles, which is necessary to model the blocking of phonon emission processes at low bias.

The flow of energy or power can, just as the flow of particles, be obtained with the NEGF formalism. In the LOE approximation, the total power dissipated into the phonon system $P^{\mathrm{LOE}} \equiv P_L + P_R$ can, after lengthy derivations, be written [58]

$$P^{\mathrm{LOE}} = \sum_\lambda \frac{(\hbar\omega_\lambda)^2}{\pi\hbar}[n_B(\hbar\omega_\lambda) - N_\lambda]\,\mathrm{Tr}\left[\mathbf{M}_\lambda\mathbf{A}\mathbf{M}_\lambda\mathbf{A}\right] \tag{1.24}$$

$$+ \quad \mathcal{P}(V,\hbar\omega_\lambda,T)\,\mathrm{Tr}\left[\mathbf{M}_\lambda\mathbf{G}\mathbf{\Gamma}_L\mathbf{G}^\dagger\mathbf{M}_\lambda\mathbf{G}\mathbf{\Gamma}_R\mathbf{G}^\dagger\right], \tag{1.25}$$

$$\mathcal{P} = \frac{\hbar\omega}{\pi\hbar}\frac{[\cosh\left(\frac{eV}{\mathrm{k_B}T}\right)-1]\coth\left(\frac{\hbar\omega}{2\mathrm{k_B}T}\right)\hbar\omega - eV\sinh\left(\frac{eV}{\mathrm{k_B}T}\right)}{\cosh\left(\frac{\hbar\omega}{\mathrm{k_B}T}\right)-\cosh\left(\frac{eV}{\mathrm{k_B}T}\right)}, \tag{1.26}$$

where the Bose-Einstein distribution $n_B(\varepsilon)$ appears naturally from the integration of the Fermi-Dirac functions of the electrons in the contacts. Here $\mathbf{G}$, $\mathbf{\Gamma}_{L/R}$, and $\mathbf{A} = \mathbf{A}_L + \mathbf{A}_R$ are the non-interacting equilibrium quantities introduced in Sec. 1.2 and evaluated at the Fermi-level E_F. The first term in Eq. (1.25) describes the equilibrium energy exchange between the vibrational and electronic degrees of freedom (electron-hole damping of the vibrations [59]), while the second term includes the emission of vibrational quanta at finite bias. Furthermore, it can be shown that

$$\mathrm{Tr}\left[\mathbf{M}_\lambda\mathbf{A}\mathbf{M}_\lambda\mathbf{A}\right] \geq 2\mathrm{Tr}\left[\mathbf{M}_\lambda\mathbf{G}\mathbf{\Gamma}_L\mathbf{G}^\dagger\mathbf{M}_\lambda\mathbf{G}\mathbf{\Gamma}_R\mathbf{G}^\dagger\right], \tag{1.27}$$

which implies that the electron-hole damping constant [rate] is at least twice the emission constant [rate] of vibrations, i.e., there exists an upper bound on the mode occupation N_λ [60]. For $eV \gg \hbar\omega$ this bound can be expressed as

$$N_\lambda \leq (eV/\hbar\omega_\lambda - 1)/2. \tag{1.28}$$

The LOE approximations, which above was applied to the power, also allow us to write the current

through the device I^{LOE} as [58, 61]

$$
\begin{aligned}
I^{\mathrm{LOE}} &= \mathrm{G}_0 V \mathrm{Tr}\left[\mathbf{G}\mathbf{\Gamma}_R\mathbf{G}^\dagger\mathbf{\Gamma}_L\right] && (1.29)\\
&\quad + \sum_\lambda \mathcal{I}^{\mathrm{sym}}(V,\hbar\omega_\lambda,T,N_\lambda)\\
&\qquad \times \mathrm{Tr}\left[\mathbf{G}^\dagger\mathbf{\Gamma}_L\mathbf{G}\left\{\mathbf{M}_\lambda\mathbf{G}\mathbf{\Gamma}_R\mathbf{G}^\dagger\mathbf{M}_\lambda + \frac{i}{2}\left(\mathbf{\Gamma}_R\mathbf{G}^\dagger\mathbf{M}_\lambda\mathbf{A}\mathbf{M}_\lambda - \mathrm{h.c.}\right)\right\}\right]\\
&\quad + \sum_\lambda \mathcal{I}^{\mathrm{asym}}(V,\hbar\omega_\lambda,T)\\
&\qquad \times \mathrm{Tr}\left[\mathbf{G}^\dagger\mathbf{\Gamma}_L\mathbf{G}\left\{\mathbf{\Gamma}_R\mathbf{G}^\dagger\mathbf{M}_\lambda\mathbf{G}\left(\mathbf{\Gamma}_R - \mathbf{\Gamma}_L\right)\mathbf{G}^\dagger\mathbf{M}_\lambda + \mathrm{h.c.}\right\}\right], && (1.30)\\
\mathcal{I}^{\mathrm{sym}} &= \frac{e}{\pi\hbar}\left(2eV\langle n_\lambda\rangle + \frac{\hbar\omega_\lambda - eV}{e^{\frac{\hbar\omega_\lambda - eV}{\mathrm{k_B}T}} - 1} - \frac{\hbar\omega_\lambda + eV}{e^{\frac{\hbar\omega_\lambda + eV}{\mathrm{k_B}T}} - 1}\right), && (1.31)\\
\mathcal{I}^{\mathrm{asym}} &= \frac{e}{2\pi\hbar}\int_{-\infty}^{\infty}\left[n_F(\varepsilon) - n_F(\varepsilon - eV)\right] && (1.32)\\
&\qquad \times \mathcal{H}_{\varepsilon'}\{n_F(\varepsilon' + \hbar\omega_\lambda) - n_F(\varepsilon' - \hbar\omega_\lambda)\}(\varepsilon)d\varepsilon,
\end{aligned}
$$

where the bias is defined via $eV = E_F^R - E_F^L$. The Hilbert transform is defined as $\mathcal{H}_x\{f\}(y) = 1/\pi\mathcal{P}\int_{-\infty}^{\infty} dx f(x)/(x-y)$. The above expression is current conserving, i.e., calculating the current at the left and right contacts give the same result.

The LOE expression Eq. (1.30) for the current contains three terms, (i) the Landauer-Büttiker term corresponding to the elastic conductance, (ii) the "symmetric" term to symmetric conductance steps at the vibrational energies, and (iii) the "asymmetric" term to peaks and dips in the conductance which are asymmetric with voltage inversion, see Fig. 1.9. For geometrically symmetric junctions, it can be shown that the asymmetric term vanishes exactly. Even for geometrically asymmetric systems we typically find that it is a very small contribution compared with the symmetric term. Furthermore, the sign of the conductance step for the symmetric term in general shows an increase (decrease) in the conductance for low (high) conducting systems, e.g., vibrations usually help electrons through molecules while they backscatter electrons in atomic wires. This is discussed further for a one-level model in Ref. [62].

The LOE approximation is computationally simple and can be applied to systems of considerable size. Although the approximation is not strictly valid for systems with energy-dependent DOS, comparison with the full SCBA calculations shows good agreement even for systems that have a slowly varying DOS (on the scale of vibrational energies), e.g., the organic molecules connected to gold electrodes described in the next section. The LOE approximation will certainly fail when sharp (compared to the vibrational energies) resonances are present within the order of phonon energies of the Fermi energy. However, in this case Coulomb blockade is expected and the electron-electron interactions dominant. It is our experience that the LOE approximation is a valuable tool that may be used (with caution) for systems with a slowly varying DOS.

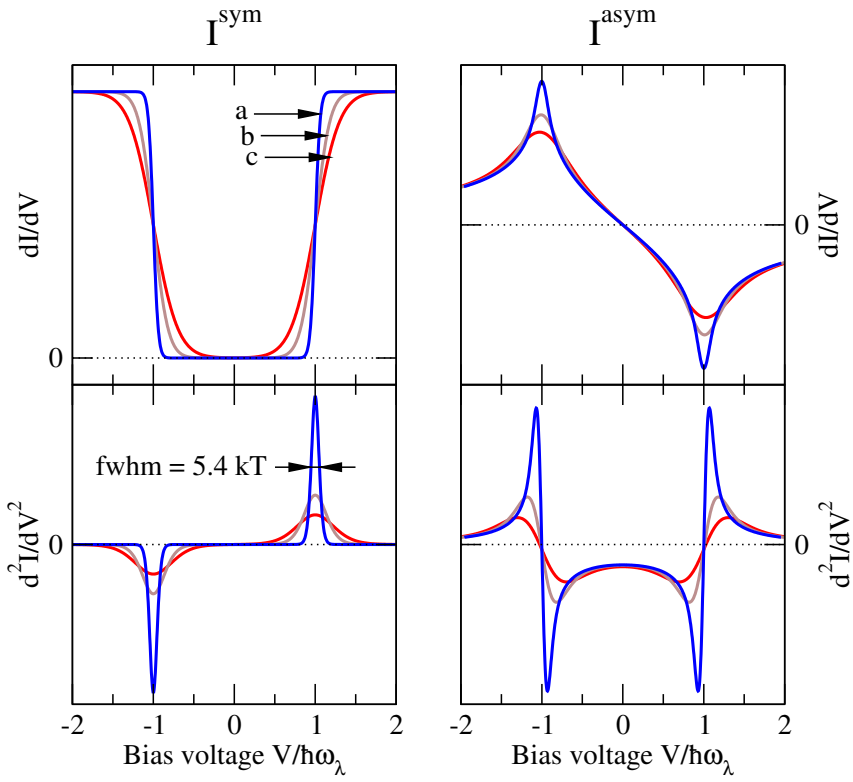


Figure 1.9: Universal functions Eq. (1.31) and (1.33) giving symmetric and asymmetric phonon contributions to the conductance, respectively. The differential conductance dI/dV and the second derivative d^2I/dV^2 are shown (in arbitrary units) for one phonon mode for three different temperatures (a) $k_BT/\hbar\omega_\lambda = 0.02$, (b) $k_BT/\hbar\omega_\lambda = 0.06$, and (c) $k_BT/\hbar\omega_\lambda = 0.10$.

1.3.2.1 Example: Hydrocarbon molecules between gold contacts

We will now illustrate the use of the LOE on molecules [60], and especially that the LOE approximation is valid for a wide range of systems even where at first glance it might fail. Measurements of the inelastic scattering signal through three different molecules connected to gold electrodes have been carried out by Kushmerick *et al.* [63] Since the number of molecules present in the experimentally realized junction is unknown it is advantageous to look at the inelastic electron tunneling spectroscopy (IETS) signal:

$$\text{IETS} \equiv \frac{d^2I/dV^2}{dI/dV} \tag{1.33}$$

which, if the current (I) simply scales with the number of molecules, is independent of the number of molecules in the junction.

Both the calculated and measured IETS for the conjugated OPE molecule (the molecule is drawn in Fig. 1.11b is shown in Fig. 1.10. For calculational details we refer to Ref. [60] As we can see the calculations reproduce the positions and relative heights of the inelastic scattering peaks. The three main peaks are given by four types of vibrations which all except the C–S stretch affect the carbon-backbone of the molecule. The dynamic atoms region used in the calculation contains 162 vibrational modes from

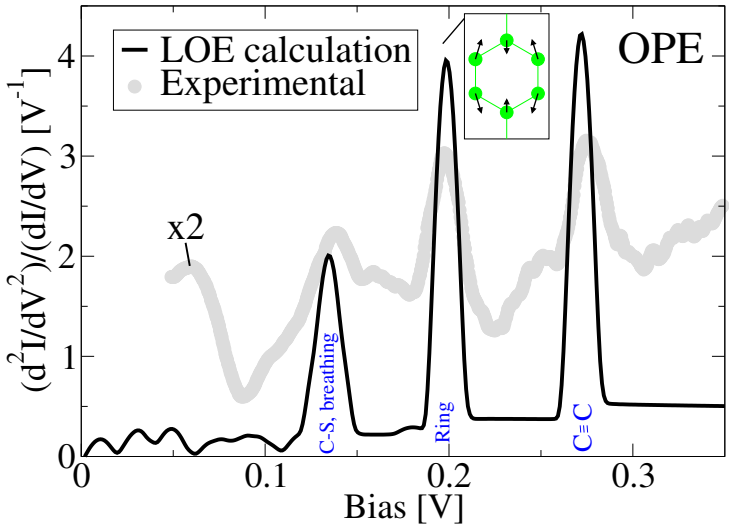


Figure 1.10: Calculated IETS compared to the experimental data from Ref. [63]. The inelastic scattering peaks arise from three different kinds of vibrations localized on the molecule. Reproduced from Ref. [60].

54 atoms, 18 Au plus the 36 atoms of the molecule. It is therefore relevant to discus selection rules, i.e., why only seven vibrations (three of the four types of vibrations are double degenerate) affect the current.

The calculations in Ref. [60] were carried out using the LOE approximation giving similar agreement with experiments for all three molecules studied as for the OPE molecule, see Fig. 1.10. However, electron conduction through these molecules is through the energy gap of the molecule where the transmission is not energy independent. The use of the LOE approximation may therefore seem inappropriate for these systems. We have therefore undertaken the tedious task to compare the LOE approximation to the full SCBA for the two conjugated molecules, see Fig. 1.11.

Since the SCBA is computationally expensive it was not possible to use the same accuracy in the calculations as in the LOE results. We therefore used only the molecule as the device subspace and dynamic atoms region, a smaller basis set (SZP) describing the OPE (OPV) molecule with 264 (280) atomic orbitals and only included the 5 (3) most important vibrational modes (selected from the LOE calculation). Calculating the current at 81 (61) bias points using an average of 9 (8) iterations to converge the SCBA on an energy grid of approximately 500 points took 40 (18) hours on 10 P4 computers in parallel. In comparison, the LOE approximation takes less than 1 minute on one processor.

The results shown in Fig. 1.11 reveal that the LOE approximation captures the inelastic scattering signal to a surprising accuracy. Also, the main discrepancy is directly related to the elastic part of the transport which can easily be corrected for without solving the full SCBA equations. We thus conclude that the LOE approximation can be used for molecular electronics with a significant degree of confidence.

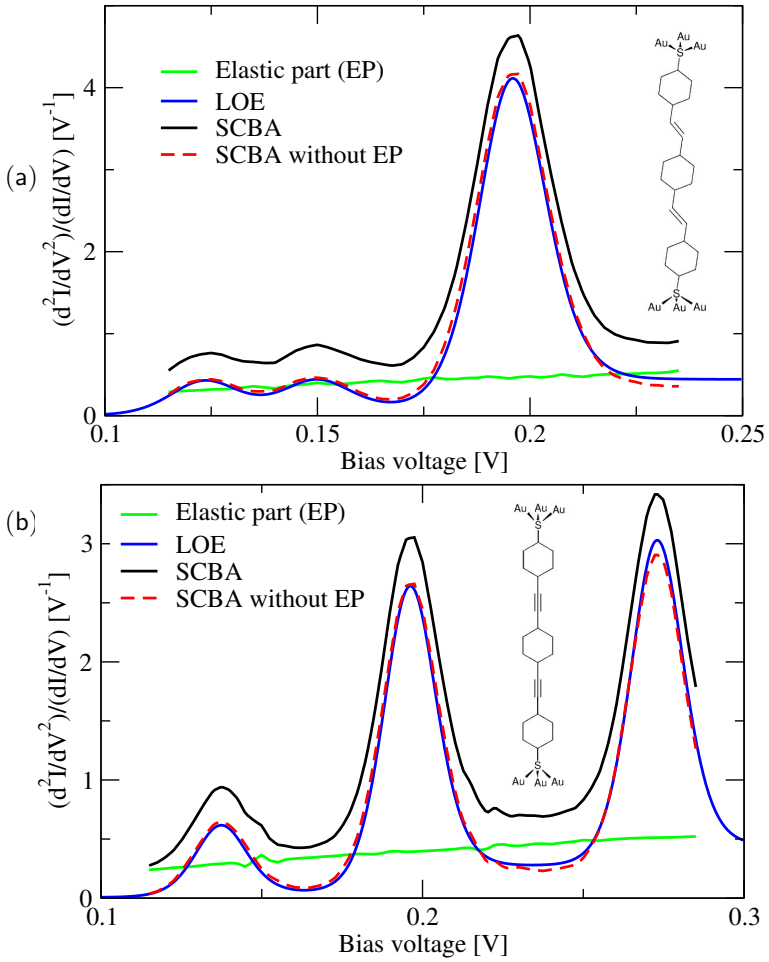


Figure 1.11: Hydrocarbon molecules OPV (upper) and OPE (lower).

1.4 Conclusions

We outlined and gave examples of atomistic transport theory which combine DFT and NEGF. From a fundamental point of view the use of the Kohn-Sham eigenfunctions of DFT as *bona fide* single-particle eigenstates for use in the transport calculation is not justified. On the other hand the reasonable agreement with experiments in many cases show that the somewhat pragmatic approach is rewarding and can contribute significantly to the understanding of these, see, e.g., the most recent Refs. [64, 65]. Moreover, the DFT seems to be a good starting-point for further theoretical developments as mentioned in the introduction, as well as new areas of applications, for example in molecular spintronics [66, 3].

The role of phonon interaction and local heating is of importance when predicting the characteristics and stability of nano-scale electronic devices, see e.g. Ref. [67] for a recent review on the theory. The complete treatment of electrons and phonons on an equal footing is yet to be developed in atomistic ab-initio calculations, i.e., taking the effects of the electrons on the phonon behavior into account in a self-consistent manner. However, as presented here, theory has been developed which takes the phonon-

interaction and its effect on the electronic current into account based on DFT combined with the NEGF formalism and the self-consistent Born approximation, or an efficient lowest order expansion applicable also for large systems. This approach works very well for weak electron-phonon coupling in molecular-scale conductors. In this type of theory the power delivered to the phonons can be taken into account and the phonon-occupation can be determined using simple rate arguments.

Bibliography

[1] M. Brandbyge, J. L. Mozos, P. Ordejon, J. Taylor, and K. Stokbro, *Density-functional method for nonequilibrium electron transport*, Phys. Rev. B **65**(16), 165401 (2002).

[2] A. Pecchia and A. Di Carlo, *Atomistic theory of transport in organic and inorganic nanostructures*, Rep. Prog. Phys. **67**(8), 1497 (2004).

[3] A. R. Rocha, V. M. Garcia-Suarez, S. W. Bailey, C. J. Lambert, J. Ferrer, and S. Sanvito, *Towards molecular spintronics*, Nature Materials **4**(4), 335–339 (2005).

[4] J. Taylor, H. Guo, and J. Wang, *Ab initio modeling of quantum transport properties of molecular electronic devices*, Phys. Rev. B **63**(24), 245407 (2001).

[5] S. Kurth, G. Stefanucci, C.-O. Almbladh, A. Rubio, and E. K. U. Gross, *Time-dependent quantum transport: A practical scheme using density functional theory*, Phys. Rev. B **72**(3), 035308 (2005).

[6] A. Pecchia, A. Di Carlo, A. Gagliardi, A. Th. Niehaus, and T. Frauenheim, *Atomistic simulation of the electronic transport in organic nanostructures: Electron-phonon and electron-electron interactions*, J. Comput. Electron. **4**, 79 (2005).

[7] C. Toher, A. Filippetti, S. Sanvito, and K. Burke, *Self-interaction errors in density-functional calculations of electronic transport*, Phys. Rev. Lett. **95**, 146402 (2005).

[8] J. J. Palacios, *Coulomb blockade in electron transport through a c-60 molecule from first principles*, Phys. Rev. B **72**(12), 125424 (2005).

[9] A. Ferretti, A. Calzolari, R. Di Felice, F. Manghi, M. J. Caldas, M. B. Nardelli, and E. Molinari, *First-principles theory of correlated transport through nanojunctions*, Phys. Rev. Lett. **94**(17), 179901 (2005).

[10] N. D. Lang, *Resistance of a one-atom contact in the scanning tunneling microscope*, Phys. Rev. B **36**(15), 8173–8176 (1987).

[11] J. Ferrer, A. Martín-Rodero, and F. Flores, *Contact resistance in the scanning tunneling microscope at very small distances*, Phys. Rev. B **38**(14), 10113–10115 (1988).

[12] W. A. Hofer and A. J Fisher, *Signature of a chemical bond in the conductance between two metal surfaces*, Phys. Rev. Lett. **91**(3), 036803 (2003).

[13] S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge, UK, 1995).

[14] H. Haug and A.-P. Jauho, *Quantum Kinetics in Transport and Optics of Semiconductors* (Springer-Verlag, 1996).

[15] D. R. Bowler, *Atomic-scale nanowires: physical and electronic structure*, J. Phys.: Condens. Matter **16**(24), R721–R754 (2004).

[16] G. Cuniberti, G. Fagas, and K. Richter, *Introducing Molecular Electronics* (Springer, 2005).

[17] A. Pecchia and A. Di Carlo, *Atomistic theory of transport in organic and inorganic nanostructures*, Rep. Prog. Phys. **67**(8), 1497–1561 (2004).

[18] J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejon, and D. Sanchez-Portal, *The siesta method for ab initio order-n materials simulation*, J. Phys.: Condens. Matter **14**(11), 2745–2779 (2002).

[19] M. Brandbyge, N. Kobayashi, and M. Tsukada, *Conduction channels at finite bias in single-atom gold contacts*, Phys. Rev. B **60**(24), 17064–17070 (1999).

[20] K. S. Thygesen, *Electron transport through an interacting region: The case of a nonorthogonal basis set*, Phys. Rev. B **73**(3), 035309 (2006).

[21] M. P. Lopez-Sancho, J. M. Lopez-Sancho, and J. Rubio, *Quick iterative scheme for the calculation of transfer matrices: application to Mo (100)*, J. Phys. F **14**, 1205 (1984).

[22] M. Brandbyge, K. Stokbro, J. Taylor, J. L. Mozos, and P. Ordejon, *Origin of current-induced forces in an atomic gold wire: A first-principles study*, Phys. Rev. B **67**(19), 193104 (2003).

[23] N. D. Lang and P. Avouris, *Understanding the variation of the electrostatic potential along a biased molecular wire*, Nano Lett. **3**(6), 737–740 (2003).

[24] J. Taylor, M. Brandbyge, and K. Stokbro, *Conductance switching in a molecular device: The role of side groups and intermolecular interactions*, Phys. Rev. B **68**(12), 121101 (2003).

[25] M. Büttiker, Y. Imry, R. Landauer, and S. Pinhas, *Generalized many-channel conductance formula with application to small rings*, Phys. Rev. B **31**(10), 6207–6215 (1985).

[26] D. S. Fisher and P. A. Lee, *Relation between conductivity and transmission matrix*, Phys. Rev. B **23**(12), 6851–6854 (1981).

[27] J. C. Cuevas, A. Levy Yeyati, and A. Martín-Rodero, *Microscopic origin of conducting channels in metallic atomic-size contacts*, Phys. Rev. Lett. **80**(5), 1066–1069 (1998).

[28] Th. Martin and R. Landauer, *Wave-packet approach to noise in multichannel mesoscopic systems*, Phys. Rev. B **45**(4), 1742–1755 (1992).

[29] M. Brandbyge, M. R. Sorensen, and K. W. Jacobsen, *Conductance eigenchannels in nanocontacts*, Phys. Rev. B **56**(23), 14956–14959 (1997).

[30] S. K. Nielsen, M. Brandbyge, K. Hansen, K. Stokbro, J. M. van Ruitenbeek, and F. Besenbacher, *Current-voltage curves of atomic-sized transition metal contacts: An explanation of why au is ohmic and pt is not*, Phys. Rev. Lett. **89**(6), 066804 (2002).

[31] L. de la Vega, A. Martin-Rodero, N. Agrait, and A. L. Yeyati, *Universal features of electron-phonon interactions in atomic wires*, Phys. Rev. B **73**(7), 075428 (2006).

[32] K. Flensberg, *Tunneling broadening of vibrational sidebands in molecular transistors*, Phys. Rev. B **68**(20), 205323 (2003).

[33] M. Galperin, A. Nitzan, and M. A.Ratner, *Resonant inelastic tunneling in molecular junctions*, Phys. Rev. B **73**(4), 045314 (2006).

[34] M. J. Montgomery, J. Hoekstra, T. N. Todorov, and A. P. Sutton, *Inelastic current-voltage spectroscopy of atomic wires*, J. Phys.: Condens. Matter **15**(4), 731–742 (2003).

[35] H. Ness and A. J. Fisher, *Vibrational inelastic scattering effects in molecular electronics*, Proc. Natl. Acad. Sci. U.S.A. **102**(25), 8826–8831 (2005).

[36] H. Ness, S. A. Shevlin, and A. J. Fisher, *Coherent electron-phonon coupling and polaronlike transport in molecular wires*, Phys. Rev. B **63**(12), 125422 (2001).

[37] Y. Asai, *Theory of inelastic electric current through single molecules*, Phys. Rev. Lett. **93**, 246102 (2004); *Erratum*, Phys. Rev. Lett. **94**, 099901(E) (2005).

[38] T. Yamamoto, K. Watanabe, and S. Watanabe, *Electronic transport in fullerene c-20 bridge assisted by molecular vibrations*, Phys. Rev. Lett. **95**(6), 065501 (2005).

[39] N. Lorente and M. Persson, *Theoretical aspects of tunneling-current-induced bond excitation and breaking at surfaces*, Discuss. Faraday Soc. **117**, 277–290 (2000).

[40] N. Lorente, M. Persson, L. J. Lauhon, and W. Ho, *Symmetry selection rules for vibrationally inelastic tunneling*, Phys. Rev. Lett. **86**, 2593 (2001).

[41] N. Lorente, R. Rurali, and H. Tang, *Single-molecule manipulation and chemistry with the stm*, J. Phys.: Condens. Matter **17**(13), S1049–S1074 (2005).

[42] H. Ueba, *Motions and reactions of single adsorbed molecules induced by vibrational excitation with stm*, Surf. Rev. Lett. **10**(5), 771–796 (2003).

[43] H. Ueba, T. Mii, N. Lorente, and B. N. J. Persson, *Adsorbate motions induced by inelastic-tunneling current: Theoretical scenarios of two-electron processes*, J. Chem. Phys. **123**(8), 084707 (2005).

[44] Y. C. Chen, M. Zwolak, and M. Di Ventra, *Local heating in nanoscale conductors*, Nano Lett. **3**(12), 1691–1694 (2003).

[45] Y. C. Chen, M. Zwolak, and M. Di Ventra, *Inelastic current-voltage characteristics of atomic and molecular junctions*, Nano Lett. **4**(9), 1709–1712 (2004).

[46] Y. C. Chen, M. Zwolak, and M. Di Ventra, *Inelastic effects on the transport properties of alkanethiols*, Nano Lett. **5**(4), 621–624 (2005).

[47] M. Di Ventra and N. D. Lang, *Transport in nanoscale conductors from first principles*, Phys. Rev. B **65**(4), 045402 (2002).

[48] A. Pecchia, A. Di Carlo, A. Gagliardi, S. Sanna, T. Frauenheim, and R. Gutierrez, *Incoherent electron-phonon scattering in octanethiols*, Nano Lett. **4**(11), 2109–2114 (2004).

[49] G. C. Solomon, A. Gagliardi, A. Pecchia, T. Frauenheim, A. Di Carlo, J. R. Reimers, and H. S. Hush, *Understanding the inelastic electron-tunneling spectra of alkanedithiols on gold*, J. Chem. Phys. **124**, 094704 (2006).

[50] N. Sergueev, D. Roubtsov, and H. Guo, *Ab initio analysis of electron-phonon coupling in molecular devices*, Phys. Rev. Lett. **95**(14), 146803 (2005).

[51] A. Troisi and M. A. Ratner, *Modeling the inelastic electron tunneling spectra of molecular wire junctions*, Phys. Rev. B **72**, 033408 (2005).

[52] J. Jiang, M. Kula, W. Lu, and Y. Luo, *First-principles simulations of inelastic electron tunneling spectroscopy of molecular electronic devices*, Nano Lett. **5**(8), 1551–1555 (2005).

[53] M. Headgordon and J. C. Tully, *Vibrational-relaxation on metal-surfaces - molecular-orbital theory and application to co/cu(100)*, J. Chem. Phys. **96**(5), 3938–3949 (1992).

[54] T. Frederiksen, M. Paulsson, M. Brandbyge, and A. P. Jauho, *Inelastic transport theory from first-principles: Methodology and applications*, to be submitted to Phys. Rev. B September 2006.

[55] T. Frederiksen, M. Brandbyge, N. Lorente, and A.-P. Jauho, *Inelastic scattering and local heating in atomic gold wires*, Phys. Rev. Lett. **93**, 256601 (2004).

[56] G. Rubio-Bollinger, S. R. Bahn, N. Agrait, K. W. Jacobsen, and S. Vieira, *Mechanical properties and formation mechanisms of a wire of single gold atoms*, Phys. Rev. Lett. **87**(2), 026101 (2001).

[57] N. Agrait, C. Untiedt, G. Rubio-Bollinger, and S. Vieira, *Onset of energy dissipation in ballistic atomic wires*, Phys. Rev. Lett. **88**(21), 216803 (2002).

[58] M. Paulsson, T. Frederiksen, and M. Brandbyge, *Modeling inelastic phonon scattering in atomic- and molecular-wire junctions*, Phys. Rev. B **72**(20), 201101 (2005).

[59] M. Persson and B Hellsing, *Electronic damping of adsorbate vibrations on metal-surfaces*, Phys. Rev. Lett **49**(9), 662–665 (1982).

[60] M. Paulsson, T. Frederiksen, and M. Brandbyge, *Inelastic transport through molecules: Comparing first-principles calculations to experiments*, Nano Lett. **6**(2), 258–262 (2006).

[61] J. K. Viljas, J. C. Cuevas, F. Pauly, and M. Hafner, *Electron-vibration interaction in transport through atomic gold wires*, Phys. Rev. B **72**(24), 245415 (2005).

[62] M. Paulsson, T. Frederiksen, and M. Brandbyge, *Phonon scattering in nanoscale systems: lowest order expansion of the current and power expressions*, Journal of Physics, Conference Series **35**, 247–254 (2005).

[63] J. G. Kushmerick, J. Lazorcik, C. H. Patterson, R. Shashidhar, D. S. Seferos, and G. C. Bazan, *Vibronic contributions to charge transport across molecular junctions*, Nano Lett. **4**(4), 639–642 (2004).

[64] J. He, O. Sankey, M. Lee, N. J. Tao, X. L. Li, and S. Lindsay, *Measuring single molecule conductance with break junctions*, Faraday Discuss. **131**, 145–154 (2006).

[65] L. Venkataraman, J. E. Klare, C. Nuckolls, M. S. Hybertsen, and M. L. Steigerwald, *Dependence of single-molecule junction conductance on molecular conformation*, Nature **442**(7105), 904–907 (2006).

[66] V. V. Maslyuk, A. Bagrets, V. Meded, A. Arnold, F. Evers, M. Brandbyge, T. Bredow, and I. Mertig. *Organometallic benzene-vanadium wire: A one-dimensional half-metallic ferromagnet*, Phys. Rev. Lett. **97**(9), 097201 (2006).

[67] A. P. Horsfield, D. R. Bowler, H. Ness, C. G. Sánchez, T. N. Todorov, and A. J. Fisher, *The transfer of energy between electrons and ions in solids*, Rep. Prog. Phys. **69**, 1195–1234 (2006).

Chapter 2

Order N Methods for Quantum Transport

S. Roche
Commissariat à l'Energie Atomique,
DRFMC/SPSMS/GT,
17 avenue des AMrtyrs, 38054 Grenoble, France
F. Triozon
CEA, LETI-Minatec,
17 avenue des Martyrs, 38054 Grenoble, France

2.1 Introduction

Today's research endeavours in molecular electronics or nanoelectronics rely on both : advanced top-down and bottom-up techniques to fabricate and engineering functional molecular based devices and circuits, but also advanced computational schemes that enable in-depth exploration of their novel properties and ultimate performances. The numerical simulation strategy of material and devices at the nanoscale is twofold. On one hand, first principles (*ab initio*) methods, such as Density Functional Theory (DFT) and its extensions push the limits of realistic simulation of the physico-chemical complexity at the atomistic scale, searching for elaborated description of electron-electron correlations, to the price of limited implementation for large scale simulations and out-of equilibrium problems. On the other hand, semi-empirical approaches provide a more suitable basis for simulating large scale systems, albeit limited to effective hamiltonians with adjustable parameters. Both directions rely however on approximations of the true many-body problem. The combination of both approaches opens new perspectives to circumvent intrinsic limitations of separated methods, enabling to explore advanced quantum transport phenomena in materials and devices with large chemical complexity. To compute current-voltage characteristics of realistic

nanodevices, efficient linear scaling computational methodologies of quantum transport at equilibrium and far for equilibrium are of crucial importance.

In this contribution, we will review our developed order N scaling transport methods, that allow to efficiently compute either the Kubo-Greenwood or the Landauer-Büttiker conductance in materials or heterojunctions of large complexity (provided we restrict to physical situations close to thermodynamic equilibrium). These methods are based on the recursive construction of either orthogonal (Lanczos-scheme) or bi-orthogonal basis in which hamiltonian matrices are first tridiagonalized, and continued-fraction expansion further used to accurately compute off-diagonal Green's function matrix elements. They turn out to be of broad range of applicability and well suited for real space calculations, so using expansion of wavefunctions in localized basis sets. This includes semi-empirical or *ab initio* methods based on localized functions and implemented within DFT method such as SIESTA [1]. Following the same philosophy out-of equilibrium methods could be further investigated.

2.2 Kubo-Greenwood and Landauer-Büttiker real space order N methods

2.2.1 Conduction mechanisms and conductance scaling

Efficient computational recursion and order N methods have been successfully developed in solid-state physics since their introduction by R. Haydock [2, 3, 4, 5]. The recursion methods are based on an eigenvalue approach of Lanczos [6], and rely on the computation of Green's functions matrix elements by continuous fraction expansion, which can be implemented either in real or reciprocal spaces. These techniques are particularly well suited for treating disorder and defect-related problems, and were successfully implemented to tackle with impurity-level calculations in semiconductors using tight-binding approximation [7], or with electronic structure investigations for amorphous semiconductors, transition metals and metallic glasses based on the linear-muffin-tin orbitals [8].

On the other hand, the general electronic transport theory in the linear response regime relies on the approach derived by R. Kubo [9]. In its zero frequency limit, it reduces to the trace of the operator $\delta(E-H)\hat{V}_x\delta(E-H)\hat{V}_x$, that relates the spectral measure operator $\delta(E-H)$ to the velocity operator $\hat{V}_x$. Electronic dc-conductivity is thus seen as a measure of autocorrelation average of wavepackets velocities. During the past decade, we have endeavoured the implementation of the recursion method in the calculation of the Kubo-Greenwood transport coefficients [10, 11, 12]. This approach, optimized over the years, is mainly based on the resolution of the time dependent Schrödinger equation (TDSE) employing an expansion of the spectral measure onto a basis of orthogonal (typically Chebyshev) polynomials $Q_n(E)$. By doing so, any operator can be expanded within the same basis as

$$|\Psi(t)\rangle = \exp(-i\frac{Ht}{\hbar})|\Psi(0)\rangle$$

$$= \sum_{n=0}^{N} \left(\int n(E) e^{-i\frac{Et}{\hbar}} Q_n(E) dE \right) Q_n(H)|\psi(0)\rangle$$

which allows an efficient and quick resolution of TDSE. Afterwards the diffusion coefficient as well as the related Kubo conductance can be straightforwardly computed as

$$\begin{aligned} D(E,t) &= \frac{1}{t}\langle (\hat{X}(t) - \hat{X}(0))^2 \rangle_E \\ G(E) &= \frac{2e^2}{L_{syst}} \lim_{t \to \tau_L} \mathrm{Tr}[\delta(E-H)D(E,t)] \end{aligned}$$

given their direct relation to the time-dependence of electronic wavepackets. The calculation of both quantities can be reduced to the computation of on-diagonal Green's function that are achieved by a continuous fraction expansion. From this methodology we have been able to analyze all the main conduction mechanisms in the quantum coherent regime in various low-dimensional systems or disordered materials. They include the

- Ballistic regime: $G(E) = (2e^2/h)N(E)$
- Diffusive regime: $G(E) = (2e^2/h)\ell_e/L_{syst}$
- Weak localization: $G(E) = \frac{2e^2}{h}\frac{\ell_e}{L_{syst}} - \delta G_{WL}$
- Strong localization $G(E) = \frac{2e^2}{h}\exp(-\ell_e/L_{syst})$

defining $N(E)$ the conducting channels number, ℓ_e the elastic mean free path associated to a given disorder model, and δG_{WL} the weak localization correction to the quantum conductance. In Fig.2.1, one shows an illustration of all those conduction mechanisms as well as the obtained conductance for the ballistic case for a metallic (10,10) nanotube with or without Anderson type disorder (taken as random fluctuations of onsite energies within an interval $[-W/2, W/2]$).

This approach was successfully applied to the study of quasiperiodic systems [10], quantum-Hall effect [11], carbon nanotubes [12] and semiconducting nanowires [13]. Hereafter, we will show a quantitative validation of the method by comparison with analytical results for a model Anderson-type disorder, and further illustrate the quantitative predictability of the methods for studying quantum transport in realistic models, such as chemically modified carbon nanotubes, effects of magnetic fields, or incommensurability in multiwalled nanotubes. Finally, the essential steps for generalizing the approach to the Landauer-Büttiker framework, essential for simulating nanoscale devices, will be drawn.

2.2.2 Elastic mean free path scaling properties in Carbon Nanotubes

For a in-depth understanding of disorder effects in low dimensional systems, the evaluation of the elastic mean free path (ℓ_e) is a first fundamental step. For sufficiently weak disorder, a perturbative treatment can be performed within the Fermi Golden Rule (FGR), giving a direct access to the elastic mean free path

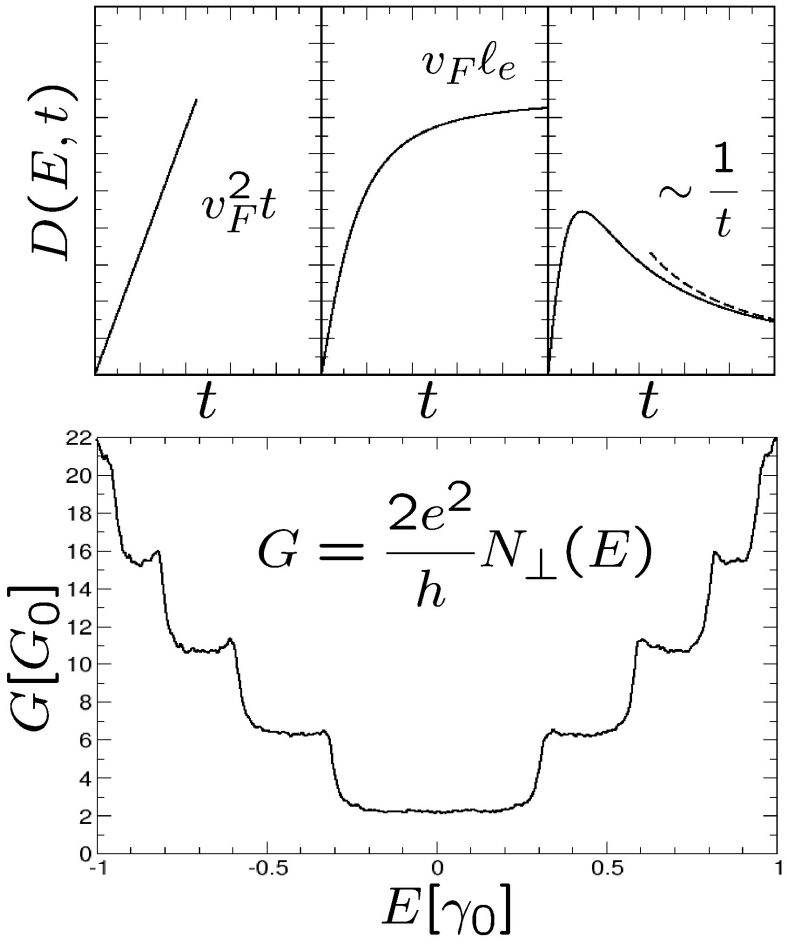


Figure 2.1: Top: time-dependent diffusion coefficient for various conduction regimes in metallic carbon nanotubes (10,10) with or without elastic disorder (Anderson-type). Bottom: quantized energy dependent Kubo conductance obtained for disorder-free nanotube.

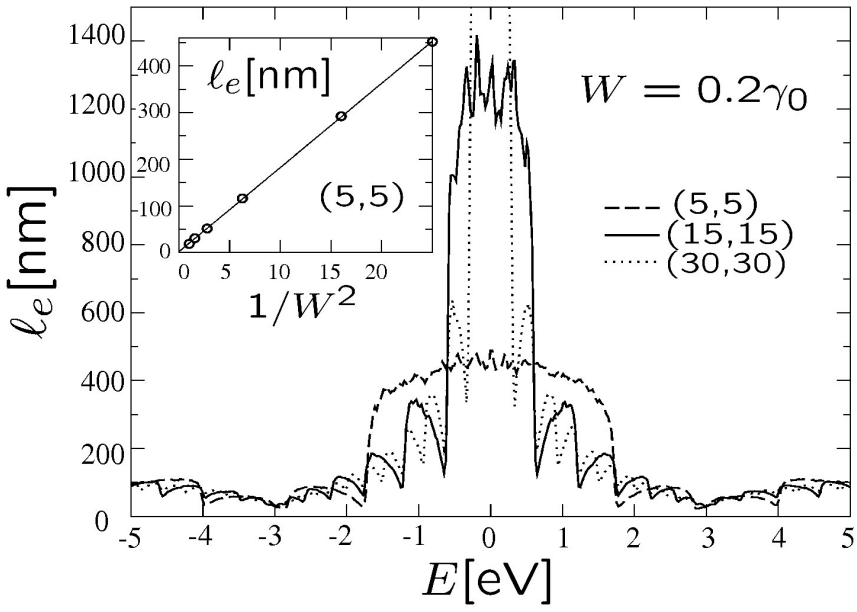


Figure 2.2: Main frame: Energy dependent mean free path as a function of diameter. Inset: $1/W^2$-scaling in agreement with Fermi golden rule. Adapted from [17].

$\ell_e = v_F\tau$. In the context of disordered metallic carbon nanotubes [14, 15], an analytical result was first derived by White and Todorov [16]. By reducing the bandstructure to a two-band approximation, and describing the disorder by the onsite Anderson-type potential (see below), ℓ_e was analytically derived, and found to linearly scale with diameter for a fixed disorder strength W, while at a fixed diameter (note that a (n,m) nanotube has a diameter $d_t = \sqrt{3}a_{cc}\sqrt{n^2+m^2+nm}/\pi$), the expected disorder scaling $\ell_e \sim 1/W^2$ was shown. The analytical derivation of such fundamental length scale requires the calculation of total DoS in the vicinity of Fermi level. The DoS can be generally written as $\rho(E) = \mathrm{Tr}[\delta(E-\mathcal{H})]$ where the trace has to be developed over a complete basis set, while the application of the FGR yields

$$\frac{1}{2\tau_e(E_F)} = \frac{2\pi}{\hbar}\left|\langle\Psi_{n1}(k_F)|\ \hat{\mathcal{U}}\ |\Psi_{n2}(-k_F)\rangle\right|^2 \rho(E_F)\times N_{\mathrm{c}}N_{\mathrm{Ring}} \tag{2.1}$$

with N_{c} and N_{Ring}, the respective number of pair atoms along the circumference and the total number of rings taken in the unit cell used for diagonalization, whereas the eigenstates at the Fermi level are rewritten as

$$\begin{aligned}
|\Psi_{n1,n2}(k_F)\rangle &= \frac{1}{\sqrt{N_{\mathrm{Ring}}}}\sum_{m=1,N_{\mathrm{Ring}}} \mathrm{e}^{\mathrm{i}mk_F}|\alpha_{n1,n2}(m)\rangle \quad \text{with} \\
|\alpha_{n1}(m)\rangle &= \frac{1}{\sqrt{2N_{\mathrm{c}}}}\sum_{n=1}^{N_{\mathrm{c}}} \mathrm{e}^{\frac{2\mathrm{i}\pi n}{N_{\mathrm{c}}}}\Big(|p_z^A(mn)\rangle + |p_z^B(mn)\rangle\Big) \\
|\alpha_{n2}(m)\rangle &= \frac{1}{\sqrt{2N_{\mathrm{c}}}}\sum_{n=1}^{N_{\mathrm{c}}} \mathrm{e}^{\frac{2\mathrm{i}\pi n}{N_{\mathrm{c}}}}\Big(|p_z^A(mn)\rangle - |p_z^B(mn)\rangle\Big)
\end{aligned} \tag{2.2}$$

while the disorder considered here is an uncorrelated white noise (Anderson-type) distribution given by

$$\begin{aligned}
\langle p_z^A(mn) \mid \hat{\mathcal{U}} \mid p_z^A(m'n')\rangle &= \varepsilon_A(m,n)\delta_{mm'}\delta_{nn'} \\
\langle p_z^B(mn) \mid \hat{\mathcal{U}} \mid p_z^b(m'n')\rangle &= \varepsilon_B(m,n)\delta_{mm'}\delta_{nn'} \\
\langle p_z^A(mn) \mid \hat{\mathcal{U}} \mid p_z^A(m'n')\rangle &= 0
\end{aligned} \tag{2.3}$$

where $\varepsilon_B(m,n)$ and $\varepsilon_A(m,n)$ are the onsite energies of electron at atoms A and B in position (m,n), randomly distributed within the interval $[-W/2, W/2]$ and following some uniform distribution with probability $\mathcal{P} = 1/W$. Then by replacing Eq.2.2 in Eq.2.1, using Eq.2.3, a straightforward calculation gives :

$$\frac{1}{\tau_e(E_F)} = \frac{\pi\rho(E_F)}{\hbar}\left(\frac{1}{\sqrt{N_c N_{\rm Ring}}}\sum_{N_c N_{\rm Ring}}\varepsilon_A^2 + \frac{1}{\sqrt{N_c N_{\rm Ring}}}\sum_{N_c N_{\rm Ring}}\varepsilon_B^2\right)$$

Hence, if the disorder is described by random fluctuations of onsite energies with uniform probability $1/W$ (W the disorder bandwidth) the mean free path can be finally analytically [16, 18] derived as

$$\ell_e = \frac{18 a_{cc}\gamma_0^2}{W^2}\sqrt{n^2+m^2+nm} \sim \left(\frac{\gamma_0}{W}\right)^2 d_t$$

For the armchair $m = n = 5$ nanotube, with disorder $W = 0.2\gamma_0$, applying the above equation, one finds $\ell_e \sim 560$ nm which is much more larger than the circumference length. As shown in Fig. 2.2, numerical studies [17] confirm the scaling law of the mean free path with the nanotube diameter close to the charge neutrality point. For semiconducting bands, the $1/W^2$ is still satisfied, but mean free paths are seen to be much smaller and do not scale with diameter, in full agreement with experimental estimates [19].

2.2.3 Weak localization and energy dependent coherence lengths

The understanding of localization effects in disordered mesoscopic systems stands as a central issue based on the quantum interference effects (QIE) on charge transport [20, 21]. These QIE between clockwise and counterclockwise backscattering paths develop in the so-called coherent regime, and yield an increase of the return probability to the origin for propagating wavepackets. The contribution of QIE is usually reduced by several inelastic scattering sources that produce decoherence of the wavepacket phase. At low temperature the main decoherence mechanisms are e-ph and electron-electron couplings. Within the framework of weak localization theory, it has been possible to derive perturbatively the relation between the measured conductance $G(E)$, its quantum correction $\delta G_{WL}(E)$ and the coherence length L_ϕ that fixes the scale beyond which QIE are destroyed. The estimation of the coherence lengths is a central issue in mesoscopic physics, and weak localization provides an elegant framework to extract the behavior of L_ϕ, that mainly depends on the dimensionality of charge transport [9]. Stojetz and coworkers [19] have recently succeeded in measuring the energy-dependence of the coherence length scale, by using an efficient back-gate electrode able to move the Fermi level position and explore the physics through

different subbands. The magnetoresistance data were well fitted by the conventional theory as [21]

$$\delta G_{WL} = -A\frac{e^2}{\pi\hbar L}\left(\frac{1}{L_\phi^2}+\frac{W^2e^2B^2}{3\hbar^2}\right)^{-1/2}, \tag{2.4}$$

with L the tube length, W the diameter and A a normalization factor. From this expression, the coherence length can be quantitatively extracted. Therefore, by using $L_\phi(T) = (GDL\hbar^2/2e^2k_B)^{1/3}T^{-1/3}$, the diffusivity was deduced $D = 100\text{cm}^2\text{s}^{-1}$ (at zero gate voltage) as well as the corresponding elastic mean free path $\ell_e = 2D/v_F \sim 20\text{nm}$. By reproducing this calculation at several values of the gate voltage, a corresponding energy-dependent pattern was revealed. The spectacular observation was the systematic decrease of L_ϕ (as well as ℓ_e) near the onsets of each new subbands (van-Hove singularities positions). If the energy dependence of ℓ_e has been revealed numerically [12], the energy dependence of L_ϕ was to date unexplored.

Finally, by studying the temperature dependence of L_ϕ, the decoherence mechanism was attributed to electron-electron scattering, in agreement with conventional theory. Electron-electron and e-ph scattering are two different sources of quantum dephasing, but within weak localization theory the resulting decoherence phenomenon is of similar nature, and for instance the derivation of the temperature dependent coherence time follows some general formal treatment.

The energy dependence of the coherence length scaling properties in carbon nanotubes can be analyzed in the situation where e-ph interaction becomes the dominant source of decoherence. The physical origin of the coherence length can be simply understood as follows. First, weak localization corrections are evaluated when computing the conductance for wavepacket propagation between two real space positions, let us say from P to Q : $G = \frac{2e^2}{h}\mathcal{P}_{P\to Q}$ which can be further expanded as $\mathcal{P}_{P\to Q} = \sum_i |\mathcal{A}_i|^2 + \sum_{i\neq j}\mathcal{A}_i\mathcal{A}_j e^{i(\alpha_i-\alpha_j)}$, where the summation over probability amplitudes (denote $\mathcal{A}_i$) includes all possible electronic pathways i (see Fig.2.3 for illustration). Here one separates the total probability between the classical and interference parts. Then averaging over (elastic) disorder configurations reduce the contribution of interference terms to a single class, namely the paths that include a finite loop which return to some initial point, let's say O, that can be further expressed analytically as $\mathcal{P}_{O\to O} = \left|\mathcal{A}_+e^{i\alpha_+} + \mathcal{A}_-e^{+i\alpha_-}\right|^2 = 4\left|\mathcal{A}_0\right|^2$, since clockwise and counter clockwise probability amplitudes have the same phase factor in case of time reversal invariance [22]. Weak localization thus results from the enhancement (doubling) of the return probability to the origin, with a consequent increase of quantum resistance. To understand decoherence, one must realize that when inelastic mechanisms come into play on top of coherent transport, then clockwise and counterclockwise pathways accumulate a different superimposed random phase factor $e^{i\alpha_+} = \langle e^{+i\phi}\rangle = \int d\phi P(\phi)e^{i\phi}$ and $e^{i\alpha_-} = \langle e^{-i\varphi}\rangle = \int d\varphi P(\varphi)e^{-i\varphi}$, respectively. These expressions are derived by considering that the coherent propagating wavepackets is coupled to some external fluctuating potential, such as a thermal bath or fluctuating electromagnetic field that respectively encode the e-ph and e-e interaction processes [22]. $P(\phi)$ denotes the probability distribution function for the phase φ, that results from the combination of inelastic and elastic scattering processes, both having a probabilistic nature along the electronic paths (Fig.2.3 (c) and (d)). Besides, if $\mathbf{u}(\mathbf{r},t)$ defines the phonon displacement field, these phase factors are physically related to the Lagrangian

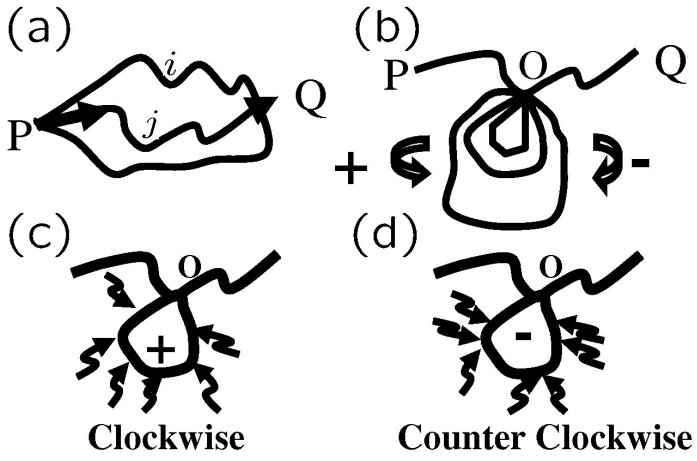


Figure 2.3: (a) Representation of electronic propagating paths between two points P and Q (a); several paths enclosing a loop returning back to some origin (b), with clockwise (-) and counterclockwise (+) parts; mechanism of decoherence on those paths (arrow indicate inelastic scattering events occurring at different points for (+) and (-) paths.

of the system, through its action along the electronic pathways, giving [22]

$$\phi, \varphi = \frac{m_C}{\hbar} \int \mathbf{v} \cdot (\mathbf{v} \cdot \nabla)\mathbf{u}(\mathbf{r}, t) - \frac{1}{d}\mathbf{v^2} \nabla \cdot \mathbf{u}(\mathbf{r}, \mathbf{t}) dt, \tag{2.5}$$

defining m_c the atom carbon mass; and d the space dimensionality, and $\mathbf{r}, \mathbf{v}$ describing the position and related velocity of the electronic pathways [22], while the summation is performed for an elapsed time required to close the relevant loop. The coherence time τ_ϕ is the time beyond which a full uncertainty of the phase difference is achieved between clockwise and counter clockwise paths (that is when the width of the distribution $P(\varphi)$ becomes in order of $\sim 2\pi$) . Accordingly, the coherence length $L_\phi(E)$ is the length associated with the loop that is closed after an elapsed time $t = \tau_\phi$. Finally, within this phenomenology, the quantum correction of conductance is given by the relevant integrated return probability to the origin: $\delta G_{WL} \sim \frac{2e^2 D}{\hbar} \int_0^\infty \mathcal{P}_{O \to O}(t)(e^{-t/\tau_\phi} - e^{-t/\tau_e})$, thus the quantitative measure of conductance correction is driven by independent calculation of the return probability to the origin in the coherent regime on one side, and estimation of the relevant coherence time on the other side. The scaling properties of the quantum correction in the coherent regime can be investigated numerically by solving the time-dependent Schrödinger equation (TDSE), whereas the coherence time will be given by the considered total elapsed time, at which the TDSE is solved [23]. By doing so, the energy-dependence of the weak localization pattern is correctly reproduced without however providing a quantitative information about the exact value of the coherence length scales, neither their temperature dependences. This point should be considered on the basis of a microscopical modeling of e-ph interaction and rigorous quantum mechanical treatment of inelastic transport. The resolution of the TDSE can be made by expanding the evolution operator $e^{-i\hat{\mathcal{H}}t}$ as a product of short-time evolution steps $e^{-i\hat{\mathcal{H}}\Delta T}$, for a total evolution time $t = n\Delta T$. Typically ΔT is one tenth of the oscillation period of the considered phonon acoustic mode.

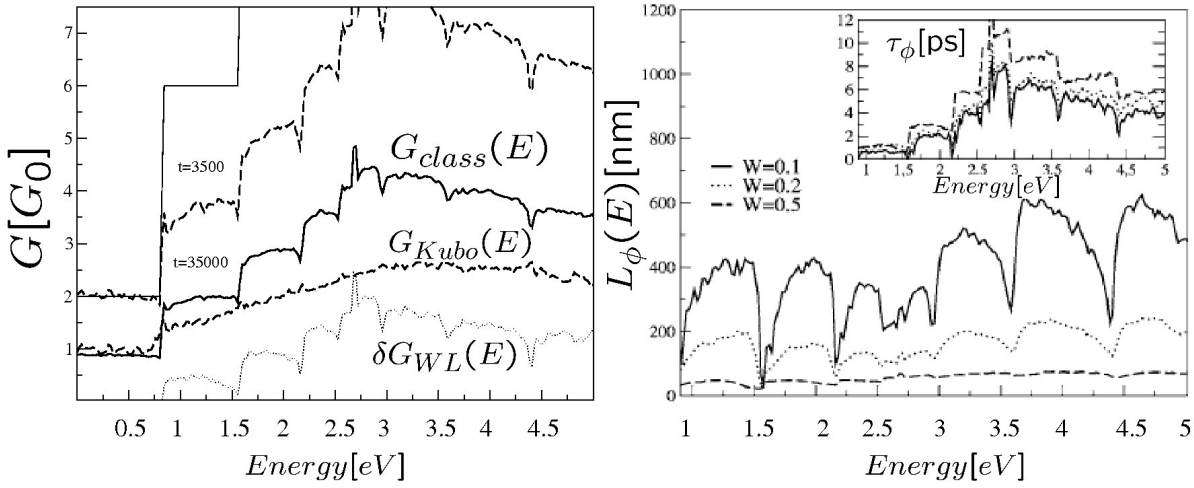


Figure 2.4: Left panel: Conductance for the disordered (10,10) tube with $W = 0.07\gamma_0$, taken at evolution times $t = 3500\hbar/\gamma_0$ and $t = 35000\hbar/\gamma_0$ (dashed curves). The bold curve gives the classical part $G_{\rm class}(E) = G_0 N_\perp \ell_e(E)/L(E,t)$, whereas the dotted curve gives the quantum correction $\delta G_{WL}(E)$ ($t = 35000\hbar/\gamma_0$). Right panel: Coherence lengths deduced from $\delta G_{WL}(E) = L_\phi(E)/L(E)$ computed at $t = 35000\hbar/\gamma_0$, and for several values of disorder potential. Inset: Corresponding $\tau_\phi(E)$ for the same parameters. In addition to the random short range disorder, a twist acoustic-mode modulation is introduced.

During each elapsed time ΔT, the hamiltonian energetics is fixed by the static part of the *Anderson potential*, whereas the time-dependent part, due to long range vibrational modes, varies in conjunction with the associated overlap integrals modulations [23].

To extract the energy-dependence of the coherence length scale, we proceed as follows. In the weak localization regime, the quantum correction of the Drude conductance is computed by solving the so-called Cooperon equation [21]. Assuming a quasi-1D geometry of the system, it has been shown that decoherence either due to e-ph or electron-electron scattering is described within the same scheme, and that the conductance reads $G_{Kubo}(E) = \frac{2e^2}{h}\left(N_\perp(E)\frac{\ell_e(E)}{L(E,t)} - \delta G_{WL}(E)\right)$ where $L(E,t)$ is the length scale that is energy-dependent due to velocity $v(E)$, and scales as $L(E,t) = \sqrt{v(E)\ell_e t}$, in the diffusive regime, whereas the term $\delta G_{WL}(E)$ gives the contribution of QIE beyond the scale of ℓ_e. Within the weak localization theory, and for quasi-1D systems, this contribution is shown to be related to the coherence length $L_\phi(E)$ as $\delta G_{WL}(E) = L_\phi(E)/L(E,t)$ whereas $\tau_\phi(E) = L_\phi^2(E)/v(E)\ell_e$. Therefore, by exploring the scaling behavior of $\delta G_{WL}(E)$, one can access relevant physical information about the fluctuations of the coherence length scales in the weak localization regime [23].

Fig. 2.4(left panel) shows the conductance computed at two different evolution times, $t = 3500\hbar/\gamma_0$ and $t = 35000\hbar/\gamma_0 \simeq 8$ps (dashed curves). One clearly sees a downscaling of the conductance with time, that comes from the classical linear downsizing in the diffusive regime $\ell_e(E)/L(E,t)$, together with the increasing QIE contribution of $\delta G_{WL}(E,t)$ with time (or equivalently length) scale. In the same figure, one also reports the classical term $G_{\rm class}(E) = N_\perp(E)G_0\ell_e(E)/L(E,t)$(bold curve), along with the quantum interference term $\delta G_{WL}(E,t) = G_{Kubo}(E,t) - G_{\rm class}(E)$ (dotted curve), at $t = 35000\hbar/\gamma_0$.

These results are obtained for $W = 0.07\gamma_0$ and no phonon dephasing, and by adding the time dependence modulations of integral overlap associated to the TW mode does not bring any appreciable changes [23].

Fig. 2.4 (right panel) gives $L_\phi(E)$ (main frame) and $\tau_\phi(E)$ (inset) for $W = \{0.1\gamma_0, 0.2\gamma_0, 0.5\gamma_0\}$ (bold, dotted and dashed curves respectively). The values range within $[10\text{nm}, 1000\text{nm}]$ in the considered energy window. One notes that for energies ≤ 0.9eV and within the considered evolution time, the quantum correction $\delta G(E) \simeq 0$, so that no meaningful information about the coherence length can be deduced since transport remains quasi-ballistic. In contrast, the coherence time shows reversed behavior, owing to the strong decrease of ℓ_e.

In the experimental situation [19], the fluctuation of $L_\phi(E)$ due to electron-electron scattering were found to scan the range $[10\text{nm}, 60\text{nm}]$, with systematic decrease near the onsets of new subbands. The values given here (for the chosen evolution time) are thus physically reasonable, since e-ph scattering is expected to lead to weaker decoherence effect [22].

2.2.4 Weak localization regimes and magnetic fields effects

Hereafter the weak localization phenomena in metallic carbon nanotubes are illustrated, on the basis of numerical results [24]. We first discuss for the metallic $(9, 0)$ nanotube, the various transport regimes under magnetic field. In Fig.2.5, the behavior of the field-dependent diffusion coefficients are shown in regards to the value of ℓ_e, that is modulated following its analytical relation to disorder strength W. Indeed, by using the Anderson-type disorder, the value of ℓ_e can thus be tuned by the disorder strength W, and the several cases of interest can be explored.

First, the weak localization regime [21] is analyzed under the condition $\ell_e < |\mathcal{C}_h| < L(\tau_\phi)$. Fig. 2.5 shows that the diffusivity increases at low fields (negative magnetoresistance) and that the periodic Aharonov-Bohm oscillations are dominated by a $\phi_0/2$ period, i.e., $\mathcal{D}(\tau_\phi, \phi+\phi_0/2) = \mathcal{D}(\tau_\phi, \phi)$ in agreement with weak localization theory. In contrast, when $\ell_e > |\mathcal{C}_h|, L(\tau_\phi < 2\ell_e)$, the system exhibits a *positive magnetoresistance* associated with $\mathcal{D}(\tau_\phi, \phi + \phi_0) = \mathcal{D}(\tau_\phi, \phi)$. For the case $\ell_e > |\mathcal{C}_h|, L(\tau_\phi > 2\ell_e)$, negative magnetoresistance and Aharonov-Bohm oscillations with period ϕ_0 are obtained. Note that with the analytical formula for the mean free path and estimates of disorder values, one gets $\ell_e \simeq 10^4 \times |\mathcal{C}_h|$, by using $|\mathcal{C}_h|$ as the circumference of the outer nanotube in the experiment of Bachtold and coworkers [25]. This leads to some inconsistency since the theoretical value of the mean free path is a priori too large to be consistent with a $\phi_0/2$ Aharonov-Bohm oscillation. Although the bandgap opening and oscillations, as well as other bandstructure changes (van-Hove singularities splitting and shifting) are likely to be smoothen by disorder, the magnetofingerprints will obviously result from a entangled situation, that goes much beyond the conventional theory of weak localization [24]. Several early experiments suggested such additional complexity in analyzing magnetotransport measurements [26, 27]. Roche and Saito [12] theoretically found that for a fixed disorder strength, magnetotransport fingerprints present strong fluctuations as a function of Fermi level position, CNT diameter, and orientation of the magnetic field with respect to the tube axis. Thanks to the engineering of efficient electrostatic gating of the nanotube, some experimental evidence of such multiple Aharonov-Bohm effects, with a $\phi_0/2$ oscillation driven by weak

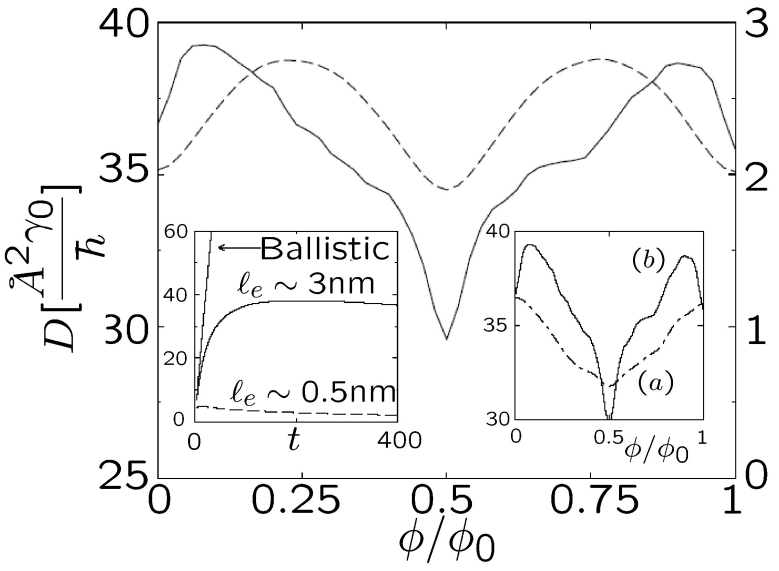


Figure 2.5: Diffusion coefficient $\mathcal{D}(\tau_\phi, \phi/\phi_0)$ (in units of $\mathring{A}^2$-$\gamma_0/\hbar$) for the (9,0) nanotube evaluated at time $\tau_\phi \gg \tau_e$, for two disorder strengths, $W/\gamma_0 = 3$ and 1, such that the mean free path ($\ell_e \sim 0.5$ and 3 nm, respectively) is either shorter (dashed line) or larger (solid line) than the nanotube circumference ($|\mathcal{C}_h| \sim 2.3$ nm). The right y-axis is for the dashed line and the left y-axis is for the solid line. Inset: $\mathcal{D}(\tau_\phi, \phi/\phi_0)$ for $\ell_e = 3$ nm and $L(\tau_\phi) < 2\ell_e$.

localization, superimposed to ϕ_0-periodic resistance fluctuations related to bandstructure modulations have been reported [28].

2.2.5 Quantum Transport and localization effects in chemically doped nanotubes

The possibility to incorporate chemical impurities as substitutions of carbon atoms has been demonstrated experimentally [29] and offers novel possibilities to investigate coherent charge transport, and magnetoresistance phenomena in chemically modified carbon nanotubes. Substitutional doping by nitrogen or boron impurities has been a very intense research topic at the theoretical level during the recent years [30, 31]. Initial works focused on the effect of a single isolated defect on electronic and transport properties, while further studies have addressed the issue of mesoscopic transport in μms long nanotubes with random distributions of impurities. These transport methods are mainly based on Kubo or Landauer-Büttiker frameworks, and mostly employed *ab initio* calculations combined with semi-empirical $\pi - \pi^*$ hamiltonian [32, 33, 34]. Such studies have allowed to explore the fundamental elastic transport length scales (elastic mean free path $\ell_e(E)$), and to investigate quantum interferences phenomena bringing the system from the weak to the strong localization regime.

To elaborate an effective tight-binding model able to describe the physics around the Fermi level, it is sufficient to properly describe the long-range scattering potential due to the chemical impurities (assuming low density approximation, i.e. no possible interferences between tails of individual impurity potentials). The long-range variations of the on-site and hopping parameters around impurities can be

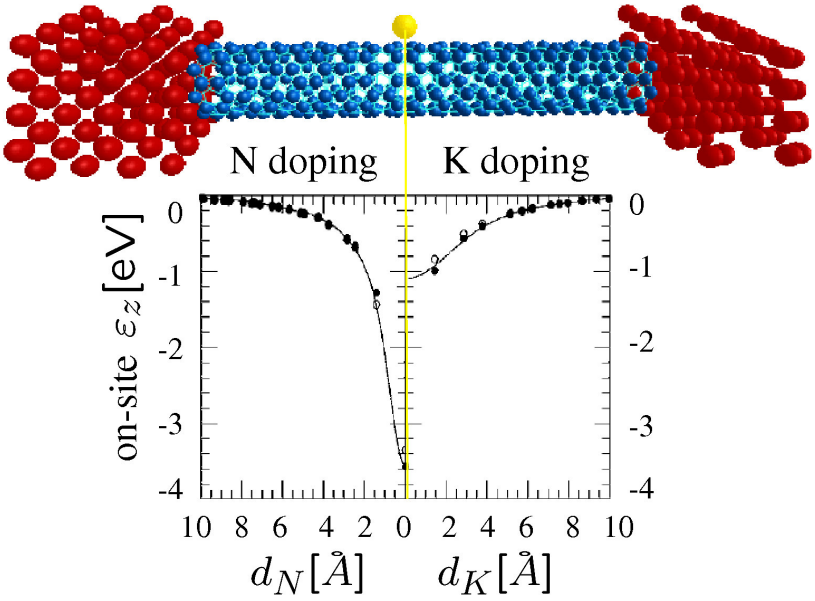


Figure 2.6: Top: representation of a single adsorbed potassium atom on a CNT connected in between two metal contacts. Bottom: long-range variations of the on-site and hopping parameters around K and N impurities derived from *ab initio* calculations [32].

directly derived from *ab initio* calculations performed using atomic-like basis [32]. The obtained tight-binding parameters allow to reproduce the position of the quasibound states in perfect agreement with the *ab initio* calculation [30]. As a matter of illustration, in Fig.6.35 the evolution of the ab initio on-site hamiltonian matrix elements associated with the p_z orbitals (perpendicular to the plane of atoms) as a function of the distance to the impurity in a doped (12×12) graphene sheet [32]. The potential well created by N in substitution is clearly much deeper that the one associated with the partially screened K+ ion. In particular, the ability of adsorbed K ions to trap electrons is significantly reduced as compared to N impurities.

In Fig.6.32, the Landauer conductance computed from the ab-initio method (left) and the tight-binding model for a single nitrogen doped $(10, 10)$ armchair nanotube are shown (right-inset). At selected energies (left inset:arrows), the conductance scaling properties are shown for a fixed impurity density $n_{\text{doping}} = 0.1\%$ (main frame). The extraction of the elastic mean free path ℓ_e is achieved by adding the contribution of the ballistic term to the diffusive one, *i.e.* $R = 1/G = R_0/N_\perp + R_0/(N_\perp L_{\text{tube}}/\ell_e)$, where $R_0 = h/2e^2$ is the resistance quantum and $N_\perp$ the number of available transverse modes at a given energy. When $\ell_e/L_{tube} \gg 1$, the statistical distribution of T is found to be narrowed and centred around $N_\perp$, in agreement with a ballistic limit $G = G_0 N_\perp$, with $G_0 = 1/R_0$ the conductance quantum. The other asymptotic case is found when $\ell_e/L_{tube} \ll 1$, where the distribution of T becomes wider with a mean value downscaling with the tube length as $\overline{T}(E) = N_\perp(E)\ell_e(E)/L_{\text{tube}}$. The conductance downscaling at a given energy exhibits a crossover from a ballistic to a diffusive regime.

Fig.2.8 shows the full energy-dependence of the elastic mean free path (ℓ_e), evaluated from the Kubo

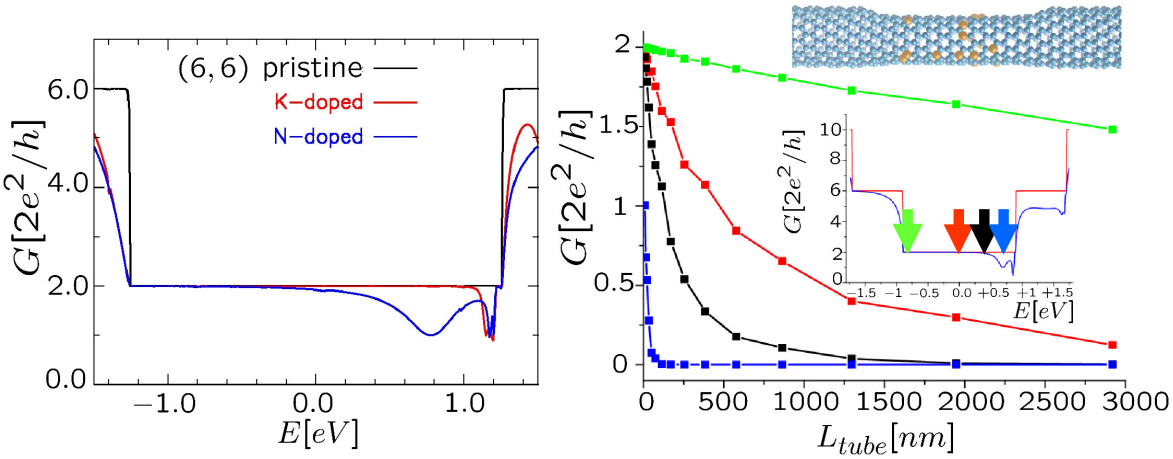


Figure 2.7: Left: Ab-initio calculation of the conductance for a single nitrogen impurity in substitution of one carbon atom in the (6,6) metallic nanotube [32]. Right: Length-dependence of the Landauer conductance for the disordered (10,10) nitrogen doped nanotube at several energies (doping is fixed to 0.1%). Inset: Conductance versus energy for the perfect (dashed line) and single-impurity (solid line) cases for a single defect. Arrows show the considered energies for the scaling analysis (main frame) [34].

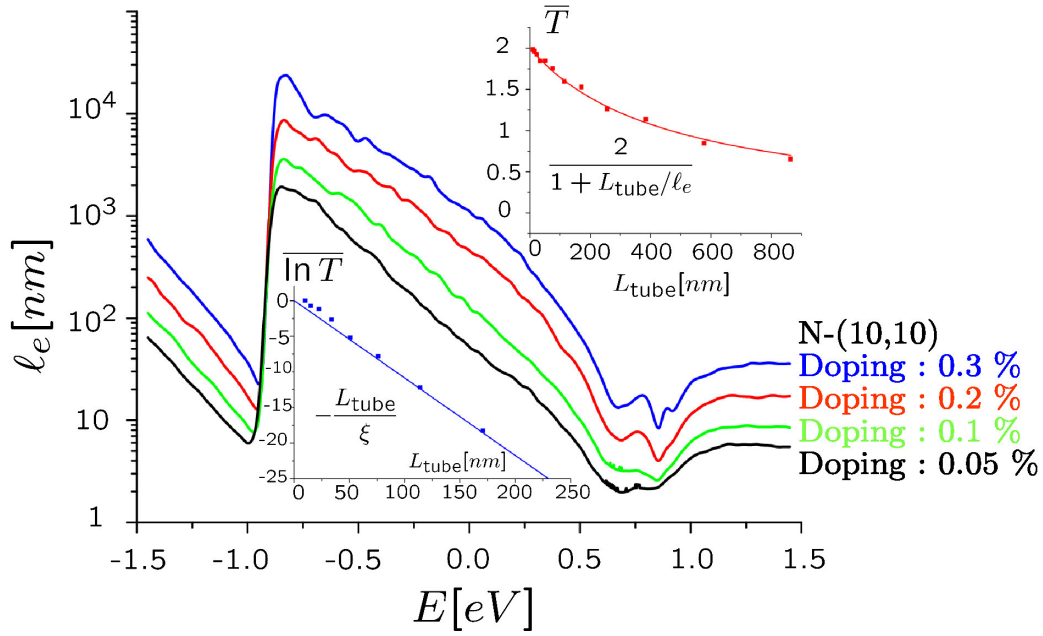


Figure 2.8: Main frame: ℓ_e for several values of n_{doping} (from lower to upper curve: 0.3%, 0.2%, 0.1%, 0.05%). Inset(top): $\overline{T}$ at CNP, for an average over 200 configurations($n_{\text{doping}} = 0.1\%$). The linear fit (dashed) directly gives ℓ_e. Inset(bottom): $\overline{\ln T}$ at $E = 0.69$eV, with linear fit (dashed) giving access to $\xi(n_{\text{doping}} = 0.1\%)$.

formula, for several values of the doping density. It is readily shown that over the whole spectral window $\ell_e \sim 1/n_{\rm doping}$ which is expected from the FGR. At charge neutrality point, the scaling of ℓ_e shows a linear increase with the radius of the tube r as found in a simpler Anderson model of disorder (not shown here). The comparison between ℓ_e extracted from both the Kubo and the Landauer formalism are in very good agreement over the full spectrum at a quantitative level [34]. For instance, the top inset of Fig.2 shows the averaged transmission as well as the fitting curve $2/(1 + L_{\rm tube}/\ell_e(E))$, that yields $\ell_e(E = 0.00) = 495.5 \pm 17.4$nm in excellent agreement with the Kubo calculation that gives ≈ 460nm. For an energy at the frontier of the first subband below the CNP, *i.e.* $E = -0.78$eV (Fig.1-a), the conductance slowly decays with length, and the regime remains quasiballistic. This is consistent with the calculated mean free path $\ell_e(E = -0.78) = 8371.6 \pm 69.4$nm which is much larger than the maximum length ($L_{tube} = 3000$nm) of the tube in between contact probes. For energies close to the nitrogen quasibound states (Fig.6.32), the impurity-induced backscattering becomes very strong, yielding a very small mean free path (see also Fig.2.8). On the "s wave" resonance, one actually finds that $\ell_e(E = 0.69) = 8.2 \pm 0.8$nm. For length larger than the mean free path, the conductance becomes exponentially reduced with length (Fig.6.35, curves c and d), defining a localized regime with ξ the length scale that quantifies the exponential decay of $\exp\left(\overline{\ln T}\right)$ [20]. Accordingly, one gets $\overline{\ln T}(E) = -L/\xi(E)$ [34].

2.2.6 Incommensurability and anomalous quantum transport in multiwalled carbon nanotubes

Thanks to the order N calculation, we were able to establish a deep connection between spectral and dynamical properties of incommensurate multi-walled nanotubes (MWNTs) [24, 35]. The MWNTs are indeed mostly intrinsic incommensurate objects since, due to registry mismatch between neighboring shells, there are very few cases in which the respective symmetries of individual shells allow a common unit cell for the whole object. In most situations, the unit cell length (along the nanotube axis) ratio between adjacent shells is an irrational number, and the MWNT taken as a whole becomes an incommensurate object.

The study of energy dependent diffusion coefficients in micron long multiwalled carbon nanotubes with intrinsic incommensurability between neighboring shells, demonstrates that the conduction regime evolves from sub-ballistic to diffusive motion as the Fermi level moves to subbands regions of large density of states, as a consequence of enhanced contribution of the underlying aperiodic potential and multiple scattering phenomena.

In incommensurate systems, such as the $(6,4)@(17,0)@(15,15)$ triple wall nanotube, the energy-dependent anomalous conduction due to geometrical incommensurability-induced aperiodic potential is particularly strong. The region around the charge neutrality point remains almost ballistic, as expected from the suppression of the intershell coupling at low energies due to helicity-determined selection rules [36]. In contrast, the rest of the electronic spectrum shows a very slow expansion of the wave packet in time. In Fig. 2.9 ones shows the time-dependent evolution of the diffusion coefficient (left) together

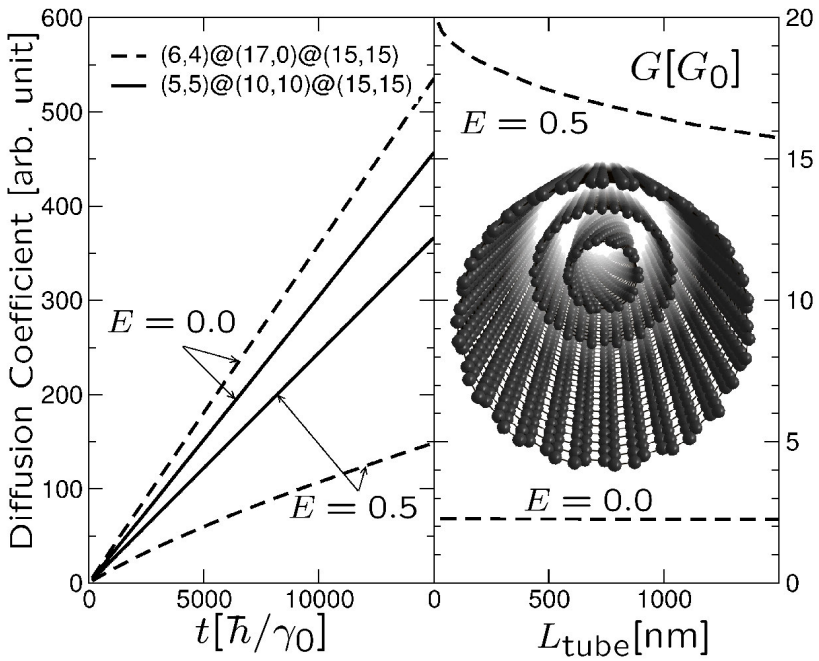


Figure 2.9: *Left*: time dependent diffusion coefficient for incommensurate/commensurate disorder-free MWNTs (with $\beta = \gamma_0/8$). *Right*: length dependence of conductance for two Fermi energies for $(6,4)@(17,0)@(15,15)$. Adapted from Triozon:PRB04

with the length dependence of the Kubo conductance (right) by defining the diffusion coefficient as

$$D(E,t) = \frac{X^2(E,t)}{t} = \frac{\mathrm{tr}\delta(E-\hat{\mathcal{H}})(\hat{\mathcal{X}}(t)-\hat{\mathcal{X}}(0))^2}{\mathrm{tr}\delta(E-\hat{\mathcal{H}})\,t}\,.$$

$D(E,t)$ either shows anomalously slow diffusion, or saturates at long times. Whenever the saturation limit is reached, a mean free path $\ell_e(E)$ for the whole object can be meaningfully extracted. The possibility to get anomalous diffusion and elastic mean free path intrinsically associated to an aperiodic potential of an otherwise clean system stands a unprecedented feature in Condensed Matter Physics. In contrast to quasiperiodic systems as model systems for quasicrystals [10, 37, 38], the simplified tight-binding description in MWNTs correctly reproduce the electronic bandstructure of real nanotubes, so a quantitative comparison between theory and experiments is possible! Despite recent efforts, the unique transport properties of MWNTs predicted theoretically remain to be experimentally confirmed.

2.3 Bi-orthogonalization process and Landauer-Büttiker conductance calculation

We have also developed a similar order N method for the calculation of Landauer-Büttiker conductance [39] in quasi-1D materials. This formula has the advantages to be general, independent of the geometrical

complexity of the system. In contrast to decimation-type approaches, this computational scheme allows order N in the direction perpendicular to the transport direction.

In brief, to express an analytical form for the conductance of the system under study, it is convenient to dissociate it from the two metallic external electrodes that serve to inject current in and out of the system, and that allow to suitably assume a well-given form for incoming and outgoing scattering states. These electrodes are further connected to measurement reservoirs, in which energy dissipation finally takes place.

We have shown that the problem can be reduced to the evaluation of $\langle\psi|G^r|\psi\rangle$, where $|\psi\rangle$ is a normalized state. To implement the recursion approach in the Landauer framework [40], a generalization to the case of *non-symmetric matrices* was performed. Indeed, the effective appearing in the Green's function is here non-symmetric hamiltonian $\mathcal{H} = \mathcal{H}_{Sys} + \Sigma_L^r + \Sigma_R^r$ because of the presence of self-energy matrix elements ($\Sigma_{L,R}^r$) describing the finite system coupled to the electrodes. A generalization of the recursion method to the computation of electronic spectra of non-symmetric hamiltonian was unsuccessfully proposed [41], but its further implementation received little consideration until the work of C. Benoit *et al.*, who applied a similar approach in a totally different context [42]. As we could demonstrate in [39], the bi-orthogonalization process can be efficiently implemented to the Landauer- Büttiker formalism, which yields a stable, efficient novel computing tool for nanoelectronics. The scheme has corrected the initial errors made in [41]. The basis equations are as follows: starting from the normalized vector $|\psi\rangle$ and from the non-hermitian matrix $\mathcal{H}$, we construct a bi-orthogonal basis $\{|\psi_n\rangle, \langle\phi_n|\}$ defined as follows

$$|\psi_{n+1}\rangle = \mathcal{H}|\psi_n\rangle - a_{n+1}|\psi_n\rangle - b_n|\psi_{n-1}\rangle \tag{2.6}$$

$$\langle\phi_{n+1}| = \langle\phi_n|\mathcal{H} - \langle\phi_n|a_{n+1} - \langle\phi_{n-1}|b_n \tag{2.7}$$

with the initial conditions $|\psi_{-1}\rangle = |\phi_{-1}\rangle = 0$, $|\psi_0\rangle = |\phi_0\rangle = |\psi\rangle$, and the bi-orthogonality condition $\langle\phi_n|\psi_m\rangle = 0$ if $n \neq m$. This last condition is equivalent to the following relations for a_n and b_n:

$$a_n = \frac{\langle\phi_n|\mathcal{H}|\psi_n\rangle}{\langle\phi_n|\psi_n\rangle} \tag{2.8}$$

$$b_n = \frac{\langle\phi_{n-1}|\mathcal{H}|\psi_n\rangle}{\langle\phi_{n-1}|\psi_{n-1}\rangle} = \frac{\langle\phi_n|\psi_n\rangle}{\langle\phi_{n-1}|\psi_{n-1}\rangle}. \tag{2.9}$$

The four equations (2.6), (2.7), (2.8), and (2.9) allow a recursive determination of the bi-orthogonal basis and of the coefficients a_n, b_n. Note that in "ket" notation, Eq. (2.7) must be understood as: $|\phi_{n+1}\rangle = \mathcal{H}^\dagger|\phi_n\rangle - a_{n+1}^*|\phi_n\rangle - b_n^*|\phi_{n-1}\rangle$. One starts from $|\phi_0\rangle = |\psi_0\rangle = |\psi\rangle$. At step 0, one computes $\mathcal{H}|\psi_0\rangle$ and $a_1 = \langle\phi_0|\mathcal{H}|\psi_0\rangle/\langle\phi_0|\psi_0\rangle$ by expanding all the amplitudes within the tight-binding localized basis. $|\psi_1\rangle$ and $|\phi_1\rangle$ are then obtained by computing $\mathcal{H}|\psi_0\rangle - a_1|\psi_0\rangle$ and $\mathcal{H}^\dagger|\phi_0\rangle - a_1^*|\phi_0\rangle$, while the first coefficient b_1 is subsequently deduced from Eq. (2.9). At step 1, $\mathcal{H}|\psi_1\rangle$ is computed together with $a_2 = \langle\phi_1|\hat{\mathcal{H}}|\psi_1\rangle/\langle\phi_1|\psi_1\rangle$. Then $|\psi_2\rangle$ and $|\phi_2\rangle$ result from the computation of vectors $\mathcal{H}|\psi_1\rangle - a_2|\psi_1\rangle - b_1|\psi_0\rangle$

and $\mathcal{H}^\dagger|\phi_1\rangle - a_2^*|\phi_1\rangle - b_1^*|\phi_0\rangle$. Finally, the coefficient b_2 is then deduced from Eq. (2.9). Steps $n \geq 2$ are fully similar to step 1. In the basis $\{|\psi_n\rangle\}$, $\mathcal{H}$ has thus a tridiagonal form:

$$H = \begin{pmatrix} a_1 & b_1 & & & \\ 1 & a_2 & b_2 & & \\ & 1 & a_3 & b_3 & \\ & & 1 & . & . \\ & & & . & . \end{pmatrix}. \tag{2.10}$$

Hence the recurrence relations (2.6) and (2.7) lead to a non-symmetric matrix and to a non-normalized bi-orthogonal basis. By choosing a different convention, a symmetric tridiagonal matrix and/or a normalized basis could be obtained. The quantity $\langle\psi|G^r(z = E \pm 0^+)|\psi\rangle = \langle\phi_0|\frac{1}{z-\mathcal{H}}|\psi_0\rangle$ can then be computed by the continued fraction method. It is equal to the first diagonal element of $(z - H)^{-1}$ where H is the tridiagonal matrix (2.10). Let us call $G_0(z)$ this matrix element and define $G_n(z)$, the first diagonal element of the matrix $(z - H_n)^{-1}$, with H_n the matrix H without its n first lines and columns:

$$H_n = \begin{pmatrix} a_n & b_n & & & \\ 1 & a_{n+1} & b_{n+1} & & \\ & 1 & a_{n+2} & b_{n+2} & \\ & & 1 & . & . \\ & & & . & . \end{pmatrix}. \tag{2.11}$$

From standard linear algebra, it can be shown that

$$G_0(z) = \frac{1}{z - a_1 - b_1 G_1(z)}, \tag{2.12}$$

and replicating such algorithm one gets a continued-fraction of $G_0(z)$:

$$G_0(z) = \cfrac{1}{z - a_1 - \cfrac{b_1}{z - a_2 - \cfrac{b_2}{\ldots}}} \tag{2.13}$$

In contrast with the standard recursion method, the recursion coefficients a_n and b_n do not show any simple behaviour for large n. In the type of applications considered here, a simple truncation of the continuous fraction at sufficiently large n gave a good convergence. This method was tested on carbon nanotube based heterojunctions [39], with perfect agreement with more traditional decimation techniques [43].

2.4 Conclusion

In conclusion, we have presented novel methodologies that can provide to scientific community efficient and optimized order N schemes, for deepening basic transport length scales in complex materials and nanodevices. Current efforts are devoted to extend the possibilities to situations far from the equilibrium, which stands as a great challenge; this include simulating nanodevices submitted to large bias voltages, and dissipation effects within the conducting channels such as those driven by electron-phonon interactions [44].

Bibliography

[1] J. M. Soler, E. Artacho, J. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, J. Phys.: Condens. Matter **14**, 2745-2779 (2002).

[2] R. Haydock, in *Solid State Physics*, edited by H. Ehrenreich, F. Seitz and D. Turnbull, (Academic Press New-York, 1980), Vol. **35**, 215.

[3] *Recursion Method and its Applications*, edited by D.G. Petitfor and D.L. Weaire, Springer Series in Solid States Sciences, Vol. **58** (Springer Verlag, Berlin, 1985). *The Recursion Method: Application to Many-Body Dynamics*, edited by V.S. Viswanath and G. Müller, Lectures Notes in Physics, Vol. **23** (Springer Verlag, Berlin, 1994).

[4] S. Goedecker, Rev. of Mod. Phys. **71**, 1085 (1999).

[5] T. Hoshi and Takeo Fujiwara, J. Phys. Soc. Jpn. **72**, 2429 (2003). R. Takayama, T. Hoshi, T. Sogabe, S.-L. Zhang, and Takeo Fujiwara, J. Phys. Soc. Jpn **73**, 1519 (2004). R. Takayama, T. Hoshi, T. Sogabe, S.-L. Zhang, and Takeo Fujiwara, Phys. Rev. B **73**, 165108, pp.1-9 (2006). Takeo Hoshi and Takeo Fujiwara, J. Phys.: Condens. Matter **18** 10787-10802 (2006).

[6] C. Lanczos, J. Res. Nat. Bur. Stand. **45**, 255 (1950).

[7] D.J. Lohrmann *et al*, Phys. Rev. B **40**, 8404 (1989).

[8] S.K. Bose, K. Winer, O.K. Andersen, Phys. Rev. B **37**, 6262 (1988). S.K. Bose *et al*, Phys. Rev. B **37**, 9955 (1988). S.K. Bose *et al*, Phys. Rev. B **41**, 7988 (1990).

[9] R. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957). D. Fisher and P.A. Lee, Phys. Rev. B **23**, 6852 (1981).

[10] S. Roche and D. Mayou, Phys. Rev. Lett. **79**, 2518 (1997).

[11] S. Roche, Phys. Rev. B **59**, 2284 (1999).

[12] S. Roche and R. Saito, Phys. Rev. Lett. **87**, 246803 (2001).

[13] A. Lherbier, M. Persson, Y.M. Niquet, F. Triozon and S. Roche, *submitted*

[14] S. Roche *t al.* in : *Understanding Carbon Nanotubes From Basics to Applications*, Series: Lecture Notes in Physics (Springer), **677** A. Loiseau, P. Launois, P. Petit, S. Roche, J.-P. Salvetat (Eds.) 2006.

[15] J.C. Charlier, X. Blase, and S. Roche, Rev. Mod. Phys. *79*, 677 (2007).

[16] C.T. White and T.N. Todorov, Nature **393**, 240 (1998).

[17] F. Triozon, S. Roche, A. Rubio, and D. Mayou, Phys. Rev. B **69**, 121410(R) (2004).

[18] S. Roche, G. Dresselhaus, M.S. Dresselhaus, and R. Saito, Phys. Rev. B **62**, 16092 (2000).

[19] B. Stojetz, C. Miko, L. Forro, and Ch. Strünk, Phys. Rev. Lett. **94**, 186802 (2005).

[20] D.J. Thouless, Phys. Rev. Lett. **39** 1167 (1977).

[21] A. Aronov, Y. Sharvin, Rev. Mod. Phys. **59**, 755 (1987).

[22] S. Chakraverty and A. Schmid, Phys. Rep. **140**, 193 (1986).

[23] S. Roche, J. Jiang, F. Triozon, and R. Saito, Phys. Rev. Lett. **95**, 076803 (2005).
S. Roche, J. Jiang, F. Triozon, and R. Saito, Phys. Rev. B **72**, 113410 (2005).

[24] S. Roche, F. Triozon, A. Rubio and D. Mayou, Phys. Rev. B **64**, 121401 (2001).

[25] A. Bachtold *et al.*, 1999, Nature **397**, 673. C. Schönenberger *et al.*, Appl. Phys. A **69** 283 (1999).

[26] A. Fujiwara *et al.*, Phys. Rev. B **60**, 13492 (1999).

[27] G. Fedorov, B. Lassagne, M. Sagnes, B. Raquet, JM. Broto, F. Triozon, S. Roche, and E. Flahaut, Phys. Rev. Lett. **94**, 066801 (2005).

[28] Ch. Strünk, B. Stojetz and S. Roche, Semicond. Sci. Technol. **21** S38-S45 (2006).

[29] Ch. Ewels and M. Glerup, J. Nanosci. Nanotech. **5**, 1354 (2005).

[30] H.J. Choi, J. Ihm, S.G. Louie and M. L. Cohen, Phys. Rev. Lett. **84**, 2917 (2000).

[31] A.H. Nevidomskyy, G. Csanyi, M.C. Payne, Phys. Rev. Lett. **91**, 105502 (2003).

[32] Ch. Adessi, S. Roche, and X. Blase, Phys. Rev. B **73**, 125414 (2006).

[33] S. Latil, S. Roche, D. Mayou, and J.C. Charlier, Phys. Rev. Lett. **92**, 256805 (2004).
S. Latil, F. Triozon, and S. Roche, Phys. Rev. Lett. **95**, 126802 (2005).

[34] R. Avriller, S. Latil, F. Triozon, X. Blase, and S. Roche, Phys. Rev. B **74**, 121406 (2006).

[35] S. Roche *et al.*, Phys. Lett. A **285**, 94 (2001). S. Wang, M. Grifoni, and S. Roche Phys. Rev. B **74**, 121407(R) (2006).

[36] S. Wang and M. Grifoni, Phys. Rev. Lett. **95**, 266802 (2005).

[37] T. Fujiwara, in *Physical Properties of Quasicrystals*, Springer Series in Solid-States Sciences Springer-Verlag, Berlin, (1998).

[38] S. Roche and Takeo Fujiwara, Phys. Rev. B **58**, 11338 (1998).

[39] F. Triozon and S. Roche, Eur. Phys. J. B 46, 427431 (2005).

[40] R. Landauer, IBM J. Res. Dev. **32**, 306 (1988). M. Büttiker, IBM J. Res. Dev. **32**, 317 (1988).

[41] R. Haydock and M. J. Kelly, J. Phys. C. **8**, L290 (1975).

[42] C. Benoit *et al*, Modelling Simul. Mater.Sci. Eng. **3**, 161 (1995). C. Benoit and G. Poussigue, Eur. Phys. J. AP **23**, 117 (2003).

[43] F. Triozon, Ph. Lambin, S. Roche, Nanotechnology **16**, 230 (2005).

[44] L.E.F. Foa Torres and S. Roche, Phys. Rev. Lett. **97**, 076804 (2006).

Chapter 3

Quantum Transport in Carbon Nanostructures

Jean-Christophe Charlier
Unité de Physico-Chimie et de Physique des Matériaux (PCPM),
Université Catholique de Louvain,
1 Place Croix du Sud, B-1348 Louvain-la-Neuve, Belgium

3.1 Introduction

For about twenty years, more and more new materials have been virtually studied in computers, and no more as traditionally using an empirical approach based on successive fittings or lucky hazards. Such a remarkable rise of the numerical simulations is probably due to the conjunction of two facts : the emergence of new theories and methodologies allowing predictive simulations on realistic systems, on one hand, and the incredible increase of the computer power, on the other hand. Physicists may calculate a priori the properties of a material which exists or not, creating a new domain of physics, the numerical-simulation experiment. Many successes validated such an approach which illuminates in a complementary way the properties of matter and leads to a highest understanding of the physical mechanisms brought into play. The development of new scientific fields or research directions at the frontier of various scientific disciplines often needs to be linked to intensive numerical simulations. For example, the initiatives launched by several countries in nanosciences include a strong component in numerical simulations.

3.2 Brief outline of the techniques

Among the various numerical techniques used to predict the electronic structure, those based on the density functional theory (DFT) [1, 2] are becoming more and more popular in physics and quantum

chemistry. Although starting from quantum mechanics first-principles, these methods allow the study of atomic systems containing a quite important number of atoms (a few hundreds) with reasonable computational means. Our group has been using this technique during several years to investigate the electronic structures of various graphites and other carbon nanostructures (nanotubes, fullerenes, hybrid systems based on nanotubes,...) [3].

In the framework of this project, we would like to help the experimentalists in their quest to create atomic or molecular systems which are able to transfer an information at the nanoscale. Presently, several groups around the world are able to integrate into electronic devices molecular systems such as benzenedithiol, nanowires, nanotubes and graphene nanoribbons, and to measure their corresponding current-voltage (I/V) characteristics. This new field of research has generated a considerable experimental activity and has raised several theoretical challenges to understand and model electronic quantum transport. Moreover, during the last years, our theoretical works have mainly focused on quantum transport using the Kubo-Greenwood formalism in real space [4] and semi-empirical approaches (tight-binding). In such a framework, the phenomena associated to the reduced dimensionality, to the quantum confinement, and to the loss of phase coherency due to defects, impurities or other aperiodic perturbations, can easily be investigated. In addition, this formalism illustrates specifically the link between the conductivity and the quantum diffusion whose value is directly connected to the average quadratic spreading of a wave-packet which diffuses inside the nanostructure. The conductivity at zero temperature can be deduced from the time evolution of this variable, as well as the conduction regime (ballistic, diffusive or localized). More specifically, we have been using this formalism to investigate the quantum transport in nanotubes with a perfect structure, doped in substitution [5] and randomly decorated with organic molecules [6].

However, these semi-empirical simulations do not allow for accurate structural relaxations to be performed and do not yield correct values for the local electronic charges in the studied systems. In other words, the chemical potential is not uniform across the entire system in equilibrium. For these reasons, it is appropriate to generalize these development on quantum transport to pure ab initio formalisms which guarantee the uniformity of the chemical potential (at least at a mean field level). Consequently, it is quite reasonable to turn to the DFT which was so successful in many applications in condensed matter. However, in a transport experiment, the system is out of equilibrium. The difficulties to use the DFT in quantum transport are due to the fact that this theory and mostly its approximations (by way of the functionals) are only well established for electronic systems in their ground state, at the equilibrium, and when the properties are not dependent on the electronic correlations. If the DFT is the undeniable technique used today to predict the ground state properties of realistic materials, the state-of-the-art for the excited sates is not fixed. Different communities are using various approaches to treat this problem : TDDFT (Time Dependent DFT), GW (many-body Green functions), DMFT (Dynamical Mean Field Theory),

The Landauer-Büttiker formalism [7] supplies a simple framework to describe electronic quantum transport in nanostructures. In particular, the central equation of this theory connects the quantum conductance $G(E) = \frac{2e^2}{h}T(E)$ to the transmission function $T(E)$. The value of this electronic transmission is essentially determined by the region where the number of channels available for conduction is the

smallest. In nanotechnology, the region of relevance is the nanostructure itself forming the nanoscopic bridge between the metallic electrodes. The electronic transmission is thus strongly dependent on the particular nanostructure, the detailed atomic arrangement of the electrodes in the contact region, and their chemical nature.

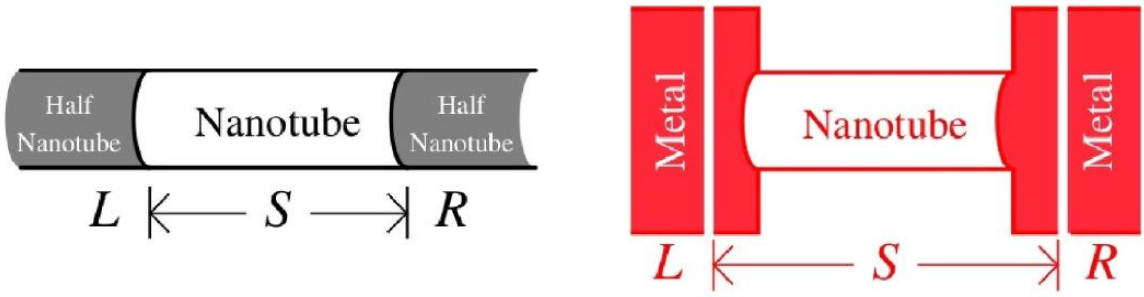


Figure 3.1: Quantum transport : modelisation of the contacts between a carbon nanostructure.

The modern theory of quantum transport gave rise to several methods to accurately calculate this transmission function whose estimation implies the good knowledge of the electronic structure of the nanoscopic system. These numerical simulations of transport experiments induce an open system when considering the nanostructure alone, or an infinite system constituted by the nanostructure and two semi-infinite electrodes which can be a prolongation of the ideal nanostructure or real metallic contacts. The electronic states of these bridged nanosystems are described by scattering states which extend over the whole region between two semi-infinite electrodes.

In a first approach, the Lippmann-Schwinger equation [8] is solved in order to obtain accurate scattering states in a plane-wave expansion or in real-space grids under finite bias voltages. However, the dimension of the matrix to be solved becomes so large as the system size increases that it is difficult to apply the method to large systems.

The Keldysh formalism [8], based on the integration of the non-equilibrium Green functions (NEGF) inside the DFT, is playing a key role. This approach consists in using the DFT to approximately describe the electronic structure of quasi-particles in nanostructures, in order to use these data as an input in the Landauer-Büttiker formalism to simulate the linear response in the limit of small differences of potential. In this framework, the investigated system is also divided into a diffusion region surrounded by two electrodes, and the conductance is expressed in terms of Green functions. In principle, in such a formalism, the non-equilibrium conduction regime, the effect of the contacts, the presence of an open system, and the effects of correlation (e-e interaction and e-phonon interaction) can be taken into account. The main problem is that NEGF is a complex theory which allows only to treat simple systems or small models. Presently, a calculation performed in the framework of the NEGF on todays computers is still a real challenge.

3.3 Challenges

In this project, we intend to use the NEGF formalism as implemented in the SIESTA [9] – SMEAGOL [10] and the ABINIT [11] – WANT [12] codes to investigate quantum transport. Both SIESTA and ABINIT are ab initio codes for the simulation of ground-state properties. SIESTA (localized-orbitals basis set, [9]), although inherently less accurate than ABINIT (plane-waves basis set, [11]), is able to treat a larger number of atoms. Moreover, two softwares specialized for the study of transport properties, named SMEAGOL [10] and WANT [12], are interfaced with SIESTA and ABINIT, respectively. The quantum transport properties will be investigated using these codes in various carbon nanostructures such as carbon nanotubes and graphene nanoribbons.

3.4 Opportunities for applications relevant in nanosciences

Their unusual electronic and structural physical properties promote carbon nanomaterials as promising candidates for a wide range of nanoscience and nanotechnology applications. Carbon is unique in possessing allotropes of each possible dimensionality and, thus, has the potential versatility of materials exhibiting different physical and chemical properties. Diamond (3D), fullerenes (0D), nanotubes (1D-CNTs), and 2D graphite platelets and graphene ribbons are selected examples. Because of their remarkable electronic properties, CNTs are expected to play an important role in the future of nanoscale electronics. Not only can nanotubes be metallic, but they are mechanically very stable and strong, and their carrier mobility is equivalent to that of good metals, suggesting that they would make ideal interconnects in nanosized devices. Further, the intrinsic semiconducting character of other tubes, as controlled by their topology, allows us to build logic devices at the nanometer scale, as already demonstrated in many laboratories.

However, like in most materials, the presence of defects in carbon nanotube has been demonstrated experimentally [13, 14] and alter this ideal situation. These atomic-scale defects may take different forms : vacancy, di-vacancy, “Stone-Wales” defect, adatom, heptagon/pentagon pair, atom in substitution, Those defects can appear at the stage of CNT growth and purification, or be deliberately created by irradiation, chemical treatment, or induced by uniderectional strains in order to achieve a desired functionality. They are known to drastically modify both the electronic and the transport properties of carbon nanotubes [15]. The presence of defects has thus become an essential part of diverse processes in carbon materials synthesis. A true realization of different devices requires a good understanding of both geometrical and electronic properties of not only the carbon nanostructures themselves, but also of their defects. For instance, structural and electronic properties of carbon nanomaterials could be tuned by such a presence in their structures. It was reported that vacancy defects created by electron irradiations could induce structural changes within carbon nanotubes. Moreover, vacancies induced the occurrence of magnetism which could be useful in some specific applications. Consequently, it is crucial to understand the properties of these defects in order to conquer their detrimental effects, but also because controlled defect introduction may be used to tune nanotube properties in a desired direction.

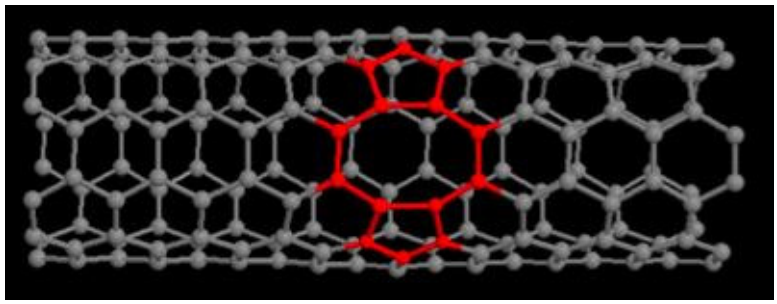


Figure 3.2: Defects in carbon nanotubes

Recently, the modifications induced by those defects in the electronic properties of the carbon hexagonal network have been investigated using first-principles calculations. Computed constant-current STM images of these defects have been calculated within a tight-binding approach in order to facilitate the interpretation of STM images of defected carbon nanostructures [16]. In addition, most of these topological defects induce localized non-dispersive electronic states close to the Fermi energy, thus modifying the intrinsic electronic properties of the carbon nanostructures. As these defects should also play a key role on the transport properties, the quantum conductance will be calculated for various topological defects incorporated in carbon nanotubes.

Due to the nano size of the systems under investigation, atomistic simulations become necessary for an accurate modelling of their transport properties. Indeed, effective bulk parameters cannot be used for the description of the electronic states since interfacial properties play a crucial role and semiclassical methods for transport calculations are not suitable at the typical scales where the device behavior is characterized by coherent tunnelling. Consequently, quantum-mechanical computations with atomic resolution can be achieved using localized basis sets for the description of the system Hamiltonian and can predict electronic and transport properties of nanostructures. In the present project, the structural and electronic properties of CNTs will be obtained via the ab initio method as implemented in the SIESTA code [9] while the quantum conductance of the system is achieved from electronic transport calculations performed with the SMEAGOL code [10]. The latter is based on the non-equilibrium Greens function (NEGF) formalism and uses the one-particle Hamiltonian obtained from the DFT calculations. Such approach, which combines NEGF and DFT, allows us to model real systems constituted by hundreds of atoms to a high degree of accuracy as mentioned in the previous section.

At last, among the different possible applications of CNTs, the present project will also focus on their use as nanosensors for gas detection thanks to the change in conductance of the nanodevice in presence of the gas to be detected. To reach this goal, the enhanced chemical reactivity of defected carbon nanotubes or tubes decorated with metal clusters will be investigated by predicting the modulation of the quantum conductance due to specific molecules adsorbed at the defect site in order to evaluate their potential sensing properties.

The field of nanotubes has further fostered much interest in related systems such as graphite platelets and graphene nanostcutures. The actual synthesis of isolated graphene sheets, or few layers graphite

materials [17] and the sudden interest generated by such a synthesis, comes nearly 15 years after that of nanotubes. This is an amazing twist of history as the properties of nanotubes are usually derived from that of graphene [3]. As in a such system the electrons are confined in 2D, the quantum transport properties of these nanostructures, measured experimentally under magnetic field, revealing a surprising and non-conventional physics due to the unique low-dimensionality of this material [18, 21]. Charge carriers around the charge neutrality point in graphene show a very high mobility with a Dirac-like behaviour, namely an energy-independent velocity [19, 20, 22, 23]. Indeed, in contrast with bulk graphite which is a 3D semimetal whose density of charge carriers strongly depends on the stacking of the its planes [24], the graphene sheet (or a few graphene layers) represents a semiconductor with a zero gap (or very small gap or semimetals with a very low density of charge carriers) [25]. In the case of graphene nanoribbons, confinement effects induce a width and orientation-dependent band-gap opening [26] and one is back to the size and helicity selection that complicates the use of nanotubes in applications. Other interesting properties, such as a half-metal behaviour, where spin-up and spin-down electrons display a different metallicity, have also been proposed [27].

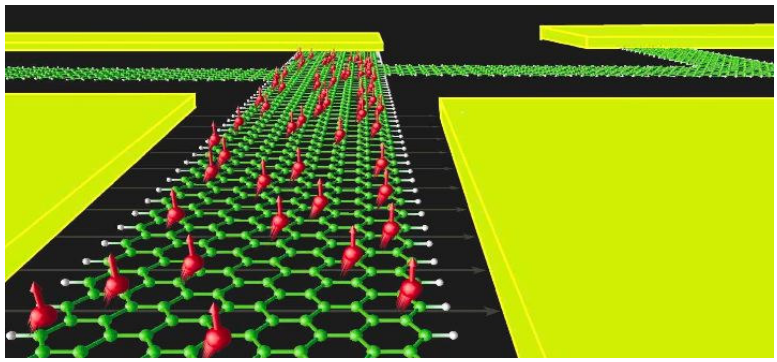


Figure 3.3: Spintronics in graphene nanoribbons, from Ref. [27].

The second part of the project will consist the quatum transport in the graphene sheet and in graphene nanoribbons where the electrons are confined in 2D. Our group plan to study electronic devices based on these graphene nanoribbons, connected to metallic electrodes in order to predict their electronic and transport properties. In this research, the first-principles simulations will mainly focus on the surface reconstruction of the graphene nanoribbon (with or without hydrogen) and also on the metallic contact - graphene sheet interface. The I/V characteristics will be predicted ab initio for several model systems.

Bibliography

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964)

[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965)

[3] J.-C. Charlier, X. Blase, and S. Roche, Rev. Mod. Phys. **79**, 677-732 (2007)

[4] S. Roche, Phys. Rev. B **59**, 2284-2291 (1999)

[5] S. Latil, S. Roche, D. Mayou, and J.-C. Charlier, Phys. Rev. Lett. **92**, 256805 (2004)

[6] S. Latil, S. Roche, and J.-C. Charlier, Nano Lett. **5**, 2216-2219 (2005)

[7] R. Landauer, IBM J. Res. Dev. **1**, 223 (1957); M. Bttiker, Phys. Rev. Lett. **57**, 1761 (1986)

[8] S. Datta, Electronic transport in mesoscopic systems (Cambridge University Press, New York, 1995)

[9] SIESTA : http: //www.uam.es/departamentos/ciencias/fismateriac/siesta

[10] SMEAGOL : http: //www.smeagol.tcd.ie

[11] ABINIT : http: //www.abinit.org

[12] WANT : http: //www.wannier-transport.org

[13] A. Hashimoto, K. Suenaga, A. Gloter, K. Urita, and S. Iijima, Nature **430**, 870-873 (2004)

[14] K. Suenaga, H. Wakabayashi, M. Koshino, Y. Sato, K. Urita, and S. Iijima, Nature Nanotechnology **2**, 358-360 (2007)

[15] C. Gomez-Navarro, P.J. De Pablo, J. Gomez-Herrero, B. Biel, F.J. Garcia-Vidal, A. Rubio, and F. Flores, Nature Materials **4**, 534-539 (2005)

[16] H. Amara, S. Latil, V. Meunier, Ph. Lambin, and J.-C. Charlier, Phys. Rev. B, in press (2007).

[17] K.S. Novoselov, A.K. Geim, S.V. Morozov, D. Jiang, S.V. Dubonos, I.V. Girgorieva, and A.A. Firsov, Science **306**, 666 (2004)

[18] K.S. Novoselov, A.K. Geim, S.V. Morozov, D. Jiang, M.I. Katsnelson, I.V. Girgorieva, S.V. Dubonos, and A.A. Firsov, Nature **438**, 197 (2005)

[19] K.S. Novoselov, E. McCann, S.V. Morozov, V.I. Fal'ko, M.I. Katsnelson, U. Zeitler, D. Jiang, F. Schedin, and A.K. Geim, Nature Physics **2**, 177 (2006)

[20] C. Berger, Z. Song, X. Li, X. Wu, N. Brown, C. Naud, D. Mayou, T. Li, J. Hass, A.N. Marchenkov, E.H. Conrad, P.N. First, and W.A. de Heer, Science **312**, 1191 (2006)

[21] Y. Zhang, Y.W. Tan, H.L. Stormer, and Ph. Kim, Nature **438**, 201 (2005)

[22] T. Ohta, A. Bodtwick, T. Seyller, K. Horn, and E. Rotenberg, Science **313**, 951 (2006)

[23] S.Y. Zhou, G.-H. Gweon, J. Graf, A.V. Fedorov, C.D. Spataru, R.D. Diehl, Y. Kopelevich, D.-H. Lee, S.G. Louie, and A. Lanzara, Nature Physics **2**, 595 (2006)

[24] J.-C. Charlier, X. Gonze, and J.-P. Michenaud, Phys. Rev. B **43**, 4579 (1991); Carbon **32**, 289 (1994)

[25] S. Latil and L. Henrard, Phys. Rev. Lett. **97**, 036803 (2006)

[26] Y.-W. Son, M.L. Cohen and S.G. Louie, Phys. Rev. Lett. **97**, 216803 (2006)

[27] Y.-W. Son, Y.-W., M.L. Cohen, and S.G. Louie, Nature **444**, 347 (2006)

Chapter 4

Atomistic Potentials for Modeling of Microelectronics Materials

S. Monaghan, A. J. de Vries, V. Bourenkov, and J. C. Greer
Tyndall National Institute,
Lee Maltings, Prospect Row,
Cork, Ireland

4.1 Introduction

Scaling of semiconductor devices has led to prototype field effect transistors with gate lengths of only a few nanometers. The persistent scaling down of transistor sizes has continued to enable more powerful microprocessors, this in turn allows for simulations of longer length and time scales. It is now possible to routinely solve the Schrödinger equation or the equations of density functional theory for several hundred atoms, and in many cases, thousands of atoms. Hence very accurate atomic scale calculations can be applied to the design of the very technologies enabling the simulations. However, even for "small" semiconductor devices, atoms can easily number into the several million, and it is desirable to be able to treat larger atomic systems with less accurate- but with the advantage of being orders of magnitude computationally less demanding compared to first principles methods- simulations.

Forcefield, empirical, atomistic, interatomic potential, analytical potential, model potential, and molecular mechanics are a few of the names given to atomic scale simulations whereby the complex interactions governed by the quantum mechanics of many-electron systems are replaced with effective interatomic potential energies. The inter-atomic potential energies (and their derivatives with respect to nuclear positions, the forces) are represented by simple analytical functions used to describe, in an approximate way, atomic interactions within solids, liquids and gases.

This survey considers interatomic potentials that can be used for modeling semiconductors and insu-

lators, the materials at the heart of modern metal-oxide-semiconductor (MOS) transistors. For semiconductor materials, the focus of this review is on silicon and germanium, and their alloys Si_xGe_{1-x}. Several crystalline polymorphs of silicon dioxide are considered. Various interatomic potentials are discussed, and their suitability for describing specific material properties is summarized.

In simulating or modelling materials on an atomic level, it is required to determine the positions and movement of all the atoms comprising a material sample. Solving Schrödinger's equation delivers an answer to this problem in principle. However, solution of these equations for macroscopic materials is intractable and even for nanoscale systems, a quantum mechanical treatment of both the nuclear and electronic motion remains a demanding computational challenge. For advanced electronic structure calculations, the electronic and nuclear motions are decoupled, and the electronic degrees of freedom are treated by accurate quantum mechanical methods, whereas the nuclear motion (i.e. the "atomic" positions) is treated as the classical mechanics of point particles moving in a potential energy field determined by the electronic degrees of freedom and the nuclear-nuclear Coulomb interactions.

Molecular dynamics is a semi-classical scheme whereby the complex quantum mechanical interactions between atoms are represented by simple analytical expressions, but the motions of atoms are governed by classical mechanics. This semi-classical approximation certainly has its limitations, but has proven to be very successful in predicting structural, thermodynamic and mechanical properties of materials. The interaction potentials between atoms may be thought of as effective potentials averaging over the detailed electronic structure of the materials, hence the method is not suitable for determining detailed electronic and optical properties of materials.

There are many detailed and interesting overviews of molecular dynamics and molecular mechanics methods and applications. In this chapter the focus is on the evaluation of analytical potential energy functions (henceforth refererred to as "analytical potentials", or simply as "potentials") in use to describe materials of great importance to electronic device technologies: silicon, germanium, silicon germanium alloys, and silicon dioxide. The purpose of this survey is to give an overview of the strengths of analytical potential methods, to give and indication of where these methods fail, and to indicate areas for further research. The survey is not exhaustive, but a wide range of potential functions and their applications are considered.

4.2 Analytical potential energy functions

Molecular dynamics simulates the position and forces on individual atoms by integrating Newton's equations for atomic positions with forces derived from analytical potential functions chosen to model the Born-Oppenheimer energy surface. Structures and material specific properties can be computed from the procedure resulting in structures and many, but not all, material properties. The potential energy functions are based on simplified, but quantum-mechanically motivated, models. The primary aim in developing an analytical potential function is to produce an approximation that enables a set of specific properties to be reproduced over a wide range of conditions; for example, for varying temperature and pressure. Ideally the analytical potential will be able to reproduce many material properties such as

viscosity and mechanical strength, and be able to treat all structural phases of a material. Clearly this a challenge for even for a full quantum mechanical treatment: the ability of analytical potential methods to describe a wide range of properties over a wide range of conditions with relatively low computational demand is the strength of the approach. Note however, due to the simplified treatment of the *net* effect of electrons used to describe chemical bond formation, detailed information such as optical excitations and electron transport properties are not accessible with the method.

In principle, any set of n interacting particles may be represented as a sum of energies depending on a single position, a sum over energy terms determined by the coordinates of pairs, a sum over interacting triplets, and so forth up to the total number of n particles within the system. For many potential models only pairs and triplets of interactions between atoms are retained (the sum over individual atoms merely defines a reference energy); the pair and triplet interactions are referred to as two-body (or pair-wise additive) and three-body terms, respectively. Interaction terms involving the positions of four or more atoms are usually neglected in the representation of a model potential for many atom systems; however, it is not always clear that the higher order terms may be *a priori* neglected. However, by far the most common analytical potential representations for semiconductors and silicon oxides include only two- and three-body terms. Thus potentials including higher-order terms are not included in the present survey.

Two general categories for potentials are identified in the comparative study of Balamane, Halicioglu and Tiller [1]: cluster potentials and cluster functionals. Within the cluster potentials, fall the Stillinger-Weber [2] and related forms. These potentials are developed from a summation over the two- and three-body terms V_2 and V_3 between all atom pairs and triplets used to express the interaction energy of the system as

$$V = \sum V_2 + \sum V_3. \tag{4.1}$$

The pairwise-interactions are often chosen to be of the form:

$$V_2 = f_c(r_{ij}) \times \left[A_1\phi_1(r_{ij}) - A_2\phi_2(r_{ij})\right], \tag{4.2}$$

where r_{ij} represents the distance between atom pairs (distance between atomic nuclei). The cut-off function f_c serves to limit the range of the interaction and is modelled in different ways, but it is generally desirable that the overall potential energy function remain continuous, or at least the cut-off should not introduce excessive errors in energy conservation and properties calculations. The cut-off is introduced to limit the range of the interactions leading to numerical efficiencies; ϕ is a decaying function of r, and its form varies for specific applications. The A_i are parameters specific to a material system. The three-body terms are likewise expressed in many different forms, but as a generalization it may be stated that they are written as a product of radial and angular functions. This class of potential is also often referred to as short-ranged.

Cluster functionals also take the nature of an atom's environment into account. They include Tersoff

and similar potential forms consisting of repulsive and attractive terms summed over atoms:

$$V = \sum_{ij} f_c(r_{ij}) \left[A_1 \phi_1(r_{ij}) - A_2 \phi_2(r_{ij}) p(\zeta_{ij}) \right]. \tag{4.3}$$

The local environment of the atoms are described by p, which represents the Pauling bond order as a function of the effective coordination number ζ.

The first test of a potential function is often a check to see if it can be used to determine the crystalline form of a material. Other potentials are developed to describe amorphous materials, and generally this is a more challenging task as the functions must be able to describe a wider range of bond lengths and angles. In general, for technology applications, crystalline and amorphorous forms of materials are important, as well as impurities, random alloys, surfaces and interfaces. Clearly this is a demanding task for analytical potential functions and below we present an overview of how different potentials may be used, and as well, where they fail.

A brief description for a selection of commonly used potential forms is given. The procedure applied to the development of the potential model is described, along with the intended application domain. When available, an evaluation of the performance of the potential function as described by the developers is summarized.

To further improve or to customize a potential model, adjustments to gain accuracy for specific properties are often undertaken. This can take the form of a re-parametrization or by the addition of new terms to the energy expression. Unfortunately, this can also lead to a decrease in accuracy for other properties or in the ability to describe other material systems: when a potential energy expression may be used to describe a variety of systems and properties, it is referred to as transferable. However, even given the loss of transferabilty, "tuned" potentials can be of great advantage when applied to specific studies. Not all such modifications for the potential below have been described, but a few illustrative examples are considered.

4.2.1 Stillinger-Weber (SW) model and related forms

The Stillinger-Weber potential energy function [2] is widely applied to the study of semiconductor materials; in particular, silicon and carbon. It consists of a sum of two- and three-body interactions:

$$\upsilon_2(r_{ij}) = \varepsilon f_2(r_{ij}/\sigma) \tag{4.4}$$

$$\upsilon_3(r_i, r_j, r_k) = \varepsilon f_3(r_i/\sigma, r_j/\sigma, r_k/\sigma) \tag{4.5}$$

A five-parameter pair potential is chosen containing a cut-off function that has no discontinuities for r on $[0, a]$:

$$f_2(r) = \begin{cases} A(Br^{-p} - r^{-q}) \exp[(r-a)^{-1}], & r < a \\ 0, & r \geq a \end{cases} \tag{4.6}$$

and for the three-body terms f_3 the form is chosen as:

$$f_3(r_i, r_j, r_k) = h(r_{ij}, r_{ik}, \theta_{jik}) + h(r_{ji}, r_{jk}, \theta_{ijk}) + h(r_{ki}, r_{kj}, \theta_{ikj}) \tag{4.7}$$

where h also contains the cut off function parameter a,

$$h(r_{ij}, r_{ik}, \theta_{jik}) = \lambda \exp[\gamma(r_{ij} - a)^{-1} + \gamma(r_{ik} - a)^{-1}] \times \left(\cos\theta_{jik} + \frac{1}{3}\right)^2 \tag{4.8}$$

With this choice, seven parameters $A, B, p, q, a, \lambda, \gamma$ are determined to reproduce the lattice stability and relative stabilities of various crystalline forms and when appropriate, melting points and correlation function in the liquid phase. When results using the the potential are compared to experimental quantities, the Stillinger-Weber potential generally performs well, *or* in the cases where it does fail, the reason for its deficiencies are well understood.

Ding and Andersen [3] have adjusted the Stillinger-Weber potential to describe crystalline and amorphorous forms of germanium. The Stillinger-Weber potential is fitted to experimental values for cohesive energy at T= 0 K and density at a pressure of 1 atmosphere. As well, the elastic constants of the crystal and the radial distribution function (rdf) for the amorphorous form (a-Ge) are reproduced. To determine the rdf, three-body terms are essential, and the rdf for amorphous silicon (a-Si) using the original Stillinger-Weber parameter set for Si does not reproduce well the experimental rdf for a-Si. The structure of amorphous germanium is simulated by quenching the liquid phase at constant volume using the Ding and Anderson parameter set for Ge. The resulting structure, rdf and phonon dispersion of a-Ge is found to be in good agreement with experiment for this "tuned" parameter set. However, the properties of liquid Ge are not modelled well, and it has been suggested that different parameter-sets should be used for different phases when applying the Stillinger-Weber model.

Laradji, Landau and Dunweg [4] adjusted the Stillinger-Weber potential for Si-Ge alloys and focused on determination of structural properties. Deduction using the formulae and parameters derived for silicon by Stillinger and Weber and for germanium by Ding and Andersen [3] results in a potential model for $Si_{1-x}Ge_x$ alloys. Monte-Carlo simulations using the new parameter set, modified to treat two atom types, were performed and alloy structures were constructed. Critical behaviour at phase transitions were simulated showing the mixing of Si and Ge atoms. Lattice properties, bond length and angles were also calculated and found to be in relatively good agreement with experiment. Thermal expansions and chemical potentials using the potentials were also calculated.

Several articles on application of the Stillinger-Weber potential may be found:

- ref. [5]: the potential form with Larandji, Landau and Dunweg parameters was successfully used for simulating Ge-growth by cluster deposition. The spreading of the Ge-atoms was modelled and the pair correlation function was calculated.

- ref. [6]: it was found that the potential works well for high temperatures, elastic constants and cohesive energy of crystalline silicon (c-Si), as well as phonon frequencies and the description of

bulk melting and liquid Si. The structures of atomic clusters smaller than 15 atoms are not well predicted. The structures of larger clusters, tested up to 480 atoms, agrees well with experiment.

- ref. [7]: Applied to Si-Si, Si-Ge and Ge-Ge interactions, using the modified form developed by Laradji, Landau and Dunweg. The growth of $Si_{1-x}Ge_x$ layers on Si(100) was simulated.

- ref. [8]: The growth of silicon clusters was simulated. Energetic order in building up the clusters and energetic and structural properties were calculated. Some improvements for the potential were suggested, and reasonable overall agreement was found.

- ref. [9]: Stability and fragmentation of Si clusters was modelled by energy monitoring. The ground-state of clusters was also calculated, some structures of particular sizes were found to have pronounced stability, the correpsonding cluster sizes are referred to as 'the magic numbers'. The structures and binding energies agree well with other studies.
  An extra term was added to simulate charged clusters, and stability for the cluster cations was calculated.

4.2.2 Watanabe

Watanabe et al [10, 11] developed an extended Stillinger-Weber potential for Si-O systems. The total interaction energy is given by:

$$\Phi = \sum_i \sum_{j>i} \varepsilon f_2(i,j) + \sum_i \sum_{j>i} \sum_{k>j} \varepsilon f_3(i,j,k) \tag{4.9}$$

where ε is the energy scale. A bond-softening function g_{ij} is introduced to the Si-O pair-interaction $f_2 = g_{ij}\, f_2^{SW}$. Bond-softening is a function of the co-ordination number of oxygen, and derived to fit to the *ab initio* cohesive energy. Included in g_{ij} is a Tersoff [12] type cut-off function, but in a less discontinious form to improve conservation of energy. In the three body term the function h from the Stillinger-Weber potential is used:

$$f_3(i,j,k) = h(r_{ij}, r_{ik}, \theta_{jik}) + h(r_{ji}, r_{jk}, \theta_{ijk}) + h(r_{ki}, r_{kj}, \theta_{ikj}). \tag{4.10}$$

The structures, lattice parameters and bond properties of several silica polymorphs were calculated and compared with *ab initio* and experimental values. The growth of a SiO_2-film was studied [13] with emphasis on growth modes at monatomic steps in the underlying substrate. The growth mode of SiO_2 on Si(001) was found to be dependent on the underlying dimer configuration and crystalline forms were predicted to become amorphous near steps, as confirmed by experiment.

4.2.3 Tersoff 1,2,3 and related forms

The Tersoff 1 potential [12, 14] takes into account the differences in bond order depending on the environment, i.e. through calculation of the number and types of neighbours. The Tersoff potential is a sum

over bond energies between all atoms:

$$V_{ij} = f_c(r_{ij})[a_{ij} f_R(r_{ij}) + b_{ij} f_A(r_{ij})] \tag{4.11}$$

consisting of a repulsive term f_R which include effects from overlapping wave functions, an attractive term f_A which represents the bonding and a cut-off function f_c:

$$f_c(r) = \begin{cases} 1, & r < R - D \\ \frac{1}{2} - \frac{1}{2}\sin\left[\frac{\pi}{2}(r-R)/D\right], & R - D < r < R + D \\ 0, & r > R + D \end{cases} \tag{4.12}$$

The repulsive and attractive terms are chosen to be exponential functions for $f_R(r) = A\exp(-\lambda_1 r)$ and $f_A(r) = -B\exp(-\lambda_2 r)$. The parameters a_{ij} and b_{ij} and their selection are motivated by theoretical considerations. The potential as a whole, is then modified to take into account the dependence of the coordination number on the distances between neighbours and on bond-angles. The parameters are defined by fitting to a small database of energies of simple bulk structures. The resulting potential is tested for several properties by comparison to experiments or higher level computations. The comparison set includes such quantities as structural energies and ordering, bonding properties, elastic constants, phonon dispersion, energies of defects and vacancies barriers, thermal expansion coefficients, and surface reconstructions. Elastic and structural properties are generally speaking not well described. The shortcomings of the potential are ascribed to two sources: the cut-off function is an *ad hoc* estimate and its explicit form is suspected to cause some discrepancies the calculation of physical quantities sensistive the the longer range of interatomic distances, and the soft bonding function can also give rise to problems in some situations as it produces too low of a strain energy, which consequently results in poor predictions for mechanical properties.

Improvements on the Tersoff 1 potential were designed to yield a better description of elastic properties, resulting in a second potential Tersoff 2 [15]. Effects for soft-bond bending are included and the cut-off function is refined, but at the cost that the modified potential generally gives less accurate results for surface geometries.

For Tersoff 2, parameters are optimised using a database of cohesive energies for bulk structures, bulk modulus, bond lengths and the elastic constants. Structures are calculated and lattice constants and structural parameters agree well with experiment. The energy predicted from the model potential is significantly higher than those calculated using density functional theory (DFT) in the local density approximation (LDA). Elastic properties and phonon frequencies are calculated well, point defect energies are satisfactory. Liquid and amorphous silicon is simulated on the resulting structures densities, coordination numbers and radial distribution function calculations are reported.

Tersoff 3 [16] extends the potential model to Si-Ge and Si-C, where one extra parameter is added as a factor weighted to account for new combination of atoms. Parameters are fitted to the heat of formation and result in a potential more accurate than averaging over known values for Si, Ge and C separately.

Srivastava et al [17] did an extensive study using this model for surfaces including: growth processes, reconstructions and dimers.

Ishimaru [18, 19] has adjusted the Tersoff 3 potential for applications on amorphous silicon. Simulating the annealing and quenching of Si-atoms results in a-Si model. Structural and dynamical properties are calculated: the radial distribution function (rdf), structure factors, defects from coordination numbers, and phonon density of states are calculated and compared to experiments, yielding good agreement. Another application of this potential [20] to $Si_{1-x}Ge_x$ gives good results for alloy structures, but results in too high melting temperatures. A distortion in the bond angle distribution is observed, and increases with increasing germanium composition.

Wang and Rockett [21] modified the Tersoff potential for Si surfaces with adatoms. The Tersoff potential is extended to longer ranges and an angle-dependent screening term is added to the attractive and repulsive terms.The pair-potential is as usual written as:

$$V_{ij} = f_c(r_{ij}) \times [A_{ij} \exp(-\lambda_1 r_{ij}) - B_{ij} \exp(-\lambda_2 r_{ij})], \tag{4.13}$$

the term A_{ij} is modified to give a cut-off function that has more continuous derivatives at the points $r = R_a \pm D_a$, using the modified values from Tersoff's R and D.

$$A_{ij} = A \times \begin{cases} 1, & r < R_a - D_a \\ B_{ij} + (1 - B_{ij})\{\frac{1}{2} - \frac{1}{2}\sin[\frac{\pi}{2(r-R_a)/D_a}]\} & R_a - D_a < r < R_a + D_a \\ B_{ij,} & r > R_a + D_a \end{cases} \tag{4.14}$$

This modified form of the potential is used to describe the energies for small clusters, where an improvement is seen due to the modification in the cut-off function. The topography of surfaces and diffusion energy barriers are given in better agreement with experimental results than Tersoff 3. The resulting potential describes the Si bulk material as well as the original model, but in the new version improvements for treating surfaces are included.

Umeno et al [22] optimised the Tersoff 1 potential, resulting in two parameter sets for potentials referred to as Umeno A&B. The original set of parameters is adjusted minimizing the differences in "root mean square" (rms) values for the forces on the atoms calculated as the difference between the potential model and *ab initio* calculations. To model Si-O interactions, some of the constants from the Tersoff potential are elevated to variables that become interaction specific.

Umeno A is yields an improved description for forces and lattice constants of Si and β-cristobalite SiO_2 crystals, however does not perform well at Si/SiO_2 interfaces development. Umeno B was then developed to remedy the problems at the interface between Si and SiO_2, giving good results at the interface as well for the Si and β-cristobalite SiO_2 crystalline forms, with errors less than 5% and significantly smaller than predicted with other forms of the Tersoff potential. Furthermore, as a transferability test, the lattice constants and force values for β-quartz SiO_2 were computed and found to be in good agreement with experimental values.

4.2.4 EDIP

A relatively new potential is the environment-dependent interatomic potential (EDIP) [23, 24, 25]. The development of the potential is based on Stillinger-Weber and Tersoff potentials and methods that have been used in attempts to improve them, using insights gained from more quantum mechanically based models such as the Pettifor and Biswas-Hamann potentials. Calculations on the elastic constants predicted by the potential were used for refinements on the functional form. The potential takes into account the environment by using an effective coordination number Z_i, which is a sum of the contributions of the neighbours over all atoms. The contributions of neighbours is approximated by a cut-off function:

$$f(r) = \begin{cases} 1, & r < c \\ \exp\left(\frac{\alpha}{1-\left(\frac{r-c}{b-c}\right)^{-3}}\right), & c < r < b \\ 0, & r > b \end{cases} \tag{4.15}$$

Following Tersoff, the potential is defined to consist of two- and three body terms focusing on pair bonding and angular forces. The total energy of formation is a sum over the energy for each atom $E_i = \sum_j V_2(r_{ij}, Z_i) + \sum_{jk} V_3(\overrightarrow{r}_{ij}, \overrightarrow{r}_{ik}, Z_i)$. The strength of the bond is defined by V_2, bond angles and angular forces are arising from V_3. The form of the Stillinger-Weber form is used for V_2, with the bond order $p(Z)$ derived from ab initio cohesive energies

$$V_2(r, Z) = A\left[\left(\frac{B}{r}\right)^{\rho} - p(Z)\right]\exp\left(\frac{\sigma}{r-a}\right). \tag{4.16}$$

The three-body term is, motivated by the elastic property calculations, chosen to be a function of a radial function $g(r)$ for each bond and an angular function $h(\theta, Z)$:

$$V_3(\overrightarrow{r}_{ij}, \overrightarrow{r}_{ik}, Z_i) = g(r_{ij})g(r_{ik})h(l_{ijk}, Z_i) \tag{4.17}$$

with $g(r) = \exp\left(\frac{\gamma}{r-b}\right)$ and $h(l, Z) = H\left(\frac{l+\tau(Z)}{w(Z)}\right)$. The functions $\tau(Z)$ and $w(Z)$ and their parameters are defined by theoretical considerations for bond-angles and energies. Thirteen parameters are then left for fitting with the least-square method to a database of *ab initio* data for cohesive energies, lattice constants, energies of formation for point defects and stacking-faults, and experimental elastic constants. The potential is then tested and compared with Stillinger-Weber, Tersoff 2, Tersoff 3, tight-binding and density functional theory (DFT) calculations. Calculations are performed to determine for cohesive energies, elastic properties and dislocations and reconstructions, liquid and amorphorous silicon, pair correlation functions and radial distribution functions. The bulk equilibrium properties, defects and dislocations are improved, other properties are described as well by the Stillinger-Weber potential.

4.2.5 Vashishta et al

Vashishta, Kalia, Rino and Ebbsjö [26] developed a two- and three body term effective potential for SiO_2. The potential has the form:

$$V = \sum V_2(r_{ij}) + \sum V_3(r_{ij}, r_{jk}, r_{ik}) \tag{4.18}$$

The two-body part V_2 is a sum of three interaction terms: steric repulsion, Coulomb interaction and charge-dipole interaction:

$$V_2 = \frac{H_{ij}}{r^{\eta_{ij}}} + \frac{Z_i Z_j}{r} - \frac{\frac{1}{2}(\alpha_i Z_j^2 + \alpha_j Z_i^2)}{r^4} \exp\left(\frac{r}{r_{4s}}\right) \tag{4.19}$$

The three-body potential V_3 consists of terms for the effects of bond stretching f, bond bending p and the strength of the interaction B, which are established using a cut-off function:

$$V_3 = B_{ijk} f(r_{ij}, r_{ik}) p(\theta_{jik}, \overline{\theta}_{jik}) \tag{4.20}$$

The parameters in the potential, 18 in total, are partly known from literature or earlier experiments, the unknown parameters where derived from experiments.

The resulting potential focuses on structural properties for a-SiO_2, some crystals and molten SiO_2 using molecular dynamics. Densities, self-diffusion constants and pair distributions were also calculated; results were compared with diffraction experiments and found to be in good agreement.

4.2.6 TTAM/Tsuneyuki et al

TTAM [27] is a potential based on Hartree-Fock calculations, developed for SiO_2 and SiO_2polymorphs. *Ab initio* calculations are used to construct the potential. The best fit is obtained for the form of the potential contains terms for modified Coulomb interaction, Born-Mayer repulsion and dispersive interaction:

$$U_{ij}(r) = U_{ij}^{Coulomb}(r) + f_0(b_i + b_j)\exp[(a_i + a_j - r)/(b_i + b_j)] - c_i c_j / r^6. \tag{4.21}$$

This is of the Buckingham-form which may be written as:

$$V_{ab}^{Buck} = A_{ab} \exp\left(\frac{-r}{B_{ab}}\right) - \frac{C_{ab}}{r^6} \tag{4.22}$$

The Coulomb interaction consists of short- and long-range terms, the modified long-range part is recalculated with the Ewald sum method and taking into account specific configurations of the atoms. The parameters are fit with a structure optimisation and to the compressibility of α-quartz. Configurations of α-quartz, α-cristobalite, coesite and stishovite were calculated with molecular dynamics, the stabilities found are used as a test for the validity of the potential. Structural parameters, binding energy and elastic properties are calculated, crystal stability and bulk moduli are found to be in reasonable agreement with experiment. For phase transitions and thermal expansions, the method is combined with total-energy

methods to obtain accurate results.

4.2.7 Flikkema-Bromley (FB)

Using a genetic algorithm in conjunction with the GULP program [28], a Buckingham-form potential for SiO_2 nanoclusters was developed [29]. Parameters are optimised by a fitting procedure to known nanostructures as calculated using DFT. The energy-order for several structures and the DFT-calculated energies are used to optimise the parameters. The potential was tested by calculating several structures of silicon-clusters, and compared with two other potentials BKS [30], TTAM [27] and a semi-empirical PM3 method. According to these simulations, this potential yields better results for small clusters than potentials derived for bulk applications. The energy r.m.s. error is smaller for the FB potential than with BKS, TTAM and PM3, the structure is better also than TTAM, but no significant differences for PM3 and BKS are seen.

4.2.8 MEAM

Using force-matching and a combination of Stillinger-Weber and EAM potential forms, the MEAM model is first fit to determine silicon properties. The form of the potential is identical to Baskes' MEAM [35], but with modified parameterizations. A large database is used to fit the potential parameters. The data consisted of *ab initio* forces and energies for clusters, liquid and amorphous silicon, experimental elastic constants and phonon frequencies and DFT calculations in the local density approximation (LDA) for vacancy and defects energies. Two methods were used to minimize the error in fitting the energy: Powell's least square method and a simulated annealing scheme. Results are compared with EDIP [23], Tersoff-3(T3) [16], the tight binding method and Stillinger-Weber(SW) [2]. In general energy and lattice results were better than or comparable to EDIP, both EDIP and Lenosky [31] are found to be better than the SW and T3 results. For thermodynamic properties, the Lenosky potential yields comparable results to T3. For dislocations and defects, EDIP and SW are less accurate. A disadvantage of the potential is poor predictions for stabilities, brittle fractures and density of some crystals, and the liquid phase is too dense.

4.2.9 Eichler

This model [32] combines *ab initio* and force field methods and is optimized for the study of SiO_2 polymorphs. The energy of a system or unit cell is divided in an inner and an outer part. The inner part is calculated using quantum-mechanical methods and the outer part is be described by an analytic potential:

$$E(\text{System}) = E_{\text{QM}}(I) + E_{\text{MM}}(O) + E_{\text{MM}}(I - O) \tag{4.23}$$

E_{MM} is calculated using an analytic potential. $E_{\text{MM}}(I - O)$ is the interaction between regions. This derived from the energy and forces in two regions resulting in equations containing both quantum-mechanical and force-field terms. Molecular dynamics is applied to calculate properties of different

clusters representing different SiO_2 polymorphs. In general, the inclusion of Hartree-Fock methods within the model is found to give better results than the use of the analytical potentials alone.

4.2.10 Biswas-Hamann A&B

The Biswas-Hamann potentials [33] include two and three body potentials for Si. The two-body term is a modified Morse-type potential with cutoff function f_c:

$$V_2(r) = (A_1 e^{-\lambda_1 r^2} + A_2 e^{-\lambda_2 r^2}) f_c(r) \tag{4.24}$$

The three-body term is dependent on the two lengths r_1, r_2 and the bond angle θ, which is expressed using Legendre polynomials:

$$V_3(r_{12}, r_{13}, \theta_1) = \sum_l C_l \phi_l(r_{12}) \phi_l(r_{13}, r) P_l(cos\theta_1) \tag{4.25}$$

Rewriting using spherical harmonics, this expression contains summations over the structure moments describing the local environment bond-order. The resulting generalized three body term for a special case of the Legendre polynomial is:

$$V_3(r_{12}, r_{13}, \theta) = [B_1 \psi_1(r_{12}) \psi_1(r_{13}) \left(cos\theta + \frac{1}{3} \right)^2 + B_2 \psi_2(r_{12}) \psi_2(r_{13}) \left(cos\theta + \frac{1}{3} \right)^3] \times f_c(r_{12}) f_c(r_{13}) \tag{4.26}$$

with the function $\psi_i = e^{-a_i r^2}$.

The potential is fitted to DFT-LDA calculations for energy and structure. Surface and bulk energy calculations were performed, phonon dispersions and frequencies and clusters were compared to DFT calculations. This potential describes high-pressure properties and surfaces well, but does not predict well interstitial and layered structures. A second potential was developed to improve these features and is suitable for tetrahedral structures, but not does not perform as well for non-tetrahedral structures.

4.2.11 BKS

BKS [30] is a potential developed for modelling SiO_2 polymorphs and aluminophosphates. Starting from the TTAM potential[27], and using a Buckingham form to describe the interaction energy Φ_{ij}, a Coulomb term and a covalent short-range term are contributing to the potential:

$$\Phi_{ij} = q_i q_j / r_{ij} + A_{ij} \exp(-b_{ij} r_{ij}) - c_{ij} / r_{ij}^6 \tag{4.27}$$

The parameters are obtained by an alternating fitting to cluster and bulk values for elastic properties, cell constants to account for short-range interactions as well as the long-range Coulomb interaction. Lattice

parameters and elastic properties are compared to TTAM, *ab initio* and empirical force field results for SiO_2 polymorphs, the results of this potential are represent a significant improvement over the TTAM model.

4.2.12 Baskes MEAM

The embedded-atom method(EAM)[34] is modified by Baskes [35] for Si, Ge and SiGe alloys. Effects from further than nearest neighbours are included, as well as bond-bending forces. The total energy is defined as in the EAM:

$$E_{tot} = \sum_i F_i(\rho_{h,i}) + \frac{1}{2}\sum_{i,j} \phi_{ij}(R_{ij}) \tag{4.28}$$

with F_i the energy needed to embed an atom, ϕ_{ij} is the pair-potential and $\rho_{h,i}$ is the host-electron density which is modified in the MEAM.

The studied properties are bulk structural energies and geometry, point defects, meta-stability of non-crystalline solids surfaces and small clusters. All are compared to DFT calculated values. Defects energies and vacancy diffusion are calculated and compared to DFT results with varying levels of agreement. Meta-stable structural energies and geometries are found to be in good agreement, within about 0.2 eV for Si and 0.4 eV for Ge. Elastic constants found by MEAM are qualitatively acceptable. The stacking fault energy is found to be too large, which is thought to be the same source of error generally for surface calculations. The potential is tested on clusters, it is advised not to use the potential in small clusters, but the results for large clusters are reasonable. And finally, the electron-density profiles are calculated and it is concluded that the modified electron-density $\rho_{h,i}$ could be further refined.

4.3 Literature comparisons and analysis

Several comparative studies of the model potentials discussed have been given in the literature, and a summary of the comparisons is given in Table 4.1. The commonly applied Stillinger-Weber potential is included in many of these studies. The Tersoff 3 model is also tested in a reasonable number of the comparsions. Other model potentials have been tested less extensively and it is difficult to make general statements regarding their applicability across a wide range of problems.

Generally speaking, the Stillinger-Weber model can be applied to a wide range of property calculations. It performs well in many cases but is not recommended for surface or cluster calculations. Othere potentials can yield better results for specific properties: Tersoff 3 results in the best predictions for point defects whereas EDIP performs better for surface defects. Ishimaru and EDIP forms are suitable for amorphous silicon. For modelling clusters, Tersoff 2 provides a reasonable description, but other potentials in the list have been developed specifically to study clusters, for example Flikkema-Bromley model, and may prove to be better suited for clusters.

The summary of the comparisons made in the literature is provided as an overview, and to help guide initial selection of potential models. Of course, each model must be tested and validated for a particular

study. The conclusion is: each model potential has its own sets of strengths and weaknesses. This summary and overview is intended to help narrow selection of a model potential for specific applications, but validation studies are always recommended to insure that a given model potential will provide meaningful results when used within a simulation.

4.4 Applications

Several crystalline forms or polymorphs of SiO_2 are known and their structures and symmetries are tabulated; for example in Wyckoff's Crystal Structures [36]. A typical feature of the crystalline forms is the SiO_4 tetrahedral units from which the different structures are built. Polymorphs differ from one another in terms of geometry, density, and stability (as a function of temperature and pressure). Transitions between polymorphs can occur when a change of temperature or pressure leads to a new region of a SiO_2 phase diagram. The most common forms of silica are quartz, tridymite and cristobalite, listed in order of increasing symmetry or decreasing density. Other forms examined for our test calculations are coesite, keatite and stishovite, which are produced only in rare circumstances of high temperature and pressure. Some forms of silica exist in two phases, labelled either α or β. The low (normal) temperature phase α transforms into the β-structure when the temperature reaches a certain point and is reversible upon cooling. The atoms within the structrures are displaced, without creation of dangling bonds, into the new equilibrium positions.

For SiO_2 polymorphs, the Buckingham form is found to produce reasonable results. Other potentials that are of the Buckingham form are BKS (4.2.11), Tsuneyuki (TTAM, 4.2.6) and Flikkema-Bromley (FB, 4.2.7). The Buckingham potential can be implemented with parameters to obtain the potentials BKS, TTAM and FB. TTAM is the original developed potential from Buckingham for SiO_2 polymorphs. BKS is TTAM optimised for SiO_2 polymorphs, FB is adjusted for SiO_2 clusters and may therefore produce good results for high strain structures.

Si and Ge crystals are of the cubic diamond structure with respective lattice constants of a = 5.43 Å and a = 5.65 Å. Random alloys of Si and Ge consists of a random distribution of the two atom types over diamond lattice sites, accompanied by a structural relaxation off site to accommodate the random environment.

The lattice constant for Si_xGe_{1-x} is to a good approximation given according to Vegard's law, which is a linear interpolation between the Si and Ge lattice constants as a function of the fractional alloy composition x.

The potentials investigated for silicon are the the Buckingham (see 4.2.6), Tersoff (4.2.3) and Stillinger-Weber (4.2.1) forms. This gives three categories of simulations that can be tested: the series of SiO_2 polymorphs, Si/Ge crystals and surfaces or interfaces of Si and SiO_2, or $Si_{1-x}Ge_x$.

The Tersoff potential is available for Si and Ge crystals and surfaces. The Umeno potential is of the Tersoff-variety and has been optimised for SiO_2 polymorphs. Umeno B can be used for Si and Si polymorphs, for both bulk and surface situations. The Stillinger-Weber potential was developed for the study of Si-crystals. Varieties described in the SW section, 4.2.1, Ding-Andersen (DA) and Laradji,Landau and

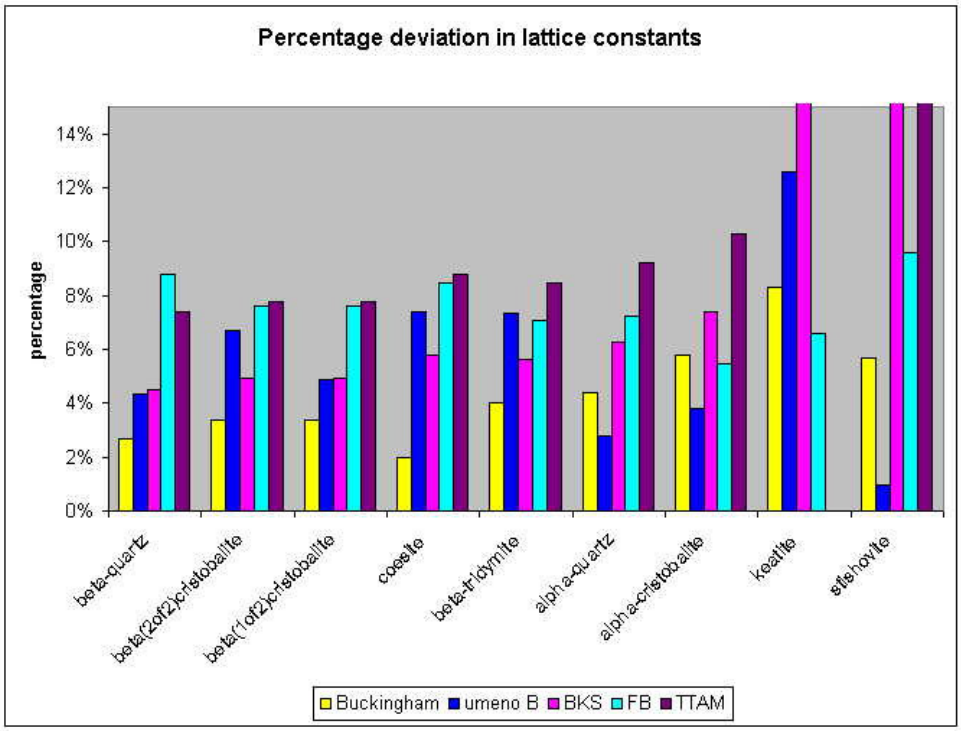


Figure 4.1: Results for lattice constants of different silica polymorph structures using model potentials

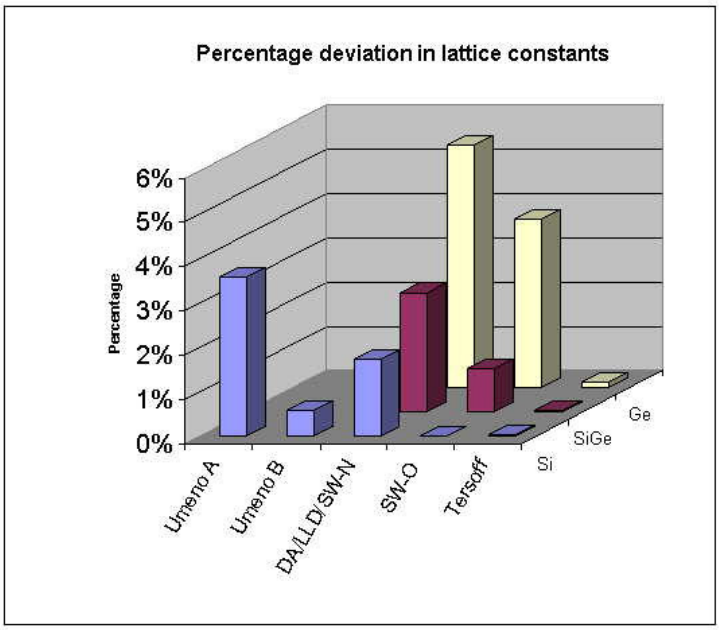


Figure 4.2: Comparsion of lattice constants for Si, Ge and $Si_{.75}Ge_{.25}$ alloy using model potentials.

Dunweg (LLD), have been optimised to treat Ge-crystals and SiGe-alloys, respectively.

LLD uses the same Ge-Ge parameters as DA, so gives the same results for Ge. SW is given in two varieties, where one is using only the two-body interaction potentials and the other using both two- and three-body interaction potentials. The Stillinger-Weber potentials were applied as well to SiGe and Ge, although not developed for them. The set of parameters were just converted to Si-Ge and Ge-Ge interactions. This was done to be able to compare SW to LLD and DA.

Stillinger-Weber is developed with parameters that agree with the Baskes MEAM (4.2.12), in an adjusted form [38]. The new form is referred to as SW-N, the original as SW-O, with an additional number 2 or 3 to indicate the two- and three body interaction potentials.

The GULP program used to optimize structures and compare geometries from DFT calculations and experiment. The results from the simulations compared are the cell dimensions and the fractional coordinates for the atomic positions. These are compared directly with experimental values. The lattice parameters are take as the most significant measure of accuracy of the potential for our comparisions on periodic structures. The series of SiO_2 polymorphs is tested with the BKS, TTAM, FB, Buckingham and Umeno B potentials. For semiconductor crystals, Si and Ge crystals and the $Si_{0.75}Ge_{0.25}$ alloy are studied with analytical potential models and compared to DFT calculations. Tests are performed using the Tersoff, Stillinger-Weber (two- and three-body, in original and modified form), the SW-varieties DA and LLD and Umeno A and B for silicon.

The Tersoff and Umeno B potentials were tested on the interfaces of Si/SiO_2, Si/SiGe and Si/Ge, using molecular dynamics and energy minimization. The interfaces are simulated through an annealing cycle, starting from 1100 K decreasing in steps of 200 K. These temperatures are not higher than typical maximum production temperatures in conventional semiconductor manufacturing. The molecular dynamics allows the structure to react to the forces within the structure, allowing structural relaxation. Lowering the temperature will reduce the atomic motion, until eventually freezing into a local potential energy minimum.

SiO_2 polymorphs For all polymorph structures, the percentage difference is calculated between experimental input values and the optimised result obtained using GULP, see Table 4.6. The average values of cell constants and angles over each of the structures except keatite and stishovite are shown in Table 4.3. In fig. 1, the averaged calculated cell constants are shown for each of the structures. The Buckingham (2-8%) and FB (6-10%) potentials work well for the high temperature and pressure SiO_2 polymorphs, coesite, keatite and stishovite. Umeno B gives a good result for stishovite. The BKS and TTAM potentials do not work well for keatite and stishovite. TTAM results in an error for keatite. Results from BKS and TTAM in some cases result in distorted cell structures.

Si and Ge Umeno A and B have been previsouly tested in literature [22] with the results for lattice constants reported. From our study. the results for Si, Ge and the $Si_{0.75}Ge_{0.25}$-alloy are summarized in fig. 4.2, reporting the percentage deviation in lattice constants between experimental and calculated values. In the figure comparable values are grouped, the exact values of lattice constant and angle deviations for

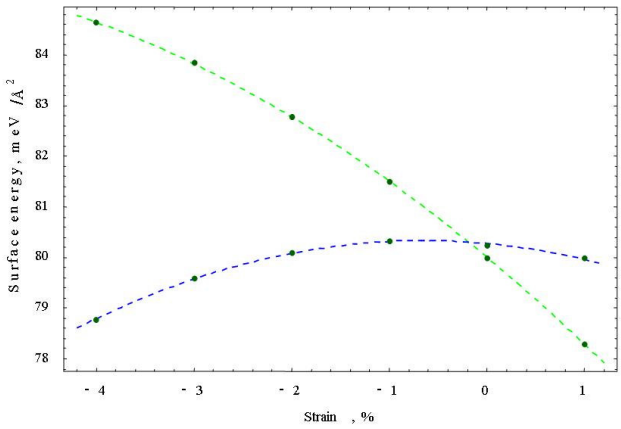


Figure 4.3: Ge surface energy as a function of biaxial strain. Green: (2x1) reconstruction of Ge (001) surface, Blue: Rebonded step (RS) reconstruction of Ge (105)

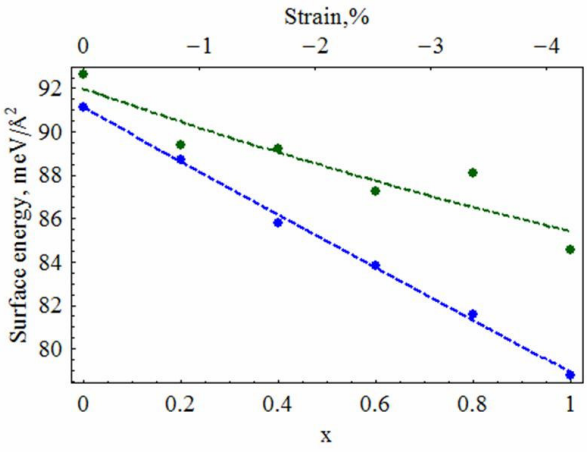


Figure 4.4: $Si_{1-x}Ge_x$ surface energy as a function of composition/biaxial strain. Green: (2x1) reconstruction of $Si_{1-x}Ge_x$ (001) surface, Blue: Rebonded step (RS) reconstruction of $Si_{1-x}Ge_x$ (105)

all potentials can be found in table 4.6. The LLD uses the same Ge-Ge interaction as from DA; for the Ge case, the results from these two models are the same. The two potentials are taken together in the results. The SW-N potential with the new parameters [38], is giving the same results as the same as LLD and DA. The original SW is giving large improvements. The SW potentials are available in two forms, with two or three body interactions. The results differ only significantly for the angles of the SiGe-alloy, which are much better calculated with the SW-3 three body potentials.

Cell parameters are calculated reasonably for all potentials considered for Si and Ge. It seems to be that the larger the percentage of Ge, the larger the error in the result, this would suggest that the potentials could use some improvement on the Si-Ge and/or Ge-Ge-parameters.

As a next evaluation of the potential models, surface energies of Si and Ge were calculated using a convergent approach, in which slabs of different thickness are used to extract surface energies from slab calculations. This approach can provide better surface energy estimates than the commonly used subtraction of bulk energies from total slab energies. Simulated annealing was performed for various silicon, germanium and alloys using an initial temperature of 1100 K, followed by steps of 900 K, 600 K, 300 K and 0 K (energy optimization step). The total time of the simulated of annealing was 28 ps with a timestep of 1 fs. After this anneal, approximately two thirds of the surface atoms for the Ge surface have formed dimers, and the dimers are not in a parallel arrangement as seen for the (2x1) reconstruction. The surface energy is calculated to be 89.7 meV/Å^2. It also should be noted that this result is not exactly reproducible if different simulation conditions (such as time of annealing or the value of timestep) are used. The (2x1) reconstruction was introduced as the initial state of the Ge (001) surface. The value of surface energy was calculated to be 82.8 meV/Å^2. In both cases, Tersoff's T3 potential is used. After the surface reconstructions are introduced, the lowest surface energy can be found by using the MD simulated annealing followed by a structure optimization. In the fig. 3, Ge surface energy versus strain for the (105) rebonded step (RS) reconstruction (energies fit to a quadratic polynomial shown in blue) and (001) (2x1) reconstruction (fit to a quadratic polynomial, shown in green) is presented. The simulations confirm that (105) RS reconstruction is energetically preferable over (001) (2x1) for strained Ge surfaces. In fig. 4 a similar study is shown whereby surface energy is investigated versus alloy composition; here it is found that the stability ordering for the two reconstruction is not altered as a function of alloy composition. However, the relative stability of the two reconstructions vary significantly for silicon-rich or germanium -rich alloys.

4.5 Conclusion

Our analysis of the selected potentials is not intended to be definitive, but is provided to illustrate the considerations required when selecting a model potential for a given application. Hence, our literature survey is intended to be representative, we have focused on potential functions used for common semiconducting materials and silicon oxides. Calculations, primarily for structure properties, have been performed to illustrate the typical steps required to benchmark and compare a given potential form.

One of the main advantages of applying analytical potential functions is the spatial and time scales

for material systems that can be treated is much larger than for other atomic scale methods. We have not presented large scale structure or molecular dynamics simulations, but it is useful to bear in mind that calculations orders of magnitude larger than those presented here are possible with atomistic models, thus increasing their relevance for modern technology design.

Acknowledgements

This work has been funded by Science Foundation Ireland. AJdV was as summer intern visiting from the University of Twente. We are grateful to Professor Julian Gale for use of the GULP program to perform the potential evaluations.

Table 4.1: Literature comparisons for selected atomic potential models

Potential:	BH		SW								T2		T3					EDIP			Ishimaru	
Source:	[1]	[37]	[1]	[39]	[40]	[41]	[42]	[43]	[44]	[37]	[45]	[1]	[37]	[40]	[1]	[39]	[40]	[42]	[43]	[45]	[39]	[43]	[45]	[18]	[20]
lattice								+											+						
crystal stability															+										
phase transform.				++								+			++							+			
elastic/vib. prop.	+/-		+							+					+/-							+			
point defects			++					+				+			+++				+			++			
liquid	-		++					+				- -			-				+						
surfaces				+	+/-	-		+								+					+				
surface reconstruction	+			++		++	++					-	-		-		- -	+							
surfaces at high temp.					+								-			+					-				
surface defects						+																			
a-Si (structure & dyn.)								+										+				++		++	
a-Si-Ge-alloys																									+
melting temp																								-	-
strain relief							+											+/-							
bulk strain energy										++									+			+			
surface displacements										++									-			++			
stacking fault energy										++									-			-			
thermomech.																			+						
clusters												++													
small clusters (< 15)									-																
mid-size clusters (16-30)									+/-																
larger clusters									+																
dispersion	-									+/-			-												

BH	Biswas-Hamann [33]
SW	Stillinger-Weber [2]
T2	Tersoff 2 [15]
T3	Tersoff 3 [16]
EDIP	[23]
Ishimaru	[18],[19]

++	potential stated to be best suited for specific property
++	claimed better than most potentials for specific property '+'
+	better results than average
+/-	average results compared to other potentials
-	not performing as well as average
- -	performs poorly for specific property

Table 4.2: Overview of applications for atomistic potential models

Pot.	*Name*	*Ref.*	*Atom-types*	*Calculated properties*	*Applications*
1	Stillinger-Weber	[2]	Si-Si (crystal, liquid, amorph)	structure, energies, lattice (thermo)dynamics, defects, film growth/clusters	general use
1.1	Ding-Andersen	[3]	Ge-Ge (c, l, a)	same applications as S-W, only with modified parameters for Ge	Ge
1.2	Laradji, Landau, Dunweg	[4]	Si-Ge	structural properties, phase-transition, lattice parameters, clusters	
2	Tersoff	[12],[15],[16]	Si crystal and liquid, Si-Ge, Si-C	structure, relaxation and formation energies, elastic properties, defects	Si-Ge, defects
2.1	Ishimaru	[18],[19]	a-Si	rdf, structure, defects	
2.2	Wang and Rockett	[21]	Si surfaces and clusters	surface topology and energy barriers, cluster energies	surfaces
2.3a	Umeno Pot.A	[22]	Si-Si, Si-O crystals	lattice constants, interatomic forces	
2.3b	Umeno B	[22]	Si-Si, Si-O, Si/SiO2 surface	pot. A plus forces at interface, Raman frequency	Si/SiO2 interfaces
3	EDIP (Bazant-Kaxiras)	[43],[23],[24]	Si	structure, rdf, pair corr. functions, defects, elastic properties	defects, a-Si
4	Watanabe	[10],[11]	Si-O	structure and lattice, rdf for a-Si, structure of interface, bond angles, growth	
5	Vashishta	[26]	SiO2, crystal and amorph	structure, energies, densities, self diffusion, bond angles, melting proves	
6	TTAM/Tsuneyuki	[27]	Si-O, polymorphs	structure parameters and stability for polymorphs	
7	Flikkema Bromley	[29]	SiO2 clusters	energy and structure for clusters	SiO2 clusters
8	Lenosky (MEAM)	[31]	Si	forces and energies, thermodyn, lattice parameters, defects, clusters	lattice and defects for Si
9	Eichler	[32]	Si-polymorphs	polymorph structures and clusters	
10a	Biswas Hamann Pot.A	[33]	Si, bulk	bulk and surface energies, high pressure properties	
10b	Biswas Hamann Pot.B	[33]	Si, tetrahedral	tetrahedral: energies and clusters, a-Si and crystal growth	
11	BKS	[30]	Si-polymorphs	lattice parameters and elastic properties for polymorphs	
12	Baskes MEAM	[35]	Si, Ge, Si-Ge	structures, surfaces, defects, clusters	

Table 4.3: Averaged percentage differences for calculated cell parameters relative to experimental values for SiO_2 polymorphs

	cell constant	angle
Buckingham	3.935%	0.005%
Umeno B	4.978%	0.000%
BKS	5.617%	0.000%
FB	7.299%	0.009%
TTAM	8.466%	0.000%

Table 4.4: Calculations and literature comparison for Umeno [22] and Tersoff [16], lattice constant in Å

		Si	β-cristobalite	β-quartz, a-axis	β-quartz, c-axis
Experiment		5.43	7.16	5.01	5.47
Umeno B	calc	5.38	7.56	5.16	5.67
	lit	5.40	7.15	5.21	5.73
Tersoff	calc	5.43	8.15	5.47	5.98
	lit	5.43	-	-	-
Umeno A	calc	5.19	-	-	-
	lit	5.24	-	-	-

Table 4.5: Energies for MD and energy minimisation of surfaces and interfaces (in eV)

non-fixed	$Si_{24}Ge_8/Si_{32}$	$Si_{32}Ge_{64}$	$Si_{64}Ge_{32}$	Si/SiO_2 bridge	Si/SiO_2 no bridge
1100 K	-289.88	-361.51	-386.06	-287.37	-273.28
900 K	-278.97	-349.77	-378.69	-469.83	-414.05
700 K	-284.53	-358.99	-381.79	-478.86	-420.95
500 K	-284.92	-363.67	-385.05	-481.58	-422.73
300 K	-285.90	-366.55	-388.97	-485.16	-425.65
En.min.	-290.00	-372.43	-395.43	-491.46	-429.84
fixed					
1100 K		-361.51	-386.06	-287.37	-273.28
900 K		-353.81	-376.08	-441.23	-385.60
700 K		-356.11	-380.93	-445.28	-394.82
500 K		-358.06	-382.81	-448.99	-396.38
300 K		-361.19	-383.52	-451.34	-398.78
En. min.		-365.72	-390.12	-467.87	-421.49

Table 4.6: Percentage difference relative to experiments for constants and angles

Lattice constants SiO_2polymorphs									
	α-quartz	α-cristobalite	β-quartz	β(1of2)cristobalite	β(2of2)cristobalite	β-tridymite	coesite	keatite	stishovite
FB	5.47%	7.22%	7.62%	7.63%	8.76%	7.079%	8.46%	6.56%	9.61%
TTAM	10.26%	9.19%	7.76%	7.76%	7.39%	8.44%	8.76%	error	76.21%
Buckingham	5.77%	4.42%	3.36%	3.36%	2.69%	4.01%	1.99%	8.29%	5.67%
BKS	7.39%	6.27%	4.95%	4.95%	4.51%	5.62%	5.80%	63.02%	69.89%
UmenoB	3.82%	2.77%	6.68%	4.89%	4.33%	7.35%	7.37%	12.57%	0.94%

Lattice angles SiO_2polymorphs									
	α-quartz	α-cristobalite	β-quartz	β(1of2)cristobalite	β(2of2)cristobalite	β-tridymite	coesite	keatite	stishovite
FB(angles)	0.04%	0.00%	0.00%	0.01%	0.00%	0.00%	0.10%	0.55%	0.00%
TTAM	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.20%	error	0.00%
Buckingham	0.00%	0.03%	0.00%	0.00%	0.00%	0.00%	0.08%	5.19%	0.00%
BKS	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.19%	40.56%	0.00%
UmenoB	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.12%	10.18%	0.00%

Lattice constants Si-Ge crystals					Lattice angles Si-Ge crystals			
	Si	Si-Ge	Ge			Si	Si-Ge	Ge
DA/LLD	1.73%	2.70%	5.48%		DA/LLD	0.00%	0.00%	0.00%
SW-new 2	1.73%	2.70%	5.48%		SW-new 2	0.00%	0.02%	0.00%
SW-new 3	1.73%	2.70%	5.48%		SW-new 3	0.00%	0.00%	0.00%
SW-original 2	0.01%	0.96%	3.82%		SW-original 2	0.00%	0.06%	0.00%
SW-original 3	0.01%	0.99%	3.82%		SW-original 3	0.00%	0.00%	0.00%
Tersoff	0.02%	0.03%	0.13%		Tersoff	0.00%	0.00%	0.00%
Umeno A	3.58%							
Umeno B	0.58%							

Bibliography

[1] H. Balamane, T. Halicioglu, W. A. Tiller, *Comparative study of silicon empirical interatomic potentials*, Phys. Rev. B **46**, 2250 (1992).

[2] F. H. Stillinger, T. A. Weber, *Computer simulation of local order in condensed phases of silicon*, Phys. Rev. B **15**, 5262 (1985).

[3] K. Ding, H. C. Andersen, *Molecular-dynamics simulation of amorphous germanium*, Phys. Rev. B **34**, 6987 (1986).

[4] M. Laradji, D. P. Landau, B. Dunweg, *Structural properties of $Si_{1-x}Ge_x$ alloys: a Monte Carlo simulation with the Stillinger-Weber potential*, Phys. Rev. B **51**, 4894 (1995).

[5] H. W. Lu, J. Y. Feng, *Molecular-dynamics simulation of Ge film growth by cluster deposition*, Modell. Simul. Mater. Sci. Eng. **8**(4), 621 (2000).

[6] M. R. Zachariah, M. J. Carrier, E. Blaisten-Barojas, *Properties of silicon nanoparticles: A molecular dynamics study*, J. Phys. Chem. **100**(36), 14856 (1996).

[7] J. L. Xu, J. Y. Feng, *Molecular-dynamics simulation of $Si_{1-x}Ge_x$ epitaxial growth on Si(100)*, Nucl. Instrum. Methods Phys. Res., Sect. B **217**, 33 (2004).

[8] E. Blaisten-Barojas, D. Levesque, *Molecular-dynamics simulation of silicon clusters*, Phys. Rev. B **34**, 3910 (1986).

[9] B. P. Feuston, R. K. Kalia, P. Vashishta, *Fragmentation of silicon microclusters: A molecular-dynamics study*, Phys. Rev. B **35**, 6222 (1987).

[10] T. Watanabe, I. Ohdomari, *Modeling of a SiO_2/Si(001) structure including step and terrace configurations*, Appl. Surf. Sci. **162-163**, 116 (2000).

[11] T. Watanabe, H. Fujiwara, N. Noguchi, T. Hoshino, I. Ohdomari, *Novel interatomic potential energy function for Si, O mixed systems*, Jpn. J. Appl. Phys. **38**, L366 (1999).

[12] J. Tersoff, *New empirical approach for the structure and energy of covalent systems*, Phys. Rev. B **37**, 6991-7000 (1988).

[13] T. Watanabe, I. Ohdomari, *Large-scale modeling of SiO2/Si(001) interface structures by using a novel inter-atomic interaction model*, in *The physics and chemistry of SiO_2 and the Si-SiO_2 Interface-4* **2000-2** (The Electrochemical Society).

[14] J. Tersoff, *New empirical model for the structural properties of silicon*, Phys. Rev. Lett. **56**, 632 (1986).

[15] J. Tersoff, *Empirical interatomic potential for silicon with improved elastic properties*, Phys. Rev. B **38**, 9902-9905 (1988).

[16] J. Tersoff, *Modeling solid-state chemistry: Interatomic potentials for multicomponent systems*, Phys. Rev. B **39**, 5566-5568 (1989).

[17] D. Srivastava, B. J. Garrison, D. W. Brenner, *Modeling the growth of semiconductor epitaxial films via nanosecond time scale molecular dynamics simulations*, Langmuir **7**(4), 683 (1991).

[18] M. Ishimaru, *Molecular-dynamics study on atomistic structures of amorphous silicon*, J. Phys.: Condens. Matter **13**(19), 4181 (2001).

[19] M. Ishimaru, *Atomistic simulations of structural relaxation processes in amorphous silicon*, J. Appl. Phys. **91**(2), 686 (2002).

[20] M. Ishimaru, M. Yamaguchi, Y. Hirotsu, *Molecular dynamics study of structural and dynamical properties of amorphous Si-Ge alloys, Phys. Rev. B* **68**, *235207 (2003).*

[21] J. Wang, A. Rockett, Simulating diffusion on Si(001) 2x1 surfaces using a modified interatomic potential, Phys. Rev. B **43**, 12571 (1991).

[22] Y. Umeno, T. Kitamura, K. Date, M. Hayashi, T. Iwasaki, *Optimization of interatomic potential for Si/SiO2 system based on force matching*, Comput. Mater. Sci. **25**, 447-456 (2002).

[23] M. Z. Bazant, E. Kaxiras, J. F. Justo, *Environment-dependent interatomic potential for bulk silicon*, Phys. Rev. B **56**, 8542 (1997).

[24] J. F. Justo, M. Z. Bazant, E. Kaxiras, V. V. Bulatov, S. Yip, *Interatomic potential for silicon defects and disordered phases*, Phys. Rev. B **58**, 2539 (1998).

[25] N. Bernstein, M. J. Aziz, E. Kaxiras, *Atomistic simulations of solid-phase epitaxial growth in silicon*, Phys. Rev. B **61**(10), 6696 (2000).

[26] P. Vashita, R. K. Kalia, J. P. Rino, I. Ebbsjo, *Interaction potential for SiO_2: A molecular-dynamics study of structural correlations*, Phys. Rev. B **41**, 12197 (1990).

[27] S. Tsuneyuki, M. Tsukada, H. Aoki, Y. Matsui, *First-Principles Interatomic Potential of Silica Applied to Molecular Dynamics*, Phys. Rev. Lett. **61**, 869 (1988).

[28] J. D. Gale, *GULP: A computer program for the symmetry-adapted simulation of solids*, J. Chem. Soc., Faraday Trans. **93**(4), 629 (1997).

[29] E. Flikkema, S. T. Bromley, *A new interatomic potential for nanoscale silica*, Chem. Phys. Lett. **378**(5-6), 622 (2003).

[30] B. W. H. van Beest, G. J. Kramer, R. A. van Santen, *Force fields for silicas and aluminophosphates based on ab initio calculations*, Phys. Rev. Lett. **64**, 1955 (1990).

[31] T. J. Lenosky, B. Sadigh, *Highly optimized empirical potential model of silicon*, Modell. Simul. Mater. Sci. Eng. **8**(6), 825 (2000).

[32] U. Eichler, C. M. Kolmel, J. Sauer, *Combining ab initio techniques with analytical potential functions for structure predictions of large systems: method and application to crystalline silica polymorphs*, J. Comput. Chem. **18**(4), 463 (1997).

[33] R. Biswas, D. R. Hamann, *New classical models for silicon structural energies*, Phys. Rev. B **36**, 6434 (1987).

[34] M. S. Daw, M. I. Baskes, *Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals*, Phys. Rev. B **29**, 6443 (1984).

[35] M. I. Baskes, J. S. Nelson, A. F. Wright, *Semiempirical modified embedded-atom potentials for silicon and germanium*, Phys. Rev. B **40**, 6085 (1989).

[36] Ralph W. G. Wyckoff, *Crystal Structures Vol.1* (Interscience, New York, 1960), p. 1-3, 253-4, 313-325.

[37] E. R. Cowley, *Lattice dynamics of Silicon with empirical many-body potentials*, Phys. Rev. Lett. **60**, 2379 (1988).

[38] B. J. Thijsse, *Relationship between the modified embedded-atom method and Stillinger-Weber potentials in calculating the structure of silicon*, Phys. Rev. B **65**, 195207 (2002).

[39] X. P. Xie, M. H. Liang, Z. M. Choo, S. Li, *A comparative simulation study of silicon (001) surface reconstruction using different interatomic potentials*, Surf. Rev. Lett. **8**(5), 471 (2001).

[40] L. Nurminen, F. Tavazza, D. P. Landau, A. Kuronen, K. Kaski, *Comparative study of Si(001) surface structure and interatomic potentials in finite-temperature simulations*, Phys. Rev. B **67**, 35405 (2003).

[41] J. H. Wilson, J. D. Todd, A. P. Sutton, *Modelling of silicon surfaces: a comparative study*, J. Phys.: Condens. Matter **2**(51), 10259 (1990).

[42] L. Nurminen, F. Tavazza, D. P. Landau, A. Kuronen, K. Kaski, *Reconstruction and intermixing in thin Ge layers on Si(001)*, Phys. Rev. B **68**, 085326 (2003).

[43] J. F. Justo, M. Z. Bazant, E. Kaxiras, V. V. Bulatov, S. Yip, *Interatomic potential for silicon defects and disordered phases*, Phys. Rev. B **58**(51), 2539 (1998).

[44] S. Yoo, X. C. Zeng, *Global geometry optimization of silicon clusters described by three empirical potentials*, J. Chem. Phys. **119**(3), 1442 (2003).

[45] J. Godet, L. Pizzagalli, S. Brochard, P. Beauchamp, *Comparison between classical potentials and ab initio methods for silicon under large shear*, J. Phys.: Condens. Matter **15**(41), 6943 (2003).

Chapter 5

The $\vec{k} \cdot \vec{p}$ method.

P. Marconcini, M. Macucci
Dipartimento di Ingegneria dell'Informazione,
Università di Pisa,
Via Caruso 16, I-56122 Pisa, Italy

5.1 Introduction

To understand the physical properties of semiconductors it is necessary to know their electronic band structure, in particular the behavior of energy as a function of the wave vector $\vec{k}$ in the reciprocal lattice of the crystal. Several numerical methods can be successfully used to find the band structures and the corresponding wave functions, such as the tight binding, the pseudopotential, the orthogonalized plane wave, the augmented plane wave, the Green's function and the cellular methods [1]. These methodologies can provide us with the desired results throughout the $\vec{k}$-space.
Actually many phenomena, for example in the study of electrical transport (due to both electrons and holes) and of optical properties (such as absorption or gain due to electronic transitions caused by an incident optical wave), involve only the top of the valence band and the bottom of the conduction band. Indeed, low energy electrons and holes are situated in these regions and also electronic transitions occur near the band edges of direct band-gap semiconductors.
The $\vec{k} \cdot \vec{p}$ method [2] is a perturbative technique which allows to obtain the band structures of materials in the regions of the reciprocal space near the band extrema, expanding the eigenvalues and eigenvectors of the single electron Hamiltonian as a function of $\vec{k}$ around the wave vector $\vec{k}_0$ corresponding to the band maximum or minimum. This method, introduced by J. Bardeen [3] and F. Seitz [4] and used by W. Shockley [5] and G. Dresselhaus, A. F. Kip and C. Kittel [6], received a general formulation with E. O. Kane [7] and J. M. Luttinger and W. Kohn [8, 9], and was later applied to strained materials (by G. E. Pikus and G. L. Bir [10]), to heterostructures (for example by G. Bastard [11], M. Altarelli [12]

and M. G. Burt [13]) and to carbon nanotubes (by M. Ajiki and T. Ando [14]), resulting a very useful and quite easy way to study the local properties of materials.

5.2 Perturbation theory and $\vec{k}\cdot\vec{p}$ method

In a pure crystal an electron is subject to a periodic potential energy

$$V(\vec{r}) = V(\vec{r}+\vec{R}) \ , \tag{5.1}$$

with $\vec{R}$ any linear combination of the lattice vectors, and thus also the Hamiltonian is invariant under translation by the lattice vectors. Therefore if $\psi^n_{\vec{k}}(\vec{r})$ is the wave function of an electron moving in the crystal also $\psi^n_{\vec{k}}(\vec{r}+\vec{R})$ will be a solution of the Schrödinger equation and therefore will coincide with $\psi^n_{\vec{k}}(\vec{r})$ apart from a constant with unit modulus (otherwise the wave function can grow to infinity if we repeat the translation $\vec{R}$ indefinitely). Thus the general form of the electron wave functions will be

$$\psi^n_{\vec{k}}(\vec{r}) = e^{i\,\vec{k}\cdot\vec{r}} u^n_{\vec{k}}(\vec{r}) \tag{5.2}$$

where $\psi^n_{\vec{k}}(\vec{r})$ is usually called "Bloch function", while $u^n_{\vec{k}}(\vec{r})$ is called "Bloch lattice function" and is periodic with the lattice periodicity:

$$u^n_{\vec{k}}(\vec{r}+\vec{R}) = u^n_{\vec{k}}(\vec{r}) \tag{5.3}$$

(Bloch's theorem).
Starting from the Schrödinger equation for $\psi^n_{\vec{k}}(\vec{r})$:

$$H^{(0)}\psi^n_{\vec{k}}(\vec{r}) = E^n_{\vec{k}}\psi^n_{\vec{k}}(\vec{r}) \tag{5.4}$$

with

$$H^{(0)} = -\frac{\hbar^2}{2\,m_e}\nabla^2 + V(\vec{r}) \tag{5.5}$$

(where m_e is the electron mass and $\hbar$ is the reduced Planck constant) and substituting to $\psi^n_{\vec{k}}(\vec{r})$ the generic expression of the Bloch function we obtain:

$$\begin{aligned}
&\left(-\frac{\hbar^2}{2\,m_e}\nabla^2 + V(\vec{r})\right) e^{i\,\vec{k}\cdot\vec{r}} u^n_{\vec{k}}(\vec{r}) = \\
&= -\frac{\hbar^2}{2\,m_e}\vec{\nabla}\cdot\left(e^{i\,\vec{k}\cdot\vec{r}}(\vec{\nabla}u^n_{\vec{k}}(\vec{r})) + (\vec{\nabla}e^{i\,\vec{k}\cdot\vec{r}})u^n_{\vec{k}}(\vec{r})\right) + V(\vec{r})e^{i\,\vec{k}\cdot\vec{r}}u^n_{\vec{k}}(\vec{r}) = \\
&= -\frac{\hbar^2}{2\,m_e}\vec{\nabla}\cdot\left(e^{i\,\vec{k}\cdot\vec{r}}(\vec{\nabla}u^n_{\vec{k}}(\vec{r}) + i\,\vec{k}u^n_{\vec{k}}(\vec{r}))\right) + V(\vec{r})e^{i\,\vec{k}\cdot\vec{r}}u^n_{\vec{k}}(\vec{r}) = \\
&= -\frac{\hbar^2}{2\,m_e}\Big(e^{i\,\vec{k}\cdot\vec{r}}\vec{\nabla}\cdot(\vec{\nabla}u^n_{\vec{k}}(\vec{r}) + i\,\vec{k}u^n_{\vec{k}}(\vec{r})) + \\
&+ (\vec{\nabla}e^{i\,\vec{k}\cdot\vec{r}})\cdot(\vec{\nabla}u^n_{\vec{k}}(\vec{r}) + i\,\vec{k}u^n_{\vec{k}}(\vec{r}))\Big) + V(\vec{r})e^{i\,\vec{k}\cdot\vec{r}}u^n_{\vec{k}}(\vec{r}) =
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{\hbar^2}{2\,m_e}\Big(e^{i\,\vec{k}\cdot\vec{r}}(\nabla^2 u^n_{\vec{k}}(\vec{r}) + i\vec{k}\cdot\vec{\nabla}u^n_{\vec{k}}(\vec{r})) + \\
&+ (i\,\vec{k}\,e^{i\,\vec{k}\cdot\vec{r}})\cdot(\vec{\nabla}u^n_{\vec{k}}(\vec{r}) + i\,\vec{k}u^n_{\vec{k}}(\vec{r}))\Big) + V(\vec{r})e^{i\,\vec{k}\cdot\vec{r}}u^n_{\vec{k}}(\vec{r}) = \\
=&-\frac{\hbar^2}{2\,m_e}e^{i\,\vec{k}\cdot\vec{r}}\Big(\nabla^2 u^n_{\vec{k}}(\vec{r}) + i\vec{k}\cdot\vec{\nabla}u^n_{\vec{k}}(\vec{r}) + i\vec{k}\cdot\vec{\nabla}u^n_{\vec{k}}(\vec{r}) - k^2 u^n_{\vec{k}}(\vec{r})\Big) + V(\vec{r})e^{i\,\vec{k}\cdot\vec{r}}u^n_{\vec{k}}(\vec{r}) = \\
&= e^{i\,\vec{k}\cdot\vec{r}}\left(\left(-\frac{\hbar^2}{2\,m_e}\nabla^2 + V(\vec{r})\right) - \frac{i\,\hbar^2}{m_e}\vec{k}\cdot\vec{\nabla} + \frac{\hbar^2k^2}{2\,m_e}\right)u^n_{\vec{k}}(\vec{r}) = \\
&= e^{i\,\vec{k}\cdot\vec{r}}(H^{(0)} + H^{(1)})u^n_{\vec{k}}(\vec{r}) = e^{i\,\vec{k}\cdot\vec{r}}E^n_{\vec{k}}u^n_{\vec{k}}(\vec{r})
\end{aligned}
\tag{5.6}
$$

and thus

$$
Hu^n_{\vec{k}}(\vec{r}) = (H^{(0)} + H^{(1)})u^n_{\vec{k}}(\vec{r}) = E^n_{\vec{k}}u^n_{\vec{k}}(\vec{r}) \tag{5.7}
$$

with

$$
H^{(1)} = -\frac{i\,\hbar^2}{m_e}\vec{k}\cdot\vec{\nabla} + \frac{\hbar^2k^2}{2\,m_e} \tag{5.8}
$$

(where $k = |\vec{k}|$). What we have just obtained is clearly an equation for the Bloch lattice functions, which needs to be solved only for a single primitive cell with the boundary condition that the function $u^n_{\vec{k}}(\vec{r})$ must be periodic with the lattice periodicity. For each value of $\vec{k}$ this equation has a periodic solution only for selected values $E^n_{\vec{k}}$ of the energy E. Noting that $H^{(1)}(\vec{r})$ reduces to zero when $\vec{k}$ approaches $\vec{0}$ and thus that this part of the Hamiltonian can be treated as a perturbation around $\vec{k} = \vec{0}$, we can locally solve this equation using the time-independent perturbation theory, assuming to know the eigenfunctions and eigenvalues of $H^{(0)}(\vec{r})$, i.e. the Bloch lattice functions and the energy band values for $\vec{k} = \vec{0}$.
For most of the semiconductors the maximum of the valence band is in the Γ-point (the center of the first Brillouin zone represented by the Wigner-Seitz method) and therefore corresponds to $\vec{k} = \vec{0}$; the minimum of the valence band instead is for $\vec{k} = \vec{0}$ only for the direct-gap semiconductors. When the extremum point of the energy band (and thus the interesting region) is for a generic $\vec{k}_0$, we can easily extend this argument observing that, if we define the value of H in $\vec{k}_0$ as

$$
H_{\vec{k_0}} = H^{(0)} - \frac{i\,\hbar^2}{m_e}\vec{k_0}\cdot\vec{\nabla} + \frac{\hbar^2{k_0}^2}{2\,m_e}\ , \tag{5.9}
$$

we have that the value of H in $\vec{k}$ is

$$
\begin{aligned}
H = H^{(0)} + H^{(1)} &= H_{\vec{k_0}} + \left[-\frac{i\,\hbar^2}{m_e}(\vec{k} - \vec{k_0})\cdot\vec{\nabla} + \frac{\hbar^2(k^2 - {k_0}^2)}{2\,m_e}\right] = \\
=H_{\vec{k_0}} + \left[-\frac{i\,\hbar^2}{m_e}(\vec{k} - \vec{k_0})\cdot\vec{\nabla} + \frac{\hbar^2(k^2 - {k_0}^2)}{2\,m_e} + \frac{\hbar}{m_e}(\vec{k} - \vec{k_0})\cdot\hbar\vec{k_0} - \frac{\hbar}{m_e}(\vec{k} - \vec{k_0})\cdot\hbar\vec{k_0}\right] &= \\
= H_{\vec{k_0}} + \left[\frac{\hbar}{m_e}(\vec{k} - \vec{k_0})\cdot(\hbar\vec{k_0} - i\,\hbar\vec{\nabla}) + \frac{\hbar^2}{2\,m_e}|\vec{k} - \vec{k_0}|^2\right]&
\end{aligned}
\tag{5.10}
$$

and for $\vec{k}$ near $\vec{k}_0$ the term between square brackets can be treated as a perturbation of $H_{\vec{k_0}}$ [15]. For the sake of simplicity, in the following we will consider $\vec{k}_0 = \vec{0}$.

An important point to notice is that, if we choose a value for $\vec{k}$, the functions $u^n_{\vec{k}}(\vec{r})$ form an orthogonal and complete set, in the restricted sense that any function with the lattice periodicity can be expanded in terms of the Bloch lattice functions for any selected $\vec{k}$.
To describe the main results of time-independent perturbation theory [16], we have to distinguish the case in which the unperturbed energy levels are non-degenerate from the case in which such a degeneration exists. Let us begin from the first case. The problem we have to solve is

$$[H^{(0)} + H^{(1)}]|n\rangle = E_n|n\rangle \tag{5.11}$$

where $H^{(0)}$ is the unperturbed Hamiltonian and $H^{(1)}$ the perturbation. If we expand the eigenvalues E_n and the eigenfunctions $|n\rangle$:

$$\begin{aligned} E_n &= E_n^{(0)} + E_n^{(1)} + E_n^{(2)} + \dots \\ |n\rangle &= |n\rangle^{(0)} + |n\rangle^{(1)} + |n\rangle^{(2)} + \dots \ , \end{aligned} \tag{5.12}$$

we insert these expressions into the eigenvalue equation and we enforce the identity between terms of the same order we find

$$\begin{aligned} H^{(0)}|n\rangle^{(0)} &= E_n^{(0)}|n\rangle^{(0)} \\ H^{(0)}|n\rangle^{(1)} + H^{(1)}|n\rangle^{(0)} &= E_n^{(0)}|n\rangle^{(1)} + E_n^{(1)}|n\rangle^{(0)} \\ H^{(0)}|n\rangle^{(2)} + H^{(1)}|n\rangle^{(1)} &= E_n^{(0)}|n\rangle^{(2)} + E_n^{(1)}|n\rangle^{(1)} + E_n^{(2)}|n\rangle^{(0)} \\ &\dots \end{aligned} \tag{5.13}$$

The first equation corresponds to the unperturbed eigenvalue equation, the solution of which, $E_n^{(0)} = E_n^0$ and $|n\rangle^{(0)} = |n0\rangle$, are assumed known. From the other equations, instead, we can obtain the corrections to these values produced by the perturbation $H^{(1)}$. In particular if we stop to the first-order corrections for the eigenfunctions and to the second-order corrections for the eigenvalues we find:

$$|n\rangle \simeq |n0\rangle + |n\rangle^{(1)} = |n0\rangle + \sum_{m\neq n}\left(|m0\rangle\frac{\langle m0|H^{(1)}|n0\rangle}{E_n^0 - E_m^0}\right) \tag{5.14}$$

(choosing $\langle n0|n\rangle^{(1)} = 0$) and

$$\begin{aligned} E_n \simeq E_n^0 + E_n^{(1)} + E_n^{(2)} &= E_n^0 + \langle n0|H^{(1)}|n0\rangle + \\ &+ \sum_{m\neq n}\left(\frac{\langle n0|H^{(1)}|m0\rangle\langle m0|H^{(1)}|n0\rangle}{E_n^0 - E_m^0}\right) \ . \end{aligned} \tag{5.15}$$

When we examine degenerate unperturbed states, the expressions we have just found diverge and thus we have to modify our treatment. In particular if the degenerate energy level $E_n^{(0)}$ corresponds to a multiplet

of degenerate states $|na0\rangle$ (with $a = 1, 2, \ldots, g_n$ where g_n is the degeneracy) and we have to solve the perturbed problem

$$H|\psi\rangle = [H^{(0)} + H^{(1)}]|\psi\rangle = E|\psi\rangle \tag{5.16}$$

we can express the new generic eigenfunction $|\psi\rangle$ as

$$|\psi\rangle = \sum_{a=1}^{g_n} |na\rangle\langle na|\psi\rangle \tag{5.17}$$

where the $|na\rangle$'s are states which are related to the unperturbed eigenvectors $|na0\rangle$'s by the perturbation matrix elements between different multiplets (as we will see). If we define

$$H^n_{ab} = \langle na|H|nb\rangle = \langle na|[H^{(0)} + H^{(1)}]|nb\rangle \tag{5.18}$$

we can express our perturbed equation in the following way:

$$\sum_{b=1}^{g_n} H^n_{ab}\langle nb|\psi\rangle = E\langle na|\psi\rangle \ . \tag{5.19}$$

Noting that the definition of the H^n_{ab}'s can be equivalently expressed in this way:

$$[H^{(0)} + H^{(1)}]|nb\rangle = \sum_{a=1}^{g_n} |na\rangle H^n_{ab} \ , \tag{5.20}$$

inserting into this equation the expansions

$$\begin{aligned} H^n_{ab} &= (H^n_{ab})^{(0)} + (H^n_{ab})^{(1)} + (H^n_{ab})^{(2)} + \ldots \\ |na\rangle &= |na\rangle^{(0)} + |na\rangle^{(1)} + |na\rangle^{(2)} + \ldots \ , \end{aligned} \tag{5.21}$$

and enforcing the identity of the terms of the same order, we find

$$\begin{aligned} H^{(0)}|nb\rangle^{(0)} &= \sum_{a=1}^{g_n} |na\rangle^{(0)}(H^n_{ab})^{(0)} \\ H^{(0)}|nb\rangle^{(1)} + H^{(1)}|nb\rangle^{(0)} &= \sum_{a=1}^{g_n} |na\rangle^{(1)}(H^n_{ab})^{(0)} + \sum_{a=1}^{g_n} |na\rangle^{(0)}(H^n_{ab})^{(1)} \\ H^{(0)}|nb\rangle^{(2)} + H^{(1)}|nb\rangle^{(1)} &= \sum_{a=1}^{g_n} |na\rangle^{(2)}(H^n_{ab})^{(0)} + \\ &+ \sum_{a=1}^{g_n} |na\rangle^{(1)}(H^n_{ab})^{(1)} + \sum_{a=1}^{g_n} |na\rangle^{(0)}(H^n_{ab})^{(2)} \\ &\ldots \end{aligned} \tag{5.22}$$

The first equation corresponds, noting that $(H^n_{ab})^{(0)} = E^0_n\delta_{ab}$, to the unperturbed eigenvalue equation, the solution of which, E^0_n and $|na\rangle^{(0)} = |na0\rangle$, are assumed known. From the other equations, instead, we can obtain the corrections to these values produced by the perturbation. In particular, if we stop to the first-order corrections for the eigenstates and to the second-order corrections for the eigenvalues, we find:

$$|nb> \simeq |nb0\rangle + |nb\rangle^{(1)} = |nb0\rangle + \sum_{m\neq n}\sum_{c=1}^{g_m}\left(|mc0\rangle\frac{\langle mc0|H^{(1)}|nb0\rangle}{E^0_n - E^0_m}\right) \tag{5.23}$$

(choosing $\langle nc0|nb\rangle^{(1)} = 0$) and

$$H^n_{cb} \simeq (H^n_{cb})^{(0)} + (H^n_{cb})^{(1)} + (H^n_{cb})^{(2)} = E^0_n\delta_{cb} + \langle nc0|H^{(1)}|nb0\rangle + \\ + \sum_{m\neq n}\sum_{a=1}^{g_m}\left(\frac{\langle nc0|H^{(1)}|ma0\rangle\langle ma0|H^{(1)}|nb0\rangle}{E^0_n - E^0_m}\right) . \tag{5.24}$$

Once the H^n_{cb} have been found, we can obtain the energy levels E solving the equation

$$\sum_{b=1}^{g_n} H^n_{ab}\langle nb|\psi\rangle = E\langle na|\psi\rangle , \tag{5.25}$$

or, equivalently, finding the eigenvalues of the matrix H^n (matrix $g_n \times g_n$ with elements H^n_{ab}) imposing

$$\det(H^n - EI) = 0$$

(with I the $g_n \times g_n$ unit matrix). We notice that, computing also the eigenvectors $\langle na|\psi\rangle$ of such a matrix and combining such results with the $|nb\rangle$ that have been computed before up to the first order, it is also possible to know the eigenfunctions $|\psi\rangle$ of the perturbed problem.
In the case of the $\vec{k}\cdot\vec{p}$ Hamiltonian that we have found before [17], we can use the $u^n_0(\vec{r})$ ($u^n_{\vec{k}}(\vec{r})$ for $\vec{k} = \vec{0}$) as $|n0\rangle$ and we have that

$$\langle m0|H^{(1)}|n0\rangle = \langle m0| - \frac{i\,\hbar^2}{m_e}(\vec{k}\cdot\vec{\nabla})|n0\rangle + \langle m0|\frac{\hbar^2k^2}{2\,m_e}|n0\rangle . \tag{5.26}$$

The second term clearly gives only diagonal matrix elements because it is equal to $(\hbar^2k^2/(2\,m_e))\delta_{nm}$. As to the first term, instead, it gives only non-diagonal matrix element because it is known [18] that

$$\langle n\vec{k}_0|(-i\hbar\vec{\nabla})|n\vec{k}_0\rangle + \hbar\vec{k}_0 = m_e\vec{v}_n = \frac{m_e}{\hbar}\vec{\nabla}_{\vec{k}}E^n_{\vec{k}} \tag{5.27}$$

(where $\vec{v}_n$ is the expectation value of the velocity of the Bloch waves, and in our considerations we are assuming $\vec{k}_0 = \vec{0}$) and $\vec{\nabla}_{\vec{k}}E^n_{\vec{k}} = \vec{0}$ in the band extrema.

Then, for non-degenerate unperturbed energy bands, we can write that

$$\begin{aligned} E^n_{\vec{k}} &= E^n_0 + \frac{\hbar^2 k^2}{2\,m_e} + \frac{\hbar^2}{{m_e}^2}\sum_{m\neq n}\frac{\langle n0|\vec{k}\cdot(-i\,\hbar\,\vec{\nabla})|m0\rangle\langle m0|\vec{k}\cdot(-i\,\hbar\,\vec{\nabla})|n0\rangle}{E^n_0 - E^m_0} = \\ &= E^n_0 + \frac{\hbar^2}{2}\sum_{\mu,\nu}\frac{k_\mu k_\nu}{m^*_{\mu\nu}} \end{aligned} \tag{5.28}$$

where $\mu,\nu = x,y,z$, while $m^*_{\mu\nu}$ is the effective-mass tensor defined by

$$\frac{1}{m^*_{\mu\nu}} = \frac{1}{m_e}\delta_{\mu\nu} + \frac{2}{{m_e}^2}\sum_{m\neq n}\frac{P^{nm}_\mu P^{mn}_\nu}{E^n_0 - E^m_0} \tag{5.29}$$

and the momentum matrix elements at the band extremum are

$$P^{nm}_\mu = \langle n0|(-i\,\hbar\,\nabla_\mu)|m0\rangle\ . \tag{5.30}$$

For degenerate unperturbed energy bands, instead, we have

$$\begin{aligned} (H^n_{\vec{k}})_{cb} &= E^n_0\delta_{cb} + \frac{\hbar^2 k^2}{2\,m_e}\delta_{cb} + \frac{\hbar}{m_e}\langle nc0|\vec{k}\cdot(-i\,\hbar\,\vec{\nabla})|nb0\rangle + \\ &+ \frac{\hbar^2}{{m_e}^2}\sum_{m\neq n}\sum_{a=1}^{g_m}\frac{\langle nc0|\vec{k}\cdot(-i\,\hbar\,\vec{\nabla})|ma0\rangle\langle ma0|\vec{k}\cdot(-i\,\hbar\,\vec{\nabla})|nb0\rangle}{E^n_0 - E^m_0} = \\ &= E^n_0\delta_{cb} + \frac{\hbar}{m_e}\sum_\mu k_\mu (P_\mu)^{nn}_{cb} + \frac{\hbar^2}{2}\sum_{\mu,\nu}\frac{k_\mu k_\nu}{m^{cb}_{\mu\nu}} \end{aligned} \tag{5.31}$$

where $\mu,\nu = x,y,z$, while $m^{cb}_{\mu\nu}$ is the effective-mass tensor defined by

$$\frac{1}{m^{cb}_{\mu\nu}} = \frac{1}{m_e}\delta_{cb}\delta_{\mu\nu} + \frac{2}{{m_e}^2}\sum_{m\neq n}\sum_{a=1}^{g_m}\frac{(P_\mu)^{nm}_{ca}(P_\nu)^{mn}_{ab}}{E^n_0 - E^m_0} \tag{5.32}$$

and the momentum matrix elements at the band extremum are

$$(P_\mu)^{nm}_{cb} = \langle nc0|(-i\,\hbar\,\nabla_\mu)|mb0\rangle\ . \tag{5.33}$$

In most of the cases all the $(P_\mu)^{nn}_{cb} = 0$ and the linear term in k_μ disappear. The energy levels will be found solving

$$\det\left(H^n_{\vec{k}} - EI\right) = 0\ . \tag{5.34}$$

Thus, to make a calculation of the energy bands we have to know the Bloch lattice functions at $\vec{k} = \vec{0}$. Most semiconductors of interest have the diamond or zinc blende crystal structure; for these materials we can choose as lattice primitive cell a Wigner-Seitz cell centered around an atomic site (the one with

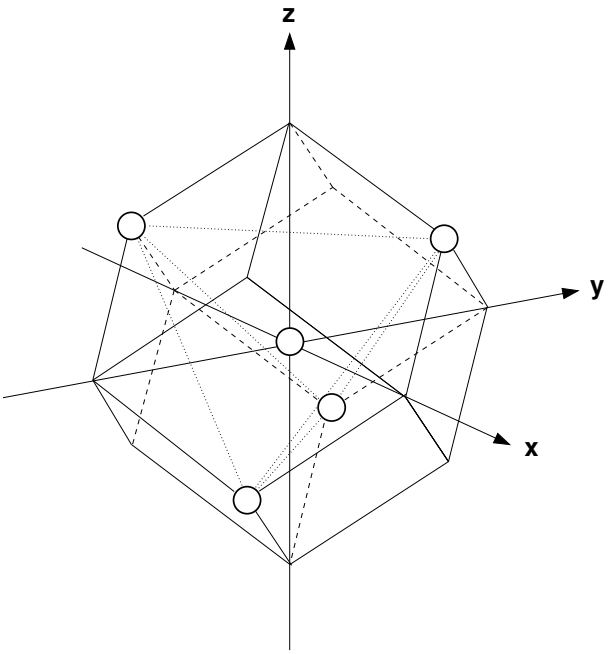


Figure 5.1: Wigner-Seitz primitive cell for the diamond or zinc blende structure.

the strongest potential in the case of the zinc blende structure, characterized by atoms that are not all identical) and with four other atoms forming a tetrahedron with the center coincident with the primitive cell center (Fig. 5.1). If we use a central force model, considering the potential inside the primitive cell as due only to the attraction of the nucleus of the central atom, shielded by its electrons, and afterwards we treat the electrostatic potential of the cores at the vertices of the primitive cell as a perturbation, we find that the Bloch lattice functions at $\vec{k} = \vec{0}$ exhibit particular symmetry properties (to the same result we can arrive using group theory). In particular the top of the valence band can be described with three degenerate states: $|vx0\rangle = \rho_v(r)x$, $|vy0\rangle = \rho_v(r)y$ and $|vz0\rangle = \rho_v(r)z$ (each one antisymmetric with respect to a coordinate and symmetric with respect to the others), while in most cases the bottom of the conductance band is described by a non-degenerate symmetric state $|c0\rangle = \rho_c(r)$ (with the important exception of silicon, where also the bottom of the conduction band is characterized by three states with symmetry properties analogous to the Bloch lattice functions of the valence band).
If we treat the conduction band as a non-degenerate band, we therefore obtain

$$E^c_{\vec{k}} = E^c_0 + \frac{\hbar^2}{2} \sum_{\mu,\nu} \frac{k_\mu k_\nu}{m^*_{\mu\nu}} \tag{5.35}$$

where $\mu, \nu = x, y, z$ and

$$\frac{1}{m^*_{\mu\nu}} = \frac{1}{m_e}\delta_{\mu\nu} + \frac{2}{{m_e}^2} \sum_{m \neq n} \frac{\langle c0|(-i\,\hbar\,\nabla_\mu)|m0\rangle\langle m0|(-i\,\hbar\,\nabla_\nu)|c0\rangle}{E^c_0 - E^m_0} \; . \tag{5.36}$$

The largest contribution to the sum comes from the bands m for which $|E_0^c - E_0^m|$ is smallest, i.e. from the three valence bands. If we compute the momentum matrix elements between the valence bands and the conduction band, we find that, due to the symmetry properties of the Bloch lattice functions,

$$\langle v\mu 0|(-i\,\hbar\,\nabla_\nu)|c0\rangle = -\langle c0|(-i\,\hbar\,\nabla_\nu)|v\mu 0\rangle = -i\,\hbar\,P\,\delta_{\mu\nu} \tag{5.37}$$

with $\mu,\nu = x,y,z$ and $P = \langle v\mu 0|\nabla_\mu|c0\rangle$ a non-zero quantity which multiplied by $\hbar$ has the dimensions of a momentum. Consequently the effective mass in the conduction band that we find is isotropic and equal to

$$\frac{1}{m^*_{\mu\nu}} = \frac{1}{m^*_c}\delta_{\mu\nu} = \left(\frac{1}{m_e} + \frac{2\,\hbar^2 P^2}{m_e^2 E_g^0}\right)\delta_{\mu\nu} \tag{5.38}$$

with $E_g^0 = E_0^c - E_0^v$.

As to the valence band, we have to use the degenerate perturbation theory and, with a motivation analogous to the one used in the study of the conduction band, we can consider only the interaction between the three degenerate valence bands and the conduction band, which is the nearest energy band. Thus, using the previous results, we will have that

$$(H^v_{\vec{k}})_{\alpha\beta} = E_0^v\delta_{\alpha\beta} + \frac{\hbar^2}{2}\sum_{\mu,\nu}\frac{k_\mu k_\nu}{m^{\alpha\beta}_{\mu\nu}} \tag{5.39}$$

with

$$\begin{aligned}
\frac{1}{m^{\alpha\beta}_{\mu\nu}} &= \frac{1}{m_e}\delta_{\alpha\beta}\delta_{\mu\nu} + \frac{2}{{m_e}^2}\sum_{m\neq v}\sum_{a=1}^{g_m}\frac{\langle v\alpha 0|(-i\,\hbar\,\nabla_\mu)|ma0\rangle\langle ma0|(-i\,\hbar\,\nabla_\nu)|v\beta 0\rangle}{E_0^v - E_0^m} = \\
&= \frac{1}{m_e}\delta_{\alpha\beta}\delta_{\mu\nu} + \frac{2}{{m_e}^2}\frac{\langle v\alpha 0|(-i\,\hbar\,\nabla_\mu)|c0\rangle\langle c0|(-i\,\hbar\,\nabla_\nu)|v\beta 0\rangle}{E_0^v - E_0^c} = \\
&= \frac{1}{m_e}\delta_{\alpha\beta}\delta_{\mu\nu} - \frac{2\,\hbar^2 P^2}{{m_e}^2 E_g^0}\delta_{\alpha\mu}\delta_{\beta\nu}
\end{aligned} \tag{5.40}$$

and thus the valence energy bands near the extremum can be obtained finding the eigenvalues of the matrix

$$H^v_{\vec{k}} = \left(E_0^v + \frac{\hbar^2 k^2}{2\,m_e}\right) I - \frac{\hbar^4 P^2}{m_e^2 E_g^0}\begin{bmatrix} k_x^2 & k_x k_y & k_x k_z \\ k_y k_x & k_y^2 & k_y k_z \\ k_z k_x & k_z k_y & k_z^2 \end{bmatrix}. \tag{5.41}$$

Till now we have not considered the effect of the so-called spin-orbit interaction, which often has a non-negligible influence on the energy bands. The physical phenomenon is the following [19]. An electron has an intrinsic magnetic moment

$$\vec{\mu} = -\gamma_e\frac{\hbar}{2}\,\vec{\sigma} = -g_e\gamma_L\frac{\hbar}{2}\,\vec{\sigma} = -g_e\frac{e}{2\,m_e}\frac{\hbar}{2}\,\vec{\sigma} = -g_e\mu_B\,\frac{\vec{\sigma}}{2} \tag{5.42}$$

where e is the modulus of the electron charge, $\vec{\sigma}$ is a vector with three components consisting in the Pauli spin matrices:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad , \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} , \tag{5.43}$$

γ_e is the intrinsic gyromagnetic ratio of the electron, γ_L is its orbital gyromagnetic ratio, $g_e = \gamma_e/\gamma_L$ is its intrinsic g-factor and $\mu_B = e\hbar/(2\,m_e)$ is the Bohr magneton. When an electron moves in a system (like the atom) where the charge distribution (for example the nucleus charge) produces an electric field $\vec{E}$, for the theory of relativity in the frame of reference of the electron this electric field will appear as a magnetic field. In particular if the electron moved uniformly, the equivalent magnetic field would be equal to $\vec{B} = -(\vec{v}\times\vec{E})/c^2$. The fact that the electron (and its frame of reference) is rotating halves such an equivalent magnetic field. Thus the Hamiltonian of the electron will have an additional part

$$H_{SO} = \mu_B\,\vec{\sigma}\cdot\left(\frac{\vec{E}\times\vec{v}}{2\,c^2}\right) = \frac{e\,\hbar}{4\,m_e c^2}\vec{\sigma}\cdot(\vec{E}\times\vec{v}) = \frac{\hbar}{4\,m_e c^2}\vec{\sigma}\cdot((\vec{\nabla}V)\times\vec{v}) \tag{5.44}$$

(with V the potential energy), which in the absence of an external magnetic field can be written also as

$$H_{SO} = \frac{\hbar}{4\,m_e^2 c^2}\vec{\sigma}\cdot((\vec{\nabla}V)\times\vec{p})\ . \tag{5.45}$$

However, if we insert this additional term in the original Schrödinger equation for the wave function $\psi^n_{\vec{k}}(\vec{r}) = e^{i\,\vec{k}\cdot\vec{r}}u^n_{\vec{k}}(\vec{r})$ we obtain

$$\begin{aligned} H_{SO}\psi^n_{\vec{k}}(\vec{r}) &= \frac{\hbar}{4\,m_e^2 c^2}\vec{\sigma}\cdot\left((\vec{\nabla}V)\times(-i\,\hbar\,\vec{\nabla})\right)e^{i\,\vec{k}\cdot\vec{r}}u^n_{\vec{k}}(\vec{r}) = \\ &= \frac{\hbar}{4\,m_e^2 c^2}\vec{\sigma}\cdot\left((\vec{\nabla}V)\times\left((\hbar\vec{k}e^{i\,\vec{k}\cdot\vec{r}})u^n_{\vec{k}}(\vec{r}) + e^{i\,\vec{k}\cdot\vec{r}}(-i\,\hbar\,\vec{\nabla}u^n_{\vec{k}}(\vec{r}))\right)\right) = \\ &= e^{i\,\vec{k}\cdot\vec{r}}\left(\frac{\hbar^2}{4\,m_e^2 c^2}\vec{\sigma}\cdot((\vec{\nabla}V)\times\vec{k}) + \frac{\hbar}{4\,m_e^2 c^2}\vec{\sigma}\cdot((\vec{\nabla}V)\times(-i\,\hbar\,\vec{\nabla}))\right)u^n_{\vec{k}}(\vec{r})\ . \end{aligned} \tag{5.46}$$

If we repeat the procedure used to move from the Schrödinger equation for the wave functions $\psi^n_{\vec{k}}(\vec{r})$ to the equation for the Bloch lattice functions $u^n_{\vec{k}}(\vec{r})$ we obtain that in the Hamiltonian of this last equation there will be two additional terms:

$$\frac{\hbar^2}{4\,m_e^2 c^2}\vec{\sigma}\cdot((\vec{\nabla}V)\times\vec{k}) + \frac{\hbar}{4\,m_e^2 c^2}\vec{\sigma}\cdot((\vec{\nabla}V)\times(-i\,\hbar\,\vec{\nabla})) = \frac{\hbar^2}{4\,m_e^2 c^2}\vec{\sigma}\cdot((\vec{\nabla}V)\times\vec{k}) + H_{SO}\ . \tag{5.47}$$

The first term near $\vec{k} = \vec{0}$ is small compared with the other term; thus only the second term is usually considered. The second term in the case of a potential energy with spherical symmetry (and thus of a

radial electric field) becomes

$$H_{SO} = \frac{e\,\hbar}{4\,m_e^2c^2}\vec{\sigma}\cdot(\vec{E}\times\vec{p}) = \frac{e\,\hbar}{4\,m_e^2c^2}\vec{\sigma}\cdot\frac{E_r}{r}(\vec{r}\times\vec{p}) = -i\left(\frac{e\,\hbar^2E_r}{4\,m_e^2c^2r}\right)\vec{\sigma}\cdot(\vec{r}\times\vec{\nabla})\ . \tag{5.48}$$

In order to calculate the influence that the spin-orbit term has on the valence bands, we need to calculate the matrix elements on the basis states $|vx0\rangle$, $|vy0\rangle$, $|vz0\rangle$ and $|c0\rangle$. Due to the symmetry proprieties of such states, we see that the only non-zero elements are the non-diagonal elements between valence band states:

$$\begin{aligned}
\langle vy0|H_{SO}|vx0\rangle &= -\langle vx0|H_{SO}|vy0\rangle = i\,\lambda\sigma_z \\
\langle vz0|H_{SO}|vy0\rangle &= -\langle vy0|H_{SO}|vz0\rangle = i\,\lambda\sigma_x \\
\langle vx0|H_{SO}|vz0\rangle &= -\langle vz0|H_{SO}|vx0\rangle = i\,\lambda\sigma_y
\end{aligned} \tag{5.49}$$

with λ an appropriate non-zero quantity obtained from the calculations.
Therefore considering also the spin-orbit coupling the matrix $H_{\vec{k}}^v$ becomes

$$H_{\vec{k}}^v = \left(E_0^v + \frac{\hbar^2k^2}{2\,m_e}\right)I - \frac{\hbar^4P^2}{m_e^2E_g^0}\begin{bmatrix} k_x^2 & k_xk_y & k_xk_z \\ k_yk_x & k_y^2 & k_yk_z \\ k_zk_x & k_zk_y & k_z^2 \end{bmatrix} + i\,\lambda\begin{bmatrix} 0 & -\sigma_z & \sigma_y \\ \sigma_z & 0 & -\sigma_x \\ -\sigma_y & \sigma_x & 0 \end{bmatrix} \tag{5.50}$$

where σ_x, σ_y and σ_z are the Pauli spin matrices, which do not commute with one another. If we consider the special case $\vec{k}\parallel\hat{z}$ we can quite easily find the eigenvalues of this matrix, arriving to a third-order equation in the energy, the solutions of which represent the dispersion relations of the three valence bands, each one degenerate with respect to the spin. In particular, if we make the approximation $(\hbar^4P^2k^2/(m_e^2E_g^0)) \ll \lambda$ we find the solutions:

$$\begin{aligned}
E_{hh} &= E_0^v + \lambda + \frac{\hbar^2}{2}\frac{1}{m_e}k^2 \\
E_{lh} &= E_0^v + \lambda + \frac{\hbar^2}{2}\frac{1}{m_e}\left(1 - \frac{4\,\hbar^2\,P^2}{3\,m_eE_g^0}\right)k^2 \\
E_{\lambda h} &= E_0^v - 2\,\lambda + \frac{\hbar^2}{2}\frac{1}{m_e}\left(1 - \frac{2\,\hbar^2\,P^2}{3\,m_eE_g^0}\right)k^2\ .
\end{aligned} \tag{5.51}$$

Thus, considering the effect of the spin-orbit interaction, we have obtained (Fig. 5.2) two valence bands (the heavy-holes band and the light-hole band) degenerate at $\vec{k}=\vec{0}$, where they have an energy $E_g = E_c^0 - (E_v^0+\lambda) = E_g^0 - \lambda$ lower than the conduction band, and one valence band (the spin-orbit band) which for $\vec{k}=\vec{0}$ has an energy $\Delta = 3\,\lambda$ lower than the other two valence bands. We notice that, while the

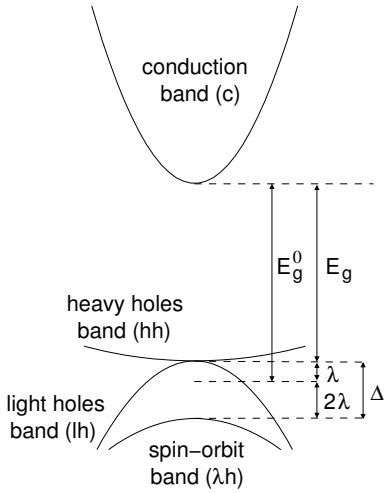


Figure 5.2: Band structure near $\vec{k} = \vec{0}$ obtained with the simplified model.

light-hole band and the spin-orbit band have a negative effective mass of the same order of magnitude of the effective mass of the electrons in the conduction band, the heavy-hole band is characterized by a much larger effective mass (the fact that the obtained effective mass is positive instead disappears with a more refined treatment: obviously the effective mass of the electrons in the valence bands has to be negative, which corresponds to a positive effective mass for the holes, which are the true charge carriers in such bands).

This simplified model is amenable to several refinements.

As to the conduction band, we can include in the calculation the spin-orbit splitting of the valence band and the effect of the higher conduction bands. In particular, with the first change we obtain a better expression for the effective mass in the conduction band:

$$\frac{1}{m_c^*} = \frac{1}{m_e} + \frac{2}{{m_e}^2}\left[\frac{2\,\hbar^2 P^2}{3\,E_g} + \frac{\hbar^2 P^2}{3(E_g + \Delta)}\right] \tag{5.52}$$

where $E_g = E_g^0 - \lambda$.

Also in the treatment of the valence bands we can consider the effect of the higher conduction bands; one of the effects is that the resulting valence bands lose their isotropy and exhibit a complex orientation dependence in the reciprocal space ("band warping").

It is important to notice that the expressions found for the band structure depend on a small number of parameters, for example E_g, Δ and m_c^* (from which we can calculate the parameter P using the expression found for the effective mass of the conduction band). These quantities are commonly obtained from "a priori" band structure calculations or, better, experimentally: in particular the bandgap values E_g and Δ are accurately known from optical experiments, while m_c^* is known from cyclotron resonance experiments.

The approach reported in this first part, based on the "traditional" perturbation theory and well described

by T. Wenckebach [17], differs from the method originally proposed by E. O. Kane [7], which uses the perturbation theory introduced by P. Löwdin [20]. Following this last method, we can divide all the bands into two sets A and B: A is the set we want to treat exactly and B contains all the other bands. At the lowest order of perturbation theory the coupling between the set A and the set B can be removed introducing the perturbed functions

$$u'_i = u_i + \sum_n^B \frac{H_{ni}u_n}{(H_{ii} - H_{nn})} \tag{5.53}$$

where i is in A and n is in B. The renormalized interactions connecting u'_i and u'_j are given by

$$H'_{ij} = H_{ij} + \sum_n^B \frac{H_{in}H_{nj}}{\left(\frac{H_{ii}+H_{jj}}{2} - H_{nn}\right)} \tag{5.54}$$

(with i, j in A). In this way we can reduce the Hamiltonian matrix, which in principle connects all the possible bands, to a Hamiltonian matrix relating only the bands of interest, in which, however, the interactions with the non considered bands are included. We notice that the Löwdin perturbation theory reduces to the ordinary perturbation theory when only a single band is considered in the set A.
As to the simplified case in which we consider only the three valence bands and the conduction band (not including the effects of the other bands) and in which $\vec{k} \parallel \hat{z}$, Kane solves exactly the Hamiltonian of the Bloch lattice functions in the presence of spin-orbit interaction [21] and finds, for small values of k^2, the following expressions for such bands (choosing the zero of energy at the top of the light-hole and heavy-hole bands and defining the various quantities as before):

$$\begin{aligned}
E_c &= E_g + \frac{\hbar^2}{2}\frac{1}{m_e}\left(1 + \frac{4\,\hbar^2\,P^2}{3\,m_e E_g} + \frac{2\,\hbar^2\,P^2}{3\,m_e(E_g+\Delta)}\right)k^2 \\
E_{hh} &= \frac{\hbar^2}{2}\frac{1}{m_e}k^2 \\
E_{lh} &= \frac{\hbar^2}{2}\frac{1}{m_e}\left(1 - \frac{4\,\hbar^2\,P^2}{3\,m_e E_g}\right)k^2 \\
E_{\lambda h} &= -\Delta + \frac{\hbar^2}{2}\frac{1}{m_e}\left(1 - \frac{2\,\hbar^2\,P^2}{3\,m_e(E_g+\Delta)}\right)k^2\ .
\end{aligned} \tag{5.55}$$

These expressions are very similar to the expressions obtained with the previously described simplified model, but clearly show the dual effect that each reciprocal interaction has on the related couple of bands. In this simple case, Kane also finds the Bloch lattice functions $u^n_{\vec{k}}(\vec{r})$ that diagonalize the Hamiltonian (i.e. the eigenfunctions of the Hamiltonian) as linear combinations of the $u^n_0(\vec{r})$ considered in the absence of spin-orbit (i.e. the functions $|c0\rangle$, $|vx0\rangle$, $|vy0\rangle$ and $|vz0\rangle$ taken with spin-up and spin-down); in particular for small k they are: $i|c0\rangle|\uparrow>$, $i|c0\rangle|\downarrow>$, $1/\sqrt{2}(|vx0\rangle + i\,|vy0\rangle)|\uparrow>$, $1/\sqrt{2}(|vx0\rangle - i\,|vy0\rangle)|\downarrow>$, $-\sqrt{2/3}|vz0\rangle|\uparrow> +1/\sqrt{6}(|vx0\rangle + i\,|vy0\rangle)|\downarrow>$, $-1/\sqrt{6}(|vx0\rangle - i\,|vy0\rangle)|\uparrow> -\sqrt{2/3}|vz0\rangle|\downarrow>$, $1/\sqrt{3}(|vx0\rangle + i\,|vy0\rangle)|\downarrow> +1/\sqrt{3}|vz0\rangle|\uparrow>$ and $-1/\sqrt{3}(|vx0\rangle - i\,|vy0\rangle)|\uparrow> +1/\sqrt{3}|vz0\rangle|\downarrow>$ (where $|\uparrow>$ and $|\downarrow>$ are

respectively the spin-up and spin-down unit spinors).
Kane treats instead the interaction of conductance and valence bands with the other bands through the previously cited Löwdin perturbation theory.

5.3 Envelope function theory and application to heterostructures

To introduce the concept of the envelope functions, we can make a very approximate calculation [22] in the hypothesis that the external potential energy $U(\vec{r})$ ("external" here meaning "not due to the periodic structure of the lattice") is slowly varying on an atomic scale and the n-th energy band that we are considering is non-degenerate (thus with unique independent Bloch lattice function $u^n_{\vec{k}}(\vec{r})$). In this case, the Schrödinger equation for the electron wave function $\psi(\vec{r})$:

$$\left(-\frac{\hbar^2}{2\,m_e}\nabla^2 + U_L(\vec{r})\right)\psi(\vec{r}) + U(\vec{r})\psi(\vec{r}) = H\psi(\vec{r}) + U(\vec{r})\psi(\vec{r}) = E\psi(\vec{r}) \tag{5.56}$$

(where $U_L(\vec{r})$ is the periodic lattice potential energy and H is the Hamiltonian in the absence of the external potential energy $U(\vec{r})$) is equivalent to the equation

$$E_n(-i\,\vec{\nabla})F(\vec{r}) + U(\vec{r})F(\vec{r}) = EF(\vec{r}) \tag{5.57}$$

where $E_n(-i\,\vec{\nabla})$ represents the operator obtained replacing, in the dispersion relation $E_n(\vec{k})$ describing the n-th energy band, each component of $\vec{k}$ with the correspondent component of $-i\,\vec{\nabla}$, and $F(\vec{r})$ is the envelope function, a slowly varying function that, when we consider only the n-th band, multiplied by the fast varying Bloch lattice function $u^n_0(\vec{r})$ (considered in $\vec{k}=\vec{0}$) gives the electron wave function. Indeed, if we expand $\psi(\vec{r})$ in the orthogonal basis set $|\nu\vec{k}\rangle = e^{i\,\vec{k}\cdot\vec{r}}u^\nu_{\vec{k}}(\vec{r})/\sqrt{V}$ (with V the crystal volume):

$$\psi(\vec{r}) = \sum_{\nu,\vec{k}} a_\nu(\vec{k})|\nu\vec{k}\rangle \tag{5.58}$$

we can re-write the Schrödinger equation in matrix form using the basis $|\nu\vec{k}\rangle$:

$$\begin{aligned}
\sum_{\nu',\vec{k}'}\left(\langle\nu\vec{k}|H+U(\vec{r})|\nu'\vec{k}'\rangle a_{\nu'}(\vec{k}')\right) &= Ea_\nu(\vec{k}) \\
E_\nu(\vec{k})a_\nu(\vec{k}) + \sum_{\nu',\vec{k}'}\left(\langle\nu\vec{k}|U(\vec{r})|\nu'\vec{k}'\rangle a_{\nu'}(\vec{k}')\right) &= Ea_\nu(\vec{k})
\end{aligned} \tag{5.59}$$

where we have used the fact that

$$\langle\nu\vec{k}|H|\nu'\vec{k}'\rangle = E_{\nu'}(\vec{k}')\langle\nu\vec{k}|\nu'\vec{k}'\rangle = E_\nu(\vec{k})\delta_{\nu\,\nu'}\delta_{\vec{k}\,\vec{k}'}\ . \tag{5.60}$$

In particular for $\nu = n$ we have that

$$E_n(\vec{k})a_n(\vec{k}) + \sum_{\nu',\vec{k}'} \left(\langle n\vec{k}|U(\vec{r})|\nu'\vec{k}'\rangle a_{\nu'}(\vec{k}') \right) = Ea_n(\vec{k}) \ . \tag{5.61}$$

If instead we expand the envelope function equation in the orthogonal set of plane waves $|\vec{k}\rangle = e^{i\,\vec{k}\cdot\vec{r}}/\sqrt{V}$:

$$F(\vec{r}) = \sum_{\vec{k}} a(\vec{k})|\vec{k}\rangle \tag{5.62}$$

we can re-write the effective mass equation in matrix form using the base $|\vec{k}\rangle$:

$$\begin{aligned} \sum_{\vec{k}'} \left(\langle \vec{k}|E_n(-i\,\vec{\nabla}) + U(\vec{r})|\vec{k}'\rangle a(\vec{k}') \right) &= Ea(\vec{k}) \\ E_n(\vec{k})a(\vec{k}) + \sum_{\vec{k}'} \left(\langle \vec{k}|U(\vec{r})|\vec{k}'\rangle a(\vec{k}') \right) &= Ea(\vec{k}) \end{aligned} \tag{5.63}$$

using the fact that

$$(-i\,\nabla_\nu)^p|\vec{k}'\rangle = (-i\,\nabla_\nu)^p(e^{i\,\vec{k}'\cdot\vec{r}}/\sqrt{V}) = (k'_\nu)^p(e^{i\,\vec{k}'\cdot\vec{r}}/\sqrt{V}) = (k'_\nu)^p|\vec{k}'\rangle \tag{5.64}$$

with $\nu = x, y, z$ and thus

$$E_n(-i\,\vec{\nabla})|\vec{k}'\rangle = E_n(\vec{k}')|\vec{k}'\rangle \tag{5.65}$$

(being $E_n(-i\,\vec{\nabla})$ an operator composed by operators of the type $(-i\,\nabla_\nu)^p$) and then exploiting the orthogonality relation $\langle\vec{k}|\vec{k}'\rangle = \delta_{\vec{k}\,\vec{k}'}$. The two equations 5.61 and 5.63, obtained from the Schrödinger equation and from the envelope function equation are exactly equal if

$$\langle n\vec{k}|U(\vec{r})|\nu'\vec{k}'\rangle = \delta_{n\,\nu'}\langle\vec{k}|U(\vec{r})|\vec{k}'\rangle \quad , \tag{5.66}$$

i.e. if the matrix elements of the external potential $U(\vec{r})$ between states from different bands are negligible. This is what happens if U is slowly varying on an atomic scale. Indeed in this case we have that (if we consider the $u^\nu_{\vec{k}}(\vec{r})$ normalized in a unit volume)

$$\begin{aligned} \langle n\vec{k}|U(\vec{r})|\nu'\vec{k}'\rangle &= \frac{1}{V}\sum_{j=1}^{N}\int_{V_j} d^3\vec{r}\, u^{n*}_{\vec{k}}(\vec{r})u^{\nu'}_{\vec{k}'}(\vec{r})e^{i\,(\vec{k}'-\vec{k})\cdot\vec{r}}U(\vec{r}) \simeq \\ &\simeq \sum_{j=1}^{N} e^{i\,(\vec{k}'-\vec{k})\cdot\vec{r}_j}U(\vec{r}_j)\frac{1}{V}\int_{V_j} d^3\vec{r}\, u^{n*}_{\vec{k}}(\vec{r})u^{\nu'}_{\vec{k}'}(\vec{r}) \simeq \\ \simeq \sum_{j=1}^{N} e^{i\,(\vec{k}'-\vec{k})\cdot\vec{r}_j}U(\vec{r}_j)\delta_{n\,\nu'}\frac{1}{N} &\simeq \delta_{n\,\nu'}\int_V d^3\vec{r}\,\frac{e^{i\,(\vec{k}'-\vec{k})\cdot\vec{r}}}{V}U(\vec{r}) = \delta_{n\,\nu'}\langle\vec{k}|U(\vec{r})|\vec{k}'\rangle \end{aligned} \tag{5.67}$$

where V the crystal volume, V_j the volume of the j-th unit cell, $\vec{r}_j$ the coordinate of its center and N the number of unit cells. We have assumed that $U(\vec{r})$ and $e^{i\,(\vec{k}'-\vec{k})\cdot\vec{r}}$ are approximately constant over a unit cell and $u^n_{\vec{k}'}(\vec{r}) \simeq u^n_{\vec{k}}(\vec{r})$ over the range of values of $|\vec{k}'-\vec{k}|$ for which $\langle\vec{k}|U(\vec{r})|\vec{k}'\rangle$ is not negligible.
If the two equations 5.61 and 5.63 are identical, they have the same solutions $a_n(\vec{k})$ and $a(\vec{k})$. Thus (assuming that $a_\nu(\vec{k})$ is non-zero only for the particular band n, coherently with our hypothesis that there is no mixing between the bands) we can write that

$$\psi(\vec{r}) = \sum_{\nu,\vec{k}} a_\nu(\vec{k})|\nu\vec{k}\rangle = \sum_{\vec{k}} a_n(\vec{k})\frac{e^{i\,\vec{k}\cdot\vec{r}}}{\sqrt{V}}u^n_{\vec{k}}(\vec{r}) \simeq$$

$$\simeq u^n_0(\vec{r})\sum_{\vec{k}} a_n(\vec{k})\frac{e^{i\,\vec{k}\cdot\vec{r}}}{\sqrt{V}} = u^n_0(\vec{r})\sum_{\vec{k}} a_n(\vec{k})|\vec{k}\rangle = u^n_0(\vec{r})F(\vec{r}) \tag{5.68}$$

where we have used the leading approximation for $u^n_{\vec{k}}(\vec{r})$.
We notice that if we express $E_n(\vec{k})$ as

$$E_n(\vec{k}) = E^n_0 + \frac{\hbar^2}{2}\sum_{\mu,\nu}\frac{k_\mu k_\nu}{m^*_{\mu\,\nu}} \tag{5.69}$$

(with $\mu,\nu = x,y,z$) the envelope function equation becomes

$$-\frac{\hbar^2}{2}\sum_{\mu,\nu}\frac{\nabla_\mu\nabla_\nu}{m^*_{\mu\,\nu}}F(\vec{r}) + (E^n_0 + U(\vec{r})) = EF(\vec{r}) \tag{5.70}$$

and when the effective mass is isotropic ($(1/m^*_{\mu\,\nu}) = (1/m^*)\delta_{\mu\,\nu}$)

$$-\frac{\hbar^2}{2\,m^*}\nabla^2 F(\vec{r}) + (E^n_0 + U(\vec{r})) = EF(\vec{r}) \tag{5.71}$$

(effective mass equation).
Luttinger and Kohn have studied the general form of the envelope function equations and in particular of the Hamiltonian that in such equations appears, with the aim to arrive at the determination of the dispersion relations of the energy bands as eigenvalues of such Hamiltonian [23]. In a famous paper [8] they start from the Schrödinger equation $(H_0+U)\psi = E_n(\vec{k})\psi$, with H_0 the Hamiltonian of the electron in the periodic lattice potential and U an additional potential which is assumed not to vary much over each unit cell. They expand, operating in the momentum space, ψ in the complete orthonormal set of functions $|n\,\vec{k}\rangle = e^{i\,\vec{k}\cdot\vec{r}}u^n_0(\vec{r})$ and re-write the Schrödinger equation accordingly. Then they remove the coupling between the n-th band and the others to the first order by means of a canonical transformation, treating accurately only the terms to the second order in $\vec{k}$. At the end they go back to the position space and they introduce the envelope function, in terms of which the resulting equation reads

$$(E_n(-i\,\vec{\nabla}) + U(\vec{r}))F_n(\vec{r}) = EF_n(\vec{r}) \tag{5.72}$$

(the envelope function equation that we have already introduced), with $E_n(-i\,\vec{\nabla})$ obtained expanding $E_n(\vec{k})$ (the dispersion relation in the absence of $U(\vec{r})$) to second order in $\vec{k}$ around $\vec{k} = \vec{0}$ with the non-degenerate perturbation theory:

$$E_n(\vec{k}) = E_0^n + \frac{\hbar^2 k^2}{2\,m_e} + \frac{\hbar^2}{{m_e}^2} \sum_{\alpha,\beta} k_\alpha k_\beta \sum_{m \neq n} \frac{P_\alpha^{nm} P_\beta^{mn}}{E_0^n - E_0^m} \tag{5.73}$$

(with $\alpha, \beta = x, y, z$) and substituting each component of $\vec{k}$ with the correspondent component of $-i\,\vec{\nabla}$. The leading term in the wave function $\psi(\vec{r})$ in general turns out to be related to the envelope function by

$$\psi = \sum_n F_n(\vec{r}) u_0^n(\vec{r}) \tag{5.74}$$

which in this case, having removed all the interband coupling, reduces to

$$\psi = F_n(\vec{r}) u_0^n(\vec{r}) \tag{5.75}$$

(the relation which we have already stated). If locally the external potential changes considerably within a cell, in that region the thus derived equation is no longer valid, but it continues to be valid in regions of space sufficiently distant from it.

Then they demonstrate that in the presence of an external magnetic field the envelope function satisfies an equation similar to the one in the absence of a magnetic field, the only difference being that the new Hamiltonian is obtained replacing, in the expansion of $E_n(\vec{k})$ to quadratic terms, each k_α by the operator $-i\nabla_\alpha + (e/\hbar)A_\alpha$ (in the MKS unit system; in the cgs system we have to divide also by c), with A_α the α-th component of the vector potential (and c the speed of light). Moreover in the expansion of $E_n(\vec{k})$ to the second order any product of noncommuting factors which arises has to be interpreted as the symmetrized product.

In the case in which the extremum is at $\vec{k} = \vec{k}_0 \neq \vec{0}$, the demonstrations (both in the absence and in the presence of an external magnetic field) can be repeated just replacing $u_0^n(\vec{r})$ (the eigenfunctions for $\vec{k} = \vec{0}$ in the absence of $U(\vec{r})$ and of external magnetic field) with $e^{i\,\vec{k}_0 \cdot \vec{r}} u_{\vec{k}_0}^n(\vec{r})$ (the eigenfunctions for $\vec{k} = \vec{k}_0$ in the absence of $U(\vec{r})$ and of external magnetic field). In this case the envelope function equation remains the same, but with E_n expanded up to the second order terms around $\vec{k}_0$ instead of around $\vec{0}$ and the relation between the wave function and the envelope function is

$$\psi = F_n(\vec{r})(e^{i\,\vec{k}_0 \cdot \vec{r}} u_{\vec{k}_0}^n(\vec{r})) \ . \tag{5.76}$$

If there are extrema at several different values of $\vec{k}_0$ within the band, we obtain an envelope function equation for each of these; if the solutions corresponding to the different $\vec{k}_0$ have different energies, the corresponding wave functions represent independent solutions of the Schrödinger equation; otherwise the correct wave function will be a linear combination of those from the different extrema which lead to the same energy.

When the band of interest is degenerate, the calculation can be done in a similar way, with the proper changes, arriving at a set of coupled second order equations which correspond to the effective mass equation found in the case of non-degenerate bands. In particular (considering the point $\vec{k}=\vec{0}$) let us assume to have r unperturbed degenerate Bloch lattice functions corresponding to the same unperturbed energy E_0^n (where "unperturbed" means for $\vec{k}=\vec{0}$ and in the absence of $U(\vec{r})$ and external magnetic field) and let us define them ϕ_j (with $j=1,\ldots,r$ where r is the degeneracy), i.e.

$$H_0\phi_j = E_0^n\phi_j \tag{5.77}$$

(notice that the ϕ_j, i.e. the u_0^n, can be seen as Bloch functions $e^{i\,\vec{k}\cdot\vec{r}}u_{\vec{k}}^n$ for $\vec{k}=\vec{0}$ and so they have to satisfy the Schrödinger equation for $\vec{k}=\vec{0}$). We will instead indicate as ϕ_i (with $i\neq 1,\ldots,r$) the other unperturbed Bloch lattice functions, not degenerate with the ϕ_j. With this convention, the result to which Luttinger and Kohn arrive is that the r envelope functions $F_j(\vec{r})$ corresponding to the originally degenerate energy bands satisfy the r coupled differential equations:

$$\sum_{j'=1}^{r}\left(\sum_{\alpha,\beta}\left(D_{jj'}^{\alpha\beta}(-i\nabla_\alpha)(-i\nabla_\beta)\right)+(E_0^n+U(\vec{r}))\delta_{j\,j'}\right)F_{j'}(\vec{r})=EF_j(\vec{r}) \tag{5.78}$$

(where in the absence of an external potential energy and setting the energy zero at E_0^n the term multiplied by $\delta_{j\,j'}$ in the equation disappears) with $\alpha,\beta=x,y,z$ and

$$D_{jj'}^{\alpha\beta}=\frac{\hbar^2}{2\,m_e}\delta_{j\,j'}\delta_{\alpha\,\beta}+\frac{\hbar^2}{{m_e}^2}\sum_i\frac{(P_\alpha)_{j\,i}(P_\beta)_{i\,j'}}{(E_0^n-E_0^i)}\ . \tag{5.79}$$

The leading term of the complete wave functions are related to the envelope functions by:

$$\psi=\sum_{j=1}^{r}F_j(\vec{r})\phi_j(\vec{r})\ . \tag{5.80}$$

We notice that the numbers $D_{jj'}^{\alpha\beta}$ play the same role in the case of degenerate bands that $\hbar^2/(2\,m_{\alpha\beta}^*)$ do for a non-degenerate band.

As before, in the presence of a magnetic field the elements of the Hamiltonian will be obtained from $\sum_{\alpha,\beta}(D_{jj'}^{\alpha\beta}k_\alpha k_\beta)$ replacing each component of $\vec{k}$ with the correspondent component of $-i\vec{\nabla}+(e/\hbar)\vec{A}$ (in the MKS unit system; in the cgs system we have to divide also by c).

In the presence of spin-orbit coupling, Luttinger and Kohn adopt the same treatment followed in its absence considering the spin-orbit contribution as part of the unperturbed Hamiltonian (the total unperturbed Hamiltonian will be H_0+H_{SO}) and the Bloch lattice functions and the correspondent energies for $\vec{k}=\vec{0}$ as known quantities. Thus the u_0^n are replaced with the $\tilde{u}_0^n$ (the spin-dependent Bloch lattice functions for $\vec{k}=\vec{0}$), $E_n(\vec{k})$ by $\tilde{E}_n(\vec{k})$ (the dispersion relation in the presence of spin-orbit) and the

$(P_\alpha)_{n\,n'}$ by

$$(\pi_\alpha)_{n\,n'} = \langle \tilde{u}_0^n | \left(-i\,\hbar\nabla_\alpha + \frac{\hbar}{4\,m_e c^2}(\vec{\sigma} \times \vec{\nabla} V)_\alpha \right) | \tilde{u}_0^{n'} \rangle \tag{5.81}$$

where the extra term arises from the fact that the spin-orbit coupling contains the differential operation $\vec{p}$. When we treat energy bands which are degenerate in the absence of spin-orbit, we have to consider that (as seen previously) the spin-orbit coupling can lift, at least partially, the degeneracy. In such a case we have to consider that the validity of the adopted theory rests on the assumption that the interband separations are large compared with the energies involved in the solution of the envelope function equation. Thus we have to evaluate if the external potential U or the magnetic field are sufficiently small to produce no appreciable mixing of the bands, the degeneracy of which has been splitted by the spin-orbit coupling. If it is sufficiently small, we can obtain a different set of coupled envelope function equations for each set of bands that have remained degenerate; otherwise we will have to deal with the full set of coupled equations for all the bands that are degenerate in the absence of spin-orbit.
We can introduce a matrix D, the elements of which are

$$D_{jj'} = \sum_{\alpha,\beta} D_{jj'}^{\alpha\beta} k_\alpha k_\beta \ . \tag{5.82}$$

If in such matrix elements we replace each component of the vector $\vec{k}$ with the correspondent component of the operator $-i\,\vec{\nabla} + (e/\hbar)\vec{A}$ (divided by c in the cgs system) we obtain the terms which appear in the envelope function coupled equations. As we have seen, the envelope function coupled equations written in the absence of an external perturbation read (if we set the energy zero at E_0^n):

$$\sum_{j'=1}^{r} \sum_{\alpha,\beta} (D_{jj'}^{\alpha\beta}(-i\nabla_\alpha)(-i\nabla_\beta)) F_{j'}(\vec{r}) = E F_j(\vec{r}) \quad . \tag{5.83}$$

If we convert them from the position representation to the momentum representation, we obtain

$$\begin{aligned} \sum_{j'=1}^{r} \sum_{\alpha,\beta} (D_{jj'}^{\alpha\beta} k_\alpha k_\beta) B_{j'}(\vec{k}) &= E B_j(\vec{k}) \\ \sum_{j'=1}^{r} D_{jj'} B_{j'}(\vec{k}) &= E B_j(\vec{k}) \\ D\vec{B} &= E\vec{B} \end{aligned} \tag{5.84}$$

(where $B_j(\vec{k})$ is the Fourier transform of $F_j(\vec{r})$), from which it is evident that the dispersion relations $E(\vec{k})$ near the extremum can be obtained finding the eigenvalues of the matrix D. We notice that this is clearly correspondent to what happens in the case of non-degeneracy, in which (as we have seen) the envelope function equation contains $E_n(-i\,\vec{\nabla})$ (the dispersion relation in the absence of external potential energy or magnetic field, in which each component of $\vec{k}$ is substituted with the correspondent component of $-i\,\vec{\nabla}$).
Using group theory, in particular considering that the Hamiltonian D should be invariant under the

operations of the cubic group (so that the Hamiltonian will give us results which transform correctly with respect to the transformations of the cubic group, which is the symmetry group of $\vec{k}$) and thus writing D as linear combination of the invariants obtained combining angular momentum matrices and components of $\vec{k}$, Luttinger [9] obtains explicit expressions for D. The elements of such a matrix are polynomials in k at most of the second-order and involve parameters characteristic of the materials, which have been experimentally found and are available for most common semiconductors [24]. For example in the case of the 4×4 matrix D corresponding to the light-holes and heavy-holes bands (the extra factor of 2 coming from spin) they are γ_1, γ_2, γ_3, κ (which is useful in the presence of an external magnetic field) and q (which approaches zero as the spin-orbit coupling does).

We notice that this method has been generalized, especially for narrow band-gap semiconductors, to include also the conduction band in the set of energy bands for which the matrix D is computed.

Bir and Pikus [10] have shown that in uniformly strained semiconductors, such that the periodicity of the structure is preserved, the strain introduces in the dispersion relation of non-degenerate bands an extra term of the kind

$$a_c(\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}) \tag{5.85}$$

and in the Hamiltonian of degenerate bands additional terms of the form

$$\sum_{\alpha,\beta} \hat{D}^{\alpha\beta}_{j\,j'} \epsilon_{\alpha\beta} \tag{5.86}$$

where $\alpha, \beta = x, y, z$ and $\epsilon_{\alpha\,\beta}$ is the generic component of the strain matrix.

Bastard [11], adopting the Kane approach, analogously to Luttinger notices that the Hamiltonian which appears in the envelope function equation in the absence of an external potential energy is equal, apart from the substitution of the components of $-i\,\vec{\nabla}$ with the correspondent components of $\vec{k}$, to the matrix the eigenvalues of which give the dispersion relations of the material. Then he uses the envelope function method to study heterostructures, for example made up of two materials A and B (Fig. 5.3). In particular he assumes that the two materials are perfectly lattice-matched and crystallize with the same crystallographic structure, so that the functions $u_0^n(\vec{r})$ in the two materials can be considered identical. With this hypothesis, if in each material the wave functions are written as

$$\psi^{(A,B)} = \sum_n F_n^{(A,B)}(\vec{r}) u_0^n(\vec{r}) \tag{5.87}$$

it is evident that, since the u_0^n are linearly independent and the wave function has to be continuous at the interface, also the envelope functions have to be continuous at the interface. As to the derivative of the envelope functions, Bastard finds, imposing the continuity of the probability current at the interface, a general condition [25], which in the simple case in which the two materials are both characterized by non-degenerate parabolic and isotropic bands but by different effective masses $m^*_{(A)}$ and $m^*_{(B)}$ reduces to imposing the continuity of

$$\frac{1}{m^*} \frac{\partial F_n}{\partial z} \tag{5.88}$$

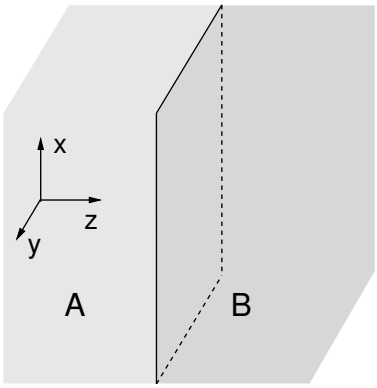


Figure 5.3: Heterojunction between two semiconductors A and B.

(where we have assumed the $\hat{z}$ axis orthogonal to the interface). This can be easily obtained enforcing in this case the continuity of the z component of the probability current density, which is equal to

$$j_z = -\frac{i\,\hbar}{2\,m}\left(\psi^*\frac{\partial\psi}{\partial z} - \psi\frac{\partial\psi^*}{\partial z}\right) \tag{5.89}$$

and noting that the continuity of the envelope function has already been enforced. As to the asymptotic behavior of the envelope functions far from the interface, it depends on the heterostructure under consideration. For example for superlattices the z-dependent part of the envelope function will be a Bloch wave, due to the periodicity of the structure in that direction, while for the bound states of a quantum well it should tend to zero for large z. Thus the envelope functions in the overall structure can be found solving the envelope function equations in the different materials, knowing the asymptotic behavior far from the interface and enforcing the correct boundary conditions at the interface. Bastard has also made an extensive analysis of the applications of this method [26].
Also M. Altarelli has given important contributions to the development of the envelope function method [27] and to its applications to the study of heterostructures [12].
M. G. Burt [13] has pointed out the errors deriving from the assumption, normally made in the application of the envelope function method to heterostructures, that the $u_0^n(\vec{r})$ in the two materials be the same and from the boundary condition enforced on the derivative of the envelope function at the interface. In a series of interesting and detailed articles he has developed an alternative envelope function theory expanding the wave function in the entire structure on the same periodic basis functions $U_n(\vec{r})$ throughout, even though they are not necessarily eigenstates of the constituent crystals, without making any hypothesis on the real eigenstates $u_0^n(\vec{r})$:

$$\psi(\vec{r}) = \sum_n F_n(\vec{r})U_n(\vec{r})\ . \tag{5.90}$$

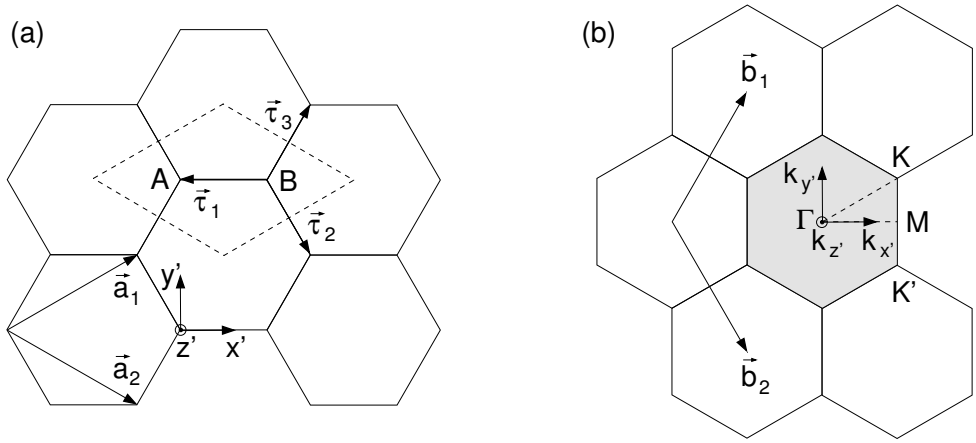


Figure 5.4: The graphene lattice in the real space (a) and in the reciprocal space (b).

The envelope functions $F_n(\vec{r})$ univocally defined in this way and all their derivatives are certainly continuous everywhere, including at the interface. Using this approach, he has first derived exact envelope function equations, then for local potentials and slowly varying envelope functions (but without any assumption on the rate of variation of the composition) has formulated approximated envelope functions equations, and finally, with the usual assumption of the dominance of one envelope function, has arrived at an effective-mass equation that includes also the effect of the differences in the $u_0^n(\vec{r})$ between the two materials. At each step the associated approximations are accurately described, so that it is possible to estimate the error.

5.4 Application of the $\vec{k} \cdot \vec{p}$ method to carbon nanotubes

In the following we show the application of the $\vec{k} \cdot \vec{p}$ method to a material which is particularly interesting for its physical properties: carbon nanotubes.

A single-wall carbon nanotube can be described as a graphene sheet rolled, along one of its lattice translational vectors, into a cylindrical shape. To find its energy dispersion relations and electron wave functions we will study the two-dimensional graphite (graphene) and at the end we will enforce periodic boundary conditions in the circumferential direction.

A graphene sheet is a hexagonal lattice of carbon atoms. In Fig. 5.4(a) we show its structure in the real space and in particular its unit cell as a dashed rhombus, containing two inequivalent carbon atoms A and B, while in Fig. 5.4(b) we show the lattice in the reciprocal space with the Brillouin zone as a shaded hexagon. The lattice unit vectors are $\vec{a}_1$ and $\vec{a}_2$ in the real space, and $\vec{b}_1$ and $\vec{b}_2$ in the reciprocal space. If we define $a = |\vec{a}_1| = |\vec{a}_2| = a_{C-C}\sqrt{3}$ (with a_{C-C} the distance between nearest-neighbor carbon atoms),

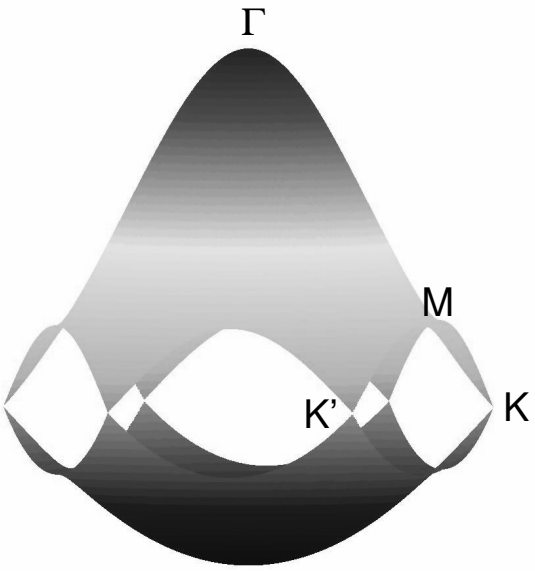


Figure 5.5: The energy dispersion relations of the graphene inside its hexagonal Brillouin zone.

the coordinates of these vectors in the represented reference frame are:

$$\vec{a}_1 = \begin{bmatrix} \frac{\sqrt{3}}{2}a \\ \frac{a}{2} \\ 0 \end{bmatrix} \quad , \quad \vec{a}_2 = \begin{bmatrix} \frac{\sqrt{3}}{2}a \\ -\frac{a}{2} \\ 0 \end{bmatrix} \quad , \quad \vec{b}_1 = \begin{bmatrix} \frac{2\pi}{\sqrt{3}a} \\ \frac{2\pi}{a} \\ 0 \end{bmatrix} \quad , \quad \vec{b}_2 = \begin{bmatrix} \frac{2\pi}{\sqrt{3}a} \\ -\frac{2\pi}{a} \\ 0 \end{bmatrix} \tag{5.91}$$

(following the conventions used by R. Saito, G. Dresselhaus and M. S. Dresselhaus [28]), which fulfill the well-know relation $\vec{a}_i \cdot \vec{b}_j = 2\pi\delta_{ij}$ between lattice unit vectors in the real space and in the reciprocal space. The most relevant graphene dispersion relations for transport and other solid state properties are the two π-bands (a upper anti-bonding band and a lower bonding band), which are degenerate at the points

$$\vec{K} = \begin{bmatrix} \frac{2\pi}{\sqrt{3}a} \\ \frac{2\pi}{3a} \\ 0 \end{bmatrix} = \frac{2\pi}{\sqrt{3}a} \begin{bmatrix} 1 \\ \frac{1}{\sqrt{3}} \\ 0 \end{bmatrix} \quad \text{and} \quad \vec{K}' = \begin{bmatrix} \frac{2\pi}{\sqrt{3}a} \\ -\frac{2\pi}{3a} \\ 0 \end{bmatrix} = \frac{2\pi}{\sqrt{3}a} \begin{bmatrix} 1 \\ -\frac{1}{\sqrt{3}} \\ 0 \end{bmatrix} \tag{5.92}$$

and obviously at their equivalents in the reciprocal space (as we can see from Fig. 5.5, which has been obtained by a nearest-neighbor tight-binding approach limited to the $2p_z$ atomic orbitals, with nonzero nearest-neighbor overlap integral).

Thus we can use the $\vec{k} \cdot \vec{p}$ method to find the dispersion relations of the graphene near these extrema points, following the approach of K. Ajiki and T. Ando [14] (however in our description we will continue to use the conventions of Ref. [28] and we will treat the behavior near $\vec{K}$ and $\vec{K}'$ separately).

We start using a simple tight-binding model, in which we consider, as atomic wave function for each of

the two atoms (A and B) of the graphene unit cell, only the $2p_z$ orbital, which is the orbital responsible for the π-bondings and thus for the above-mentioned two π-bands. In this case the generic eigenfunction in the solid can be expressed [29] as a linear combination (with coefficients C_j functions of $\vec{k}$) of the two functions obtained summing the $2p_z$ atomic wave functions φ centered on each j-like atom (with $j = A, B$) and weighted by a phase factor $e^{i\vec{k}\cdot\vec{R}_j}$ (note that $1/\sqrt{N_\Omega}$ is the normalization factor, with N_Ω the number of unit cells):

$$\begin{aligned}\psi(\vec{r}) &= \sum_{j=A,B} C_j(\vec{k}) \left(\frac{1}{\sqrt{N_\Omega}} \sum_{\vec{R}_j}^{N_\Omega} e^{i\vec{k}\cdot\vec{R}_j} \varphi(\vec{r}-\vec{R}_j) \right) = \\ &= \sum_{j=A,B} \sum_{\vec{R}_j} \left[C_j(\vec{k}) \frac{1}{\sqrt{N_\Omega}} e^{i\vec{k}\cdot\vec{R}_j} \right] \varphi(\vec{r}-\vec{R}_j) = \\ = \sum_{j=A,B} \sum_{\vec{R}_j} \psi_j(\vec{R}_j)\varphi(\vec{r}-\vec{R}_j) &= \sum_{\vec{R}_A} \psi_A(\vec{R}_A)\varphi(\vec{r}-\vec{R}_A) + \sum_{\vec{R}_B} \psi_B(\vec{R}_B)\varphi(\vec{r}-\vec{R}_B) \quad , \end{aligned} \tag{5.93}$$

where we have defined $\psi_j(\vec{R}_j)$ (with $j = A, B$) as the quantity between square brackets. Using the definition of the Hamiltonian operator

$$H|\psi\rangle = E|\psi\rangle \tag{5.94}$$

we have that

$$\langle\psi|H|\psi\rangle = E\langle\psi|\psi\rangle \tag{5.95}$$

and thus

$$\begin{aligned} E(\vec{k}) &= \frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle} = \\ &= \frac{\left\langle \sum_{j=A,B}\sum_{\vec{R}_j} \psi_j(\vec{R}_j)\varphi(\vec{r}-\vec{R}_j) \middle| H \middle| \sum_{j'=A,B}\sum_{\vec{R}_{j'}} \psi_{j'}(\vec{R}_{j'})\varphi(\vec{r}-\vec{R}_{j'}) \right\rangle}{\left\langle \sum_{j=A,B}\sum_{\vec{R}_j} \psi_j(\vec{R}_j)\varphi(\vec{r}-\vec{R}_j) \middle| \sum_{j'=A,B}\sum_{\vec{R}_{j'}} \psi_{j'}(\vec{R}_{j'})\varphi(\vec{r}-\vec{R}_{j'}) \right\rangle} = \\ &= \frac{\sum_{j,j'=A,B}\sum_{\vec{R}_j}\sum_{\vec{R}_{j'}} \psi_j(\vec{R}_j)\psi_{j'}(\vec{R}_{j'})\langle\varphi(\vec{r}-\vec{R}_j)|H|\varphi(\vec{r}-\vec{R}_{j'})\rangle}{\sum_{j,j'=A,B}\sum_{\vec{R}_j}\sum_{\vec{R}_{j'}} \psi_j(\vec{R}_j)\psi_{j'}(\vec{R}_{j'})\langle\varphi(\vec{r}-\vec{R}_j)|\varphi(\vec{r}-\vec{R}_{j'})\rangle} = \\ &= \frac{\sum_{j,j'=A,B}\sum_{\vec{R}_j}\sum_{\vec{R}_{j'}} \psi_j(\vec{R}_j)\psi_{j'}(\vec{R}_{j'})h_{jj'}}{\sum_{j,j'=A,B}\sum_{\vec{R}_j}\sum_{\vec{R}_{j'}} \psi_j(\vec{R}_j)\psi_{j'}(\vec{R}_{j'})s_{jj'}} \end{aligned} \tag{5.96}$$

where we have introduced the transfer integrals $h_{jj'}$ and the overlap integrals $s_{jj'}$ between atomic wave functions. When we set these values for a given $\vec{k}$ value, we can minimize $E(\vec{k})$ (to obtain the actual

physical state) enforcing (for both $j = A$ and $j = B$):

$$\frac{\partial E(\vec{k})}{\partial \psi_j(\vec{R}_j)} = \frac{\sum_{j'=A,B}\sum_{\vec{R}_{j'}} \psi_{j'}(\vec{R}_{j'})h_{jj'}}{\sum_{j,j'=A,B}\sum_{\vec{R}_j}\sum_{\vec{R}_{j'}} \psi_j(\vec{R}_j)\psi_{j'}(\vec{R}_{j'})s_{jj'}} -$$
$$- \frac{\sum_{j,j'=A,B}\sum_{\vec{R}_j}\sum_{\vec{R}_{j'}} \psi_j(\vec{R}_j)\psi_{j'}(\vec{R}_{j'})h_{jj'}}{\left(\sum_{j,j'=A,B}\sum_{\vec{R}_j}\sum_{\vec{R}_{j'}} \psi_j(\vec{R}_j)\psi_{j'}(\vec{R}_{j'})s_{jj'}\right)^2} \sum_{j'=A,B}\sum_{\vec{R}_{j'}} \psi_{j'}(\vec{R}_{j'})s_{jj'} = 0 \quad . \tag{5.97}$$

Multiplying both members by the appropriate quantity and rearranging, we find:

$$\sum_{j'=A,B}\sum_{\vec{R}_{j'}} \psi_{j'}(\vec{R}_{j'})h_{jj'} =$$
$$= \frac{\sum_{j,j'=A,B}\sum_{\vec{R}_j}\sum_{\vec{R}_{j'}} \psi_j(\vec{R}_j)\psi_{j'}(\vec{R}_{j'})h_{jj'}}{\sum_{j,j'=A,B}\sum_{\vec{R}_j}\sum_{\vec{R}_{j'}} \psi_j(\vec{R}_j)\psi_{j'}(\vec{R}_{j'})s_{jj'}} \sum_{j'=A,B}\sum_{\vec{R}_{j'}} \psi_{j'}(\vec{R}_{j'})s_{jj'} \tag{5.98}$$

and recognizing in the right-hand side the expression of $E(\vec{k})$:

$$\sum_{j'=A,B}\sum_{\vec{R}_{j'}} \psi_{j'}(\vec{R}_{j'})h_{jj'} = E(\vec{k}) \sum_{j'=A,B}\sum_{\vec{R}_{j'}} \psi_{j'}(\vec{R}_{j'})s_{jj'} \quad \text{with} \quad j = A, B. \tag{5.99}$$

Expanding what we have found for $j = A$ and $j = B$ we find:

$$\begin{cases} \sum_{\vec{R}_A} \psi_A(\vec{R}_A)h_{AA} + \sum_{\vec{R}_B} \psi_B(\vec{R}_B)h_{AB} = E\left(\sum_{\vec{R}_A} \psi_A(\vec{R}_A)s_{AA} + \sum_{\vec{R}_B} \psi_B(\vec{R}_B)s_{AB}\right) \\ \sum_{\vec{R}_A} \psi_A(\vec{R}_A)h_{BA} + \sum_{\vec{R}_B} \psi_B(\vec{R}_B)h_{BB} = E\left(\sum_{\vec{R}_A} \psi_A(\vec{R}_A)s_{BA} + \sum_{\vec{R}_B} \psi_B(\vec{R}_B)s_{BB}\right) \end{cases} . \tag{5.100}$$

If we consider non negligible only the integrals between each atom and itself and its nearest neighbors (which for an A atom are the nearest three B atoms, while for a B atom are the nearest three A atoms), we can reduce to:

$$\begin{cases} \psi_A(\vec{R}_A)h_{AA} + \sum_{l=1}^{3} \psi_B(\vec{R}_{B_l})h_{AB_l} = E\left(\psi_A(\vec{R}_A)s_{AA} + \sum_{l=1}^{3} \psi_B(\vec{R}_{B_l})s_{AB_l}\right) \\ \sum_{l=1}^{3} \psi_A(\vec{R}_{A_l})h_{BA_l} + \psi_B(\vec{R}_B)h_{BB} = E\left(\sum_{l=1}^{3} \psi_A(\vec{R}_{A_l})s_{BA_l} + \psi_B(\vec{R}_B)s_{BB}\right) \end{cases} . \tag{5.101}$$

In particular we consider:

$$h_{jj'} = \begin{cases} \epsilon_j = 0 & \text{if } j = j' \\ -\gamma_0 & \text{if } j \neq j' \end{cases} \quad ; \quad s_{jj'} = \delta_{jj'} \tag{5.102}$$

(where ϵ_j is the orbital energy of the $2p_z$ level, that we take as zero of the energy, and γ_0 is the modulus of the nearest neighbor transfer integral). Thus the tight-binding relations become:

$$\begin{cases} -\gamma_0 \sum_{l=1}^{3} \psi_B(\vec{R}_{B_l}) = E\,\psi_A(\vec{R}_A) \\ -\gamma_0 \sum_{l=1}^{3} \psi_A(\vec{R}_{A_l}) = E\,\psi_B(\vec{R}_B) \end{cases} . \tag{5.103}$$

If we introduce the vectors (Fig. 5.4(a)):

$$\vec{\tau}_1 = \frac{a}{\sqrt{3}} \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \quad , \quad \vec{\tau}_2 = \frac{a}{\sqrt{3}} \begin{bmatrix} \frac{1}{2} \\ -\frac{\sqrt{3}}{2} \\ 0 \end{bmatrix} \quad , \quad \vec{\tau}_3 = \frac{a}{\sqrt{3}} \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \\ 0 \end{bmatrix} \tag{5.104}$$

we can write the positions of the nearest-neighbor atoms in this way:

$$\begin{aligned} \vec{R}_{B_1} &= \vec{R}_A - \vec{\tau}_1 \\ \vec{R}_{B_2} &= \vec{R}_A - \vec{\tau}_2 \\ \vec{R}_{B_3} &= \vec{R}_A - \vec{\tau}_3 \\ \vec{R}_{A_1} &= \vec{R}_B + \vec{\tau}_1 \\ \vec{R}_{A_2} &= \vec{R}_B + \vec{\tau}_2 \\ \vec{R}_{A_3} &= \vec{R}_B + \vec{\tau}_3 \end{aligned} \tag{5.105}$$

and thus we can rewrite the tight-binding relations in the following form:

$$\begin{cases} -\gamma_0 \sum_{l=1}^{3} \psi_B(\vec{R}_A - \vec{\tau}_l) = E\,\psi_A(\vec{R}_A) \\ -\gamma_0 \sum_{l=1}^{3} \psi_A(\vec{R}_B + \vec{\tau}_l) = E\,\psi_B(\vec{R}_B) \end{cases} . \tag{5.106}$$

Now let us consider what happens near the points $\vec{K}$ and $\vec{K}'$. It is important to notice that all the results we will obtain are true also if we consider (instead of $\vec{K}$ and $\vec{K}'$) points of the reciprocal space equivalent to $\vec{K}$ and $\vec{K}'$, which are obtained adding to $\vec{K}$ and $\vec{K}'$ a linear combination with integer coefficients of the reciprocal space unit vectors $\vec{b}_1$ and $\vec{b}_2$.

Near the point $\vec{K}$ it is useful to rewrite the tight-binding expression of the wave function in the following

way:

$$\begin{aligned}
\psi(\vec{r}) &= \sum_{j=A,B} C_j(\vec{k}) \left[\frac{1}{\sqrt{N_\Omega}} \sum_{\vec{R}_j}^{N_\Omega} e^{i\vec{k}\cdot\vec{R}_j} \varphi(\vec{r}-\vec{R}_j) \right] = \\
&= \sum_{j=A,B} \sum_{\vec{R}_j} \left\{ \left[C_j(\vec{k}) \frac{1}{\sqrt{N_\Omega}} e^{i(\vec{k}-\vec{K})\cdot\vec{R}_j} \right] e^{i\vec{K}\cdot\vec{R}_j} \right\} \varphi(\vec{r}-\vec{R}_j) = \\
&= \sum_{j=A,B} \sum_{\vec{R}_j} \left\{ \left[C_j(\vec{k}) \frac{1}{\sqrt{N_\Omega}} e^{i\vec{\kappa}\cdot\vec{R}_j} \right] e^{i\vec{K}\cdot\vec{R}_j} \right\} \varphi(\vec{r}-\vec{R}_j) = \\
&= \sum_{j=A,B} \sum_{\vec{R}_j} \left\{ \tilde{F}_j^{\vec{K}}(\vec{R}_j) e^{i\vec{K}\cdot\vec{R}_j} \right\} \varphi(\vec{r}-\vec{R}_j) = \\
&= \sum_{\vec{R}_A} \left\{ \tilde{F}_A^{\vec{K}}(\vec{R}_A) e^{i\vec{K}\cdot\vec{R}_A} \right\} \varphi(\vec{r}-\vec{R}_A) + \sum_{\vec{R}_B} \left\{ \tilde{F}_B^{\vec{K}}(\vec{R}_B) e^{i\vec{K}\cdot\vec{R}_B} \right\} \varphi(\vec{r}-\vec{R}_B) = \\
&= \sum_{\vec{R}_A} \psi_A(\vec{R}_A)\varphi(\vec{r}-\vec{R}_A) + \sum_{\vec{R}_B} \psi_B(\vec{R}_B)\varphi(\vec{r}-\vec{R}_B) \quad .
\end{aligned} \tag{5.107}$$

We notice that $\tilde{F}_j^{\vec{K}}(\vec{r})$, defined as $C_j(\vec{k})(1/\sqrt{N_\Omega})e^{i\vec{\kappa}\cdot\vec{r}}$ (with $\vec{\kappa}=\vec{k}-\vec{K}$), for $\vec{k}$ near $\vec{K}$ (and thus for small $\vec{\kappa}$) is a slowly-varying function of $\vec{r}$. Incidentally, for $\vec{k}$ near $\vec{K}$ we can write that

$$\begin{aligned}
\psi(\vec{r}) &= \sum_{i=A,B} \sum_{\vec{R}_i} \tilde{F}_i^{\vec{K}}(\vec{R}_i)\, e^{i\vec{K}\cdot\vec{R}_i} \varphi(\vec{r}-\vec{R}_i) \simeq \sum_{i=A,B} \sum_{\vec{R}_i} \tilde{F}_i^{\vec{K}}(\vec{r})\, e^{i\vec{K}\cdot\vec{R}_i} \varphi(\vec{r}-\vec{R}_i) = \\
&= \sum_{i=A,B} \tilde{F}_i^{\vec{K}}(\vec{r}) e^{i\vec{K}\cdot\vec{r}} \left[\sum_{\vec{R}_i} \varphi(\vec{r}-\vec{R}_i)\, e^{-i\vec{K}\cdot(\vec{r}-\vec{R}_i)} \right] = \sum_{i=A,B} \tilde{F}_i^{\vec{K}}(\vec{r})\, e^{i\vec{K}\cdot\vec{r}}\, \tilde{u}_{\vec{K}}^i(\vec{r})
\end{aligned} \tag{5.108}$$

(where we have substituted $\tilde{F}_i^{\vec{K}}(\vec{r})$ to $\tilde{F}_i^{\vec{K}}(\vec{R}_i)$ using the fact that $\tilde{F}_i^{\vec{K}}$ is a slowly-varying function of $\vec{r}$, while φ is a function concentrated around the atom position). Since the quantity between square brackets (that we have called here $\tilde{u}_{\vec{K}}^i$) is periodic with the periodicity of the lattice and $\vec{K}$ is an extremum point (different from $\vec{0}$) of the dispersion relations, from the relation between $\psi(\vec{r})$, $\tilde{u}_{\vec{K}}^i(\vec{r})$ and $\tilde{F}_i^{\vec{K}}(\vec{r})$ we see that $\tilde{F}_i^{\vec{K}}$ can be seen as the electron envelope function corresponding to the extremum point $\vec{K}$.
To simplify the notation we can define the quantity $\omega = e^{i\frac{2\pi}{3}}$, which satisfies the relations:

$$\begin{aligned}
1+\omega+\omega^{-1} &= 0 \\
1+\omega &= -\omega^{-1} \\
1-\omega &= i\sqrt{3}\omega^{-1} \\
1-\omega^{-1} &= -i\sqrt{3}\omega \\
\omega-\omega^{-1} &= i\sqrt{3} \quad .
\end{aligned} \tag{5.109}$$

To obtain a simpler final result we can introduce the slightly modified functions:

$$\begin{cases} F_A^{\vec{K}}(\vec{R}_A) = \tilde{F}_A^{\vec{K}}(\vec{R}_A) \\ F_B^{\vec{K}}(\vec{R}_B) = -i\omega e^{-i\theta'}\tilde{F}_B^{\vec{K}}(\vec{R}_B) \end{cases} \tag{5.110}$$

(with θ' a properly chosen angle, as we will see). Consequently we can write that

$$\begin{aligned} \psi(\vec{r}) &= \sum_{\vec{R}_A}\left\{F_A^{\vec{K}}(\vec{R}_A)e^{i\vec{K}\cdot\vec{R}_A}\right\}\varphi(\vec{r}-\vec{R}_A) + \\ &+ \sum_{\vec{R}_B}\left\{\left(-\frac{\omega^{-1}}{i}e^{i\theta'}F_B^{\vec{K}}(\vec{R}_B)\right)e^{i\vec{K}\cdot\vec{R}_B}\right\}\varphi(\vec{r}-\vec{R}_B) = \\ &= \sum_{\vec{R}_A}\psi_A(\vec{R}_A)\varphi(\vec{r}-\vec{R}_A) + \sum_{\vec{R}_B}\psi_B(\vec{R}_B)\varphi(\vec{r}-\vec{R}_B) \end{aligned} \tag{5.111}$$

and thus that

$$\begin{cases} \psi_A(\vec{R}_A) = e^{i\vec{K}\cdot\vec{R}_A}F_A^{\vec{K}}(\vec{R}_A) \\ \psi_B(\vec{R}_B) = -\frac{\omega^{-1}}{i}e^{i\theta'}e^{i\vec{K}\cdot\vec{R}_B}F_B^{\vec{K}}(\vec{R}_B) \end{cases} . \tag{5.112}$$

If we introduce these equivalences into the tight-binding equations:

$$\begin{cases} E\,\psi_A(\vec{R}_A) = -\gamma_0\sum_l \psi_B(\vec{R}_A-\vec{\tau}_l) \\ E\,\psi_B(\vec{R}_B) = -\gamma_0\sum_l \psi_A(\vec{R}_B+\vec{\tau}_l) \end{cases} \tag{5.113}$$

we find

$$\begin{cases} E\left(e^{i\vec{K}\cdot\vec{R}_A}F_A^{\vec{K}}(\vec{R}_A)\right) = -\gamma_0\sum_l\left(-\frac{\omega^{-1}}{i}e^{i\theta'}e^{i\vec{K}\cdot(\vec{R}_A-\vec{\tau}_l)}F_B^{\vec{K}}(\vec{R}_A-\vec{\tau}_l)\right) \\ E\left(-\frac{\omega^{-1}}{i}e^{i\theta'}e^{i\vec{K}\cdot\vec{R}_B}F_B^{\vec{K}}(\vec{R}_B)\right) = -\gamma_0\sum_l\left(e^{i\vec{K}\cdot(\vec{R}_B+\vec{\tau}_l)}F_A^{\vec{K}}(\vec{R}_B+\vec{\tau}_l)\right) \end{cases} . \tag{5.114}$$

We can introduce a smoothing function $g(\vec{r})$ which varies smoothly in the range $|\vec{r}| \sim a$, decays rapidly for $|\vec{r}| \gg a$ and satisfies the conditions

$$\sum_{\vec{R}_A} g(\vec{r}-\vec{R}_A) = \sum_{\vec{R}_B} g(\vec{r}-\vec{R}_B) = 1 \tag{5.115}$$

and

$$\int d\vec{r}\, g(\vec{r}-\vec{R}_A) = \int d\vec{r}\, g(\vec{r}-\vec{R}_B) = \Omega_0 \quad , \tag{5.116}$$

where Ω_0 is the area of a graphene unit cell, equal to $\Omega_0 = \sqrt{3}a^2/2$. When this function is multiplied by a generic smooth function $f(\vec{r})$ (like the functions F before defined) clearly we have that

$$f(\vec{r})g(\vec{r}-\vec{R}) \simeq f(\vec{R})g(\vec{r}-\vec{R}) \quad . \tag{5.117}$$

If we multiply the first tight-binding equation by $g(\vec{r}-\vec{R}_A)e^{-i\vec{K}\cdot\vec{R}_A}$ and we sum it over $\vec{R}_A$ we find

$$E\sum_{\vec{R}_A} g(\vec{r}-\vec{R}_A)F_A^{\vec{K}}(\vec{R}_A) = -\gamma_0\sum_l\left(-\frac{\omega^{-1}}{i}e^{i\theta'}e^{-i\vec{K}\cdot\vec{\tau}_l}\right)\sum_{\vec{R}_A} g(\vec{r}-\vec{R}_A)F_B^{\vec{K}}(\vec{R}_A-\vec{\tau}_l) \tag{5.118}$$

and, using the last property of the function g,

$$E\left[\sum_{\vec{R}_A} g(\vec{r}-\vec{R}_A)\right]F_A^{\vec{K}}(\vec{r}) = \gamma_0\frac{\omega^{-1}}{i}e^{i\theta'}\sum_l e^{-i\vec{K}\cdot\vec{\tau}_l}\left[\sum_{\vec{R}_A} g(\vec{r}-\vec{R}_A)\right]F_B^{\vec{K}}(\vec{r}-\vec{\tau}_l) \quad . \tag{5.119}$$

Using the first property of the function g and then expanding the smooth quantity $F_B^{\vec{K}}(\vec{r}-\vec{\tau}_l)$ to the first order in $\vec{\tau}_l$ we can write:

$$\begin{aligned} E\,F_A^{\vec{K}}(\vec{r}) &= \gamma_0\frac{\omega^{-1}}{i}e^{i\theta'}\sum_l e^{-i\vec{K}\cdot\vec{\tau}_l}F_B^{\vec{K}}(\vec{r}-\vec{\tau}_l) \simeq \\ &\simeq \gamma_0\frac{\omega^{-1}}{i}e^{i\theta'}\sum_l e^{-i\vec{K}\cdot\vec{\tau}_l}\left[F_B^{\vec{K}}(\vec{r}) - \left(\vec{\tau}_l\cdot\frac{\partial}{\partial\vec{r}}\right)F_B^{\vec{K}}(\vec{r})\right] = \\ &= \gamma_0\frac{\omega^{-1}}{i}e^{i\theta'}\left\{\left(\sum_l e^{-i\vec{K}\cdot\vec{\tau}_l}\right)F_B^{\vec{K}}(\vec{r}) - \left[\sum_l e^{-i\vec{K}\cdot\vec{\tau}_l}\left(\vec{\tau}_l\cdot\frac{\partial}{\partial\vec{r}}\right)\right]F_B^{\vec{K}}(\vec{r})\right\} \quad . \end{aligned} \tag{5.120}$$

Let us now calculate the value of the sums which appear in the previous expression:

$$\begin{aligned} &\sum_l e^{-i\vec{K}\cdot\vec{\tau}_l} = e^{-i\left(-\frac{2\pi}{3}\right)} + e^{-i\frac{2\pi}{3}\left(\frac{1}{2}-\frac{1}{2}\right)} + e^{-i\frac{2\pi}{3}\left(\frac{1}{2}+\frac{1}{2}\right)} = \omega + 1 + \omega^{-1} = 0 \ ; \\ &\sum_l e^{-i\vec{K}\cdot\vec{\tau}_l}\left(\vec{\tau}_l\cdot\frac{\partial}{\partial\vec{r}}\right) = \\ &= \omega\frac{a}{\sqrt{3}}\left(-\frac{\partial}{\partial x'}\right) + 1\frac{a}{\sqrt{3}}\left(\frac{1}{2}\frac{\partial}{\partial x'} - \frac{\sqrt{3}}{2}\frac{\partial}{\partial y'}\right) + \omega^{-1}\frac{a}{\sqrt{3}}\left(\frac{1}{2}\frac{\partial}{\partial x'} + \frac{\sqrt{3}}{2}\frac{\partial}{\partial y'}\right) = \\ &= \frac{a}{\sqrt{3}}\left(\left(-\omega + \frac{1}{2} + \frac{1}{2}\omega^{-1}\right)\frac{\partial}{\partial x'} + \left(-\frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{2}\omega^{-1}\right)\frac{\partial}{\partial y'}\right) \quad . \end{aligned} \tag{5.121}$$

For the properties of the quantity ω we have that

$$-\omega + \frac{1}{2} + \frac{1}{2}\omega^{-1} = -\omega + \frac{1}{2}(1+\omega^{-1}) = -\omega - \frac{1}{2}\omega = -\frac{3}{2}\omega \tag{5.122}$$

and

$$-\frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{2}\omega^{-1} = -\frac{\sqrt{3}}{2}(1-\omega^{-1}) = i\frac{3}{2}\omega \quad , \tag{5.123}$$

thus

$$\sum_l e^{-i\vec{K}\cdot\vec{\tau}_l}\left(\vec{\tau}_l\cdot\frac{\partial}{\partial\vec{r}}\right) = -\frac{a}{\sqrt{3}}\frac{3}{2}\omega\left(\frac{\partial}{\partial x'} - i\frac{\partial}{\partial y'}\right) =$$
$$= -\frac{\sqrt{3}}{2}a\omega(i\kappa_{x'} + \kappa_{y'}) = -i\frac{\sqrt{3}}{2}a\omega(\kappa_{x'} - i\kappa_{y'}) \quad , \tag{5.124}$$

where we have used the fact that $F_j^{\vec{K}}(\vec{r}) \propto e^{i\vec{\kappa}\cdot\vec{r}}$ and therefore

$$-i\frac{\partial}{\partial x'}F_j^{\vec{K}}(\vec{r}) = \kappa_{x'}F_j^{\vec{K}}(\vec{r}) \qquad \text{and} \qquad -i\frac{\partial}{\partial y'}F_j^{\vec{K}}(\vec{r}) = \kappa_{y'}F_j^{\vec{K}}(\vec{r}) \quad . \tag{5.125}$$

Substituting such results, we find that

$$E\,F_A^{\vec{K}}(\vec{r}) = \gamma_0\frac{\omega^{-1}}{i}e^{i\theta'}i\frac{\sqrt{3}}{2}a\omega(\kappa_{x'} - i\kappa_{y'})F_B^{\vec{K}}(\vec{r}) =$$
$$= \frac{\sqrt{3}}{2}\gamma_0 a(\kappa_x - i\kappa_y)F_B^{\vec{K}}(\vec{r}) = \gamma(\kappa_x - i\kappa_y)F_B^{\vec{K}}(\vec{r}) \quad , \tag{5.126}$$

where we have passed from the original reference frame (x', y') to a new frame (x, y), rotated around the origin by an angle θ' with respect to the original one (Fig. 5.6) and we have used the fact that

$$e^{i\theta'}(\kappa_{x'} - i\kappa_{y'}) = (\cos\theta' + i\sin\theta')(\kappa_{x'} - i\kappa_{y'}) =$$
$$= (\cos\theta'\kappa_{x'} + \sin\theta'\kappa_{y'}) - i(\cos\theta'\kappa_{y'} - \sin\theta'\kappa_{x'}) = \kappa_x - i\kappa_y \tag{5.127}$$

(due to the relations between old and new coordinates). We have also defined the quantity $\gamma = (\sqrt{3}/2)\gamma_0 a$.

Then we can multiply by $g(\vec{r} - \vec{R}_B)(-i\,\omega\,e^{-i\theta'}e^{-i\vec{K}\cdot\vec{R}_B})$ the second tight-binding equation, sum it over $\vec{R}_B$ and proceed in an analogous way:

$$E\sum_{\vec{R}_B} g(\vec{r} - \vec{R}_B)F_B^{\vec{K}}(\vec{R}_B) = -\gamma_0\sum_l\left(-i\omega e^{-i\theta'}e^{i\vec{K}\cdot\vec{\tau}_l}\right)\sum_{\vec{R}_B} g(\vec{r} - \vec{R}_B)F_A^{\vec{K}}(\vec{R}_B + \vec{\tau}_l) \quad ;$$
$$E\left[\sum_{\vec{R}_B} g(\vec{r} - \vec{R}_B)\right]F_B^{\vec{K}}(\vec{r}) = \gamma_0 i\omega e^{-i\theta'}\sum_l e^{i\vec{K}\cdot\vec{\tau}_l}\left[\sum_{\vec{R}_B} g(\vec{r} - \vec{R}_B)\right]F_A^{\vec{K}}(\vec{r} + \vec{\tau}_l) \quad ;$$
$$EF_B^{\vec{K}}(\vec{r}) = \gamma_0 i\omega e^{-i\theta'}\sum_l e^{i\vec{K}\cdot\vec{\tau}_l}F_A^{\vec{K}}(\vec{r} + \vec{\tau}_l) \simeq$$
$$\simeq \gamma_0 i\omega e^{-i\theta'}\sum_l e^{i\vec{K}\cdot\vec{\tau}_l}\left[F_A^{\vec{K}}(\vec{r}) + \left(\vec{\tau}_l\cdot\frac{\partial}{\partial\vec{r}}\right)F_A^{\vec{K}}(\vec{r})\right] =$$
$$= \gamma_0 i\omega e^{-i\theta'}\left\{\left(\sum_l e^{i\vec{K}\cdot\vec{\tau}_l}\right)F_A^{\vec{K}}(\vec{r}) + \left[\sum_l e^{i\vec{K}\cdot\vec{\tau}_l}\left(\vec{\tau}_l\cdot\frac{\partial}{\partial\vec{r}}\right)\right]F_A^{\vec{K}}(\vec{r})\right\} \quad . \tag{5.128}$$

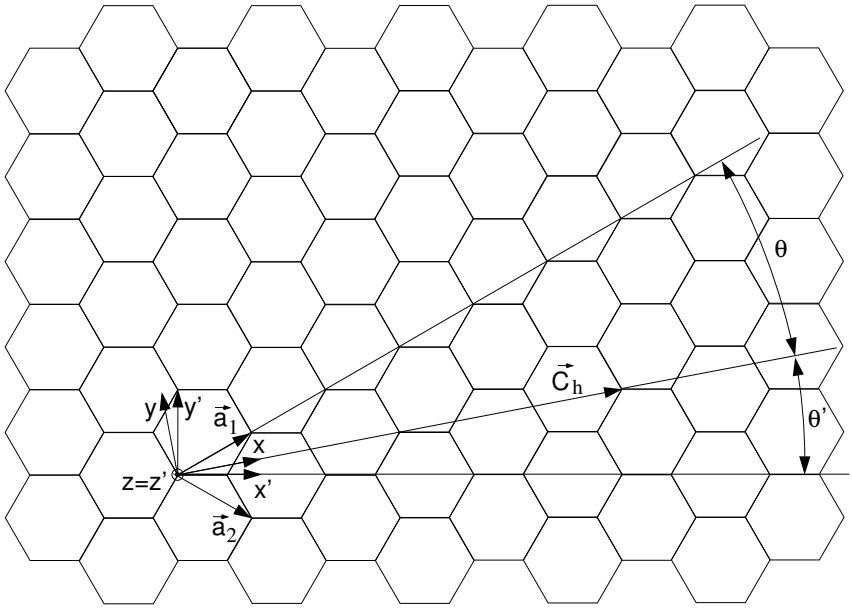


Figure 5.6: The reference frames used in the calculations.

In this case the values of the sums which appear in the equations are:

$$\sum_l e^{i\vec{K}\cdot\vec{\tau}_l} = e^{i\left(-\frac{2\pi}{3}\right)} + e^{i\frac{2\pi}{3}\left(\frac{1}{2}-\frac{1}{2}\right)} + e^{i\frac{2\pi}{3}\left(\frac{1}{2}+\frac{1}{2}\right)} = \omega^{-1} + 1 + \omega = 0 \quad ;$$

$$\sum_l e^{i\vec{K}\cdot\vec{\tau}_l}\left(\vec{\tau}_l\cdot\frac{\partial}{\partial\vec{r}}\right) =$$

$$= \omega^{-1}\frac{a}{\sqrt{3}}\left(-\frac{\partial}{\partial x'}\right) + 1\frac{a}{\sqrt{3}}\left(\frac{1}{2}\frac{\partial}{\partial x'} - \frac{\sqrt{3}}{2}\frac{\partial}{\partial y'}\right) + \omega\frac{a}{\sqrt{3}}\left(\frac{1}{2}\frac{\partial}{\partial x'} + \frac{\sqrt{3}}{2}\frac{\partial}{\partial y'}\right) =$$

$$= \frac{a}{\sqrt{3}}\left(\left(-\omega^{-1} + \frac{1}{2} + \frac{1}{2}\omega\right)\frac{\partial}{\partial x'} + \left(-\frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{2}\omega\right)\frac{\partial}{\partial y'}\right) \quad , \tag{5.129}$$

which, being

$$-\omega^{-1} + \frac{1}{2} + \frac{1}{2}\omega = -\omega^{-1} + \frac{1}{2}(1+\omega) = -\omega^{-1} - \frac{1}{2}\omega^{-1} = -\frac{3}{2}\omega^{-1} \tag{5.130}$$

and

$$-\frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{2}\omega = -\frac{\sqrt{3}}{2}(1-\omega) = -i\frac{3}{2}\omega^{-1} \quad , \tag{5.131}$$

becomes

$$\sum_l e^{i\vec{K}\cdot\vec{\tau}_l}\left(\vec{\tau}_l\cdot\frac{\partial}{\partial\vec{r}}\right) = -\frac{a}{\sqrt{3}}\frac{3}{2}\omega^{-1}\left(\frac{\partial}{\partial x'} + i\frac{\partial}{\partial y'}\right) =$$

$$= -\frac{\sqrt{3}}{2}a\omega^{-1}(i\kappa_{x'} - \kappa_{y'}) = -i\frac{\sqrt{3}}{2}a\omega^{-1}(\kappa_{x'} + i\kappa_{y'}) \quad . \tag{5.132}$$

Substituting these values, we find:

$$E\,F_B^{\vec{K}}(\vec{r}) = \gamma_0\omega e^{-i\theta'}\frac{\sqrt{3}}{2}a\omega^{-1}(\kappa_{x'}+i\kappa_{y'})F_A^{\vec{K}}(\vec{r}) =$$
$$= \frac{\sqrt{3}}{2}\gamma_0 a(\kappa_x+i\kappa_y)F_A^{\vec{K}}(\vec{r}) = \gamma(\kappa_x+i\kappa_y)F_A^{\vec{K}}(\vec{r}) \quad , \tag{5.133}$$

where we have made use of the relation

$$e^{-i\theta'}(\kappa_{x'}+i\kappa_{y'}) = (\cos\theta' - i\sin\theta')(\kappa_{x'}+i\kappa_{y'}) =$$
$$= (\cos\theta'\kappa_{x'}+\sin\theta'\kappa_{y'}) + i(\cos\theta'\kappa_{y'}-\sin\theta'\kappa_{x'}) = \kappa_x+i\kappa_y \quad . \tag{5.134}$$

In conclusion, we have found near $\vec{K}$ the equations:

$$\begin{cases} E\,F_A^{\vec{K}}(\vec{r}) = \gamma(\kappa_x - i\kappa_y)F_B^{\vec{K}}(\vec{r}) \\ E\,F_B^{\vec{K}}(\vec{r}) = \gamma(\kappa_x + i\kappa_y)F_A^{\vec{K}}(\vec{r}) \end{cases} , \tag{5.135}$$

which can also be written in matrix form as:

$$\begin{bmatrix} 0 & \gamma(\kappa_x - i\kappa_y) \\ \gamma(\kappa_x+i\kappa_y) & 0 \end{bmatrix}\begin{bmatrix} F_A^{\vec{K}}(\vec{r}) \\ F_B^{\vec{K}}(\vec{r}) \end{bmatrix} = E\begin{bmatrix} F_A^{\vec{K}}(\vec{r}) \\ F_B^{\vec{K}}(\vec{r}) \end{bmatrix} \tag{5.136}$$

or (using the Pauli spin matrices)

$$\gamma(\kappa_x\sigma_x+\kappa_y\sigma_y)\vec{F}^{\vec{K}}(\vec{r}) = \gamma(\vec{\kappa}\cdot\vec{\sigma})\vec{F}^{\vec{K}}(\vec{r}) = E\vec{F}^{\vec{K}}(\vec{r}) \quad . \tag{5.137}$$

In this way we have obtained the $\vec{k}\cdot\vec{p}$ Hamiltonian matrix which has as eigenvalues the dispersion relations of the two degenerate energy bands $E_s^{\vec{K}}(\vec{\kappa})$ and as eigenvectors the correspondent electron envelope functions $F_{s\vec{\kappa}}^{\vec{K}}(\vec{r})$.
In particular if we enforce

$$\det\left\{\begin{bmatrix} 0 & \gamma(\kappa_x - i\kappa_y) \\ \gamma(\kappa_x+i\kappa_y) & 0 \end{bmatrix} - E\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right\} = 0 \tag{5.138}$$

we find the dispersion relations

$$E_s^{\vec{K}}(\vec{\kappa}) = s\gamma|\vec{\kappa}| \quad , \tag{5.139}$$

where s can assume the values $+1$ or -1.
If we define the angle α in such a way that

$$\kappa_x + i\kappa_y = |\vec{\kappa}|e^{i(\frac{\pi}{2}+\alpha)} \tag{5.140}$$

and thus

$$\kappa_x - i\kappa_y = (\kappa_x + i\kappa_y)^* = |\vec{\kappa}|e^{i(-\frac{\pi}{2}-\alpha)} \tag{5.141}$$

we have that the correspondent envelope functions are

$$\vec{F}^{\vec{K}}_{s\vec{\kappa}}(\vec{r}) = \frac{1}{\sqrt{\Sigma}}e^{i\vec{\kappa}\cdot\vec{r}}F^{\vec{K}}_{s\vec{\kappa}} \tag{5.142}$$

with

$$\vec{F}^{\vec{K}}_{s\vec{\kappa}} = e^{i\phi_s(\vec{\kappa})}R(-\alpha(\vec{\kappa}))|s\rangle \tag{5.143}$$

and

$$|s\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} -is \\ 1 \end{bmatrix} \quad , \tag{5.144}$$

where Σ is the considered surface area, $\phi_s(\vec{\kappa})$ is an arbitrary phase factor and $R(\alpha)$ is given by

$$R(\alpha) = \begin{bmatrix} e^{i\frac{\alpha}{2}} & 0 \\ 0 & e^{-i\frac{\alpha}{2}} \end{bmatrix} \quad . \tag{5.145}$$

This can be easily verified noting that

$$\begin{aligned}
&\gamma\begin{bmatrix} 0 & \kappa_x - i\kappa_y \\ \kappa_x + i\kappa_y & 0 \end{bmatrix}\vec{F}^{\vec{K}}_{s\vec{\kappa}}(\vec{r}) = \gamma\begin{bmatrix} 0 & |\vec{\kappa}|e^{i(-\frac{\pi}{2}-\alpha)} \\ |\vec{\kappa}|e^{i(\frac{\pi}{2}+\alpha)} & 0 \end{bmatrix}\vec{F}^{\vec{K}}_{s\vec{\kappa}}(\vec{r}) = \\
&= \gamma\begin{bmatrix} 0 & -i|\vec{\kappa}|e^{-i\alpha} \\ i|\vec{\kappa}|e^{i\alpha} & 0 \end{bmatrix}\left(\frac{1}{\sqrt{\Sigma}}e^{i\vec{\kappa}\cdot\vec{r}}e^{i\phi_s}\begin{bmatrix} e^{-i\frac{\alpha}{2}} & 0 \\ 0 & e^{i\frac{\alpha}{2}} \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} -is \\ 1 \end{bmatrix}\right) = \\
&= \frac{1}{\sqrt{2\Sigma}}\gamma e^{i\vec{\kappa}\cdot\vec{r}}e^{i\phi_s}\begin{bmatrix} 0 & -i|\vec{\kappa}|e^{-i\frac{\alpha}{2}} \\ i|\vec{\kappa}|e^{i\frac{\alpha}{2}} & 0 \end{bmatrix}\begin{bmatrix} -is \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2\Sigma}}\gamma e^{i(\vec{\kappa}\cdot\vec{r}+\phi_s)}\begin{bmatrix} -i|\vec{\kappa}|e^{-i\frac{\alpha}{2}} \\ |\vec{\kappa}|se^{i\frac{\alpha}{2}} \end{bmatrix}
\end{aligned} \tag{5.146}$$

and also

$$\begin{aligned}
&E^{\vec{K}}_s\vec{F}^{\vec{K}}_{s\vec{\kappa}}(\vec{r}) = s\gamma|\vec{\kappa}|\left(\frac{1}{\sqrt{\Sigma}}e^{i\vec{\kappa}\cdot\vec{r}}e^{i\phi_s}\begin{bmatrix} e^{-i\frac{\alpha}{2}} & 0 \\ 0 & e^{i\frac{\alpha}{2}} \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} -is \\ 1 \end{bmatrix}\right) = \\
&= s\gamma|\vec{\kappa}|\frac{1}{\sqrt{2\Sigma}}e^{i(\vec{\kappa}\cdot\vec{r}+\phi_s)}\begin{bmatrix} -ise^{-i\frac{\alpha}{2}} \\ e^{i\frac{\alpha}{2}} \end{bmatrix} = \frac{1}{\sqrt{2\Sigma}}\gamma e^{i(\vec{\kappa}\cdot\vec{r}+\phi_s)}\begin{bmatrix} -is^2|\vec{\kappa}|e^{-i\frac{\alpha}{2}} \\ |\vec{\kappa}|se^{i\frac{\alpha}{2}} \end{bmatrix} = \\
&= \frac{1}{\sqrt{2\Sigma}}\gamma e^{i(\vec{\kappa}\cdot\vec{r}+\phi_s)}\begin{bmatrix} -i|\vec{\kappa}|e^{-i\frac{\alpha}{2}} \\ |\vec{\kappa}|se^{i\frac{\alpha}{2}} \end{bmatrix}
\end{aligned} \tag{5.147}$$

(where we have used the fact that $s^2 = (\pm 1)^2 = 1$). From these functions $F^{\vec{K}}_A$ and $F^{\vec{K}}_B$ then we can find the functions ψ_A and ψ_B and thus the electron wave function ψ, using the previously written relations. Now we can perform similar calculations near the point $\vec{K}'$. In particular we can rewrite the tight-binding

expression of the wave function in this way:

$$\begin{aligned}
\psi(\vec{r}) &= \sum_{j=A,B} C_j(\vec{k}) \left[\frac{1}{\sqrt{N_\Omega}} \sum_{\vec{R}_j}^{N_\Omega} e^{i\vec{k}\cdot\vec{R}_j} \varphi(\vec{r}-\vec{R}_j) \right] = \\
&= \sum_{j=A,B} \sum_{\vec{R}_j} \left\{ \left[C_j(\vec{k}) \frac{1}{\sqrt{N_\Omega}} e^{i(\vec{k}-\vec{K}')\cdot\vec{R}_j} \right] e^{i\vec{K}'\cdot\vec{R}_j} \right\} \varphi(\vec{r}-\vec{R}_j) = \\
&= \sum_{j=A,B} \sum_{\vec{R}_j} \left\{ \left[C_j(\vec{k}) \frac{1}{\sqrt{N_\Omega}} e^{i\vec{\kappa}\cdot\vec{R}_j} \right] e^{i\vec{K}'\cdot\vec{R}_j} \right\} \varphi(\vec{r}-\vec{R}_j) = \\
&= \sum_{j=A,B} \sum_{\vec{R}_j} \left\{ \tilde{F}_j^{\vec{K}'}(\vec{R}_j) e^{i\vec{K}'\cdot\vec{R}_j} \right\} \varphi(\vec{r}-\vec{R}_j) = \\
&= \sum_{\vec{R}_A} \left\{ \tilde{F}_A^{\vec{K}'}(\vec{R}_A) e^{i\vec{K}'\cdot\vec{R}_A} \right\} \varphi(\vec{r}-\vec{R}_A) + \sum_{\vec{R}_B} \left\{ \tilde{F}_B^{\vec{K}'}(\vec{R}_B) e^{i\vec{K}'\cdot\vec{R}_B} \right\} \varphi(\vec{r}-\vec{R}_B) = \\
&= \sum_{\vec{R}_A} \psi_A(\vec{R}_A)\varphi(\vec{r}-\vec{R}_A) + \sum_{\vec{R}_B} \psi_B(\vec{R}_B)\varphi(\vec{r}-\vec{R}_B)
\end{aligned} \tag{5.148}$$

where we notice that now $\kappa = \vec{k} - \vec{K}'$.
Then we introduce the functions:

$$\begin{cases} F_A^{\vec{K}'}(\vec{R}_A) = i\omega^{-1} e^{-i\theta'} \tilde{F}_A^{\vec{K}'}(\vec{R}_A) \\ F_B^{\vec{K}'}(\vec{R}_B) = \tilde{F}_B^{\vec{K}'}(\vec{R}_B) \end{cases}, \tag{5.149}$$

so that we can write that

$$\begin{aligned}
\psi(\vec{r}) &= \sum_{\vec{R}_A} \left\{ \left(\frac{\omega}{i} e^{i\theta'} \right) F_A^{\vec{K}'}(\vec{R}_A) e^{i\vec{K}'\cdot\vec{R}_A} \right\} \varphi(\vec{r}-\vec{R}_A) + \\
&\quad + \sum_{\vec{R}_B} \left\{ F_B^{\vec{K}'}(\vec{R}_B) e^{i\vec{K}'\cdot\vec{R}_B} \right\} \varphi(\vec{r}-\vec{R}_B) = \\
&= \sum_{\vec{R}_A} \psi_A(\vec{R}_A)\varphi(\vec{r}-\vec{R}_A) + \sum_{\vec{R}_B} \psi_B(\vec{R}_B)\varphi(\vec{r}-\vec{R}_B)
\end{aligned} \tag{5.150}$$

and thus that

$$\begin{cases} \psi_A(\vec{R}_A) = \dfrac{\omega}{i} e^{i\theta'} e^{i\vec{K}'\cdot\vec{R}_A} F_A^{\vec{K}'}(\vec{R}_A) \\ \psi_B(\vec{R}_B) = e^{i\vec{K}'\cdot\vec{R}_B} F_B^{\vec{K}'}(\vec{R}_B) \end{cases}. \tag{5.151}$$

Introducing these expressions into the tight-binding equations:

$$\begin{cases} E\,\psi_A(\vec{R}_A) = -\gamma_0 \sum_l \psi_B(\vec{R}_A - \vec{\tau}_l) \\ E\,\psi_B(\vec{R}_B) = -\gamma_0 \sum_l \psi_A(\vec{R}_B + \vec{\tau}_l) \end{cases} \tag{5.152}$$

we find

$$\begin{cases} E\left(\frac{\omega}{i}e^{i\theta'}e^{i\vec{K}'\cdot\vec{R}_A}F_A^{\vec{K}'}(\vec{R}_A)\right)=-\gamma_0\sum_l\left(e^{i\vec{K}'\cdot(\vec{R}_A-\vec{\tau}_l)}F_B^{\vec{K}'}(\vec{R}_A-\vec{\tau}_l)\right) \\ E\left(e^{i\vec{K}'\cdot\vec{R}_B}F_B^{\vec{K}'}(\vec{R}_B)\right)=-\gamma_0\sum_l\left(\frac{\omega}{i}e^{i\theta'}e^{i\vec{K}'\cdot(\vec{R}_B+\vec{\tau}_l)}F_A^{\vec{K}'}(\vec{R}_B+\vec{\tau}_l)\right) \end{cases} . \tag{5.153}$$

If we multiply the first equation by $g(\vec{r}-\vec{R}_A)\left(i\omega^{-1}e^{-i\theta'}e^{-i\vec{K}'\cdot\vec{R}_A}\right)$ and we sum it over $\vec{R}_A$ we obtain (proceeding in a similar way to what we have done near the point $\vec{K}$):

$$\begin{aligned} E\sum_{\vec{R}_A}g(\vec{r}-\vec{R}_A)F_A^{\vec{K}'}(\vec{R}_A)&=-\gamma_0\sum_l\left(i\omega^{-1}e^{-i\theta'}e^{-i\vec{K}'\cdot\vec{\tau}_l}\right)\sum_{\vec{R}_A}g(\vec{r}-\vec{R}_A)F_B^{\vec{K}'}(\vec{R}_A-\vec{\tau}_l)\ ; \\ E\left[\sum_{\vec{R}_A}g(\vec{r}-\vec{R}_A)\right]F_A^{\vec{K}'}(\vec{r})&=-\gamma_0 i\omega^{-1}e^{-i\theta'}\sum_l e^{-i\vec{K}'\cdot\vec{\tau}_l}\left[\sum_{\vec{R}_A}g(\vec{r}-\vec{R}_A)\right]F_B^{\vec{K}'}(\vec{r}-\vec{\tau}_l)\ ; \\ E\,F_A^{\vec{K}'}(\vec{r})&=-\gamma_0 i\omega^{-1}e^{-i\theta'}\sum_l e^{-i\vec{K}'\cdot\vec{\tau}_l}F_B^{\vec{K}'}(\vec{r}-\vec{\tau}_l)\simeq \\ &\simeq-\gamma_0 i\omega^{-1}e^{-i\theta'}\sum_l e^{-i\vec{K}'\cdot\vec{\tau}_l}\left[F_B^{\vec{K}'}(\vec{r})-\left(\vec{\tau}_l\cdot\frac{\partial}{\partial\vec{r}}\right)F_B^{\vec{K}'}(\vec{r})\right]= \\ =-\gamma_0 i\omega^{-1}e^{-i\theta'}&\left\{\left(\sum_l e^{-i\vec{K}'\cdot\vec{\tau}_l}\right)F_B^{\vec{K}'}(\vec{r})-\left[\sum_l e^{-i\vec{K}'\cdot\vec{\tau}_l}\left(\vec{\tau}_l\cdot\frac{\partial}{\partial\vec{r}}\right)\right]F_B^{\vec{K}'}(\vec{r})\right\}\ . \end{aligned} \tag{5.154}$$

The values of the sums which appear in the previous expression are:

$$\begin{aligned} \sum_l e^{-i\vec{K}'\cdot\vec{\tau}_l}&=e^{-i\left(-\frac{2\pi}{3}\right)}+e^{-i\frac{2\pi}{3}\left(\frac{1}{2}+\frac{1}{2}\right)}+e^{-i\frac{2\pi}{3}\left(\frac{1}{2}-\frac{1}{2}\right)}=\omega+\omega^{-1}+1=0\quad ; \\ &\sum_l e^{-i\vec{K}'\cdot\vec{\tau}_l}\left(\vec{\tau}_l\cdot\frac{\partial}{\partial\vec{r}}\right)= \\ =\omega\frac{a}{\sqrt{3}}\left(-\frac{\partial}{\partial x'}\right)+\omega^{-1}\frac{a}{\sqrt{3}}&\left(\frac{1}{2}\frac{\partial}{\partial x'}-\frac{\sqrt{3}}{2}\frac{\partial}{\partial y'}\right)+1\frac{a}{\sqrt{3}}\left(\frac{1}{2}\frac{\partial}{\partial x'}+\frac{\sqrt{3}}{2}\frac{\partial}{\partial y'}\right)= \\ &=\frac{a}{\sqrt{3}}\left(\left(-\omega+\frac{1}{2}\omega^{-1}+\frac{1}{2}\right)\frac{\partial}{\partial x'}+\left(-\frac{\sqrt{3}}{2}\omega^{-1}+\frac{\sqrt{3}}{2}\right)\frac{\partial}{\partial y'}\right)\quad . \end{aligned} \tag{5.155}$$

In particular, being

$$-\omega+\frac{1}{2}\omega^{-1}+\frac{1}{2}=-\omega+\frac{1}{2}(1+\omega^{-1})=-\omega-\frac{1}{2}\omega=-\frac{3}{2}\omega \tag{5.156}$$

and

$$-\frac{\sqrt{3}}{2}\omega^{-1}+\frac{\sqrt{3}}{2}=\frac{\sqrt{3}}{2}(1-\omega^{-1})=-i\frac{3}{2}\omega \tag{5.157}$$

we have that

$$\sum_l e^{-i\vec{K}'\cdot\vec{\tau}_l}\left(\vec{\tau}_l\cdot\frac{\partial}{\partial\vec{r}}\right) = -\frac{a}{\sqrt{3}}\frac{3}{2}\omega\left(\frac{\partial}{\partial x'}+i\frac{\partial}{\partial y'}\right) =$$
$$= -\frac{\sqrt{3}}{2}a\omega(i\kappa_{x'}-\kappa_{y'}) = -i\frac{\sqrt{3}}{2}a\omega(\kappa_{x'}+i\kappa_{y'}) \quad , \tag{5.158}$$

where we have used the fact that $F_j^{\vec{K}'}(\vec{r}) \propto e^{i\vec{\kappa}\cdot\vec{r}}$ and thus

$$-i\frac{\partial}{\partial x'}F_j^{\vec{K}'}(\vec{r}) = \kappa_{x'}F_j^{\vec{K}'}(\vec{r}) \quad \text{and} \quad -i\frac{\partial}{\partial y'}F_j^{\vec{K}'}(\vec{r}) = \kappa_{y'}F_j^{\vec{K}'}(\vec{r}) \quad . \tag{5.159}$$

Substituting such results, we find that

$$E\,F_A^{\vec{K}'}(\vec{r}) = \gamma_0\omega^{-1}e^{-i\theta'}\frac{\sqrt{3}}{2}a\omega(\kappa_{x'}+i\kappa_{y'})F_B^{\vec{K}'}(\vec{r}) =$$
$$= \frac{\sqrt{3}}{2}\gamma_0 a(\kappa_x+i\kappa_y)F_B^{\vec{K}'}(\vec{r}) = \gamma(\kappa_x+i\kappa_y)F_B^{\vec{K}'}(\vec{r}) \quad , \tag{5.160}$$

where we have made the same change of reference frame and we have used the fact that (as we have already seen) $e^{-i\theta'}(\kappa_{x'}+i\kappa_{y'}) = \kappa_x+i\kappa_y$.
If we multiply the second tight-binding equation by $g(\vec{r}-\vec{R}_B)e^{-i\vec{K}'\cdot\vec{R}_B}$, we sum over $\vec{R}_B$ and we proceed analogously, we can rewrite it in the following way:

$$E\sum_{\vec{R}_B} g(\vec{r}-\vec{R}_B)F_B^{\vec{K}'}(\vec{R}_B) = -\gamma_0\sum_l\left(\frac{\omega}{i}e^{i\theta'}e^{i\vec{K}'\cdot\vec{\tau}_l}\right)\sum_{\vec{R}_B} g(\vec{r}-\vec{R}_B)F_A^{\vec{K}'}(\vec{R}_B+\vec{\tau}_l) \quad ;$$
$$E\left[\sum_{\vec{R}_B} g(\vec{r}-\vec{R}_B)\right]F_B^{\vec{K}'}(\vec{r}) = -\gamma_0\frac{\omega}{i}e^{i\theta'}\sum_l e^{i\vec{K}'\cdot\vec{\tau}_l}\left[\sum_{\vec{R}_B} g(\vec{r}-\vec{R}_B)\right]F_A^{\vec{K}'}(\vec{r}+\vec{\tau}_l) \quad ;$$
$$EF_B^{\vec{K}'}(\vec{r}) = -\gamma_0\frac{\omega}{i}e^{i\theta'}\sum_l e^{i\vec{K}'\cdot\vec{\tau}_l}F_A^{\vec{K}'}(\vec{r}+\vec{\tau}_l) \simeq$$
$$\simeq -\gamma_0\frac{\omega}{i}e^{i\theta'}\sum_l e^{i\vec{K}'\cdot\vec{\tau}_l}\left[F_A^{\vec{K}'}(\vec{r})+\left(\vec{\tau}_l\cdot\frac{\partial}{\partial\vec{r}}\right)F_A^{\vec{K}'}(\vec{r})\right] =$$
$$= -\gamma_0\frac{\omega}{i}e^{i\theta'}\left\{\left(\sum_l e^{i\vec{K}'\cdot\vec{\tau}_l}\right)F_A^{\vec{K}'}(\vec{r})+\left[\sum_l e^{i\vec{K}'\cdot\vec{\tau}_l}\left(\vec{\tau}_l\cdot\frac{\partial}{\partial\vec{r}}\right)\right]F_A^{\vec{K}'}(\vec{r})\right\} \quad . \tag{5.161}$$

In this case the values of the sums are:

$$\sum_l e^{i\vec{K}'\cdot\vec{\tau}_l} = e^{i\left(-\frac{2\pi}{3}\right)} + e^{i\frac{2\pi}{3}\left(\frac{1}{2}+\frac{1}{2}\right)} + e^{i\frac{2\pi}{3}\left(\frac{1}{2}-\frac{1}{2}\right)} = \omega^{-1} + \omega + 1 = 0 \quad ;$$

$$\sum_l e^{i\vec{K}'\cdot\vec{\tau}_l}\left(\vec{\tau}_l\cdot\frac{\partial}{\partial\vec{r}}\right) =$$

$$= \omega^{-1}\frac{a}{\sqrt{3}}\left(-\frac{\partial}{\partial x'}\right) + \omega\frac{a}{\sqrt{3}}\left(\frac{1}{2}\frac{\partial}{\partial x'} - \frac{\sqrt{3}}{2}\frac{\partial}{\partial y'}\right) + 1\frac{a}{\sqrt{3}}\left(\frac{1}{2}\frac{\partial}{\partial x'} + \frac{\sqrt{3}}{2}\frac{\partial}{\partial y'}\right) =$$

$$= \frac{a}{\sqrt{3}}\left(\left(-\omega^{-1} + \frac{1}{2}\omega + \frac{1}{2}\right)\frac{\partial}{\partial x'} + \left(-\frac{\sqrt{3}}{2}\omega + \frac{\sqrt{3}}{2}\right)\frac{\partial}{\partial y'}\right) \quad , \tag{5.162}$$

which, being

$$-\omega^{-1} + \frac{1}{2}\omega + \frac{1}{2} = -\omega^{-1} + \frac{1}{2}(1+\omega) = -\omega^{-1} - \frac{1}{2}\omega^{-1} = -\frac{3}{2}\omega^{-1} \tag{5.163}$$

and

$$-\frac{\sqrt{3}}{2}\omega + \frac{\sqrt{3}}{2} = \frac{\sqrt{3}}{2}(1-\omega) = i\frac{3}{2}\omega^{-1} \quad , \tag{5.164}$$

becomes

$$\sum_l e^{i\vec{K}'\cdot\vec{\tau}_l}\left(\vec{\tau}_l\cdot\frac{\partial}{\partial\vec{r}}\right) = -\frac{a}{\sqrt{3}}\frac{3}{2}\omega^{-1}\left(\frac{\partial}{\partial x'} - i\frac{\partial}{\partial y'}\right) =$$

$$= -\frac{\sqrt{3}}{2}a\omega^{-1}(i\kappa_{x'} + \kappa_{y'}) = -i\frac{\sqrt{3}}{2}a\omega^{-1}(\kappa_{x'} - i\kappa_{y'}) \quad . \tag{5.165}$$

Substituting the found values, we find:

$$E\,F_B^{\vec{K}'}(\vec{r}) = \gamma_0\frac{\omega}{i}e^{i\theta'}i\frac{\sqrt{3}}{2}a\omega^{-1}(\kappa_{x'} - i\kappa_{y'})F_A^{\vec{K}'}(\vec{r}) =$$

$$= \frac{\sqrt{3}}{2}\gamma_0 a(\kappa_x - i\kappa_y)F_A^{\vec{K}'}(\vec{r}) = \gamma(\kappa_x - i\kappa_y)F_A^{\vec{K}'}(\vec{r}) \quad , \tag{5.166}$$

where we used the previously found relation $e^{i\theta'}(\kappa_{x'} - i\kappa_{y'}) = \kappa_x - i\kappa_y$.
Therefore the equations we have found near K' are:

$$\begin{cases} E\,F_A^{\vec{K}'}(\vec{r}) = \gamma(\kappa_x + i\kappa_y)F_B^{\vec{K}'}(\vec{r}) \\ E\,F_B^{\vec{K}'}(\vec{r}) = \gamma(\kappa_x - i\kappa_y)F_A^{\vec{K}'}(\vec{r}) \end{cases} \tag{5.167}$$

which can be written in matrix form:

$$\begin{bmatrix} 0 & \gamma(\kappa_x + i\kappa_y) \\ \gamma(\kappa_x - i\kappa_y) & 0 \end{bmatrix}\begin{bmatrix} F_A^{\vec{K}'}(\vec{r}) \\ F_B^{\vec{K}'}(\vec{r}) \end{bmatrix} = E\begin{bmatrix} F_A^{\vec{K}'}(\vec{r}) \\ F_B^{\vec{K}'}(\vec{r}) \end{bmatrix} \tag{5.168}$$

or equivalently (using the Pauli spin matrices):

$$\gamma(\kappa_x\sigma_x-\kappa_y\sigma_y)\vec{F}^{\vec{K}'}(\vec{r})=\gamma\left(\begin{bmatrix}\kappa_x\\-\kappa_y\\0\end{bmatrix}\cdot\vec{\sigma}\right)\vec{F}^{\vec{K}'}(\vec{r})=E\vec{F}^{\vec{K}'}(\vec{r})\quad . \tag{5.169}$$

If we enforce

$$\det\left\{\begin{bmatrix}0&\gamma(\kappa_x+i\kappa_y)\\\gamma(\kappa_x-i\kappa_y)&0\end{bmatrix}-E\begin{bmatrix}1&0\\0&1\end{bmatrix}\right\}=0 \tag{5.170}$$

we find the dispersion relations

$$E_s^{\vec{K}'}(\vec{\kappa})=s\gamma|\vec{\kappa}|\quad , \tag{5.171}$$

where s can assume the values $+1$ or -1.
The correspondent envelope functions are

$$\vec{F}^{\vec{K}'}_{s\vec{\kappa}}(\vec{r})=\frac{1}{\sqrt{\Sigma}}e^{i\vec{\kappa}\cdot\vec{r}}\vec{F}^{\vec{K}'}_{s\vec{\kappa}} \tag{5.172}$$

with

$$\vec{F}^{\vec{K}'}_{s\vec{\kappa}}=e^{i\phi_s(\vec{\kappa})}R(\alpha(\vec{\kappa}))|s\rangle \tag{5.173}$$

($\phi_s(\vec{\kappa})$ is an arbitrary phase factor) and

$$|s\rangle=\frac{1}{\sqrt{2}}\begin{bmatrix}is\\1\end{bmatrix}\quad . \tag{5.174}$$

This result is easily verified noting that

$$\begin{aligned}\gamma\begin{bmatrix}0&\kappa_x+i\kappa_y\\\kappa_x-i\kappa_y&0\end{bmatrix}\vec{F}^{\vec{K}'}_{s\vec{\kappa}}(\vec{r})&=\gamma\begin{bmatrix}0&|\vec{\kappa}|e^{i(\frac{\pi}{2}+\alpha)}\\|\vec{\kappa}|e^{i(-\frac{\pi}{2}-\alpha)}&0\end{bmatrix}\vec{F}^{\vec{K}'}_{s\vec{\kappa}}(\vec{r})=\\&=\gamma\begin{bmatrix}0&i|\vec{\kappa}|e^{i\alpha}\\-i|\vec{\kappa}|e^{-i\alpha}&0\end{bmatrix}\left(\frac{1}{\sqrt{\Sigma}}e^{i\vec{\kappa}\cdot\vec{r}}e^{i\phi_s}\begin{bmatrix}e^{i\frac{\alpha}{2}}&0\\0&e^{-i\frac{\alpha}{2}}\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}is\\1\end{bmatrix}\right)=\\&=\frac{1}{\sqrt{2\Sigma}}\gamma e^{i\vec{\kappa}\cdot\vec{r}}e^{i\phi_s}\begin{bmatrix}0&i|\vec{\kappa}|e^{i\frac{\alpha}{2}}\\-i|\vec{\kappa}|e^{-i\frac{\alpha}{2}}&0\end{bmatrix}\begin{bmatrix}is\\1\end{bmatrix}=\frac{1}{\sqrt{2\Sigma}}\gamma e^{i(\vec{\kappa}\cdot\vec{r}+\phi_s)}\begin{bmatrix}i|\vec{\kappa}|e^{i\frac{\alpha}{2}}\\|\vec{\kappa}|se^{-i\frac{\alpha}{2}}\end{bmatrix}\end{aligned} \tag{5.175}$$

and that also

$$E_s^{\vec{K}'}\vec{F}_{s\vec{\kappa}}^{\vec{K}'}(\vec{r}) = s\gamma|\vec{\kappa}|\left(\frac{1}{\sqrt{\Sigma}}e^{i\vec{\kappa}\cdot\vec{r}}e^{i\phi_s}\begin{bmatrix} e^{i\frac{\alpha}{2}} & 0 \\ 0 & e^{-i\frac{\alpha}{2}} \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} is \\ 1 \end{bmatrix}\right) =$$

$$= s\gamma|\vec{\kappa}|\frac{1}{\sqrt{2\Sigma}}e^{i(\vec{\kappa}\cdot\vec{r}+\phi_s)}\begin{bmatrix} ise^{i\frac{\alpha}{2}} \\ e^{-i\frac{\alpha}{2}} \end{bmatrix} = \frac{1}{\sqrt{2\Sigma}}\gamma e^{i(\vec{\kappa}\cdot\vec{r}+\phi_s)}\begin{bmatrix} is^2|\vec{\kappa}|e^{i\frac{\alpha}{2}} \\ |\vec{\kappa}|se^{-i\frac{\alpha}{2}} \end{bmatrix} =$$

$$= \frac{1}{\sqrt{2\Sigma}}\gamma e^{i(\vec{\kappa}\cdot\vec{r}+\phi_s)}\begin{bmatrix} i|\vec{\kappa}|e^{i\frac{\alpha}{2}} \\ |\vec{\kappa}|se^{-i\frac{\alpha}{2}} \end{bmatrix} \quad . \tag{5.176}$$

We notice that the energy dispersion relations that we have found in this way near $\vec{K}$ and $\vec{K}'$ are identical to those that one can obtain calculating the dispersion relations with the nearest-neighbor tight-binding technique and then expanding them near the extrema points.

As we have said, a single-wall carbon nanotube can be obtained rolling a monoatomic layer of graphite to form a cylinder. In particular it is completely specified by the chiral vector $\vec{C}_h$, which corresponds to a section of the nanotube perpendicular to the nanotube axis and thus has a length equal to the nanotube circumference and connects two points of the graphite sheet which coincide in the nanotube. This vector can be expressed as a linear combination of the real space unit vectors of the graphene with integer coefficients n and m:

$$\vec{C}_h = n\vec{a}_1 + m\vec{a}_2 = na\begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix} + ma\begin{bmatrix} \frac{\sqrt{3}}{2} \\ -\frac{1}{2} \\ 0 \end{bmatrix} = a\begin{bmatrix} \frac{\sqrt{3}}{2}(n+m) \\ \frac{1}{2}(n-m) \\ 0 \end{bmatrix} \quad . \tag{5.177}$$

The correspondent carbon nanotube will be indicated as (n, m).

If we define the chiral angle of the nanotube θ (with $-\pi/6 < \theta \leq \pi/6$, due to the hexagonal symmetry of graphene lattice) as the angle (positive in the clockwise direction) between $\vec{a}_1$ and $\vec{C}_h$ (or equivalently as the tilt angle of the edges of the hexagons constituting the graphite sheet with respect to the direction of the nanotube axis), it can be found from the values of n and m noting that

$$\cos\theta = \frac{\vec{C}_h\cdot\vec{a}_1}{|\vec{C}_h||\vec{a}_1|} = \frac{2n+m}{2\sqrt{n^2+m^2+nm}} \tag{5.178}$$

and

$$\sin\theta = \frac{(\vec{C}_h\times\vec{a}_1)\cdot\hat{z}'}{|\vec{C}_h||\vec{a}_1|} = \frac{\sqrt{3}m}{2\sqrt{n^2+m^2+nm}} \quad , \tag{5.179}$$

where $\hat{z}'$ is chosen in such a way to form a right-hand reference frame $(\hat{x}', \hat{y}', \hat{z}')$ with the unit vectors $\hat{x}'$ and $\hat{y}'$ which identify the frame already used in the calculations on graphene. In the successive expressions we will identify the previously introduced angle θ' with $\theta' = (\pi/6) - \theta$ (the angle between $\vec{C}_h$ and $\hat{x}'$), as shown in Fig. 5.6, and thus we will take $\hat{x}$ parallel to $\vec{C}_h$.

To find the dispersion relations and the electron wave functions of a carbon nanotube from the ones of

graphite, we have to enforce for the electron wave function the periodic boundary condition:

$$\psi(\vec{r}+\vec{C}_h)=\psi(\vec{r}) \quad . \tag{5.180}$$

Remembering that using the tight-binding technique the electron wave function can be expressed as

$$\psi(\vec{r})=\sum_{\vec{R}_A}\psi_A(\vec{R}_A)\varphi(\vec{r}-\vec{R}_A)+\sum_{\vec{R}_B}\psi_B(\vec{R}_B)\varphi(\vec{r}-\vec{R}_B) \tag{5.181}$$

the boundary condition can be written

$$\begin{aligned}
\psi(\vec{r}+\vec{C}_h)&=\sum_{\vec{R}_A}\psi_A(\vec{R}_A)\varphi((\vec{r}+\vec{C}_h)-\vec{R}_A)+\sum_{\vec{R}_B}\psi_B(\vec{R}_B)\varphi((\vec{r}+\vec{C}_h)-\vec{R}_B)=\\
&=\sum_{\vec{R}_A}\psi_A(\vec{R}_A)\varphi(\vec{r}-(\vec{R}_A-\vec{C}_h))+\sum_{\vec{R}_B}\psi_B(\vec{R}_B)\varphi(\vec{r}-(\vec{R}_B-\vec{C}_h))=\\
=\sum_{\vec{R}_A}\psi_A((\vec{R}_A-\vec{C}_h)+\vec{C}_h)\varphi(\vec{r}-(\vec{R}_A-\vec{C}_h))&+\sum_{\vec{R}_B}\psi_B((\vec{R}_B-\vec{C}_h)+\vec{C}_h)\varphi(\vec{r}-(\vec{R}_B-\vec{C}_h))=\\
&=\sum_{\vec{R}^*_A}\psi_A(\vec{R}^*_A+\vec{C}_h)\varphi(\vec{r}-\vec{R}^*_A)+\sum_{\vec{R}^*_B}\psi_B(\vec{R}^*_B+\vec{C}_h)\varphi(\vec{r}-\vec{R}^*_B)=\\
&=\psi(\vec{r})=\sum_{\vec{R}^*_A}\psi_A(\vec{R}^*_A)\varphi(\vec{r}-\vec{R}^*_A)+\sum_{\vec{R}^*_B}\psi_B(\vec{R}^*_B)\varphi(\vec{r}-\vec{R}^*_B)
\end{aligned} \tag{5.182}$$

(where we have used the fact that, being $\vec{C}_h$ a linear combination with integer coefficients of the real space lattice unit vectors, also $\vec{R}_A-\vec{C}_h$ and $\vec{R}_B-\vec{C}_h$ are atom positions $\vec{R}^*_A$ and $\vec{R}^*_B$). Thus the boundary condition is equivalent to the two conditions:

$$\begin{cases}
\psi_A(\vec{R}^*_A+\vec{C}_h)=\psi_A(\vec{R}^*_A)\\
\psi_B(\vec{R}^*_B+\vec{C}_h)=\psi_B(\vec{R}^*_B)
\end{cases} \quad . \tag{5.183}$$

If we enforce these two conditions for the expressions of $\psi_A(\vec{r})$ and $\psi_B(\vec{r})$ that we have previously used near $\vec{K}$, we have

$$\begin{cases}
e^{i\vec{K}\cdot(\vec{R}_A+\vec{C}_h)}F^{\vec{K}}_A(\vec{R}_A+\vec{C}_h)=e^{i\vec{K}\cdot\vec{R}_A}F^{\vec{K}}_A(\vec{R}_A)\\
-\dfrac{\omega^{-1}}{i}e^{i\theta'}e^{i\vec{K}\cdot(\vec{R}_B+\vec{C}_h)}F^{\vec{K}}_B(\vec{R}_B+\vec{C}_h)=-\dfrac{\omega^{-1}}{i}e^{i\theta'}e^{i\vec{K}\cdot\vec{R}_B}F^{\vec{K}}_B(\vec{R}_B)
\end{cases} \quad . \tag{5.184}$$

Multiplying the first equation by $g(\vec{r}-\vec{R}_A)e^{-i\vec{K}\cdot\vec{R}_A}$, summing it over $\vec{R}_A$ and then using the properties of the function g, we find:

$$\begin{aligned}
\sum_{\vec{R}_A} g(\vec{r}-\vec{R}_A)e^{i\vec{K}\cdot\vec{C}_h}F_A^{\vec{K}}(\vec{R}_A+\vec{C}_h) &= \sum_{\vec{R}_A} g(\vec{r}-\vec{R}_A)F_A^{\vec{K}}(\vec{R}_A) \quad ; \\
e^{i\vec{K}\cdot\vec{C}_h}\left[\sum_{\vec{R}_A} g(\vec{r}-\vec{R}_A)\right]F_A^{\vec{K}}(\vec{r}+\vec{C}_h) &= \left[\sum_{\vec{R}_A} g(\vec{r}-\vec{R}_A)\right]F_A^{\vec{K}}(\vec{r}) \quad ; \\
e^{i\vec{K}\cdot\vec{C}_h}F_A^{\vec{K}}(\vec{r}+\vec{C}_h) &= F_A^{\vec{K}}(\vec{r}) \quad . \qquad (5.185)
\end{aligned}$$

If we calculate the scalar product between $\vec{K}$ and $\vec{C}_h$ we obtain

$$\begin{aligned}
\vec{K}\cdot\vec{C}_h &= \frac{2\pi}{\sqrt{3}}\left(\frac{\sqrt{3}}{2}(n+m)+\frac{1}{2\sqrt{3}}(n-m)\right) = \frac{2\pi}{3}\left(\frac{3}{2}(n+m)+\frac{1}{2}(n-m)\right) = \\
&= \frac{2\pi}{3}\frac{4n+2m}{2} = \frac{2\pi}{3}(2n+m) = 2\pi\tilde{N}+\frac{2\pi\nu}{3} \quad , \qquad (5.186)
\end{aligned}$$

where $2n+m = 3\tilde{N}+\nu$ with $\nu = 0$ or ± 1 and $\tilde{N}$ a proper integer. Therefore we have that

$$e^{i\vec{K}\cdot\vec{C}_h} = e^{i2\pi\tilde{N}}e^{i\frac{2\pi\nu}{3}} = e^{i\frac{2\pi\nu}{3}} \qquad (5.187)$$

and thus the first boundary condition near $\vec{K}$ is

$$e^{i\frac{2\pi\nu}{3}}F_A^{\vec{K}}(\vec{r}+\vec{C}_h) = F_A^{\vec{K}}(\vec{r}) \qquad (5.188)$$

or equivalently

$$F_A^{\vec{K}}(\vec{r}+\vec{C}_h) = e^{-i\frac{2\pi\nu}{3}}F_A^{\vec{K}}(\vec{r}) \quad . \qquad (5.189)$$

Multiplying the second equation by $g(\vec{r}-\vec{R}_B)(-i\omega e^{-i\theta'}e^{-i\vec{K}\cdot\vec{R}_B})$, summing it over $\vec{R}_B$ and then using the properties of the function g, we find analogously

$$\begin{aligned}
\sum_{\vec{R}_B} g(\vec{r}-\vec{R}_B)e^{i\vec{K}\cdot\vec{C}_h}F_B^{\vec{K}}(\vec{R}_B+\vec{C}_h) &= \sum_{\vec{R}_B} g(\vec{r}-\vec{R}_B)F_B^{\vec{K}}(\vec{R}_B) \quad ; \\
e^{i\vec{K}\cdot\vec{C}_h}\left[\sum_{\vec{R}_B} g(\vec{r}-\vec{R}_B)\right]F_B^{\vec{K}}(\vec{r}+\vec{C}_h) &= \left[\sum_{\vec{R}_B} g(\vec{r}-\vec{R}_B)\right]F_B^{\vec{K}}(\vec{r}) \quad ; \\
e^{i\vec{K}\cdot\vec{C}_h}F_B^{\vec{K}}(\vec{r}+\vec{C}_h) &= F_B^{\vec{K}}(\vec{r}) \quad . \qquad (5.190)
\end{aligned}$$

Substituting the value of $e^{i\vec{K}\cdot\vec{C}_h}$, we can rewrite the second boundary condition near $\vec{K}$ in the form

$$e^{i\frac{2\pi\nu}{3}}F_B^{\vec{K}}(\vec{r}+\vec{C}_h) = F_B^{\vec{K}}(\vec{r}) \qquad (5.191)$$

or equivalently

$$F_B^{\vec{K}}(\vec{r}+\vec{C}_h) = e^{-i\frac{2\pi\nu}{3}} F_B^{\vec{K}}(\vec{r}) \quad . \tag{5.192}$$

Thus the periodic boundary condition near $\vec{K}$ is

$$\begin{bmatrix} F_A^{\vec{K}}(\vec{r}+\vec{C}_h) \\ F_B^{\vec{K}}(\vec{r}+\vec{C}_h) \end{bmatrix} = e^{-i\frac{2\pi\nu}{3}} \begin{bmatrix} F_A^{\vec{K}}(\vec{r}) \\ F_B^{\vec{K}}(\vec{r}) \end{bmatrix} \quad , \tag{5.193}$$

which can be written in this compact way:

$$\vec{F}^{\vec{K}}(\vec{r}+\vec{C}_h) = e^{-i\frac{2\pi\nu}{3}} \vec{F}^{\vec{K}}(\vec{r}) \quad . \tag{5.194}$$

But, as we have previously seen, the envelope functions have the following form:

$$\vec{F}^{\vec{K}}_{s\vec{\kappa}}(\vec{r}) = \frac{1}{\sqrt{L\ell}} e^{i\vec{\kappa}\cdot\vec{r}} \vec{F}^{\vec{K}}_{s\vec{\kappa}} = \frac{\vec{F}^{\vec{K}}_{s\vec{\kappa}}}{\sqrt{L\ell}} e^{i(\kappa_x x+\kappa_y y)} \quad , \tag{5.195}$$

with the surface area $\Sigma = L\ell$, where $L = |\vec{C}_h|$ and ℓ is the length of the nanotube. Thus the periodic boundary condition becomes

$$\frac{\vec{F}^{\vec{K}}_{s\vec{\kappa}}}{\sqrt{L\ell}} e^{i\vec{\kappa}\cdot(\vec{r}+\vec{C}_h)} = e^{-i\frac{2\pi\nu}{3}} \frac{\vec{F}^{\vec{K}}_{s\vec{\kappa}}}{\sqrt{L\ell}} e^{i\vec{\kappa}\cdot\vec{r}} \tag{5.196}$$

or equivalently

$$e^{i\vec{\kappa}\cdot\vec{C}_h} = e^{-i\frac{2\pi\nu}{3}} \quad . \tag{5.197}$$

This condition can be written also in the following way:

$$e^{i\kappa_x L} = e^{-i\frac{2\pi\nu}{3}} 1 = e^{-i\frac{2\pi\nu}{3}} e^{i2\pi\tilde{n}} \tag{5.198}$$

or equivalently

$$\kappa_x L = -\frac{2\pi\nu}{3} + 2\pi\tilde{n} \tag{5.199}$$

and thus

$$\kappa_x = \frac{2\pi}{L}\left(\tilde{n} - \frac{\nu}{3}\right) = \kappa_\nu(\tilde{n}) \tag{5.200}$$

with $\tilde{n}$ integer.

This condition on κ_x can be obtained also in a different way, imposing the boundary condition on the overall wave vector $\vec{k}$. To do so, we have to notice that, considering only the periodic lattice potential inside the graphene sheet, obviously the wave function $\psi(\vec{r})$ has to be a Bloch function $u(\vec{k},\vec{r})e^{i\vec{k}\cdot\vec{r}}$, where $u(\vec{k},\vec{r})$ has the periodicity of the lattice.

Incidentally, we notice that also the previously written tight-binding expression has this form, as we can

easily check:

$$\psi(\vec{r}) = \sum_{j=A,B} C_j(\vec{k}) \left[\frac{1}{\sqrt{N_\Omega}} \sum_{\vec{R}_j}^{N_\Omega} e^{i\vec{k}\cdot\vec{R}_j} \varphi(\vec{r}-\vec{R}_j) \right] =$$

$$= \left[\sum_{j=A,B} \sum_{\vec{R}_j} C_j(\vec{k}) \frac{1}{\sqrt{N_\Omega}} e^{-i\vec{k}\cdot(\vec{r}-\vec{R}_j)} \varphi(\vec{r}-\vec{R}_j) \right] e^{i\vec{k}\cdot\vec{r}} = u(\vec{k},\vec{r}) e^{i\vec{k}\cdot\vec{r}} \quad , \tag{5.201}$$

where $u(\vec{k},\vec{r})$ has the periodicity of the lattice because

$$\begin{aligned} u(\vec{k},\vec{r}+\vec{a}_i) &= \sum_{j=A,B} \sum_{\vec{R}_j} C_j(\vec{k}) \frac{1}{\sqrt{N_\Omega}} e^{-i\vec{k}\cdot(\vec{r}+\vec{a}_i-\vec{R}_j)} \varphi(\vec{r}+\vec{a}_i-\vec{R}_j) = \\ &= \sum_{j=A,B} \sum_{\vec{R}_j} C_j(\vec{k}) \frac{1}{\sqrt{N_\Omega}} e^{-i\vec{k}\cdot(\vec{r}-(\vec{R}_j-\vec{a}_i))} \varphi(\vec{r}-(\vec{R}_j-\vec{a}_i)) = \\ &= \sum_{j=A,B} \sum_{\vec{R}_j^*} C_j(\vec{k}) \frac{1}{\sqrt{N_\Omega}} e^{-i\vec{k}\cdot(\vec{r}-\vec{R}_j^*)} \varphi(\vec{r}-\vec{R}_j^*) = u(\vec{k},\vec{r}) \end{aligned} \tag{5.202}$$

(here we have used the fact that translating the position of an atom of a lattice by a lattice unit vector we obtain the position of another equivalent atom).
Thus the boundary condition

$$\psi(\vec{r}+\vec{C}_h) = \psi(\vec{r}) \tag{5.203}$$

is equivalent to

$$u(\vec{k},\vec{r}+\vec{C}_h) e^{i\vec{k}\cdot(\vec{r}+\vec{C}_h)} = u(\vec{k},\vec{r}) e^{i\vec{k}\cdot\vec{r}} \quad . \tag{5.204}$$

Since we know that $u(\vec{k},\vec{r})$ has the lattice periodicity and thus $u(\vec{k},\vec{r}+\vec{C}_h) = u(\vec{k},\vec{r})$ ($\vec{C}_h$ being a linear combination with integer coefficients of the lattice unit vectors) the boundary condition can also be written:

$$e^{i\vec{k}\cdot\vec{C}_h} = 1 \tag{5.205}$$

or equivalently

$$\vec{k}\cdot\vec{C}_h = 2\pi\tilde{m} \quad . \tag{5.206}$$

Thus the boundary condition is (being $\hat{C}_h = \vec{C}_h/|\vec{C}_h| = \vec{C}_h/L$):

$$\vec{k}\cdot\hat{C}_h = \vec{k}\cdot\hat{x} = k_x = (\vec{K})_x + \kappa_x = \frac{2\pi}{L}\tilde{m} \tag{5.207}$$

and (using the expression for $\vec{K}\cdot\vec{C}_h$ that we have previously written):

$$\begin{aligned} \kappa_x &= \frac{2\pi}{L}\tilde{m} - (\vec{K})_x = \frac{2\pi}{L}\tilde{m} - \frac{\vec{K}\cdot\vec{C}_h}{L} = \frac{2\pi}{L}\tilde{m} - \frac{2\pi}{L}\tilde{N} - \frac{2\pi}{3L}\nu = \\ &= \frac{2\pi}{L}\left(\tilde{m} - \tilde{N} - \frac{\nu}{3}\right) = \frac{2\pi}{L}\left(\tilde{n} - \frac{\nu}{3}\right) = \kappa_\nu(\tilde{n}) \quad , \end{aligned} \tag{5.208}$$

which is equal to the previously found expression.
If we substitute this condition on κ_x inside the dispersion relations of the graphite we find:

$$E^{\vec{K}}_{s,\nu}(\kappa_y) = s\gamma|\vec{\kappa}| = s\gamma\sqrt{\kappa_x^2+\kappa_y^2} = s\gamma\sqrt{\kappa_\nu(\tilde{n})^2+\kappa_y^2} \quad , \tag{5.209}$$

where $s=+1$ and $s=-1$ represent the conduction and valence bands, respectively.
We notice that k_y now is the wave vector k of the nanotube, which, being a substantially unidimensional material, has a one-dimensional Brillouin zone with width $2\pi/T$ (where T is the length of the unit cell of the nanotube, along its axis, which can be easily found from the numbers n and m characterizing the nanotube [30]). Correspondingly κ_y is the difference between the wave vector k of the nanotube and the component of $\vec{K}$ along $\hat{y}$.
As to the envelope functions near $\vec{K}$, if we rewrite the expression found for the graphene, then we choose as value of the arbitrary phase $\phi_s = -\alpha/2$ and finally we enforce the condition on κ_x, we can write:

$$\begin{aligned} F^{\vec{K}}_{s\vec{\kappa}}(\vec{r}) &= \frac{1}{\sqrt{L\ell}} e^{i\vec{\kappa}\cdot\vec{r}} e^{i\phi_s} \begin{bmatrix} e^{-i\frac{\alpha}{2}} & 0 \\ 0 & e^{i\frac{\alpha}{2}} \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} -is \\ 1 \end{bmatrix} = \\ &= \frac{1}{\sqrt{2L\ell}} e^{i(\kappa_x x+\kappa_y y)} e^{i\phi_s} \begin{bmatrix} -ise^{-i\frac{\alpha}{2}} \\ e^{i\frac{\alpha}{2}} \end{bmatrix} = \frac{1}{\sqrt{2L\ell}} e^{i(\kappa_x x+\kappa_y y)} \begin{bmatrix} -ise^{-i\alpha} \\ 1 \end{bmatrix} = \\ &= \frac{1}{\sqrt{2L\ell}} \begin{bmatrix} se^{-i(\frac{\pi}{2}+\alpha)} \\ 1 \end{bmatrix} e^{i\kappa_x x+i\kappa_y y} = \frac{1}{\sqrt{2L\ell}} \begin{bmatrix} sb_\nu(\tilde{n},\kappa) \\ 1 \end{bmatrix} e^{i\kappa_\nu(\tilde{n})x+i\kappa_y y} \quad . \end{aligned} \tag{5.210}$$

The function $b_\nu(\tilde{n},\kappa) = e^{-i(\frac{\pi}{2}+\alpha)}$ can be found noting that α has been defined in such a way that

$$\vec{\kappa} = |\vec{\kappa}| e^{i\left(\frac{\pi}{2}+\alpha\right)} \quad ; \tag{5.211}$$

this means that

$$e^{i\left(\frac{\pi}{2}+\alpha\right)} = \frac{\vec{\kappa}}{|\vec{\kappa}|} = \frac{\kappa_x+i\kappa_y}{\sqrt{\kappa_x^2+\kappa_y^2}} \tag{5.212}$$

and thus

$$\begin{aligned} b_\nu(\tilde{n},\kappa) &= e^{-i\left(\frac{\pi}{2}+\alpha\right)} = \left(e^{i\left(\frac{\pi}{2}+\alpha\right)}\right)^* = \\ &= \left(\frac{\kappa_x+i\kappa_y}{\sqrt{\kappa_x^2+\kappa_y^2}}\right)^* = \frac{\kappa_x-i\kappa_y}{\sqrt{\kappa_x^2+\kappa_y^2}} = \frac{\kappa_\nu(\tilde{n})-i\kappa_y}{\sqrt{\kappa_\nu(\tilde{n})^2+\kappa_y^2}} \quad . \end{aligned} \tag{5.213}$$

Analogously, if we impose the two boundary conditions

$$\begin{cases} \psi_A(\vec{R}_A+\vec{C}_h) = \psi_A(\vec{R}_A) \\ \psi_B(\vec{R}_B+\vec{C}_h) = \psi_B(\vec{R}_B) \end{cases} \tag{5.214}$$

to the expressions of $\psi_A(\vec{r})$ and $\psi_B(\vec{r})$ that we have previously used near $\vec{K}'$, we have

$$\begin{cases} \frac{\omega}{i}e^{i\theta'}e^{i\vec{K}'\cdot(\vec{R}_A+\vec{C}_h)}F_A^{\vec{K}'}(\vec{R}_A+\vec{C}_h) = \frac{\omega}{i}e^{i\theta'}e^{i\vec{K}'\cdot\vec{R}_A}F_A^{\vec{K}'}(\vec{R}_A) \\ e^{i\vec{K}'\cdot(\vec{R}_B+\vec{C}_h)}F_B^{\vec{K}'}(\vec{R}_B+\vec{C}_h) = e^{i\vec{K}'\cdot\vec{R}_B}F_B^{\vec{K}'}(\vec{R}_B) \end{cases} \quad . \tag{5.215}$$

Multiplying the first equation by $g(\vec{r}-\vec{R}_A)(i\omega^{-1}e^{-i\theta'}e^{-i\vec{K}'\cdot\vec{R}_A})$, summing it over $\vec{R}_A$ and then using the properties of the function g, we find:

$$\begin{aligned} \sum_{\vec{R}_A} g(\vec{r}-\vec{R}_A)e^{i\vec{K}'\cdot\vec{C}_h}F_A^{\vec{K}'}(\vec{R}_A+\vec{C}_h) &= \sum_{\vec{R}_A} g(\vec{r}-\vec{R}_A)F_A^{\vec{K}'}(\vec{R}_A) \quad ; \\ e^{i\vec{K}'\cdot\vec{C}_h}\left[\sum_{\vec{R}_A} g(\vec{r}-\vec{R}_A)\right]F_A^{\vec{K}'}(\vec{r}+\vec{C}_h) &= \left[\sum_{\vec{R}_A} g(\vec{r}-\vec{R}_A)\right]F_A^{\vec{K}'}(\vec{r}) \quad ; \\ e^{i\vec{K}'\cdot\vec{C}_h}F_A^{\vec{K}'}(\vec{r}+\vec{C}_h) &= F_A^{\vec{K}'}(\vec{r}) \quad . \end{aligned} \tag{5.216}$$

The scalar product between $\vec{K}'$ and $\vec{C}_h$ is equal to

$$\begin{aligned} \vec{K}'\cdot\vec{C}_h &= \frac{2\pi}{\sqrt{3}}\left(\frac{\sqrt{3}}{2}(n+m)-\frac{1}{2\sqrt{3}}(n-m)\right) = \frac{2\pi}{3}\left(\frac{3}{2}(n+m)-\frac{1}{2}(n-m)\right) = \\ &= \frac{2\pi}{3}\left(3(n+m)-\frac{3}{2}(n+m)-\frac{1}{2}(n-m)\right) = 2\pi(n+m)-\frac{2\pi}{3}\frac{4n+2m}{2} = \\ &= 2\pi(n+m)-\frac{2\pi}{3}(2n+m) = 2\pi(n+m-\tilde{N})-\frac{2\pi\nu}{3} \quad , \end{aligned} \tag{5.217}$$

where we have used the previously introduced relation $2n+m = 3\tilde{N}+\nu$ with $\nu = 0$ or ± 1 and $\tilde{N}$ a proper integer. Thus we have that

$$e^{i\vec{K}'\cdot\vec{C}_h} = e^{i2\pi(n+m-\tilde{N})}e^{-i\frac{2\pi\nu}{3}} = e^{-i\frac{2\pi\nu}{3}}$$

and consequently the boundary condition near K' is

$$e^{-i\frac{2\pi\nu}{3}}F_A^{\vec{K}'}(\vec{r}+\vec{C}_h) = F_A^{\vec{K}'}(\vec{r}) \tag{5.218}$$

or equivalently

$$F_A^{\vec{K}'}(\vec{r}+\vec{C}_h) = e^{i\frac{2\pi\nu}{3}}F_A^{\vec{K}'}(\vec{r}) \quad . \tag{5.219}$$

On the other hand, multiplying the second equation by $g(\vec{r}-\vec{R}_B)e^{-i\vec{K}'\cdot\vec{R}_B}$, summing it over $\vec{R}_B$ and then using the properties of the function g, we find

$$\begin{aligned}
\sum_{\vec{R}_B} g(\vec{r}-\vec{R}_B)e^{i\vec{K}'\cdot\vec{C}_h}F_B^{\vec{K}'}(\vec{R}_B+\vec{C}_h) &= \sum_{\vec{R}_B} g(\vec{r}-\vec{R}_B)F_B^{\vec{K}'}(\vec{R}_B) \quad ; \\
e^{i\vec{K}'\cdot\vec{C}_h}\left[\sum_{\vec{R}_B} g(\vec{r}-\vec{R}_B)\right]F_B^{\vec{K}'}(\vec{r}+\vec{C}_h) &= \left[\sum_{\vec{R}_B} g(\vec{r}-\vec{R}_B)\right]F_B^{\vec{K}'}(\vec{r}) \quad ; \\
e^{i\vec{K}'\cdot\vec{C}_h}F_B^{\vec{K}'}(\vec{r}+\vec{C}_h) &= F_B^{\vec{K}'}(\vec{r}) \quad . \qquad (5.220)
\end{aligned}$$

Substituting the value of $e^{i\vec{K}'\cdot\vec{C}_h}$, we can rewrite this second boundary condition near $\vec{K}'$ in the form

$$e^{-i\frac{2\pi\nu}{3}}F_B^{\vec{K}'}(\vec{r}+\vec{C}_h) = F_B^{\vec{K}'}(\vec{r}) \qquad (5.221)$$

or equivalently

$$F_B^{\vec{K}'}(\vec{r}+\vec{C}_h) = e^{i\frac{2\pi\nu}{3}}F_B^{\vec{K}'}(\vec{r}) \quad . \qquad (5.222)$$

Thus the overall periodic boundary condition near $\vec{K}'$ is

$$\begin{bmatrix} F_A^{\vec{K}'}(\vec{r}+\vec{C}_h) \\ F_B^{\vec{K}'}(\vec{r}+\vec{C}_h) \end{bmatrix} = e^{i\frac{2\pi\nu}{3}}\begin{bmatrix} F_A^{\vec{K}'}(\vec{r}) \\ F_B^{\vec{K}'}(\vec{r}) \end{bmatrix} \quad , \qquad (5.223)$$

which can be written compactly:

$$\vec{F}^{\vec{K}'}(\vec{r}+\vec{C}_h) = e^{i\frac{2\pi\nu}{3}}\vec{F}^{\vec{K}'}(\vec{r}) \quad . \qquad (5.224)$$

Substituting the form that the envelope functions have near $\vec{K}'$

$$\vec{F}_{s\vec{\kappa}}^{\vec{K}'}(\vec{r}) = \frac{1}{\sqrt{L\ell}}e^{i\vec{\kappa}\cdot\vec{r}}\vec{F}_{s\vec{\kappa}}^{\vec{K}'} = \frac{\vec{F}_{s\vec{\kappa}}^{\vec{K}'}}{\sqrt{L\ell}}e^{i(\kappa_x x+\kappa_y y)} \qquad (5.225)$$

the periodic boundary condition becomes

$$\frac{\vec{F}_{s\vec{\kappa}}^{\vec{K}'}}{\sqrt{L\ell}}e^{i\vec{\kappa}\cdot(\vec{r}+\vec{C}_h)} = e^{i\frac{2\pi\nu}{3}}\frac{\vec{F}_{s\vec{\kappa}}^{\vec{K}'}}{\sqrt{L\ell}}e^{i\vec{\kappa}\cdot\vec{r}} \qquad (5.226)$$

or equivalently

$$e^{i\vec{\kappa}\cdot\vec{C}_h} = e^{i\frac{2\pi\nu}{3}} \quad . \qquad (5.227)$$

This can be rewritten in the form

$$e^{i\kappa_x L} = e^{i\frac{2\pi\nu}{3}}1 = e^{i\frac{2\pi\nu}{3}}e^{i2\pi\tilde{n}} \qquad (5.228)$$

or equivalently

$$\kappa_x L = \frac{2\pi\nu}{3} + 2\pi\tilde{n} \qquad (5.229)$$

and thus

$$\kappa_x = \frac{2\pi}{L}\left(\tilde{n}+\frac{\nu}{3}\right) = \kappa_\nu(\tilde{n}) \tag{5.230}$$

with $\tilde{n}$ integer.

Analogously to what we have done near $\vec{K}$, this condition on κ_x can be found also imposing that

$$e^{i\vec{k}\cdot\vec{C}_h} = 1 \tag{5.231}$$

or equivalently

$$\vec{k}\cdot\hat{C}_h = k_x = (\vec{K}')_x + \kappa_x = \frac{2\pi}{L}\tilde{m} \quad , \tag{5.232}$$

which using the expression for $\vec{K}'\cdot\vec{C}_h$ becomes

$$\begin{aligned}\kappa_x &= \frac{2\pi}{L}\tilde{m} - (\vec{K}')_x = \frac{2\pi}{L}\tilde{m} - \frac{\vec{K}'\cdot\vec{C}_h}{L} = \frac{2\pi}{L}\tilde{m} - \frac{2\pi}{L}(n+m-\tilde{N}) + \frac{2\pi}{3L}\nu = \\ &= \frac{2\pi}{L}\left(\tilde{m}-n-m+\tilde{N}+\frac{\nu}{3}\right) = \frac{2\pi}{L}\left(\tilde{n}+\frac{\nu}{3}\right) = \kappa_\nu(\tilde{n}) \quad .\end{aligned} \tag{5.233}$$

If we substitute this condition on κ_x inside the dispersion relations of the graphite we find:

$$E^{\vec{K}'}_{s,\nu}(\kappa_y) = s\gamma|\vec{\kappa}| = s\gamma\sqrt{\kappa_x^2+\kappa_y^2} = s\gamma\sqrt{\kappa_\nu(\tilde{n})^2+\kappa_y^2} \quad , \tag{5.234}$$

where k_y now is the wave vector k of the nanotube and κ_y is the difference between the wave vector k of the nanotube and the component of $\vec{K}'$ along $\hat{y}$.

On the other hand, if we rewrite the expression of the electron envelope functions of the graphite near $\vec{K}'$, we choose as arbitrary phase $\phi_s = \alpha/2$ and finally we enforce the condition on κ_x, we find as envelope functions in the carbon nanotube near $\vec{K}'$:

$$\begin{aligned}\vec{F}^{\vec{K}'}_{s\vec{\kappa}}(\vec{r}) &= \frac{1}{\sqrt{L\ell}}e^{i\vec{\kappa}\cdot\vec{r}}e^{i\phi_s}\begin{bmatrix} e^{i\frac{\alpha}{2}} & 0 \\ 0 & e^{-i\frac{\alpha}{2}}\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} is \\ 1\end{bmatrix} = \\ &= \frac{1}{\sqrt{2L\ell}}e^{i(\kappa_x x+\kappa_y y)}e^{i\phi_s}\begin{bmatrix} ise^{i\frac{\alpha}{2}} \\ e^{-i\frac{\alpha}{2}}\end{bmatrix} = \frac{1}{\sqrt{2L\ell}}e^{i(\kappa_x x+\kappa_y y)}\begin{bmatrix} ise^{i\alpha} \\ 1\end{bmatrix} = \\ &= \frac{1}{\sqrt{2L\ell}}\begin{bmatrix} se^{i(\frac{\pi}{2}+\alpha)} \\ 1\end{bmatrix}e^{i\kappa_x x+i\kappa_y y} = \frac{1}{\sqrt{2L\ell}}\begin{bmatrix} sb_\nu(\tilde{n},\kappa_y) \\ 1\end{bmatrix}e^{i\kappa_\nu(\tilde{n})x+i\kappa_y y}\end{aligned} \tag{5.235}$$

where for the definition of the angle α we have that

$$b_\nu(\tilde{n},\kappa_y) = e^{i\left(\frac{\pi}{2}+\alpha\right)} = \frac{\vec{\kappa}}{|\vec{\kappa}|} = \frac{\kappa_x+i\kappa_y}{\sqrt{\kappa_x^2+\kappa_y^2}} = \frac{\kappa_\nu(\tilde{n})+i\kappa_y}{\sqrt{\kappa_\nu(\tilde{n})^2+\kappa_y^2}} \quad . \tag{5.236}$$

Let us notice again that also if we consider points of the reciprocal space equivalent to $\vec{K}$ and $\vec{K}'$ we obtain the same final expressions, with $\tilde{n}$ integer.

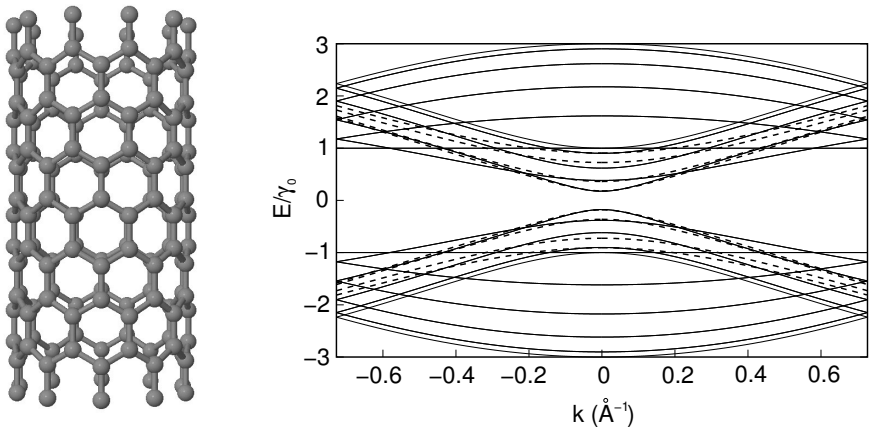


Figure 5.7: The nanotube (10,0) and its dispersion relations, obtained both by the tight-binding method (solid lines) and (for the bands correspondent to the smallest values of $|\kappa_x|$) by the $\vec{k} \cdot \vec{p}$ method (dashed lines).

If $2n + m$ is a multiple of 3 and thus $\nu = 0$, for $\tilde{n} = 0$ we have that $\kappa_x = \kappa_\nu(\tilde{n}) = 0$ and consequently $E_{s,\nu} = s\gamma|\kappa_y|$, that vanishes for $\kappa_y = 0$, so that $E_{+,\nu} = E_{-,\nu} = 0$. This means that when $2n + m$ is a multiple of 3 among the values of $\vec{k}$ which are allowed by the periodic boundary condition there are also the points $\vec{K}$ and $\vec{K}'$, where the upper and lower bands of the graphite are degenerate, and thus the nanotube is metallic.

Instead, if $2n+m$ is not a multiple of 3 and thus $\nu = \pm 1$, the allowed $\vec{k}$'s nearest to $\vec{K}$ and $\vec{K}'$ correspond to $\tilde{n} = 0$, for which $\kappa_x = \kappa_\nu(\tilde{n}) = \pm 2\pi/(3L)$ and consequently

$$E_{s,\nu} = s\gamma\sqrt{\left(\frac{2\pi}{3L}\right)^2 + \kappa_y^2} \quad . \tag{5.237}$$

In particular, the minimum and maximum values of the nanotube bands are obtained with the further position $\kappa_y = 0$ and therefore are equal to

$$E_{s,\nu} = s\gamma\frac{2\pi}{3L} \quad ; \tag{5.238}$$

thus the bandgap of the nanotube is

$$E_g = E_{+,\nu} - E_{-,\nu} = 2\gamma\frac{2\pi}{3L} = \frac{4\pi\gamma}{3L} = \frac{4\pi}{3L}\frac{\sqrt{3}a\gamma_0}{2} = 2\frac{\pi}{L}\frac{a}{\sqrt{3}}\gamma_0 = \frac{2\gamma_0\, a_{C-C}}{d_t} \quad , \tag{5.239}$$

where $d_t = L/\pi$ is the nanotube diameter. Therefore we have that the bandgap of the nanotube depends upon the reciprocal nanotube diameter.

We notice that the approximate approach for the computation of the density of states in carbon nanotubes proposed by J. W. Mintmire and C. T. White [31], being based on a linear approximations of the dispersion relations of the graphene near the extrema points, can be consequently seen as a consequence of a $\vec{k} \cdot \vec{p}$ study of the nanotube energy bands.

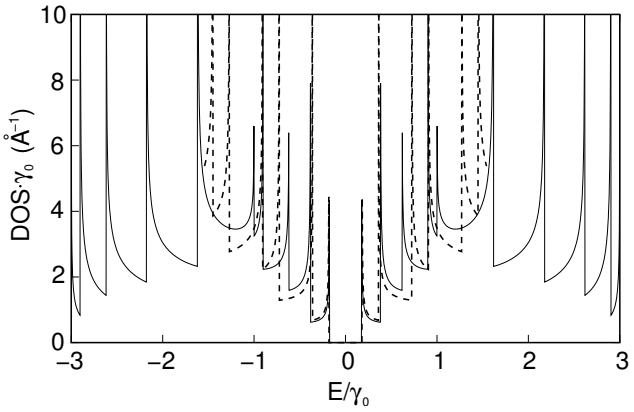


Figure 5.8: The density of states for unit length of the nanotube (10,0), obtained both by the tight-binding method (solid lines) and (in a smaller region around $E = 0$) by the $\vec{k} \cdot \vec{p}$ method (dashed lines).

In Fig. 5.7 we compare the dispersion relations that we have obtained for the same carbon nanotube using the nearest-neighbor tight-binding method and the $\vec{k} \cdot \vec{p}$ method. We see that the $\vec{k} \cdot \vec{p}$ method gives a good approximation for the parts of the nanotube dispersion relations obtained from regions of the graphene reciprocal space near $\vec{K}$ and $\vec{K}'$.
In Fig. 5.8 instead we show for the same nanotube both the density of states that we have obtained properly differentiating the dispersion relations obtained with the tight-binding approach and the density of states obtained using the Mintmire-White approach. We see that this last approximation gives good results near $E = 0$, which actually is the region that derives from the parts of the graphite energy bands near $\vec{K}$ and $\vec{K}'$.

5.5 Conclusion

The $\vec{k} \cdot \vec{p}$ method and the related envelope function method are widely used to study the physical properties of materials. They have been developed in many and sometimes quite different ways by several authors and have been successfully applied to a multitude of different problems. This explains the great variety and inhomogeneity of the related literature. In this review we have briefly described the basics of these methodologies, dwelling upon the treatments that we have considered more useful for an easy comprehension. For a detailed explanation of the different approaches, the interested reader can resort to the many papers and books on the topic, some of which are listed in the references.

Acknowledgements

We would like to thank Prof. Tsuneya Ando and Prof. Paolo Lugli for useful discussions and suggestions.

Bibliography

[1] See for example J. Callaway, *Energy band theory* (Academic Press, New York, 1964).

[2] A detailed description of the $\vec{k} \cdot \vec{p}$ method can be found in the following books:
W. T. Wenckebach, *Essentials of semiconductor physics* (J. Wiley & Sons, Chichester, 1999);
S. L. Chuang, *Physics of optoelectronic devices* (J. Wiley & Sons, New York, 1995);
I. M. Tsidilkovski, *Band Structure of semiconductors (International series on the science of solid state, vol.19)* (Pergamon Press, Oxford, 1982) (translation of *Zonnaia struktura poluprovodnikov*);
P. Y. Yu, M. Cardona, *Fundamentals of semiconductors* (Springer-Verlag, Berlin, 1995);
Ming-Fu Li, *Modern semiconductor quantum physics* (World Scientific, Singapore, 1994);
E. O. Kane, *Band structure of narrow gap semiconductors*, in W. Zawadzki, *Narrow gap semiconductors: physics and applications (Lecture notes in physics, vol.133)* (Springer-Verlag, Berlin, 1980), p.13;
E. O. Kane, *The $\vec{k} \cdot \vec{p}$ method*, in R. K. Willardson, A. C. Beer, *Semiconductors and semimetals, vol.1: Physics of III-V compounds* (Academic Press, New York, 1966), p.75;
E. O. Kane, *Energy band theory*, in T. S. Moss, W.Paul, *Handbook on semiconductors, vol 1: Band theory and transport properties* (North-Holland Publishing Company, Amsterdam, 1982), p.193;
C. R. Pidgeon, *Free carrier optical properties of semiconductors*, in T. S. Moss, M. Balkanski, *Handbook on semiconductors, vol 2: Optical properties of solids* (North-Holland Publishing Company, Amsterdam, 1980), p.223;
G. L. Bir, G. E. Pikus, *Symmetry and strain-induced effects in semiconductors* (J. Wiley & Sons, New York / Keter Publishing House, Jerusalem, 1974) (translation of *Simmetriya i deformatsionnye effekty v poluprovodnikakh*, Moskow, Izdatel'stvo "Nauka", 1972);
A. I. Anselm, *Introduction to semiconductor theory* (MIR Publishers-Moscow, Moscow, 1978; English translation:Prentice Hall, 1981);
J. Callaway, *Quantum theory of the solid state* (Academic Press, San Diego, 1974);
J. Singh, *Electronic and optoelectronic properties of semiconductor structures* (Cambridge University Press, Cambridge, 2003);
D. Long, *Energy bands in semiconductors* (J. Wiley & Sons, New York, 1968);
H. Kroemer, *Quantum mechanics* (Prentice Hall, Englewood Cliffs, 1994);
F. Bassani, G. Pastori Parravicini, *Electronic states and optical transitions in solids* (Pergamon Press, Oxford, 1975);
R. Enderlein, N. J. M. Horing, *Fundamentals of semiconductor physics and devices* (World Scientific, Singapore, 1997);
H. Haug, S. W. Koch, *Quantum theory of the optical and electronic properties of semiconductors* (World Scientific, Singapore, 1990);
W.A.Harrison, *Electronic structure and the properties of solids* (Dover Publications, New York, 1980).

[3] J. Bardeen, *An Improved Calculation of the Energies of Metallic Li and Na*, J. Chem. Phys. **6**, 367 (1938).

[4] F. Seitz, *The modern theory of solids* (McGraw Hill, New York, 1940), p.352.

[5] W. Shockley, *Energy band structures in semiconductors*, Phys. Rev. **78**, 173 (1950).

[6] G. Dresselhaus, A. F. Kip, C. Kittel, *Cyclotron resonance of electrons and holes in silicon and germanium crystals*, Phys. Rev. **98**, 368 (1955).

[7] E. O. Kane, *Energy band structure in p-type germanium and silicon*, J. Phys. Chem. Solids **1**, 82 (1956);
E. O. Kane, *Energy band of indium antimonide*, J. Phys. Chem. Solids **1**, 249 (1957);
E. O. Kane, *The semi-empirical approach to band structure*, J. Phys. Chem. Solids **8**, 38 (1959);
E. O. Kane, *Band structure of narrow gap semiconductors*, in W. Zawadzki, *Narrow gap semiconductors: physics and applications (Lecture notes in physics, vol.133)* (Springer-Verlag, Berlin, 1980), p.13;
E. O. Kane, *The $\vec{k} \cdot \vec{p}$ method*, in R. K. Willardson, A. C. Beer, *Semiconductors and semimetals, vol.1: Physics of III-V compounds* (Academic Press, New York, 1966), p.75;
E. O. Kane, *Energy band theory*, in T. S. Moss, W.Paul, *Handbook on semiconductors, vol 1: Band theory and transport properties* (North-Holland Publishing Company, Amsterdam, 1982), p.193.

[8] J. M. Luttinger, W. Kohn, *Motion of electrons and holes in perturbed periodic fields*, Phys. Rev. **97**, 869 (1955).

[9] J. M. Luttinger, *Quantum theory of cyclotron resonance in semiconductors: general theory*, Phys. Rev. **102**, 1030 (1956).

[10] G. L. Bir, G. E. Pikus, *Symmetry and strain-induced effects in semiconductors* (J. Wiley & Sons, New York / Keter Publishing House, Jerusalem, 1974) (translation of *Simmetriya i deformatsionnye effekty v poluprovodnikakh*, Moskow, Izdatel'stvo "Nauka", 1972).

[11] G. Bastard, *Wave mechanics applied to semiconductor heterostructures* (Les editions de physique, Les Ulis Cedex, 1992).

[12] M. Altarelli, *Electronic structure of semiconductor superlattices*, in *Application of high magnetic fields in semiconductor physics (Lecture notes in physics, vol.177)* (Springer-Verlag, Berlin, 1983), p.174;
M. Altarelli, *Electronic structure and semiconductor-semimetal transition in InAs-GaSb superlattices*, Phys. Rev. B **28**, 842 (1983);
M. Altarelli, *Electronic structure of semiconductor superlattices*, Physica **117B & 118B**, 747 (1983).

[13] M. G. Burt, *An exact formulation of the envelope function method for the determination of electronic states in semiconductor microstructures*, Semicond. Sci. Technol. **2**, 460 (1987);

M. G. Burt, *A new effective-mass equation for microstructures*, Semicond. Sci. Technol. **3**, 1224 (1988);
M. G. Burt, *On the validity and range of applicability of the particle in a box model*, Appl. Phys. Lett. **65**, 717 (1994);
M. G. Burt, *An exact formulation of the envelope function method for the determination of electronic states in semiconductor microstructures*, Semicond. Sci. Technol. **3**, 739 (1988);
M. G. Burt, *The justification for applying the effective-mass approximation to microstructures*, J. Phys.: Condens. Matter **4**, 6651 (1992).

[14] M. Ajiki, T. Ando, *Electronic States of Carbon Nanotubes*,
J. Phys. Soc. Jpn. **62**, 1255 (1993);
T. Ando, *Theory of Electronic States and Transport in Carbon Nanotubes*, J. Phys. Soc. Jpn. **74**, 777 (2005).

[15] E. O. Kane, *Band structure of narrow gap semiconductors*, in W. Zawadzki, *Narrow gap semiconductors: physics and applications (Lecture notes in physics, vol.133)* (Springer-Verlag, Berlin, 1980), p.13

[16] For a brief but very clear explanation of the perturbation theory, also in the degenerate case, see the Appendix C of W. T. Wenckebach, *Essentials of semiconductor physics* (J. Wiley & Sons, Chichester, 1999); see also for example H. Kroemer, *Quantum mechanics* (Prentice Hall, Englewood Cliffs, 1994).

[17] See the Chapter 4 of W. T. Wenckebach, *Essentials of semiconductor physics* (J. Wiley & Sons, Chichester, 1999).

[18] See for example H. Kroemer, *Quantum mechanics* (Prentice Hall, Englewood Cliffs, 1994), p.434.

[19] For a simplified explanation see H. Kroemer, *Quantum mechanics* (Prentice Hall, Englewood Cliffs, 1994), p.449.

[20] P. Löwdin, *A note on the quantum-mechanical perturbation theory*,
J. Chem. Phys. **19**, 1396 (1951);
see also S. Datta, *Quantum phenomena* (Addison-Wesley, Reading, 1989), p.68 and S. L. Chuang, *Physics of optoelectronic devices* (J. Wiley & Sons, New York, 1995), p.114.

[21] E. O. Kane, *Energy band of indium antimonide*, J. Phys. Chem. Solids **1**, 249 (1957).

[22] S. Datta, *Quantum phenomena* (Addison-Wesley, Reading, 1989), p.186;
see also S. L. Chuang, *Physics of optoelectronic devices* (J. Wiley & Sons, New York, 1995), Appendix B and V. V. Mitin, V. A. Kochelap, M. A. Stroscio, *Quantum Heterostructures* (Cambridge University Press, Cambridge, 1999).

[23] Notice that there is also an alternative approach to the envelope function theory using the definition of Wannier (G. H. Wannier, *The structure of electronic excitation levels in insulating crystals*, Phys. Rev. **52**, 191 (1937)) and Slater (J. C. Slater, *Electrons in perturbed periodic lattices*,

Phys. Rev. **76**, 1592 (1949)) based on Wannier orbitals. See also K. Young, *Position-dependent effective mass for inhomogeneous semiconductors*, Phys. Rev. B **39**, 13434 (1989) and M. G. Burt, *The justification for applying the effective-mass approximation to microstructures*, J. Phys.: Condens. Matter **4**, 6651 (1992).

[24] P. Lawaetz, *Valence-Band Parameters in Cubic Semiconductors*, Phys. Rev. B **4**, 3460 (1971).

[25] See also R. I. Taylor, M. G. Burt, *Continuity conditions for envelope functions at semiconductor interfaces derived from the Bastard model*, Semicond. Sci. Technol. **2**, 485 (1987).

[26] G. Bastard, J. A. Brum, R. Ferreira, *Electronic states in semiconductor heterostructures*, in H. Ehrenreich, D. Turnbull, *Solid state physics, vol. 44* (Academic Press, New York, 1991), p.229.

[27] M. Altarelli, F. Bassani, *Impurity states: theoretical*, in T. S. Moss, W.Paul, *Handbook on semiconductors, vol 1: Band theory and transport properties* (North-Holland Publishing Company, Amsterdam, 1982), p.269.

[28] R. Saito, G. Dresselhaus, M. S. Dresselhaus, *Physical Properties of Carbon Nanotubes* (Imperial College Press, London, 1998).

[29] See the Chapter 2 of R. Saito, G. Dresselhaus, M. S. Dresselhaus, *Physical Properties of Carbon Nanotubes* (Imperial College Press, London, 1998).

[30] R. Saito, G. Dresselhaus, M. S. Dresselhaus, *Physical Properties of Carbon Nanotubes* (Imperial College Press, London, 1998), pp. 39-40.

[31] J. W. Mintmire, C. T. White, *Universal Density of States for Carbon Nanotubes*, Phys. Rev. Lett. **81**, 2506 (1998).

Chapter 6

Green function techniques in the treatement of quantum transport at the molecular scale

D. A. Ryndyk[1], R. Gutiérrez[1,2], B. Song[1,2], G. Cuniberti[1,2]

[1] Institute for Theoretical Physics, University of Regensburg, D-93040, Regensburg, Germany

[2] Max Bergmann Center for Biomaterials, Dresden University of Technology, D-01062, Dresden, Germany

Abstract. The theoretical investigation of charge (and spin) transport at nanometer length scales requires the use of advanced and powerful techniques able to deal with the dynamical properties of the relevant physical systems, to explicitly include out-of-equilibrium situations typical for electrical/heat transport as well as to take into account interaction effects in a systematic way. Equilibrium Green function techniques and their extension to non-equilibrium situations via the Keldysh formalism build one of the pillars of current state-of-the-art approaches to quantum transport which have been implemented in both model Hamiltonian formulations and first-principle methodologies. In this article we offer an overview of the applications of Green functions to deal with some fundamental aspects of charge transport at the nanoscale, mainly focusing on applications to model Hamiltonian formulations.

6.1 Introduction

The natural limitations that are expected to arise by the further miniaturization attempts of semiconductor-based electronic devices have led in the past two decades to the emergence of the new field of molecular electronics, where electronic functions are going to be performed at the single-molecule level, see for a recent overview Ref. [1]. The original conception which lies at the bottom of this fascinating field can be

traced back to the paper by Ari Aviram and Mark Ratner in 1974 [2], where a single-molecule rectifying diode was proposed. Obviously, one of the core issues at stake in molecular electronics is to clarify the question whether single molecules (or more complex molecular aggregates) can support an electric current. To achieve this goal, extremely refined experimental techniques are required in order to probe the response of such a nano-object to external fields. The meanwhile paradigmatic situation is that of a single molecule contacted by two metallic electrodes between which a bias voltage is applied. Enormous progress has been achieved in the experimental realization of such nano-devices [3, 4, 7, 8, 5, 6, 9, 10, 11, 12, 13, 14, 15]; we only mention the development of mechanically controllable break-junctions [3, 4] and scanning tunneling microscopy-based techniques [11, 12, 13, 14, 15]. With their help, a plethora of interesting phenomena like rectification [7], negative differential conductance [16], Coulomb blockade [10], and Kondo effect [6], among others, have been demonstrated. Apart from single molecules, carbon nanotubes have also found extensive applications and have been the target of experimental and theoretical studies over the last years, see Ref. [17] for a very recent review. The expectations to realize electronic at the molecular scale also reached into the domain of bio-molecular systems, thus opening new perspectives for the field due to the specific self-recognition and self-assembling properties of biomolecules. For instance, DNA oligomers have been already used as templates in molecular electronic circuits [18]. Much less clear is, however, if bio-molecules, and more specifically short DNA oligomers could also act as wiring systems. Their electrical response properties are much harder to disclose and there is still much controversy about the factors that determine charge migration through such systems [20, 21, 22].

The theoretical treatment of transport at the nanoscale requires the combined use of different techniques [23, 24, 25, 27, 26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47] which range from minimal model Hamiltonians, passing through semi-empirical methods up to full first-principle methodologies. Model Hamiltonians can in a straightforward way select, out of the many variables that can control charge migration those which are thought to be the most relevant ones for a specific molecule-electrode set-up. They contain, however, in a sometimes not well-controlled way, many free parameters; hence, they can point at generic effects, but they must be complemented with other methodologies able to yield microscopic specific information. Semi-empirical methods can deal with rather large systems due to the use of special subsets of electronic states to construct molecular Hamiltonians as well as to the approximate treatment of interactions, but often have the drawback of not being transferable. *Ab initio* approaches, finally, can deal in a very precise manner with the electronic and atomic structure of the different constituents of a molecular junction (metallic electrodes, molecular wire, the interface) but it is not *apriori* evident that they can also be applied to strong non-equilibrium situations (density-functional-theory is a ground state theory and e.g. the transmission calculated using static DFT eigenvalues will display peaks at the Kohn-Sham excitation energies, which in general do not coincide with the true excitation energies). Extensions to include excited states as in time-dependent density-functional theory, though very promising, are not fully developed up to date [48].

From a more formal standpoint, there are roughly two main theoretical frameworks that can be used to study quantum transport: generalized master equations (GME) [49] and Green function (GF) techniques [50, 51, 52, 53]. The former also lead to the more simple rate equations in the case where (i)

the electrode-system coupling can be considered as a weak perturbation, and (ii) off-diagonal elements of the reduced density matrix (coherences) can be neglected due to very short decoherence times. Both approaches, the GME and GF techniques, can yield formally exact expressions for many observables. For non-interacting systems, one can even solve analytically many models. However, once interactions are introduced - and these are the most interesting cases containing a very rich physics - different approximation schemes have to be introduced to make the problems tractable.

In this chapter, we will review the technique of non-equilibrium Keldysh Green functions (NEGF). This approach is able to deal with a very broad variety of physical problems related to quantum transport at the molecular scale. It can deal with strong non-equilibrium situations via an extension of the conventional GF formalism to the Schwinger-Keldysch contour [51] and it can also include interaction effects (electron-electron, electron-vibron, etc) in a systematic way (diagrammatic perturbation theory, equation of motion techniques). Moreover, as we will show later on, it can reproduce results obtained within the master equation approach in the weak coupling limit to the electrodes (Coulomb blockade), but it can also go *beyond* this limit and cover intermediate coupling (Kondo effect) and strong coupling (Fabry-Perot) domains. It thus offer the possibility of dealing with different physical regimes in a unified way.

In Sections **II** we will first introduce the Green functions for non-interacting systems, and present few examples of transport through non-interacting regions. Then we review the master equation approach and its application to describe Coulomb blockade and vibron-mediated Franck-Condon blockade. In Section **III** the Keldysh NEGF technique is developed in detail. In equilibrium situations or within the linear response regime, dynamic response and static correlation functions are related via the fluctuation-dissipation theorem. Thus, solving Dyson's equation for the retarded GF is enough to obtain the correlation functions. In strong out-of-equilibrium situations, however, dynamic response and correlation functions have to be calculated simultaneously and are not related by fluctuation-dissipation theorems. The Kadanof-Baym-Keldysh approach yield a compact, powerful formulation to derive Dyson and kinetic equations for non-equilibrium systems. In Sec. **IV** we present different applications of the Green function techniques. We show how Coulomb blockade can be described within the Anderson-Hubbard model, once an appropriate truncation of the equation of motion hierarchy is performed (Sec. IV.A). Further, the paradigmatic case of transport through a single electronic level coupled to a local vibrational mode is discussed in detail within the context of the self-consistent Born approximation. It is shown that already this simple model can display non-trivial physics (Sec. IV.B). Finally, the case of an electronic system interacting with a bosonic bath is discussed in Sec. IV.C where it is shown that the presence of an environment with a continuous spectrum can modified the low-energy analytical structure of the Green function and lead to dramatic changes in the electrical response of the system. We point at the relevance of this situation to discuss transport experiments in short DNA oligomers. We have not addressed the problem of the (equilibrium or non-equilibrium) Kondo effect, since this issue alone would require a chapter on its own due to the non-perturbative character of the processes leading to the formation of the Kondo resonance [54, 55, 56]

In view of the broadness of the topic, the authors were forced to do a very subjective selection of the topics to be included in this review as well as of the most relevant literature. We thus apologize for the

omission of many interesting studies which could not be dealt with in the restricted space at our disposal. We refer the interested reader to the other contributions in this book.

6.2 From coherent transport to sequential tunneling (basics)

6.2.1 Coherent transport: single-particle Green functions

Nano-scale and molecular-scale systems are naturally described by the discrete-level models, for example eigenstates of quantum dots, molecular orbitals, or atomic orbitals. But the leads are very large (infinite) and have a continuous energy spectrum. To include the lead effects systematically, it is reasonable to start from the discrete-level representation for the whole system. It can be made by the tight-binding (TB) model, which was proposed to describe quantum systems in which the localized electronic states play an essential role, it is widely used as an alternative to the plane wave description of electrons in solids, and also as a method to calculate the electronic structure of molecules in quantum chemistry.

A very effective method to describe scattering and transport is the Green function (GF) method. In the case of non-interacting systems and coherent transport single-particle GFs are used. In this section we consider the matrix Green function method for coherent transport through discrete-level systems.

(i) Matrix (tight-binding) Hamiltonian The main idea of the method is to represent the wave function of a particle as a linear combination of some known *localized* states $\psi_\alpha(\mathbf{r},\sigma)$, where α denote the set of quantum numbers, and σ is the spin index (for example, atomic orbitals, in this particular case the method is called LCAO – linear combination of atomic orbitals)

$$\psi(\xi) = \sum_\alpha c_\alpha \psi_\alpha(\xi), \tag{6.1}$$

here and below we use $\xi \equiv (\mathbf{r},\sigma)$ to denote both spatial coordinates and spin.

Using the Dirac notations $|\alpha\rangle \equiv \psi_\alpha(\xi)$ and assuming that $\psi_\alpha(\xi)$ are orthonormal functions $\langle\alpha|\beta\rangle = \delta_{\alpha\beta}$ we can write the *single-particle matrix (tight-binding) Hamiltonian* in the Hilbert space formed by $\psi_\alpha(\xi)$

$$\hat{H} = \sum_\alpha (\epsilon_\alpha + e\varphi_\alpha)|\alpha\rangle\langle\alpha| + \sum_{\alpha\beta} t_{\alpha\beta}|\alpha\rangle\langle\beta|, \tag{6.2}$$

the first term in this Hamiltonian describes the states with energies ϵ_α, φ_α is the electrical potential, the second term should be included if the states $|\alpha\rangle$ are not eigenstates of the Hamiltonoian. In the TB model $t_{\alpha\beta}$ is the hopping matrix element between states $|\alpha\rangle$ and $|\beta\rangle$, which is nonzero, as a rule, for nearest neighbor cites. The two-particle interaction is described by the Hamiltonian

$$\hat{H} = \sum_{\alpha\beta,\delta\gamma} V_{\alpha\beta,\delta\gamma}|\alpha\rangle|\beta\rangle\langle\delta|\langle\gamma|, \tag{6.3}$$

in the two-particle Hilbert space, and so on.

The energies and hopping matrix elements in this Hamiltomian can be calculated, if the single-particle

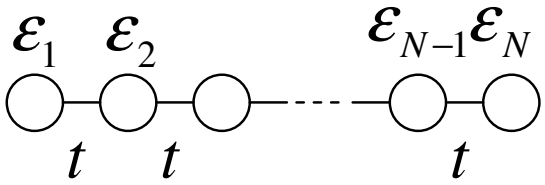


Figure 6.1: A linear chain of sites.

real-space Hamiltomian $\hat{h}(\xi)$ is known:

$$\epsilon_\alpha \delta_{\alpha\beta} + t_{\alpha\beta} = \int \psi^*_\alpha(\xi) \hat{h}(\xi) \psi_\beta(\xi) d\xi. \tag{6.4}$$

This approach was developed originally as an approximate method, if the wave functions of isolated atoms are taken as a basis wave functions $\psi_\alpha(\xi)$, but also can be formulated exactly with the help of Wannier functions. Only in the last case the expansion (6.1) and the Hamiltonian (6.2) are exact, but some extension to the arbitrary basis functions is possible. In principle, the TB model is reasonable only when *local* states can be orthogonalized. The method is useful to calculate the conductance of complex quantum systems in combination with *ab initio* methods. It is particular important to describe small molecules, when the atomic orbitals form the basis.

In the mathematical sense, the TB model is a discrete (grid) version of the continuous Schrödinger equation, thus it is routinely used in numerical calculations.

To solve the single-particle problem it is convenient to introduce a new representation, where the coefficients c_α in the expansion (6.1) are the components of a vector wave function (we assume here that all states α are numerated by integers)

$$\Psi = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix}, \tag{6.5}$$

and the eigenstates Ψ_λ are to be found from the matrix Schrödinger equation

$$\mathbf{H}\Psi_\lambda = E_\lambda \Psi_\lambda, \tag{6.6}$$

with the matrix elements of the single-particle Hamiltonian

$$H_{\alpha\beta} = \begin{cases} \epsilon_\alpha + e\varphi_\alpha, & \alpha = \beta, \\ t_{\alpha\beta}, & \alpha \neq \beta. \end{cases} \tag{6.7}$$

Now let us consider some typical systems, for which the matrix method is appropriate starting point. The simplest example is a single quantum dot, the basis is formed by the *eigenstates*, the corresponding

Hamiltonian is diagonal

$$\mathbf{H} = \begin{pmatrix} \epsilon_1 & 0 & 0 & \cdots & 0 \\ 0 & \epsilon_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \epsilon_{N-1} & 0 \\ 0 & \cdots & 0 & 0 & \epsilon_N \end{pmatrix}. \tag{6.8}$$

The next typical example is a linear chain of single-state sites with only nearest-neighbor couplings (Fig. 6.1)

$$\mathbf{H} = \begin{pmatrix} \epsilon_1 & t & 0 & \cdots & 0 \\ t & \epsilon_2 & t & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & t & \epsilon_{N-1} & t \\ 0 & \cdots & 0 & t & \epsilon_N \end{pmatrix}. \tag{6.9}$$

The method is applied as well to consider the semi-infinite leads. Although the matrices are formally infinitely-dimensional in this case, we shall show below, that the problem is reduced to the finite-dimensional problem for the quantum system of interest, and the semi-infinite leads can be integrated out.

Finally, in the second quantized form the tight-binding Hamiltonian is

$$\hat{H} = \sum_{\alpha} \left(\epsilon_\alpha + e\varphi_\alpha\right) c_\alpha^\dagger c_\alpha + \sum_{\alpha \neq \beta} t_{\alpha\beta} c_\alpha^\dagger c_\beta. \tag{6.10}$$

(ii)Matrix Green functions and contact self-energies The solution of single-particle quantum problems, formulated with the help of a matrix Hamiltonian, is possible along the usual line of finding the wave-functions on a lattice, solving the Schrödinger equation (6.6). The other method, namely matrix Green functions, considered in this section, was found to be more convenient for transport calculations, especially when interactions are included.

The retarded *single-particle* matrix Green function $\mathbf{G}^R(\epsilon)$ is determined by the equation

$$\left[(\epsilon + i\eta)\mathbf{I} - \mathbf{H}\right] \mathbf{G}^R = \mathbf{I}, \tag{6.11}$$

where η is an infinitesimally small positive number $\eta = 0^+$.

For an isolated noninteracting system the Green function is simply obtained after the matrix inversion

$$\mathbf{G}^R = \left[(\epsilon + i\eta)\mathbf{I} - \mathbf{H}\right]^{-1}. \tag{6.12}$$

Let us consider the trivial example of a two-level system with the Hamiltonian

$$\mathbf{H} = \begin{pmatrix} \epsilon_1 & t \\ t & \epsilon_2 \end{pmatrix}. \tag{6.13}$$

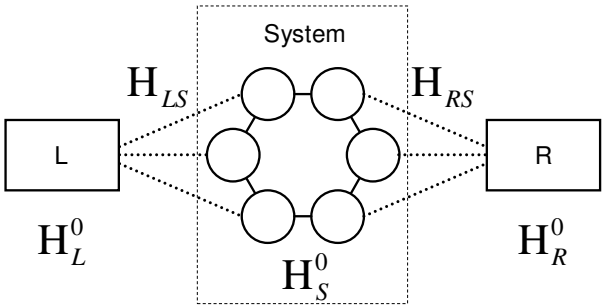


Figure 6.2: A quantum system coupled to the left and right leads.

The retarded GF is easy found to be ($\tilde{\epsilon} = \epsilon + i\eta$)

$$\mathbf{G}^R(\epsilon) = \frac{1}{(\tilde{\epsilon} - \epsilon_1)(\tilde{\epsilon} - \epsilon_2) + t^2} \begin{pmatrix} \tilde{\epsilon} - \epsilon_2 & t \\ t & \tilde{\epsilon} - \epsilon_1 \end{pmatrix}. \tag{6.14}$$

Now let us consider the case, when the system of interest is coupled to two contacts (Fig. 6.2). We assume here that the contacts are also described by the tight-binding model and by the matrix GFs. Actually, the semi-infinite contacts should be described by the matrix of infinite dimension. We shall consider the semi-infinite contacts in the next section.

Let us present the full Hamiltonian of the considered system in a following block form

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_L^0 & \mathbf{H}_{LS} & 0 \\ \mathbf{H}_{LS}^\dagger & \mathbf{H}_S^0 & \mathbf{H}_{RS}^\dagger \\ 0 & \mathbf{H}_{RS} & \mathbf{H}_R^0 \end{pmatrix}, \tag{6.15}$$

where $\mathbf{H}_L^0$, $\mathbf{H}_S^0$, and $\mathbf{H}_R^0$ are Hamiltonians of the left lead, the system, and the right lead separately. And the off-diagonal terms describe system-to-lead coupling. The Hamiltonian should be hermitian, so that

$$\mathbf{H}_{SL} = \mathbf{H}_{LS}^\dagger, \quad \mathbf{H}_{SR} = \mathbf{H}_{RS}^\dagger. \tag{6.16}$$

The Eq. (6.11) can be written as

$$\begin{pmatrix} \mathbf{E} - \mathbf{H}_L^0 & -\mathbf{H}_{LS} & 0 \\ -\mathbf{H}_{LS}^\dagger & \mathbf{E} - \mathbf{H}_S^0 & -\mathbf{H}_{RS}^\dagger \\ 0 & -\mathbf{H}_{RS} & \mathbf{E} - \mathbf{H}_R^0 \end{pmatrix} \begin{pmatrix} \mathbf{G}_L & \mathbf{G}_{LS} & 0 \\ \mathbf{G}_{SL} & \mathbf{G}_S & \mathbf{G}_{SR} \\ 0 & \mathbf{G}_{RS} & \mathbf{G}_R \end{pmatrix} = \mathbf{I}, \tag{6.17}$$

where we introduce the matrix $\mathbf{E} = (\epsilon + i\eta)\mathbf{I}$, and represent the matrix Green function in a convenient form, the notation of retarded function is omitted in intermediate formulas. Now our first goal is to find

the system Green function $\mathbf{G}_S$ which defines all quantities of interest. From the matrix equation (6.17)

$$\left(\mathbf{E}-\mathbf{H}_L^0\right)\mathbf{G}_{LS}-\mathbf{H}_{LS}\mathbf{G}_S=0, \tag{6.18}$$

$$-\mathbf{H}_{LS}^\dagger\mathbf{G}_{LS}+\left(\mathbf{E}-\mathbf{H}_S^0\right)\mathbf{G}_S-\mathbf{H}_{RS}^\dagger\mathbf{G}_{RS}=\mathbf{I}, \tag{6.19}$$

$$-\mathbf{H}_{RS}\mathbf{G}_S+\left(\mathbf{E}-\mathbf{H}_R^0\right)\mathbf{G}_{RS}=0. \tag{6.20}$$

From the first and the third equations one has

$$\mathbf{G}_{LS}=\left(\mathbf{E}-\mathbf{H}_L^0\right)^{-1}\mathbf{H}_{LS}\mathbf{G}_S, \tag{6.21}$$

$$\mathbf{G}_{RS}=\left(\mathbf{E}-\mathbf{H}_R^0\right)^{-1}\mathbf{H}_{RS}\mathbf{G}_S, \tag{6.22}$$

and substituting it into the second equation we arrive at the equation

$$\left(\mathbf{E}-\mathbf{H}_S^0-\mathbf{\Sigma}\right)\mathbf{G}_S=\mathbf{I}, \tag{6.23}$$

where we introduce the *contact self-energy* (which should be also called retarded, we omit the index in this chapter)

$$\mathbf{\Sigma}=\mathbf{H}_{LS}^\dagger\left(\mathbf{E}-\mathbf{H}_L^0\right)^{-1}\mathbf{H}_{LS}+\mathbf{H}_{RS}^\dagger\left(\mathbf{E}-\mathbf{H}_R^0\right)^{-1}\mathbf{H}_{RS}. \tag{6.24}$$

Finally, we found, that the retarded GF of a nanosystem coupled to the leads is determined by the expression

$$\mathbf{G}_S^R(\epsilon)=\left[(\epsilon+i\eta)\mathbf{I}-\mathbf{H}_S^0-\mathbf{\Sigma}\right]^{-1}, \tag{6.25}$$

the effects of the leads are included through the self-energy.

Here we should stress the important property of the self-energy (6.24), it is determined only by the coupling Hamiltonians and the retarded GFs of the *isolated* leads $\mathbf{G}_i^{0R}=\left(\mathbf{E}-\mathbf{H}_R^0\right)^{-1}$ $(i=L,R)$

$$\mathbf{\Sigma}_i=\mathbf{H}_{iS}^\dagger\left(\mathbf{E}-\mathbf{H}_i^0\right)^{-1}\mathbf{H}_{iS}=\mathbf{H}_{iS}^\dagger\mathbf{G}_i^{0R}\mathbf{H}_{iS}, \tag{6.26}$$

it means, that the contact self-energy is independent of the state of the nanosystem itself and describes completely the influence of the leads. Later we shall see that this property conserves also for interacting system, if the leads are noninteracting.

Finally, we should note, that the Green functions considered in this section, are *single-particle* GFs, and can be used only for noninteracting systems.

(iii) Semi-infinite leads Let us consider now a nanosystem coupled to a semi-infinite lead (Fig. 6.3). The direct matrix inversion can not be performed in this case. The spectrum of a semi-infinite system is continuous. We should transform the expression (6.26) into some other form.

To proceed, we use the relation between the Green function and the eigenfunctions Ψ_λ of a system, which are solutions of the Schrödinger equation (6.6). Let us define $\Psi_\lambda(\alpha)\equiv c_\lambda$ in the eigenstate $|\lambda\rangle$ in

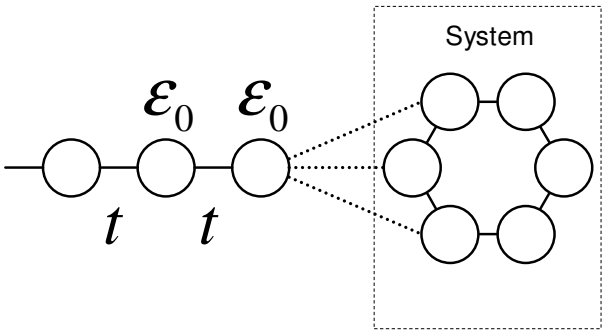


Figure 6.3: A quantum system coupled to a semi-infinite 1D lead.

the sense of definition (6.5), then

$$G^R_{\alpha\beta}(\epsilon) = \sum_\lambda \frac{\Psi_\lambda(\alpha)\Psi^*_\lambda(\beta)}{\epsilon + i\eta - E_\lambda}, \tag{6.27}$$

where α is the TB state (site) index, λ denotes the eigenstate, E_λ is the energy of the eigenstate. The summation in this formula can be easy replaced by the integration in the case of a continuous spectrum. It is important to notice, that the eigenfunctions $\Psi_\lambda(\alpha)$ should be calculated for the separately taken semi-infinite lead, because the Green function of isolated lead is substituted into the contact self-energy.

For example, for the semi-infinite 1D chain of single-state sites ($n, m = 1, 2, ...$)

$$G^R_{nm}(\epsilon) = \int_{-\pi}^{\pi} \frac{dk}{2\pi} \frac{\Psi_k(n)\Psi^*_k(m)}{\epsilon + i\eta - E_k}, \tag{6.28}$$

with the eigenfunctions $\Psi_k(n) = \sqrt{2}\sin kn$, $E_k = \epsilon_0 + 2t\cos k$.

Let us consider a simple situation, when the nanosystem is coupled only to the end site of the 1D lead (Fig. 6.3). From (6.26) we obtain the matrix elements of the self-energy

$$\Sigma_{\alpha\beta} = V^*_{1\alpha} V_{1\beta} G^{0R}_{11}, \tag{6.29}$$

where the matrix element $V_{1\alpha}$ describes the coupling between the end site of the lead ($n = m = 1$) and the state $|\alpha\rangle$ of the nanosystem.

To make clear the main physical properties of the lead self-energy, let us analyze in detail the semi-infinite 1D lead with the Green function (6.28). The integral can be calculated analytically (Datta II, p. 213, [57])

$$G^R_{11}(\epsilon) = \frac{1}{\pi}\int_{-\pi}^{\pi} \frac{\sin^2 k dk}{\epsilon + i\eta - \epsilon_0 - 2t\cos k} = -\frac{\exp(iK(\epsilon))}{t}, \tag{6.30}$$

$K(\epsilon)$ is determined from $\epsilon = \epsilon_0 + 2t\cos K$. Finally, we obtain the following expressions for the real and

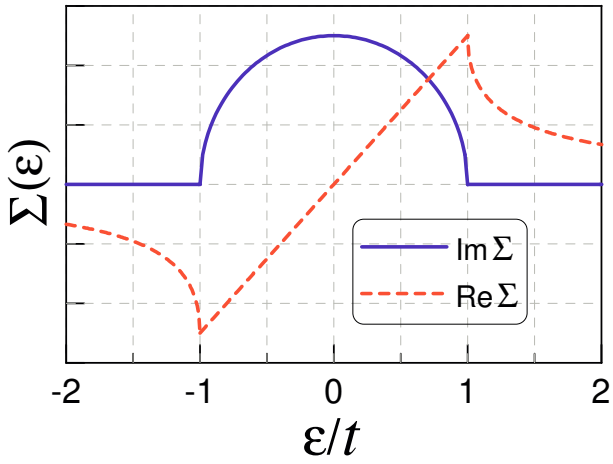


Figure 6.4: Real and imaginary parts of the contact self-energy as a function of energy for a one-band one-dimensional lead.

imaginary part of the self-energy

$$\mathrm{Re}\Sigma_{\alpha\alpha} = \frac{|V_{1\alpha}|^2}{t}\left(\kappa - \sqrt{\kappa^2-1}\,[\theta(\kappa-1) - \theta(-\kappa-1)]\right), \tag{6.31}$$

$$\mathrm{Im}\Sigma_{\alpha\alpha} = -\frac{|V_{1\alpha}|^2}{t}\sqrt{1-\kappa^2}\theta(1-|\kappa|), \tag{6.32}$$

$$\kappa = \frac{\epsilon - \epsilon_0}{2t}. \tag{6.33}$$

The real and imaginary parts of the self-energy, given by these expressions, are shown in Fig. 6.4. There are several important general conclusion that we can make looking at the formulas and the curves.

(a) The self-energy is a complex function, the real part describes the energy shift of the level, and the imaginary part describes broadening. The *finite* imaginary part appears as a result of the continuous spectrum in the leads. The broadening is described traditionally by the matrix

$$\mathbf{\Gamma} = i\left(\mathbf{\Sigma} - \mathbf{\Sigma}^\dagger\right), \tag{6.34}$$

called *level-width function*.

(b) In the wide-band limit ($t \to \infty$), at the energies $\epsilon - \epsilon_0 \ll t$, it is possible to neglect the real part of the self-energy, and the only effect of the leads is level broadening. So that the self-energy of the left (right) lead is

$$\mathbf{\Sigma}_{L(R)} = -i\frac{\mathbf{\Gamma}_{L(R)}}{2}. \tag{6.35}$$

(iv) Transmission, conductance, current After all, we want again to calculate the current through the nanosystem. We assume, as before, that the contacts are equilibrium, and there is the voltage V applied between the left and right contacts. The calculation of the current in a general case is more

convenient to perform using the full power of the nonequilibrium Green function method. Here we present a simplified approach, valid for noninteracting systems only, following Paulsson [58].

Let us come back to the Schrödinger equation (6.6) in the matrix representation, and write it in the following form

$$\begin{pmatrix} \mathbf{H}_L^0 & \mathbf{H}_{LS} & 0 \\ \mathbf{H}_{LS}^\dagger & \mathbf{H}_S^0 & \mathbf{H}_{RS}^\dagger \\ 0 & \mathbf{H}_{RS} & \mathbf{H}_R^0 \end{pmatrix} \begin{pmatrix} \Psi_L \\ \Psi_S \\ \Psi_R \end{pmatrix} = E \begin{pmatrix} \Psi_L \\ \Psi_S \\ \Psi_R \end{pmatrix}, \tag{6.36}$$

where Ψ_L, Ψ_S, and Ψ_R are vector wave functions of the left lead, the nanosystem, and the right lead correspondingly.

Now we find the solution in the scattering form (which is difficult to call true scattering because we do not define explicitly the geometry of the leads). Namely, in the left lead $\Psi_L = \Psi_L^0 + \Psi_L^1$, where Ψ_L^0 is the eigenstate of $\mathbf{H}_L^0$, and is considered as known initial wave. The "reflected" wave Ψ_L^1, as well as the transmitted wave in the right lead Ψ_R, appear only as a result of the interaction between subsystems. The main trick is, that we find a *retarded* solution.

Solving the equation (6.36) with these conditions, the solution is

$$\Psi_L = \left(1 + \mathbf{G}_L^{0R}\mathbf{H}_{LS}\mathbf{G}_S^R\mathbf{H}_{LS}^\dagger\right)\Psi_L^0, \tag{6.37}$$

$$\Psi_R = \mathbf{G}_R^{0R}\mathbf{H}_{RS}\mathbf{G}_S^R\mathbf{H}_{LS}^\dagger\Psi_L^0 \tag{6.38}$$

$$\Psi_S = \mathbf{G}_S^R\mathbf{H}_{LS}^\dagger\Psi_L^0. \tag{6.39}$$

The physical sense of this expressions is quite transparent, they describe the quantum amplitudes of the scattering processes. Three functions Ψ_L, Ψ_S, and Ψ_R are equivalent together to the scattering state in the Landauer-Büttiker theory. Note, that $\mathbf{G}_S^R$ here is the full GF of the nanosystem including the lead self-energies.

Now the next step. We want to calculate the current. The partial (for some particular eigenstate $\Psi_{L\lambda}^0$) current from the lead to the system is

$$j_{i=L,R} = \frac{ie}{\hbar}\left(\Psi_i^\dagger\mathbf{H}_{iS}\Psi_S - \Psi_S^\dagger\mathbf{H}_{iS}^\dagger\Psi_i\right). \tag{6.40}$$

To calculate the total current we should substitute the expressions for the wave functions (6.37)-(6.39), and summarize all contributions [58]. As a result the Landauer formula is obtained. We present the calculation for the transmission function. First, after substitution of the wave functions we have for the partial current going through the system

$$\begin{aligned} j_\lambda = j_L = -j_R = &- \frac{ie}{\hbar}\left(\Psi_R^\dagger\mathbf{H}_{RS}\Psi_S - \Psi_S^\dagger\mathbf{H}_{RS}^\dagger\Psi_R\right) = \\ &- \frac{ie}{\hbar}\left(\Psi_L^{0\dagger}\mathbf{H}_{LS}\mathbf{G}_S^A\mathbf{H}_{RS}^\dagger\left(\mathbf{G}_R^{0\dagger} - \mathbf{G}_R^0\right)\mathbf{H}_{RS}\mathbf{G}_S^R\mathbf{H}_{LS}^\dagger\Psi_L^0\right) = \\ &\frac{e}{\hbar}\left(\Psi_L^{0\dagger}\mathbf{H}_{LS}\mathbf{G}_S^A\mathbf{\Gamma}_R\mathbf{G}_S^R\mathbf{H}_{LS}^\dagger\Psi_L^0\right). \end{aligned} \tag{6.41}$$

The full current of all possible left eigenstates is given by

$$I = \sum_\lambda j_\lambda = \sum_\lambda \frac{e}{\hbar} \left(\Psi^{0\dagger}_{L\lambda} \mathbf{H}_{LS} \mathbf{G}^A_S \mathbf{\Gamma}_R \mathbf{G}^R_S \mathbf{H}^\dagger_{LS} \Psi^0_{L\lambda} \right) f_L(E_\lambda), \tag{6.42}$$

the distribution function $f_L(E_\lambda)$ describes the population of the left states, the distribution function of the right lead is absent here, because we consider only the current from the left to the right.

The same current is given by the Landauer formula through the transmission function $\bar{T}(E)$

$$I = \frac{e}{\hbar} \int_{-\infty}^{\infty} \overline{T}(E) f_L(E) dE. \tag{6.43}$$

If one compares these two expressions for the current, the transmission function at some energy is obtained as

$$\begin{aligned} \overline{T}(E) &= 2\pi \sum_\lambda \delta(E - E_\lambda) \left(\Psi^{0\dagger}_{L\lambda} \mathbf{H}_{LS} \mathbf{G}^A_S \mathbf{\Gamma}_R \mathbf{G}^R_S \mathbf{H}^\dagger_{LS} \Psi^0_{L\lambda} \right) \\ &= 2\pi \sum_\lambda \sum_\delta \delta(E - E_\lambda) \left(\Psi^{0\dagger}_{L\lambda} \mathbf{H}_{LS} \Psi_\delta \right) \left(\Psi^\dagger_\delta \mathbf{G}^A_S \mathbf{\Gamma}_R \mathbf{G}^R_S \mathbf{H}^\dagger_{LS} \Psi^0_{L\lambda} \right) \\ &= \sum_\delta \left(\Psi^\dagger_\delta \mathbf{G}^A_S \mathbf{\Gamma}_R \mathbf{G}^R_S \mathbf{H}^\dagger_{LS} \left(2\pi \sum_\lambda \delta(E - E_\lambda) \Psi^0_{L\lambda} \Psi^{0\dagger}_{L\lambda} \right) \mathbf{H}_{LS} \Psi_\delta \right) \\ &= \mathrm{Tr} \left(\mathbf{\Gamma}_L \mathbf{G}^A_S \mathbf{\Gamma}_R \mathbf{G}^R_S \right). \end{aligned} \tag{6.44}$$

To evaluate the sum in brackets we used the eigenfunction expansion (6.27) for the left contact.

We obtained the new representation for the transmission formula, which is very convenient for numerical calculations

$$\overline{T} = \mathrm{Tr} \left(\hat{t} \hat{t}^\dagger \right) = \mathrm{Tr} \left(\mathbf{\Gamma}_L \mathbf{G}^A \mathbf{\Gamma}_R \mathbf{G}^R \right). \tag{6.45}$$

Finally, one important remark, at finite voltage the diagonal energies in the Hamiltonians $\mathbf{H}^0_L$, $\mathbf{H}^0_S$, and $\mathbf{H}^0_R$ are shifted $\epsilon_\alpha \to \epsilon_\alpha + e\varphi_\alpha$. Consequently, the energy dependencies of the self-energies defined by (6.26) are also changed and the lead self-energies are voltage dependent. However, it is convenient to define the self-energies using the Hamiltonians at zero voltage, in that case the voltage dependence should be explicitly shown in the transmission formula

$$\overline{T}(E) = \mathrm{Tr} \left[\mathbf{\Gamma}_L(E - e\varphi_L) \mathbf{G}^R(\epsilon) \mathbf{\Gamma}_R(E - e\varphi_R) \mathbf{G}^A(\epsilon) \right], \tag{6.46}$$

where φ_R and φ_L are electrical potentials of the right and left leads.

With known transmission function, the current I at finite voltage V can be calculated by the usual Landauer-Bütiker formulas (without spin degeneration, otherwise it should be multiplied additionally by 2)

$$I(V) = \frac{e}{\hbar} \int_{-\infty}^{\infty} \overline{T}(E) \left[f_L(E) - f_R(E) \right] dE, \tag{6.47}$$

where the equilibrium distribution functions of the contacts should be written with corresponding chemical potentials μ_i, and electrical potentials φ_i

$$f_L(E) = \frac{1}{\exp\left(\frac{E-\mu_L-e\varphi_L}{T}\right)+1}, \quad f_R(E) = \frac{1}{\exp\left(\frac{E-\mu_R-e\varphi_R}{T}\right)+1}. \tag{6.48}$$

The zero-voltage conductance G is

$$G = \left.\frac{dI}{dV}\right|_{V=0} = -\frac{e^2}{h}\int_{-\infty}^{\infty}\overline{T}(E)\frac{\partial f^0(E)}{\partial E}dE, \tag{6.49}$$

where $f^0(E)$ is the equilibriumfermi-function

$$f^0(E) = \frac{1}{\exp\left(\frac{E-\mu}{T}\right)+1}. \tag{6.50}$$

6.2.2 Interacting nanosystems and master equation method

The single-particle matrix Green function methods, considered in the previous section, can be applied only in the case of *noninteracting* electrons and without inelastic scattering. In the case of interacting systems, the other approach, known as the method of tunneling (or transfer) Hamiltonian (TH), plays an important role, and is widely used to describe tunneling in superconductors, in ferromagnets, effects in small tunnel junctions such as Coulomb blockade (CB), etc. The main advantage of this method is that it is easely combined with powerful methods of many-body theory. Besides, it is very convenient even for noninteracting electrons, when the coupling between subsystems is weak, and the tunneling process can be described by rather simple matrix elements.

(i) Tunneling (transfer) Hamiltonian The main idea is to represent the Hamiltonian of the system (we consider first a single contact between two subsystems) as a sum of three parts: "left" $\hat{H}_L$, "right" $\hat{H}_R$, and "tunneling" $\hat{H}_T$

$$\hat{H} = \hat{H}_L + \hat{H}_R + \hat{H}_T, \tag{6.51}$$

$\hat{H}_L$ and $\hat{H}_R$ determine "left" $|Lk\rangle$ and "right" $|Rq\rangle$ states

$$\hat{H}_L\psi_k(\xi) = E_k\psi_k(\xi), \tag{6.52}$$

$$\hat{H}_R\psi_q(\xi) = E_q\psi_q(\xi), \tag{6.53}$$

below in this lecture we use the index k for left states and the index q for right states. $\hat{H}_T$ determines "transfer" between these states and is *defined* through matrix elements $V_{kq} = \langle Lk|\hat{H}_T|Rq\rangle$. With these definitions the single-particle tunneling Hamiltonian is

$$\hat{H} = \sum_{k\in L} E_k|k\rangle\langle k| + \sum_{q\in R} E_q|q\rangle\langle q| + \sum_{kq}\left[V_{qk}|q\rangle\langle k| + V^*_{qk}|k\rangle\langle q|\right]. \tag{6.54}$$

The method of the tunneling Hamiltonian was introduced by Bardeen [59], developed by Harrison [60], and formulated in most familiar second quantized form by Cohen, Falicov, and Phillips [61]. In spite of many very successful applications of the TH method, it was many times criticized for it's phenomenological character and incompleteness, beginning from the work of Prange [62]. However, in the same work Prange showed that the tunneling Hamiltonian is well defined in the sense of the perturbation theory. These developments and discussions were summarized by Duke [63]. Note, that the formulation equivalent to the method of the tunneling Hamiltonian can be derived exactly from the tight-binding approach.

Indeed, the tight-binding model assumes that the left and right states can be clearly separated, also when they are orthogonal. The difference with the continuous case is, that we restrict the Hilbert space introducing the tight-binding model, so that the solution is not exact in the sense of the continuous Schrödinger equation. But, in fact, we only consider physically relevant states, neglecting high-energy states not participating in transport.

Compare the tunneling Hamiltonian (6.54) and the tight-binding Hamiltonian (6.2), divided into left and right parts

$$\hat{H} = \sum_{\alpha\beta\in L} \tilde{\epsilon}_{\alpha\beta}|\alpha\rangle\langle\beta| + \sum_{\delta\gamma\in R} \tilde{\epsilon}_{\delta\gamma}|\delta\rangle\langle\gamma| + \sum_{\alpha\in L,\,\delta\in R} \left[V_{\delta\alpha}|\delta\rangle\langle\alpha| + V^*_{\delta\alpha}|\alpha\rangle\langle\delta|\right]. \tag{6.55}$$

The first two terms are the Hamiltonians of the left and right parts, the third term describes the left-right (tunneling) coupling. The equivalent matrix representation of this Hamiltonian is

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}^0_L & \mathbf{H}_{LR} \\ \mathbf{H}^\dagger_{LR} & \mathbf{H}^0_R \end{pmatrix}. \tag{6.56}$$

The Hamiltonians (6.54) and (6.55) are essentially the same, only the first one is written in the eigenstate basis $|k\rangle$, $|q\rangle$, while the second in the tight-binding basis $|\alpha\rangle$, $|\beta\rangle$ of the left lead and $|\delta\rangle$, $|\gamma\rangle$ of the right lead. Now we want to transform the TB Hamiltonian (6.55) into the eigenstate representation.

Canonical transformations from the tight-binding (atomic orbitals) representation to the eigenstate (molecular orbitals) representation play an important role, and we consider it in detail. Assume, that we find two unitary matrices $\mathbf{S}_L$ and $\mathbf{S}_R$, such that the Hamiltonians of the left part $\mathbf{H}^0_L$ and of the right part $\mathbf{H}^0_R$ can be diagonalized by the canonical transformations

$$\bar{\mathbf{H}}^0_L = \mathbf{S}^{-1}_L \mathbf{H}^0_L \mathbf{S}_L, \tag{6.57}$$

$$\bar{\mathbf{H}}^0_R = \mathbf{S}^{-1}_R \mathbf{H}^0_R \mathbf{S}_R. \tag{6.58}$$

The left and right eigenstates can be written as

$$|k\rangle = \sum_\alpha S_{Lk\alpha}|\alpha\rangle, \tag{6.59}$$

$$|q\rangle = \sum_\delta S_{Rq\delta}|\delta\rangle, \tag{6.60}$$

and the first two free-particle terms of the Hamiltonian (6.54) are reproduced. The tunneling terms are transformed as

$$\bar{\mathbf{H}}_{LR} = \mathbf{S}_L^{-1}\mathbf{H}_{LR}\mathbf{S}_R, \tag{6.61}$$

$$\bar{\mathbf{H}}_{LR}^{\dagger} = \mathbf{S}_R^{-1}\mathbf{H}_{LR}^{\dagger}\mathbf{S}_L, \tag{6.62}$$

or explicitly

$$\sum_{\alpha\in L,\,\delta\in R} V_{\delta\alpha}|\delta\rangle\langle\alpha| = \sum_{kq} V_{qk}|q\rangle\langle k|, \tag{6.63}$$

where

$$V_{qk} = \sum_{\alpha\in L,\,\delta\in R} V_{\delta\alpha}S_{L\alpha k}S_{R\delta q}. \tag{6.64}$$

The last expression solve the problem of transformation of the tight-binding matrix elements into tunneling matrix elements.

For applications the tunneling Hamiltonian (6.54) should be formulated in the second quantized form. We introduce creation and annihilation *Schrödinger* operators $c_{Lk}^{\dagger}$, c_{Lk}, $c_{Rq}^{\dagger}$, c_{Rq}. Using the usual rules we obtain

$$\hat{H} = \hat{H}_L\left(\left\{c_k^{\dagger}; c_k\right\}\right) + \hat{H}_R\left(\left\{c_q^{\dagger}; c_q\right\}\right) + \hat{H}_T\left(\left\{c_k^{\dagger}; c_k; c_q^{\dagger}; c_q\right\}\right), \tag{6.65}$$

$$\hat{H} = \sum_k(\epsilon_k + e\varphi_L(t))c_k^{\dagger}c_k + \sum_q(\epsilon_q + e\varphi_R(t))c_q^{\dagger}c_q + \sum_{kq}\left[V_{qk}c_q^{\dagger}c_k + V_{qk}^{*}c_k^{\dagger}c_q\right]. \tag{6.66}$$

It is assumed that left c_k and right c_q operators describe independent states and are anticommutative. For nonorthogonal states of the Hamiltonian $\hat{H}_l + \hat{H}_R$ it is not exactly so. But if we consider $\hat{H}_L$ and $\hat{H}_R$ as two independent Hamiltonians with independent Hilbert spaces we resolve this problem. Thus we again should consider (6.66) not as a true Hamiltonian, but as the formal expression describing the current between left and right states. In the weak coupling case the small corrections to the commutation relations are of the order of $|V_{qk}|$ and can be neglected. If the tight-binding formulation is possible, (6.66) is exact within the framework of this formulation. In general the method of tunneling Hamiltonian can be considered as a *phenomenological* microscopic approach, which was proved to give reasonable results in many cases, e.g. in description of tunneling between superconductors and Josephson effect.

(ii) Tunneling current The current from the state k into the state q is given by the golden rule

$$J_{k\to q} = e\Gamma_{qk} = \frac{2\pi e}{\hbar}|V_{qk}|^2 f_L(k)\,(1 - f_R(q))\,\delta(E_k - E_q), \tag{6.67}$$

the probability $(1 - f_R(E_q))$ that the right state is unoccupied should be included, it is different from the scattering approach because left and right states are two independent states!

Then we write the total current as the sum of all partial currents from left states to right states and

vice versa (note that the terms $f_L(k)f_R(q)$ are cancelled)

$$J = \frac{2\pi e}{\hbar} \sum_{kq} |V_{qk}|^2 \left[f(k) - f(q) \right] \delta(E_q - E_k). \tag{6.68}$$

For tunneling between two equilibrium leads distribution functions are simply Fermi-Dirac functions (6.48) and current can be finally written in the well known form (To do this one should multiply the integrand on $1 = \int \delta(E - E_q) dE$.)

$$J = \frac{e}{h} \int_{-\infty}^{\infty} T(E, V) \left[f_L(E) - f_R(E) \right] dE, \tag{6.69}$$

with

$$T(E, V) = (2\pi)^2 \sum_{qk} |V_{kq}|^2 \delta(E - E_k - e\varphi_L) \delta(E - E_q - e\varphi_R). \tag{6.70}$$

This expression is equivalent to the Landauer formula (6.47), but the transmission function is related now to the tunneling matrix element.

Now let us calculate the tunneling current as the time derivative of the number of particles operator in the left lead $\hat{N}_L = \sum_k c_k^\dagger c_k$. Current from the left to right contact is

$$J(t) = -e \left\langle \left(\frac{dN_L}{dt} \right) \right\rangle_S = -\frac{ie}{\hbar} \left\langle \left[\hat{H}_T, N_L \right]_- \right\rangle_S, \tag{6.71}$$

where $\langle ... \rangle_S$ is the average over time-dependent Schrödinger state. $\hat{N}_L$ commute with both left and right Hamiltonians, but not with the tunneling Hamiltonian

$$\left[\hat{H}_T, N_L \right]_- = \sum_{k'} \sum_{kq} \left[\left(V_{qk} c_q^\dagger c_k + V_{qk}^* c_q c_k^\dagger \right) c_{k'}^\dagger c_{k'} \right]_-, \tag{6.72}$$

using commutation relations

$$c_k c_{k'}^\dagger c_{k'} - c_{k'}^\dagger c_{k'} c_k = c_k c_{k'}^\dagger c_{k'} + c_{k'}^\dagger c_k c_{k'} = (c_k c_{k'}^\dagger + \delta_{kk'} - c_k c_{k'}^\dagger) c_{k'} = \delta_{kk'} c_k,$$

we obtain

$$J(t) = \frac{ie}{\hbar} \sum_{kq} \left[V_{qk} \left\langle c_q^\dagger c_k \right\rangle_S - V_{qk}^* \left\langle c_k^\dagger c_q \right\rangle_S \right]. \tag{6.73}$$

Now we switch to the Heisenberg picture, and average over initial time-independent *equilibrium* state

$$\left\langle \hat{O}(t) \right\rangle = Sp \left(\hat{\rho}_{eq} \hat{O}(t) \right), \quad \hat{\rho}_{eq} = \frac{e^{-H_{eq}/T}}{Sp \left(e^{-H_{eq}/T} \right)}. \tag{6.74}$$

One obtains

$$J(t) = \frac{ie}{\hbar} \sum_{kq} \left[V_{qk} \left\langle c_q^\dagger(t) c_k(t) \right\rangle - V_{qk}^* \left\langle c_k^\dagger(t) c_q(t) \right\rangle \right]. \tag{6.75}$$

It can be finally written as

$$J(t) = \frac{2e}{\hbar}\mathrm{Im}\left(\sum_{kq} V_{qk}\rho_{kq}(t)\right) = \frac{2e}{\hbar}\mathrm{Re}\left(\sum_{kq} V_{qk}G^{<}_{kq}(t,t)\right).$$

We define "left-right" density matrix or more generally lesser Green function

$$G^{<}_{kq}(t_1,t_2) = i\left\langle c^{\dagger}_q(t_2)c_k(t_1)\right\rangle.$$

Later we show that these expressions for the tunneling current give the same answer as was obtained above by the golden rule in the case of noninteracting leads.

(iii) Sequential tunneling and the master equation Let us come back to our favorite problem – transport through a quantum system. There is one case (called *sequential tunneling*), when the simple formulas discussed above can be applied even in the case of resonant tunneling

Assume that a noninteracting nanosystem is coupled weakly to a thermal bath (in addition to the leads). The effect of the thermal bath is to break phase coherence of the electron inside the system during some time τ_{ph}, called *decoherence or phase-breaking time*. τ_{ph} is an important time-scale in the theory, it should be compared with the so-called "tunneling time" – the characteristic time for the electron to go from the nanosystem to the lead, which can be estimated as an inverse level-width function Γ^{-1}. So that the criteria of sequential tunneling is

$$\Gamma\tau_{ph} \ll 1. \tag{6.76}$$

The finite decoherence time is due to some inelastic scattering mechanism inside the system, but typically this time is shorter than the energy relaxation time τ_ϵ, and the distribution function of electrons inside the system can be nonequilibrium (if the finite voltage is applied), this transport regime is well known in semiconductor superlattices and quantum-cascade structures.

In the sequential tunneling regime the tunneling events between the left lead and the nanosystem and between the left lead and the nanosystem are independent and the current from the left (right) lead to the nanosystem is given by the golden rule expression (6.68). Let us modify it to the case of tunneling from the lead to a *single level* $|\alpha\rangle$ of a quantum system

$$J = \frac{2\pi e}{\hbar}\sum_k |V_{\alpha k}|^2\left[f(k) - P_\alpha\right]\delta(E_\alpha - E_k), \tag{6.77}$$

where we introduce the probability P_α to find the electron in the state $|\alpha\rangle$ with the energy E_α.

(iv) Rate equations for noninteracting systems Rate equation method is a simple approach base on the balance of incoming and outgoing currents. Assuming that the contacts are equilibrium we obtain for the left and right currents

$$J_{i=L(R)} = e\Gamma_{i\alpha}\left[f^0_i(E_\alpha) - P_\alpha\right], \tag{6.78}$$

where

$$\Gamma_{i\alpha} = \frac{2\pi}{\hbar} \sum_k |V_{\alpha k}|^2 \delta(E_\alpha - E_k). \tag{6.79}$$

In the stationary state $J = J_L = -J_R$, and from this condition the level population P_α is found to be

$$P_\alpha = \frac{\Gamma_{L\alpha} f_L^0(E_\alpha) + \Gamma_{R\alpha} f_R^0(E_\alpha)}{\Gamma_{L\alpha} + \Gamma_{R\alpha}}, \tag{6.80}$$

with the current

$$J = e \frac{\Gamma_{L\alpha}\Gamma_{R\alpha}}{\Gamma_{L\alpha} + \Gamma_{R\alpha}} \big(f_L^0(E_\alpha) - f_R^0(E_\alpha) \big). \tag{6.81}$$

It is interesting to note that this expression is exactly the same, as one can obtain for the resonant tunneling through a single level without any scattering. It should be not forgotten, however, that we did not take into account additional level broadening due to scattering.

(v) Master equation for interacting systems Now let us formulate briefly a more general approach to transport through interacting nanosystems weakly coupled to the leads in the sequential tunneling regime, namely the master equation method. Assume, that the system can be in several states $|\lambda\rangle$, which are the eigenstates of an isolated system and introduce the distribution function P_λ – the probability to find the system in the state $|\lambda\rangle$. Note, that these states are *many-particle* states, for example for a two-level quantum dot the possible states are $|\lambda\rangle = |00\rangle$, $|10\rangle$, $01|\rangle$, and $|11\rangle$. The first state is empty dot, the second and the third with one electron, and the last one is the double occupied state. The other non-electronic degrees of freedom can be introduce on the same ground in this approach. The only restriction is that some full set of eigenstates should be used

$$\sum_\lambda P_\lambda = 1. \tag{6.82}$$

The next step is to treat tunneling as a perturbation. Following this idea, the transition rates $\Gamma^{\lambda\lambda'}$ from the state λ' to the state λ are calculated using the Fermi golden rule

$$\Gamma^{fi} = \frac{2\pi}{\hbar} \left| \left\langle f | \hat{H}_T | i \right\rangle \right|^2 \delta(E_f - E_i). \tag{6.83}$$

Then, the kinetic (master) equation can be written as

$$\frac{dP_\lambda}{dt} = \sum_{\lambda'} \Gamma^{\lambda\lambda'} P_{\lambda'} - \sum_{\lambda'} \Gamma^{\lambda'\lambda} P_\lambda, \tag{6.84}$$

where the first term describes tunneling transition *into the state* $|\lambda\rangle$, and the second term – tunneling transition *out of the state* $|\lambda\rangle$.

In the stationary case the probabilities are determined from

$$\sum_{\lambda'} \Gamma^{\lambda\lambda'} P_{\lambda'} = \sum_{\lambda'} \Gamma^{\lambda'\lambda} P_{\lambda}. \tag{6.85}$$

For noninteracting electrons the transition rates are determined by the single-electron tunneling rates, and are nonzero only for the transitions between the states with the number of electrons different by one. For example, transition from the state $|\lambda'\rangle$ with empty electron level α into the state $|\lambda\rangle$ with filled state α is described by

$$\Gamma^{n_\alpha=1\ n_\alpha=0} = \Gamma_{L\alpha} f_L^0(E_\alpha) + \Gamma_{R\alpha} f_R^0(E_\alpha), \tag{6.86}$$

where $\Gamma_{L\alpha}$ and $\Gamma_{R\alpha}$ are left and right level-width functions (6.79).

For interacting electrons the calculation is a little bit more complicated. One should establish the relation between *many-particle* eigenstates of the system and *single-particle* tunneling. To do this, let us note, that the states $|f\rangle$ and $|i\rangle$ in the golden rule formula (6.83) are actually the states of the whole system, including the leads. We denote the initial and final states as

$$|i\rangle = |\hat{k}_i, \lambda'\rangle = |\hat{k}_i\rangle|\lambda'\rangle, \tag{6.87}$$

$$|f\rangle = |\hat{k}_f, \lambda\rangle = |\hat{k}_f\rangle|\lambda\rangle, \tag{6.88}$$

where $\hat{k}$ is the occupation of the single-particle states in the lead. The parameterization is possible, because we apply the perturbation theory, and isolated lead and nanosystem are independent.

The important point is, that the leads are actually in the equilibrium mixed state, the single electron states are populated with probabilities, given by the Fermi-Dirac distribution function. Taking into account all possible single-electron tunneling processes, we obtain the incoming tunneling rate

$$\Gamma_{in}^{\lambda\lambda'} = \frac{2\pi}{\hbar} \sum_{ik\sigma} f_i^0(E_{ik\sigma}) \left| \left\langle i\bar{k}, \lambda \left| \bar{H}_T \right| ik, \lambda' \right\rangle \right|^2 \delta(E_{\lambda'} + E_{ik\sigma} - E_\lambda), \tag{6.89}$$

where we use the short-hand notations: $|ik, \lambda'\rangle$ is the state with occupied k-state in the $i-$th lead, while $|i\bar{k}, \lambda\rangle$ is the state with unoccupied k-state in the $i-$th lead, and all other states are assumed to be unchanged, E_λ is the energy of the state λ .

To proceed, we introduce the following Hamiltonian, describing single electron tunneling and charging of the nanosystem state

$$\hat{H}_T = \sum_{k\lambda\lambda'} \left[V_{\lambda\lambda' k} c_k X^{\lambda\lambda'} + V^*_{\lambda\lambda' k} c_k^\dagger X^{\lambda'\lambda} \right], \tag{6.90}$$

the Hubbard operators $X^{\lambda\lambda'} = |\lambda\rangle\langle\lambda'|$ describe transitions between eigenstates of the nanosystem.

Substituting this Hamiltonian one obtains

$$\Gamma_{in}^{\lambda\lambda'} = \frac{2\pi}{\hbar} \sum_{ik\sigma} f_i^0(E_{ik\sigma}) \left| V_{ik\sigma} \right|^2 \left| V_{\lambda\lambda' k} \right|^2 \delta(E_{\lambda'} + E_{ik\sigma} - E_\lambda). \tag{6.91}$$

In the important limiting case, when the matrix element $V_{\lambda\lambda' k}$ is k-independent, the sum over k can be performed, and finally

$$\Gamma_{in}^{\lambda\lambda'} = \sum_{i=L,R} \Gamma_i(E_\lambda - E_{\lambda'}) \left|V_{\lambda\lambda'}\right|^2 f_i^0(E_\lambda - E_{\lambda'}). \tag{6.92}$$

Similarly, the outgoing rate is

$$\Gamma_{out}^{\lambda\lambda'} = \sum_{i=L,R} \Gamma_i(E_{\lambda'} - E_\lambda) \left|V_{\lambda\lambda'}\right|^2 \left(1 - f_i^0(E_{\lambda'} - E_\lambda)\right). \tag{6.93}$$

The current (from the left or right lead to the system) is

$$J_{i=L,R}(t) = e \sum_{\lambda\lambda'} \left(\Gamma_{i\,in}^{\lambda\lambda'} P_{\lambda'} - \Gamma_{i\,out}^{\lambda\lambda'} P_{\lambda'}\right). \tag{6.94}$$

This system of equations solves the transport problem in the sequential tunneling regime.

6.2.2.1 Electron-electron interaction and Coulomb blockade

(i) Anderson-Hubbard and constant-interaction models To take into account both discrete energy levels of a system and the electron-electron interaction, it is convenient to start from the general Hamiltonian

$$\hat{H} = \sum_{\alpha\beta} \tilde{\epsilon}_{\alpha\beta} d_\alpha^\dagger d_\beta + \frac{1}{2} \sum_{\alpha\beta\gamma\delta} V_{\alpha\beta,\gamma\delta} d_\alpha^\dagger d_\beta^\dagger d_\gamma d_\delta. \tag{6.95}$$

The first term of this Hamiltonian is a free-particle discrete-level model (6.10) with $\tilde{\epsilon}_{\alpha\beta}$ including electrical potentials. And the second term describes all possible interactions between electrons and is equivalent to the real-space Hamiltonian

$$\hat{H}_{ee} = \frac{1}{2} \int d\xi \int d\xi' \hat{\psi}^\dagger(\xi) \hat{\psi}^\dagger(\xi') V(\xi, \xi') \hat{\psi}(\xi') \hat{\psi}(\xi), \tag{6.96}$$

where $\hat{\psi}(\xi)$ are field operators

$$\hat{\psi}(\xi) = \sum_\alpha \psi_\alpha(\xi) d_\alpha, \tag{6.97}$$

$\psi_\alpha(\xi)$ are the basis single-particle functions, we remind, that spin quantum numbers are included in α, and spin indices are included in $\xi \equiv \mathbf{r}, \sigma$ as variables.

The matrix elements are defined as

$$V_{\alpha\beta,\gamma\delta} = \int d\xi \int d\xi' \psi_\alpha^*(\xi) \psi_\beta^*(\xi') V(\xi, \xi') \psi_\gamma(\xi) \psi_\delta(\xi'). \tag{6.98}$$

For pair Coulomb interaction $V(|\mathbf{r}|)$ the matrix elements are

$$V_{\alpha\beta,\gamma\delta} = \sum_{\sigma\sigma'} \int d\mathbf{r} \int d\mathbf{r}' \psi_\alpha^*(\mathbf{r}, \sigma) \psi_\beta^*(\mathbf{r}', \sigma') V(|\mathbf{r} - \mathbf{r}'|) \psi_\gamma(\mathbf{r}, \sigma) \psi_\delta(\mathbf{r}', \sigma'). \tag{6.99}$$

Assume now, that the basis states $|\alpha\rangle$ are the states with definite spin quantum number σ_α. It means, that only one spin component of the wave function, namely $\psi_\alpha(\sigma_\alpha)$ is nonzero, and $\psi_\alpha(\bar{\sigma}_\alpha) = 0$. In this case the only nonzero matrix elements are those with $\sigma_\alpha = \sigma_\gamma$ and $\sigma_\beta = \sigma_\delta$, they are

$$V_{\alpha\beta,\gamma\delta} = \int d\mathbf{r} \int d\mathbf{r}' \psi^*_\alpha(\mathbf{r}) \psi^*_\beta(\mathbf{r}') V(|\mathbf{r} - \mathbf{r}'|) \psi_\gamma(\mathbf{r}) \psi_\delta(\mathbf{r}'). \tag{6.100}$$

In the case of delocalized basis states $\psi_\alpha(\mathbf{r})$, the main matrix elements are those with $\alpha = \gamma$ and $\beta = \delta$, because the wave functions of two different states with the same spin are orthogonal in real space and their contribution is small. It is also true for the systems with localized wave functions $\psi_\alpha(\mathbf{r})$, when the overlap between two different states is weak. In these cases it is enough to replace the interacting part by the Anderson-Hubbard Hamiltonian, describing only density-density interaction

$$\hat{H}_{AH} = \frac{1}{2} \sum_{\alpha \neq \beta} U_{\alpha\beta} \hat{n}_\alpha \hat{n}_\beta. \tag{6.101}$$

with the Hubbard interaction defined as

$$U_{\alpha\beta} = \int d\mathbf{r} \int d\mathbf{r}' |\psi_\alpha(\mathbf{r})|^2 |\psi_\beta(\mathbf{r}')|^2 V(|\mathbf{r} - \mathbf{r}'|). \tag{6.102}$$

In the limit of a single-level quantum dot (which is, however, a two-level system because of spin degeneration) we get the Anderson impurity model (AIM)

$$\hat{H}_{AIM} = \sum_{\sigma = \uparrow\downarrow} \epsilon_\sigma d^\dagger_\sigma d_\sigma + U \hat{n}_\uparrow \hat{n}_\downarrow. \tag{6.103}$$

The other important limit is the constant interaction model (CIM), which is valid when many levels interact with similar energies, so that approximately, assuming $U_{\alpha\beta} = U$ for any states α and β

$$\hat{H}_{AH} = \frac{1}{2} \sum_{\alpha \neq \beta} U_{\alpha\beta} \hat{n}_\alpha \hat{n}_\beta \approx \frac{U}{2} \left(\sum_\alpha \hat{n}_\alpha \right)^2 - \frac{U}{2} \left(\sum_\alpha \hat{n}^2_\alpha \right) = \frac{U\hat{N}(\hat{N} - 1)}{2}. \tag{6.104}$$

where we used $\hat{n}^2 = \hat{n}$.

Thus, the CIM reproduces the charging energy considered above, and the Hamiltonian of an isolated system is

$$\hat{H}_{CIM} = \sum_{\alpha\beta} \tilde{\epsilon}_{\alpha\beta} d^\dagger_\alpha d_\beta + E(N). \tag{6.105}$$

Note, that the equilibrium compensating charge density can be easily introduced into the AH Hamiltonian

$$\hat{H}_{AH} = \frac{1}{2} \sum_{\alpha \neq \beta} U_{\alpha\beta} \left(\hat{n}_\alpha - \bar{n}_\alpha \right) \left(\hat{n}_\beta - \bar{n}_\beta \right). \tag{6.106}$$

(ii) Coulomb blockade in quantum dots Here we want to consider the Coulomb blockade in intermediate-size quantum dots, where the typical energy level spacing $\Delta\epsilon$ is not too small to neglect it completely, but the number of levels is large enough, so that one can use the constant-interaction model (6.105), which we write in the eigenstate basis as

$$\hat{H}_{CIM} = \sum_{\alpha} \tilde{\epsilon}_{\alpha} d_{\alpha}^{\dagger} d_{\alpha} + E(n), \tag{6.107}$$

where the charging energy $E(n)$ is determined in the same way as previously, for example by the expression (6.104). Note, that for quantum dots the usage of classical capacitance is not well established, although for large quantum dots it is possible. Instead, we shift the energy levels in the dot $\tilde{\epsilon}_{\alpha} = \epsilon_{\alpha} + e\varphi_{\alpha}$ by the electrical potential

$$\varphi_{\alpha} = V_G + V_R + \eta_{\alpha}(V_L - V_R), \tag{6.108}$$

where η_{α} are some coefficients, dependent on geometry. This method can be easily extended to include any self-consistent effects on the mean-field level by the help of the Poisson equation (instead of classical capacitances). Besides, if all η_{α} are the same, our approach reproduce again the the classical expression

$$\hat{E}_{CIM} = \sum_{\alpha} \epsilon_{\alpha} n_{\alpha} + E(n) + en\varphi_{ext}. \tag{6.109}$$

The *addition energy* now depends not only on the charge of the molecule, but also on the state $|\alpha\rangle$, in which the electron is added

$$\Delta E_{n\alpha}^{+}(n, n_{\alpha} = 0 \to n+1, n_{\alpha} = 1) = E(n+1) - E(n) + \epsilon_{\alpha}, \tag{6.110}$$

we can assume in this case, that the single particle energies are additive to the charging energy, so that the full quantum eigenstate of the system is $|n, \hat{n}\rangle$, where the set $\hat{n} \equiv \{n_{\alpha}\}$ shows weather the particular single-particle state $|\alpha\rangle$ is empty or occupied. Some arbitrary state $\hat{n}$ looks like

$$\hat{n} \equiv \{n_{\alpha}\} \equiv \big(n_1, n_2, n_3, n_4, n_5, ...\big) = \big(1, 1, 0, 1, 0, ...\big). \tag{6.111}$$

Note, that the distribution $\hat{n}$ defines also $n = \sum_{\alpha} n_{\alpha}$. It is convenient, however, to keep notation n to remember about the charge state of a system, below we use both notations $|n, \hat{n}\rangle$ and short one $|\hat{n}\rangle$ as equivalent.

The other important point is that the distribution function $f_n(\alpha)$ in the charge state $|n\rangle$ is not assumed to be equilibrium, as previously (this condition is not specific to quantum dots with discrete energy levels, the distribution function in metallic islands can also be nonequilibrium. However, in the parameter range, typical for classical Coulomb blockade, the tunneling time is much smaller than the energy relaxation time, and quasiparticle nonequilibrium effects are usually neglected).

With these new assumptions, the theory of sequential tunneling is quite the same, as was considered

in the previous section. The master equation is [64, 65, 66, 67]

$$\frac{dp(n,\hat{n},t)}{dt} = \sum_{\hat{n}'} \left(\Gamma^{n\,n-1}_{\hat{n}\hat{n}'} p(n-1,\hat{n}',t) + \Gamma^{n\,n+1}_{\hat{n}\hat{n}'} p(n+1,\hat{n}',t)\right) - \sum_{\hat{n}'} \left(\Gamma^{n-1\,n}_{\hat{n}'\hat{n}} + \Gamma^{n+1\,n}_{\hat{n}'\hat{n}}\right) p(n,\hat{n},t) + I\left\{p(n,\hat{n},t)\right\}, \tag{6.112}$$

where $p(n,\hat{n},t)$ is now the probability to find the system in the state $|n,\hat{n}\rangle$, $\Gamma^{n\,n-1}_{\hat{n}\hat{n}'}$ is the transition rate from the state with $n-1$ electrons and single level occupation $\hat{n}'$ into the state with n electrons and single level occupation $\hat{n}$. The sum is over all states $\hat{n}'$, which are different by one electron from the state $\hat{n}$. The last term is included to describe possible inelastic processes inside the system and relaxation to the equilibrium function $p_{eq}(n,\hat{n})$. In principle, it is not necessary to introduce such type of dissipation in calculation, because the current is in any case finite. But the dissipation may be important in large systems and at finite temperatures. Besides, it is necessary to describe the limit of classical single-electron transport, where the distribution function of qausi-particles is assumed to be equilibrium. Below we shall not take into account this term, assuming that tunneling is more important.

While all considered processes are, in fact, single-particle tunneling processes, we arrive at

$$\frac{dp(\hat{n},t)}{dt} = \sum_{\beta} \left(\delta_{n_\beta 1}\Gamma^{n\,n-1}_{\beta} p(\hat{n},n_\beta=0,t) + \delta_{n_\beta 0}\Gamma^{n\,n+1}_{\beta} p(\hat{n},n_\beta=1,t)\right) - \sum_{\beta} \left(\delta_{n_\beta 1}\Gamma^{n-1\,n}_{\beta} + \delta_{n_\beta 0}\Gamma^{n+1\,n}_{\beta}\right) p(\hat{n},t), \tag{6.113}$$

where the sum is over single-particle states. The probability $p(\hat{n},n_\beta=0,t)$ is the probability of the state equivalent to $\hat{n}$, but without the electron in the state β. Consider, for example, the first term in the right part. Here the delta-function $\delta_{n_\beta 1}$ shows, that this term should be taken into account only if the single-particle state β in the many-particle state $\hat{n}$ is occupied, $\Gamma^{n\,n-1}_{\beta}$ is the probability of tunneling from the lead to this state, $p(\hat{n},n_\beta=0,t)$ is the probability of the state $\hat{n}'$, from which the system can come into the state $\hat{n}$.

The transitions rates are defined by the same golden rule expressions, as before, but with explicitly shown single-particle state α

$$\Gamma^{n+1\,n}_{L\alpha} = \frac{2\pi}{\hbar}\left|\left\langle n+1,n_\alpha=1|\hat{H}_{TL}|n,n_\alpha=0\right\rangle\right|^2 \delta(E_i-E_f) = \frac{2\pi}{\hbar}\sum_k |V_{k\alpha}|^2 f_k \delta(\Delta E^+_{n\alpha} - E_k), \tag{6.114}$$

$$\Gamma^{n-1\,n}_{L\alpha} = \frac{2\pi}{\hbar}\left|\left\langle n-1,n_\alpha=0|\hat{H}_{TL}|n,n_\alpha=1\right\rangle\right|^2 \delta(E_i-E_f) = \frac{2\pi}{\hbar}\sum_k |V_{k\alpha}|^2 (1-f_k)\, \delta(\Delta E^+_{n-1\,\alpha} - E_k), \tag{6.115}$$

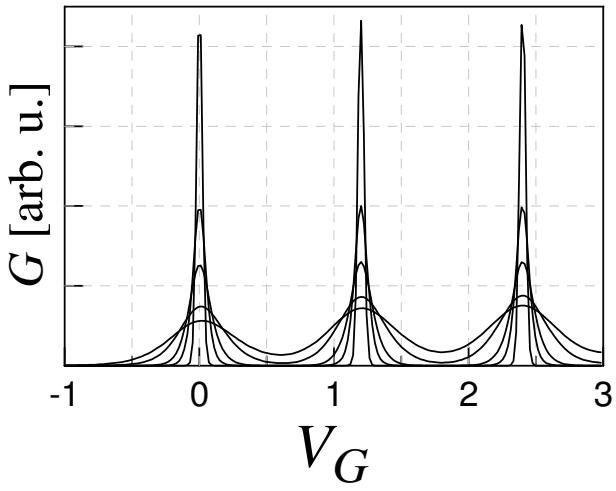


Figure 6.5: Linear conductance of a QD as a function of the gate voltage at different temperatures $T = 0.01E_C$, $T = 0.03E_C$, $T = 0.05E_C$, $T = 0.1E_C$, $T = 0.15E_C$ (lower curve).

there is no occupation factors $(1 - f_\alpha)$, f_α because this state is assumed to be empty in the sense of the master equation (6.113). The energy of the state is now included into the addition energy.

Using again the level-width function

$$\Gamma_{i=L,R\,\alpha}(E) = \frac{2\pi}{\hbar}\sum_k |V_{ik,\alpha}|^2 \delta(E - E_k). \tag{6.116}$$

we obtain

$$\Gamma_\alpha^{n+1\,n} = \Gamma_{L\alpha} f_L^0(\Delta E_{n\alpha}^+) + \Gamma_{R\alpha} f_R^0(\Delta E_{n\alpha}^+), \tag{6.117}$$

$$\Gamma_\alpha^{n-1\,n} = \Gamma_{L\alpha}\left(1 - f_L^0(\Delta E_{n-1\,\alpha}^+)\right) + \Gamma_{R\alpha}\left(1 - f_R^0(\Delta E_{n-1\,\alpha}^+)\right). \tag{6.118}$$

Finally, the current from the left or right contact to a system is

$$J_{i=L,R} = e\sum_\alpha \sum_{\hat{n}} p(\hat{n})\Gamma_{i\alpha}\left(\delta_{n_\alpha 0} f_i^0(\Delta E_{n\alpha}^+) - \delta_{n_\alpha 1}(1 - f_i^0(\Delta E_{n\alpha}^+))\right). \tag{6.119}$$

The sum over α takes into account all possible single particle tunneling events, the sum over states $\hat{n}$ summarize probabilities $p(\hat{n})$ of these states.

(iii) Linear conductance The linear conductance can be calculated analytically [64, 66]. Here we present the final result:

$$G = \frac{e^2}{T}\sum_\alpha \sum_{n=1}^{\infty} \frac{\Gamma_{L\alpha}\Gamma_{R\alpha}}{\Gamma_{L\alpha} + \Gamma_{R\alpha}} P_{eq}(n, n_\alpha = 1)\left[1 - f^0(\Delta E_{n-1\,\alpha}^+)\right], \tag{6.120}$$

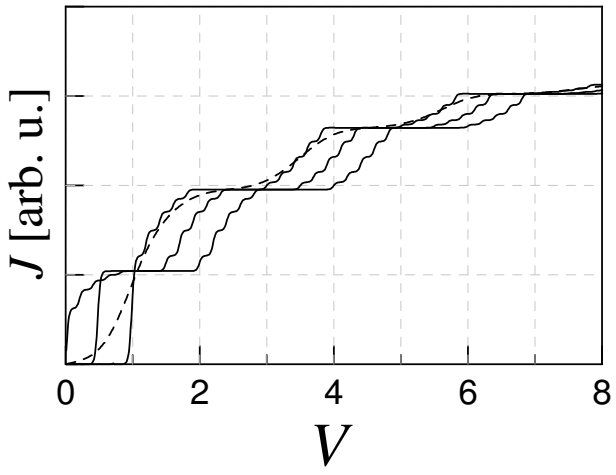


Figure 6.6: Coulomb staircase.

where $P_{eq}(n, n_\alpha = 1)$ is the joint probability that the quantum dot contains n electrons and the level α is occupied

$$P_{eq}(n, n_\alpha = 1) = \sum_{\hat{n}} p_{eq}(\hat{n}) \delta \left(n - \sum_{\beta} n_\beta \right) \delta_{n_\alpha 1}, \tag{6.121}$$

and the equilibrium probability (distribution function) is determined by the Gibbs distribution in the grand canonical ensemble:

$$p_{eq}(\hat{n}) = \frac{1}{Z} \exp \left[-\frac{1}{T} \left(\sum_{\alpha} \tilde{\epsilon}_\alpha + E(n) \right) \right]. \tag{6.122}$$

A typical behaviour of the conductance as a function of the gate voltage at different temperatures is shown in Fig. 6.5. In the resonant tunneling regime at low temperatures $T \ll \Delta\epsilon$ the peak height is strongly temperature-dependent. It is changed by classical temperature dependence (constant height) at $T \gg \Delta\epsilon$.

(iv)Transport at finite bias voltage At finite bias voltage we find new manifestations of the interplay between single-electron tunneling and resonant free-particle tunneling.

Now, let us consider the current-voltage curve of the differential conductance (Fig. 6.7). First of all, Coulomb staircase is reproduced, which is more pronounced, than for metallic islands, because the density of states is limited by the available single-particle states and the current is saturated. Besides, small additional steps due to discrete energy levels appear. This characteristic behaviour is possible for large enough dots with $\Delta\epsilon \ll E_C$. If the level spacing is of the oder of the charging energy $\Delta\epsilon \sim E_C$, the Coulomb blockade steps and discrete-level steps look the same, but their statistics (position and height distribution) is determined by the details of the single-particle spectrum and interactions [67].

Finally, let us consider the contour plot of the differential conductance (Fig. 6.7). Ii is essentially different from those for the metallic island. First, it is not symmetric in the gate voltage, because the energy

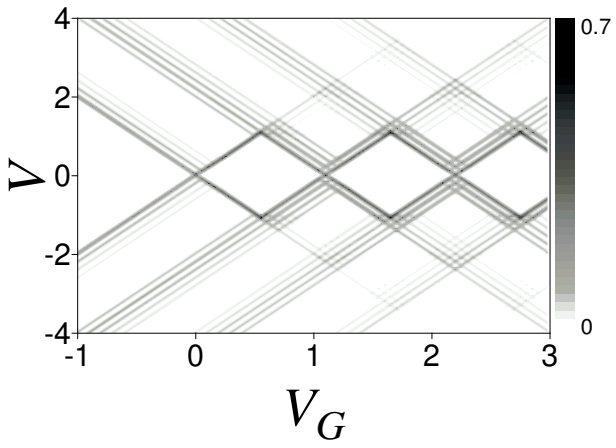


Figure 6.7: Contour plot of the differential conductance.

spectrum is restricted from the bottom, and at negative bias all the levels are above the Fermi-level (the electron charge is negative, and a negative potential means a positive energy shift). Nevertheless, existing stability patterns are of the same origin and form the same structure. The qualitatively new feature is additional lines correspondent to the additional discrete-level steps in the voltage-current curves.In general, the current and conductance of quantum dots demonstrate all typical features of discrete-level systems: current steps, conductance peaks. Without Coulomb interaction the usual picture of resonant tunneling is reproduced. In the limit of dense energy spectrum $\Delta\epsilon \to 0$ the sharp single-level steps are merged into the smooth Coulomb staircase.

6.2.2.2 Vibrons and Franck-Condon blockade

(i) Linear vibrons Vibrons are quantum local vibrations of nanosystems (Fig. 6.8), especially important in flexible molecules. In the linear regime the small displacements of the system can be expressed as linear combinations of the coordinates of the normal modes x_q, which are described by a set of independent

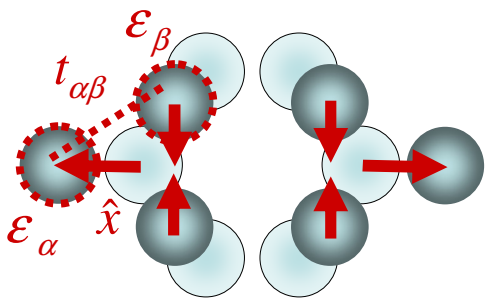


Figure 6.8: A local molecular vibration. The empty circles show the equilibrium positions of the atoms. The energies ϵ_α, ϵ_β and the overlap integral $t_{\alpha\beta}$ are perturbed.

linear oscillators with the Hamiltonian

$$\hat{H}_V^{(0)} = \sum_q \left(\frac{\hat{p}_q^2}{2m_q} + \frac{1}{2} m_q \omega_q^2 \hat{x}_q^2 \right). \tag{6.123}$$

The parameters m_q are determined by the microscopic theory, and $\hat{p}_q$ ($\hat{p}_q = -i\hbar \frac{\partial}{\partial x_q}$ in the x-representation) is the momentum conjugated to $\hat{x}_q$, $[\hat{x}_q, \hat{p}_q]_- = i\hbar$.

Let us outline briefly a possible way to calculate the normal modes of a molecule, and the relation between the positions of individual atoms and collective variables. We assume, that the atomic configuration of a system is determined mainly by the elastic forces, which are insensitive to the *transport* electrons. The dynamics of this system is determined by the *atomic* Hamiltonian

$$\hat{H}_{at} = \sum_n \frac{P_n^2}{2M_n} + W\left(\{\mathbf{R}_n\}\right), \tag{6.124}$$

where $W\left(\{\mathbf{R}_n\}\right)$ is the elastic energy, which includes also the static external forces and can be calculated by some *ab initio* method. Now define new generalized variables q_i with corresponding momentum p_i (as the generalized coordinates not only atomic positions, but also any other convenient degrees of freedom can be considered, for example, molecular rotations, center-of-mass motion, etc.)

$$\hat{H}_{at} = \sum_i \frac{p_i^2}{2m_i} + W\left(\{q_i\}\right), \tag{6.125}$$

"masses" m_i should be considered as some parameters. The equilibrium coordinates q_i^0 are defined from the energy minimum, the set of equations is

$$\frac{\partial W\left(\{q_i^0\}\right)}{\partial q_i} = 0. \tag{6.126}$$

The equations for linear oscillations are obtained from the next order expansion in the deviations $\Delta q_i = q_i - q_i^0$

$$\hat{H}_{at} = \sum_i \frac{p_i^2}{2m_i} + \sum_{ij} \frac{\partial^2 W\left(\{q_j^0\}\right)}{\partial q_i \partial q_j} \Delta q_i \Delta q_j. \tag{6.127}$$

This Hamiltonian describes a set of coupled oscillators. Finally, applying the canonical transformation from Δq_i to new variables x_q (q is now the index of independent modes)

$$x_q = \sum_i C_{qi} q_i \tag{6.128}$$

we derive the Hamiltonian (6.123) together with the frequencies ω_q of vibrational modes.

It is useful to introduce the creation and annihilation operators

$$a_q^\dagger = \frac{1}{\sqrt{2}}\left(\sqrt{\frac{m_q\omega_q}{\hbar}}\hat{x}_q + \frac{i}{\sqrt{m_q\omega_q\hbar}}\hat{p}_q\right), \tag{6.129}$$

$$a_q = \frac{1}{\sqrt{2}}\left(\sqrt{\frac{m_q\omega_q}{\hbar}}\hat{x}_q - \frac{i}{\sqrt{m_q\omega_q\hbar}}\hat{p}_q\right), \tag{6.130}$$

in this representation the Hamiltonian of free vibrons is ($\hbar = 1$)

$$\hat{H}_V^{(0)} = \sum_q \omega_q a_q^\dagger a_q. \tag{6.131}$$

(ii) Electron-vibron Hamiltonian A system without vibrons is described as before by a basis set of states $|\alpha\rangle$ with energies ϵ_α and inter-state overlap integrals $t_{\alpha\beta}$, the model Hamiltonian of a noninteracting system is

$$\hat{H}_S^{(0)} = \sum_\alpha \left(\epsilon_\alpha + e\varphi_\alpha(t)\right) d_\alpha^\dagger d_\alpha + \sum_{\alpha\neq\beta} t_{\alpha\beta} d_\alpha^\dagger d_\beta, \tag{6.132}$$

where $d_\alpha^\dagger$,d_α are creation and annihilation operators in the states $|\alpha\rangle$, and $\varphi_\alpha(t)$ is the (self-consistent) electrical potential (6.108). The index α is used to mark single-electron states (atomic orbitals) including the spin degree of freedom.

To establish the Hamiltonian describing the interaction of electrons with vibrons in nanosystems, we can start from the generalized Hamiltonian

$$\hat{H}_S = \sum_\alpha \tilde{\epsilon}_\alpha\left(\{x_q\}\right) d_\alpha^\dagger d_\alpha + \sum_{\alpha\neq\beta} t_{\alpha\beta}\left(\{x_q\}\right) d_\alpha^\dagger d_\beta, \tag{6.133}$$

where the parameters are some functions of the vibronic normal coordinates x_q. Note that we consider now only the electronic states, which were excluded previously from the Hamiltonian (6.124), it is important to prevent double counting.

Expanding to the first order near the equilibrium state we obtain

$$\hat{H}_{ev} = \sum_\alpha \sum_q \frac{\partial\tilde{\epsilon}_\alpha(0)}{\partial x_q} x_q d_\alpha^\dagger d_\alpha + \sum_{\alpha\neq\beta}\sum_q \frac{\partial t_{\alpha\beta}(0)}{\partial x_q} x_q d_\alpha^\dagger d_\beta, \tag{6.134}$$

where $\tilde{\epsilon}_\alpha(0)$ and $t_{\alpha\beta}(0)$ are unperturbed values of the energy and the overlap integral. In the quantum limit the normal coordinates should be treated as operators, and in the second-quantized representation the interaction Hamiltonian is

$$\hat{H}_{ev} = \sum_{\alpha\beta}\sum_q \lambda_{\alpha\beta}^q (a_q + a_q^\dagger) d_\alpha^\dagger d_\beta. \tag{6.135}$$

This Hamiltonian is similar to the usual electron-phonon Hamiltonian, but the vibrations are like localized phonons and q is an index labeling them, not the wave-vector. We include both diagonal coupling, which describes a change of the electrostatic energy with the distance between atoms, and the off-diagonal

coupling, which describes the dependence of the matrix elements $t_{\alpha\beta}$ over the distance between atoms.

The full Hamiltonian

$$\hat{H} = \hat{H}_S^0 + \hat{H}_V + \hat{H}_L + \hat{H}_R + \hat{H}_T \tag{6.136}$$

is the sum of the noninteracting Hamiltonian $\hat{H}_S^0$, the Hamiltonians of the leads $\hat{H}_{R(L)}$, the tunneling Hamiltonian $\hat{H}_T$ describing the system-to-lead coupling, the vibron Hamiltonian $\hat{H}_V$ including electron-vibron interaction and coupling of vibrations to the environment (describing dissipation of vibrons).

Vibrons and the electron-vibron coupling are described by the Hamiltonian ($\hbar = 1$)

$$\hat{H}_V = \sum_q \omega_q a_q^\dagger a_q + \sum_{\alpha\beta}\sum_q \lambda_{\alpha\beta}^q (a_q + a_q^\dagger) d_\alpha^\dagger d_\beta + \hat{H}_{env}. \tag{6.137}$$

The first term represents free vibrons with the energy $\hbar\omega_q$. The second term is the electron-vibron interaction. The rest part $\hat{H}_{env}$ describes dissipation of vibrons due to interaction with other degrees of freedom, we do not consider the details in this chapter.

The Hamiltonians of the right (R) and left (L) leads read as usual

$$\hat{H}_{i=L(R)} = \sum_{k\sigma} (\epsilon_{ik\sigma} + e\varphi_i) c_{ik\sigma}^\dagger c_{ik\sigma}, \tag{6.138}$$

φ_i are the electrical potentials of the leads. Finally, the tunneling Hamiltonian

$$\hat{H}_T = \sum_{i=L,R}\sum_{k\sigma,\alpha} \left(V_{ik\sigma,\alpha} c_{ik\sigma}^\dagger d_\alpha + V_{ik\sigma,\alpha}^* d_\alpha^\dagger c_{ik\sigma} \right) \tag{6.139}$$

describes the hopping between the leads and the molecule. A direct hopping between two leads is neglected.

The simplest example of the considered model is a single-level model (Fig. 6.9) with the Hamiltonian

$$\hat{H} = \tilde{\epsilon}_0 d^\dagger d + \omega_0 a^\dagger a + \lambda \left(a^\dagger + a\right) d^\dagger d + \sum_{ik} \left[\tilde{\epsilon}_{ik} c_{ik}^\dagger c_{ik} + V_{ik} c_{ik}^\dagger d + h.c. \right], \tag{6.140}$$

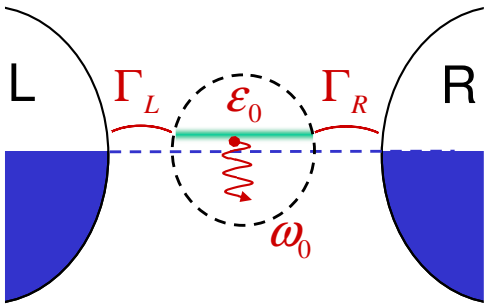


Figure 6.9: Single-level electron-vibron model.

where the first and the second terms describe free electron state and free vibron, the third term is electron-vibron interaction, and the rest is the Hamiltonian of the leads and tunneling coupling ($i = L, R$ is the lead index).

The other important case is a center-of-mass motion of molecules between the leads (Fig. 6.10). Here not the internal overlap integrals, but the coupling to the leads $V_{ik\sigma,\alpha}(x)$ is fluctuating. This model is easily reduced to the general model (6.137), if we consider additionaly two not flexible states in the left and right leads (two atoms most close to a system), to which the central system is coupled (shown by the dotted circles).

Tunneling Hamiltonian includes x-dependent matrix elements, considered in linear approximation

$$H_T = \sum_{i=L,R} \sum_{k\sigma,\alpha} \left(V_{ik\sigma,\alpha}(\hat{x}) c^\dagger_{ik\sigma} d_\alpha + h.c. \right), \tag{6.141}$$

$$V_{L,R}(x) = V_0 e^{\mp \hat{x}/L} \approx V_0 \left(1 \mp \frac{\hat{x}}{L} \right). \tag{6.142}$$

Consider now a single-level molecule ($\alpha \equiv 0$) and extend our system, including two additional states from the left ($\alpha \equiv l$) and right ($\alpha \equiv r$) sides of a molecule, which are coupled to the central state through x-dependent matrix elements, and to the leads in a usual way through $\Gamma_{L(R)}$. Then the Hamiltonian is of linear electron-vibron type

$$\begin{aligned} \hat{H}_{M+V} = & \sum_{\alpha=l,0,r} (\epsilon_\alpha + e\varphi_\alpha)\, d^\dagger_\alpha d_\alpha + t_l (d^\dagger_l d_0 + h.c.) + t_r (d^\dagger_r d_0 + h.c.) + \\ & + \omega_0 a^\dagger a + (a + a^\dagger) \left(\lambda_0 d^\dagger_0 d_0 - \lambda_l (d^\dagger_l d_0 + h.c.) + \lambda_r (d^\dagger_r d_0 + h.c.) \right). \end{aligned} \tag{6.143}$$

(iii) Local polaron and canonical transformation Now let us start to consider the situation, when the electron-vibron interaction is strong. For an isolated system with the Hamiltonian, including only

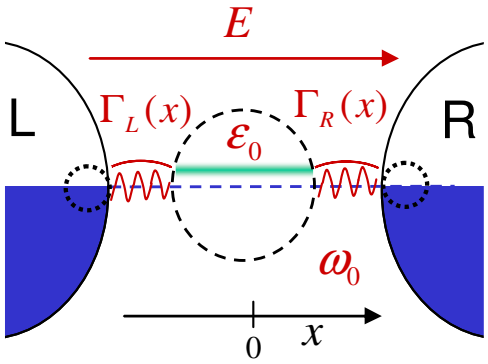


Figure 6.10: A center-of-mass vibration.

diagonal terms,

$$\hat{H}_{S+V} = \sum_{\alpha} \tilde{\epsilon}_{\alpha} d_{\alpha}^{\dagger} d_{\alpha} + \sum_{q} \omega_q a_q^{\dagger} a_q + \sum_{\alpha} \sum_{q} \lambda_{\alpha}^{q} (a_q + a_q^{\dagger}) d_{\alpha}^{\dagger} d_{\alpha}, \tag{6.144}$$

the problem can be solved exactly. This solution, as well as the method of the solution (canonical transformation), plays an important role in the theory of electron-vibron systems, and we consider it in detail.

Let's start from the simplest case. The single-level electron-vibron model is described by the Hamiltonian

$$\hat{H}_{S+V} = \tilde{\epsilon}_0 d^{\dagger} d + \omega_0 a^{\dagger} a + \lambda \left(a^{\dagger} + a\right) d^{\dagger} d, \tag{6.145}$$

where the first and the second terms describe free electron state and free vibron, and the third term is the electron-vibron interaction.

This Hamiltonian is diagonalized by the canonical transformation (called "Lang-Firsov" or "polaron") [69, 70, 71]

$$\bar{H} = \hat{S}^{-1} \hat{H} \hat{S}, \tag{6.146}$$

with

$$\hat{S} = \exp\left[-\frac{\lambda}{\omega_0} \left(a^{\dagger} - a\right) d^{\dagger} d\right], \tag{6.147}$$

the Hamiltonian (6.145) is transformed as

$$\bar{H}_{S+V} = \hat{S}^{-1} \hat{H}_{S+V} \hat{S} = \tilde{\epsilon}_0 \bar{d}^{\dagger} \bar{d} + \omega_0 \bar{a}^{\dagger} \bar{a} + \lambda \left(\bar{a}^{\dagger} + \bar{a}\right) \bar{d}^{\dagger} \bar{d}, \tag{6.148}$$

it has the same form as (6.145) with new operators, it is a trivial consequence of the general property

$$\hat{S}^{-1} \left(\hat{f}_1 \hat{f}_2 \hat{f}_3 ...\right) \hat{S} = (\hat{S}^{-1} \hat{f}_1 \hat{S})(\hat{S}^{-1} \hat{f}_2 \hat{S})(\hat{S}^{-1} \hat{f}_3 \hat{S}) ... = \bar{f}_1 \bar{f}_2 \bar{f}_3 ... \tag{6.149}$$

and new single-particle operators are

$$\bar{a} = \hat{S}^{-1} a \hat{S} = a - \tfrac{\lambda}{\omega_0} d^{\dagger} d, \tag{6.150}$$

$$\bar{a}^{\dagger} = \hat{S}^{-1} a^{\dagger} \hat{S} = a^{\dagger} - \tfrac{\lambda}{\omega_0} d^{\dagger} d, \tag{6.151}$$

$$\bar{d} = \hat{S}^{-1} d \hat{S} = \exp\left[-\tfrac{\lambda}{\omega_0} \left(a^{\dagger} - a\right)\right] d, \tag{6.152}$$

$$\bar{d}^{\dagger} = \hat{S}^{-1} d^{\dagger} \hat{S} = \exp\left[\tfrac{\lambda}{\omega_0} \left(a^{\dagger} - a\right)\right] d^{\dagger}. \tag{6.153}$$

Substituting these expressions into (6.148) we get finally

$$\bar{H}_{S+V} = \left(\tilde{\epsilon}_0 - \frac{\lambda^2}{\omega_0}\right) d^{\dagger} d + \omega_0 a^{\dagger} a. \tag{6.154}$$

We see that the electron-vibron Hamiltonian (6.145) is equivalent to the free-particle Hamiltonian (6.154). This equivalence means that any quantum state $|\bar{\psi}_{\lambda}\rangle$, obtained as a solution of the Hamiltonian

(6.154) is one-to-one equivalent to the state $|\psi_\lambda\rangle$ as a solution of the initial Hamiltonian (6.145), with the same matrix elements for any operator

$$\langle\bar{\psi}_\lambda|\bar{f}|\bar{\psi}_\lambda\rangle = \langle\psi_\lambda|\hat{f}|\psi_\lambda\rangle, \tag{6.155}$$

$$\bar{f} = \hat{S}^{-1}\hat{f}\hat{S}, \tag{6.156}$$

$$|\bar{\psi}_\lambda\rangle = \hat{S}^{-1}|\psi_\lambda\rangle. \tag{6.157}$$

It follows immediately that the eigenstates of the free-particle Hamiltonian are

$$|\bar{\psi}_{nm}\rangle = |n = 0, 1; m = 0, 1, 2, ...\rangle = (d^\dagger)^n \frac{(a^\dagger)^m}{\sqrt{m!}}|0\rangle, \tag{6.158}$$

and the eigen-energies are

$$E(n, m) = \left(\tilde{\epsilon}_0 - \frac{\lambda^2}{\omega_0}\right) n + \omega_0 m. \tag{6.159}$$

The eigenstates of the *initial* Hamiltonian (6.145) are

$$|\psi_{nm}\rangle = \hat{S}|\bar{\psi}_{nm}\rangle = e^{-\frac{\lambda}{\omega_0}(a^\dagger - a)d^\dagger d}(d^\dagger)^n \frac{(a^\dagger)^m}{\sqrt{m!}}|0\rangle, \tag{6.160}$$

with the same quantum numbers (n, m) and the same energies (6.159). This representation of the eigenstates demonstrates clearly the collective nature of the excitations, but it is inconvenient for practical calculations.

Now let us consider the polaron transformation (6.146)-(6.147) applied to the tunneling Hamiltonian

$$\hat{H}_T = \sum_{i=L,R}\sum_{k\sigma}\left(V_{ik\sigma}c^\dagger_{ik\sigma}d + V^*_{ik\sigma}d^\dagger c_{ik\sigma}\right) \tag{6.161}$$

The electron operators in the left and right leads $c_{ik\sigma}$ are not changed by this operation, but the dot operators d_α, $d^\dagger_\alpha$ are changed in accordance with (6.152) and (6.153). So that transformed Hamiltonian is

$$\bar{H}_T = \sum_{i=L,R}\sum_{k\sigma}\left(V_{ik\sigma}e^{-\frac{\lambda}{\omega_0}(a^\dagger - a)}c^\dagger_{ik\sigma}d + V^*_{ik\sigma}e^{\frac{\lambda}{\omega_0}(a^\dagger - a)}d^\dagger c_{ik\sigma}\right). \tag{6.162}$$

Now we see clear the problem: while the new dot Hamiltonian (6.154) is very simple and exactly solvable, the new tunneling Hamiltonian (6.162) is complicated. Moreover, instead of one linear electron-vibron interaction term, the exponent in (6.162) produces all powers of vibronic operators. Actually, we simply remove the complexity from one place to the other. This approach works well, if the tunneling can be considered as a perturbation, we consider it in the next section. In the general case the problem is quite difficult, but in the single-particle approximation it can be solved exactly [72, 73, 74, 75].

To conclude, after the canonical transformation we have two equivalent models: (1) the initial model (6.145) with the eigenstates (6.160); and (2) the *fictional* free-particle model (6.154) with the eigenstates

(6.158). We shall call this second model *polaron representation*. The relation between the models is established by (6.155)-(6.157). It is also clear from the Hamiltonian (6.148), that the operators $\bar{d}^\dagger$, $\bar{d}$, $\bar{a}^\dagger$, and $\bar{a}$ describe the initial electrons and vibrons in the fictional model.

(iv) Inelastic tunneling in the single-particle approximation In this section we consider a special case of a *single particle transmission* through an electron-vibron system. It means that we consider a system coupled to the leads, but without electrons in the leads. This can be considered equivalently as the limit of large electron level energy ϵ_0 (far from the Fermi surface in the leads).

The inelastic *transmission matrix* $T(\epsilon', \epsilon)$ describes the probability that an electron with energy ϵ, incident from one lead, is transmitted with the energy ϵ' into a second lead. The transmission function can be defined as the total transmission probability

$$T(\epsilon) = \int T(\epsilon', \epsilon) d\epsilon'. \tag{6.163}$$

For a noninteracting single-level system the transmission matrix is

$$T^0(\epsilon', \epsilon) = \frac{\Gamma_R(\epsilon)\Gamma_L(\epsilon)\delta(\epsilon - \epsilon')}{(\epsilon - \epsilon_0 - \Lambda(\epsilon))^2 + (\Gamma(\epsilon)/2)^2}, \tag{6.164}$$

where $\Gamma(\epsilon) = \Gamma_L(\epsilon) + \Gamma_R(\epsilon)$ is the level-width function, and $\Lambda(\epsilon)$ is the real part of the self-energy.

We can do some general conclusions, based on the form of the tunneling Hamiltonian (6.162). Expanding the exponent in the same way as before, we get

$$\bar{H}_T = \sum_{i=L,R} \sum_{k\sigma} \left(V_{ik\sigma} c^\dagger_{ik\sigma} d \left[\alpha_0 + \sum_{m=1}^{\infty} \alpha_m \left((a^\dagger)^m + a^m \right) \right] + h.c. \right), \tag{6.165}$$

with the coefficients

$$\alpha_m = \left(-\frac{\lambda}{\omega_0} \right)^m \frac{e^{-(\lambda/\omega_0)^2/2}}{m!}. \tag{6.166}$$

This complex Hamiltonian has very clear interpretation, the tunneling of one electron from the right to the left lead is accompanied by the excitation of vibrons. The energy conservation implies that

$$\epsilon - \epsilon' = \pm m\omega_0, \tag{6.167}$$

so that the inelastic tunneling with emission or absorption of vibrons is possible.

The exact solution is possible in the wide-band limit. [72, 73, 74, 75]

It is convenient to introduce the dimensionless electron-vibron coupling constant

$$g = \left(\frac{\lambda}{\omega_0} \right)^2. \tag{6.168}$$

At zero temperature the solution is

$$T(\epsilon',\epsilon) = \Gamma_L\Gamma_R e^{-2g}\sum_{m=0}^{\infty}\frac{g^m}{m!}\delta(\epsilon-\epsilon'-m\omega_0) \times\left|\sum_{j=0}^{m}(-1)^j\frac{m!}{j!(m-j)!}\sum_{l=0}^{\infty}\frac{g^l}{l!}\frac{1}{\epsilon-\epsilon_0+g\omega_0-(j+l)\omega_0+i\Gamma/2}\right|^2, \tag{6.169}$$

the total transmission function $T(\epsilon)$ is trivially obtain by integration over ϵ'. The representative results are presented in Figs. 6.11 and 6.12.

At finite temperature the general expression is too cumbersome, and we present here only the expression for the total transmission function

$$T(\epsilon) = \frac{\Gamma_L\Gamma_R}{\Gamma}e^{-g(1+2n_\omega)}\int_{-\infty}^{\infty}dt \times\exp\left(-\frac{\Gamma}{2}|t|+i(\epsilon-\epsilon_0+g\omega_0)t-g\left[(1+n_\omega)e^{-i\omega_0 t}+n_\omega e^{i\omega_0 t}\right]\right), \tag{6.170}$$

where n_ω is the equilibrium number of vibrons.

(v) Master equation When the system is weakly coupled to the leads, the polaron representation (6.154), (6.162) is a convenient starting point. Here we consider how the sequential tunneling is modified by vibrons.

The master equation for the probability $p(n,m,t)$ to find the system in one of the polaron eigenstates (6.158) can be written as

$$\frac{dp(n,m)}{dt} = \sum_{n'm'}\Gamma^{nn'}_{mm'}p(n',m') - \sum_{n'm'}\Gamma^{n'n}_{m'm}p(n,m) + I^V[p], \tag{6.171}$$

where the first term describes tunneling transition *into the state* $|n,m\rangle$, and the second term – tunneling transition *out of the state* $|n,m\rangle$, $I^V[p]$ is the vibron scattering integral describing the relaxation to equilibrium. The transition rates $\Gamma^{nn'}_{mm'}$ should be found from the Hamiltonian (6.162).

Taking into account all possible single-electron tunneling processes, we obtain the incoming tunneling rate

$$\begin{aligned}\Gamma^{10}_{mm'} &= \frac{2\pi}{\hbar}\sum_{ik\sigma}f_i^0(E_{ik\sigma})\left|\left\langle i\bar{k},1,m\left|\bar{H}_T\right|ik,0,m'\right\rangle\right|^2\delta(E_{0m'}+E_{ik\sigma}-E_{1m})\\ &= \frac{2\pi}{\hbar}\sum_{ik\sigma}f_i^0(E_{ik\sigma})\left|V_{ik\sigma}\right|^2\left|\left\langle m\left|e^{\frac{\lambda}{\omega_0}(a^\dagger-a)}\right|m'\right\rangle\right|^2\delta(E_{0m'}+E_{ik\sigma}-E_{1m})\\ &= \sum_{i=L,R}\Gamma_i(E_{1m}-E_{0m'})\left|M_{mm'}\right|^2 f_i^0(E_{1m}-E_{0m'}),\end{aligned} \tag{6.172}$$

where

$$M_{mm'} = \left\langle m\left|e^{\frac{\lambda}{\omega_0}(a^\dagger-a)}\right|m'\right\rangle \tag{6.173}$$

is the Franck-Condon matrix element. We use usual short-hand notations: $|ik, n, m\rangle$ is the state with occupied k-state in the $i-$th lead, n electrons, and m vibrons, while $|i\bar{k}, n, m\rangle$ is the state with unoccupied k-state in the $i-$th lead, E_{nm} is the polaron energy (6.159).

Similarly, the outgoing rate is

$$\Gamma^{01}_{mm'} = \sum_{i=L,R} \Gamma_i(E_{1m'} - E_{0m}) \left|M_{mm'}\right|^2 \left(1 - f^0_i(E_{1m'} - E_{0m})\right). \tag{6.174}$$

The current (from the left or right lead to the system) is

$$J_{i=L,R}(t) = e \sum_{mm'} \left(\Gamma^{10}_{imm'} p(0, m') - \Gamma^{01}_{imm'} p(1, m')\right). \tag{6.175}$$

The system of equations (6.171)-(6.175) solves the transport problem in the sequential tunneling regime.

(v) Franck-Condon blockade Now let us consider some details of the tunneling at small and large values of the electro-vibron coupling parameter $g = \left(\frac{\lambda}{\omega_0}\right)^2$.

The matrix element (6.173) can be calculated analytically, it is symmetric in $m - m'$ and for $m < m'$ is

$$M_{m<m'} = \sum_{l=0}^{m} \frac{(-g)^l \sqrt{m!m'!} e^{-g/2} g^{(m'-m)/2}}{l!(m-l)!(l+m'-m)!}. \tag{6.176}$$

The lowest order elements are

$$M_{0m} = e^{-g/2} \frac{g^{m/2}}{\sqrt{m!}}, \tag{6.177}$$

$$M_{11} = (1-g)e^{-g/2}, \tag{6.178}$$

$$M_{12} = \sqrt{2g}\left(1 - \frac{g}{2}\right) e^{-g/2} ... \tag{6.179}$$

The characteristic feature of these matrix elements is so-called Franck-Condon blockade [87, 88], illustrated in Fig. 6.13 for the matrix element M_{0m}. From the picture, as well as from the analytical formulas, it is clear, that in the case of strong electron-vibron interaction the tunneling with small change of the vibron quantum number is suppressed exponentially, and only the tunneling through high-energy states is possible, which is also suppressed at low bias voltage and low temperature. Thus, the electron transport through a system (linear conductance) is very small.

There are several interesting manifestations of the Franck-Condon blockade.

The *life-time* of the state $|n, m\rangle$ is determined by the sum of the rates of all possible processes which change this state in the assumption that all other states are empty

$$\tau^{-1}_{nm} = \sum_{n'm'} \Gamma^{n'n}_{m'm}. \tag{6.180}$$

As an example, let us calculate the life-time of the neutral state $|0, 0\rangle$, which has the energy higher

than the charged ground state $|1,0\rangle$.

$$\tau_{00}^{-1} = \sum_{n'm'} \Gamma_{m'0}^{n'0} = \sum_{m} \sum_{i=L,R} \Gamma_i(E_{1m} - E_{00}) \left|M_{m0}\right|^2 f_i^0(E_{1m} - E_{00}). \tag{6.181}$$

In the wide-band limit we obtain the simple analytical expression

$$\tau_{00}^{-1} = \Gamma \sum_{m} e^{-g} \frac{g^m}{m!} f^0 \left(\tilde{\epsilon}_0 - \frac{\lambda^2}{\omega_0} + \omega_0 m \right). \tag{6.182}$$

The corresponding expression for the life-time of the charged state (which can be excited by thermal fluctuations) is

$$\tau_{10}^{-1} = \Gamma \sum_{m} e^{-g} \frac{g^m}{m!} f^0 \left(-\tilde{\epsilon}_0 + \frac{\lambda^2}{\omega_0} + \omega_0 m \right). \tag{6.183}$$

The result of the calculation is shown in Fig. 6.14, it is clear seen that the tunneling from the state $|0,0\rangle$ to the charged state and from the state $|1,0\rangle$ to the neutral state is exponentially suppressed in comparison with the bare tunneling rate Γ at large values of the electron-vibron interaction constant λ. This *polaron memory effect* can be used to create nano-memory and nano-switches. At finite voltage the switching between two states is easy accessible through the excited vibron states. It can be used to switch between memory states.

The other direct manifestation of the Franck-Condon blockade, – suppression of the linear conductance, was considered in Refs. [87, 88].

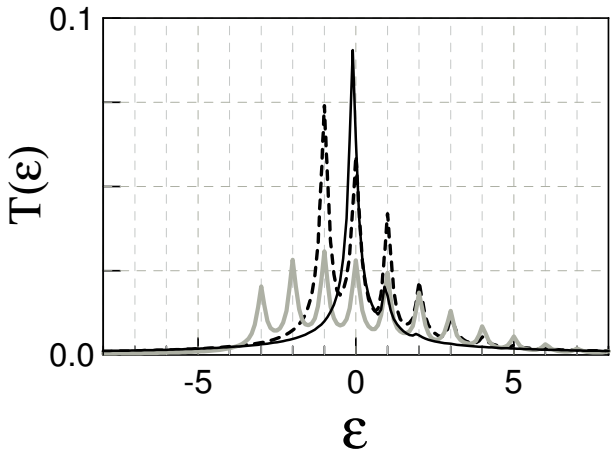


Figure 6.11: Transmission function as a function of energy at different electron-vibron coupling: $g = 0.1$ (thin solid line), $g = 1$ (dashed line), and $g = 3$ (thick solid line), at $\Gamma = 0.1$.

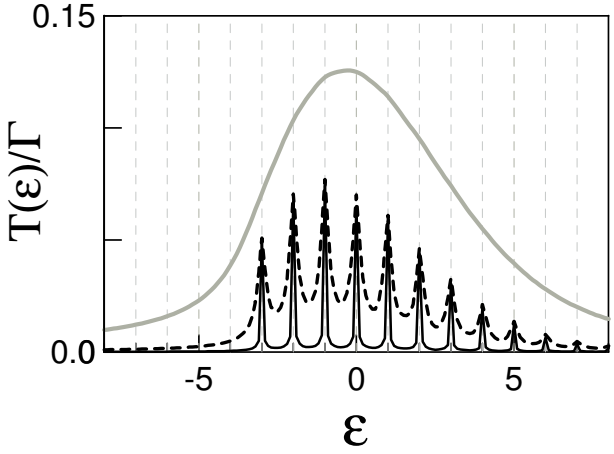


Figure 6.12: Transmission function as a function of energy at different coupling to the leads: $\Gamma = 0.01$ (thin solid line), $\Gamma = 0.1$ (dashed line), and $\Gamma = 1$ (thick solid line), at $g = 3$.

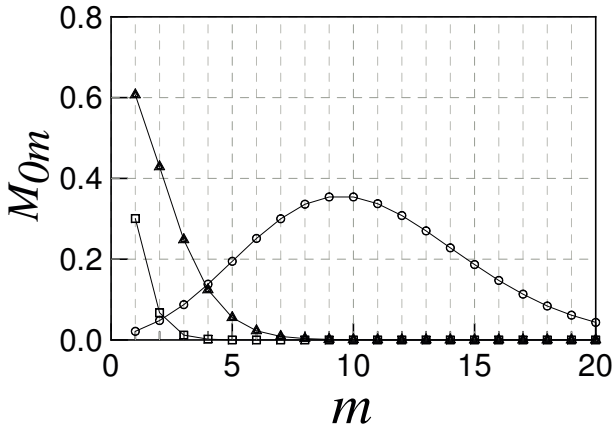


Figure 6.13: Franck-Condon matrix elements M_{0m} for weak ($g = 0.1$, squares), intermediate ($g = 1$, triangles), and strong ($g = 10$, circles) electron-vibron interaction. Lines are the guides for eyes.

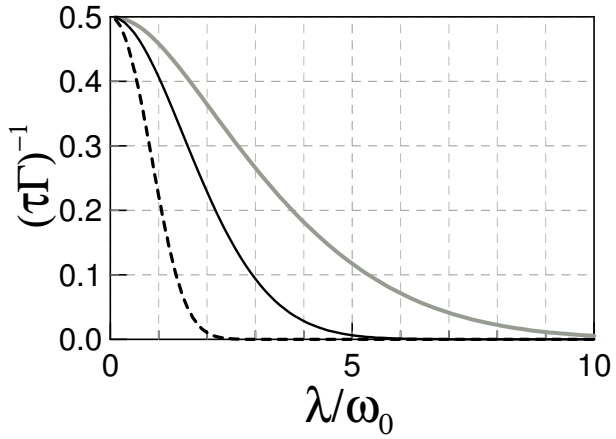


Figure 6.14: The inverse life-time $(\tau\Gamma)^{-1}$ as a function of λ/ω_0 at optimal electron level position $\epsilon_0 = \lambda^2/2\omega_0$ for neutral state (thin solid line), and for the charged state (dashed line), and for the neutral state at other level position $\epsilon_0 = \lambda^2/4\omega_0$ (thick solid line).

6.3 Nonequilibrium Green function theory of transport

6.3.1 Standard transport model: a nanosystem between ideal leads

First of all, we formulate a standard discrete-level model to describe nanoscale interacting quantum systems (quantum dot, system of quantum dots, molecule, below "nanosystem", "central system", or simply "system") coupled to free conduction electrons in the leads. We include the Coulomb interaction with the help of the Anderson-Hubbard Hamiltonan to be able to describe correlation effects, such as Coulomb blockade and Kondo effect, which could dominate at low temperatures. At high temperatures or weak interaction the self-consistent mean-field effects are well reproduced by the same model. Furthermore, electrons are coupled to vibrational modes, below we use the electron-vibron model introduced previously.

(i) The model Hamiltonian The full Hamiltonian is the sum of the free system Hamiltonian $\hat{H}_S^{(0)}$, the inter-system electron-electron interaction Hamiltonian $\hat{H}_C$, the vibron Hamiltonian $\hat{H}_V$ including the electron-vibron interaction and coupling of vibrations to the environment (dissipation of vibrons), the Hamiltonians of the leads $\hat{H}_{R(L)}$, and the tunneling Hamiltonian $\hat{H}_T$ describing the system-to-lead coupling

$$\hat{H} = \hat{H}_S + \hat{H}_C + \hat{H}_V + \hat{H}_L + \hat{H}_R + \hat{H}_T. \tag{6.184}$$

An isolated noninteracting nanosystem is described as a set of discrete states $|\alpha\rangle$ with energies ϵ_α and inter-orbital overlap integrals $t_{\alpha\beta}$ by the following model Hamiltonian:

$$\hat{H}_S^{(0)} = \sum_\alpha \left(\epsilon_\alpha + e\varphi_\alpha(t)\right) d_\alpha^\dagger d_\alpha + \sum_{\alpha\neq\beta} t_{\alpha\beta} d_\alpha^\dagger d_\beta, \tag{6.185}$$

where $d_\alpha^\dagger$,d_α are creation and annihilation operators in the states $|\alpha\rangle$, and $\varphi_\alpha(t)$ is the effective (self-consistent) electrical potential. The index α is used to mark single-electron states (e.g. atomic orbitals) including the spin degree of freedom. In the eigenstate (molecular orbital) representation the second term is absent and the Hamiltonian is diagonal.

For molecular transport the parameters of a model are to be determined by *ab initio* methods or considered as semi-empirical. This is a compromise, which allows us to consider complex molecules with a relatively simple model.

The Hamiltonians of the right (R) and left (L) leads are

$$\hat{H}_{i=L(R)} = \sum_{k\sigma} (\epsilon_{ik\sigma} + e\varphi_i(t)) c_{ik\sigma}^\dagger c_{ik\sigma}, \tag{6.186}$$

$\varphi_i(t)$ are the electrical potentials of the leads, the index k is the wave vector, but can be considered as representing an other conserved quantum number, σ is the spin index, but can be considered as a generalized channel number, describing e.g. different bands or subbands in semiconductors. Alternatively, the tight-binding model can be used also for the leads, then (6.186) should be considered as a result of the Fourier transformation. The leads are assumed to be noninteracting and equilibrium.

The tunneling Hamiltonian

$$\hat{H}_T = \sum_{i=L,R} \sum_{k\sigma,\alpha} \left(V_{ik\sigma,\alpha} c^\dagger_{ik\sigma} d_\alpha + V^*_{ik\sigma,\alpha} d^\dagger_\alpha c_{ik\sigma} \right) \tag{6.187}$$

describes the hopping between the leads and the system. The direct hopping between two leads is neglected (relatively weak molecule-to-lead coupling case). Note, that the direct hoping between equilibrium leads can be easy taken into account as an additional independent current channel.

The Coulomb interaction inside a system is described by the Anderson-Hubbard Hamiltonian

$$\hat{H}_C = \frac{1}{2} \sum_{\alpha \neq \beta} U_{\alpha\beta} \hat{n}_\alpha \hat{n}_\beta. \tag{6.188}$$

This Hamiltonian is used usually only for the short-range part of Coulomb interaction. The long-range interactions can be better introduced through the self-consistent electrical potential φ_α, which is determined by the Poison equation with the average electron density.

Vibrations and the electron-vibron coupling are described by the Hamiltonian

$$\hat{H}_V = \sum_q \hbar\omega_q a^\dagger_q a_q + \sum_{\alpha\beta} \sum_q \lambda^q_{\alpha\beta} (a_q + a^\dagger_q) d^\dagger_\alpha d_\beta + \hat{H}_e. \tag{6.189}$$

Here vibrations are considered as localized phonons and q is the index labeling them, not the wave-vector. The first term describes free vibrons with the energy $\hbar\omega_q$. The second term represents the electron-vibron interaction. The third term describes the coupling to the environment and the dissipation of vibrons. We include both diagonal coupling, which originates from a change of the electrostatic energy with the distance between atoms, and the off-diagonal coupling, which can be obtained from the dependence of the matrix elements $t_{\alpha\beta}$ over the distance between atoms.

(ii) Nonequilibrium current and charge To connect the microscopic description of a system with the macroscopic (electrodynamic) equations and calculate the observables, we need the expressions for the nonequilibrium electrical charge of the system and the current between the system and the leads.

The charge in a nonequilibrium state is given by (Q_0 is the background charge)

$$Q_S(t) = e \sum_\alpha \left\langle d^\dagger_\alpha d_\alpha \right\rangle - Q_0. \tag{6.190}$$

To calculate the current we find the time evolution of the particle number operator $\hat{N}_S = \sum_\alpha d^\dagger_\alpha d_\alpha$ due to tunneling from the left ($i = L$) or right ($i = R$) contact.

The current *from* the left ($i = L$) or right ($i = R$) contact *to* the nanosystem is determined by (note, that we consider e as the charge of the electron (negative) or the hole (positive))

$$J_i(t) = -e \left\langle \left(\frac{dN_S}{dt} \right)_i \right\rangle = -\frac{ie}{\hbar} \left\langle [H^{(i)}_T, N_S] \right\rangle, \tag{6.191}$$

where

$$H_T^{(i)} = \sum_{k\sigma,\alpha} \left(V_{ik\sigma,\alpha} c_{ik\sigma}^\dagger d_\alpha + V_{ik\sigma,\alpha}^* d_\alpha^\dagger c_{ik\sigma} \right) \tag{6.192}$$

is the Hamiltonian of the coupling to the corresponding contact. The current is determined by this only part of the full Hamiltonian (6.136), because all other terms commute with $\hat{N}_S$.

Applying the commutation relation

$$\begin{aligned} \left[d_\alpha, d_\beta^\dagger d_\beta \right] = d_\alpha d_\beta^\dagger d_\beta - d_\beta^\dagger d_\beta d_\alpha = & d_\alpha d_\beta^\dagger d_\beta + d_\beta^\dagger d_\alpha d_\beta = \\ & (d_\alpha d_\beta^\dagger + \delta_{\alpha\beta} - d_\alpha d_\beta^\dagger) d_\beta = \delta_{\alpha\beta} d_\alpha, \end{aligned} \tag{6.193}$$

one obtains finally

$$J_i(t) = \frac{ie}{\hbar} \sum_{k\sigma,\alpha} \left[V_{ik\sigma,\alpha} \left\langle c_{ik\sigma}^\dagger d_\alpha \right\rangle - V_{ik\sigma,\alpha}^* \left\langle d_\alpha^\dagger c_{ik\sigma} \right\rangle \right]. \tag{6.194}$$

(iii) Density matrix and NGF The averages of the operators in Eqs. (6.190) and (6.194) are the elements of the density matrix in the single-particle space

$$\rho_{\alpha\alpha}(t) = \left\langle d_\alpha^\dagger(t) d_\alpha(t) \right\rangle, \tag{6.195}$$

$$\rho_{\alpha,ik\sigma}(t) = \left\langle c_{ik\sigma}^\dagger(t) d_\alpha(t) \right\rangle. \tag{6.196}$$

It is possible, also, to express it as a two-time Green function at equal times

$$Q_S(t) = e \sum_\alpha \rho_{\alpha\alpha}(t) = -ie \sum_\alpha G_{\alpha\alpha}^<(t,t), \tag{6.197}$$

$$J_i(t) = \frac{2e}{\hbar} \mathrm{Im} \left(\sum_{k\sigma,\alpha} V_{ik\sigma,\alpha} \rho_{\alpha,ik\sigma}(t) \right) = \frac{2e}{\hbar} \mathrm{Re} \left(\sum_{k\sigma,\alpha} V_{ik\sigma,\alpha} G_{\alpha,ik\sigma}^<(t,t) \right), \tag{6.198}$$

where we define the system-to-lead lesser Green function

$$G_{\alpha,ik\sigma}^<(t_1,t_2) = i \left\langle c_{ik\sigma}^\dagger(t_2) d_\alpha(t_1) \right\rangle, \tag{6.199}$$

while nonequilibrium charge distribution of the molecule is determined by the system lesser function

$$G_{\alpha\beta}^<(t_1,t_2) = i \left\langle d_\beta^\dagger(t_2) d_\alpha(t_1) \right\rangle. \tag{6.200}$$

One can ask: what is the advantage to use the more complex two-time Green functions instead of density matrices? There are several reasons. First of all, NGF give, as we shall see below, a clear description of both density of states and distribution of particles over this states. Then, the equations of motion including interactions and the influence of environment can be obtained with the help of

a diagrammatic technique, and (very important) all diagrammatic results of *equilibrium* theory can be easily incorporated. Retardation effects are conveniently taken into account by two-time Green functions. And, ... finally, one can always go back to the density matrix when necessary.

It is important to note, that the *single-particle* density matrix (6.195) should not be mixed up with the density matrix in the basis of *many-body eigenstates*.

In these review we consider different methods. The density matrix can be determined from the master equation. For Green functions the EOM method or Keldysh method can be applied. Traditionally, the density matrix is used in the case of very weak system-to-lead coupling, while the NGF methods are more successful in the description of strong and intermediate coupling to the leads. The convenience of one or other method is determined essentially by the type of interaction. Our aim is to combine the advantages of both methods.

6.3.2 Nonequilibrium Green functions: definition and properties

In the previous section we found, that the current through a system (as well as other observables) can be expressed through nonequilibrium Green functions. Here we give the definitions of retarded, advanced, lesser, and greater Green functions and consider some simple examples. We also introduce a very important concept of the Schwinger-Keldysh closed-time contour, and define contour Green functions. This section is a little bit technical, but we need these definitions in the next sections.

6.3.2.1 Spectral - retarded (G^R) and advanced (G^A) functions

Definition Retarded Green function for fermions is defined as

$$G^R_{\alpha\beta}(t_1,t_2) = -i\theta(t_1-t_2)\left\langle \left[c_\alpha(t_1), c^\dagger_\beta(t_2)\right]_+ \right\rangle, \tag{6.201}$$

where $c^\dagger_\alpha(t)$, $c_\alpha(t)$ are creation and annihilation time-dependent (Heisenberg) operators, $[c,d]_+ = cd+dc$ is the anti-commutator, $\langle ... \rangle$ denotes averaging over equilibrium state.

We use notations α, β, ... to denote single-particle quantum states, the other possible notation is more convenient for bulk systems

$$G^R(x_1,x_2) = -i\theta(t_1-t_2)\left\langle \left[c(x_1), c^\dagger(x_2)\right]_+ \right\rangle, \tag{6.202}$$

where $x \equiv \mathbf{r}, t, \sigma, ...$ or $x \equiv \mathbf{k}, t, \sigma, ...$, etc. Some other types of notations can be found in the literature, they are equivalent to (6.201).

The advanced function for fermions is defined as

$$G^A_{\alpha\beta}(t_1,t_2) = i\theta(t_2-t_1)\left\langle \left[c_\alpha(t_1), c^\dagger_\beta(t_2)\right]_+ \right\rangle. \tag{6.203}$$

Finally, retarded and advanced functions for *bosons* can be defined as

$$\tilde{G}^R_{\alpha\beta}(t_1,t_2) = -i\theta(t_1-t_2)\left\langle \left[a_\alpha(t_1), a^\dagger_\beta(t_2)\right]_- \right\rangle, \tag{6.204}$$

$$\tilde{G}^A_{\alpha\beta}(t_1,t_2) = i\theta(t_2-t_1)\left\langle \left[a_\alpha(t_1), a^\dagger_\beta(t_2)\right]_- \right\rangle, \tag{6.205}$$

where $a^\dagger_\alpha(t)$, $a_\alpha(t)$ are creation and annihilation boson operators, $[a,b]_- = ab - ba$ is the commutator.

Discussion of averaging The average value of any operator $\hat{O}$ can be written as $\langle\hat{O}\rangle = \langle t|\hat{O}^S|t\rangle$ in the Schrödinger representation or $\langle\hat{O}\rangle = \langle 0|\hat{O}^H(t)|0\rangle$ in the Heisenberg representation, where $|0\rangle$ is some initial state. This initial state is in principle arbitrary, but in many-particle problems it is convenient to take this state as an equilibrium state, consequently without time-dependent perturbation we obtain usual equilibrium Green functions.

In accordance with this definition the Heisenberg operators $c_\alpha(t)$, $c^\dagger_\beta(t)$, etc. are equal to the time-independent Schrödinger operators at some initial time t_0: $c_\alpha(t_0) = c_\alpha$, etc. Density matrix of the system is assumed to be equilibrium at this time $\hat{\rho}(t_0) = \hat{\rho}_{eq}$. Usually we can take $t_0 = 0$ for simplicity, but if we want to use $t_0 \neq 0$ the transformation to Heisenberg operators should be written as

$$\hat{f}^H(t) = e^{i\hat{H}(t-t_0)}\hat{f}^S e^{-i\hat{H}(t-t_0)}. \tag{6.206}$$

In fact, the initial conditions are not important because of dissipation (the memory about the initial state is completely lost after the relaxation time). However, in some pathological cases, for example for free noninteracting particles, the initial state determines the state at all times. Note also, that the initial conditions can be more convenient formulated for Green functions itself, instead of corresponding initial conditions for operators or wave functions.

Nevertheless, thermal averaging is widely used and we define it here explicitly. If we introduce the basis of exact time-independent many-particle states $|n\rangle$ with energies E_n, the averaging over equilibrium state can be written as

$$\langle\hat{O}\rangle = \frac{1}{Z}\sum_n e^{-E_n/T}\left\langle n\left|\hat{O}^H(t)\right|n\right\rangle, \quad Z = \sum_n e^{-E_n/T}. \tag{6.207}$$

In the following when we use notations like $\left\langle\hat{O}\right\rangle$ or $\left\langle\Psi\left|\hat{O}(t)\right|\Psi\right\rangle$, we assume the averaging with density matrix (density operator) $\hat{\rho}$

$$\left\langle\hat{O}\right\rangle = Sp\left(\hat{\rho}\hat{O}\right), \tag{6.208}$$

for equilibrium density matrix and Heisenberg operators it is equivalent to (6.207).

Free-particle retarded function for fermions Now consider the simplest possible example – retarded Green function for free particles (fermions).

The free-particle Hamiltonian has equivalent form if one uses Schrödinger or Heisenberg operators

$$\hat{H} = \sum_\alpha \epsilon_\alpha c_\alpha^\dagger c_\alpha = \sum_\alpha \epsilon_\alpha c_\alpha^\dagger(t) c_\alpha(t), \tag{6.209}$$

because (here we assume $t_0 = 0$)

$$\begin{aligned} c_\alpha^\dagger(t) c_\alpha(t) &= e^{i\hat{H}t} c_\alpha^\dagger e^{-i\hat{H}t} e^{i\hat{H}t} c_\alpha e^{-i\hat{H}t} \\ &= e^{i\hat{H}t} c_\alpha^\dagger c_\alpha e^{-i\hat{H}t} = c_\alpha^\dagger c_\alpha, \end{aligned} \tag{6.210}$$

where we used that $c_\alpha^\dagger c_\alpha$ is commutative with the Hamiltonian $\hat{H} = \sum_\alpha \epsilon_\alpha c_\alpha^\dagger c_\alpha$.

From the definitions (6.201) and (6.207)

$$\begin{aligned} &\left\langle \left[c_\alpha(t_1), c_\beta^\dagger(t_2) \right]_+ \right\rangle = \left\langle c_\alpha(t_1) c_\beta^\dagger(t_2) + c_\beta^\dagger(t_2) c_\alpha(t_1) \right\rangle = \\ &= \left\langle e^{i\hat{H}t_1} c_\alpha(t_1) e^{-i\hat{H}t_1} e^{i\hat{H}t_2} c_\beta^\dagger(t_2) e^{-i\hat{H}t_2} + e^{i\hat{H}t_2} c_\beta^\dagger(t_2) e^{-i\hat{H}t_2} e^{i\hat{H}t_1} c_\alpha(t_1) e^{-i\hat{H}t_1} \right\rangle = \\ &= e^{i\epsilon_\beta t_2 - i\epsilon_\alpha t_1} \left\langle c_\alpha c_\beta^\dagger + c_\beta^\dagger c_\alpha \right\rangle = e^{-i\epsilon_\alpha(t_1 - t_2)} \delta_{\alpha\beta}, \end{aligned} \tag{6.211}$$

$$\begin{aligned} G_{\alpha\beta}^R(t_1, t_2) &= -i\theta(t_1 - t_2) \left\langle \left[c_\alpha(t_1), c_\beta^\dagger(t_2) \right]_+ \right\rangle \\ &= -i\theta(t_1 - t_2) e^{-i\epsilon_\alpha(t_1 - t_2)} \delta_{\alpha\beta}, \end{aligned} \tag{6.212}$$

where we used some obvious properties of the creation and annihilation operators and commutation relations.

We consider also the other method, based on the equations of motion for operators. From Liuville – von Neuman equation we find (all c-operators are Heisenberg operators in the formula below, (t) is omitted for shortness)

$$\begin{aligned} i\frac{dc_\alpha(t)}{dt} &= [c_\alpha(t), H]_- = \sum_\beta \epsilon_\beta \left[c_\alpha, c_\beta^\dagger c_\beta \right]_- \\ &= \sum_\beta \epsilon_\beta \left(c_\alpha c_\beta^\dagger c_\beta - c_\beta^\dagger c_\beta c_\alpha \right) = \sum_\beta \epsilon_\beta \left(c_\alpha c_\beta^\dagger c_\beta + c_\beta^\dagger c_\alpha c_\beta \right) \\ &= \sum_\beta \epsilon_\beta \left(c_\alpha c_\beta^\dagger + c_\beta^\dagger c_\alpha \right) c_\beta = \sum_\beta \epsilon_\beta \delta_{\alpha\beta} c_\beta = \epsilon_\alpha c_\alpha(t), \end{aligned} \tag{6.213}$$

so that Heisenberg operators for free fermions are

$$c_\alpha(t) = e^{-i\epsilon_\alpha t} c_\alpha(0), \quad c_\alpha^\dagger(t) = e^{i\epsilon_\alpha t} c_\alpha^\dagger(0). \tag{6.214}$$

Substituting these expressions into (6.201) we obtain again (6.212). Note also that if we take $t_0 \neq 0$,

then Heisenberg operators for free fermions are

$$c_\alpha(t) = e^{-i\epsilon_\alpha(t-t_0)}c_\alpha(t_0), \quad c^\dagger_\alpha(t) = e^{i\epsilon_\alpha(t-t_0)}c^\dagger_\alpha(t_0), \tag{6.215}$$

but the result for the Green functions is just the same, because

$$\begin{aligned}\left\langle \left[c_\alpha(t_1), c^\dagger_\beta(t_2)\right]_+ \right\rangle &= \left\langle c_\alpha(t_1)c^\dagger_\beta(t_2) + c^\dagger_\beta(t_2)c_\alpha(t_1)\right\rangle = \\ &= e^{i\epsilon_\beta(t_2-t_0)-i\epsilon_\alpha(t_1-t_0)}\left\langle c_\alpha c^\dagger_\beta + c^\dagger_\beta c_\alpha \right\rangle = e^{-i\epsilon_\alpha(t_1-t_2)}\delta_{\alpha\beta}.\end{aligned} \tag{6.216}$$

It is interesting to make Fourie-transform of this function. In equilibrium two-time function $G^R_{\alpha\beta}(t_1,t_2)$ is a function of the time difference only, so that we define transform over time difference (t_1-t_2)

$$G^R(\epsilon) = \int_0^\infty G^R(t_1-t_2)e^{i(\epsilon+i0)(t_1-t_2)}d(t_1-t_2), \tag{6.217}$$

we add infinitely small positive complex part to ϵ to make this integral well defined in the upper limit (this is necessary for free particles without dissipation because function (6.212) oscillates at large times $\tau = t_1 - t_2$ and the integral (6.217) can not be calculated without $i0$ term. Then we obtain

$$G^R_{\alpha\beta}(\epsilon) = \frac{\delta_{\alpha\beta}}{\epsilon - \epsilon_\alpha + i0}. \tag{6.218}$$

More generally, transformation (6.217) can be considered as the Laplas transformation with complex argument $z = \epsilon + i\eta$.

For advanced function

$$G^A_{\alpha\beta}(t_1,t_2) = i\theta(t_2-t_1)e^{-i\epsilon_\alpha(t_1-t_2)}\delta_{\alpha\beta}, \tag{6.219}$$

the Fourier transform is given by

$$G^A(\epsilon) = \int_{-\infty}^0 G^A(t_1-t_2)e^{i(\epsilon-i0)(t_1-t_2)}d(t_1-t_2), \tag{6.220}$$

with other sign of the term $i0$.

Spectral function Finally, we introduce the important combination of retarded and advanced functions known as *spectral* or *spectral weight* function

$$A_{\alpha\beta}(\epsilon) = i\left(G^R_{\alpha\beta}(\epsilon) - G^A_{\alpha\beta}(\epsilon)\right), \tag{6.221}$$

in equilibrium case Fourie-transformed retarded and advanced functions are complex conjugate $G^A(\epsilon) = \left(G^R(\epsilon)\right)^*$, and $A_{\alpha\beta}(\epsilon) = -2\mathrm{Im}G^R_{\alpha\beta}(\epsilon)$.

For free fermions the spectral function is

$$A_{\alpha\beta}(\epsilon) = -2\mathrm{Im}\left(\frac{\delta_{\alpha\beta}}{\epsilon - \epsilon_\alpha + i0}\right) = 2\pi\delta(\epsilon - \epsilon_\alpha)\delta_{\alpha\beta}. \tag{6.222}$$

The result is transparent – the function $A_{\alpha\beta}(\epsilon)$ is nonzero only at particle eigen-energies, so that

$$\rho(\epsilon) = \frac{1}{2\pi}\mathrm{Sp}A_{\alpha\beta}(\epsilon) = \frac{1}{2\pi}\sum_\alpha A_{\alpha\alpha}(\epsilon) = \sum_\alpha \delta(\epsilon - \epsilon_\alpha) \tag{6.223}$$

is the usual energy density of states. Note that the imaginary part $i0$ is necessary to obtain this result, thus it is not only mathematical trick, but reflects the physical sense of the retarded Green function.

If we introduce finite relaxation time

$$G^R_{\alpha\beta}(\tau) = -i\theta(\tau)e^{-i\epsilon_\alpha\tau - \gamma\tau}\delta_{\alpha\beta}, \tag{6.224}$$

then the spectral function has familiar Lorentzian form

$$A_{\alpha\beta}(\epsilon) = \frac{2\gamma\delta_{\alpha\beta}}{(\epsilon - \epsilon_\alpha)^2 + \gamma^2}. \tag{6.225}$$

Finally, spectral function has a special property, so-called *sum rule*, namely

$$\int_{-\infty}^{\infty} A_{\alpha\beta}(\epsilon)\frac{d\epsilon}{2\pi} = \delta_{\alpha\beta}. \tag{6.226}$$

6.3.2.2 Kinetic - lesser ($G^<$) and greater ($G^>$) functions

Definition Spectral functions, described before, determine single-particle properties of the system, such as quasiparticle energy, broadening of the levels (life-time), and density of states. These functions can be modified in nonequilibrium state, but most important *kinetic* properties, such as distribution function, charge, and current, are determined by lesser Green function

$$G^<_{\alpha\beta}(t_1, t_2) = i\left\langle c^\dagger_\beta(t_2)c_\alpha(t_1)\right\rangle. \tag{6.227}$$

Indeed, density matrix is the same as equal-time lesser function

$$\rho_{\alpha\beta}(t) = \left\langle c^\dagger_\beta(t)c_\alpha(t)\right\rangle = -iG^<_{\alpha\beta}(t, t). \tag{6.228}$$

the number of particles in state $|\alpha\rangle$ (distribution function) is

$$n_\alpha(t) = \left\langle c^\dagger_\alpha(t)c_\alpha(t)\right\rangle = -iG^<_{\alpha\alpha}(t, t), \tag{6.229}$$

the tunneling current is

$$J(t) = \frac{ie}{\hbar}\sum_{kq}\left[V_{qk}\left\langle c_q^\dagger(t)c_k(t)\right\rangle - V_{qk}^*\left\langle c_k^\dagger(t)c_q(t)\right\rangle\right]$$
$$= \frac{2e}{\hbar}\mathrm{Re}\left(\sum_{kq}V_{qk}G_{kq}^<(t,t)\right). \tag{6.230}$$

In addition to the lesser the other (greater) function is used

$$G_{\alpha\beta}^>(t_1,t_2) = -i\left\langle c_\alpha(t_1)c_\beta^\dagger(t_2)\right\rangle. \tag{6.231}$$

For bosons lesser and greater functions are defined as

$$\tilde{G}_{\alpha\beta}^<(t_1,t_2) = -i\left\langle a_\beta^\dagger(t_2)a_\alpha(t_1)\right\rangle, \tag{6.232}$$

$$\tilde{G}_{\alpha\beta}^>(t_1,t_2) = -i\left\langle a_\alpha(t_1)a_\beta^\dagger(t_2)\right\rangle. \tag{6.233}$$

The name "lesser" originates from the time-ordered Green function, the main function in equilibrium theory, which can be calculated by diagrammatic technique

$$G_{\alpha\beta}(t_1,t_2) = -i\left\langle T\left(c_\alpha(t_1)c_\beta^\dagger(t_2)\right)\right\rangle, \tag{6.234}$$

$$G_{\alpha\beta}(t_1,t_2) = \begin{cases} -i\left\langle c_\alpha(t_1)c_\beta^\dagger(t_2)\right\rangle & \text{if} \quad t_1 > t_2 \quad \Rightarrow \quad G_{\alpha\beta}\equiv G_{\alpha\beta}^>, \\ i\left\langle c_\beta^\dagger(t_2)c_\alpha(t_1)\right\rangle & \text{if} \quad t_1 < t_2 \quad \Rightarrow \quad G_{\alpha\beta}\equiv G_{\alpha\beta}^<, \end{cases} \tag{6.235}$$

here additional sing minus appears for interchanging of fermionic creation-annihilation operators. Lesser means that $t_1 < t_2$.

From the definitions it is clear that the retarded function can be combined from lesser and greater functions

$$G_{\alpha\beta}^R(t_1,t_2) = \theta(t_1-t_2)\left[G_{\alpha\beta}^>(t_1,t_2) - G_{\alpha\beta}^<(t_1,t_2)\right]. \tag{6.236}$$

Free-particle lesser function for fermions Now let us consider again free fermions. Heisenberg operators for free fermions are ($t_0 = 0$)

$$c_\alpha(t) = e^{-i\epsilon_\alpha t}c_\alpha(0), \quad c_\alpha^\dagger(t) = e^{i\epsilon_\alpha t}c_\alpha^\dagger(0). \tag{6.237}$$

Lesser function is

$$G_{\alpha\beta}^<(t_1,t_2) = i\left\langle c_\beta^\dagger(t_2)c_\alpha(t_1)\right\rangle = ie^{i\epsilon_\beta t_2 - i\epsilon_\alpha t_1}\left\langle c_\beta^\dagger c_\alpha\right\rangle$$
$$= ie^{-i\epsilon_\alpha(t_1-t_2)}f^0(\epsilon_\alpha)\delta_{\alpha\beta}, \tag{6.238}$$

one sees that contrary to the retarded function, the lesser function is proportional to the distribution function, in equilibrium this is Fermi distribution function

$$f^0(\epsilon) = \frac{1}{e^{\frac{\epsilon-\mu}{T}}+1}. \tag{6.239}$$

It is interesting to compare this answer with the result for *nonthermal* initial conditions. Assume that initial state is described by the density matrix $\rho^0_{\alpha\beta} = \left\langle c^\dagger_\beta c_\alpha \right\rangle$, now with nonzero off-diagonal elements. Time dependence of the density matrix is given by

$$\rho_{\alpha\beta}(t) = e^{i(\epsilon_\beta - \epsilon_\alpha)t}\rho^0_{\alpha\beta}. \tag{6.240}$$

We obtain the well known result that off-diagonal elements oscillate in time.

Now define Fourier-transform for lesser function ($\tau = t_1 - t_2$)

$$G^<(\epsilon) = \int_{-\infty}^{\infty} G^<(\tau) e^{i[\epsilon + i0\mathrm{sign}(\tau)]\tau} d\tau, \tag{6.241}$$

note that here we use Fourie-transform with complicated term $i0\mathrm{sign}(\tau)$, which makes this transformation consistent with previously introduced transformations (6.217) for retarded ($\tau > 0$) and (6.220) advanced ($\tau < 0$) functions.

Applying this transformation to (6.238) we obtain

$$\begin{aligned} G^<_{\alpha\beta}(\epsilon) =& i f^0(\epsilon_\alpha)\delta_{\alpha\beta} \int_{-\infty}^{\infty} e^{+i[\epsilon - \epsilon_\alpha + i0\mathrm{sign}(\tau)]\tau} d\tau \\ &= 2\pi i f^0(\epsilon_\alpha)\delta(\epsilon - \epsilon_\alpha)\delta_{\alpha\beta}. \end{aligned} \tag{6.242}$$

For free fermion greater function one obtaines

$$G^>_{\alpha\beta}(t_1, t_2) = -ie^{-i\epsilon_\alpha(t_1 - t_2)}(1 - f^0(\epsilon_\alpha))\delta_{\alpha\beta}, \tag{6.243}$$

$$G^>_{\alpha\beta}(\epsilon) = -2\pi i(1 - f^0(\epsilon_\alpha))\delta(\epsilon - \epsilon_\alpha)\delta_{\alpha\beta}. \tag{6.244}$$

Equilibrium case. Fluctuation-dissipation theorem Now we want to consider some general properties of interacting systems. In equilibrium the lesser function is not independent and is simply related to the spectral function by the relation

$$G^<_{\alpha\beta}(\epsilon) = iA_{\alpha\beta}(\epsilon) f^0(\epsilon). \tag{6.245}$$

This relation is important because establish equilibrium initial condition for nonequilibrium lesser function, and propose useful Ansatz if equilibrium distribution function $f^0(\epsilon)$ is replaced by some unknown nonequilibrium function.

Here we prove this relation using *Lehmann representation* – quite useful method in the theory of

Green functions. The idea of the method is to use exact many-particle eigenstates $|n\rangle$, even if they are not explicitly known.

Consider first the greater function. Using states $|n\rangle$ we represent this function as

$$G^{>}_{\alpha\beta}(t_1,t_2) = -i\left\langle c_\alpha(t_1)c^\dagger_\beta(t_2)\right\rangle = -\frac{i}{Z}\sum_n \left\langle n\left|e^{-\hat{H}/T}c_\alpha(t_1)c^\dagger_\beta(t_2)\right|n\right\rangle = \\ = -\frac{i}{Z}\sum_{nm} e^{-E_n/T}\langle n|c_\alpha|m\rangle\langle m|c^\dagger_\beta|n\rangle e^{i(E_n-E_m)(t_1-t_2)}. \tag{6.246}$$

In Fourie representation

$$G^{>}_{\alpha\beta}(\epsilon) = -\frac{2\pi i}{Z}\sum_{nm} e^{-E_n/T}\langle n|c_\alpha|m\rangle\langle m|c^\dagger_\beta|n\rangle\delta(E_n-E_m+\epsilon). \tag{6.247}$$

Similarly, for the lesser function we find

$$G^{<}_{\alpha\beta}(\epsilon) = \frac{2\pi i}{Z}\sum_{nm} e^{-E_m/T}\langle n|c^\dagger_\beta|m\rangle\langle m|c_\alpha|n\rangle\delta(E_m-E_n+\epsilon). \tag{6.248}$$

Now we can use these expressions to obtain some general properties of Green functions without explicit calculation of the matrix elements. Exchanging indices n and m in the expression (6.248) and taking into account that $E_m = E_n - \epsilon$ because of delta-function, we see that

$$G^{>}_{\alpha\beta}(\epsilon) = -e^{-\epsilon/T}G^{<}_{\alpha\beta}(\epsilon). \tag{6.249}$$

From this expression and relation (6.236), which can be written as

$$A_{\alpha\beta}(\epsilon) = i\left[G^{>}_{\alpha\beta}(\epsilon) - G^{<}_{\alpha\beta}(\epsilon)\right] \tag{6.250}$$

we derive (6.245).

6.3.2.3 Interaction representation

In the previous lectures we found that nonequilibrium Green functions can be quite easy calculated for free particles, and equations of motion for one-particle Green functions (the functions which are the averages of two creation-annihilation operators) can be formulated if we add interactions and time-dependent perturbations, but these equations include high-order Green functions (the averages of three, four, and larger number of operators). The equations can be truncated and formulated in terms of one-particle Green functions in some simple approximations. However, systematic approach is needed to proceed with perturbation expansion and self-consistent methods (all together is known as *diagrammatic approach*). The main idea of the diagrammatic approach is to start from some "simple" Hamiltonian (usually for free particles) and, treating interactions and external fields as a perturbation, formulate perturbation expansion, and summarize all most important terms (diagrams) *in all orders of perturbation theory*. The

result of such procedure gives, in principle, *nonperturbative* description (ordinary mean-field theory is the simplest example). The starting point of the method is so-called *interaction representation*.

Let us consider the full Hamiltonian $\hat{H}$ as the sum of a *free-particle* time-independent part $\hat{H}_0$ and (possibly time-dependent) perturbation $\hat{V}(t)$ (note that this "perturbation" should not be necessarily small)

$$\hat{H} = \hat{H}_0 + \hat{V}(t). \tag{6.251}$$

We define new operators in *interaction representation* by

$$\hat{f}^I(t) = e^{i\hat{H}_0 t}\hat{f}^S e^{-i\hat{H}_0 t}, \tag{6.252}$$

where $\hat{f}^S$ is the time-independent Schrödinger operator. This is equivalent to the time-dependent Heisenberg operator, defined by the part $\hat{H}_0$ of the Hamiltonian. For a free-particle Hamiltonian $\hat{H}_0$ the operators $\hat{f}^I(t)$ can be calculated exactly.

A new wave function corresponding to (6.252) is

$$\Psi^I(t) = e^{i\hat{H}_0 t}\Psi^S(t). \tag{6.253}$$

It is easy to see that transformation (6.252), (6.253) is unitary transformation and conserves the average value of any operator

$$\langle\Psi^S|\hat{f}^S|\Psi^S\rangle = \langle\Psi^I|\hat{f}^I|\Psi^I\rangle. \tag{6.254}$$

Substituting (6.253) into the ordinary Schrödinger equation, we derive the equation

$$i\frac{\partial\Psi^I}{\partial t} = \hat{V}^I(t)\Psi^I, \tag{6.255}$$

where $\hat{V}^I(t) = e^{i\hat{H}_0 t}\hat{V}^S(t)e^{-i\hat{H}_0 t}$ is in the intreraction representation.

Equation (6.255) seems to be quite simple, however the operator nature of $\hat{V}$ makes this problem nontrivial. Indeed, consider a small time-step Δt. Then

$$\Psi(t+\Delta t) = \left[1 - i\hat{V}^S(t)\Delta t\right]\Psi(t) = \exp^{-i\hat{V}^S(t)\Delta t}\Psi(t), \tag{6.256}$$

linear in Δt term can be transformed into the exponent if we understand the exponential function of the operator in the usual way

$$\exp^{\hat{A}} = 1 + \hat{A} + \frac{1}{2!}\hat{A}^2 + ... + \frac{1}{n!}\hat{A}^n + ..., \tag{6.257}$$

and assume that only linear term should be taken at $\Delta t \to 0$.

If we now repeat this procedure at times t_i with step Δt, we obtain finally

$$\Psi^I(t) = \hat{S}(t, t_0)\Psi^I(t_0), \tag{6.258}$$

with

$$\hat{S}(t,t_0) = \prod_{t_i=t_0}^{t} \exp\left(-i\hat{V}^I(t_i)\Delta t\right), \tag{6.259}$$

this product, however, is not simply $\exp\left(-i\int_{t_0}^{t}\hat{V}^I(t')dt'\right)$ in the limit $\Delta t \to 0$, because operators $\hat{V}^I(t')$ are not commutative at different times, and for two noncommutative operators $\hat{A}$ and $\hat{B}$ $e^{\hat{A}+\hat{B}} \neq e^{\hat{A}}e^{\hat{B}}$.

In the product (6.259) operators at earlier times should be applied first, before operators at later times. In the limit $\Delta t \to 0$ we obtain

$$\hat{S}(t,t_0) = T\exp\left(-i\int_{t_0}^{t}\hat{V}^I(t')dt'\right), \tag{6.260}$$

where T is the time-ordering operator ("-" for fermionic operators)

$$T\left(\hat{A}(t_1)\hat{B}(t_2)\right) = \begin{cases} \hat{A}(t_1)\hat{B}(t_2) & \text{if} \quad t_1 > t_2, \\ \pm\hat{B}(t_2)\hat{A}(t_1) & \text{if} \quad t_1 < t_2. \end{cases} \tag{6.261}$$

Of cause, expression (6.260) is defined only in the sense of expansion (6.257). Consider for example the second-order term in the time-ordered expansion.

$$\begin{aligned} T\left[\int_{t_0}^{t}\hat{V}^I(t')dt'\right]^2 &= T\left[\int_{t_0}^{t}\hat{V}^I(t')dt'\int_{t_0}^{t}\hat{V}^I(t'')dt''\right] = \\ &= \int_{t_0}^{t}dt'\int_{t_0}^{t'}dt''\hat{V}^I(t')\hat{V}^I(t'') + \int_{t_0}^{t}dt''\int_{t_0}^{t''}dt'\hat{V}^I(t'')\hat{V}^I(t'). \end{aligned} \tag{6.262}$$

If we exchange t' and t'' in the second integral, we see finally that

$$T\left[\int_{t_0}^{t}\hat{V}^I(t')dt'\right]^2 = 2\int_{t_0}^{t}dt'\int_{t_0}^{t'}dt''\hat{V}^I(t')\hat{V}^I(t''). \tag{6.263}$$

Properties of $\hat{S}(t,t_0)$ $\hat{S}$ is the unitary operator and

$$\hat{S}^{-1}(t,t_0) = \hat{S}^{\dagger}(t,t_0) = \tilde{T}\exp\left(i\int_{t_0}^{t}\hat{V}^I(t')dt'\right), \tag{6.264}$$

where $\tilde{T}$ is time-anti-ordering operator. Some other important properties are

$$\hat{S}^{-1}(t,t_0) = \hat{S}(t_0,t), \tag{6.265}$$

$$\hat{S}(t_3,t_2)\hat{S}(t_2,t_1) = \hat{S}(t_3,t_1), \tag{6.266}$$

$$\hat{S}^{-1}(t_2,t_1)\hat{S}^{-1}(t_3,t_2) = \hat{S}^{-1}(t_3,t_1). \tag{6.267}$$

Finally, we need the expression of a Heisenberg operator, defined by the full Hamiltonian $\hat{H} = \hat{H}_0 + \hat{V}(t)$,

through an operator in the interaction representation. The transformation, corresponding to (6.258), is given by

$$\hat{f}^H(t) = e^{-i\hat{H}_0 t_0}\hat{S}^{-1}(t,t_0)\hat{f}^I(t)\hat{S}(t,t_0)e^{i\hat{H}_0 t_0}, \tag{6.268}$$

and the state $\Psi^I(t_0)$ is related to the Heisenberg time-independent wave function by

$$\Psi^I(t_0) \equiv e^{i\hat{H}_0 t_0}\Psi^S(t_0) = e^{i\hat{H}_0 t_0}\Psi^H, \tag{6.269}$$

in accordance with our previous discussion of averaging we assume that at time $t = t_0$ Heisenberg operators coincide with time-independent Schrödinger operators $\hat{f}^H(t_0) = \hat{f}^S$, and Schrödinger wave function coincides at the same time with Heisenberg time-independent wave function $\Psi^S(t_0) = \Psi^H$. To avoid these additional exponents in (6.268) we can redefine the transformation to the interaction representation as

$$\hat{f}^I(t) = e^{i\hat{H}_0(t-t_0)}\hat{f}^S e^{-i\hat{H}_0(t-t_0)}, \tag{6.270}$$

in accordance with the transformation (6.206) for time-independent Hamiltonian. Previously we showed that free-particle Green functions are not dependent on t_0 for equilibrium initial condition, if we want to consider some nontrivial initial conditions, it is easier to formulate these conditions directly for Green functions. Thus below we shall use relations

$$\hat{f}^H(t) = \hat{S}^{-1}(t,t_0)\hat{f}^I(t)\hat{S}(t,t_0), \tag{6.271}$$

and

$$\Psi^I(t_0) \equiv \Psi^S(t_0) = \Psi^H. \tag{6.272}$$

Green functions in the interaction representation Consider, for example, the lesser function

$$G^<_{\alpha\beta}(t_1,t_2) = i\left\langle c^\dagger_\beta(t_2)c_\alpha(t_1)\right\rangle = i\left\langle \Psi^H\left|c^\dagger_\beta(t_2)c_\alpha(t_1)\right|\Psi^H\right\rangle, \tag{6.273}$$

c-operators here are Heisenberg operators and they should be replaced by operators $c^I(t) \equiv \tilde{c}(t)$ in the interaction representation:

$$G^<_{\alpha\beta}(t_1,t_2) = i\left\langle \Psi^H\left|\hat{S}^{-1}(t_2,t_0)\tilde{c}^\dagger_\beta(t_2)\hat{S}(t_2,t_0)\hat{S}^{-1}(t_1,t_0)\tilde{c}_\alpha(t_1)\hat{S}(t_1,t_0)\right|\Psi^H\right\rangle. \tag{6.274}$$

Using properties of $\hat{S}$ operators, we rewrite this expression as

$$G^<_{\alpha\beta}(t_1,t_2) = i\left\langle \hat{S}(t_0,t_2)\tilde{c}^\dagger_\beta(t_2)\hat{S}(t_2,t_1)\tilde{c}_\alpha(t_1)\hat{S}(t_1,t_0)\right\rangle. \tag{6.275}$$

6.3.2.4 Schwinger-Keldysh time contour and contour functions

Closed time-path integration Now let us introduce one useful trick, so-called *closed time-path contour of integration*. First, note that the expression of the type

$$\begin{aligned}\hat{f}^H(t) = \hat{S}^{-1}(t,t_0)\hat{f}^I(t)\hat{S}(t,t_0)\\ = \tilde{T}e^{i\int_{t_0}^{t}\hat{V}^I(t')dt'}\hat{f}^I(t)Te^{-i\int_{t_0}^{t}\hat{V}^I(t')dt'},\end{aligned} \tag{6.276}$$

can be written as

$$\hat{f}^H(t) = T_{C_t}\exp\left(-i\int_{C_t}\hat{V}^I(t')dt'\right)\hat{f}^I(t), \tag{6.277}$$

where the integral is taken along closed time contour from t_0 to t and then back from t to t_0

$$\int_{C_t}dt' = \int_{t_0}^{t}dt' + \int_{t}^{t_0}dt', \tag{6.278}$$

contour time-ordering operator T_{C_t} works along the contour C_t, it means that for times $t^{\rightarrow}$ it is usual time-ordering operator T, and for times $t^{\leftarrow}$ it is anti-time-ordering operator $\tilde{T}$. Symbolically

$$T_{C_t}\int_{C_t}dt' = T\int_{\rightarrow}dt' + \tilde{T}\int_{\leftarrow}dt'. \tag{6.279}$$

Consider now the application of this closed time-path contour to calculation of Green functions. It is convenient to start from the time-ordered function at $t_2 > t_1$

$$\left\langle T\left(\hat{B}(t_2)\hat{A}(t_1)\right)\right\rangle = \left\langle \hat{S}(t_0,t_2)\tilde{B}(t_2)\hat{S}(t_2,t_1)\tilde{A}(t_1)\hat{S}(t_1,t_0)\right\rangle, \tag{6.280}$$

here $\hat{A}(t)$ and $\hat{B}(t)$ are Heisenberg operators, $\tilde{A}(t)$ and $\tilde{B}(t)$ are operators in the interaction representation, in the case of fermionic operators the additional minus should be added for any permutation of two operators.

Using the properties of the $\hat{S}$-operator, we transform this expression as

$$\begin{aligned}\left\langle \hat{S}(t_0,t_2)\tilde{B}(t_2)\hat{S}(t_2,t_1)\tilde{A}(t_1)\hat{S}(t_1,t_0)\right\rangle = \left\langle \hat{S}^{-1}(t_2,t_0)\tilde{B}(t_2)\hat{S}(t_2,t_1)\tilde{A}(t_1)\hat{S}(t_1,t_0)\right\rangle =\\ = \left\langle \hat{S}^{-1}(\infty,t_0)\hat{S}(\infty,t_2)\tilde{B}(t_2)\hat{S}(t_2,t_1)\tilde{A}(t_1)\hat{S}(t_1,t_0)\right\rangle = \left\langle \hat{S}^{-1}T\left(\tilde{B}(t_2)\tilde{A}(t_1)\hat{S}\right)\right\rangle,\end{aligned} \tag{6.281}$$

where we defined operator

$$\hat{S} = \hat{S}(\infty,t_0). \tag{6.282}$$

Using contour integration, it can be written as

$$\left\langle T\left(\hat{B}(t_2)\hat{A}(t_1)\right)\right\rangle = \left\langle T_C\left(\hat{S}_C\tilde{B}(t_2^{\rightarrow})\tilde{A}(t_1^{\rightarrow})\right)\right\rangle, \tag{6.283}$$

$$\hat{S}_C = T_C \exp\left(-i\int_C \hat{V}^I(t')dt'\right), \tag{6.284}$$

contour C goes from t_0 trough t_1 and t_2, and back to t_0. If $t_2 > t_1$ it is obvious that contour ordering along $C^{\rightarrow}$ gives the terms from $\hat{S}(t_1,t_0)$ to $\hat{B}(t_2)$ in (6.280). The integral over the back path $C^{\leftarrow}$ gives

$$\begin{aligned} T_C \exp\left(-i\int_{\leftarrow} \hat{V}^I(t')dt'\right) &= \tilde{T}\exp\left(-i\int_{t_2}^{t_0} \hat{V}^I(t')dt'\right) = \\ = \tilde{T}\exp\left(i\int_{t_0}^{t_2} \hat{V}^I(t')dt'\right) &= \hat{S}^{-1}(t_2,t_0) = \hat{S}(t_0,t_2). \end{aligned} \tag{6.285}$$

For $t_2 < t_1$ the operators in (6.280) are reordered by T-operator and we again obtain (6.283).

The lesser and greater functions are not time-ordered and arguments of the operators are not affected by time-ordering operator. Nevertheless we can write such functions in the same form (6.283). The trick is to use one time argument from the forward contour and the other from the backward contour, for example

$$\left\langle \hat{B}(t_2)\hat{A}(t_1)\right\rangle = \left\langle T_C\left(\hat{S}_C\tilde{B}(t_2^{\leftarrow})\tilde{A}(t_1^{\rightarrow})\right)\right\rangle, \tag{6.286}$$

here the time t_1 is always before t_2.

Contour (contour-ordered) Green function Now we are able to define *contour* or *contour-ordered* Green function – the useful tool of Keldysh diagrammatic technique. The definition is similar to the previous one

$$G^C_{\alpha\beta}(\tau_1,\tau_2) = -i\left\langle T_C\left(c_\alpha(\tau_1)c^\dagger_\beta(\tau_2)\right)\right\rangle, \tag{6.287}$$

where, however, τ_1 and τ_2 are contour times. This function includes all nonequilibrium Green functions introduced before. Indeed, depending on contour position of times we obtain lesser, greater, or time-ordered functions (below we give different notations used in the literature)

$$G^C_{\alpha\beta}(\tau_1,\tau_2) = \begin{cases} \tau_1,\tau_2 \in C^{\rightarrow}: & -i\left\langle Tc_\alpha(t_1)c^\dagger_\beta(t_2)\right\rangle \implies G^{--} \quad or \quad G^T(t_1,t_2), \\ \tau_1 \in C^{\leftarrow}, \tau_2 \in C^{\rightarrow}: & -i\left\langle c_\alpha(t_1)c^\dagger_\beta(t_2)\right\rangle \implies G^{+-} \quad or \quad G^{>}(t_1,t_2), \\ \tau_1 \in C^{\rightarrow}, \tau_2 \in C^{\leftarrow}: & i\left\langle c^\dagger_\beta(t_2)c_\alpha(t_1)\right\rangle \implies G^{-+} \quad or \quad G^{<}(t_1,t_2), \\ \tau_1,\tau_2 \in C^{\leftarrow}: & -i\left\langle \tilde{T}c_\alpha(t_1)c^\dagger_\beta(t_2)\right\rangle \implies G^{++} \quad or \quad G^{\tilde{T}}(t_1,t_2). \end{cases} \tag{6.288}$$

These four functions are not independent, from definitions it follows that

$$G^< + G^> = G^T + G^{\tilde{T}}, \tag{6.289}$$

and anti-hermitian relations

$$G^T_{\alpha\beta}(t_1,t_2) = -G^{T^*}_{\beta\alpha}(t_2,t_1), \tag{6.290}$$

$$G^<_{\alpha\beta}(t_1,t_2) = -G^{<^*}_{\beta\alpha}(t_2,t_1), \tag{6.291}$$

$$G^>_{\alpha\beta}(t_1,t_2) = -G^{>^*}_{\beta\alpha}(t_2,t_1). \tag{6.292}$$

It is more convenient to use retarded and advanced functions instead of time-ordered functions. There is a number of ways to express G^R and G^A through above defined functions

$$G^R = \theta(t_1-t_2)\left[G^> - G^<\right] = G^T - G^< = G^> - G^{\tilde{T}}, \tag{6.293}$$

$$G^A = \theta(t_2-t_1)\left[G^< - G^>\right] = G^T - G^> = G^< - G^{\tilde{T}}. \tag{6.294}$$

Contour Green function in the interaction representation In the interaction representation one should repeat the calculations performed before and given the expressions (6.275), (6.280), and then replace usual times by contour times τ, so we obtain

$$\begin{aligned}&\left\langle T_C\left(c_\alpha(\tau_1)c^\dagger_\beta(\tau_2)\right)\right\rangle \\ &= \left\langle T_C\left(\hat{S}(\tau_0,\tau_2)\tilde{c}^\dagger_\beta(\tau_2)\hat{S}(\tau_2,\tau_1)\tilde{c}_\alpha(\tau_1)\hat{S}(\tau_1,\tau_0)\right)\right\rangle.\end{aligned} \tag{6.295}$$

Using contour integration, it can be written as

$$\begin{aligned}G^C_{\alpha\beta}(\tau_1,\tau_2) &= -i\left\langle T_C\left(c_\alpha(\tau_1)c^\dagger_\beta(\tau_2)\right)\right\rangle \\ &= -i\left\langle T_C\left(\hat{S}_C\tilde{c}_\alpha(\tau_1)\tilde{c}^\dagger_\beta(\tau_2)\right)\right\rangle,\end{aligned} \tag{6.296}$$

$$\hat{S}_C = T_C\exp\left(-i\int_C \hat{V}^I(t')dt'\right), \tag{6.297}$$

6.3.3 Current through a nanosystem: Meir-Wingreen-Jauho formula

Now we consider the central point of the NGF transport theory through nanosystems - the Meir-Wingreen-Jauho current formula [106, 107, 53]. This important expression shows that the current can be calculated, if the spectral and kinetic Green functions of the central system are known, and it is exact in the case of noninteracting leads. The details of the derivation can be found in the above cited papers, so we only briefly outline it.

(i) Derivation by the NGF method In the absence of interactions in the leads (besides the tunneling) one can derive the following exact expression for the lead-system function:

$$G^<_{\alpha,ik\sigma}(\epsilon) = \sum_\beta V^*_{ik\sigma,\beta}\left[G^R_{\alpha\beta}(\epsilon)g^<_{ik\sigma}(\epsilon) + G^<_{\alpha\beta}(\epsilon)g^A_{ik\sigma}(\epsilon)\right], \tag{6.298}$$

where $g^{<}_{ik\sigma}(\epsilon)$ and $g^{A}_{ik\sigma}(\epsilon)$ are Green functions of *isolated* leads, Substituting it into (6.198), we obtain for the current

$$J_i(t) = \frac{2e}{\hbar}\int\frac{d\epsilon}{2\pi}\mathrm{Re}\left[\sum_{k\sigma,\alpha\beta} V_{ik\sigma,\alpha}V^{*}_{ik\sigma,\beta}\left[G^{R}_{\alpha\beta}(\epsilon)g^{<}_{ik\sigma}(\epsilon) + G^{<}_{\alpha\beta}(\epsilon)g^{A}_{ik\sigma}(\epsilon)\right]\right]. \tag{6.299}$$

For equilibrium right or left lead Green functions we obtain directly

$$g^{<}_{k\sigma}(t_1-t_2) = i\left\langle c^{\dagger}_{k\sigma}(t_2)c_{k\sigma}(t_1)\right\rangle = if^{0}_{\sigma}(\epsilon_{k\sigma})e^{-i(\epsilon_{k\sigma}+e\varphi)(t_1-t_2)}, \tag{6.300}$$

$$g^{R}_{k\sigma}(t_1-t_2) = -i\theta(t_1-t_2)\left\langle\left[c_{k\sigma}(t_1), c^{\dagger}_{k\sigma}(t_2)\right]_{+}\right\rangle = -i\theta(t_1-t_2)e^{-i(\epsilon_{k\sigma}+e\varphi)(t_1-t_2)}, \tag{6.301}$$

$$g^{A}_{k\sigma}(t_1-t_2) = i\theta(t_2-t_1)\left\langle\left[c_{k\sigma}(t_1), c^{\dagger}_{k\sigma}(t_2)\right]_{+}\right\rangle = i\theta(t_2-t_1)e^{-i(\epsilon_{k\sigma}+e\varphi)(t_1-t_2)}, \tag{6.302}$$

or after the Fourier transform

$$g^{<}_{k\sigma}(\epsilon) = \int g^{<}_{k\sigma}(t_1-t_2)e^{i\epsilon(t_1-t_2)}d(t_1-t_2) = 2\pi i f^{0}_{\sigma}(\epsilon_{k\sigma})\delta(\epsilon-\epsilon_{k\sigma}-e\varphi), \tag{6.303}$$

$$g^{>}_{k\sigma}(\epsilon) = -2\pi i[1-f^{0}_{\sigma}(\epsilon_{k\sigma})]\delta(\epsilon-\epsilon_{k\sigma}-e\varphi), \tag{6.304}$$

$$g^{R}_{k\sigma}(\epsilon) = \frac{1}{\epsilon-\epsilon_{k\sigma}-e\varphi+i0}, \tag{6.305}$$

$$g^{A}_{k\sigma}(\epsilon) = \frac{1}{\epsilon-\epsilon_{k\sigma}-e\varphi-i0}, \tag{6.306}$$

$$f^{0}_{\sigma}(\epsilon) = \frac{1}{\exp\left(\frac{\epsilon-\mu_\sigma}{T}\right)+1}. \tag{6.307}$$

Using the level-width function (below without *spin polarization* of the leads)

$$\mathbf{\Gamma}_{i=L(R)}(\epsilon) \equiv \Gamma_{i\alpha\beta}(\epsilon) = 2\pi\sum_{k\sigma} V_{ik\sigma,\beta}V^{*}_{ik\sigma,\alpha}\delta(\epsilon-\epsilon_{ik\sigma}) = 2\pi\sum_{\sigma}\rho_{i\sigma}(\epsilon)V_{i\sigma,\beta}(\epsilon)V^{*}_{i\sigma,\alpha}(\epsilon), \tag{6.308}$$

and changing the momentum summation to the energy integration $\sum_k \Rightarrow \int \rho(\epsilon_k)d\epsilon_k$, we obtain the following expression for the current

$$J_{i=L,R} = \frac{ie}{\hbar}\int\frac{d\epsilon}{2\pi}\mathrm{Tr}\left\{\mathbf{\Gamma}_i(\epsilon-e\varphi_i)\left(\mathbf{G}^{<}(\epsilon) + f^{0}_{i}(\epsilon-e\varphi_i)\left[\mathbf{G}^{R}(\epsilon)-\mathbf{G}^{A}(\epsilon)\right]\right)\right\}, \tag{6.309}$$

where f^{0}_{i} is the equilibrium Fermi distribution function with chemical potential μ_i. Thus, we obtain the well-known Meir-Wingreen formula. Note, that we use explicitly the electrical potential of the leads in this expression. It is important to mention, that at finite voltage the arguments of the left and right level-width functions are changed in a different way, which means, in particular, that the known condition of proportional coupling $\mathbf{\Gamma}_L = \lambda\mathbf{\Gamma}_R$ can be fulfilled only in the wide-band limit, when both functions are energy independent.

(ii) Different forms of the MWJ formula In a stationary state $J_R = -J_L = J$ and one can use the symmetric form

$$J = \frac{ie}{2\hbar}\int \frac{d\epsilon}{2\pi}\mathrm{Tr}\Big\{\Big[\mathbf{\Gamma}_L(\epsilon - e\varphi_L) - \mathbf{\Gamma}_R(\epsilon - e\varphi_R)\Big]\mathbf{G}^<(\epsilon) + \\ + \Big[\mathbf{\Gamma}_L(\epsilon - e\varphi_L)f_L^0(\epsilon - e\varphi_L) - \mathbf{\Gamma}_R(\epsilon - e\varphi_R)f_R^0(\epsilon - e\varphi_R)\Big]\Big[\mathbf{G}^R(\epsilon) - \mathbf{G}^A(\epsilon)\Big]\Big\}. \tag{6.310}$$

For the proportional coupling $\mathbf{\Gamma}_L(\epsilon) = \lambda\mathbf{\Gamma}_R(\epsilon)$ in *linear response* (φ_i dependence of Γ_i is ignored!)

$$J = \frac{2e}{\hbar}\int \frac{d\epsilon}{4\pi}\left[f_L^0(\epsilon - e\varphi_L) - f_R^0(\epsilon - e\varphi_R)\right]\mathrm{Tr}\left(\frac{\mathbf{\Gamma}_L(\epsilon)\mathbf{\Gamma}_R(\epsilon)}{\mathbf{\Gamma}_L(\epsilon) + \mathbf{\Gamma}_R(\epsilon)}\mathbf{A}(\epsilon)\right). \tag{6.311}$$

$\mathbf{A} = i(\mathbf{G}^R - \mathbf{G}^A)$ is the spectral function. This expression is valid for *nonlinear response* if the energy dependence of $\mathbf{\Gamma}$ can be neglected (wide band limit).

(iii) Noninteracting case Finally, in the noninteracting case it is possible to obtain the usual Landauer-Büttikier formula with the transmission function

$$T(\epsilon) = \mathrm{Tr}\left[\mathbf{\Gamma}_L(\epsilon - e\varphi_L)\mathbf{G}^R(\epsilon)\mathbf{\Gamma}_R(\epsilon - e\varphi_R)\mathbf{G}^A(\epsilon)\right]. \tag{6.312}$$

This expression is equivalent to the one derived earlier by the single-particle Green function method.

We should stress once more that this formula is valid for finite voltage. Therefore, the voltage dependence of the level-width functions is important.

6.3.4 Nonequilibrium equation of motion method

Now we start to consider the case of interacting nanosystems. Although the MWJ current formula is exact, the problem to find the Green functions of the central region is sometimes highly nontrivial. At the present time there are several techniques developed to solve this problem.

Nonequilibrium equation of motion (NEOM) method is the simplest approximate approach. In spite of its simplicity, it is very useful in many cases, and is very convenient for numerical implementation. In this section we consider only a general formulation, some particular examples are considered further.

We start from the general definition of a Green function as the average of two Heisenberg operators $\hat{A}(t)$ and $\hat{B}(t)$, denoted as

$$\left\langle\!\left\langle \hat{A}(t_1), \hat{B}(t_2)\right\rangle\!\right\rangle^{R,A,<}.$$

The particular definitions of the averages for spectral and kinetic functions are

$$\left\langle\!\left\langle \hat{A}(t_1), \hat{B}(t_2)\right\rangle\!\right\rangle^{R} = -i\theta(t_1 - t_2)\left\langle\left[\hat{A}(t_1), \hat{B}(t_2)\right]_{\mp}\right\rangle, \tag{6.313}$$

where upper sing here and below is for boson functions, lower sing for fermions,

$$\left\langle\!\left\langle \hat{A}(t_1), \hat{B}(t_2) \right\rangle\!\right\rangle^{<} = -i\Big\langle \hat{A}(t_1), \hat{B}(t_2) \Big\rangle. \tag{6.314}$$

The equations of motion for NGF are obtained from the Heisenberg equation of motion for operators

$$i\frac{\partial \hat{A}}{\partial t} = \Big[\hat{A}, \hat{H}\Big]_- = \hat{A}\hat{H} - \hat{H}\hat{A}, \tag{6.315}$$

for any Heisenberg operator $\hat{A}(t)$. Here and below all Hamiltonians are *time-independent*. We consider the *stationary problem*.

(iv) Spectral (retarded and advanced) functions Let us start from a retarded function

$$\left\langle\!\left\langle \hat{A}(t_1), \hat{B}(t_2) \right\rangle\!\right\rangle^{R} = -i\theta(t_1 - t_2)\Big\langle \Big[\hat{A}(t_1), \hat{B}(t_2)\Big]_{\mp} \Big\rangle. \tag{6.316}$$

Taking the time derivative we obtain

$$i\frac{\partial}{\partial t_1}\left\langle\!\left\langle \hat{A}(t_1), \hat{B}(t_2) \right\rangle\!\right\rangle^{R} = \delta(t_1 - t_2)\Big\langle \Big[\hat{A}(t_1), \hat{B}(t_1)\Big]_{\mp} \Big\rangle + \left\langle\!\left\langle \Big[\hat{A}(t_1), \hat{H}\Big]_-, \hat{B}(t_2) \right\rangle\!\right\rangle^{R}, \tag{6.317}$$

where the first term originates from the time-derivative of the θ-function, and the equation (6.315) is used in the second term.

In the stationary case the Fourier transform can be used

$$(\epsilon + i\eta)\left\langle\!\left\langle \hat{A}, \hat{B} \right\rangle\!\right\rangle_{\epsilon}^{R} = \Big\langle \Big[\hat{A}, \hat{B}\Big]_{\mp} \Big\rangle + \left\langle\!\left\langle \Big[\hat{A}, \hat{H}\Big]_-, \hat{B} \right\rangle\!\right\rangle_{\epsilon}^{R}. \tag{6.318}$$

Now let us assume that the Hamiltonian can be divided into "free particle" and "interaction" parts $\hat{H} = \hat{H}_0 + \hat{H}_1$, and $[\hat{A}, \hat{H}_0]_- = \hat{\epsilon}_0\hat{A}$. (The simple example. For the free particle Hamiltonian $\hat{H}_0 = \sum_\beta \epsilon_\beta d_\beta^\dagger d_\beta$ and the operator $\hat{A} = d_\alpha^\dagger$ one has $[\hat{A}, \hat{H}_0]_- = \sum_\beta \epsilon_\beta [d_\alpha^\dagger, d_\beta^\dagger d_\beta]_- = \epsilon_\alpha d_\alpha^\dagger$, $\hat{\epsilon}_0 = \epsilon_\alpha$ is simply a number. In general, $\hat{\epsilon}_0$ is some time-independent operator). So that

$$(\epsilon + i\eta - \hat{\epsilon}_0)\left\langle\!\left\langle \hat{A}, \hat{B} \right\rangle\!\right\rangle_{\epsilon}^{R} = \Big\langle \Big[\hat{A}, \hat{B}\Big]_{\mp} \Big\rangle + \left\langle\!\left\langle \Big[\hat{A}, \hat{H}_1\Big]_-, \hat{B} \right\rangle\!\right\rangle_{\epsilon}^{R}, \tag{6.319}$$

the second term includes interaction and can not be easy simplified.

It is convenient now to introduce the "free particle" function $\hat{g}_\epsilon^R$ as a solution of the equation

$$(\epsilon + i\eta - \hat{\epsilon}_0)\hat{g}_\epsilon^R = 1. \tag{6.320}$$

Now we multiply the right and left parts of (6.319) by $\hat{g}_\epsilon^R$. Using the function $\hat{g}^R(t) = \int \hat{g}_\epsilon^R e^{-i\epsilon t}\frac{d\epsilon}{2\pi}$

we can write the time-dependent solution of (6.317) as

$$\left\langle\!\left\langle \hat{A}(t_1), \hat{B}(t_2) \right\rangle\!\right\rangle^R = \hat{g}^R(t_1 - t_2)\left\langle \left[\hat{A}(t_1), \hat{B}(t_1)\right]_{\mp} \right\rangle + \int \hat{g}^R(t_1 - t')\left\langle\!\left\langle \left[\hat{A}(t'), \hat{H}_1\right]_{-}, \hat{B}(t_2) \right\rangle\!\right\rangle^R dt'. \tag{6.321}$$

(v) EOM on the Schwinger-Keldysh contour The calculation of the lesser functions by the EOM technique requires some care. To demonstrate it let us compare the EOM for retarded and lesser functions of free particles.

The equation for $g^R_{\alpha\beta}$ is (assuming the diagonal matrix $\tilde{\epsilon}_{\alpha\beta}$)

$$(\epsilon + i\eta - \tilde{\epsilon}_\alpha)\, g^R_{\alpha\beta} = \delta_{\alpha\beta}, \tag{6.322}$$

from which the free-particle Green function is easily obtained.

At the same time for the lesser function we have the equation

$$(\epsilon - \tilde{\epsilon}_\alpha)\, g^<_{\alpha\beta} = 0, \tag{6.323}$$

from which, however, the free-particle lesser function $g^<_{\alpha\beta} = 2\pi f_0(\epsilon)\delta(\epsilon - \epsilon_\alpha)\delta_{\alpha\beta}$ can not be obtained.

The problem can be generally resolved by using the EOM on the Schwinger-Keldysh time contour. Contour-ordered Green function is defined as

$$\left\langle\!\left\langle \hat{A}(\tau_1), \hat{B}(\tau_2) \right\rangle\!\right\rangle^C = -i\left\langle T_c\Big(\hat{A}(\tau_1), \hat{B}(\tau_2)\Big) \right\rangle, \tag{6.324}$$

where $\hat{A}(\tau_1)$ and $\hat{B}(\tau_2)$ are two Heisenberg operators, defined along the contour.

Taking the time derivative we obtain the equation

$$i\frac{\partial}{\partial\tau_1}\left\langle\!\left\langle \hat{A}(\tau_1), \hat{B}(\tau_2) \right\rangle\!\right\rangle^C = \delta^c(\tau_1 - \tau_2)\left\langle \left[\hat{A}(\tau_1), \hat{B}(\tau_1)\right]_{\mp} \right\rangle + \left\langle\!\left\langle \left[\hat{A}(\tau_1), \hat{H}\right]_{-}, \hat{B}(\tau_2) \right\rangle\!\right\rangle^C, \tag{6.325}$$

in the stationary case this equation can be formally solved if one applies the Fourier transform along the contour, or perturbation expansion in the interaction representation (Niu et al. 1999). Using the free particle solution $\hat{g}^C(\tau_1 - \tau_2)$ we can write the time-dependent solution as

$$\left\langle\!\left\langle \hat{A}(\tau_1), \hat{B}(\tau_2) \right\rangle\!\right\rangle^C = \hat{g}^C(\tau_1 - \tau_2)\left\langle \left[\hat{A}(\tau_1), \hat{B}(\tau_1)\right]_{\mp} \right\rangle + \int \hat{g}^C(\tau_1 - \tau')\left\langle\!\left\langle \left[\hat{A}(\tau'), \hat{H}_1\right]_{-}, \hat{B}(\tau_2) \right\rangle\!\right\rangle^C d\tau'. \tag{6.326}$$

(vi) Kinetic (lesser) function Applying now the Langreth theorem, which shows, that from

$$C(\tau_1, \tau_2) = \int_C A(\tau_1, \tau_3)B(\tau_3, \tau_2)d\tau_3 \tag{6.327}$$

it follows

$$C^R(t_1,t_2) = \int A^R(t_1,t_3)B^R(t_3,t_2)dt_3, \tag{6.328}$$

$$C^<(t_1,t_2) = \int \left(A^R(t_1,t_3)B^R(t_3,t_2) + A^<(t_1,t_3)B^A(t_3,t_2)\right) dt_3, \tag{6.329}$$

we get (6.321) for the retarded function, and

$$\left\langle\!\left\langle \hat{A}(t_1), \hat{B}(t_2) \right\rangle\!\right\rangle^< = \hat{g}^<(t_1-t_2)\left\langle \left[\hat{A}(t_1), \hat{B}(t_1)\right]_\mp \right\rangle + \int \hat{g}^R(t_1-t')\left\langle\!\left\langle \left[\hat{A}(t'), \hat{H}_1\right]_-, \hat{B}(t_2) \right\rangle\!\right\rangle^< dt' + \int \hat{g}^<(t_1-t')\left\langle\!\left\langle \left[\hat{A}(t'), \hat{H}_1\right]_-, \hat{B}(t_2) \right\rangle\!\right\rangle^A dt' \tag{6.330}$$

for the lesser function. And the Fourier transform is

$$\left\langle\!\left\langle \hat{A}, \hat{B} \right\rangle\!\right\rangle_\epsilon^< = \hat{g}_\epsilon^<\left\langle \left[\hat{A}, \hat{B}\right]_\mp \right\rangle + \hat{g}_\epsilon^R\left\langle\!\left\langle \left[\hat{A}, \hat{H}_1\right]_-, \hat{B} \right\rangle\!\right\rangle_\epsilon^< + \hat{g}_\epsilon^<\left\langle\!\left\langle \left[\hat{A}, \hat{H}_1\right]_-, \hat{B} \right\rangle\!\right\rangle_\epsilon^A. \tag{6.331}$$

6.3.5 Kadanoff-Baym-Keldysh method

Now we review briefly the other approach. Kadanoff-Baym-Keldysh (KBK) method systematically extends the equilibrium many-body theory to the nonequilibrium case. Potentially, it is the most powerful approach. Below we give a simple introduction into the method, which is currently actively developed.

(i) Perturbation expansion and diagrammatic rules for contour functions We found that Green functions can be written in the interaction representation with a help of the $\hat{S}$-operator. For example, time-ordered fermionic Green function is

$$G^T_{\alpha\beta}(t_1,t_2) = -i\left\langle T\left(c_\alpha(t_1)c^\dagger_\beta(t_2)\right)\right\rangle \tag{6.332}$$

$$= -i\left\langle \hat{S}^{-1}T\left(\tilde{c}_\alpha(t_1)\tilde{c}^\dagger_\beta(t_2)\hat{S}\right)\right\rangle, \tag{6.333}$$

using "usual" $\hat{S}$-operator

$$\hat{S} = \hat{S}(\infty,t_0) = T\exp\left(-i\int_{t_0}^{\infty}\hat{V}^I(t')dt'\right), \tag{6.334}$$

or

$$G^T_{\alpha\beta}(t_1,t_2) = -i\left\langle T_C\left(\tilde{c}_\alpha(t_1^{\rightarrow})\tilde{c}^\dagger_\beta(t_2^{\rightarrow})\hat{S}_C\right)\right\rangle, \tag{6.335}$$

using "contour" $\hat{S}_C$-operator

$$\hat{S}_C = T_C\exp\left(-i\int_C\hat{V}^I(t')dt'\right). \tag{6.336}$$

We first consider the zero temperature case, when one can set $t_0 = -\infty$,

$$\hat{S} = \hat{S}(\infty,-\infty) = T\exp\left(-i\int_{-\infty}^{\infty}\hat{V}^I(t')dt'\right), \tag{6.337}$$

and assume that interaction is switched on and switched off at $t \to +\infty$ *adiabatically*. This condition

is necessary to prevent excitation of the system from its ground state. The other necessary condition is that the perturbation is time-independent in the Schrödinger representation. In this case if the initial state $|\Psi(t=-\infty)\rangle = |\Psi_0\rangle$ is the ground state (of free particles), then the final state $|\Psi(t=+\infty)\rangle = \hat{S}|\Psi^0\rangle = e^{i\theta}|\Psi^0\rangle$ is also the ground state, only the phase can be changed. Now, using the average value of the $\hat{S}$-operator

$$\langle\hat{S}\rangle = \langle\Psi^0|\hat{S}|\Psi^0\rangle = e^{i\theta}\langle\Psi^0|\Psi^0\rangle = e^{i\theta}, \tag{6.338}$$

we obtain

$$\hat{S}|\Psi^0\rangle = \langle\hat{S}\rangle|\Psi^0\rangle, \tag{6.339}$$

and

$$\langle\Psi^0|\hat{S}^{-1} = \frac{\langle\Psi^0|}{\langle\hat{S}\rangle}. \tag{6.340}$$

So that (6.332) can be written as

$$G^T_{\alpha\beta}(t_1,t_2) = -i\frac{\left\langle T\left(\tilde{c}_\alpha(t_1)\tilde{c}^\dagger_\beta(t_2)\hat{S}\right)\right\rangle}{\langle\hat{S}\rangle}. \tag{6.341}$$

Now we can expand the exponent (note that S-operator is defined only in the sense of this expansion)

$$\hat{S} = T\exp\left(-i\int_{-\infty}^{\infty}\hat{V}^I(t')dt'\right) = T\sum_{n=0}^{\infty}\frac{(-i)^n}{n!}\int_{-\infty}^{\infty}dt'_1...\int_{-\infty}^{\infty}dt'_n\ \hat{V}^I(t'_1)...\hat{V}^I(t'_n), \tag{6.342}$$

and numerator and denominator of the expression (6.341) are

$$\left\langle T\left(\tilde{c}_\alpha(t_1)\tilde{c}^\dagger_\beta(t_2)\hat{S}\right)\right\rangle = \sum_{n=0}^{\infty}\frac{(-i)^n}{n!}\int_{-\infty}^{\infty}dt'_1...\int_{-\infty}^{\infty}dt'_n\left\langle T\tilde{c}_\alpha(t_1)\tilde{c}^\dagger_\beta(t_2)\hat{V}^I(t'_1)...\hat{V}^I(t'_n)\right\rangle, \tag{6.343}$$

$$\langle\hat{S}\rangle = \sum_{n=0}^{\infty}\frac{(-i)^n}{n!}\int_{-\infty}^{\infty}dt'_1...\int_{-\infty}^{\infty}dt'_n\left\langle T\hat{V}^I(t'_1)...\hat{V}^I(t'_n)\right\rangle. \tag{6.344}$$

These expressions are used to produce the perturbation series.

The main quantity to be calculated is the contour Green function

$$G(1,2) \equiv G^C_{\alpha\beta}(\tau_1,\tau_2) = -i\left\langle T_C\left(c_\alpha(\tau_1)c^\dagger_\beta(\tau_2)\right)\right\rangle, \tag{6.345}$$

where τ_1 and τ_2 are contour times. Here $1_c \equiv \alpha, \tau_1$.

The general diagrammatic rules for contour Green functions are exactly the same as in the usual zero-temperature technique (we call it standard rules). The correspondence between diagrams and analytical expressions is established in the following way.

1. Open bare electron line is $iG_0(1,2)$.

2. Closed bare electron line is $n_0(1) \equiv n^{(0)}_\alpha(\tau_1)$.

3. Bare interaction line is $-iv(1,2)$.

4. Self-energy is $-i\Sigma(1,2)$.

5. Integration over internal vertices, and other standard rules.

(ii) Langreth rules Although the basic equations and diagrammatic rules are formulated for contour Green functions, the solution of these equation and final results are much more transparent when represented by real-time spectral and kinetic functions.

As in the ordinary diagrammatic technique, the important role is played by the integration (summation) over space and contour-time arguments of Green functions, which is denoted as

$$\int d1_c \equiv \sum_\alpha \int_C d\tau_1. \tag{6.346}$$

After application of the Langreth rules, for real-time functions these integrals become

$$\int d1 \equiv \sum_\alpha \int_{-\infty}^{\infty} dt_1. \tag{6.347}$$

The Langreth rules show, for example, that from

$$C(\tau_1,\tau_2) = \int_C A(\tau_1,\tau_3)B(\tau_3,\tau_2)d\tau_3 \tag{6.348}$$

it follows

$$C^R(t_1,t_2) = \int A^R(t_1,t_3)B^R(t_3,t_2)dt_3, \tag{6.349}$$

$$C^<(t_1,t_2) = \int \big(A^R(t_1,t_3)B^<(t_3,t_2) + A^<(t_1,t_3)B^A(t_3,t_2)\big)dt_3. \tag{6.350}$$

The other important rules are: from

$$C(\tau_1,\tau_2) = A(\tau_1,\tau_2)B(\tau_1,\tau_2) \tag{6.351}$$

it follows

$$C^R(t_1,t_2) = A^R(t_1,t_2)B^R(t_1,t_2) + A^R(t_1,t_2)B^<(t_1,t_2) + A^<(t_1,t_2)B^R(t_1,t_2), \tag{6.352}$$

$$C^<(t_1,t_2) = A^<(t_1,t_2)B^<(t_1,t_2), \tag{6.353}$$

and from

$$C(\tau_1,\tau_2) = A(\tau_1,\tau_2)B(\tau_2,\tau_1) \tag{6.354}$$

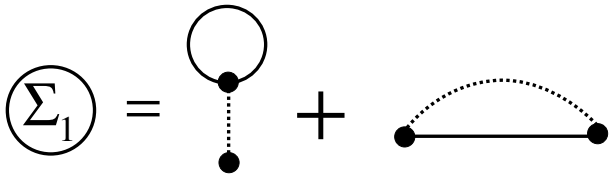


Figure 6.15: Diagrammatic representation of the first-order self-energy.

it follows

$$C^R(t_1,t_2) = A^R(t_1,t_2)B^<(t_2,t_1) + A^<(t_1,t_2)B^A(t_2,t_1), \tag{6.355}$$

$$C^<(t_1,t_2) = A^<(t_1,t_2)B^>(t_2,t_1). \tag{6.356}$$

(ii) First-order self-energy and polarization operator Consider, as an example, the first order expression for the self-energy, shown in Fig. 6.15. Following the diagrammatic rules, we find

$$\Sigma_1(1,2) = \delta(1-2)\int v(1,3)n_0(3)d3 + iv(1,2)G_0(1,2), \tag{6.357}$$

where the first term is the Hartree contribution, which can be included into the unperturbed Green function $G_0(1,2)$. This expression is actually symbolic, and translation from contour (Keldysh-time) to real-time functions is necessary. Using the Langreth rules, one obtains

$$\begin{aligned}\Sigma_1^R(1,2) &= \delta(1^+-2)\int v^R(1,3)n_0(3,3)d3 + iv^R(1,2)G_0^R(1,2)\\ &+ iv^<(1,2)G_0^R(1,2) + iv^R(1,2)G_0^<(1,2),\end{aligned} \tag{6.358}$$

$$\Sigma_1^<(1,2) = iv^<(1,2)G_0^<(1,2). \tag{6.359}$$

There is no Hartree term for lesser function, because the times τ_1 and τ_2 are always at the different branches of the Keldysh contour, and the δ-function $\delta(\tau_1-\tau_2)$ is zero.

In the stationary case and using explicit matrix indices, we have, finally ($\tau = t_1 - t_2$!, not to mix with the Keldysh time)

$$\begin{aligned}\Sigma_{\alpha\beta}^{R(1)}(\tau) = \delta(\tau^+)\delta_{\alpha\beta}\textstyle\sum_\gamma \tilde{v}_{\alpha\gamma}^R(0)n_\gamma^{(0)} + iv_{\alpha\beta}^R(\tau)G_{\alpha\beta}^{R(0)}\tau)\\ +iv_{\alpha\beta}^<(\tau)G_{\alpha\beta}^{R(0)}(\tau) + iv_{\alpha\beta}^R(\tau)G_{\alpha\beta}^{<(0)}(\tau),\end{aligned} \tag{6.360}$$

$$\Sigma_{\alpha\beta}^{<(1)}(\tau) = iv_{\alpha\beta}^<(\tau)G_{\alpha\beta}^{<(0)}(\tau), \tag{6.361}$$

and we define the Fourier transform of the bare interaction

$$\tilde{v}_{\alpha\gamma}^R(0) = \int v_{\alpha\gamma}^R(\tau)d\tau. \tag{6.362}$$

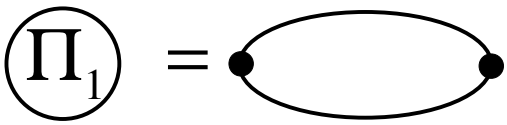


Figure 6.16: Diagrammatic representation of the first-order polarization operator.

Finally, the Fourier transforms are

$$\Sigma^{R(1)}_{\alpha\beta}(\epsilon) = \delta_{\alpha\beta}\sum_{\gamma}\tilde{v}^{R}_{\alpha\gamma}(0)n^{(0)}_{\gamma} + i\int\frac{d\epsilon'}{2\pi}\left[v^{R}_{\alpha\beta}(\epsilon')G^{R(0)}_{\alpha\beta}(\epsilon-\epsilon') + v^{<}_{\alpha\beta}(\epsilon')G^{R(0)}_{\alpha\beta}(\epsilon-\epsilon') + v^{R}_{\alpha\beta}(\epsilon')G^{<(0)}_{\alpha\beta}(\epsilon-\epsilon')\right], \tag{6.363}$$

$$\Sigma^{<(1)}_{\alpha\beta}(\epsilon) = i\int\frac{d\epsilon'}{2\pi}v^{<}_{\alpha\beta}(\epsilon')G^{<(0)}_{\alpha\beta}(\epsilon-\epsilon'). \tag{6.364}$$

The second important function is the polarization operator ("self-energy for interaction"), showing in Fig. 6.16. Following the diagrammatic rules, we find

$$\Pi_1(1,2) = -iG_0(1,2)G_0(2,1), \tag{6.365}$$

note the order of times in this expression.

Using the Langreth rules,

$$\Pi^{R}_1(1,2) = iG^{R}_0(1,2)G^{<}_0(2,1) + iG^{<}_0(1,2)G^{A}_0(2,1), \tag{6.366}$$

$$\Pi^{<}_1(1,2) = iG^{<}_0(1,2)G^{>}_0(2,1). \tag{6.367}$$

And in the stationary case, restoring the matrix indices

$$\Pi^{R(1)}_{\alpha\beta}(\epsilon) = -i\left[G^{R(0)}_{\alpha\beta}(\tau)G^{<(0)}_{\beta\alpha}(-\tau) + G^{<(0)}_{\alpha\beta}(\tau)G^{A(0)}_{\beta\alpha}(-\tau)\right], \tag{6.368}$$

$$\Pi^{<(1)}_{\alpha\beta}(\epsilon) = -iG^{<(0)}_{\alpha\beta}(\tau)G^{>(0)}_{\beta\alpha}(-\tau). \tag{6.369}$$

In the Fourier representation

$$\Pi^{R(1)}_{\alpha\beta}(\tau) = -i\int\frac{d\epsilon'}{2\pi}\left[G^{R(0)}_{\alpha\beta}(\epsilon')G^{<(0)}_{\beta\alpha}(\epsilon'-\epsilon) + G^{<(0)}_{\alpha\beta}(\epsilon')G^{A(0)}_{\beta\alpha}(\epsilon'-\epsilon)\right], \tag{6.370}$$

$$\Pi^{<(1)}_{\alpha\beta}(\tau) = -i\int\frac{d\epsilon'}{2\pi}G^{<(0)}_{\alpha\beta}(\epsilon')G^{>(0)}_{\beta\alpha}(\epsilon'-\epsilon). \tag{6.371}$$

These expressions are quite general and can be used for both electron-electron and electron-vibron interaction.

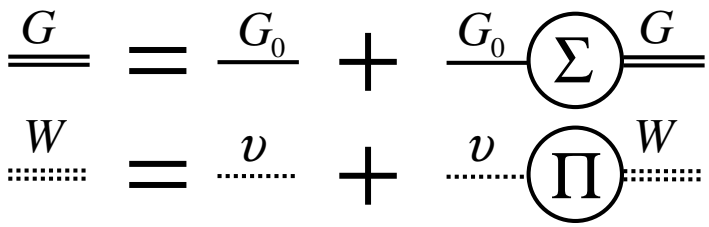


Figure 6.17: Diagrammatic representation of the Dyson equations.

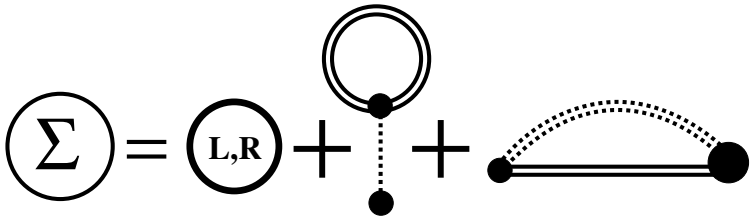


Figure 6.18: Diagrammatic representation of the full self-energy.

For Coulomb interaction the bare interaction is is $v(1,2) \equiv U_{\alpha\beta}\delta(\tau_1^+ - \tau_2)$, so that

$$v^R(1,2) \equiv U_{\alpha\beta}\delta(t_1^+ - t_2), \tag{6.372}$$

$$v^<(1,2) = 0. \tag{6.373}$$

(iv) Self-consistent equations The diagrams can be partially summed in all orders of perturbation theory. The resulting equations are known as Dyson equations for the dressed Green function $G(1,2)$ and the effective interaction $W(1,2)$ (Fig. 6.17). Analytically these equations are written as

$$G(1,2) = G_0(1,2) + \iint G_0(1,3)\Sigma(3,4)G(4,2)d3d4, \tag{6.374}$$

$$W(1,2) = v(1,2) + \iint v(1,3)\Pi(3,4)W(4,2)d3d4. \tag{6.375}$$

In the perturbative approach the first order (or higher order) expressions for the self-energy and the polarization operator are used. The other possibility is to summarize further the diagrams and obtain the self-consistent approximations (Figs. 6.18,6.19), which include, however, a new unknown function,

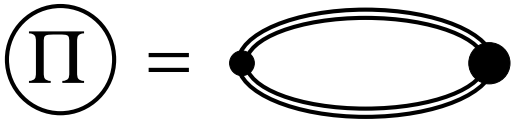


Figure 6.19: Diagrammatic representation of the full polarization operator.

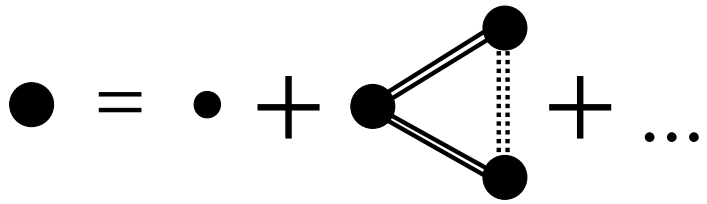


Figure 6.20: Diagrammatic representation of the vertex function.

called vertex function. We shall write these expressions analytically, including the Hartree-Fock part into unperturbed Green function $G_0(1,2)$.

$$\Sigma'(1,2) = i \iint W(1,3)G(1,4)\Gamma(3;4,2)d3d4, \tag{6.376}$$

$$\Pi(1,2) = -i \iint G(1,3)G(4,1)\Gamma(2;3,4)d3d4. \tag{6.377}$$

The equation for the vertex function can not be closed diagrammatically (Fig. 6.20). Nevertheless, it is possible to write close set of equations (*Hedin's equations*), which are exact equations for full Green functions written through a functional derivative. Hedin's equations are equations (6.374)-(6.377) and the equation for the vertex function

$$\Gamma(1;2,3) = \delta(1,2)\delta(1,3) + \iiiint G(4,6)G(7,5)\Gamma(1;6,7)\frac{\delta\Sigma(2,3)}{\delta G(4,5)}d4d5d6d7. \tag{6.378}$$

6.4 Applications

6.4.1 Coulomb blockade and non-equilibrium Green functions

In Section II we have seen that Coulomb blockade phenomena mediated by electron-electron interactions on a quantum dot can be dealt with in a straightforward way by using master equation (ME) approaches, which are based on Fermi's Golden Rule. [64, 65, 66, 67, 108, 110, 109, 111] However, due to its intrinsic perturbative character in the lead-dot coupling, ME techniques cannot cover the whole interaction range from weak-coupling (Coulomb blockade), intermediate coupling (Kondo physics), up to strong coupling (Fabry-Perot physics). It is thus of methodological and practical interest to develop schemes which allow, in a systematic way, to describe the three mentioned regimes also in out-of-equilibrium situations. As stated in the introduction, we believe that Green function techniques are such a tool; in this section we will show how a non-equilibrium treatment of the Hubbard-Anderson model together with appropriate approximations allow us to reproduce the well-known Coulomb blockade stability diagrams obtained with the master equation approach (see also Section II). For the sake of simplicity we will deal with the problem of single and double-site dots in the CB regime, although the method can be straightforwardly extended to multi-level systems. Our purpose is to study the problem of a two site donor/acceptor molecule in the CB regime within the NEGF as a first step to deal with the phenomenology of a rigid multilevel island. The nuclear dynamics (vibrations) always present in molecular junctions could be then modularly included in this theory. Our method can be calibrated on the well-studied double quantum dot problem [109, 112] and could be possibly integrated in the density functional theory based approaches to molecular conductance. The Kondo regime would require a separate treatment involving more complex decoupling schemes and will be thus left out of this review.

The *linear conductance* properties of a single site junction (SSJ) with Coulomb interactions (Anderson impurity model), have been extensively studied by means of the EOM approach in the cases related to CB [113, 114] and the Kondo effect. [115] Later the same method was applied to some two-site models. [116, 117, 118] Multi-level systems were started to be considered only recently. [119, 120] For out-of-equilibrium situations (finite applied bias), there are some methodological unclarified issues for calculating correlation functions using EOM techniques. [121, 122, 123] We have developed an EOM-based method which allows to deal with the finite-bias case in a self-consistent way. [124]

6.4.1.1 Nonequilibrium EOM formalism

The Anderson-Hubbard Hamiltonian We consider the following model Hamiltonian (which can be called the multi-level Anderson impurity model, the Hubbard model, or the quantum cluster model)

$$\hat{H} = \sum_{\alpha\beta} \tilde{\epsilon}_{\alpha\beta} d_\alpha^\dagger d_\beta + \frac{1}{2} \sum_{\alpha\beta} U_{\alpha\beta} \hat{n}_\alpha \hat{n}_\beta + \sum_{ik\sigma} \tilde{\epsilon}_{ik\sigma} c_{ik\sigma}^\dagger c_{ik\sigma} + \sum_{ik\sigma,\alpha} \left(V_{ik\sigma,\alpha} c_{ik\sigma}^\dagger d_\alpha + h.c. \right), \tag{6.379}$$

electrical potentials are included into the energies $\tilde{\epsilon}_{ik\sigma} = \epsilon_{ik\sigma} + e\varphi_i(t)$ and $\tilde{\epsilon}_{\alpha\alpha} = \epsilon_{\alpha\alpha} + e\varphi_\alpha(t)$.

This model is quite universal, describing a variety of correlated electron systems coupled to the leads:

the Anderson impurity model, the multilevel quantum dot with diagonal noninteracting Hamiltonian $\tilde{\epsilon}_{\alpha\beta}$, a system (cluster) of several quantum dots, when the off-diagonal matrix elements of $\tilde{\epsilon}_{\alpha\beta}$ describe hopping between individual dots, and, finally, the 1D and 2D quantum point contacts.

EOM for Heisenberg operators Using the Hamiltonian (6.379) one derives

$$i\frac{\partial c_{ik\sigma}}{\partial t} = \left[c_{ik\sigma}, \hat{H}\right]_- = \tilde{\epsilon}_{ik\sigma}c_{ik\sigma} + \sum_{\alpha} V_{ik\sigma,\alpha}d_{\alpha}, \tag{6.380}$$

$$i\frac{\partial c^{\dagger}_{ik\sigma}}{\partial t} = -\tilde{\epsilon}_{ik\sigma}c^{\dagger}_{ik\sigma} - \sum_{\alpha} V^{*}_{ik\sigma,\alpha}d^{\dagger}_{\alpha}, \tag{6.381}$$

$$i\frac{\partial d_{\alpha}}{\partial t} = \sum_{\beta}\tilde{\epsilon}_{\alpha\beta}d_{\beta} + \sum_{\beta\neq\alpha} U_{\alpha\beta}\hat{n}_{\beta}d_{\alpha} + \sum_{ik\sigma} V^{*}_{ik\sigma,\alpha}c_{ik\sigma}, \tag{6.382}$$

$$i\frac{\partial d^{\dagger}_{\alpha}}{\partial t} = -\sum_{\beta}\tilde{\epsilon}_{\alpha\beta}d^{\dagger}_{\beta} - \sum_{\beta\neq\alpha} U_{\alpha\beta}\hat{n}_{\beta}d^{\dagger}_{\alpha} - \sum_{ik\sigma} V_{ik\sigma,\alpha}c^{\dagger}_{ik\sigma}, \tag{6.383}$$

$$i\frac{\partial \hat{n}_{\alpha}}{\partial t} = \sum_{ik\sigma}\left[-V_{ik\sigma,\alpha}c^{\dagger}_{ik\sigma}d_{\alpha} + V^{*}_{ik\sigma,\alpha}d^{\dagger}_{\alpha}c_{ik\sigma}\right]. \tag{6.384}$$

These equations look like a set of ordinary differential equations, but are, in fact, much more complex. The first reason is, that there are the equations for *operators*, and special algebra should be used to solve it. Secondly, the number of $c_{ik\sigma}$ operators is infinite! Because of that, the above equations are not all sufficient, but are widely used to obtain the equations for Green functions.

Spectral (retarded and advanced) functions Now we follow the general NEOM method described in the Section III. Using (6.382), we get the equation for $G^{R}_{\alpha\beta} = -i\left\langle\left[d_{\alpha}, d^{\dagger}_{\beta}\right]_{+}\right\rangle_{\epsilon}$

$$(\epsilon + i\eta)G^{R}_{\alpha\beta} - \sum_{\gamma}\tilde{\epsilon}_{\alpha\gamma}G^{R}_{\gamma\beta} = \delta_{\alpha\beta} + \sum_{\gamma\neq\alpha} U_{\alpha\gamma}G^{(2)R}_{\alpha\gamma,\beta} + + \sum_{ik\sigma} V^{*}_{ik\sigma,\alpha}G^{R}_{ik\sigma,\beta} \tag{6.385}$$

which includes two new functions: $G^{(2)R}_{\alpha\gamma,\beta}$ and $G^{R}_{ik\sigma,\beta}$.

The equation for $G^{R}_{ik\sigma,\beta}$ is closed (includes only the function $G^{R}_{\alpha\beta}$ introduced before)

$$(\epsilon + i\eta - \tilde{\epsilon}_{ik\sigma})G^{R}_{ik\sigma,\beta} = \sum_{\delta} V_{ik\sigma,\delta}G^{R}_{\delta\beta}. \tag{6.386}$$

The equation for

$$G^{(2)R}_{\alpha\gamma,\beta}(t_1 - t_2) = -i\theta(t_1 - t_2)\left\langle\left[d_{\alpha}(t_1)\hat{n}_{\gamma}(t_1), d^{\dagger}_{\beta}(t_2)\right]_{+}\right\rangle$$

is more complicated ($\gamma \neq \alpha$)

$$(\epsilon + i\eta)G^{(2)R}_{\alpha\gamma,\beta} - \sum_{\delta} \tilde{\epsilon}_{\alpha\delta} G^{(2)R}_{\delta\gamma,\beta} = \langle \hat{n}_\gamma \rangle \delta_{\alpha\beta} + i \int \frac{d\epsilon}{2\pi} G^{<}_{\alpha\beta} \delta_{\beta\gamma}$$
$$+ \sum_{\delta \neq \alpha} U_{\alpha\delta} \left\langle\!\!\left\langle \hat{n}_\delta d_\alpha \hat{n}_\gamma ; d^\dagger_\beta \right\rangle\!\!\right\rangle^R + \sum_{ik\sigma} V^*_{ik\sigma,\alpha} \left\langle\!\!\left\langle c_{ik\sigma} n_\gamma ; d^\dagger_\beta \right\rangle\!\!\right\rangle^R$$
$$+ \sum_{ik\sigma} V^*_{ik\sigma,\gamma} \left\langle\!\!\left\langle d_\alpha d^\dagger_\gamma c_{ik\sigma} ; d^\dagger_\beta \right\rangle\!\!\right\rangle^R - \sum_{ik\sigma} V_{ik\sigma,\gamma} \left\langle\!\!\left\langle d_\alpha c^\dagger_{ik\sigma} d_\gamma ; d^\dagger_\beta \right\rangle\!\!\right\rangle^R . \tag{6.387}$$

We also present here one of the next order equations, for the function $\left\langle\!\!\left\langle c_{ik\sigma} n_\gamma ; d^\dagger_\beta \right\rangle\!\!\right\rangle^R$

$$(\epsilon + i\eta - \tilde{\epsilon}_{ik\sigma}) \left\langle\!\!\left\langle c_{ik\sigma} n_\gamma ; d^\dagger_\beta \right\rangle\!\!\right\rangle^R = i \int \frac{d\epsilon}{2\pi} G^{<}_{ik\sigma,\beta} \delta_{\beta\gamma}$$
$$+ \sum_{\alpha} V_{ik\sigma,\alpha} G^{(2)R}_{\alpha\gamma,\beta} + \sum_{ik'\sigma'} V^*_{ik'\sigma',\gamma} \left\langle\!\!\left\langle c_{ik\sigma} d^\dagger_\gamma c_{ik'\sigma'} ; d^\dagger_\beta \right\rangle\!\!\right\rangle^R - \sum_{ik'\sigma'} V_{ik'\sigma',\gamma} \left\langle\!\!\left\langle c_{ik\sigma} c^\dagger_{ik'\sigma'} d_\gamma ; d^\dagger_\beta \right\rangle\!\!\right\rangle^R . \tag{6.388}$$

The equation (6.388) is not closed again and produces new Green functions of higher order. And so on. These sequence of equations can not be closed in the general case and should be truncated at some point. Below we consider some possible approximations. The other important point is, that average populations and lesser Green functions should be calculated self-consistently. In equilibrium (linear response) these functions are easy related to the spectral functions. But at finite voltage it should be calculated independently.

Kinetic (lesser) function Following the same way, as for the retarded functions (using only the definitions of NGF and Heisenberg equations of motion) one derives instead of (6.385)-(6.388)

$$\epsilon G^{<}_{\alpha\beta} - \sum_{\gamma} \tilde{\epsilon}_{\alpha\gamma} G^{<}_{\gamma\beta} = \sum_{\gamma \neq \alpha} U_{\alpha\gamma} G^{(2)<}_{\alpha\gamma,\beta} + \sum_{ik\sigma} V^*_{ik\sigma,\alpha} G^{<}_{ik\sigma,\beta}, \tag{6.389}$$

$$(\epsilon - \tilde{\epsilon}_{ik\sigma}) G^{<}_{ik\sigma,\beta} = \sum_{\delta} V_{ik\sigma,\delta} G^{<}_{\delta\beta}, \tag{6.390}$$

$$\epsilon G^{(2)<}_{\alpha\gamma,\beta} - \sum_{\delta} \tilde{\epsilon}_{\alpha\delta} G^{(2)<}_{\delta\gamma,\beta} = \sum_{\delta \neq \alpha} U_{\alpha\delta} \left\langle\!\!\left\langle \hat{n}_\delta d_\alpha \hat{n}_\gamma ; d^\dagger_\beta \right\rangle\!\!\right\rangle^<$$
$$+ \sum_{ik\sigma} V^*_{ik\sigma,\alpha} \left\langle\!\!\left\langle c_{ik\sigma} n_\gamma ; d^\dagger_\beta \right\rangle\!\!\right\rangle^< + \sum_{ik\sigma} V^*_{ik\sigma,\gamma} \left\langle\!\!\left\langle d_\alpha d^\dagger_\gamma c_{ik\sigma} ; d^\dagger_\beta \right\rangle\!\!\right\rangle^< - \sum_{ik\sigma} V_{ik\sigma,\gamma} \left\langle\!\!\left\langle d_\alpha c^\dagger_{ik\sigma} d_\gamma ; d^\dagger_\beta \right\rangle\!\!\right\rangle^< , \tag{6.391}$$

$$(\epsilon - \tilde{\epsilon}_{ik\sigma}) \left\langle\!\!\left\langle c_{ik\sigma} n_\gamma ; d^\dagger_\beta \right\rangle\!\!\right\rangle^< = \sum_{\alpha} V_{ik\sigma,\alpha} G^{(2)<}_{\alpha\gamma,\beta}$$
$$+ \sum_{ik'\sigma'} V^*_{ik'\sigma',\gamma} \left\langle\!\!\left\langle c_{ik\sigma} d^\dagger_\gamma c_{ik'\sigma'} ; d^\dagger_\beta \right\rangle\!\!\right\rangle^< - \sum_{ik'\sigma'} V_{ik'\sigma',\gamma} \left\langle\!\!\left\langle c_{ik\sigma} c^\dagger_{ik'\sigma'} d_\gamma ; d^\dagger_\beta \right\rangle\!\!\right\rangle^< . \tag{6.392}$$

To find $G^{<}_{ik\sigma,\beta}$ and $\left\langle\!\left\langle c_{ik\sigma}n_\gamma; d^\dagger_\beta \right\rangle\!\right\rangle^{<}$ we should divide the right parts by $(\epsilon - \tilde{\epsilon}_{ik\sigma})$, which is not well defined at $\epsilon = \tilde{\epsilon}_{ik\sigma}$. In the section III we considered the general prescription to avoid this problem, we use the equation (6.331), and instead of (6.390) and (6.392) we obtain

$$G^{<}_{ik\sigma,\beta} = g^R_{ik\sigma}\sum_\delta V_{ik\sigma,\delta}G^{<}_{\delta\beta} + g^{<}_{ik\sigma}\sum_\delta V_{ik\sigma,\delta}G^A_{\delta\beta}, \tag{6.393}$$

$$\begin{aligned}\left\langle\!\left\langle c_{ik\sigma}n_\gamma; d^\dagger_\beta \right\rangle\!\right\rangle^{<} &= g^R_{ik\sigma}\sum_\alpha V_{ik\sigma,\alpha}G^{(2)<}_{\alpha\gamma,\beta} + g^{<}_{ik\sigma}\sum_\alpha V_{ik\sigma,\alpha}G^{(2)A}_{\alpha\gamma,\beta} + \\ +g^R_{ik\sigma}\sum_{ik'\sigma'} V^*_{ik'\sigma',\gamma}\left\langle\!\left\langle c_{ik\sigma}d^\dagger_\gamma c_{ik'\sigma'}; d^\dagger_\beta \right\rangle\!\right\rangle^{<} &- g^R_{ik\sigma}\sum_{ik'\sigma'} V_{ik'\sigma',\gamma}\left\langle\!\left\langle c_{ik\sigma}c^\dagger_{ik'\sigma'}d_\gamma; d^\dagger_\beta \right\rangle\!\right\rangle^{<} \\ +g^{<}_{ik\sigma}\sum_{ik'\sigma'} V^*_{ik'\sigma',\gamma}\left\langle\!\left\langle c_{ik\sigma}d^\dagger_\gamma c_{ik'\sigma'}; d^\dagger_\beta \right\rangle\!\right\rangle^{A} &- g^{<}_{ik\sigma}\sum_{ik'\sigma'} V_{ik'\sigma',\gamma}\left\langle\!\left\langle c_{ik\sigma}c^\dagger_{ik'\sigma'}d_\gamma; d^\dagger_\beta \right\rangle\!\right\rangle^{A}.\end{aligned} \tag{6.394}$$

The equations (6.389) and (6.391) can be used without modifications because they include the imaginary parts (dissipation) from the lead terms.

At this point we stop the general consideration, and introduce a powerful *Ansatz* for the NEGF which is related both to the equation-of-motion (EOM) method and to the Dyson equation approach. [124] From the knowledge of the Green function (GF) we then calculate the transport observables. For clarity, we first describe our method in the more familiar problem of a single site junction, which is the well-known Anderson impurity model. Then we apply it to a double quantum dot. The equations obtained below by the heuristic mapping method can be obtained straightforward from the general NEOM equations derived in this section using the same approximations as in the mapping method.

6.4.1.2 Anderson impurity model (single site)

The Anderson impurity model is used to describe the Coulomb interaction on a single site:

$$H = H_{\rm D} + \sum_\alpha (H_\alpha + H_{\alpha{\rm D}}),$$

where

$$\begin{aligned}H_{\rm D} &= \sum_\sigma \epsilon_\sigma d^\dagger_\sigma d_\sigma + \frac{1}{2}U n_\sigma n_{\bar\sigma}, \\ H_\alpha &= \sum_{k,\sigma}\epsilon^\alpha_{k,\sigma}c^\dagger_{\alpha,k,\sigma}c_{\alpha,k,\sigma}, \\ H_{\alpha{\rm D}} &= \sum_{k,\sigma}\left(V_{\alpha,k,\sigma}c^\dagger_{\alpha,k,\sigma}d_\sigma + V^*_{\alpha,k,\sigma}d^\dagger_\sigma c_{\alpha,k,\sigma}\right),\end{aligned}$$

where d and c are the operators for electrons on the dot and on the left ($\alpha = {\rm L}$) and the right ($\alpha = {\rm R}$) lead, U is the Coulomb interaction parameter, ϵ_σ is the σ level of the quantum dot, while $\epsilon^\alpha_{k,\sigma}$ is the spin σ level of lead α in k space, $\sigma = \uparrow, \downarrow$. With the help of the EOM and the truncation approximation, we

can get a closed set of equations for the retarded and advanced GFs $G^{\mathrm{r/a}}_{\sigma,\tau}$ [114, 53]

$$(\omega - \epsilon_\sigma - \Sigma^{\mathrm{r/a}}_\sigma) G^{\mathrm{r/a}}_{\sigma,\tau} = \delta_{\sigma,\tau} + U G^{(2)\mathrm{r/a}}_{\sigma,\tau}, \tag{6.395a}$$

$$(\omega - \epsilon_\sigma - U - \Sigma^{\mathrm{r/a}}_\sigma) G^{(2)\mathrm{r/a}}_{\sigma,\tau} = \langle n_{\bar\sigma}\rangle \delta_{\sigma,\tau}, \tag{6.395b}$$

where $G^{\mathrm{r/a}}_{\sigma,\tau} = \langle\langle d_\sigma | d^\dagger_\tau\rangle\rangle^{\mathrm{r/a}}$, $G^{(2)\mathrm{r/a}}_{\sigma,\tau} = \langle\langle n_{\bar\sigma} d_\sigma | d^\dagger_\tau\rangle\rangle^{\mathrm{r/a}}$ and

$$\Sigma^{\mathrm{r/a}}_\sigma(\omega) = \Sigma^{\mathrm{r/a}}_{\mathrm{L},\sigma} + \Sigma^{\mathrm{r/a}}_{\mathrm{R},\sigma} = \sum_{\alpha,k} \frac{|V_{\alpha,k,\sigma}|^2}{\omega - \epsilon^\alpha_{k,\sigma} \pm \mathrm{i}0^+} \tag{6.396}$$

are the electron self-energies.

Mapping on retarded Green functions For retarded GFs, from the EOM method, and with the help of Eqs. (6.395a) and (6.395b), we can get

$$G^{\mathrm{r}} \;=\; G^{\mathrm{r}}_0 + G^{\mathrm{r}}_0 U G^{(2)\mathrm{r}} = G^{\mathrm{r}}_0 + G^{\mathrm{r}}_0 \Sigma^{\mathrm{EOM}} G^{(1)\mathrm{r}},$$

where G^{r} is single-particle GF matrix

$$G^{\mathrm{r}} = \begin{pmatrix} G^{\mathrm{r}}_{\uparrow,\uparrow} & G^{\mathrm{r}}_{\uparrow,\downarrow} \\ G^{\mathrm{r}}_{\downarrow,\uparrow} & G^{\mathrm{r}}_{\downarrow,\downarrow} \end{pmatrix},$$

and $G^{(1)\mathrm{r}}_{\sigma,\tau} = G^{(2)\mathrm{r}}_{\sigma,\tau}/\langle n_{\bar\sigma}\rangle$. G^{r}_0 describes the single-particle spectrum without Coulomb interaction, but including the effects from the electrodes. $\Sigma^{\mathrm{EOM}}_{\sigma,\tau} = U\langle n_{\bar\sigma}\rangle$ is the Hartree-like self-energy of our model. Since there is only Coulomb interaction on the site with the levels ϵ_σ, the Fock-like self-energy is vanishing.

Alternatively, by means of the Dyson equation and the second-order truncation approximation, taking Hartree-like self-energies $\Sigma^{\mathrm{H}}_{\sigma,\tau} = U\langle n_{\bar\sigma}\rangle$ $(= \Sigma^{\mathrm{EOM}}_{\sigma,\tau})$, we can also get the retarded GFs as follows

$$G^{\mathrm{r}} = G^{\mathrm{r}}_0 + G^{\mathrm{r}}_0 \Sigma^{\mathrm{H}} G^{\mathrm{r}}_1, \tag{6.397}$$

where $G^{\mathrm{r}}_1 = G^{\mathrm{r}}_0 + G^{\mathrm{r}}_0 \Sigma^{\mathrm{H}} G^{\mathrm{r}}_0$ is the first-order truncation GF.

Within the level of the second-order truncation approximation, we see that there is a map between the EOM results and the Dyson results:

$$G^{\mathrm{r}} \;=\; G^{\mathrm{r}}_0 \;+\; G^{\mathrm{r}}_0 \;\Sigma^{\mathrm{H}}\; G^{(1)\mathrm{r}} \qquad \text{(EOM)}, \tag{6.398a}$$

$$\updownarrow \qquad\qquad\qquad \updownarrow$$

$$G^{\mathrm{r}} \;=\; G^{\mathrm{r}}_0 \;+\; G^{\mathrm{r}}_0 \;\Sigma^{\mathrm{H}}\; G^{\mathrm{r}}_1 \qquad \text{(Dyson)}. \tag{6.398b}$$

Eqs. (6.398) prompts a way to include further many-particle effects into the Dyson equation, Eq. (6.398b), by replacing the *Dyson-first-order* retarded Green function G^{r}_1 with the EOM $G^{(1)r}$. Then one obtains already the correct results to describe CB while keeping the Hartree-like self-energy.

Mapping on contour and lesser Green functions Introducing now the contour GF $\check{G}$, we can get the Dyson equation as follows [50, 51, 52, 53]

$$\check{G} = \check{G}_0 + \check{G}_0\check{\Sigma}\check{G}, \tag{6.399}$$

where $\check{\Sigma}$ is the self-energy matrix. [53]

According to the approximation for the retarded GF in Eq. (6.397), we take the second-order truncation on Eq. (6.399), and then get

$$\check{G} = \check{G}_0 + \check{G}_0\check{\Sigma}^{\mathrm{H}}\check{G}_1,$$

where $\check{G}_1 = \check{G}_0 + \check{G}_0\check{\Sigma}^{\mathrm{H}}\check{G}_0$ is the first-order contour GF, and $\check{G}_0$ has already included the lead broadening effects.

Similar to the mapping in Eq. (6.398), we perform an *Ansatz* consisting in substituting the *Dyson-first-order* $G_1^{\mathrm{r/a/<}}$ with the EOM one $G^{(1)\mathrm{r/a/<}}$ to consider more many-particle correlations, while the EOM self-energy is used for the *Dyson* equation for consistency:

$$\begin{array}{ccccccccc} \check{G} & = & \check{G}_0 & + & \check{G}_0 & \check{\Sigma}^{\mathrm{H}} & \check{G}_1 & & \text{(Dyson)}, \\ \updownarrow & & & & & & \uparrow & & \\ \check{G} & & & & & & \check{G}^{(1)} & & \text{(EOM)}. \end{array} \tag{6.400}$$

Then, using the Langreth theorem [53] we get the lesser GF,

$$\begin{aligned} G^{<} &= G_0^{<} + G_0^{\mathrm{r}}\Sigma^{\mathrm{H,r}}G^{(1)<} + G_0^{<}\Sigma^{\mathrm{H,a}}G^{(1)\mathrm{a}} \\ &= G_0^{<} + G_0^{\mathrm{r}}UG^{(2)<} + G_0^{<}UG^{(2)\mathrm{a}} \end{aligned} \tag{6.401}$$

where $G_0^{\mathrm{r/a/<}}$ are GFs for $U = 0$, *but* including the lead broadening effects, *i.e.*

$$\begin{aligned} G_0^{<} &= g_0^{<} + g_0^{\mathrm{r}}\Sigma^{<}G_0^{\mathrm{a}} + g_0^{<}\Sigma^{\mathrm{a}}G_0^{\mathrm{a}} + g_0^{\mathrm{r}}\Sigma^{\mathrm{r}}G_0^{<}, \\ G_0^{\mathrm{r/a}} &= g_0^{\mathrm{r/a}} + g_0^{\mathrm{r/a}}\Sigma^{\mathrm{r/a}}G_0^{\mathrm{r/a}}, \end{aligned}$$

with $g_0^{\mathrm{r/a/<}}$ the free electron GF, and

$$\Sigma^{\mathrm{r/a/<}} = \begin{pmatrix} \Sigma_{\uparrow}^{\mathrm{r/a/<}} & 0 \\ 0 & \Sigma_{\downarrow}^{\mathrm{r/a/<}} \end{pmatrix},$$

$\Sigma_\sigma^{<} = \mathrm{i}\sum_\alpha \Gamma_\alpha f_\alpha(\omega)$, and $\Gamma_\alpha = \mathrm{i}(\Sigma_\alpha^{\mathrm{r}} - \Sigma_\alpha^{\mathrm{a}})$, $f_\alpha(\omega) = f(\omega - \mu_\alpha)$, f is the equilibrium Fermi function and μ_α is the electro-chemical potential in lead α; $\Sigma_\alpha^{\mathrm{r/a}}$ are the retarded/advanced electron self-energies from Eq. (6.396) and $G_{\sigma,\tau}^{(1)\mathrm{r/a/<}} = G_{\sigma,\tau}^{(2)\mathrm{r/a/<}}/\langle n_{\bar{\sigma}}\rangle$. Performing the same *Ansatz* on the double-particle GF,

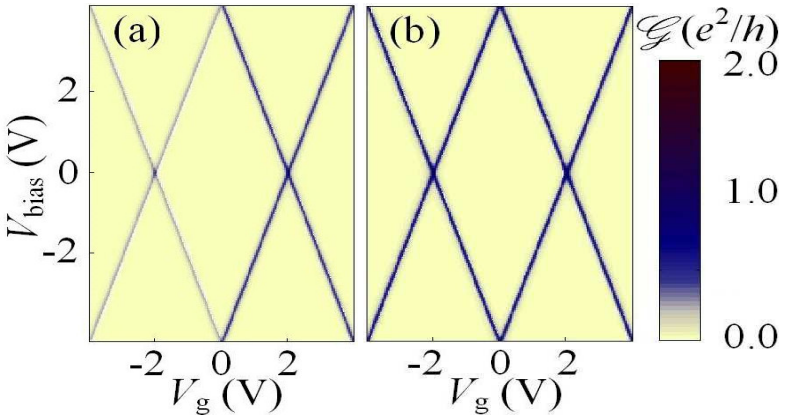


Figure 6.21: (Color online) The stability diagram of a SSJ with $\epsilon_\sigma = 2.0$ eV, $U = 4.0$ eV, $\Gamma_{\rm L} = \Gamma_{\rm R} = 0.05$ eV. (a) The uncorrect result obtained by means of the widely used formula in Eq. (6.403) for the lesser GF is not symmetric for levels ϵ_σ and $\epsilon_\sigma + U$. (b) Results obtained by means of our *Ansatz* in Eq. (6.401) shows correctly symmetric for levels ϵ_σ and $\epsilon_\sigma + U$.

from Eq. (6.395b) we can get

$$G^{(2)<} = G^{(2)\rm r}\Sigma^{(2)<}G^{(2)\rm a}, \tag{6.402}$$

with $\Sigma_\sigma^{(2)<} = \Sigma_\sigma^< / \langle n_{\bar\sigma}\rangle$.

The lesser GFs in Eq. (6.401) can also be obtained directly from the general formula [53]

$$G^< = G_0^< + G_0^{\rm r}\Sigma^{\rm r}G^< + G_0^{\rm r}\Sigma^<G^{\rm a} + G_0^<\Sigma^{\rm a}G^{\rm a},$$

with the help of the *Ansatz* in Eq. (6.400). It should be noted that Eq. (6.401) is very different from the lesser GF formula,

$$G^< = G^{\rm r}\Sigma^<G^{\rm a}, \tag{6.403}$$

with the self-energy $\Sigma^<$ containing *only* contributions from the electrodes. The equation (6.403) is widely used for both first-principle [35, 45, 41] and model Hamiltonian calculations. [117]

The numerical calculation results of conductance dependence on the bias and gate voltages by the two different NEGF Eqs. (6.401) and (6.403) are shown in Fig. 6.21. As we can see in the left panel, the adoption of Eq. (6.403) results in an incorrectly symmetry-breaking in the gate potential. This wrong behavior is corrected in the right panel where Eq. (6.401) has been used.

Note, that the expressions for the retarded and lesser functions, described above, can be obtained in a more formal way by the EOM method formulated on the Keldysh contour.

Comparison with the master equation result In the single site model with two (spin-up and spin-down) levels it is possible to make the direct comparison between our *Ansatz* and the master equation

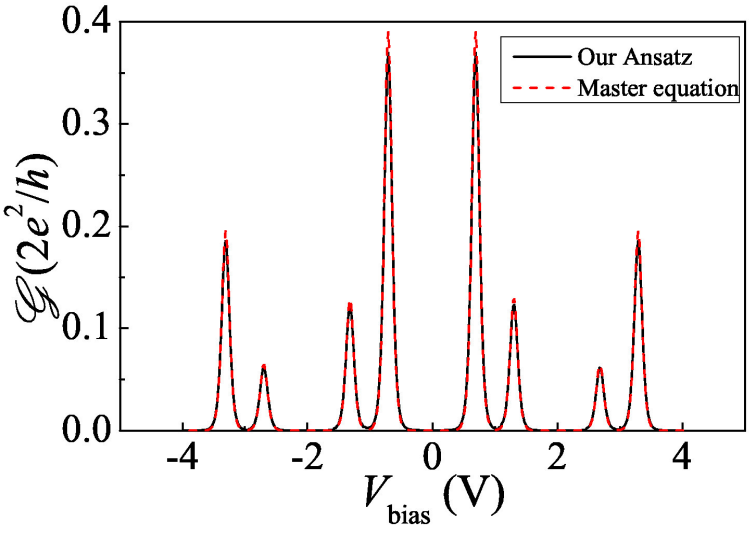


Figure 6.22: (Color online) The comparison of the master equation method and our *Ansatz* for the differential conductance of the two level model with $\epsilon_\uparrow = -0.35$ eV, $\epsilon_\downarrow = -0.65$ eV, $U = 1.0$ eV, $V_g = 1.0$ V, $\Gamma_L = \Gamma_R = 0.05$ eV.

methods. For the latter, we used the well known master equations for quantum dots [64, 65].

In the Fig. 6.22 the typical curves of the differential conductance as a function of the bias voltage at fixed gate voltage obtained by the two methods are shown together: there is basically no difference in the results obtained by these two methods. In the Fig. 6.23 the contour plot of the differential conductance obtained by our *Ansatz* is shown. We do not present here the contour plot obtained by the master equation method because it looks exactly the same.

It is quite clear from the presented figures that our *Ansatz* and the master equation method give essentially the same results in the limit of weak coupling to the leads. The systematic investigation of the deviations between the two methods at stronger tunneling will be presented in a separate publication.

It is important that our *Ansatz* can be applied straightforwardly to multilevel systems in the case when the exact eigenstates of an isolated system are unknown and the usage of the master equation method is not easy. In this paper we consider the simplest example of such a system, namely a double site case.

6.4.1.3 Double quantum dot (two sites)

We now return to the investigation of the DSJ system (Fig. 6.24) with Coulomb interaction on each site. The Hamiltonian is expressed as follows,

$$H = H_D + H_t + \sum_\alpha (H_\alpha + H_{\alpha D}),$$

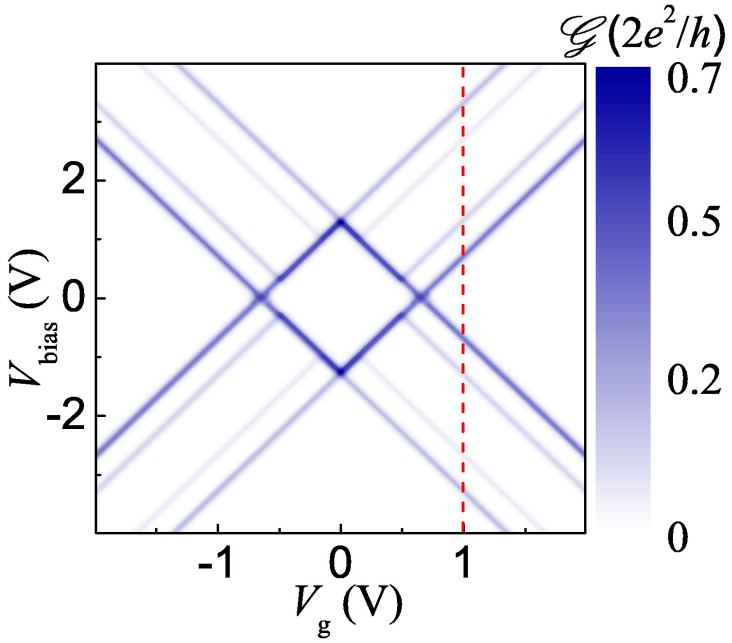


Figure 6.23: (Color online) The stability diagram (the contour plot of the differential conductance) calculated by our *Ansatz* for the two level model with parameters as in Fig. 6.22. The latter is indicated with a dash line at $V_g = 1.0$ V. $\epsilon_\downarrow = -0.65$ eV, $U = 1.0$ eV,

where

$$H_{\rm D} = \sum_{i,\sigma} \epsilon_{i,\sigma} d^{\dagger}_{i,\sigma} d_{i,\sigma} + \frac{U_i}{2} n_{i,\sigma} n_{i,\bar{\sigma}},$$
$$H_t = \sum_{i \neq j,\sigma} \frac{t}{2} (d^{\dagger}_{i,\sigma} d_{j,\sigma} + d^{\dagger}_{j,\sigma} d_{i,\sigma}),$$
$$H_{\alpha,\sigma} = \sum_{k,\sigma} \epsilon^{(\alpha)}_{k,\sigma} c^{\dagger}_{\alpha,k,\sigma} c_{\alpha,k,\sigma},$$
$$H_{\alpha{\rm D},\sigma} = \sum_{k,\sigma} \left(V_{\alpha,k,\sigma} c^{\dagger}_{\alpha,k,\sigma} d_{i,\sigma} + V^{*}_{\alpha,k,\sigma} d^{\dagger}_{i,\sigma} c_{\alpha,k,\sigma} \right),$$

with $i, j = 1, 2$ indicate the site, t is the constant for electron hopping between different sites.

With the help of the EOM, and by means of the truncation approximation on the double-particle GFs, we obtain the closed form for the retarded GFs as follows

$$(\omega - \epsilon_{i,\sigma} - \Sigma^{\rm r}_{i,\sigma}) G^{(U,t){\rm r}}_{i,\sigma;j,\tau} = \delta_{i,j}\delta_{\sigma,\tau} + U_i G^{(2)(U,t){\rm r}}_{i,\sigma;j,\tau} + t\ G^{(U,t){\rm r}}_{i,\sigma;j,\tau}, \tag{6.404a}$$
$$(\omega - \epsilon_{i,\sigma} - U_i - \Sigma^{\rm r}_{i,\sigma}) G^{(2)(U,t){\rm r}}_{i,\sigma;j,\tau} = \langle n_{i,\bar{\sigma}} \rangle \delta_{i,j}\delta_{\sigma,\tau} + t\ n_{i,\bar{\sigma}} G^{(U,t){\rm r}}_{i,\sigma;j,\tau}, \tag{6.404b}$$

where the DSJ retarded GFs are defined as

$$G^{(U,t){\rm r}}_{i,j;\sigma,\tau} = \langle\langle d_{i,\sigma} | d^{\dagger}_{j,\tau} \rangle\rangle^{\rm r}, \tag{6.405}$$
$$G^{(2)(U,t){\rm r}}_{i,j;\sigma,\tau} = \langle\langle n_{i,\bar{\sigma}} d_{i,\sigma} | d^{\dagger}_{j,\tau} \rangle\rangle^{\rm r}. \tag{6.406}$$

Here $\bar{i}$ means 'NOT i', and $\Sigma^{\rm r}_{i,\sigma}$ are the electron self-energy from leads.

From Eqs. (6.404a), (6.404b) and performing the same *Ansatz* as in the case of SSJ, we can obtain the DSJ lesser GFs with Coulomb-interaction effects as follows

$$G^{(U,t)<}(\omega) = (1 + G^{(U,t){\rm r}}\Sigma^{\rm r}_t) G^{(U)<} (1 + \Sigma^{\rm a}_t G^{(U,t){\rm a}}) + G^{(U,t){\rm r}} \Sigma^{<}_t G^{(U,t){\rm a}}, \tag{6.407}$$

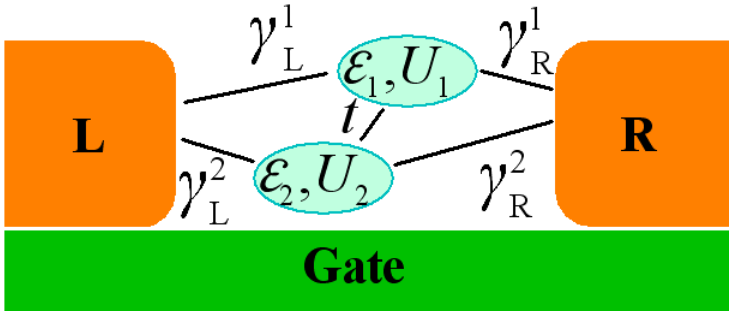


Figure 6.24: (Color online) The general configuration of a double site junction. The levels $\epsilon_{1,2}$ with charging energies $U_{1,2}$ are connected via t and coupled to the electrodes via the linewidth injection rates γ^i_α.

with

$$\Sigma_t^{\mathrm{r}} = \Sigma_t^{\mathrm{a}} = \begin{pmatrix} 0 & t & 0 & 0 \\ t & 0 & 0 & 0 \\ 0 & 0 & 0 & t \\ 0 & 0 & t & 0 \end{pmatrix},$$

and $\Sigma_t^{<} = 0$. $G^{(U)<}$ is the DSJ lesser GF with the same form as Eq. (6.401), but taking

$$U = \begin{pmatrix} U_1 & 0 & 0 & 0 \\ 0 & U_2 & 0 & 0 \\ 0 & 0 & U_1 & 0 \\ 0 & 0 & 0 & U_2 \end{pmatrix}, \ \Gamma_\alpha = \begin{pmatrix} \gamma_\alpha^1 & 0 & 0 & 0 \\ 0 & \gamma_\alpha^2 & 0 & 0 \\ 0 & 0 & \gamma_\alpha^1 & 0 \\ 0 & 0 & 0 & \gamma_\alpha^2 \end{pmatrix}, \tag{6.408}$$

where γ_α^i indicates the line width function of lead α to site i, and U_i is the charging energy at site i. $G^{\mathrm{r/a}}$ and $G^{(2)\mathrm{r/a}}$ are the GF matrix from Eqs. (6.404a) and (6.404b). Here, in order to distinguish different GFs, we introduce the subscript '(U,t)' for the one with both Coulomb interaction U and inter-site hopping t, while '(U)' for the one only with Coulomb interaction.

For our models, the lesser GFs in Eq. (6.401), (6.402) and (6.407), which are obtained with help of our *Ansatz*, can also be obtained by the EOM NEGF formula (6.331) within the same truncation approximation.

The current can be generally written as [106]

$$J = \frac{\mathrm{i}e}{2\hbar} \int \frac{d\epsilon}{2\pi} \mathrm{Tr}\{(\Gamma_{\mathrm{L}} - \Gamma_{\mathrm{R}}) G^{(U,t)<} + [f_{\mathrm{L}}(\omega)\Gamma_{\mathrm{L}} - f_{\mathrm{R}}(\omega)\Gamma_{\mathrm{R}}](G^{(U,t)\mathrm{r}} - G^{(U,t)\mathrm{a}})\},$$

where the lesser GF is given by Eq. (6.407). The differential conductance is defined as

$$\mathcal{G} = \frac{\partial J}{\partial V_{\mathrm{bias}}},$$

where the bias voltage is defined as $V_{\mathrm{bias}} = (\mu_{\mathrm{R}} - \mu_{\mathrm{L}})/e$.

Serial configuration By taking $\gamma_L^2 = \gamma_R^1 = 0$, we obtain a serial DSJ, which could describe the kind of molecular quantum junctions like the ones studied in Ref. [125]. First, at small bias voltages, the conductance with the two gate voltages V_{g_1} and V_{g_2} was calculated, and the relative stability diagram was obtained as shown in Fig. 6.25. Because of the double degeneracy (spin-up and spin-down) considered for each site and electrons hopping between the dots, there are eight resonance-tunnelling regions. This result is consistent with the master-equation approach. [109]

Further, we studied the nonequilibrium current for large bias-voltages (Fig. 6.26). Because $\epsilon_{1,\sigma}$ and $\epsilon_{2,\sigma}$ are taken as asymmetric, for the case without Coulomb interaction, the I-V curve is asymmetric for $\pm V_{\mathrm{bias}}$, and there are one step and one maximum for the current. The step contributes to one peak for

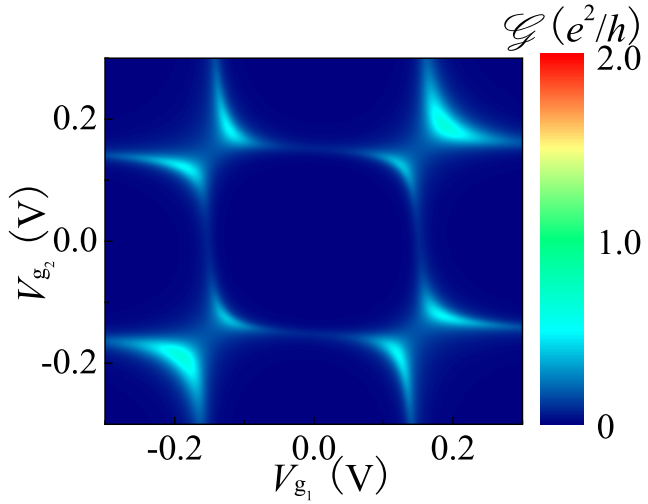


Figure 6.25: (Color online) The stability diagram of a serial DSJ with $\epsilon_{1,\sigma} = \epsilon_{2,\sigma} = -0.15$ eV, $U_1 = U_2 = 0.3$ eV, $t = 0.05$ eV, $\gamma_{\rm L}^1 = \gamma_{\rm R}^2 = 0.02$ eV, $\gamma_{\rm L}^2 = \gamma_{\rm R}^1 = 0$,$V_{\rm bias} = 0.005V$. The maximums of conductance are observed when the levels of the first site ($\epsilon_{1,\sigma}$ or $\epsilon_{1,\sigma} + U$) are overlapped with the levels of the second site ($\epsilon_{2,\sigma}$ or $\epsilon_{2,\sigma} + U$), and with the Fermi energy in the leads. The splitting of the four maximums is due to the hopping between the dots.

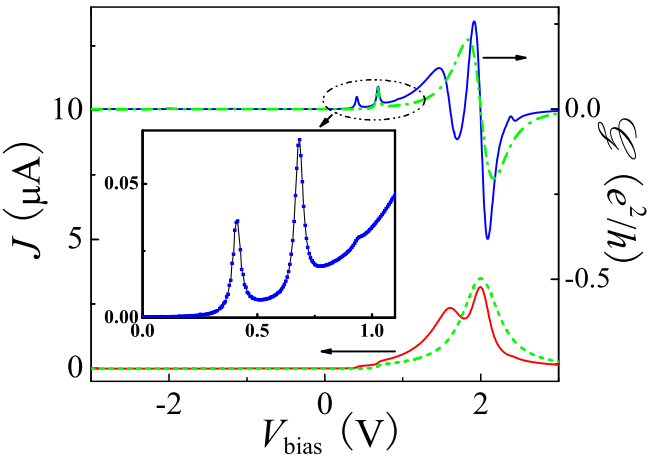


Figure 6.26: (Color online) Current and conductance vs. bias-voltage of a DSJ far from equilibrium with parameters $\epsilon_{1,\sigma} = 0.5$ eV, $\epsilon_{2,\sigma} = -0.5$ eV, $U_1 = U_2 = U = 0.2$ eV, $t = 0.07$ eV, $\gamma_{\rm L}^1 = \gamma_{\rm R}^2 = 0.03$ eV, $V_{g_2} = -V_{g_1} = V_{\rm bias}/4$ and $V_{\rm R} = -V_{\rm L} = V_{\rm bias}/2$. The red curve represents the current, while the blue the conductance. The inset is the blow-up for the conductance peak split. The dash and dot-dash curves are for current and conductance with $U = 0$, respectively.

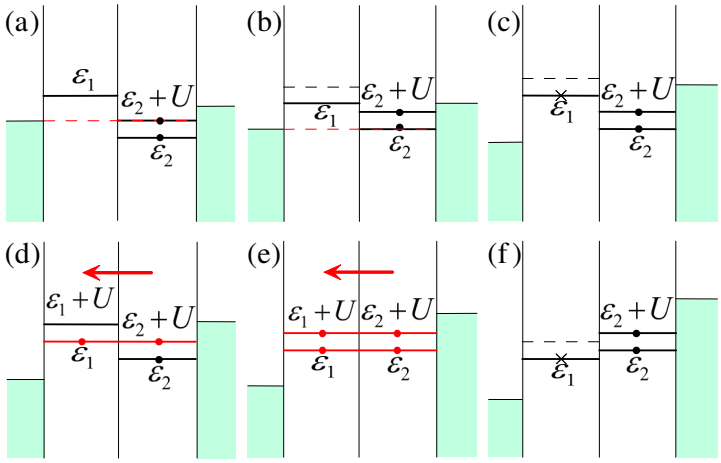


Figure 6.27: (Color online) The processes involved in the transport characteristics in figure 6.26. $\epsilon_1 \equiv \epsilon_{1,\sigma}$, $\epsilon_2 \equiv \epsilon_{2,\sigma}$, The red line indicates electron resonant-tunnelling. a) The first conductance peak. b) The second conductance peak. c) The pseudo-peak of conductance. d) The first current maximum, and the red line indicates resonant tunnelling of electrons. e) The second current maximum for electron resonant tunnelling. f) The dip of conductance.

the conductance. When we introduce the Coulomb interaction to the system, the one conductance peak is split into several: two peaks, one pseudo-peak and one dip, while the current maximum comes to be double split (see Fig. 6.26). The origin of this is in the effective splitting of the degenerate level when one of the spin states is occupied and the other is empty. When both spin states are occupied, the degeneracy is restored.

This process can be illustrated by the help of Fig. 6.27. At zero bias-voltage, $\epsilon_{2,\sigma}$ is occupied and $\epsilon_{1,\sigma}$ is empty. Then we start to increase the bias voltage. a) The level $\epsilon_{2,\sigma}+U$ is first opened for transport. It will contribute the first peak for conductance. b) Further, the levels $\epsilon_{2,\sigma}$ and $\epsilon_{1,\sigma}$ come into the transport window between the left and the right Fermi levels, resulting in the second peak. c) When the level $\epsilon_{1,\sigma}+U$ comes into play, only a pseudo-peak appears. This is because there is only a little possibility for electrons to occupy the level $\epsilon_{1,\sigma}$ under positive bias voltage. d) Levels $\epsilon_{2,\sigma}+U$ and $\epsilon_{1,\sigma}$ meet, which results in electron resonant-tunnelling and leads to the first maximum of the current. Then a new level $\epsilon_{1,\sigma}+U$ appears over the occupied $\epsilon_{1,\sigma}$ due to the Coulomb interaction. e) The meeting of $\epsilon_{2,\sigma}$ and $\epsilon_{1,\sigma}$ results in electron resonant tunnelling. It means that $\epsilon_{1,\sigma}$ will be occupied, which leads to the appearance of a new level $\epsilon_{1,\sigma}+U$. Then $\epsilon_{2,\sigma}+U$ meets $\epsilon_{1,\sigma}+U$ and another resonant tunnelling channel is opened for electrons. The two channels result in the second current maximum. f) finally, the level $\epsilon_{1,\sigma}+U$ disappears if the level $\epsilon_{1,\sigma}$ is empty. This means that a dip appears in the conductance.

It should be noted that the characteristics of serial DSJ in Fig. 6.26 have showed some reasonable similarities to experiments of a single-molecule diode. [125]

Parallel configuration If on the other hand, the two sites are symmetrically connected to the electrodes, possibly with a small inter-dot hopping, but with charging energies U_1 and U_2 fixed to different scales for transport. The resulting stability diagram contains both interference effects for parallel path-

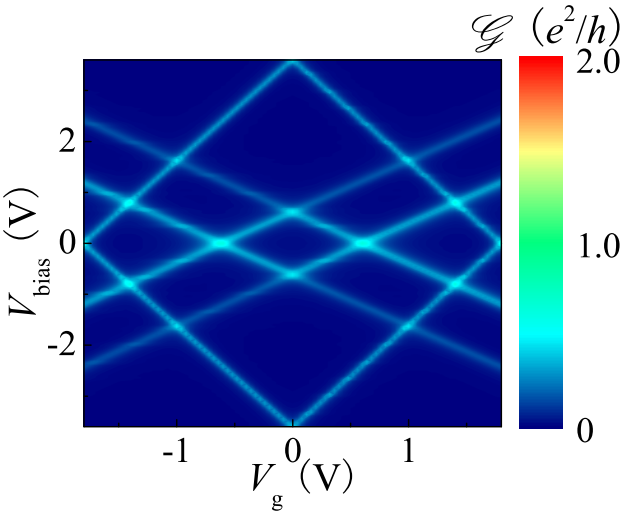


Figure 6.28: (Color online) Nested stability diagram of a parallel DSJ with parameters $\epsilon_{1,\sigma} = -1.8$ eV, $\epsilon_{2,\sigma} = -0.3$ eV, $U_1 = 3.6$ eV, $U_2 = 0.6$ eV $t = 0.001$ eV, $\gamma_{\rm L}^1 = \gamma_{\rm R}^1 = 0.04$ eV, $\gamma_{\rm L}^2 = \gamma_{\rm R}^2 = 0.05$ eV, $V_{\rm g_2} = V_{\rm g_1}/2 = V_{\rm g}/2$ and $V_{\rm R} = -V_{\rm L} = V_{\rm bias}/2$. See discussion in the text.

ways and an overlap of U_1 and U_2 stability diagrams, which we refer to a nesting characteristic. (see Fig. 6.28).

The physics of the *weak* lines in the figure can be understood by the help of charging effects. For simplicity, here we would ignore the site index i. In the region of large positive gate voltage at zero bias voltage, $\epsilon_\uparrow$ and $\epsilon_\downarrow$ are all empty, which means that the two levels are degenerate. Therefore adding a bias voltage, first, there will be two channels ($\epsilon_\uparrow$ and $\epsilon_\downarrow$) opened for current (thick lines). After then, one level ϵ_σ (spin-up or spin-down) is occupied, while the other obtains a shift for Coulomb interaction: $\epsilon_{\bar\sigma} \to \epsilon_{\bar\sigma} + U$. Therefore, when the bias voltage is further increased to make the Fermi-window boundary meeting level $\epsilon_{\bar\sigma} + U$, only one channel is opened for the current, which results in the *weak* lines in Fig. 6.28, which is the characteristic of CB. The similar case appears in the region of large negative gate voltages.

Finally, we here introduced a powerful *Ansatz* for the lesser Green function, which is consistent with both the Dyson-equation approach and the equation-of-motion approach. By using this *Ansatz* together with the standard equation-of-motion technique for the retarded and advanced Green functions, we obtained the NEGF for both the single and the double site junctions in the Coulomb blockade regime *at finite voltages* and calculated the transport observables. The method can be applied to describe self-consistently transport through single molecules with strong Coulomb interaction and arbitrary coupling to the leads.

To test our method, we here analyzed the CB stability diagrams for a SSJ and a DSJ. Our results are all consistent with the results of experiments and the master-equation approach. We showed, that the improved lesser Green function gives better results for weak molecule-to-contact couplings, where a

comparison with the master equation approach is possible.

For the serial configuration of a DSJ, such as a donor/acceptor rectifier, the I-V curves maintain a diode-like behavior, as it can be already inferred by coherent transport calculations. [126] Besides, we predict that as a result of charging effects, one conductance peak will be split into three peaks and one dip, and one current maximum into two. For a DSJ parallel configuration, due to different charging energies on the two dot sites, the stability diagrams show peculiar nesting characteristics.

6.4.2 Nonequilibrium vibrons

Though the electron-vibron model described in the Section II has a long history, the many questions it implies are not answered up to now. While the isolated electron-vibron model can be solved exactly by the so-called polaron or Lang-Firsov transformation [69, 70, 71], the coupling to the leads produces a true many-body problem. The inelastic resonant tunneling of *single* electrons through the localized state coupled to phonons was first considered in Refs. [72, 73, 74, 75]. There, the exact solution in the single-particle approximation was derived, ignoring completely the Fermi sea in the leads. At strong electron-vibron couplings and weak couplings to the leads, satellites of the main resonant peak are formed in the spectral function (Fig. 6.11). The number of the relevant side-bands is determined by the well known Huang-Rhys factor [76] $g = (\lambda/\omega_0)^2$. The question which remains is whether these side-bands can be observed in the differential conductance, when the coupling to all electrons in the leads should be taken into account simultaneously. New theoretical treatments were presented recently in Refs. [23, 24, 25, 26, 27, 28, 29, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 90, 87, 88, 89, 91, 92, 93]. In parallel, the theory of inelastic resonant tunneling in scanning tunneling spectroscopy was developed [94, 95, 96, 97, 98, 99]. For a recent review of the electron-vibron problem and its relation to charge transport at the molecular scale see Ref. [27].

Many interesting results by the investigation of quantum transport in the strong electron-vibron coupling limit has been achieved with the help of the master equation approach [80, 83, 87, 88, 89]. This method, however, is valid only in the limit of very weak molecule-to-lead coupling and neglects all spectral effects, which are the most important at finite coupling to the leads.

6.4.2.1 Nonequilibrium Dyson-Keldysh method

The model electron-vibron Hamiltonian We use the minimal transport model described in the previous sections. For convenience, we present the Hamiltonian here once more. The full Hamiltonian is the sum of the molecular Hamiltonian $\hat{H}_M$, the Hamiltonians of the leads $\hat{H}_{R(L)}$, the tunneling Hamiltonian $\hat{H}_T$ describing the molecule-to-lead coupling, the vibron Hamiltonian $\hat{H}_V$ including electron-vibron interaction and coupling of vibrations to the environment (describing dissipation of vibrons)

$$\hat{H} = \hat{H}_M + \hat{H}_V + \hat{H}_L + \hat{H}_R + \hat{H}_T. \tag{6.409}$$

A molecule is described by a set of localized states $|\alpha\rangle$ with energies ϵ_α and inter-orbital overlap integrals $t_{\alpha\beta}$ by the following model Hamiltonian:

$$\hat{H}_M^{(0)} = \sum_\alpha \left(\epsilon_\alpha + e\varphi_\alpha(t)\right) d_\alpha^\dagger d_\alpha + \sum_{\alpha\neq\beta} t_{\alpha\beta} d_\alpha^\dagger d_\beta. \tag{6.410}$$

Vibrations and the electron-vibron coupling are described by the Hamiltonian [23, 24, 25, 29] ($\hbar = 1$)

$$\hat{H}_V = \sum_q \omega_q a_q^\dagger a_q + \sum_{\alpha\beta}\sum_q \lambda_{\alpha\beta}^q (a_q + a_q^\dagger) d_\alpha^\dagger d_\beta. \tag{6.411}$$

Here vibrations are considered as localized phonons and q is an index labeling them, not the wave-vector. The first term describes free vibrons with the energy ω_q. The second term represents the electron-vibron interaction. We include both diagonal coupling, which describes a change of the electrostatic energy with the distance between atoms, and the off-diagonal coupling, which describes the dependence of the matrix elements $t_{\alpha\beta}$ over the distance between atoms.

The Hamiltonians of the right (R) and left (L) leads read

$$\hat{H}_{i=L(R)} = \sum_{k\sigma} (\epsilon_{ik\sigma} + e\varphi_i) c^\dagger_{ik\sigma} c_{ik\sigma}, \tag{6.412}$$

$\varphi_i(t)$ are the electrical potentials of the leads. Finally, the tunneling Hamiltonian

$$\hat{H}_T = \sum_{i=L,R} \sum_{k\sigma,\alpha} \left(V_{ik\sigma,\alpha} c^\dagger_{ik\sigma} d_\alpha + \text{h.c.} \right) \tag{6.413}$$

describes the hopping between the leads and the molecule. A direct hopping between two leads is neglected.

Keldysh-Dyson equations and self-energies We use the nonequilibrium Green function (NGF) method, as introduced in Section III. The current in the left ($i = L$) or right ($i = R$) contact to the molecule is described by the expression

$$J_{i=L,R} = \frac{ie}{\hbar} \int \frac{d\epsilon}{2\pi} \text{Tr} \left\{ \mathbf{\Gamma}_i(\epsilon - e\varphi_i) \left(\mathbf{G}^<(\epsilon) + f^0_i(\epsilon - e\varphi_i) \left[\mathbf{G}^R(\epsilon) - \mathbf{G}^A(\epsilon) \right] \right) \right\}, \tag{6.414}$$

where $f^0_i(\epsilon)$ is the equilibrium Fermi distribution function with chemical potential μ_i, and the level-width function is

$$\mathbf{\Gamma}_{i=L(R)}(\epsilon) = \Gamma_{i\alpha\beta}(\epsilon) = 2\pi \sum_{k\sigma} V_{ik\sigma,\beta} V^*_{ik\sigma,\alpha} \delta(\epsilon - \epsilon_{ik\sigma}).$$

The lesser (retarded, advanced) Green function matrix of a nonequilibrium molecule $\mathbf{G}^{<(R,A)} \equiv G^{<(R,A)}_{\alpha\beta}$ can be found from the Dyson-Keldysh equations in the integral form

$$\mathbf{G}^R(\epsilon) = \mathbf{G}^R_0(\epsilon) + \mathbf{G}^R_0(\epsilon) \mathbf{\Sigma}^R(\epsilon) \mathbf{G}^R(\epsilon), \tag{6.415}$$

$$\mathbf{G}^<(\epsilon) = \mathbf{G}^R(\epsilon) \mathbf{\Sigma}^<(\epsilon) \mathbf{G}^A(\epsilon), \tag{6.416}$$

or from the corresponding equations in the differential form (see e.g. Refs. [28, 29] and references therein).

Here

$$\mathbf{\Sigma}^{R,<} = \mathbf{\Sigma}^{R,<(T)}_L + \mathbf{\Sigma}^{R,<(T)}_R + \mathbf{\Sigma}^{R,<(V)} \tag{6.417}$$

is the total self-energy of the molecule composed of the tunneling (coupling to the left and right leads) self-energies

$$\mathbf{\Sigma}^{R,<(T)}_{j=L,R} \equiv \Sigma^{R,<(T)}_{j\alpha\beta} = \sum_{k\sigma} \left\{ V^*_{jk\sigma,\alpha} G^{R,<}_{jk\sigma} V_{jk\sigma,\beta} \right\}, \tag{6.418}$$

and the vibronic self-energy $\mathbf{\Sigma}^{R,<(V)} \equiv \Sigma^{R,<(V)}_{\alpha\beta}$.

For the retarded tunneling self-energy $\mathbf{\Sigma}^{R(T)}_i$ one obtains

$$\mathbf{\Sigma}^{R(T)}_i(\epsilon) = \mathbf{\Lambda}_i(\epsilon - e\varphi_i) - \frac{\mathrm{i}}{2}\mathbf{\Gamma}_i(\epsilon - e\varphi_i), \tag{6.419}$$

where $\mathbf{\Lambda}_i$ is the real part of the self-energy, which usually can be included in the single-particle Hamiltonian $\hat{H}^{(0)}_M$, and $\mathbf{\Gamma}_i$ describes level broadening due to coupling to the leads. For the corresponding lesser function one finds

$$\mathbf{\Sigma}^{<(T)}_i(\epsilon) = \mathrm{i}\mathbf{\Gamma}_i(\epsilon - e\varphi_i) f^0_i(\epsilon - e\varphi_i). \tag{6.420}$$

In the standard self-consistent Born approximation, using the Keldysh technique, one obtains for the vibronic self-energies [83, 23, 24, 25, 84, 85, 27, 29]

$$\mathbf{\Sigma}^{R(V)}(\epsilon) = \frac{i}{2}\sum_q \int \frac{d\omega}{2\pi}\left(\mathbf{M}^q \mathbf{G}^R_{\epsilon-\omega}\mathbf{M}^q D^K_{q\omega} + \right.$$

$$\left. + \mathbf{M}^q \mathbf{G}^K_{\epsilon-\omega}\mathbf{M}^q D^R_{q\omega} - 2D^R_{q\omega=0}\mathbf{M}^q \mathrm{Tr}\left[\mathbf{G}^<_\omega \mathbf{M}^q\right]\right), \tag{6.421}$$

$$\mathbf{\Sigma}^{<(V)}(\epsilon) = i\sum_q \int \frac{d\omega}{2\pi}\mathbf{M}^q \mathbf{G}^<_{\epsilon-\omega}\mathbf{M}^q D^<_{q\omega}, \tag{6.422}$$

where $\mathbf{G}^K = 2\mathbf{G}^< + \mathbf{G}^R - \mathbf{G}^A$ is the Keldysh Green function, and $\mathbf{M}^q \equiv M^q_{\alpha\beta}$.

If vibrons are noninteracting, in equilibrium, and non-dissipative, then the vibronic Green functions write:

$$D^R_0(q,\omega) = \frac{1}{\omega - \omega_q + \mathrm{i}0^+} - \frac{1}{\omega + \omega_q + \mathrm{i}0^+}, \tag{6.423}$$

$$D^<_0(q,\omega) = -2\pi\mathrm{i}\left[(f^0_B(\omega_q)+1)\delta(\omega+\omega_q)\right.$$

$$\left. + f^0_B(\omega_q)\delta(\omega-\omega_q)\right], \tag{6.424}$$

where the equilibrium Bose distribution function is

$$f^0_B(\omega) = \frac{1}{\exp(\omega/T) - 1}. \tag{6.425}$$

In the Migdal model the retarded vibron function is calculated from the Dyson-Keldysh equation

$$D^R(q,\omega) = \frac{2\omega_q}{\omega^2 - \omega_q^2 - 2\omega_q \Pi^R(q,\omega)}, \tag{6.426}$$

where $\Pi(q,\omega)$ is the polarization operator (boson self-energy). The equation for the lesser function (quantum kinetic equation in the integral form) is

$$(\Pi^R_{q\omega} - \Pi^A_{q\omega})D^<_{q\omega} - (D^R_{q\omega} - D^A_{q\omega})\Pi^<_{q\omega} = 0, \tag{6.427}$$

this equation in the stationary case considered here is algebraic in the frequency domain.

The polarization operator is the sum of two parts, environmental and electronic: $\Pi^{R,<}_{q\omega} = \Pi^{R,<(\mathrm{env})}_{q\omega} + \Pi^{R,<(\mathrm{el})}_{q\omega}$.

The environmental equilibrium part of the polarization operator can be approximated by the simple expressions

$$\Pi^{R(\mathrm{env})}(q,\omega) = -\frac{i}{2}\gamma_q \mathrm{sign}(\omega), \tag{6.428}$$

$$\Pi^{<(\mathrm{env})}(q,\omega) = -i\gamma_q f^0_B(\omega)\mathrm{sign}(\omega), \tag{6.429}$$

where γ_g is the vibronic dissipation rate, and $f^0_B(\omega)$ is the equilibrium Bose-Einstein distribution function.

The electronic contribution to the polarization operator within the SCBA is

$$\Pi^{R(\mathrm{el})}(q,\omega) = -i\int \frac{d\epsilon}{2\pi}\mathrm{Tr}\left(\mathbf{M}^q\mathbf{G}^<_\epsilon\mathbf{M}^q\mathbf{G}^A_{\epsilon-\omega} + \mathbf{M}^q\mathbf{G}^R_\epsilon\mathbf{M}^q\mathbf{G}^<_{\epsilon-\omega}\right), \tag{6.430}$$

$$\Pi^{<(\mathrm{el})}(q,\omega) = -i\int \frac{d\epsilon}{2\pi}\mathrm{Tr}\left(\mathbf{M}^q\mathbf{G}^<_\epsilon\mathbf{M}^q\mathbf{G}^>_{\epsilon-\omega}\right). \tag{6.431}$$

We obtained the full set of equations, which can be used for numerical calculations.

6.4.2.2 Single-level model: spectroscopy of vibrons

The isolated single-level electron-vibron model is described by the Hamiltonian

$$\hat{H}_{M+V} = (\epsilon_0 + e\varphi_0)d^\dagger d + \omega_0 a^\dagger a + \lambda\left(a^\dagger + a\right)d^\dagger d, \tag{6.432}$$

where the first and the second terms describe the free electron state and the free vibron, and the third term is electron-vibron minimal coupling interaction.

The electrical potential of the molecule φ_0 plays an important role in transport at finite voltages. It describes the shift of the molecular level by the bias voltage, which is divided between the left lead (tip), the right lead (substrate), and the molecule as $\varphi_0 = \varphi_R + \eta(\varphi_L - \varphi_R)$ [100]. We assume the simplest linear dependence of the molecular potential ($\eta = const$), but its nonlinear dependence [101] can be easily included in our model.

Here we assume, that the vibrons are in equilibrium and are not excited by the current, so that the self-consistent Born approximation is a good starting point. The vibron Green function are assumed to be equilibrium with the broadening defined by the external thermal bath, see for details Refs. [24, 84, 85, 27, 29].

For the single-level model all equations are significantly simplified. Combining J_L and J_R the expression for the current can be written for energy independent $\Gamma_{L(R)}$ (wide-band limit) as

$$J = \frac{e}{h}\frac{\Gamma_L\Gamma_R}{\Gamma_R+\Gamma_L}\int d\epsilon A(\epsilon)\left[f^0(\epsilon - e\varphi_L) - f^0(\epsilon - e\varphi_R)\right]. \tag{6.433}$$

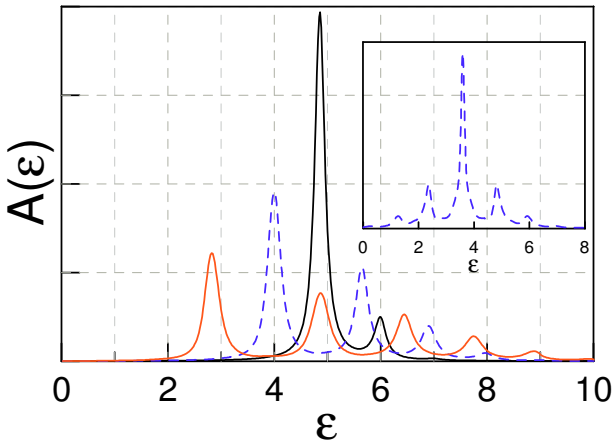


Figure 6.29: (Color online) Spectral function at different electron-vibron couplings: $\lambda/\omega_0 = 0.4$ (black), $\lambda/\omega_0 = 1.2$ (blue, dashed), and $\lambda/\omega_0 = 2$ (red); at $\epsilon_0/\omega_0 = 5$, $\Gamma_L/\omega_0 = \Gamma_R/\omega_0 = 0.1$. In the insert the spectral function at $\lambda/\omega_0 = 1.2$ is shown at finite voltage, when the level is partially filled. Energies are in units of $\hbar\omega_0$.

It looks as simple as the Landauer-Büttiker formula, but it is not trivial, because the spectral density $A(\epsilon) = -2\mathrm{Im}G^R(\epsilon)$ now depends on the distribution function of the electrons in the fluctuating molecule and hence the applied voltage, $\varphi_L = -\varphi_R = V/2$ [28]. Indeed, $G^R(\epsilon)$ can be found from (6.201)

$$G^R(\epsilon) = \frac{1}{\epsilon - \tilde{\epsilon}_0 - \Sigma^{R(V)}(\epsilon) + i(\Gamma_L + \Gamma_R)/2}, \tag{6.434}$$

where $\Sigma^{R(V)}(\epsilon)$ is a functional of the electron distribution function inside a molecule. Actually, the lesser function $G^<(\epsilon)$ is used in the quantum kinetic formalism as a distribution function. In the single-level case the usual distribution function can be introduced through the relation

$$G^<(\epsilon) = iA(\epsilon)f(\epsilon). \tag{6.435}$$

Note the essential difference between symmetric ($\Gamma_L = \Gamma_R$) and asymmetric junctions. It is clear from the noninteracting solution of the transport problem. Neglecting for a moment the vibron self-energies, we obtain the noninteracting distribution function

$$f(\epsilon) = \frac{\Gamma_L f_L^0(\epsilon - e\varphi_L) + \Gamma_R f_R^0(\epsilon - e\varphi_R)}{\Gamma_L + \Gamma_R}. \tag{6.436}$$

For strongly asymmetric junctions (e.g. $\Gamma_L \ll \Gamma_R$) the distribution function remains close to the equilibrium function in the right lead $f_R^0(\epsilon - e\varphi_R)$, thus essentially simplifying the solution. While for symmetric junctions the distribution function has the double-step form and is very different from the equilibrium one.

A typical example of the spectral function at zero voltage is shown in Fig. 6.29. At finite voltage it should be calculated self-consistently. In the insert the spectral function of the symmetric junction at

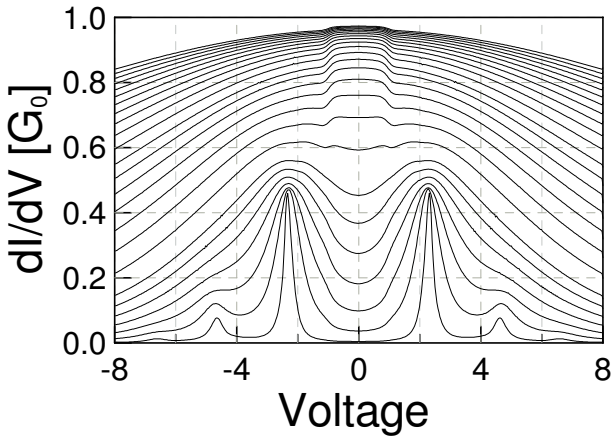


Figure 6.30: Differential conductance of a *symmetric* junction ($\eta = 0.5$, $\Gamma_R = \Gamma_L$) at different molecule-to-lead coupling, from $\Gamma_L/\omega_0 = 0.1$ (lower curve) to $\Gamma_L/\omega_0 = 10$ (upper curve), $\lambda/\omega_0 = 1$, $\epsilon_0/\omega_0 = 2$. Voltage is in the units of $\hbar\omega_0/e$.

finite voltage is shown, it is changed essentially because the distribution function is changed.

Let us discuss a general picture of the vibronic transport in symmetric and asymmetric single-molecule junctions, provided in experiments with the molecular bridges and STM-to-molecule junctions, respectively. The differential conductance, calculated at different molecule-to-lead coupling, is shown in Fig. 6.30 (symmetric) and Fig. 6.31 (asymmetric). At weak coupling, the vibronic side-band peaks are observed, reproducing the corresponding peaks in the spectral function. At strong couplings the broadening of the electronic state hides the side-bands, and new features become visible. In the symmetric junction, a suppression of the conductance at $V \simeq \pm\hbar\omega_0$ takes place as a result of inelastic scattering of the coherently transformed from the left lead to the right lead electrons. In the asymmetric junction (Fig. 6.31), the usual IETS increasing of the conductance is observed at a negative voltage $V \simeq -\hbar\omega_0$, this feature is weak and can be observed only in the incoherent tail of the resonant conductance. We conclude, that the vibronic contribution to the conductance can be distinguished clearly in both coherent and tunneling limits.

Now let us discuss the particular situation of STS experiments [102, 103, 104, 105]. Here we concentrate mainly on the dependence on the tip-to-molecule distance [103]. When the tip (left lead in our notations) is far from the molecule, the junction is strongly asymmetric: $\Gamma_L \ll \Gamma_R$ and $\eta \to 0$, and the conductance is similar to that shown in Fig. 6.31. When the tip is close to the molecule, the junction is approximately symmetric: $\Gamma_L \approx \Gamma_R$ and $\eta \approx 0.5$, and the conductance curve is of the type shown in Fig. 6.30. We calculated the transformation of the conductance from the asymmetric to symmetric case (Fig. 6.32). It is one new feature appeared in asymmetric case due to the fact that we started from a finite parameter $\eta = 0.2$ (in the Fig. 6.31 $\eta = 0$), namely a single peak at negative voltages, which is shifted to smaller voltage in the symmetric junction. The form and behavior of this peak is in agreement with experimental results [103].

In conclusion, at weak molecule-to-lead (tip, substrate) coupling the usual vibronic side-band peaks

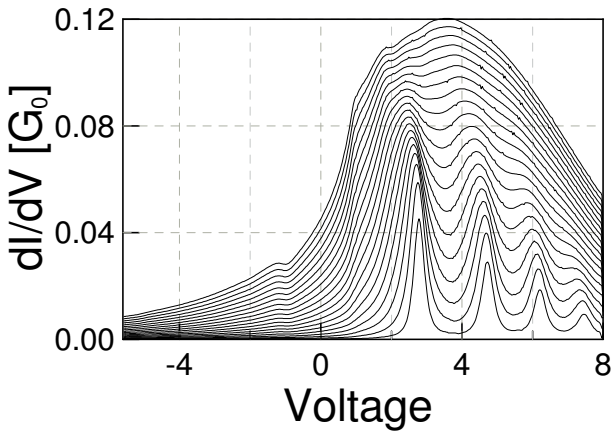


Figure 6.31: Differential conductance of an *asymmetric* junction ($\eta = 0$, $\Gamma_R = 20\Gamma_L$) at different molecule-to-lead coupling, from $\Gamma_R/\omega_0 = 0.2$ (lower curve) to $\Gamma_R/\omega_0 = 4$ (upper curve), $\lambda/\omega_0 = 2$, $\epsilon_0/\omega_0 = 5$. The voltage is in the units of $\hbar\omega_0/e$

in the differential conductance are observed; at stronger coupling to the leads (broadening) these peaks are transformed into step-like features. A vibronic-induced decreasing of the conductance with voltage is observed in high-conductance junctions. The usual IETS feature (increasing of the conductance) can be observed only in the case of low off-resonant conductance. By changing independently the bias voltage and the tip position, it is possible to determine the energy of molecular orbitals and the spectrum of molecular vibrations. In the multi-level systems with strong electron-electron interaction further effects, such as Coulomb blockade and Kondo effect, could dominate over the physics which we address here; these effects have to be included in a subsequent step.

6.4.2.3 Multi-level model: nonequilibrium vibrons

Basically there are two main nonequilibrium effects: the electronic spectrum modification and excitation of vibrons (quantum vibrations). In the weak electron-vibron coupling case the spectrum modification is usually small (which is dependent, however, on the vibron dissipation rate, temperature, etc.) and the main possible nonequilibrium effect is the excitation of vibrons at finite voltages. We have developed an analytical theory for this case [29]. This theory is based on the self-consistent Born approximation (SCBA), which allows to take easily into account and calculate nonequilibrium distribution functions of electrons and vibrons.

If the mechanical degrees of freedom are coupled strongly to the environment (dissipative vibron), then the dissipation of molecular vibrations is determined by the environment. However, if the coupling of vibrations to the leads is weak, we should consider the case when the vibrations are excited by the current flowing through a molecule, and the dissipation of vibrations is also determined essentially by the coupling to the electrons. Here , we show that the effects of vibron emission and vibronic instability are important especially in the case of electron-vibron resonance.

We simplify the equations and obtain some analytical results in the *vibronic quasiparticle approxima-*

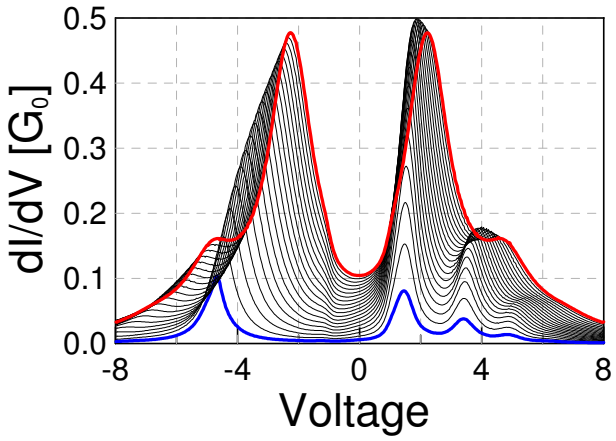


Figure 6.32: (Color online) Differential conductance at different molecule-to-STM coupling (see the text), from *asymmetric* junction with $\Gamma_L/\omega_0 = 0.025$, $\Gamma_R/\omega_0 = 0.5$ and $\eta = 0.2$ (lower curve, blue thick line) to *symmetric* junction with $\Gamma_L/\omega_0 = \Gamma_R/\omega_0 = 0.5$ and $\eta = 0.5$ (upper curve, red thick line), $\lambda/\omega_0 = 1$, $\epsilon_0/\omega_0 = 2$. Voltage is in the units of $\hbar\omega_0/e$

tion, which assumes weak electron-vibron coupling limit and weak external dissipation of vibrons:

$$\gamma_q^* = \gamma_q - 2\mathrm{Im}\Pi^R(\omega_q) \ll \omega_q. \tag{6.437}$$

So that the spectral function of vibrons can be approximated by the Dirac δ, and the lesser function reads

$$D^<(q,\omega) = -2\pi i \left[(N_q+1)\delta(\omega+\omega_q) + N_q\delta(\omega-\omega_q)\right], \tag{6.438}$$

where N_q is (nonequilibrium) number of vibrations in the q-th mode. So, in this approximation the spectrum modification of vibrons is not taken into account, but the possible excitation of vibrations is described by the nonequilibrium N_q. The dissipation of vibrons is neglected in the spectral function, but is taken into account later in the kinetic equation for N_q. A similar approach to the single-level problem was considered recently in [96, 97, 98, 99, 83, 84, 85]. The more general case with broadened equilibrium vibron spectral function seems to be not very interesting, because in this case vibrons are not excited. Nevertheless, in the numerical calculation it can be easy taken into consideration.

From the general quantum kinetic equation for vibrons, we obtain in this limit

$$N_q = \frac{\gamma_q N_q^0 - \mathrm{Im}\Pi^<(\omega_q)}{\gamma_q - 2\mathrm{Im}\Pi^R(\omega_q)}. \tag{6.439}$$

This expression describes the number of vibrons N_q in a nonequilibrium state, $N_q^0 = f_B^0(\omega_q)$ is the equilibrium number of vibrons. In the linear approximation the polarization operator is independent of N_q and $-2\mathrm{Im}\Pi^R(\omega_q)$ describes additional dissipation. Note that in equilibrium $N_q \equiv N_q^0$ because $\mathrm{Im}\Pi^<(\omega_q) = 2\mathrm{Im}\Pi^R(\omega_q) f_B^0(\omega_q)$. See also detailed discussion of vibron emission and absorption rates in Refs. [96, 97, 98, 99].

For weak electron-vibron coupling the number of vibrons is close to equilibrium and is changed because of *vibron emission* by nonequilibrium electrons, N_q is roughly proportional to the number of such electrons, and the distribution function of nonequilibrium electrons is not change essentially by the interaction with vibrons (perturbation theory can be used). The situation changes, however, if nonequilibrium dissipation $-2\mathrm{Im}\Pi^R(\omega_q)$ is *negative*. In this case the number of vibrons can be essentially larger than in the equilibrium case (*vibronic instability*), and the change of electron distribution function should be taken into account self-consistently.

In the stationary state the *nonlinear* dissipation rate

$$\gamma_q^* = \gamma_q - 2\mathrm{Im}\Pi^R(\omega_q) \tag{6.440}$$

is positive, but the nonequilibrium contribution to dissipation $-2\mathrm{Im}\Pi^R(\omega_q)$ remains negative.

Additionally to the vibronic quasiparticle approximation, the *electronic quasiparticle approximation* can be used when the coupling to the leads is weak. In this case the lesser function can be parameterized through the number of electrons F_η *in the eigenstates of the noninteracting molecular Hamiltonian* $H_M^{(0)}$

$$G^<_{\alpha\beta} = i\sum_{\gamma\eta} A_{\alpha\gamma} S_{\gamma\eta} F_\eta S^{-1}_{\eta\beta}, \tag{6.441}$$

we introduce the unitary matrix $\mathbf{S}$, which transfer the Hamiltonian $\mathbf{H} \equiv H^{(0)}_{M\alpha\beta}$ into the diagonal form $\tilde{\mathbf{H}} = \mathbf{S}^{-1}\mathbf{H}\mathbf{S}$, so that the spectral function of this diagonal Hamiltonian is

$$\tilde{A}_{\delta\eta}(\epsilon) = 2\pi\delta(\epsilon - \tilde{\epsilon}_\delta)\delta_{\delta\eta}, \tag{6.442}$$

where $\tilde{\epsilon}_\delta$ are the eigenenergies.

Note that in the calculation of the self-energies and polarization operators we can not use δ-approximation for the spectral function (this is too rough and results in the absence of interaction out of the exact electron-vibron resonance). So that in the calculation we use actually (6.441) with broadened equilibrium spectral function. This approximation can be systematically improved by including nonequilibrium corrections to the spectral function, which are important near the resonance. It is important to comment that for stronger electron-vibron coupling *vibronic side-bands* are observed in the spectral function and voltage-current curves at energies $\tilde{\epsilon}_\delta \pm n\omega_q$, we do not consider these effects in the rest of our paper and concentrate on resonance effects.

After correspondingly calculations we obtain finally

$$N_q = \frac{\gamma_q N_q^0 - \sum_{\eta\delta}\kappa_{\eta\delta}(\omega_q)F_\eta(F_\delta - 1)}{\gamma_q - \sum_{\eta\delta}\kappa_{\eta\delta}(\omega_q)(F_\eta - F_\delta)}, \tag{6.443}$$

where coefficients $\kappa_{\eta\delta}$ are determined by the spectral function and electron-vibron coupling in the diagonal representation

$$\kappa_{\eta\delta}(\omega_q) = \int \frac{d\epsilon}{2\pi} \tilde{M}^q_{\eta\delta}\tilde{A}_{\delta\delta}(\epsilon - \omega_q)\tilde{M}^q_{\delta\eta}\tilde{A}_{\eta\eta}(\epsilon), \tag{6.444}$$

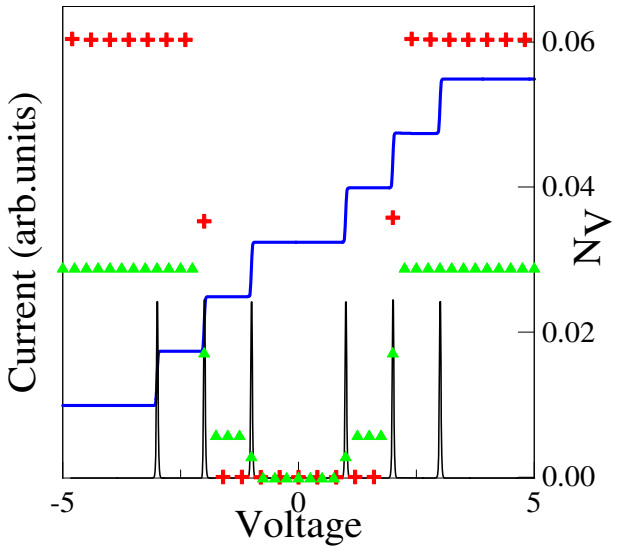


Figure 6.33: (Color online) Vibronic emission in the symmetric multilevel model: voltage-current curve, differential conductance, and the number of excited vibrons in the off-resonant (triangles) and resonant (crosses) cases (details see in the text).

$$F_\eta = \frac{\tilde{\Gamma}_{L\eta\eta} f^0_{L\eta} + \tilde{\Gamma}_{R\eta\eta} f^0_{R\eta} + \sum_{q\eta} \left[\zeta^{-q}_{\eta\delta} F_\delta N_q + \zeta^{+q}_{\eta\delta} F_\delta (1+N_q) \right]}{\tilde{\Gamma}_{L\eta\eta} + \tilde{\Gamma}_{R\eta\eta} + \sum_{q\eta} \left[\zeta^{-q}_{\eta\delta} (1-F_\delta+N_q) + \zeta^{+q}_{\eta\delta} (F_\delta+N_q) \right]}, \tag{6.445}$$

$$\zeta^{\pm q}_{\eta\delta} = \tilde{M}^q_{\eta\delta} \tilde{A}_{\delta\delta}(\tilde{\epsilon}_\eta \pm \omega_q) \tilde{M}^q_{\delta\eta}, \tag{6.446}$$

here $\tilde{\Gamma}_{i\eta\eta}$ and $f^0_{i\eta}$ are the level width matrix in the diagonal representation and Fermi function at energy $\tilde{\epsilon}_\eta - e\varphi_i$.

These kinetic equations are similar to the usual golden rule equations, but are more general.

Now let us consider several examples of vibron emission and vibronic instability.

First we consider the most simple case, when the instability is not possible and only vibron emission takes place. This corresponds to a negative imaginary part of the electronic polarization operator: $\mathrm{Im}\Pi^{R)}(\omega_q) < 0$. From the Eq. (6.444) one can see that for any two levels with the energies $\tilde{\epsilon}_\eta > \tilde{\epsilon}_\delta$ the coefficient $\kappa_{\eta\delta}$ is larger than $\kappa_{\delta\eta}$, because the spectral function $\tilde{A}_{\delta\delta}(\epsilon)$ has a maximum at $\epsilon = \tilde{\epsilon}_\delta$. The contribution of $\kappa_{\eta\delta}(\omega_q)(F_\eta - F_\delta)$ is negative if $F_\eta < F_\delta$. This takes place in equilibrium, and in nonequilibrium for transport through *symmetric* molecules, when higher energy levels are populated after lower levels. The example of such a system is shown in Fig. 6.33. Here we consider a simple three-level system ($\tilde{\epsilon}_1 = 1$, $\tilde{\epsilon}_2 = 2$, $\tilde{\epsilon}_3 = 3$) coupled symmetrically to the leads ($\Gamma_{L\eta} = \Gamma_{R\eta} = 0.01$). The current-voltage curve is the same with and without vibrations in the case of symmetrical coupling to the leads and in the weak electron-vibron coupling limit (if we neglect change of the spectral function). The figure shows how vibrons are excited, the number of vibrons N_V in the mode with frequency ω_0 is presented in two cases. In the off-resonant case (green triangles) N_V is very small comparing with the resonant case ($\omega_0 = \tilde{\epsilon}_2 - \tilde{\epsilon}_1$, red crosses, the vertical scale is changed for the off-resonant points). In fact, if the number

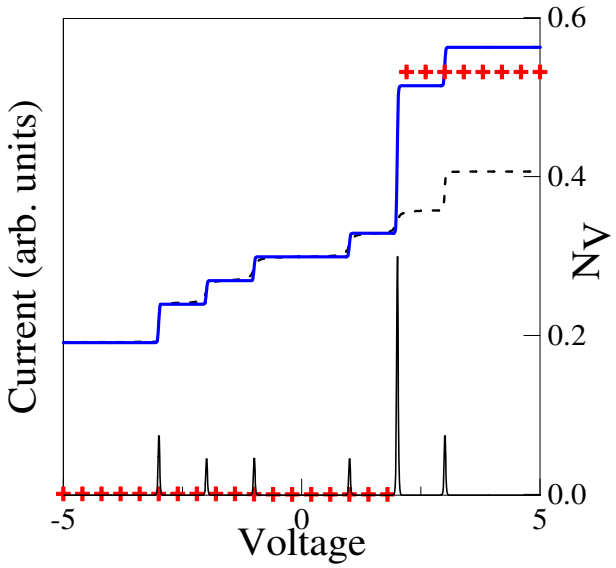


Figure 6.34: (Color online) Vibronic instability in an asymmetric multilevel model: voltage-current curve, differential conductance, and the number of excited vibrons (crosses). Dashed line show the voltage-current curve without vibrons (details see in the text).

of vibrons is very large, the spectral function and voltage-current curve are changed. We shall consider this in a separate publication.

Now let us consider the situation when the imaginary part of the electronic polarization operator can be positive: $\mathrm{Im}\Pi^R(\omega_q) > 0$. Above we considered the normal case when the population of higher energy levels is smaller than lower levels. The opposite case $F_2 > F_1$ is known as inversion in laser physics. Such a state is unstable if the total dissipation γ_q^* (6.440) is negative, which is possible only in the nonstationary case. As a result of the instability, a large number of vibrons is excited, and in the stationary state γ_q^* is positive. This effect can be observed for transport through *asymmetric* molecules, when higher energy levels are populated *before* lower levels. The example of a such system is shown in Fig. 6.34. It is the same three-level system as before, but the first and second levels are coupled not symmetrically to the leads ($\Gamma_{L1} = 0.001$, $\Gamma_{R1} = 0.1$, $\Gamma_{L2} = 0.1$, $\Gamma_{R2} = 0.001$). The vibron couple resonantly these levels ($\omega_q = \tilde{\epsilon}_2 - \tilde{\epsilon}_1$). The result is qualitatively different from the symmetrical case. The voltage-current curve is now asymmetric, a large *step* corresponds to the resonant level with inverted population.

Note the importance of the off-diagonal electron-vibron coupling for the resonant effects. If the matrix $\tilde{\mathbf{M}}$ in the eigen-state representation is diagonal, there is no resonant coupling between different electronic states.

Finally, let us consider the important case, when initially symmetric molecule becomes asymmetric when the external voltage is applied. The reason for such asymmetry is simply that in the external electric field left and right atoms feel different electrical potentials and the position of the levels $\epsilon_\alpha = \epsilon_\alpha^{(0)} + e\varphi_\alpha$ is changed (float) with the external voltage. The example of a such system is shown in Fig. 6.35. Here we consider a two-level system, one level is coupled electrostatically to the left lead $\tilde{\epsilon}_1 \propto \varphi_L$, the other level to the right lead $\tilde{\epsilon}_2 \propto \varphi_R$, the tunneling coupling to the leads also is not symmetrical ($\Gamma_{L1} = 0.1$,

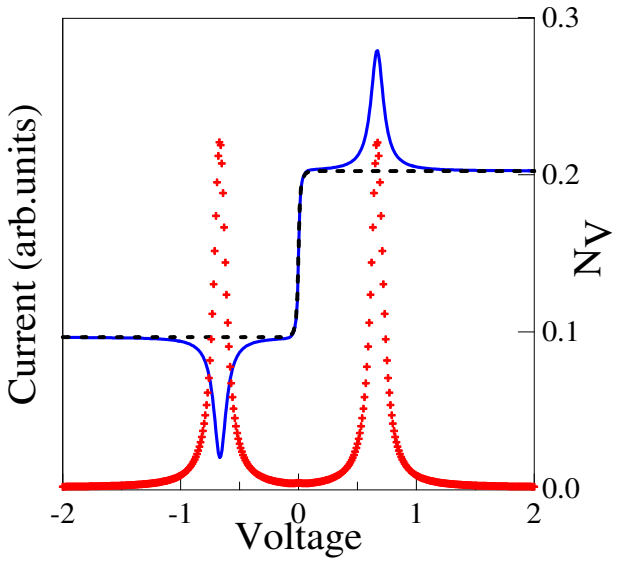


Figure 6.35: (Color online) Floating level resonance: voltage-current curve and the number of excited vibrons (crosses). Dashed line show the voltage-current curve without vibrons (details see in the text).

$\Gamma_{R1} = 0.001$, $\Gamma_{L2} = 0.001$, $\Gamma_{R2} = 0.1$). The frequency of the vibration, coupling these two states, is $\omega_0 = 1$. When we sweep the voltage, a *peak* in the voltage-current curve is observed when the energy difference $\tilde{\epsilon}_1 - \tilde{\epsilon}_2 \propto eV$ is going through the resonance $\tilde{\epsilon}_1 - \tilde{\epsilon}_2 \approx \omega_0$.

6.4.3 Coupling to a vibrational contiuum: dissipation and renormalization

6.4.3.1 The model Hamiltonian

In the previous section we have dealt with a simple, but nevertheless physically rich, model describing the interaction of an electronic level with some specific vibrational mode confined to the quantum dot. We have seen how to apply in this case the Keldysh non-equilibrium techniques described in Section III within the self-consistent Born and Migdal approximations. The latter are however appropriate for the weak coupling limit to the vibrational degrees of freedom. In the opposite case of strong coupling, different techniques must be applied. For equilibrium problems, unitary transformations combined with variational approaches can be used, in non-equilibrium only recently some attempts were made to deal with the problem. [90]

In this section we will consider the case of a multi-level electronic system in interaction with a bosonic bath [30, 31]. We will use unitary transformation techniques to deal with the problem, but will only focus on the low-bias transport, so that strong non-equilibrium effects can be disregarded. Our interest is to explore how the *qualitative* low-energy properties of the electronic system are modified by the interaction with the bosonic bath. We will see that the existence of a continuum of vibrational excitations (up to some cut-off frequency) dramatically changes the analytic properties of the electronic Green function and may lead in some limiting cases to a qualitative modification of the low-energy electronic spectrum. As a result, the I-V characteristics at low bias may display "metallic" behavior (finite current) even if the isolated electronic system does exhibit a band gap. The model to be discussed below has been motivated by the very exciting electrical transport measurements on short poly(dG)-poly(dC) DNA molecular wires carried out at the group of N. Tao some time ago [127]. Peculiar in these experiments was the large measured currents -up to 150 nA at 0.8 V- at low voltages, which stood in strong contrast to the usually accepted view that DNA should behave as an insulator at low applied bias. Further, a power-law length scaling of the linear conductance with increasing wire length was demonstrated, indicating that long-range charge transport was possible. Since the experiments were carried out in an aqueous solution, the possibility of a solvent-induced modification of the low-energy transport properties of the wire lied at hand, although additional factors like internal vibrations could also play a role.

The proposed model is based on an earlier work [128] and assumes, within a minimal tight-binding picture, that the DNA electronic states can be qualitatively classified into extended (conducting) and localized (non-conducting) states. The former may correspond e.g. to the π-orbital stack of the base pairs, the latter to energetically deeper lying (w.r.t. the frontier orbitals) base-pair states or sugar-phosphate backbone states. A further assumption is that any modification of the conducting states through the environment only takes place through a coupling to the non-conducting set. The tight-binding electronic

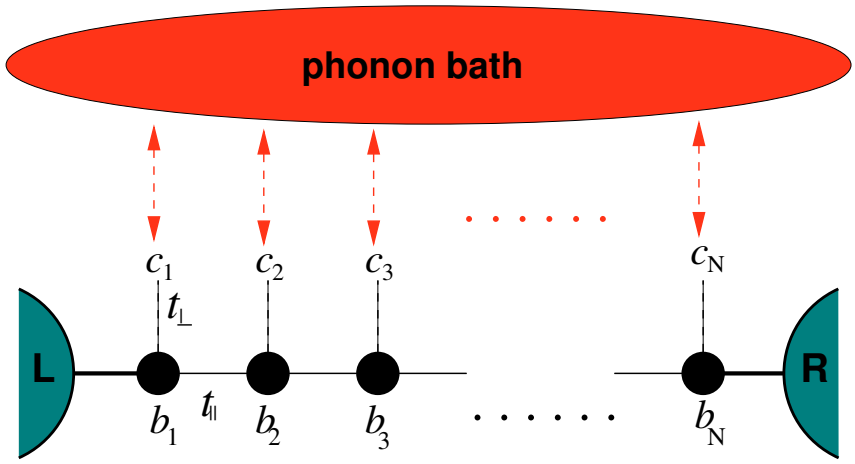


Figure 6.36: Schematic drawing of a DNA molecular wire in contact with a dissipative environment. The central chain (extended states) with N sites is connected to semiinfinite left (L) and right (R) electronic reservoirs. The bath only interacts with the side chain sites (c), which we call for simplicity backbone sites, but which collectively stay for non-conducting, localized electronic states. The Hamiltonian associated with this model is given by Eqs. (6.447), (6.448), and (6.449) in the main text.

Hamiltonian for N sites can then be written as (see also Fig. 6.36):

$$\begin{aligned}\mathcal{H}_{\rm el} &= \epsilon_b \sum_j b_j^\dagger b_j - t_{||} \sum_j \left[b_j^\dagger b_{j+1} + \text{H.c.} \right] + \epsilon \sum_j c_j^\dagger c_j \\ &- t_\perp \sum_j \left[b_j^\dagger c_j + \text{H.c.} \right] = \mathcal{H}_{\rm C} + \mathcal{H}_{\rm b} + \mathcal{H}_{\text{C-b}}. \end{aligned} \tag{6.447}$$

Hereby $\mathcal{H}_{\rm C}$ and $\mathcal{H}_{\rm b}$ are the Hamiltonians of the extended and localized states (called in what follows "backbone" states for simplicity), respectively, and $\mathcal{H}_{\text{C-b}}$ is the coupling between them. $t_{||}$ and $t_\perp$ are hopping integrals along the central chain (extended states) and between the localized states and the central chain, respectively. If not stated otherwise, the on-site energies will be later set equal to zero to simplify the calculations. Notice that this model displays a gap in the electronic spectrum roughly proportional to the transversal coupling $t_\perp$. This can be easely seen by looking at the limit $N \to \infty$ which leads to a periodic system. In this case, the Hamiltonian can be analytically diagonalized and two energy dispersion curves are obtained, which are given by $E_\pm(k) = t_{||}\cos(k) \pm \sqrt{t_\perp^2 + t_{||}^2 \cos^2(k)}$. The direct gap between the two bands is simply $\delta = 2\sqrt{t_\perp^2 + t_{||}^2}$. Since this model further shows electron-hole symmetry, two electronic manifolds (bands in the limit of $N \to \infty$) containing N states each, are symmetrically situated around the Fermi level, which is taken as the zero of energy.

The gap is obviously temperature independent and furthermore it is expected that transport at energies $E < \delta$ will be strongly suppressed due to the absence of electronic states to suport charge propagation. As a result, the linear conductance should display a strong exponential dependence as a function of the chain length N. In view of this behavior, an immediate issue that arises is how stable this electronic structure, *i.e.* two electronic manifolds separated by a gap, is against the coupling to an environment. This is

an issue which reaches farther than the problem of charge transport in DNA wires, since it addresses the interaction of an open quantum mechanical system with a countable number of electronic energy levels to a continuum of states ("universe"). A generic example of such a situation is the measurement process in quantum mechanics. It is well-known that the interaction with complex environments is a source of dissipation and decoherence in quantum mechanical systems. [129] Concerning more specifically the case of DNA (and proteins), there is broad experimental evidence that the molecule dynamics follows the solvent dynamics over a broad temperature range. Especially, conformational changes, low-energy vibrational excitations and the corresponding temperature dependences turn out to be very sensitive to the solvents dynamics. [130] We will thus consider the vibrational degrees of freedom of counterions and hydration shells of the solvent as a dynamical bath able to break the electronic phase coherence and additionally to act as a dissipative environment. We do not consider specific features of the environment but represent it in a generic way by a bosonic bath of M harmonic oscillators. Then, the previous Hamiltonian can be extended to:

$$\mathcal{H}_{\rm W} = \mathcal{H}_{\rm el} + \sum_{\alpha} \Omega_\alpha B_\alpha^\dagger B_\alpha + \sum_{\alpha,j} \lambda_\alpha c_j^\dagger c_j (B_\alpha + B_\alpha^\dagger) = \mathcal{H}_{\rm el} + \mathcal{H}_{\rm B} + \mathcal{H}_{\text{c-B}}, \tag{6.448}$$

where $\mathcal{H}_{\rm B}$ and $\mathcal{H}_{\text{c-B}}$ are the phonon bath Hamiltonian and the (localized) state-bath interaction, respectively. B_α is a bath phonon operator and λ_α denotes the electron-phonon coupling. Note that we assume a local coupling of the bath modes to the electronic density at the side chain. Later on, the thermodynamic limit ($M \to \infty$) in the bath degrees of freedom will be carried out and the corresponding bath spectral density introduced, so that at this stage we do not need to further specify the set of bath frequencies Ω_α and coupling constants λ_α. Obviously, the bath can be assumed to be in thermal equilibrium and be described by a canonical partition function.

To complete the formulation of the model, we have to include the interaction of the system with electronic reservoirs in order to describe charge transport along the same lines as before. We assume, as usual, a tunnel-type Hamiltonian with the form:

$$\begin{aligned} \mathcal{H} &= \mathcal{H}_{\rm W} + \sum_{\mathbf{k}\in \text{L,R},\sigma} \epsilon_{\mathbf{k}\sigma} d_{\mathbf{k}\sigma}^\dagger d_{\mathbf{k}\sigma} + \sum_{\mathbf{k}\in \text{L},\sigma} (V_{\mathbf{k},1}\, d_{\mathbf{k}\sigma}^\dagger\, b_1 + \text{H.c.}) \\ &\quad + \sum_{\mathbf{k}\in R,\sigma} (V_{\mathbf{k},N}\, d_{\mathbf{k}\sigma}^\dagger\, b_N + \text{H.c.}) = \mathcal{H}_{\rm W} + \mathcal{H}_{\text{L/R}} + \mathcal{H}_{\text{L-C}} + \mathcal{H}_{\text{R-C}} \end{aligned} \tag{6.449}$$

The Hamiltonian of Eq. (6.449) is the starting point of our investigation. For a weak charge-bath coupling, a perturbative approach similar to the second order Born approximation, as described in the previous section can be applied. We expect, however, qualitative new effects rather in the opposite limit of strong coupling to the bath. To deal with this problem, a unitary transformation, the Lang-Firsov (LF) transformation, can be performed on the Hamiltonian of Eq. (6.449), which allows to eliminate the linear charge-vibron interaction $\mathcal{H}_{\text{c-B}}$. In the limiting case of an isolated system with a single electron (or hole) this transformation becomes exact and allows for a full decoupling of electronic and vibronic propagators, see e.g. Ref. [71]. In the present case, this transformation is not exact and further approximations have

to be introduced in order to make the problem tractable.

The generator of the LF transformation is given by $S = \sum_{\alpha,j}(\lambda_\alpha/\Omega_\alpha)c_j^\dagger c_j(B_\alpha - B_\alpha^\dagger)$ and $S^\dagger = -S$. In the transformed Hamiltonian $\bar{\mathcal{H}} = \mathrm{e}^{\,S}\mathcal{H}\mathrm{e}^{\,-S}$ the linear coupling to the bath is eliminated. One should notice that in $\bar{\mathcal{H}}$ only the "backbone" part of the Hamiltonian is modified since the conducting state operators b_ℓ as well as the lead operators $d_{\mathbf{k}\sigma}$ are invariant with respect to the above transformation. The new Hamiltonian reads:

$$\begin{aligned}\bar{\mathcal{H}} &= \mathcal{H}_{\mathrm{C}} + \mathcal{H}_{\mathrm{L/R}} + \mathcal{H}_{\mathrm{B}} + \mathcal{H}_{\mathrm{L/R\text{-}C}} \\ &+ (\epsilon - \Delta)\sum_j c_j^\dagger c_j - t_\perp \sum_j \left[b_j^\dagger c_j \mathcal{X} + \mathrm{H.c.}\right], \\ \mathcal{X} &= \exp\left[\sum_\alpha \frac{\lambda_\alpha}{\Omega_\alpha}(B_\alpha - B_\alpha^\dagger)\right], \quad \Delta = \sum_\alpha \frac{\lambda_\alpha^2}{\Omega_\alpha}.\end{aligned} \tag{6.450}$$

As a result of the LF we get a shift of the onsite energies (polaron shift or reorganization energy in electron transfer theory) and a renormalization of both the tunneling and of the transversal coupling Hamiltonian via the bosonic operators $\mathcal{X}$. There is also an additional electron-electron interaction term which we will not be concerned with in the remaining of this section and is thus omitted. Since we are mainly interested in qualitative statements, we will assume the wide-band approximation in the coupling to the electrodes which is equivalent to substituting the electrode self-energies by a purely imaginary constant, i.e. $\Sigma_{\mathrm{L,R}} \approx -\mathrm{i}\,\Gamma_{\mathrm{L,R}}$. We are thus not interested in specific features of the electrode electronic structure.

To further proceed, let us now introduce two kinds of retarded thermal Green functions related to the central chain $G_{j\ell}(t)$ and to the "backbones" $P_{j\ell}(t)$, respectively (taking $\hbar = 1$):

$$\begin{aligned}G_{j\ell}(t,t') &= -i\Theta(t-t')\left\langle\left[b_j(t), b_\ell^\dagger(t')\right]_+\right\rangle, \\ P_{j\ell}(t,t') &= -i\Theta(t-t')\left\langle\left[c_j(t)\mathcal{X}(t), c_\ell^\dagger(t')\mathcal{X}^\dagger(t')\right]_+\right\rangle,\end{aligned} \tag{6.451}$$

where Θ is the Heaviside function. Notice that the P-Green function doe not have a pure electronic character but also contains the bath operators $\mathcal{X}$. For a full out-of-equilibrium calculation, the full Keldysh formalism including lesser- and greater-GF would also be needed. However, as we will briefly show below, the final expression for the electrical current at low applied voltages and for small transversal coupling $t_\perp$ will only include the retarded propagators.

We now use the equation of motion technique (EOM) to obtain an expression for the GF $G_{j\ell}(t)$. We first remark that in the time domain two EOM can be written, depending on which time argument in the double-time GF the time derivative will act. One thus obtains in general:

$$\begin{aligned}\mathrm{i}\,\partial_t G(t,t') &= \left\langle\left[b(t), b^\dagger(t')\right]_+\right\rangle \delta(t-t') + (([b(t), H]\,|b^\dagger(t'))). \\ G(t,t')[-\mathrm{i}\,\partial_{t'}] &= \left\langle\left[b(t), b^\dagger(t')\right]_+\right\rangle \delta(t-t') - ((b(t)|\left[b^\dagger(t'), H\right])).\end{aligned}$$

The EOM for the GF $G_{j\ell}(t)$ reads then in the energy space:

$$\sum_n \left[G_0^{-1}(E)\right]_{\ell n} G_{nj}(E) = \delta_{\ell j} - t_\perp((c_\ell \mathcal{X}|b_j^\dagger)) \tag{6.452}$$

$$\begin{aligned}\left[G_0^{-1}(E)\right]_{\ell n} &= (E-\epsilon_b)\delta_{n\ell} + t_{||}(\delta_{n,\ell+1}+\delta_{n,\ell-1}) - \Sigma_{\rm L}\delta_{\ell 1}\delta_{n1} - \Sigma_{\rm R}\delta_{\ell N}\delta_{nN} \\ \Sigma_{\rm L(R)} &= \sum_{\mathbf{k}\in L(R)} \frac{|V_{\mathbf{k},1(N)}|^2}{E-\epsilon_{\mathbf{k}}+{\rm i}\,0^+} \approx -{\rm i}\,\Gamma_{\rm L,R}\end{aligned}$$

In the next step, EOM for the "right" time argument t' of the GF $Z^{\mathcal{X}}_{\ell j}(t,t')((c_\ell(t)\mathcal{X}(t)|b_j^\dagger(t')))$ can be written. This leads to:

$$\sum_m Z^{\mathcal{X}}_{\ell m}(E)\left[G_0^{-1}(E)\right]_{mj} = -t_\perp((c_\ell\mathcal{X}|c_j^\dagger\mathcal{X}^\dagger)) = -t_\perp P_{\ell j}(E) \tag{6.453}$$

Inserting Eq. (6.453) into Eq. (6.452) we arrive at the matrix equation:

$$\mathbf{G}(E) = \mathbf{G}_0(E) + \mathbf{G}_0(E)\mathbf{\Sigma}_{\rm B}(E)\mathbf{G}_0(E),$$

which can be transformed into a Dyson-like equation when introducing the irreducible part $\mathbf{\Sigma}_{\rm B}(E) = \mathbf{\Sigma}^{\rm irr}_{\rm B}(E) + \mathbf{\Sigma}^{\rm irr}_{\rm B}(E)\mathbf{G}_0(E)\mathbf{\Sigma}^{\rm irr}_{\rm B}(E) + \ldots$:

$$\mathbf{G}(E) = \mathbf{G}_0(E) + \mathbf{G}_0(E)\mathbf{\Sigma}^{\rm irr}_{\rm B}(E)\mathbf{G}(E), \tag{6.454}$$

or equivalently:

$$\begin{aligned}\mathbf{G}^{-1}(E) &= \mathbf{G}_0^{-1}(E) - t_\perp^2\mathbf{P}(E) \\ \mathbf{G}_0^{-1}(E) &= E\mathbf{1} - \mathcal{H}_{\rm C} - \Sigma_{\rm L}(E) - \Sigma_{\rm R}(E).\end{aligned} \tag{6.455}$$

$\mathbf{\Sigma}^{\rm irr}_{\rm B}(E) = t_\perp^2\mathbf{P}(E)$ is the crucial contribution to the GF since it contains the influence of the bosonic bath. Note that $\mathbf{\Sigma}^{\rm irr}_{\rm B}(E)$ includes the transversal hopping $t_\perp$ to all orders, the leading one being $t_\perp^2$.

In the next step, an expression for the electrical current flowing through the system must be derived. Using the results of Sec. 2, we can directly write the following expression:

$$\begin{aligned}I &= \frac{2e}{h}\int dE\,{\rm Tr}(f_{\rm L}(E) - f_{\rm R}(E))\,t(E) \\ &+ t_\perp^2\frac{2e}{h}\int dE\,\left\{{\rm Tr}[\mathbf{\Sigma}_{\rm L}^{>}\mathbf{P}^{<} - \mathbf{\Sigma}_{\rm L}^{<}\mathbf{P}^{>}] - (L\leftrightarrow R)\right\}.\end{aligned} \tag{6.456}$$

The first summand has the same form as Landauer's expression for the current with an effective transmission function $t(E) = {\rm Tr}[\mathbf{G}^\dagger\mathbf{\Gamma}_{\rm R}\mathbf{G}\mathbf{\Gamma}_{\rm R}]$. However, the reader should keep in mind that the GFs appearing in this expression do contain the full dressing by the bosonic bath and hence, $t(E)$ does not

describe elastic transport. The remaining terms contain explicitly contributions from the bath. It can be shown after some transformations that the leading term is proportional to $(t_\perp^2)^2$ so that within a perturbative approach in $t_\perp$ and at low bias it can be approximately neglected. We therefore remian with the exression $I = \frac{2e}{h}\int dE\,\mathrm{Tr}(f_\mathrm{L}(E) - f_\mathrm{R}(E))\,t(E)$ to obtain the current.

To remain consistent with this approximation, the bath selfenergy should also be treated to order $t_\perp^2$, more explicitly:

$$\begin{aligned} P_{\ell j}(t,t') &= ((c_\ell(t)\mathcal{X}(t)|c_j^\dagger(t')\mathcal{X}^\dagger(t'))) \\ &\approx -\mathrm{i}\,\theta(t-t')\left\{\left\langle c_\ell(t)c_j^\dagger(t')\right\rangle\left\langle\mathcal{X}(t)\mathcal{X}^\dagger(t')\right\rangle + \left\langle c_j^\dagger(t')c_\ell(t)\right\rangle\left\langle\mathcal{X}^\dagger(t')\mathcal{X}(t)\right\rangle\right\} \\ &\approx -\mathrm{i}\,\delta_{\ell j}\theta(t-t')\left\{\left\langle c_j(t)c_j^\dagger(t')\right\rangle\left\langle\mathcal{X}(t)\mathcal{X}^\dagger(t')\right\rangle + \left\langle c_j^\dagger(t')c_j(t)\right\rangle\left\langle\mathcal{X}^\dagger(t')\mathcal{X}(t)\right\rangle\right\} \\ &= -\mathrm{i}\,\delta_{\ell j}\theta(t-t')\mathrm{e}^{-\mathrm{i}\,(\epsilon-\Delta)\,t}\left\{(1-f_\mathsf{c})\mathrm{e}^{-\Phi(t)} + f_\mathsf{c}\mathrm{e}^{-\Phi(-t)}\right\}. \end{aligned} \tag{6.457}$$

In the previous expression we have replaced the full averages of the "backbone" operators by their zero order values (free propagators). $\mathrm{e}^{-\Phi(t)} = \left\langle\mathcal{X}(t)\mathcal{X}^\dagger(0)\right\rangle_\mathrm{B}$ is a dynamical bath correlation function to be specified later on. The average $\langle\cdot\rangle_\mathrm{B}$ is performed over the bath degrees of freedom. f_c is the Fermi function at the backbone sites. In what follows we consider the case of empty sites by setting $f_\mathsf{c} = 0$. The Fourier transform $P_{\ell j}(E)$ reads then:

$$P_{\ell j}(E) = -\mathrm{i}\,\delta_{\ell j}\int_0^\infty \mathrm{d}t\,\mathrm{e}^{\mathrm{i}\,(E+\mathrm{i}\,0^+)t}\,\mathrm{e}^{-\mathrm{i}\,(\epsilon-\Delta)\,t}\,\left[(1-f_\mathsf{c})\mathrm{e}^{-\Phi(t)} + f_\mathsf{c}\mathrm{e}^{-\Phi(-t)}\right] \tag{6.458}$$

In order to get closed expressions for the bath thermal averages it is appropriate to introduce a bath spectral density [129] defined by :

$$J(\omega) = \sum_\alpha \lambda_\alpha^2\delta(\omega-\Omega_\alpha) = J_0(\frac{\omega}{\omega_\mathrm{c}})^s\mathrm{e}^{-\omega/\omega_\mathrm{c}}\Theta(\omega), \tag{6.459}$$

where ω_c is a cut-off frequency related to the bath memory time $\tau_\mathrm{c} \sim \omega_\mathrm{c}^{-1}$. It is easy to show that the limit $\omega_\mathrm{c} \to \infty$ corresponds to a Markovian bath, *i.e.* $J(t) \sim J_0\delta(t)$. Using this *Ansatz*, $\Phi(t)$ can be written as:

$$\Phi(t) = \int_0^\infty \mathrm{d}\omega\,\frac{J(\omega)}{\omega^2}\left[1 - \mathrm{e}^{-\mathrm{i}\,\omega t} + 2\frac{1-\cos\omega t}{\mathrm{e}^{\beta\omega}-1}\right]. \tag{6.460}$$

Although the integral can be performed analytically [129], we consider $\Phi(t)$ in some limiting cases where it is easier to work directly with Eq. (6.460).

6.4.3.2 Limiting cases

We use now the results of the foregoing section to discuss the electronic transport properties of our model in some limiting cases for which analytic expressions can be derived. We will discuss the mean-field approximation and the weak-coupling regime in the electron-bath interaction as well as to elaborate on

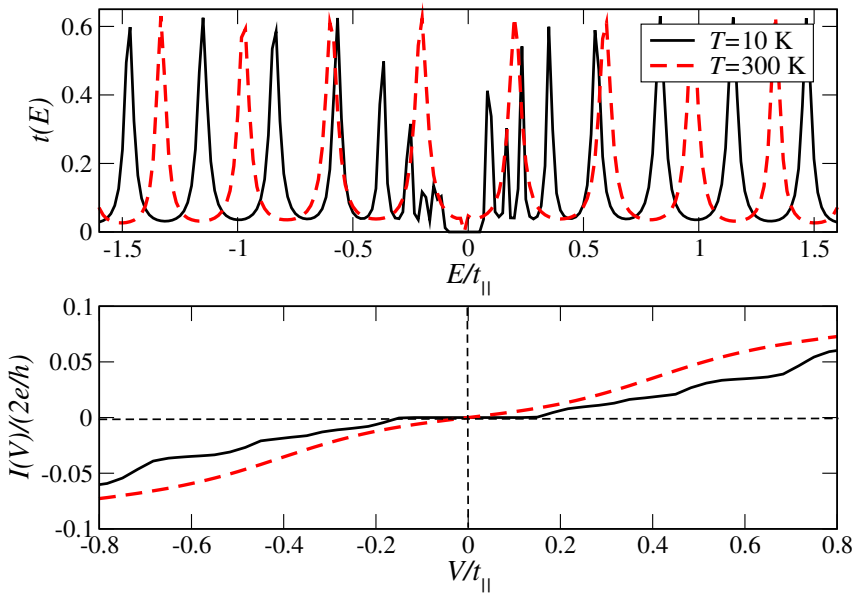


Figure 6.37: (Color online) Electronic transmission and corresponding current in the mean-field approximation for two different temperatures. Parameters: $N = 20, J_0/\omega_c = 0.12, t_\perp/t_{||} = 0.5, \Gamma_{L/R}/t_{||} = 0.5$.

the strong-coupling limit. Farther, the cases of ohmic ($s = 1$) and superohmic ($s = 3$) spectral densities are treated.

(i) Mean-field approximation The mean-field approximation is the simplest one and neglects bath fluctuations contained in $P(E)$. The MFA can be introduced by writing the phonon operator $\mathcal{X}$ as $\langle\mathcal{X}\rangle_B + \delta\mathcal{X}$ in $\mathcal{H}_{C\text{-}c}$ in Eq. (6.450), *i.e.* $\mathcal{H}^{MF}_{C\text{-}b} = -t_\perp \sum_j \left[b_j^\dagger c_j \langle\mathcal{X}\rangle_B + \text{H.c.}\right] + O(\delta\mathcal{X})$. As a result a real, static and temperature dependent term in Eq. (6.455) is found:

$$\mathbf{G}^{-1}(E) = \mathbf{G}_0^{-1}(E) - t_\perp^2 \frac{|\langle\mathcal{X}\rangle_B|^2}{E - \epsilon + \Delta + i0^+}\mathbf{1}, \tag{6.461}$$

where $|\langle\mathcal{X}\rangle_B|^2 = e^{-2\kappa(T)}$ and $\kappa(T)$ is given by:

$$\kappa(T) = \int_0^\infty \frac{d\omega}{\omega^2} J(\omega) \coth\frac{\omega}{2k_B T}. \tag{6.462}$$

The effect of the MF term is thus to scale the bare transversal hopping $t_\perp$ by the exponential temperature dependent factor $e^{-\kappa(T)}$.

In the case of an ohmic bath, $s = 1$, the integrand in $\kappa(T)$ scales as $1/\omega^p$, $p = 1, 2$ and has thus a logarithmic divergence at the lower integration limit. Thus, the MF contribution would vanish. In other words, no gap would exist on this approximation level.

In the superohmic case ($s = 3$) all integrals are regular. One obtains $\Delta = \int d\omega\, \omega^{-1} J(\omega) = \Gamma(s-1)J_0 =$

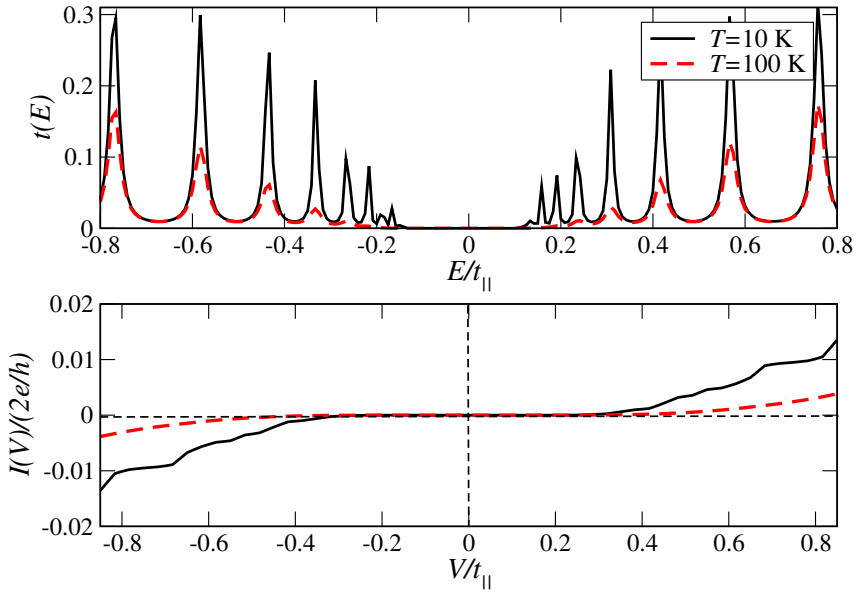


Figure 6.38: Electronic transmission and corresponding current in the weak-coupling limit with ohmic dissipation ($s = 1$) in the bath. Parameters: $N = 20, J_0/\omega_c = 0.2, t_\perp/t_{||} = 0.6, \Gamma_{L/R}/t_{||} = 0.5$

$2J_0$, with $\Gamma(s)$ being the Gamma function and $\kappa(T)$ reads:

$$\kappa(T) = \frac{2J_0}{\omega_c}\left[2\left(\frac{k_B T}{\omega_c}\right)^2 \zeta_H\left(2, \frac{k_B T}{\omega_c}\right) - 1\right]. \tag{6.463}$$

$\zeta_H(s,z) = \sum_{n=0}^{\infty}(n+z)^{-s}$ is the Hurwitz ζ-function, a generalization of the Riemann ζ-function. [131]

It follows from Eq. (17) that $\kappa(T)$ behaves like a constant for low temperatures ($k_B T/\omega_c < 1$), $\kappa(T) \sim J_0/\omega_c$, while it scales linear with T in the high-temperature limit ($k_B T/\omega_c > 1$), $\kappa(T) \sim J_0/\omega_c(1+ 2k_B T/\omega_c)$).

For $J_0 \neq 0$ and at zero temperature the hopping integral is roughly reduced to $t_\perp e^{-\frac{J_0}{\omega_c}}$ which is similar to the renormalization of the hopping in Holstein's polaron model [132], though here it is $t_\perp$ rather than $t_{||}$ the term that is rescaled. At high temperatures $t_\perp$ is further reduced ($\kappa(T) \sim T$) so that the gap in the electronic spectrum finally collapses and the system becomes "metallic", see Fig. 6.37. An appreciable temperature dependence can only be observed in the limit $J_0/\omega_c < 1$; otherwise the gap would collapse already at zero temperature due to the exponential dependence on J_0. We further remark that the MFA is only valid in this regime ($J_0/\omega_c < 1$), since for $J_0/\omega_c \gg 1$ multiphonon processes in the bath, which are not considered in the MFA, become increasingly relevant and thus a neglection of bath fluctuations is not possible.

(ii) Beyond MF: weak-coupling limit As a first step beyond the mean-field approach let's first consider the weak-coupling limit in $\mathbf{P}(E)$. For $J_0/\omega_c < 1$ and not too high temperatures ($k_B T/\omega_c \lesssim 1$) the main contribution to the integral in Eq. (6.460) comes from long times $t \gg \omega_c^{-1}$. 3 With the change

of variables $z = \omega t$, $\Phi(t)$ can be written as:

$$\begin{aligned}\Phi(t) &= J_0\omega_c^{-s}t^{1-s}\int_0^\infty \mathrm{d}z\, z^{s-2}\mathrm{e}^{-\frac{z}{\omega_c t}} \\ &\quad\times\left(1-\mathrm{e}^{-\mathrm{i}\,z}+2\frac{1-\cos z}{\mathrm{e}^{z\frac{\beta\omega_c}{\omega_c t}}-1}\right).\end{aligned} \tag{6.464}$$

As far as $\omega_c t \gg \beta\omega_c$ this can be simplified to:

$$\begin{aligned}\Phi(t) &\approx J_0\omega_c^{-s}t^{1-s}\int_0^\infty \mathrm{d}x\, z^{s-2}\mathrm{e}^{-\frac{z}{\omega_c t}} \\ &\quad\times\left(1-\mathrm{e}^{-\mathrm{i}\,z}+2\frac{\beta\omega_c}{\omega_c t}\frac{1-\cos z}{z}\right).\end{aligned} \tag{6.465}$$

Since in the long-time limit the low-frequency bath modes are giving the most important contribution we may expect some qualitative differences in the ohmic and superohmic regimes. For $s = 1$ we obtain $\Phi(t) \sim \pi\frac{J_0}{\omega_c}\frac{k_B T}{\omega_c}(\omega_c t)$ which leads to ($\Delta(s=1) = J_0$):

$$\mathbf{G}^{-1}(E) = \mathbf{G}_0^{-1}(E) - t_\perp^2\frac{1}{E + J_0 + \mathrm{i}\,\pi\frac{J_0}{\omega_c}k_B T}\mathbf{1}, \tag{6.466}$$

i.e. there is only a pure imaginary contribution from the bath. For the simple case of $N = 1$ (a two-states model) one can easily see that the gap approximately scales as $\sqrt{k_B T}$; thus it grows with increasing temperature. This is shown in Fig. 6.38, where we also see that the intensity of the transmission resonances strongly goes down with increasing temperature. The gap enhancement is induced by the suppression of the transmission peaks of the frontier orbitals, i. e. those closest to the Fermi energy.

For $s = 3$ and $k_B T/\omega_c \lesssim 1$, $\Phi(t)$ takes a nearly temperature independent value proportional to J_0/ω_c. As a result the gap is slightly reduced ($t_\perp \to t_\perp \mathrm{e}^{-J_0/\omega_c}$) but, because of the weak-coupling condition, the effect is rather small.

From this discussion we can conclude that in the weak-coupling limit ohmic dissipation in the bath induces an enhancement of the electronic gap while superohmic dissipation does not appreciably affect it. In the high-temperature limit $k_B T/\omega_c > 1$ a short-time expansion can be performed which yields similar results to those of the strong-coupling limit (see next section), [133] so that we do not need to discuss them here. Note farther that the gap obtained in the weak-coupling limit is an "intrinsic" property of the electronic system; it is only quantitatively modified by the interaction with the bath degrees of freedom. We thus trivially expect a strong exponential dependence of $t(E = E_F)$, typical of virtual tunneling through a gap. Indeed, we find $t(E = E_F) \sim \exp(-\beta L)$ with $\beta \sim 2 - 3\,\mathring{A}^{-1}$.

(iii) Beyond MF: strong coupling limit (SCL) In this section we elaborate on the strong-coupling regime, as defined by the condition $J_0/\omega_c > 1$. In the SCL the main contribution to the time integral in Eq. (6.460) arises from short times. Hence a short-time expansion of $\Phi(t)$ may already give reasonable

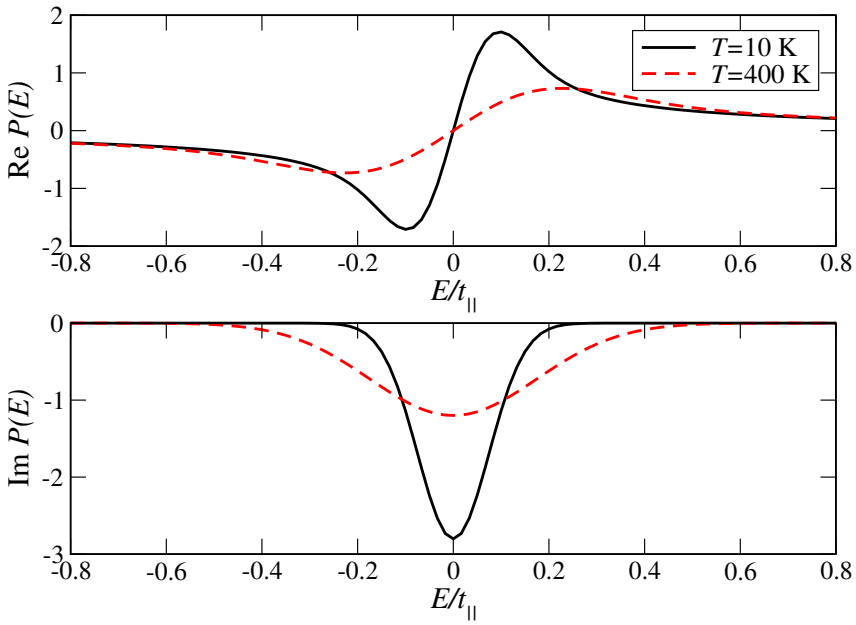


Figure 6.39: Temperature dependence of the real and imaginary parts of $P(E)$ for $N = 20$, $J_0/\omega_c = 10, t_\perp/t_{||} = 0.4, \Gamma_{L/R}/t_{||} = 0.5$. With increasing temperature the slope of the real part near $E = 0$ decreases and the imaginary part broadens and loses intensity. A similar qualitative dependence on J_0 was found (not shown).

results and it allows, additionally, to find an analytical expression for $\mathbf{P}(E)$. At $t \ll \omega_c^{-1}$ we find,

$$\begin{aligned}
\Phi(t) &\approx \mathrm{i}\,\Delta\, t + (\omega_c t)^2\,\kappa_0(T) \qquad (6.467)\\
P_{\ell j}(E) &= -\mathrm{i}\,\delta_{\ell j}\int_0^\infty \mathrm{d}t\,\mathrm{e}^{\,\mathrm{i}\,(E-\epsilon+\mathrm{i}\,0^+)t}\,\mathrm{e}^{\,-(\omega_c t)^2\kappa_0(T)}\\
&= -\mathrm{i}\,\delta_{\ell j}\,\frac{\sqrt{\pi}}{2}\frac{1}{\omega_c\sqrt{\kappa_0(T)}}\,\exp\left(-\frac{(E-\epsilon+\mathrm{i}\,0^+)^2}{4\omega_c^2\kappa_0(T)}\right)\\
&\quad\times\left(1+\mathrm{erf}\left[\frac{\mathrm{i}\,(E-\epsilon+\mathrm{i}\,0^+)}{2\omega_c\sqrt{\kappa_0(T)}}\right]\right),\\
\kappa_0(T) &= \frac{1}{2\omega_c^2}\int_0^\infty \mathrm{d}\omega J(\omega)\coth\frac{\omega}{2k_BT}.
\end{aligned}$$

result. regimes, since spectral for all values of the exponent Before presenting the results for the electronic transmission, it is useful to first consider the dependence of the real and imaginary parts of $\mathbf{P}(E)$ on temperature and on the reduced coupling constant J_0/ω_c. Both functions are shown in Fig. 6.39. We see that around the Fermi level at $E = 0$ the real part is approximately linear, $\mathrm{Re}\,P(E) \sim E$ while the imaginary part shows a Lorentzian-like behavior. The imaginary part loses intensity and becomes broadened with increasing temperature or J_0, while the slope in the real part decreases when k_BT or J_0 are increased. If we neglect for the moment the imaginary part (the dissipative influence of the bath), we can understand the consequences of the real part being nonzero around the Fermi energy, *i.e.* in the gap region of the model of Ref. [128]. The solutions of the non-linear equation $\det|(E - t_\perp^2\,\mathrm{Re}\,P(E))\mathbf{1} - \mathcal{H}_C| = 0$

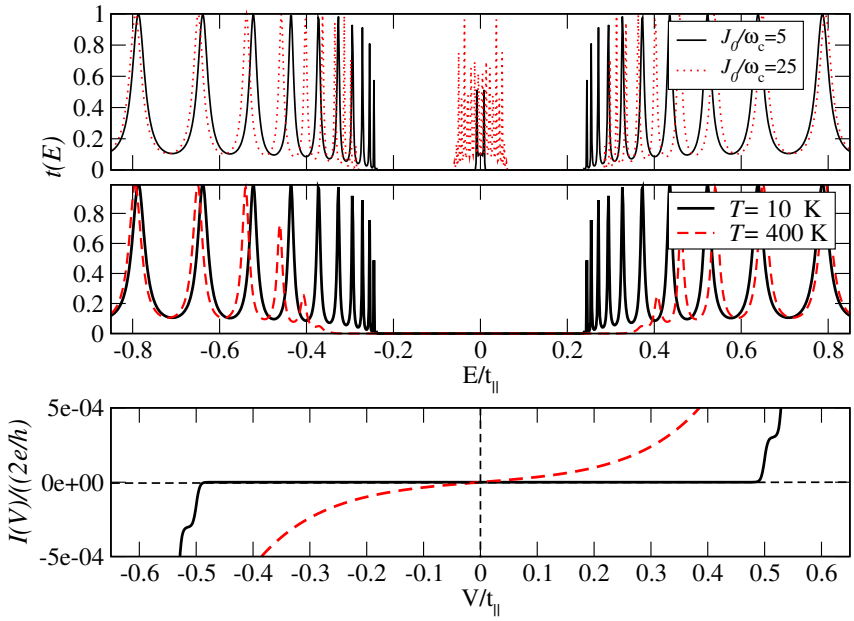


Figure 6.40: Upper panel: $t(E)$ with $\mathrm{Im}\, P(E) = 0$; the intensity of the resonances on the central narrow band is strongly dependent on J_0/ω_c and $k_\mathrm{B}T$ (not shown). Temperature dependence of $t(E)$ with full inclusion of $P(E)$ (middle panel) and corresponding current (lower panel) for $N = 20$, $J_0/\omega_c = 5, t_\perp/t_{||} = 0.5, \Gamma_{\mathrm{L/R}}/t_{||} = 0.2$. The pseudo-gap increases with temperature.

give the new poles of the Green function of the system in presence of the phonon bath. For comparison, the equation determining the eigenstates *without* the bath is simply $\det|(E - t_\perp^2/E)\mathbf{1} - \mathcal{H}_\mathrm{C}| = 0$. It is just the $1/E$ dependence near $E = 0$ that induces the appearance of two electronic bands of states separated by a gap. In our present study, however, $\mathrm{Re}\, P(E \to 0)$ has no singular behavior and additional poles of the Green function may be expected to appear in the low-energy sector. This is indeed the case, as shown in Fig. 6.40 (upper panel). We find a third band of states around the Fermi energy, which we may call a polaronic band because it results from the strong interaction between an electron and the bath modes. The intensity of this band as well as its band width strongly depend on temperature and on J_0. When $k_\mathrm{B}T$ (or J_0) become large enough, these states spread out and eventually merge with the two other side bands. This would result in a transmission spectrum similar of a gapless system.

This picture is nevertheless not complete since the imaginary component of $P(E)$ has been neglected. Its inclusion leads to a dramatic modification of the spectrum, as shown in Fig. 6.40 (middle panel). We now only see two bands separated by a gap which basically resembles the semiconducting-type behavior of the original model. The origin of this gap or rather *pseudo-gap* (see below) is however quite different. It turns out that the imaginary part of $P(E)$, being peaked around $E = 0$, strongly suppresses the transmission resonances belonging to the third band. Additionally, the frontier orbitals on the side bands, *i.e.* orbitals closest to the gap region, are also strongly damped, this effect becoming stronger with increasing temperature ($\mathrm{Im}\, P(E)$ broadens). This latter effect has some similarities with the previously discussed weak-coupling regime. Note, however, that the new electronic manifold around the Fermi energy does not appear in the weak-coupling regime. We further stress that the density of states around

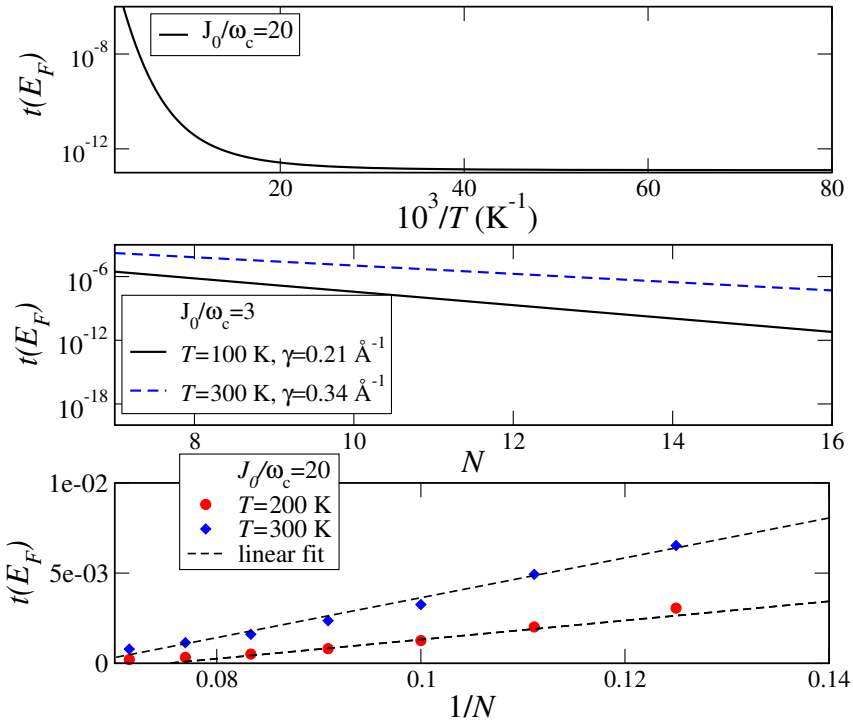


Figure 6.41: Upper panel: Arrhenius plot for $t(E_F)$. Parameters: $N = 20$, $t_{||} = 0.6\,eV$, $t_\perp/t_{||} = 0.2$, $\Gamma_{\mathrm{L/R}}/t_{||} = 0.3$. Middle and lower panels: Length dependence of $t(E_F)$ at different temperatures for two different strengths of the electron-bath coupling J_0/ω_c. The electronic coupling parameters are the same as in the upper panel.

the Fermi level is not exactly zero (hence the term pseudo-gap); the states on the polaronic manifold, although strongly damped, contribute nevertheless with a finite temperature dependent background to the transmission. As a result, with increasing temperature, a crossover from "semiconducting" to "metallic" behavior in the low-voltage region of the I-V characteristics takes place, see Fig. 6.40 (lower panel). The slope in the I-V plot becomes larger when $t_\perp$ is reduced, since the side bands approach each other and the effect of $\mathrm{Im}\,P(E)$ is reinforced.

In Fig. 6.41 (top panel) an Arrhenius plot of the transmission at the Fermi energy is shown, which suggests that activated transport is the physical mechanism for propagation at low energies. Increasing the coupling to the phonon bath makes the suppression of the polaronic band around $E = 0$ less effective ($\mathrm{Im}\,P(E \sim 0)$ decreases) so that the density of states around this energy becomes larger. Hence the absolute value of the transmission will also increase. On the other side, increasing $t_\perp$ leads to a reduction of the transmission at the Fermi level, since the energetic separation of the side bands increases with $t_\perp$.

A controversial issue in transport through molecular wires is the actual length dependence of the electron transfer rates or correspondingly, of the linear conductance. This is specially critical in the case of DNA nanowires [[134, 135, 136]]. Different functional dependences have been found in charge transfer experiments ranging from strong exponential behavior related to superexchange mediated electron transfer [[135]] to algebraic dependences typical of thermal activated hopping [[134, 136]]. As far as transport experiments are concerned, the previously mentioned experiments at the group of N. Tao [127] reported an

algebraic length dependence of the conductance for poly(GC) oligomers in solution. We have investigated the length dependence of $t(E_F)$ and found for the strong dissipative regime $J_0/\omega_c > 1$, an exponential law for energies close to E_F, $t(E_F) \sim \exp(-\gamma N)$. At the first sight, this might be not surprising since a gap in the spectrum does exist. Indeed, in the absence of the bath, *i.e.* with an intrinsic gap, we get decay lengths γ_{coh} of the order of 2 $Å^{-1}$. However, as soon as the interaction with the bath is included, we find values of γ much smaller than expected for virtual tunneling, ranging from 0.15 $Å^{-1}$ to 0.4 $Å^{-1}$. Additionally, γ is strongly dependent on the strength of the electron-bath coupling J_0/ω_c as well as on temperature; γ is reduced when J_0/ω_c or k_BT increases, since in both cases the density of states within the pseudo-gap increases. Remarkably, a further increase of the electron-bath coupling eventually leads to an algebraic length dependence, see lower panel of Fig. 6.41.

The studies presented in this section indicate that the presence of a complex environment, which induces decoherence and dissipation, can dramatically modify the electronic response of a nanowire coupled to electrodes. Electron transport on the low-energy sector of the transmission spectrum is supported by the formation of (virtual) polaronic states. Though strongly damped, these states manifest nonetheless with a finite density of states inside the bandgap and mediate thermally activated transport.

6.5 Conclusions and Perspectives

In this chapter we have reviewed the method of nonequilibrium Green functions and few selected applications to problems related with charge transport at the molecular scale. Hereby we have only focused on minimal model Hamiltonian formulations which build a very appropriate starting point to illustrate the power and range of validity of such techniques. We have showed how this approach can be used to deal with a variety of physical systems, covering both noninteracting and interacting cases. Thus, so different issues as coherent transport, Coulomb blockade phenomena, charge-vibron interaction, coupling to dissipative environments, and the Kondo effect (not addressed in this review) can be in principle treated on the same footing. Specially, the existence of well-developed diagrammatic techniques allows for a systematic treatment of interactions in nanoscale quantum systems. For the sake of space, we did not deal with applications of NEGF techniques to spin-dependent transport [137, 138, 45], a field that has been increasingly attracting the attention of the physical community in the past years due to its potential applications in quantum information theory and quantum computation [139, 140]. For the same reason, the implementation of NEGF into first-principle based approaches was not discussed neither [43, 142, 143, 23, 39, 35, 141]. This is nevertheless a crucial methodological issue, since system-specific and realistic information about molecule-metal contact details, charge transfer effects, modifications of the molecular electronic structure and configuration upon contacting, the electrostatic potential distribution in a device, etc can only be obtained via a full *ab initio* description of transport. For charge transport through noninteracting systems this has been accomplished some years ago by combining NEGF with DFT methods [143, 39, 141, 41]. The inclusion of interactions, however, represents a much stronger challenge and has been mainly carried out, within the self-consistent Born-approximation, for the case of tunneling charges coupling to vibrational excitations in the molecular region [23, 144, 40]. Much harder and till the present not achieved at all is the inclusion of electronic correlation effects -responsible for many-particle effects like Coulomb blockade or the Kondo effect- in a non-equilibrium transport situation. DFT-based techniques, being essentially mean-field theories, cannot deal in a straightforward way with such problems and have to be improved, e.g. within the LDA+U approaches [145]. For the case of equilibrium transport, a generalization of the Landauer formula including correlations has been recently formulated [43] as well as first attempts to go beyond the linear response regime [44]; for strong out-of-equilibrium situations this will be, in our view, one of the most demanding issues that the theoretical "transport" community will be facing in the coming years.

Acknowledgements

We appreciate very useful and illuminating discussions with Andrea Donarini, Pino D'Amico, Dietrich Foerster, Milena Grifoni, Joachim Keller, Abraham Nitzan, Norbert Nemec, Danny Porath, Florian Pump, Klaus Richter, Eugen Starikow, and Juyeon Yi. The authors acknowledge further the collaboration with Miriam del Valle, Marieta Gheorghe, Michael Hartung, Sudeep Mandal, and Soumya Mohapatra, with whom part of the work reviewed here was done. Funding by the EU (grants IST-021285-2 and IST-

029192-2), by the Volkswagen Stiftung (grant I/78 340), by the DFG (grants CU 44/5-1 and CU 44/3-2), by the DFG Graduated School "Nonlinearity and Nonequilibrium in Condensed Matter" GRK638, by the DFG Priority Program "Quantum Transport at the Molecular Scale" SPP1243, and by the DFG Collaborative Research Center " Spin phenomena in reduced dimensions" SFB689 is acknowledged. We also thank the Vielberth Foundation, the Minerva Foundation, the German Exchange Program (DAAD), the Humboldt foundation, and the German-Israeli foundation (GIF) for financial support.

Bibliography

[1] G. Cuniberti, G. Fagas, K. Richter (Eds.), *Introducing Molecular Electronics*, Lecture Notes in Physics, Vol. 680, Springer (2005).

[2] A. Aviram and M. A. Ratner, Chem. Phys. Lett. **29**, 277 (1974).

[3] M. A. Reed, C. Zhou, C. J. Muller, T. P. Burgin, and J. M. Tour, Science **278**, 252 (1997).

[4] R. H. M. Smit, Y. Noat, C. Untiedt, N. D. Lang, M. C. van Hemert, and J. M. van Ruitenbeek, Nature **419**, 906 (2002).

[5] H. Park, J. Park, A. K. L. Lim, E. H. Anderson, A. P. Alivisatos and P. L. McEuen, Nature **407**, 57 (2000).

[6] J. Park, A. N. Pasupathy, J. I. Goldsmith, C. Chang, Y. Yaish, J. R. Petta, M. Rinkoski, J. P. Sethna, H. D. Abruna, P. L. McEuen, and D. Ralph, Nature **417**, 722 (2002).

[7] J. G. Kushmerick, J. Lazorcik, C. H. Patterson, S. Shashidhar, D. S. Seferos, and G. C. Bazan, Nano Lett. **4**, 639 (2004).

[8] J. Chen, M. A. Reed, A. M. Rawlett, and J. M. Tour, Science **286**, 1550 (1999).

[9] W. Liang, M. P. Shores, M. Bockrath, J. R. Long, and H. Park, Nature **417**, 725 (2002).

[10] L. H. Yu and D. Natelson, Nano Lett **4**, 79 (2004); L. H. Yu, Z. K. Keane, J. W. Ciszek, L. Cheng, M. P. Stewart, J. M. Tour, and D. Natelson, Phys. Rev. Lett. **93**, 266802 (2004).

[11] B.-Y. Choi, S.-J. Kahng, S. Kim, H. Kim, H. W. Kim, Y. J. Song, J. Ihm, and Y. Kuk, Phys. Rev. Lett. **96**, 156106 (2006).

[12] M. Martin, M. Lastapis, D. Riedel, G. Dujardin, M. Mamatkulov, L. Stauffer, and Ph. Sonnet, Phys. Rev. Lett. **97**, 216103 (2006).

[13] S.-W. Hla, K.-F. Braun, B. Wassermann, and K.-H. Rieder, Phys. Rev. Lett. **93**, 208302 (2004).

[14] Y. Kim, T. Komeda, and M. Kawai, Phys. Rev. Lett. **89**, 126104 (2002).

[15] J. R. Hahn and W. Ho, Phys. Rev. Lett. **87**, 196102 (2001)

[16] J. Repp, G. Meyer, S. M. Stojkovic, A. Gourdon, and Ch. Joachim, Phys. Rev. Lett. **94**, 026803 (2005).

[17] J.-Ch. Charlier, X. Blase, and S. Roche, Rev. Mod. Phys. **79**, 677 (2007).

[18] K. Keren, R. S. Berman, E. Buchstab, U. Sivan, and E. Braun, Science **302**, 1380 (2002).

[19] M. Mertig, L. Colombi Ciacchi, R. Seidel, W. Pompe, A. De Vita, NanoLetters **2**, 841 (2002); J. Richter, M. Mertig, W. Pompe, I. Mönch, H. K. Schackert, Appl. Phys. Lett. **78**, 536 (2001).

[20] R. Gutierrez, R. Bulla, and G. Cuniberti, in *Modern methods for theoretical physical chemistry of biopolymers*, ed. by S. Tanaka, J. Lewis, and E. Starikow, Elsevier (2006).

[21] R. Gutierrez, D. Porath, and G. Cuniberti, in *Charge Transport in Disordered Solids with Applications in Electronics*, ed. by S. Baranovski, Wiley-CH (2006).

[22] D. Porath, G. Cuniberti, and R. Di Felice, Topics in Current Chemistry **237**, 183 (2004).

[23] T. Frederiksen, M. Brandbyge, N. Lorente, and A.-P. Jauho, Phys. Rev. Lett. **93**, 256601 (2004).

[24] T. Frederiksen, Master's thesis, Technical University of Denmark (2004).

[25] M. Hartung, Master's thesis, University of Regensburg (2004).

[26] M. Cízek, M. Thoss, and W. Domcke, Phys. Rev. B **70**, 125406 (2004).

[27] M. Galperin, M. A. Ratner, and A. Nitzan, J. Phys.: Condens. Matter **19**, 103201 (2007).

[28] D. A. Ryndyk and J. Keller, Phys. Rev. B **71**, 073305 (2005).

[29] D. A. Ryndyk, M. Hartung, and G. Cuniberti, Phys. Rev. B **73**, 045420 (2006).

[30] R. Gutierrez, S. Mandal, and G. Cuniberti, Nano Letters **5**, 1093 (2005).

[31] R. Gutierrez, S. Mandal, and G. Cuniberti, Phys. Rev. B **71**, 235116 (2005).

[32] R. Gutierrez, G. Fagas, K. Richter, F. Grossmann and R. Schmidt, Europhys. Lett. **62**, 90 (2003).

[33] M. del Valle, R. Gutierrez, C. Tejedor, and G. Cuniberti, Nature Nanotech. **2**, 176 (2007).

[34] Th. Frederiksen, M. Paulsson, M. Brandbyge, and A.-P. Jauho, Phys. Rev. B **75**, 205413 (2007)

[35] M. Brandbyge, J.-L. Mozos, P. Ordejon, J. Taylor, and K. Stokbro, Phys. Rev. B **65**, 165401 (2002).

[36] T. Frauenheim, G. Seifert, M. Elstner, Z. Hajnal, G. Jungnickel, D. Porezag, S. Suhai, and R. Scholz, Phys. Stat. Sol. (b) **217**, 41 (2000).

[37] A. D. Carlo, M. Gheorghe, P. Lugli, M. Sternberg, G. Seifert, and T. Frauenheim, Physica B **314**, 86 (2002).

[38] T. Frauenheim, G. Seifert, M. Elstner, T. Niehaus, C. K. M. Amkreutz, M. Sternberg, Z. Hajnal, A. D. Carlo, and S. Suhai, J. Phys.-Condens. Matter **14**, 3015 (2002).

[39] J. Taylor, M. Brandbyge, and K. Stokbro, Phys. Rev. B **68**, 121101 (2003).

[40] A. Pecchia, G. Romano, and A. Di Carlo, Phys. Rev. B **75**, 035401 (2007).

[41] A. Pecchia and A. Di Carlo, Reports on Progress in Physics **67**, 1497 (2004).

[42] S.-H. Ke, H. U. Baranger, and W. Yang, Phys. Rev. B **71**, 113401 (2005).

[43] A. Ferretti, A. Calzolari, R. Di Felice, and F. Manghi, Phys. Rev. B **72**, 125114 (2005).

[44] Giorgos Fagas, Paul Delaney, and James C. Greer, Phys. Rev. B **73**, 241314 (2006).

[45] A. R. Rocha, V. M. Garcia-Suarez, S. Bailey, C. Lambert, J. Ferrer, and S. Sanvito, Phys. Rev. B **73**, 085414 (2006).

[46] M. Gheorghe, R. Gutierrez, N. Ranjan, A. Pecchia, A. Di Carlo, and G. Cuniberti, Europhys. Lett. **71**, 438 (2005).

[47] N. Nemec, D. Tomanek, and G. Cuniberti, Phys. Rev. Lett. **96**, 076802 (2006).

[48] S. Kurth, G. Stefanucci, C.-O. Almbladh, A. Rubio, and E. K. Gross, Phys. Rev. B **72**, 035308 (2005).

[49] H. P. Breuer and F. Petruccione, *The theory of open quantum systems*, (Oxford University Press, Oxford, 2002).

[50] L. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (Benjamin, New York, 1962).

[51] L. V. Keldysh, Zh. Eksp. Teor. Fiz. **47**, 1515 (1964) [Sov. Phys. JETP **20**, 1018 (1965)].

[52] H. S. J. Rammer, Rev. Mod. Phys. **58**, 323 (1986).

[53] H. Haug and A.-P. Jauho, *Quantum Kinetics and Optics of Semiconductors* **123** of *Springer Series in Solid-State Sciences* (Springer, 1996).

[54] Tae-Suk Kim and S. Hershfield, Phys. Rev. B **63**, 245326 (2001)

[55] M. Plihal and J. W. Gadzuk, Phys. Rev. B **63**, 085404 (2001)

[56] R. Aguado and D. C. Langreth, Phys. Rev. Lett. **85**, 1946 (2000)

[57] G. Cuniberti, G. Fagas, and K. Richter, Chem. Phys. **281,** 465 (2002).

[58] M. Paulsson, cond-mat/0210519 (2002).

[59] J. Bardeen, Phys. Rev. Lett. **6,** 57 (1961).

[60] W. A. Harrison, Phys. Rev. **123,** 85 (1961).

[61] M. H. Cohen, L. M. Falicov, and J. C. Phillips, Phys. Rev. Lett. **8,** 316 (1962).

[62] R. E. Prange, Phys. Rev. **131,** 1083 (1963).

[63] C. B. Duke, *Tunneling in solids* (Academic Press, New York, 1969), chapter VII.

[64] C. W. J. Beenakker, Phys. Rev. B **44,** 1646 (1991).

[65] D. V. Averin, A. N. Korotkov, and K. K. Likharev, Phys. Rev. B **44**, 6199 (1991).

[66] H. van Houten, C. W. J. Beenakker, and A. A. M. Staring, in *Single charge tunneling*, Vol. 294 of *NATO ASI Series B*, H. Grabert and M. H. Devoret, eds., (Plenum, 1992), Chap. 5, p. 167.

[67] J. von Delft and D. C. Ralph, Phys. Rep. **345,** 61 (2001).

[68] I. G. Lang and Y. A. Firsov, Sov. Phys. JETP **16**, 1301 (1963).

[69] I. G. Lang and Y. A. Firsov, Sov. Phys. JETP **16**, 1301 (1963).

[70] A. C. Hewson and D. M. Newns, Japan. J. Appl. Phys. **Suppl. 2, Pt. 2**, 121 (1974).

[71] G. Mahan, *Many-Particle Physics* (Plenum, New York, 1990, 2nd ed.).

[72] L. I. Glazman and R. I. Shekhter, Zh. Eksp. Teor. Fiz. **94**, 292 (1988), [Sov. Phys. JETP **67**, 163 (1988)].

[73] N. S. Wingreen, K. W. Jacobsen, and J. W. Wilkins, Phys. Rev. Lett. **61**, 1396–1399 (1988).

[74] N. S. Wingreen, K. W. Jacobsen, and J. W. Wilkins, Phys. Rev. B **40**, 11834 (1989).

[75] M. Jonson, Phys. Rev. B **39**, 5924 (1989).

[76] K. Huang and A. Rhys, Proc. R. Soc. London Ser. A **204**, 406 (1950).

[77] P. Král, Phys. Rev. B **56**, 7293 (1997).

[78] U. Lundin and R. H. McKenzie, Phys. Rev. B **66**, 075303 (2002).

[79] J.-X. Zhu and A. V. Balatsky, Phys. Rev. B **67**, 165326 (2003).

[80] S. Braig and K. Flensberg, Phys. Rev. B **68**, 205324 (2003).

[81] K. Flensberg, Phys. Rev. B **68**, 205323 (2003).

[82] V. Aji, J. E. Moore, and C. M. Varma, cond-mat/0302222 (2003).

[83] A. Mitra, I. Aleiner, and A. J. Millis, Phys. Rev. B **69**, 245302 (2004).

[84] M. Galperin, M. A. Ratner, and A. Nitzan, Nano Lett. **4**, 1605 (2004).

[85] M. Galperin, M. A. Ratner, and A. Nitzan, J. Phys. Chem. **121**, 11965 (2004).

[86] M. Galperin, M. A. Ratner, and A. Nitzan, Nano Lett. **5**, 125 (2005).

[87] J. Koch and F. von Oppen, Phys. Rev. Lett. **94**, 206804 (2005).

[88] J. Koch, M. Semmelhack, F. von Oppen, and A. Nitzan, Phys. Rev. B **73**, 155306 (2006).

[89] J. Koch, F. von Oppen, and A. V. Andreev, Phys. Rev. B **74**, 205438 (2006).

[90] M. Galperin, M. A. Ratner, and A. Nitzan, Phys. Rev. B **73**, 045314 (2006).

[91] P. I. Arseyev and N. S. Maslova, JETP Lett. **82**, 297 (2005).

[92] P. I. Arseyev and N. S. Maslova, JETP Lett. **84**, 93 (2006).

[93] R. Gutiérrez, S. Mohapatra, H. Cohen, D. Porath, and G. Cuniberti, Phys. Rev. B **74**, 235105 (2006).

[94] B. N. J. Persson and A. Baratoff, Phys. Rev. Lett. **59**, 339 (1987).

[95] M. A. Gata and P. R. Antoniewicz, Phys. Rev. B **47**, 13797 (1993).

[96] S. Tikhodeev, M. Natario, K. Makoshi, T. Mii, and H. Ueba, Surf. Sci. **493**, 63 (2001).

[97] T. Mii, S. Tikhodeev, and H. Ueba, Surf. Sci. **502**, 26 (2002).

[98] T. Mii, S. Tikhodeev, and H. Ueba, Phys. Rev. B **68**, 205406 (2003).

[99] S. Tikhodeev and H. Ueba, Phys. Rev. B **70**, 125414 (2004).

[100] S. Datta, W. Tian, S. Hong, R. Reifenberger, J. I. Henderson, and C. P. Kubiak, Phys. Rev. Lett. **79**, 2530 (1997).

[101] T. Rakshit, G.-C. Liang, A. W. Gosh, M. C. Hersam, and S. Datta, Phys. Rev. B **72**, 125305 (2005).

[102] X. H. Qiu, G. V. Nazin, and W. Ho, Phys. Rev. Lett. **92**, 206102 (2004).

[103] S. W. Wu, G. V. Nazin, X. Chen, X. H. Qiu, and W. Ho, Phys. Rev. Lett. **93**, 236802 (2004).

[104] J. Repp, G. Meyer, S. M. Stojković, A. Gourdon, and C. Joachim, Phys. Rev. Lett. **94**, 026803 (2005).

[105] J. Repp, G. Meyer, S. Paavilainen, F. E. Olsson, and M. Persson, Phys. Rev. Lett. **95**, 225503 (2005).

[106] Y. Meir and N. S. Wingreen, Phys. Rev. Lett. **68,** 2512 (1992).

[107] A.-P. Jauho, N. S. Wingreen, and Y. Meir, Phys. Rev. B **50**, 5528 (1994).

[108] E. Bonet, M. M. Deshmukh, and D. C. Ralph, Phys. Rev. B **65,** 045317 (2002).

[109] W. G. van der Wiel, S. D. Franceschi, J. M. Elzerman, T. Fujisawa, S. Tarucha, and L. P. Kouwenhoven, Rev. Mod. Phys. **75**, 1 (2003).

[110] M. H. Hettler, W. Wenzel, M. R. Wegewijs, and H. Schoeller, Phys. Rev. Lett. **90**, 076805 (2003).

[111] B. Muralidharan, A. W. Ghosh, and S. Datta, Phys. Rev. B **73**, 155410 (2006).

[112] J. N. Pedersen, B. Lassen, A. Wacker, and M. H. Hettler (2007).

[113] C. Lacroix, J. Phys. F: Metal Phys. **11**, 2389 (1981).

[114] Y. Meir, N. S. Wingreen, and P. A. Lee, Phys. Rev. Lett. **66**, 3048 (1991).

[115] Y. Meir, N. S. Wingreen, and P. A. Lee, Phys. Rev. Lett. **70**, 2601 (1993).

[116] C. Niu, L.-J. Liu, and T.-H. Lin, Phys. Rev. B **51**, 5130 (1995).

[117] P. Pals and A. Mackinnon, J. Phys.-Condens. Matter **8**, 5401 (1996).

[118] S. Lamba and S. K. Joshi, Phys. Rev. B **62**, 1580 (2000).

[119] J. J. Palacios, L. Liu, and D. Yoshioka, Phys. Rev. B **55**, 15735 (1997).

[120] L. Yi and J.-S. Wang, Phys. Rev. B **66**, 085105 (2002).

[121] C. Niu, D. L. Lin, and T.-H. Lin, J. Phys.-Condens. Matter **11**, 1511 (1999).

[122] R. Swirkowicz, J. Barnas, and M. Wilczynski, Phys. Rev. B **68**, 195318 (2003).

[123] B. R. Bułka and T. Kostyrko, Phys. Rev. B **70**, 205333 (2004).

[124] B. Song, D. Ryndyk, and G. Cuniberti, Phys. Rev. B **76**, 045408 (2007).

[125] M. Elbing, R. Ochs, M. Koentopp, M. Fischer, C. von Hanisch, F. Weigend, F. Evers, H. B. Weber, and M. Mayor, Proc. Natl. Acad. Sci. USA **102**, 8815 (2005).

[126] F. Pump and G. Cuniberti, Surface Science, to appear (2007).

[127] B. Xu, P. Zhang, X. Li, and N. Tao, Nano Letters **4**, 1105 (2004).

[128] G. Cuniberti, L. Craco, D. Porath, and C. Dekker, Phys. Rev. B **65**, 241314(R) (2002).

[129] U. Weiss, *Quantum Dissipative Systems*, vol. **10** of *Series in Modern Condensed Matter Physics* (World Scientific, 1999).

[130] G. Caliscan, D. Mechtani, J. H. Roh, A. Kisliuk, A. P. Sokolov, S. Azzam, M. T. Cicerone, S. Lin-Gibson, and I. Peral, J. Chem. Phys. **121**, 1978 (2004).

[131] I. S. Gradshteyn and I. M. Ryzhik, Academic Press (2000).

[132] T. Holstein, Ann. Phys. N. Y. **8**, 325 (1959).

[133] P. Ao, S. Grundberg, and J. Rammer, Phys. Rev. B **53**, 10042 (1996).

[134] C. R. Treadway, M. G. Hill, and J. K. Barton, Chem. Phys. **281**, 409 (2002).

[135] E. Meggers, M. E. Michel-Beyerle, and B. Giese, J. Am. Chem. Soc. **120**, 12950 (1998).

[136] S. O. Kelley and J. K. Barton, Science **283**, 375 (1999).

[137] W. J. M. Naber, S. Faez, and W. G. van der Wiel, J. Phys. D: Appl. Phys. **40**, R205 (2007)

[138] S. Sanvito and A. R. Rocha, Journal of Computational and Theoretical Nanoscience **3**, 624 (2006).

[139] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).

[140] D. Bouwmeester, A. K. Ekert, and A. Zeilinger, *The Physics of Quantum Information: Quantum Cryptography, Quantum Teleportation, Quantum Computation* (Springer, 2000).

[141] A. Pecchia and A. Di Carlo, *Microscopic description of molecular devices*, in *Proceedings of Fermi School in Physics From Nanostructures to Nanosensing Applications*, A. D'Amico, G. Balestrino and A. Paoletti Eds., Elsevier and SIF (2005).

[142] Y. J. Lee, M. Brandbyge, M. J.'Puska, J. Taylor, K. Stokbro, and R. M. Nieminen, Phys. Rev. B **69**, 125409 (2004)

[143] M. Brandbyge, M. R. Sorensen, and K. W. Jacobsen, Phys. Rev. B **56**, 14956 (1997)

[144] A. Pecchia, A. Di Carlo, A. Gagliardi, S. Sanna, Th. Frauenheim, and R. Gutierrez, Nano Letters **4**, 2109 (2004).

[145] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **44**, 943 (1991).

Chapter 7

Numerical methods for the calculation of shot noise in nanoscale devices.

M. Macucci
Dipartimento di Ingegneria dell'Informazione,
Università di Pisa,
Via Caruso 16, I-56122 Pisa, Italy

7.1 Introduction

Shot noise was first analyzed by Schottky[1], who explained its origin on the basis of the discreteness of charge carriers and the variance associated with the Poisson process that results for independent charges crossing the device. In the case of independent carriers, i.e. device traversal events that are completely uncorrelated, the noise current power spectral density is proportional to the average value of the current ($S_I = 2qI$) through twice the electron charge q or, in general, the charge associated with the specific carrier. This is a direct result of the fact that in a Poisson process the variance equals the average value[1]. Indeed, recent experiments have demonstrated that shot noise in superconductors has a power spectral density $4qI$[2], since in such a case current is carried by Cooper pairs, and that in the fractional quantum Hall regime it is proportional to a fractional charge[3].

Schottky's theorem breaks down in the presence of correlations between carriers, which in general originate form Coulomb interaction or from the Pauli exclusion principle. The action of such correlations leads to a reduction of the shot noise power spectral density if the act is such a way as to make the charge flow more "regular," i.e. they lead to an "anti-bunching" effect, or to an enhancement, when more

than one carrier tend to be "bunched" together. A detailed discussion of such effects can be found in the comprehensive review by Blanter and Büttiker[4], from which the usefulness of shot noise as a probe for correlations in mesoscopic systems and nanoelectronic devices is apparent.

In this chapter we shall discuss approaches for modeling shot noise in nanoscale devices on the basis of the formalism developed by Büttiker and with a detailed discussion of the techniques to be used in the presence of a finite temperature and of finite applied bias. Our main focus will be on the numerical approaches and on their optimization for dealing with structures in which a very large number of transverse modes propagate. A few examples will also be presented, with the calculation of the shot noise power spectral density in diffusive wires and in mesoscopic cavities.

7.2 The Büttiker formalism

While in a classical device shot noise originates substantially from the injection of charges with a thermal velocity distribution from the contacts, so that a continuous transition between thermal and shot noise can be observed as a function of the applied bias and of the ratio of the device length to the mean free path [5], in quantum devices there is a further source of randomness originating from the probabilistic nature of the transmission.

Let us first discuss the case of shot noise in a mesoscopic device in the limit of zero temperature. In this condition there is no thermally generated randomness in the emission from the reservoirs, and, in the case of perfect transmission through the device, shot noise is completely suppressed. This has been verified experimentally for the case of the conductance plateaus of a quantum point contacts [6] and, from a semiclassical point of view, is justified by Thomas and Landauer [7]. Kluhs [8] and Lesovik [9] provided the first insight into the problem of evaluating the low-frequency shot noise current power spectral density for a device with nonunitary transmission. They considered the particular case in which the transmission matrix is diagonal, therefore each $i-$th impinging mode is transmitted into a single corresponding mode at the output, with a transmission probability T_i. In such a case, the shot noise power spectral density is given by

$$S_I = 4\,\frac{q^3}{h}\,|V|\,\sum_i T_i\,(1-T_i)\,, \tag{7.1}$$

where h is Planck's constant and V is the applied voltage. This result can be easily explained on the basis of the analysis of a stochastic process in which multiple attempts are performed (in the specific case multiple attempts to cross the barrier) and can be understood also with relationship to the classical case of partition noise [10].

The issue becomes somewhat more complex in the case of a transmission matrix that is not diagonal, when there is interaction between the different modes and therefore the simple picture described above does not apply any more. The general case is treated by Büttiker in Ref. [11]: starting from the transmission and reflection matrices t and r, the noise current power spectral density can be expressed as

$$S_I = 4\frac{q^2}{h}|qV|\mathrm{Tr}\,\left(r^\dagger r t^\dagger t\right), \tag{7.2}$$

(notice that there is a factor 2 difference with respect to Ref. [11], because here we include spin degeneracy). Expanding the trace, we obtain

$$S_I = 4\frac{q^2}{h}|qV|\sum_i\sum_m\left(\sum_k r^*_{ki}r_{km}\right)\left(\sum_h t^*_{hm}t_{hi}\right). \tag{7.3}$$

Due to the unitarity of the scattering matrix, we could also write

$$S_I = 4\frac{q^2}{h}|qV|\mathrm{Tr}\left[t^\dagger t(I-t^\dagger t)\right], \tag{7.4}$$

and, as a result of the invariance properties of the trace of a matrix for rotations, we get

$$S_I = 4\frac{q^2}{h}|qV|\sum_i T_i(1-T_i), \tag{7.5}$$

where T_i is now the $i-$th eigenvector of the matrix $t^\dagger t$.

In many cases we are interested in the o called Fano factor, i.e. the ratio of the shot noise current power spectral density S_I to that which would be predicted by Schottky's for the given average value of the current ($S_I S = 2qI$). The expression of the average current I can be easily obtained from the Landauer-Büttiker formula for conductance:

$$G = 2\frac{q^2}{h}\sum_{ij}|t_{ij}|^2, \tag{7.6}$$

which, if the eigenvalues T_i of the matrix $t^\dagger t$ have been computed, can also be written as

$$G = 2\frac{q^2}{h}\sum_i T_i. \tag{7.7}$$

Thus the average value of the current reads

$$|I| = G|V| = 2\frac{q^2}{h}|V|\sum_i T_i, \tag{7.8}$$

and thus the expression for the Fano factor γ is

$$\gamma = \frac{S_I}{S_I S} = \frac{4\frac{q^2}{h}|qV|\sum_i T_i(1-T_i)}{4q\frac{q^2}{h}|V|\sum_i T_i} = \frac{\sum_i T_i(1-T_i)}{\sum_i T_i}. \tag{7.9}$$

In the case in which, at least in one section of the device, there are few propagating modes, we must take into account the presence of fluctuations, which are of the order of one conductance quantum (and therefore become quite significant if only a few modes are propagating). While in the presence of many modes such fluctuations are effectively averaged over modes, for devices with low conductance the only averaging that takes place is the one over energy as a result of the presence of a finite temperature.

Particular care must be exercised when reproducing such averaging in simulations. First of all, let us state that in most experiments a single device is measured, and therefore in the experimental data there is no ensemble average: this implies that also in the numerical simulations a single device must be considered and not an ensemble of them with small variations (for example with slightly different defect distributions).

While in many instances, when evaluating conductance, a situation with infinitesimal applied bias is considered and the effect of a finite temperature is consistently included by energy averaging of the transmission eigenvalues, using the derivative of the Fermi function as a weighting factor, in the case of shot noise we are in general in the opposite regime. Indeed, in order to make shot noise prevalent with respect to thermal noise, a DC bias of a least about $10kT/q$ (where k is the Boltzmann constant and T is the absolute temperature) must be applied. Usually this is also the maximum applied bias, to avoid significant heating of the electron gas, which would, in turn, raise the thermal noise. If the applied bias is of the order of $10kT/q$, the Fermi functions expressing the occupancy of the states in the source and drain reservoirs vary from 0 to 1 over an interval much smaller than the chemical potential difference between the reservoirs, thus, for the purposes of our calculation, they can be approximated with step functions. With this approximation, energy averaging becomes very simple, because it is just a uniform averaging over the chemical potential of the source and that of the drain.

An additional note of caution must be added about the order in which averaging of the numerator and the denominator of Eq.(7.9) is performed. The correct procedure consists in first computing the average of the numerator and of the denominator and then taking the ratio of the averages [12]. This is consistent with the experimental procedure, in which the noise current power spectral density and the current are separately measured and averaged. Indeed, a different and incorrect result would be obtained if the averaging were performed after taking the ratio of $S_I S$ and S_I for each energy value.

Summarizing, for the numerical evaluation of the Fano factor for a given device structure it is necessary to compute the transmission matrix t for a number of energy values in the interval between the chemical potential of the source reservoir and that of the drain reservoir, then the eigenvalues T_i of the matrix $t^\dagger t$ are computed for each energy value and are plugged into Eq.(7.9), which is then averaged according to the above discussed procedure. The evaluation of the transmission matrix can be performed with one of the many available numerical techniques, such as recursive Green's functions [13, 14, 15], recursive scattering matrix [16, 17, 18], mode matching [19].

7.3 Calculation of shot noise in diffusive wires

In 1992 Beenakker and Büttiker made the remarkable prediction, based on random matrix theory, that the shot noise current power spectral density in diffusive conductors is suppressed down to 1/3 of the value that would expected from Schottky's theorem. Such a prediction was later experimentally confirmed, as far as disordered metal wires are concerned, by Henny *et al.* [20], while the only experiment involving a semiconductor structure [21] did not lead to really conclusive results, exhibiting a shot noise suppression that depended on the bias voltage applied to the depletion gates used to define the wire

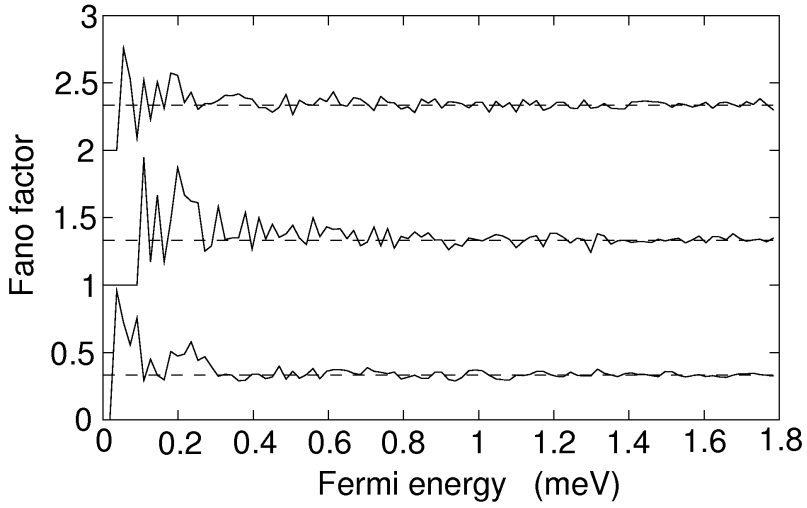


Figure 7.1: Fano factor as a function of the Fermi energy for a disordered quantum wire. The top plot is for a 7.72 μm wire with 600 square 25×25 nm^2 obstacles, the middle plot for a wire 3.85 μm long with 600 square 25×25 nm^2 obstacles and the bottom plot for a a 7.72 μm wire, with 600 square 12.5×12.5 μm obstacles.

in a 2-dimensional electron gas created by modulation doping in a gallium arsenide/aluminum gallium arsenide heterostructure.

A better understanding into the reasons for the difficulty in achieving the diffusive regime in semiconductor nanostructures can be achieved through numerical simulation. A first, simplified approach consists in modeling the scattering by means of randomly located hard-wall obstacles [22] located within a 2-dimensional model quantum wire, and computing the transmission matrix with one of the standard approaches, for example recursive Green's functions (which is simpler to apply in the case of hard walls than other approaches that require a relatively complex check at each interface [19]). Once the transmission matrix is available, the Fano factor can be quickly obtained with the previously described procedure. Results from Ref. [22] are shown in Fig. 7.1, where the Fano factor is reported as a function of the Fermi energy for three wires 5 μm wide: the traces are vertically shifted by one unit to make the graph more readable. The dashed lines represent the 1/3 diffusive limit; the top plot represents the Fano factor for a wire that is 7.72 μm long with 600 square 25×25 nm^2 hard-wall obstacles; the plot in the middle for a for a wire 3.85 μm long with 600 square 25×25 nm^2 hard-wall obstacles; the plot at the bottom for a wire 7.72 μm long with 600 square 12.5×12.5 nm^2 hard-wall obstacles. It is apparent that, as soon as the Fermi energy is large enough to allow a relatively large number of propagating modes (about 50), the Fano factor converges exactly to the diffusive limit of 1/3. In such a situation the inequality $l < L < Nl$ [23] is verified, where l is the elastic mean free path, N is the number of propagating modes, and L is the length of the wire. It turns out that the diffusive regime is achieved as soon as, with $L \gg l$, $Nl \approx 3L$. In a real semiconductor, however, elastic scattering is mainly due to the random potential associated with dopants, which, particularly for a 2-dimensional electron gas obtained by modulation doping,

are not located where transport actually takes place. Therefore the action of dopant atoms cannot be realistically represented with a hard-wall model, and a detailed description would require the solution of the Poisson equation on an extremely fine grid, since the decay of the potential due to each donor is quite sharp. Such a grid would be prohibitively large for problems of practical interest, since a domain of a few square microns should be covered with a step size of 1-2 nm. An approximate approach has been developed [24], based on an initial solution of the Poisson equation computed assuming that the charge density due to the donors is uniform, followed by the addition of the contribution of the discrete donors (whose positions are generated with a uniform random distribution), which is computed with a semi-analytical model that takes into account screening from the carriers in the 2DEG. In order to avoid double counting of the average component, the contribution of the donors is added to the solution of the Poisson equation only after removing the average value.

The transmission matrix of the device and, therefore, the shot noise behavior are then computed with the recursive Green's function technique, adopting a discretization fine enough to resolve the details of the potential fluctuations. The case of a quantum wire defined in GaAs/AlGaAs heterostructure by means of depletion gates 1 μm long and with a 400 nm gap[24] yields the results reported in Fig. 7.2(a): the Fano factor varies as a function of the voltage applied to the gates used to define the quantum wire, but does not stabilize at 1/3, i.e. in a diffusive regime. This result is consistent with the existing experimental data by Liefrink *et al.* [21], in which a similar dependence of the Fano factor on gate voltage is observed. If the calculation is repeated increasing the length of the wire to 3 μm [Fig. 7.2(b)] or 5 μm [Fig. 7.2(c)], the Fano factor rises, but no definite plateau at 1/3 appears. These data can be interpreted as evidence that, as a result of the limited amplitude of the potential fluctuations induced by the donors with respect to the Fermi energy, a diffusive transport condition cannot be easily achieved in a semiconductor nanostructure. Mesoscopic cavities (ofter referred to as "chaotic cavities" are relatively large (a few microns by a few microns) structures delimited, however by narrow constrictions with a width of the order of a few tens of nanometers. A sketch of a mesoscopic cavity is represented in Fig. 7.3: if the constriction are narrow enough, the region inside the cavity behaves as a quasi-reservoir, which leads to quite a peculiar behavior in terms of conductance and noise. In a seminal paper from 1993 Jalabert *et al.* [25] predicted, on the basis of random matrix theory, that a symmetric cavity would exhibit a suppression by a factor 1/4 with respect to full shot noise. Such a result was later also justified in semiclassical terms [26] and found experimental confirmation [27]. Although the basic effect of shot noise suppression in a mesoscopic cavity is well established, there has been much debate about its origin, with several authors that were attributing it to the classically chaotic shape of the cavity or even to scattering off a disordered potential. In this respect, numerical simulations have played an important role, allowing the detailed investigation of cavities with various shapes. In particular, a rectangular (and thereby classically regular cavity) has been shown [12] to yield the very same Fano factor behavior as predicted for cavities with a chaotic shape. In particular, the dependence of the Fano factor on the number of transverse modes propagating through the apertures has been computed for a rectangular cavity 8 μm wide and 5 μm long, exhibiting a behavior [12] that is in good agreement with experimental data [28]. Results are presented in Fig. 7.4, where circles indicate the numerical results and squares correspond to the experimental data: while for few propagating modes

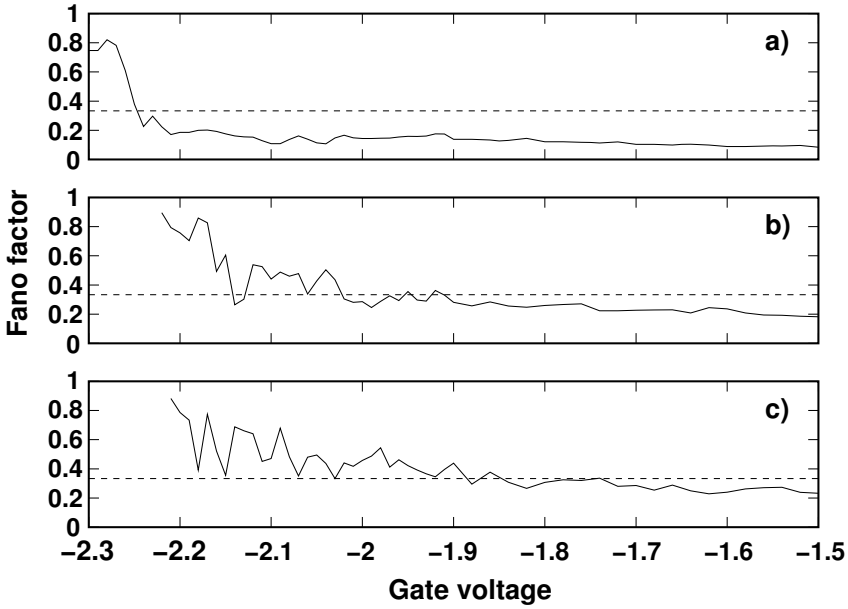


Figure 7.2: Fano factor in a semiconductor quantum wire as a function of the voltage applied to the depletion gates defining it. The top plot is for a wire defined by 1 μm long gates with a gap of 400 nm, the middle and the bottom plots are for a wires defined by gates 3 μm and 5 μm long, respectively.

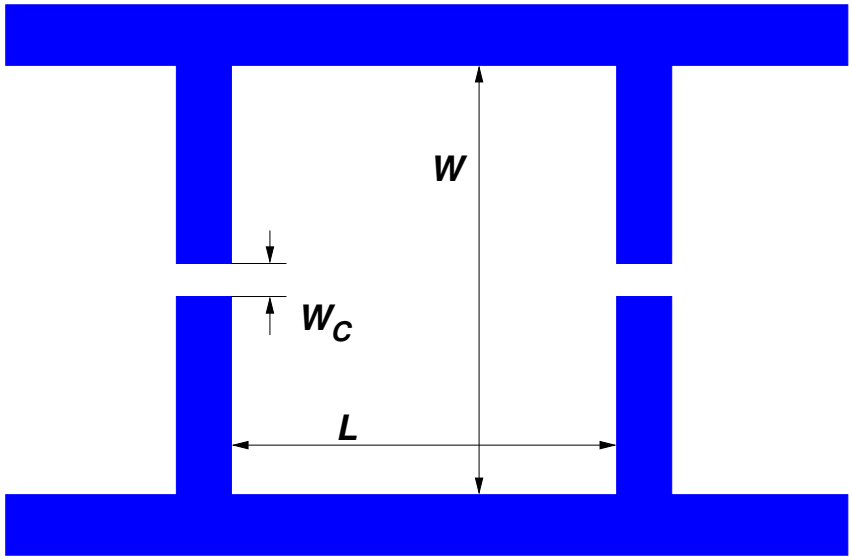


Figure 7.3: Sketch of a mesoscopic cavity, defined by constriction with a width W_C.

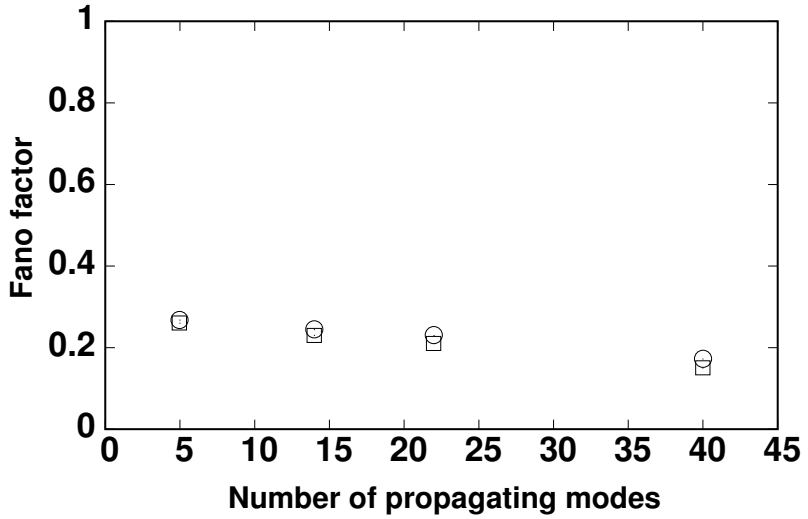


Figure 7.4: Fano factor in a symmetric mesoscopic cavity as a function of the number of modes propagating in the constrictions: the circles indicate the numerical results and the squares represent the experimental values.

the Fano factor is about 1/4, as the constrictions are widened, the Fano factor decreases, finally vanishing when the width of the constrictions equals that of the cavity (as expected, since in this case we have a quantum wire with integer transmission). Analogous results have been obtained for a stadium cavity (a classically chaotic shape): the actual origin of the observed noise behavior lies in the diffraction of electron waves at the apertures. Indeed, in semiclassical terms diffraction is equivalent to generating many different trajectories from a single incoming one, which leads to chaotic behavior. Thus, there is no need for classical chaotic properties, since the origin of the chaotic motion is purely quantum, being a result of diffraction. Further confirmation of the irrelevance of the cavity shape has been obtained by Rotter *et al.* [29], who have compared the Fano factor and the transmission eigenvalue distribution [30] for a few different cavity shapes, such as a stadium, a semicircle, and a rectangle, without finding any significant difference.

Another relevant issue on which modeling has provided a useful input is represented by the effect of an orthogonal magnetic field on the Fano factor for a mesoscopic cavity. Experiments [28] have shown a decrease of the Fano factor with increasing magnetic field and initial explanations of the effect were based on a very simple semiclassical model attributing the shot noise suppression to a reduced portion of the cavity which is explored in the presence of a magnetic field, as a consequence of gradual edge state formation. Consequently, they obtain that the relevant parameter would be represented by the ratio of the cyclotron radius to the cavity size. However, numerical simulations performed with a recursive scattering matrix technique [12] for rectangular cavities of different sizes or with different aperture widths have clearly shown that the relevant quantity is the ratio of the cyclotron radius to the aperture width. If such a ratio is large, electrons will be repeatedly scattered by the constrictions and a behavior similar to that in the absence of a magnetic field will be observed; if, instead, the cyclotron radius will become

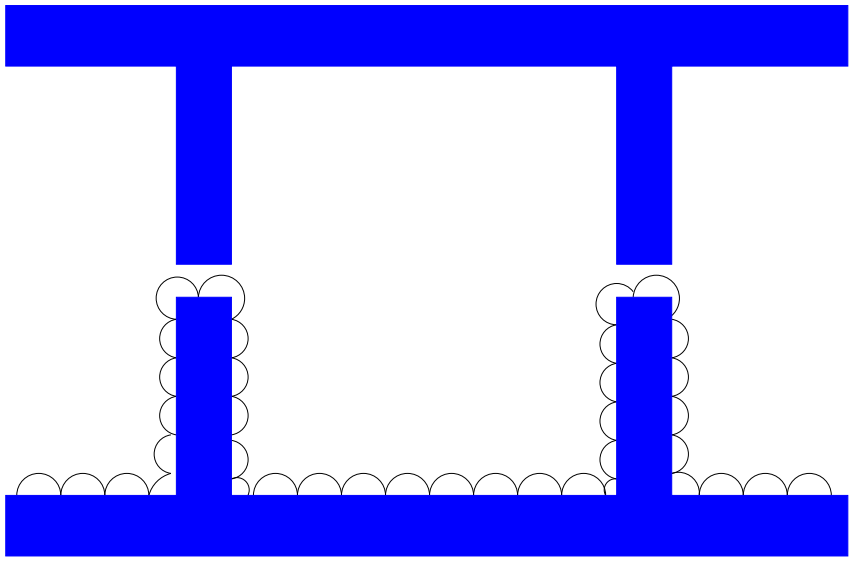


Figure 7.5: Semiclassical representation of edge state propagation regime, in which, as a result of the cyclotron diameter being less than the constriction width, almost no scattering occurs.

comparable to the aperture width or smaller, it will be more and more likely that electrons will follow skipping orbits along the walls without undergoing diffraction (Fig. 7.5). As we get channels with unitary transmission probability, shot noise will drop to zero, as expected. This is a typical example in which collection of experimental data on a variety of structures is very time consuming and it is not possible to have a direct control on all experimental parameters. On the other hand, the analytical models contain significant approximations and are based on assumptions that are not fully justified. Therefore the availability of numerical models becomes essential for the validation of the approximate analytical approaches. In Fig. 7.6(a) we report the dependence of the Fano factor on the orthogonal magnetic field, for a few values of the constriction widths. It is possible to notice that the experimental results, indicated with diamonds, are rather close to the curve for 60 nm apertures. In Fig. 7.6(b) we report the same data with a different representation: we plot the Fano factor as a function of the ratio of the cyclotron diameter D_C (twice the cyclotron radius) to the constriction width W_C. It is apparent that the different curves, corresponding to different constriction widths, are superimposed, which confirms the interpretation based on the interplay between the cyclotron radius and the constriction width. Such an interpretation is further reinforced by another result from numerical calculations: while the semianalytical model of Ref. [28] would imply that, in the presence of a magnetic field, there is a strong dependence of the Fano factor on the length (and also the width) of the cavity, numeric results do not exhibit any such dependence, supporting the conclusion that the noise behavior depends just on the ratio of the cyclotron radius to the constriction width. In Fig. 7.7 we report the Fano factor as a function of the cavity length and width, for given values of the magnetic field: it is apparent that there is no significant dependence on either quantity. Therefore the techniques for the calculation of the low-frequency shot noise power spectral density that we have described in this chapter can be fruitfully applied to a variety of problems in

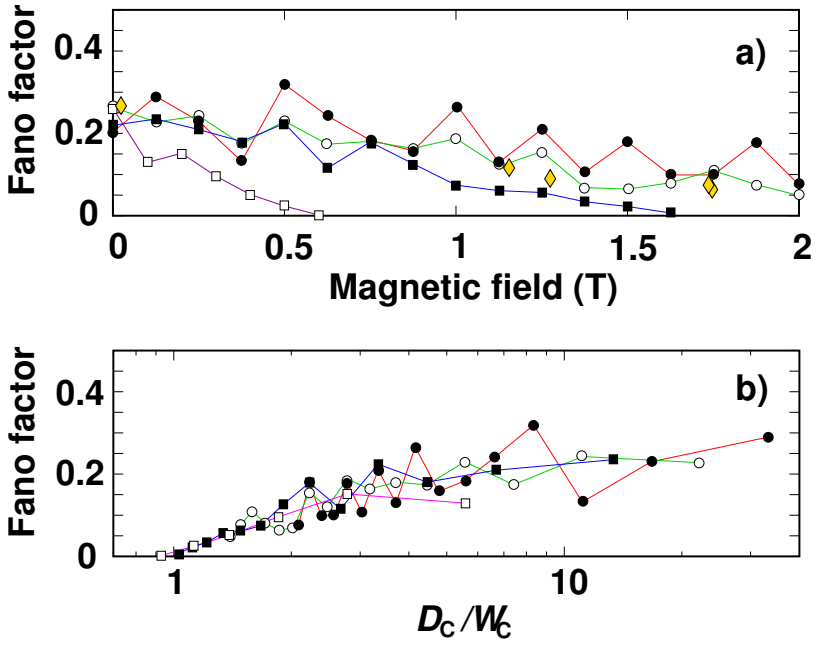


Figure 7.6: a) Fano factor for cavities with 40 nm (solid circles), 60 nm (empty circles), 100 nm (solid squares), 300 nm (empty squares) constrictions as a function of the applied perpendicular magnetic field (diamonds correspond to available experimental data); b) Same data as in a), but represented as a function of the ratio of the cyclotron diameter to the constriction width.

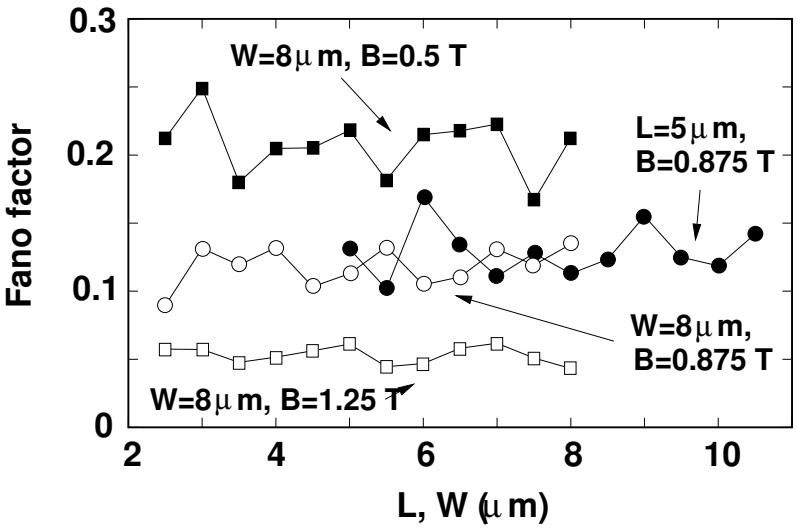


Figure 7.7: Fano factor as a function of the cavity length (width), for several choices of the width (length) and of the perpendicular magnetic field.

mesoscopic physics, providing an efficient tool for the exploration of the parameter space, an exploration that would be extremely difficult and time-consuming to achieve with a direct experimental investigation. The examples we have discussed demonstrate the good predictive capabilities of numerical models for shot noise and how they can be applied to the interpretation of experimental results as well as to the validation of theoretical conjectures.

Bibliography

[1] W. Schottky, Ann. d. Physik **57**, 541 (1918).

[2] X. Jehl, M. Sanquer, R. Calemczuk, D. Mailly, Nature **405**, 50 (2000).

[3] T. G. Griffiths, E. Comforti, M. Heiblum, A. Stern, V. Umansky, Phys. Rev. Lett. **85**, 3918 (2000).

[4] Ya. M. Blanter and m. Büttiker, Phys. Rep. **336**, 1 (2000).

[5] C. J. Stanton, J. W. Wilkins, Physica **134B**, 255 (1985).

[6] A. Kumar, L. Saminadayar, D. C. Glattli, Y. Jin, and B. Etienne, Phys. Rev. Lett. **76**, 2778 (1996).

[7] T. Martin and R. Landauer, Phys. Rev. B **45**, 1742 (1992).

[8] V. A. Khlus, Sov. Phys. JETP **66**, 1243 (1987).

[9] G. B. Lesovik, JETP Lett. **49**, 592 (1989).

[10] A. van der Ziel, *Noise in Solid State Devices and Circuits* (Wiley-Interscience, 1986), p. 18.

[11] M. Büttiker, Phys. Rev. Lett. **65**, 2901 (1990).

[12] P. Marconcini, M. Macucci, G. Iannaccone, B. Pellegrini and G. Marola, Europhys. Lett. **73**, 574 (2006).

[13] D. J. Thouless, S. Kirkpatrick, J. Phys. C **14**, 235 (1981).

[14] F. Sols, M. Macucci, U. Ravaioli, Karl Hess, J. Appl. Phys. **66**, 3892 (1989).

[15] M. Macucci, A. Galick, U. Ravaioli, Phys. Rev. B **52**, 5210 (1995).

[16] R. L. Schult, H. W. Wyld, and D. G. Ravenhall, Phys. Rev. B **41**, 12760 (1990).

[17] S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge, 1995), p. 117.

[18] P. Marconcini, M. Macucci, Journal of Computational Electronics **2**, 387 (2003).

[19] M. Macucci and Karl Hess, Phys. Rev. B **46**, 15364 (1992).

[20] M. Henny, S. Oberholzer, C. Strunk, and C. Schönenberger, **59**, 2871 (1999).

[21] F. Liefrink, J. I. Dijkhuis, M. J. M. de Jong, L. W. Molenkamp, and H. van Houten, Phys. Rev. B **49**, 14066 (1994).

[22] M. Macucci, G. Iannaccone, G. Basso, and B. Pellegrini, Phys. Rev. B **67**, 115339 (2003).

[23] C. W. J. Beenakker and M. Büttiker, Phys. Rev. B **46**, 1889 (1992).

[24] L. Bonci, G. Fiori, M. Macucci, G. Iannaccone, S. Roddaro, P. Pingue, V. Piazza, M. Cecchini and F. Beltram, Physica E **19**, 107 (2003).

[25] R. A. Jalabert, J. L. Pichard, C. W. J. Beenakker, Europhys. Lett. **27**, 255 (1994).

[26] Ya. M. Blanter, E. V. Sukhorukov, Phys. Rev. Lett. **84**, 1280 (2000).

[27] S. Oberholzer, E. V. Sukhorukov, C. Strunk, C. Schönenberger, T. Heinzel, M. Holland, Phys. Rev. Lett. **86**, 2114 (2001).

[28] S. Oberholzer, E. V. Sukhorukov, C. Schönenberger, Nature **415**, 765 (2002).

[29] F. Aigner, S. Rotter, J. Burgdörfer, Phys. Rev. Lett. **94**, 216801 (2005).

[30] S. Rotter, F. Aigner, J. Burgdörfer, Phys. Rev. B **75**, 125312 (2007).

Chapter 8

Models of gated circular Quantum Dots

J. Martorell
Dept. d'Estructura i Constituents de la Materia, Facultat Física,
University of Barcelona, Barcelona 08028, Spain
D. W. L. Sprung
Department of Physics and Astronomy, McMaster University
Hamilton, Ontario L8S 4M1 Canada

Abstract. For gated circular quantum dots, we show that a suitably defined *two-dimensional* Thomas-Fermi model is in excellent agreement with more elaborate Poisson-Schrödinger calculations. Further we show that using the Shikin density profile one can derive accurate semi-analytic approximations for the solutions of the Thomas-Fermi equation.

8.1 Introduction

Quantum dots can be formed by applying a negative potential to metallic gates deposited on the surface of a GaAs heterojunction. When the applied potential is sufficiently negative, a 3D dot forms near the plane of the GaAs/AlGaAs junction. In this paper we describe a sequence of models for the properties of such dots. For simplicity we will restrict our discussion to dots with axial symmetry, formed by a gate that covers the entire heterojunction surface except for a circular hole. Although usually the gates cover only a portion of the exposed surface, our assumed gate geometry is close to experimental conditions, because it can be checked (from the expressions given below, eqs. 8.1-8.3), that by suitably readjusting the applied voltage, the confining potential of our gate can be made similar to that of an annular gate. We describe the equilibrium properties of this system of electrons, using conventional methods that we have

previously applied to linear split-gate devices [1, 2]. We model the electrons in the dot in the envelope function approximation, so that the hamiltonian of the many electron system will include three kinds of terms: *i*) the kinetic energy of each electron, estimated under the assumption that in the semiconductor their effective mass, m^*, is constant; *ii*) the external electrostatic potentials due to the gate and the donor layer, to be written in eqs. 8.3 and 8.4 below, and *iii*) the inter-electronic potential written in eq. 8.5. In addition we assume that the dot is in thermodynamic equilibrium with the electrons trapped in the surface states in the ungated part of the surface, so that the Fermi level is the same at the surface and in the dot, being pinned by the surface states.

We first consider the Hartree approximation to the dot ground state, and determine it by iterative solution of the 3D Poisson and Schrödinger equations. This model requires a sizeable numerical effort and does not easily connect the predicted properties of the dot to the physical parameters of the device. Therefore, to have a better understanding of the physics of the dot, we introduce successive approximations which lead to a much simpler model: that of a strictly two dimensional dot described in the Thomas-Fermi approximation. We study the quantitative effect of these approximations and show that a good qualitative understanding of the properties of the dot can still be achieved with this much simpler model. By introducing suitable expansions and a parabolic approximation for the confining potential we derive analytic results for the radial charge distribution and for the inter-electronic potential. In zeroth order these turn out to be very similar to those of Shikin *et al.* [3, 4] for a schematic classical model of a dot. We can therefore extend the simple analytic approximations found by these authors and show how they apply to a more realistic model for the dot.

8.2 Microscopic models

The heterojunction that we take as reference is the one already studied by us for the case of a linear split-gate device. It is similar to that originally introduced by Laux *et al.* [5], [6], [7] for these devices and consists of successive planar layers of a) a GaAs substrate, that we assume to be free of acceptors and behaving as infinitely thick, b) a spacer layer of undoped AlGaAs, of thickness s, c) a donor layer of AlGaAs, of thickness d and uniformly n-doped with a donor concentration ρ_d, that we assume throughout this paper to be fully ionized; and d) an undoped cap layer of GaAs of thickness c. The metallic gate is deposited on top of that. We choose the center of coordinates at the center of the circular hole in the gate, with the z axis orthogonal to the surface, and denote by $\vec{r}$, $\vec{s}$ distances to that axis in a $z = const.$ plane or distances between points in that same plane. The numerical values for the geometrical parameters introduced above, and for the dielectric constants, effective masses, etc, are the same as in ref.[1].

Following the methods of [1], we construct the total electrostatic potential acting on the electrons in the conduction band by adding the contributions from a) the gates, b) the ionized donor layer, and c) the electrons in the dot. The potential due to a metallic gate with a circular hole can be easily derived

from expressions given in Sect. 3.6 of [8] for a point located at a distance $\vec{r}$ from the z axis:

$$eV_g(\vec{r},z) = \frac{1}{2\pi}\int d\vec{s}\; eV_g(\vec{s},0)\; \frac{|z|}{(z^2+|\vec{r}-\vec{s}|^2)^{3/2}} \tag{8.1}$$

and using the boundary condition:

$$eV_g(\vec{s},0) = eV_g\;\Theta(s-S_0)\;, \tag{8.2}$$

(Dirichlet b.c. at the exposed surface, consistent with the assumed Fermi level pinning), one finds for $z>0$:

$$\begin{aligned} eV_g(r,z) &= eV_g\frac{2}{\pi}z\int_{S_0}^{\infty} s\;ds\;E\left(\frac{4rs}{z^2+(r+s)^2}\right) \\ &\quad \frac{1}{z^2+(r-s)^2}\frac{1}{(z^2+(r+s)^2)^{1/2}}\;. \end{aligned} \tag{8.3}$$

Here $E(k^2)$ is the complete elliptic integral of the second kind as defined in [9]. To this one must add the electrostatic potential due to the donor layer, including the corresponding mirror term which maintains the boundary condition at the surface. As in [1], in the substrate and spacer layers this contribution reduces to a constant additive term:

$$eV_d(z\geq z_2) = -\frac{e^2}{\varepsilon}\rho_d\;d(c+\frac{d}{2})\quad . \tag{8.4}$$

Finally the third contribution is the potential of the Coulomb interaction between the electrons in the dot. Including the mirror terms it is:

$$\begin{aligned} eV_e(\vec{r},z;\vec{r}',z') &= \frac{e^2}{\varepsilon_r}\left[\frac{1}{\sqrt{|\vec{r}-\vec{r}'|^2+(z-z')^2}}\right. \\ &- \left.\frac{1}{\sqrt{|\vec{r}-\vec{r}'|^2+(z+z')^2}}\right] \end{aligned} \tag{8.5}$$

Appropriate band offsets in the AlGaAs layers are also added. For brevity we omit them in the expressions written below. Choosing the Fermi level as the origin of energies and denoting by eV_s the binding energy of the surface states with respect to the conduction band edge, the latter will be located at:

$$eV_{g+d}(r,z) = eV_s + eV_g(r,z) + eV_d(z)\;. \tag{8.6}$$

We write the hamiltonian for N electrons in the dot as:

$$\begin{aligned} H &= \sum_{i=1}^{N}\left(\frac{\hat{p}_i^2}{2m^*}+eV_{g+d}(r_i,z_i)\right) \\ &\quad + \sum_{i<j=1}^{N} eV_e(\vec{r}_i,z_i;\vec{r}_j,z_j)\;, \end{aligned} \tag{8.7}$$

which we will solve in the Hartree approximation.

8.2.1 3D Hartree approximation

We take a trial wave function for the many electron system of the form:

$$\Psi(\vec{r}_1, z_1; ....; \vec{r}_N, z_N) = \prod_{i=1}^{N} \psi_i(\vec{r}_i, z_i) \tag{8.8}$$

and minimize the trial energy under constraints that guarantee orthonormality of the ψ_i. The value of N is chosen to guarantee equilibrium with the surface states, i.e, so that all single particle states of negative energy are occupied. Imposing the variational condition one finds:

$$\begin{aligned} \left[\frac{\hat{p}_i^2}{2m^*} + eV_{g+d}(r_i, z_i) \quad + \quad U_H(\vec{r}_i, z_i)\right]\psi_i &= \varepsilon_i \psi_i(\vec{r}_i, z_i) \\ E_T \quad = \quad \sum_{i=1}^{N} \varepsilon_i - \frac{1}{2} &< \Psi | U_H | \Psi > \end{aligned} \tag{8.9}$$

where the Hartree potential is given by:

$$U_H(\vec{r}, z) = \int d\vec{r}' dz' \; eV_e(\vec{r}, z; \vec{r}', z') \; \rho(\vec{r}', z') \tag{8.10}$$

For an even number of electrons there is a solution with axial symmetry. For it,

$$\rho(\vec{r}, z) = \rho(r, z) \tag{8.11}$$

and:

$$U_H(\vec{r}, z) = U_H(r, z) \; . \tag{8.12}$$

Therefore the solutions of eq. 8.9 take the form:

$$\psi_{nl}(\vec{r}, z) = \frac{u_{nl}(r, z)}{\sqrt{r}} \; \frac{e^{il\phi}}{2\pi} \; , \tag{8.13}$$

where the index i corresponds to the quantum numbers n, ℓ, and we have omitted for simplicity the spin component of the wavefunction. Furthermore, eq. 8.9 becomes a two-dimensional Schrödinger equation:

$$\begin{aligned} -\frac{\hbar^2}{2m^*}\left(\frac{\partial^2}{\partial r^2} + \frac{\partial^2}{\partial z^2}\right) u_{nl}(r, z) + \\ \left(eV_{g+d}(r, z) + U_H(r, z) + \frac{\hbar^2}{2m^*}\frac{l^2 - \frac{1}{4}}{r^2}\right) u_{nl} \quad = \\ \varepsilon_{nl} u_{nl}(r, z) \quad . \end{aligned} \tag{8.14}$$

8.2.2 Factorization ansatz

Guided by the success of a similar *ansatz* in the description of the electron wavefunctions in linear wires we employ it here for dots. This will be the first step towards providing quantitative support to the conventional models that treat this system of electrons as a strictly two-dimensional confined gas.

We start by introducing the following approximation to the two potential terms in the above Schrödinger equation, eq. 8.9:

$$\begin{aligned} eV_{g+d}(r,z) &+ U_H(\vec{r},z) = \\ &= eV_{g+d}(0,z) + e\Delta V_{g+d}(r,z) + U_H(\vec{r},z) \\ &\simeq eV_{g+d}(0,z) + \overline{\Delta U}_c(r) + \overline{U}_H(\vec{r}) \end{aligned} \tag{8.15}$$

where the external confining, $\overline{\Delta U}_c$, and two-dimensional Hartree, $\bar{U}_H$, potentials are defined as:

$$\begin{aligned} \overline{\Delta U}_c(r) &= \int dz A^2(z)(eV_{g+d}(r,z) - eV_{g+d}(0,z)) \\ \overline{U}_H(\vec{r}) &= \int dz\ A^2(z) U_H(\vec{r},z) \end{aligned} \tag{8.16}$$

with a weight function $A^2(z)$ to be specified below. Substituting these approximations in the Schrödinger equation, 8.9, have:

$$\left[\frac{\hat{p}^2}{2m^*} + eV_{g+d}(0,z) + \overline{\Delta U}_c(r) + \overline{U}_H(\vec{r})\right]\psi_i^{(a)} = \varepsilon_i^{(a)}\psi_i^{(a)} \tag{8.17}$$

whose solutions are separable:

$$\psi_i^{(a)}(\vec{r},z) = A(z)\phi_i(\vec{r})\ , \tag{8.18}$$

with separately normalized $A(z)$ and ϕ_i, which satisfy corresponding Schrödinger equations with eigenvalues E_z and e_i:

$$\left[\frac{\hat{p}_z^2}{2m^*} + eV_{g+d}(0,z)\right]A(z) = E_z A(z) \tag{8.19}$$

$$\left[\frac{\hat{p}_\perp^2}{2m^*} + \overline{\Delta U}_c(r) + \overline{U}_H(\vec{r})\right]\phi_i(\vec{r}) = e_i\phi_i(\vec{r}) \tag{8.20}$$

Note that we have chosen the previously unspecified weight function $A(z)$ to be the common longitudinal component of the wavefunctions in 8.18. This guarantees consistency, within this approximation, with the 3D expressions for expectation values of the potential energy. We then have:

$$\begin{aligned} \varepsilon_i^{(a)} &= E_z + e_i \\ E_T^{(a)} &= NE_z\ + \sum_{i=1}^{N} e_i\ - \frac{1}{2}\sum_{i=1}^{n} <\phi_i|\overline{U}_H|\phi_i> \\ &= NE_z + \sum_{i=1}^{n} <\phi_i\ |\frac{\hat{p}_\perp^2}{2m^*} + \overline{\Delta U}_c + \frac{1}{2}\overline{U}_H\ |\phi_i> \end{aligned} \tag{8.21}$$

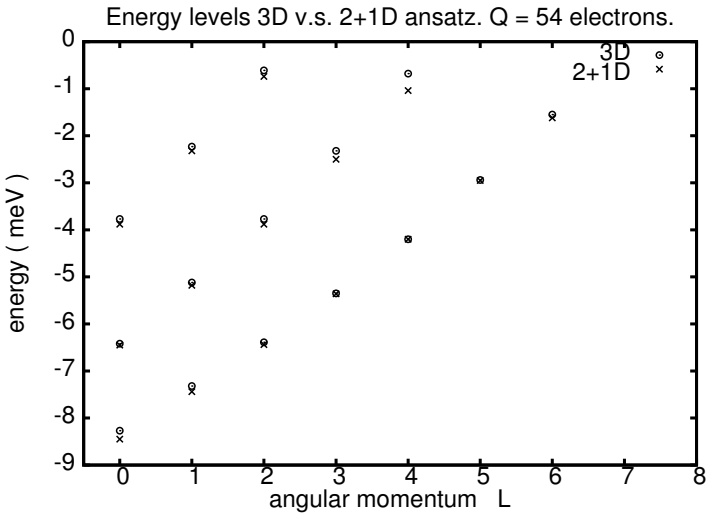


Figure 8.1: Energies of single electron levels. $Q = 54$, corresponding to $V_0 = -1.27$ eV and $S_0 = 210$ nm.

The numerical process starts by constructing $eV_{g+d}(0, z)$ and solving the Schrödinger eq. for $A(z)$. After that the weight function is fixed and one has to solve self-consistently the two-dimensional Hartree problem, constructing the corresponding potentials from eqs. 8.16, and solving the Schrödinger equation 8.20 for the $\phi_i(\vec{r})$, with $i = 1, ...N$.

8.2.3 Results and discussion

For definiteness we present results for a dot containing 54 electrons, which is large enough that the Hartree approximation should be fairly reliable. In addition the semiclassical models to be described later are also expected to work for a large number of particles. In figure 8.1 we show the electron level energies as a function of angular momentum. (Remember that in our convention the Fermi level is located at zero energy.) One can see three families of levels ($n = 0, 1$ and 2) ranging from $\ell = 0$ to $\ell = 6$. Given the apparent arbitrariness of the factorization ansatz it is surprising that the agreement between the exact (labelled 3 D) and the approximate (labelled 2+1 D) energies is so good. For later discussion, we note also that $E_z = -0.104$ eV for this configuration, so that the effective Fermi level for the 2D levels e_i is located at $E_F^{(2D)} = -E_z$, rather than at zero energy.

In figure 8.2 we show the electron density $\rho(r, z)$ of the Hartree approximation. The dot thickness along the z axis is fairly large, of the order of ten nanometers, and in this direction the variation of $\rho(r, z)$ is rather smooth. Radially, the density shows oscillations due to filling of the various levels. This is examined in further detail in figure 8.3 where we have plotted:

$$\sigma_H(r) \equiv \int_{-\infty}^{\infty} \rho(r, z)\ dz. \tag{8.22}$$

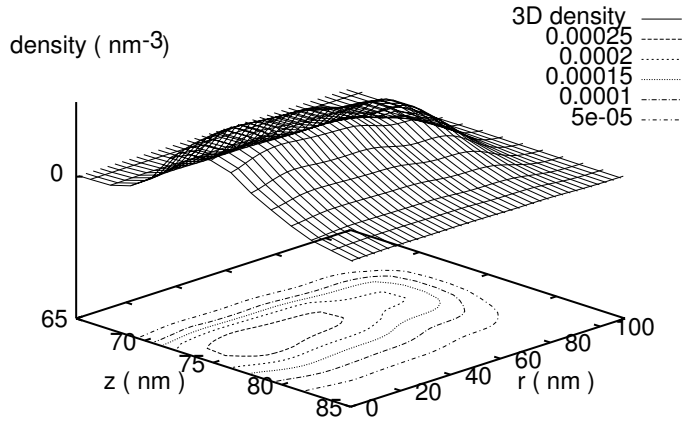


Figure 8.2: Electron density $\rho(r, z)$ computed with the 3D model. $Q = 54$, corresponding to $V_0 = -1.27$ eV and $S_0 = 210$ nm.

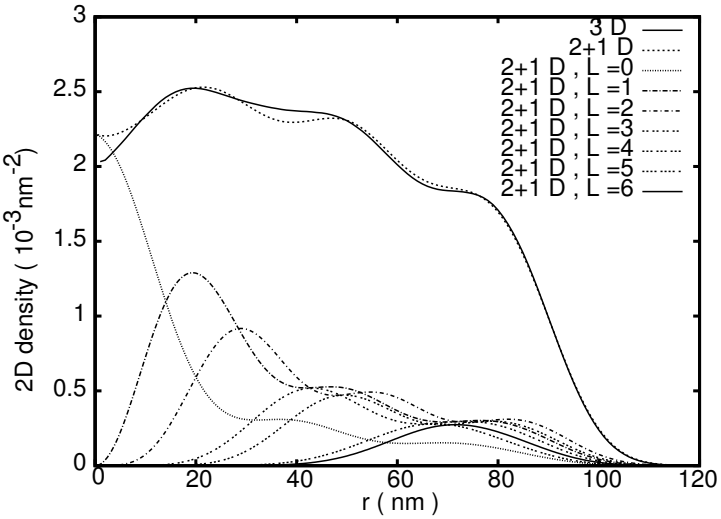


Figure 8.3: 2D electron densities $\sigma(r, z)$ computed with the 3D and the 2+1 D models. $Q = 54$, corresponding to $V_0 = -1.27$ eV and $S_0 = 210$ nm.

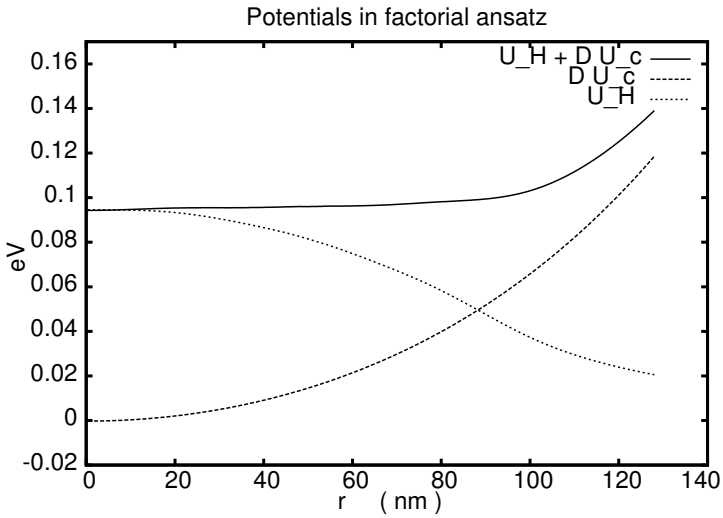


Figure 8.4: 2D potentials $\Delta\bar{U}_c(r)$, $\bar{U}_h(r)$ and their sum . $Q = 54$, corresponding to $V_0 = -1.27$ eV and $S_0 = 210$ nm.

and the corresponding quantity in the factorized approximation, which then reduces to

$$\sigma_H^{(a)}(r) = \sum_i |\phi_i(r)|^2 \ . \tag{8.23}$$

Again, the agreement between the Hartree (3D) and the factorized densities is very good. We also show the individual contributions of levels with different ℓ. Clearly, the inner oscillations in $\sigma_H(r)$ are a quantal effect due to the contributions from different angular momenta being peaked at different r, as expected. At the surface practically all ℓ contribute, giving a smooth decrease that cannot be attributed to any single state or subset of them.

Figure 8.4 shows the Hartree and confinement potentials. One sees that the confinement potential is nearly parabolic, and that the Hartree potential is also nearly so for the values of r corresponding to the area occupied by the dot ($r \leq 100$ nm.) The sum of the two is practically constant in this range of r. This will be later used to introduce the classical models of Shikin *et al.* Finally, by performing the corresponding integrals one finds: $E_{kinetic} = 0.189$ eV, $E_{conf} = 1.338$ eV, $E_{coul} = 1.949$ eV, which add up to a total of $E_{tot} = 3.476$ eV. Note the smallness of the kinetic energy.

Figure 8.4 has the clue to understanding why the factorization ansatz is so accurate: the total potential seen by the electrons is almost flat inside the dot and rises sharply outside it. A similar plot of the potential obtained from the exact "3D" calculation, $eV_{g+d}(r,z) + U_H(r,z)$ shows a substantial z dependence, but again it is almost independent of r inside the dot. As is well known, in the limit of cylindrical infinitely sharp walls, and a z dependent potential inside, the solutions of the single electron Schödinger equation are trivially factorable and in the x, y plane they are Bessel functions. Therefore, to the extent that the rise of the potential outside the dot region is sufficiently sharp, we can expect the factorization ansatz to

be a good approximation. Similarly, for a rectangular dot of sides w_x x w_y and infinitely sharp walls, with a potential inside that depends only on z, the exact solutions of the single electron Schrödinger equation are factorable in the form $\psi_{n,m}(x,y,z) = \mathcal{N}\sin(n\pi x/w_x)\sin(m\pi y/w_y)\ A(z)$. Thus we can expect the factorization ansatz to be reliable for dots that *i)* contain a large number of electrons, so that the total potential inside the region occupied by the dot is fairly independent of x and y, and *ii)* the potential has a sharp rise at the boundaries of the dot. Later, when we discuss the classical models of dots, we will return to the reasons why the total 2D potential shown in figure 8.4 has to be almost constant in the dot region.

8.3 Semiclassical approximations: Thomas-Fermi

When the number of electrons in the dot is rather large, a Thomas-Fermi model would seem to be appropriate if one is more interested in the global properties of the system, like its charging energy and capacitance, rather than in the detail of individual electron energies and wavefunctions. Besides requiring much less numerical effort (than solving the Schrödinger equations for all the ϕ_i), this simpler model will allow us to obtain several accurate analytic approximations that clarify the role of the different physical parameters of the system.

8.3.1 The 2D Thomas-Fermi model

We will now replace the two-dimensional density constructed from the solutions of eq. 8.20 by its circular symmetric Thomas-Fermi approximation, $\sigma_H(r) \to \sigma_{TF}(r)$. By definition, it is determined from a variational condition on a suitably simplified expression for the total energy. Starting from eq. 8.21 we write:

$$E_T = NE_z + E_k + E_c + E_e \tag{8.24}$$

with:

$$\begin{aligned} N &= \int \sigma_{TF}(r)\ d\vec{r} \\ E_k &= \frac{\hbar^2}{2m^*}\pi \int \sigma_{TF}^2(r)\ d\vec{r} \\ E_c &= \int \overline{\Delta U}_c(r)\ \sigma_{TF}(r)\ d\vec{r} \\ E_e &= \frac{e^2}{2\varepsilon_r}\int (\mathcal{V}(|\vec{r}-\vec{r}'|) + \mathcal{V}_m(|\vec{r}-\vec{r}'|))\,\sigma_{TF}(r) \\ &\quad .\ \ \sigma_{TF}(r')\ d\vec{r}\ d\vec{r}'\ , \end{aligned} \tag{8.25}$$

where we have used the well known expression for the kinetic energy of a 2D Thomas-Fermi gas, and

$$\frac{\varepsilon_r}{e^2}\mathcal{V}(r) \equiv \int\int A^2(z)A^2(z')\frac{dz\ dz'}{\sqrt{(z-z')^2+r^2}}\ , \tag{8.26}$$

and a similar expression for the mirror term. Defining now:

$$E(\mu) = E_T - \mu N \tag{8.27}$$

with μ a Lagrange multiplier to fix the number of electrons, the variational condition leads to:

$$\mu = E_z + \frac{\hbar^2}{m^*}\pi\sigma_{TF}(r) + \overline{\Delta U}_c(r) + \overline{U}_{H,TF}(r) \tag{8.28}$$

where:

$$\overline{U}_{H,TF}(r) = \frac{e^2}{\varepsilon_r}\int \left(\mathcal{V}(|\vec{r}-\vec{r}'|) + \mathcal{V}_m(|\vec{r}-\vec{r}'|)\right)\sigma_{TF}(r')\ d\vec{r}' \ , \tag{8.29}$$

This pair of equations defines the iterative Thomas-Fermi approach: the process is started by solving the Schrödinger eq. 8.19 for $A(z)$ to determine E_z. Then $\overline{\Delta U}_c$ given in eq. 8.16, and $\mathcal{V}$, $\mathcal{V}_m$ are computed. The choice previously made for the origin of energies fixes $\mu = 0$. By iteratively solving eqs. 8.28 and 8.29 we determine σ_{TF} (and N), starting from an appropriate initial guess.

Remarks:

i) note that we do take into account the longitudinal energy E_z in a realistic way, and therefore the 2D Fermi level of the dot is not 0 but $-E_z$. If we had chosen it to be zero the results would be completely different.

ii) in a Thomas-Fermi model the number of electrons is non-integer in general; it varies continuously as a function of e.g.: the radius of the gate S_0, or the gate potential V_g.

8.3.2 A strictly two dimensional model

As a further simplification that allows one to make the connection between the above Thomas-Fermi model and the phenomenological models where the electron gas is assumed to be strictly two-dimensional, we now study the case where the $A(z)$ used as input in the previous model is replaced by a delta function:

$$A(z) \simeq \delta(z - \bar{z})\,. \tag{8.30}$$

Then, writing $\sigma_{2D}(r)$ for the corresponding Thomas-Fermi density, the expressions in eq. 8.25 simplify:

$$\begin{aligned}
N &= \int \sigma_{2D}(r)\ d\vec{r} \\
E_k &= \frac{\hbar^2}{2m^*}\pi\int \sigma_{2D}^2(r)\ d\vec{r} \\
E_c &= \int \overline{\Delta U}_c(r)\ \sigma_{2D}(r)\ d\vec{r} \\
E_e &= \frac{e^2}{2\varepsilon_r}\int\left(\frac{1}{|\vec{r}-\vec{r}'|} - \frac{1}{\sqrt{|\vec{r}-\vec{r}'|^2+4\bar{z}^2}}\right)\sigma_{2D}(r) \\
&\quad .\quad \sigma_{2D}(r')\ d\vec{r}\ d\vec{r}' \ ,
\end{aligned} \tag{8.31}$$

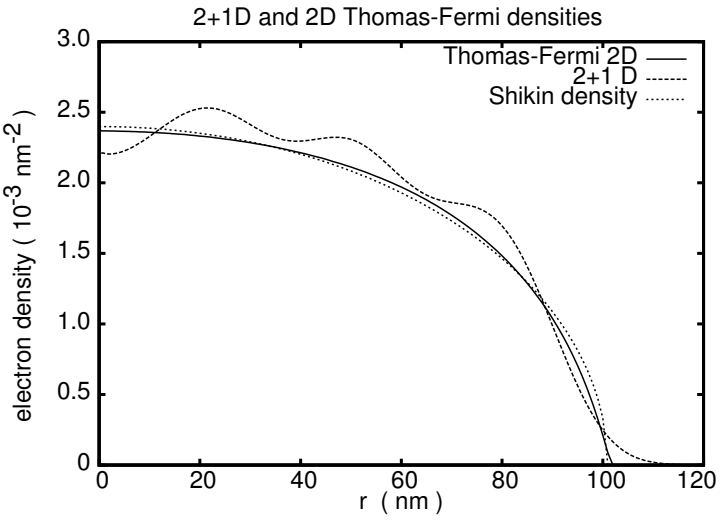


Figure 8.5: Densities predicted by 2D Thomas-Fermi and the Shikin model compared to the 2+1 D given by the factorization ansatz.

and similarly in eq. 8.29 the $\mathcal{V}$'s are replaced by unsmoothed direct plus mirror terms. With these alterations, one has only to repeat the same iterative process to determine the corresponding solutions for σ_{2D}.

Results Note that the value of $\bar{z}$ in eq. 8.30 has not been yet specified. One way to do that is to compare $A(z)$ to the Airy solution for a linear potential, which leads to:

$$\bar{z} = z_m - \frac{\xi_0' - \xi_0}{\alpha} - \frac{2}{3}\frac{\xi_0}{\alpha} \ , \tag{8.32}$$

where z_m is the location of the maximum of $A(z)$, $\xi_0 = -2.33810$ and $\xi_0' = -1.01879$ are the locations of the first zero and the first maximum of the Airy function, and $\alpha = (2m^*\gamma/\hbar^2)^{1/3}$, with γ the slope of the potential $eV_{g+d}(0,z)$ at $z = z_m$.
Starting from the expression for $\Delta\bar{U}_c$ in eq. 8.16 and approximating again $A(z)$ by an Airy function, we arrive at:

$$\begin{aligned} \overline{\Delta U}_c(r) &= (eV_{g+d}(r,z_3) - eV_{g+d}(0,z_3)) \\ &\cdot \quad \frac{\bar{z}}{z_3}\sqrt{\frac{S_0^2 + z_3^2}{S_0^2 + \bar{z}^2}} \ , \end{aligned} \tag{8.33}$$

where z_3 is the location of the plane separating the spacer and the substrate.
Results: For the same gate parameters $V_g = -1.27$ V, $S_0 = 210$ nm we then find $Q \simeq 51$ electrons instead of the 54 of the 3D calculation. This is rather satisfactory given the simplicity of the Thomas-Fermi approximation and the further simplification introduced by eq. 8.30. In figure 8.5 we compare the

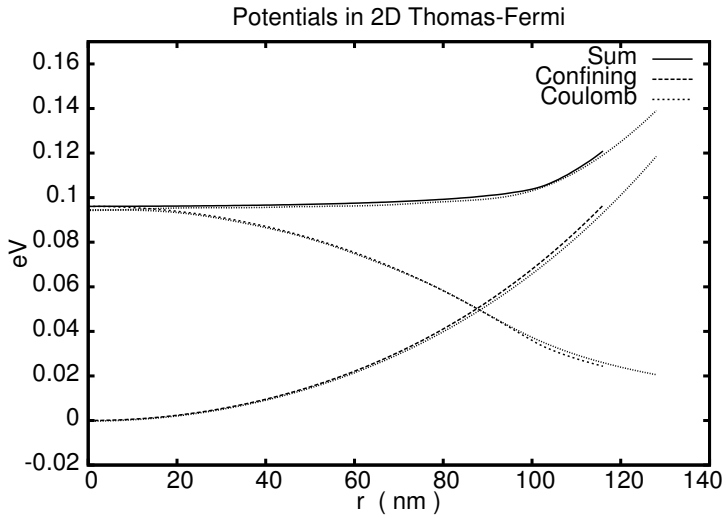


Figure 8.6: Potentials by 2D Thomas-Fermi compared to the 2+1 D given by the factorization ansatz.

predicted $\sigma_{2D}(r)$ to the one from the factorized ansatz (2+1 D). As can be seen the overall extension and density of the dot are fairly well predicted by Thomas-Fermi, but the quantal oscillations are missing and this affects not only the inner density profile of the dot but also the surface. Note that by adjusting $\bar{z}$ we could increase the number of electrons from 51 to 54 and raise the Thomas-Fermi profile to bring it into better quantitative agreement with the 2+1 D density, though the oscillations would still be missing.
Figure 8.6 shows the confinement and Coulomb fields, and their sum. In the same figure the corresponding fields of the factorization ansatz are shown as dotted lines: they are practically indistinguishable from Thomas-Fermi. In that sense Thomas-Fermi is a particularly good approximation [2]. These fields could then be inserted into the Schrödinger equation which allows one to reproduce the 2+1 D results and recover quantal effects after a single iteration.

8.3.3 Shikin model

The 2D Thomas-Fermi model above is based on the semiclassical approximation for the kinetic energy. Some time ago, in a somewhat different context, Shikin *et al.* [3, 4] proposed a more drastic approximation: a classical model where the kinetic energy contribution was neglected in writing the total energy of the dot, and no mirror term was explicitly included. In addition they assumed that the net confining potential was parabolic:

$$\overline{\Delta U}_c^S(r) = U_{0c} + \frac{1}{2}kr^2 \ . \tag{8.34}$$

With these simplifications they obtained an analytic solution for the charge distribution:

$$\sigma_S(r) = \sigma_0\sqrt{1 - \frac{r^2}{R_0^2}} \tag{8.35}$$

with σ_0 and R_0 given analytically in terms of the parameters defining the potentials. We shall now extend this method including appropriate approximations for the contributions of the mirror potential and the kinetic energy. We will show that one still obtains analytic expressions and that our extension leads to results that compare very well with the 2D Thomas-Fermi predictions of the previous section.

Model S0: Our starting point is to suppose that the mirror part of the electron-electron interaction

$$\overline{U}_{2D,m}(r) = -\frac{e^2}{\varepsilon_r}\int \frac{\sigma_{2D}(r')}{\sqrt{|\vec{r}-\vec{r}'|^2+4\bar{z}^2}}\; d\vec{r}' \;, \tag{8.36}$$

can be combined with the confining potential and these together can be approximated by the parabolic form given in eq. 8.34, with the values of U_{0c} and k fitted to the curve computed from eqs. 8.33 and 8.36. Then eq. 8.28 becomes:

$$\mu = E_z + U_{0c} + \frac{\hbar^2\pi}{m^*}\sigma_{2D}(r) + \frac{1}{2}kr^2 + U_{2D,e}(r) \tag{8.37}$$

with

$$U_{2D,e}(r) = \frac{e^2}{\varepsilon_r}\int \frac{1}{|\vec{r}-\vec{r}'|}\sigma_{2D}(r')\; d\vec{r}' \;. \tag{8.38}$$

It is now easy to prove[2] that Shikin's form, eq. 8.35, with appropriate parameters, is a solution of the above equation if we approximate the small kinetic energy contribution by its truncated quadratic expansion:

$$\frac{\hbar^2\pi}{m^*}\sigma_{2D} \;\simeq\; \frac{\hbar^2\pi}{m^*}\sigma_0\left(1-\frac{r^2}{2R_0^2}\right) \;. \tag{8.39}$$

Using the identity

$$\int d\vec{r}' \;\frac{\sqrt{1-\dfrac{r'^2}{R^2}}\,\Theta\left(1-\dfrac{r'^2}{R^2}\right)}{|\vec{r}-\vec{r}'|} \;\equiv\; \frac{1}{2}\pi^2 R\left(1-\frac{r^2}{2R^2}\right), \quad 0<r<R\,, \tag{8.40}$$

we get

$$U_{2D,e}(r) = \frac{\pi^2 e^2\sigma_0 R_0}{2\varepsilon_r}\left(1-\frac{r^2}{2R_0^2}\right) \;, \tag{8.41}$$

and therefore eq. 8.37 becomes:

$$\begin{aligned}\mu &= E_z + U_{0c} + \frac{\hbar^2\pi}{m^*}\sigma_0 + \frac{e^2\pi^2}{2\varepsilon_r}\sigma_0 R_0 \\ &+ \frac{r^2}{2R_0^2}\left(kR_0^2 - \frac{\hbar^2\pi}{m^*}\sigma_0 - \frac{e^2\pi^2}{2\varepsilon_r}\sigma_0 R_0\right) \;,\end{aligned} \tag{8.42}$$

so that to satisfy it, one has to choose:

$$\begin{aligned} R_0 &= \sqrt{\frac{\mu - E_z - U_{0c}}{k}} \\ \sigma_0 &= \frac{\mu - E_z - U_{0c}}{\frac{\hbar^2\pi}{m^*} + \frac{\pi^2 e^2}{2\varepsilon_r} R_0} \end{aligned} \tag{8.43}$$

Knowing these, direct integration gives the total number of electrons:

$$N = \frac{2}{3}\pi\sigma_0 R_0^2 \tag{8.44}$$

The only flaw in this scheme, is that U_{0c} is the value of the mirror potential at the origin, which depends on N, and thus iterations are required. Therefore we formulate below a new version of the model which treats the mirror term explicitly, but the price will be that the equation determining R_0 has to be solved numerically. Therefore we pass to

Model S1: whose ingredients are as follows:
a) we approximate *only* the gates plus donors potential, eq. 8.16 by a parabolic term:

$$\overline{\Delta U}_c(r) \simeq U_{2D,c}(r) \equiv \frac{1}{2}k_c r^2 \ . \tag{8.45}$$

The value of k can be easily extracted from the curve computed with eq. 8.33.
b) we assume that the density is still of the form of eq. 8.34, and therefore we find that the direct potential is given by eq. 8.41.
c) the mirror term can be expressed in a simplified form which is still sufficiently accurate: writing $D = 2\bar{z}$ for the distance between the plane of the dot and its mirror, one has:

$$\begin{aligned} U_{2D,m}(r) &= -\frac{e^2}{\varepsilon_r}\int d\vec{r}' \frac{\sigma_{2D}(r')}{\sqrt{D^2 + |\vec{r} - \vec{r}'|^2}} \\ &\simeq -\frac{e^2}{\varepsilon_r}\int d\vec{r}' \frac{\sigma_{2D}(r')}{\sqrt{D^2 + \frac{R_u^2}{2} + r^2}} \\ &= -\frac{e^2}{\varepsilon_r}\frac{N}{\sqrt{D^2 + \frac{R_u^2}{2} + r^2}} \ . \end{aligned} \tag{8.46}$$

Here we have replaced the $|\vec{r} - \vec{r}'|^2$ by its average over r' on a uniform disk of radius R_u. We have checked that this approximation is very accurate if R_u is chosen to be equal to $\sqrt{2/3}R_0$, i.e. it is the radius of the uniform disk containing the same charge as Shikin's density of radius R_0. Therefore

$$U_{2D,m}(r) = -\frac{e^2}{\varepsilon_r}\frac{N}{\sqrt{L^2 + r^2}} \ , \tag{8.47}$$

where $L^2 = D^2 + R_0^2/3$.

Starting again from the expression for the chemical potential

$$\mu = E_z + \frac{\hbar^2\pi}{m^*}\sigma_S(r) + U_{2D,c}(r) + U_{2D,e}(r) + U_{2D,m}(r) \ . \tag{8.48}$$

we see clearly that this cannot hold for all r when the density is taken to be of Shikin form, unless parabolic approximations for the kinetic energy and mirror terms are introduced. To avoid these we instead make the alternative approximation of requiring that eq. 8.48 be satisfied only in two cases: 1) when $r = 0$:

$$\begin{aligned} \mu &= E_z + \frac{\hbar^2\pi}{m^*}\sigma_0 + \frac{\pi^2 e^2}{2\varepsilon_r}\sigma_0 R_0 - \frac{e^2}{\varepsilon_r}\frac{N}{L} \\ &= E_z + \sigma_0\big[\frac{\hbar^2\pi}{m^*} + \frac{\pi^2 e^2}{2\varepsilon_r}R_0(1 - \frac{4}{3\pi}\frac{R_0}{L})\big] \ , \end{aligned} \tag{8.49}$$

and 2) as an uniform average over a disk of radius R_0: Integrating both sides of 8.48 and dividing by πR_0^2 we find:

$$\begin{aligned} \mu &= E_z + \frac{2}{3}\frac{\hbar^2\pi}{m^*}\sigma_0 + \frac{3}{4}\frac{\pi^2 e^2}{2\varepsilon_r}\sigma_0 R_0 + \frac{1}{4}k_c R_0^2 \\ &- \frac{4\pi e^2}{3\varepsilon_r}\sigma_0\left(\sqrt{L^2 + R_0^2} - |L|\right) \ . \end{aligned} \tag{8.50}$$

From 8.49 and 8.50 we can eliminate σ_0 and write an analytic equation with R_0 as the only unknown. This equation is solved numerically for R_0, using e.g. a Newton-Raphson algorithm.

Results The predicted density is also shown in figure 8.5 (dotted line.) As can be seen it is very close to the Thomas-Fermi prediction. This confirms the validity of the additional approximations introduced. The numerical values found are $\sigma_0 = 0.00240$ nm^{-2}, $R_0 = 101.0$ nm and $N = 51.22$ electrons.

Acknowledgements

We are grateful to NSERC-Canada for Discovery Grant RGPIN-3198 (DWLS) and to DGES-Spain for continued support through grants BFM2001-3710 and FIS2004-03156 (JM).

Bibliography

[1] J. Martorell, Hua Wu and D. W. L. Sprung, Phys. Rev. B **50** , 17298-308 (1994).

[2] D. W. L. Sprung and J. Martorell, *Semi-classical treatment of electrons in quantum wires*, Solid State Commun. **99**, 701-6 (1996).

[3] V. Shikin, S. Nazin, D. Heitmann and T. Demel, Phys. Rev. B **43**, 11903-7 (1991).

[4] V. Shikin, T. Demel and D. Heitmann, Sov. Phys. JETP **69**, 797-803 (1989).

[5] S. E. Laux, J. D. Franck and F. Stern, Surf. Sci. **196**, 101 (1988).

[6] A. Kumar, S. E. Laux and F. Stern, Appl. Phys. Letters **54**, 1270 (1989).

[7] A. Kumar, S. E. Laux and F. Stern, Phys. Rev. B **42**, 5166 (1990).

[8] J. H. Davies, Semicond. Sci. Technol. **3**, 995 (1988).

[9] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover Publications, New York, 1965), Ch. 17.

Chapter 9

Quantum Monte Carlo Methods

R. Nieminen
Laboratory of Physics,
Helsinki University of Technology,
P.O.Box 1100, FIN-02015 HUT, Finland

9.1 Brief outline of the techniques

Quantum Monte Carlo (QMC) methods enable accurate solutions of the Schrödinger equation for many-particle systems. This is particularly useful in such cases where the interactions between particles are strong enough to make mean-field-type solutions erroneous, sometimes even qualitatively. Examples of such systems are strongly correlated electrons in condensed matter or discrete-lattice Hamiltonians with quantum degrees of freedom.

In *variational* QMC, one postulates a wavefunction for the system and calculates the physically interesting expectation values using Metropolis techniques. For example, the ground state energy of a N-particle system is

$$E_{VCM} = \frac{\langle\Psi|H|\Psi\rangle}{\langle\Psi|\Psi\rangle} = \int d\mathbf{R}^{3N} \left(\frac{H\Psi}{\Psi}\right)_{\mathbf{R}} \frac{|\Psi(\mathbf{R})|^2}{\int d\mathbf{R}^{3N}|\Psi(\mathbf{R})|^2} = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{H\Psi}{\Psi}\right)_{\mathbf{R}_i} \tag{9.1}$$

where H is the Hamiltonian and $\mathbf{R}$ defines a random walk in $3N$ dimensions ($\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$). Variational QMC is basically a trick to calculate the multidimensional integral, and the essential physics is contained in the wavefunction. Its choice and variational freedom in minimizing EVMC are therefore crucial.

In *diffusion* QMC, one realizes an iterative projection by solving the Schrödinger equation in imaginary time:

$$|\Psi^{(n+1)}\rangle = e^{-\tau\hat{H}}|\Psi^n\rangle \quad \xrightarrow{n\to\infty} \quad |\Psi_0\rangle \quad . \tag{9.2}$$

Here, τ is an arbitrary small constant and $|\Psi_0\rangle$ an arbitratry initial state. The quantum-statistical symmetry of the wavefunction is important. For fermions, the total wavefunction is antisymmetric. Usually, the node structure is preserved during diffusion QMC iteration (the fixed-node approximation). In this case, the method finds the best ground-state wavefunction with the same nodes as the initial (trial) wavefunction. The trial wavefunction thus enforces the fermionic antisymmetry and selects the state.

For *excited* states, diffusion QMC finds the exact solution provided that the nodal structure is exact.

A popular choice for QMC is the fermion trial wavefunction of the Jastrow-Slater form

$$\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N) = J(\mathbf{r}_1, \ldots, \mathbf{r}_N) \times \sum_i c_i D_i(\mathbf{r}_1, \ldots, \mathbf{r}_N) \tag{9.3}$$

where J is the Jastrow factor and D_i are determinants made of single-particle wavefunctions ("orbitals") for the electrons. The Jastrow factor depends on the inter-electronic distances and accounts for their repulsion, is positively definite, and also takes care of divergences of the inter-electronic (Coulomb) potential. The summation is over a few (Slater) determinants, which defines the nodal surface. Note that there are *not* millions of determinants, as in multiconfiguration quantum-chemistry calculations.

An important current theme is the *optimization* of the trial wavefunction. This is usually best done by not minimizing the trial energy directly, but minimizing the variance of the "local" energy

$$\sigma^2[\alpha] = \sum_{i=1}^{N_{conf}} \left(\frac{H\Psi(\mathbf{R}_i, \{\alpha\})}{\Psi(\mathbf{R}_i, \{\alpha\})} - \bar{E} \right)^2 \tag{9.4}$$

depending on a set of parameters α.

QMC methods have proven most useful for obtaining accurate values for binding energies, bond distances and angles, vibrational and torsional excitations of molecules and clusters. They are also increasingly used in solid state physics problems, including the core electrons of atoms treated within a pseudopotential construction. They can also be generalized to *finite temperatures* through the *path-integral* formulation. Techniques are being developed for accurate evaluation of interatomic Hellmann-Feynman forces from the QMC total energies, which would enable *first-principles molecular-dynamics* simulations in the QMC context.

9.2 Challenges

QMC methods offer in general a powerful approach for quantitative modeling of systems where quantum many-body correlations cannot be modeled accurately enough by mean-field techniques, such as those based on density-functional theory (DFT) or (for bosons) the Gross-Pitaevskii theory. QMC methods are best suited for benchmarking ground-state properties, and they allow system sizes well beyond those relevant for exact-diagonalization techniques (typically less than 10 electrons). Another strength of QMC methods (of the path-integral type) is that they enable the inclusion of quantum and classical degrees of

freedom on equal footing and at finite temperatures.

Yet the QMC methods in their standard formulations are computer-intensive, which has slowed down their application in real-world materials problems, as opposed to "clean" quantum liquids and gases and model solid-state problems, such as a few electrons confined in a parabolic potential well.

An important challenge for the QMC methods is thus to *develop new algorithms and implementations with better scaling as the number of degrees of freedom is increased.* There has been considerable recent progress in developing linear-scaling QMC algorithms, which could open the door applications to system sizes with hundreds rather than tens of electrons, for example.

A second challenge is the solidification of the QMC techniques as applied to *excited states and the energy hypersurfaces away from the Born-Oppenheimer ground state.* Such nonadiabatic energy contours are very important in understanding and modeling situations studied by pump-and-probe spectroscopies and relevant for reaction and relaxation dynamics.

A third challenge is provided by the improved treatment of *interatomic forces*, both in the ground and excited states. The ultimate goal should be the capability to simulate the full nonadiabatic dynamics of a coupled electron-ion system at finite temperatures and in real time, a situation most relevant for e.g. time-resolved x-ray imaging of nanoscale structures at future synchrotrons, such as those based on free-electron lasers. A specific question in this context is whether QMC can offer an alternative way to model the weak *van der Waals-type interactions*, induced by nonlocal correlations between atomic-scale objects. These are notoriously difficult to model by standard techniques.

9.3 Opportunities for applications relevant in nanosciences

The application areas for QMC methods are wide and expanding, especially if the pace of methodological development can be maintained. Apart from the traditional application areas of QMC methods to quantum fluids (both bulk, surface and interface properties) such as electron gas, helium and hydrogen etc., the field of *confined quantum gases* continues to pose interesting problems. These low-density, low-temperature systems are clean, controllable quantum systems where the role of quantum statistics, interparticle interactions, and optical/magnetic confinement can be systematically studied. Periodic optical lattices can be generated in one, two or three dimensions, and the behavior can be tuned between a Fermi or Bose gas and a Hubbard-like lattice.

Semiconductor quantum dots continue to attract modelling interest, which seems to be moving from "toy" problems of a few electrons in a featureless trap towards realistic aggregates of thousands of atoms. The subtle interplay between many-body effects and others (band structure, spin-orbit coupling etc.) can be explored by systematic application of improving QMC techniques.

For modeling of realistic nanostructures, *atomic-scale defects* are important but notoriously difficult to model accurately in semiconductors and insulators. The defect-associated electronic levels control their properties, but are difficult to place accurately due to the inaccuracy of standard DFT methods in the bandgap region. QMC could come to rescue here and enable much better prediction of defect-induced properties.

QMC-based molecular dynamics in the ground and excited states, already described above, is a major possibility for simulation of nanoscale structures, for example the real-time melting and dissociation dynamics of clusters and aggregates, either unsupported or depeosited on a substrate. The range of opportunities in modeling the processing of various nanoscale objects, including *biologically relevant molecular systems*, is enormous.

Nanoscale *quantum transport* attracts much interest among the theory and modeling communities. The popular approaches to nanoscale and molecular electronics transport include the Landauer-Bttiker-type mean-field methods, time-dependent DFT, and the quantum master equation. Yet the role of strong inter-electron correlations in the narrow channels or constrictions remains unclear. QMC methods, with their explicit inclusion of many-body correlations, could offer interesting alternatives for accurate modeling.

These are just a few examples of the opportunites that the QMC approach provides for theoretical and computational nanoscience. It is a robust but flexible framework. *Standardised, well-tested software* with new functionalities and utilities has appeared, which has quickly increased the popularity of the method. The range of applications has increased substantially, especially in the nanoscience area. QMC not only complements and benchmarks the widespread mean-field-type approaches, but also provides alternatives for situations where those simply fail. One can foresee the emergence of QMC acomparable to DFT as a popular platform for modeling nanoscale quantum systems.

Chapter 10

Molecular Dynamics Simulations

R. Nieminen
Laboratory of Physics,
Helsinki University of Technology,
P.O.Box 1100, FIN-02015 HUT, Finland

10.1 Brief outline

Molecular dynamics (MD) simulation has become the tool of choice to model physical properties of complex systems. The basic idea is to explicitly solve the equations of motion of the atoms and molecules in the system, interacting with specified forces and under given external constraints. The description of the instantaneous forces can vary from empirically fitted force fields to "tight-binding" or Hückel treatments of electronic degrees of freedom and eventually to full, "*ab initio*" MD where the forces are calculated by minimizing the electronic quantum energy functional at each time step. An alternative formulation of the latter is the famous Car-Parrinello approach, where the electronic degrees of freedom are assigned a fictitious classical mass and are allowed to relax to the adiabatic energy surface in the course of the simulation.

MD simulation has become a huge industry with wide-range applications in physics, chemistry, materials science and molecular biology. The large increase in computing power during the last 50 years has pushed the attainable timescales from nanoseconds to milliseconds and the system sizes from tens of atoms to billions of atoms. However, even if the CPU power and memory capacities continue to increase at this speed, we can cover only a tiny fraction of the experimentally relevant length and time scales. The capability gap can only be bridged by the development and systematic application of new computational techniques.

In the nanoscience realm, there are several outstanding issues to be resolved via direct modeling and simulation. These are often related to the *manufacturing and processing* of nanoscale objects through

a complex set steps, including physical and chemical deposition, annealing, *in situ* chemical reactions, diffusion on a substrate, evaporation and desorption, etc. In many cases, these process steps involve *non-equilibrium* and nonadiabatic processes and non-equilibrium thermodynamics. The timescales range from femtoseconds to seconds and beyond, and the system sizes range from a few atoms up to mesoscale (micrometer) objects. Biologically inspired nanostructures are rapidly increasing in importance, as the self-assembly, self-organisation and other bottom-up techniques gain popularity. Biological systems are typically far too large for their force fields to be treated fully quantum mechanically.

Many of the interesting phenomena currently investigated in nanometer-size objects are related either to their electronic and optical properties (e.g. in quantum dots, wires and rings and in carbon nanotubes) or to their peculiar structural and mechanical properties (e.g. in carbon nanotubes and composites and multiferroic thin films). Systems with unique chemical properties have been discovered, such as the strongly enhanced catalytic activity in small gold clusters. Magnetic properties can be tuned as well. In the future, it is likely that new, functionalized materials and devices will be built by combining different nanoscale objects into aggregates with truly novel properties. Metamaterials with completely new electromagnetic properties (such as negative dielectric and magnetic permittivity and refractive index over a wide range of frequencies) are a distinct possibility. MD simulation is an indispensable tool in the design of the assembly process of such materials.

10.2 Challenges for molecular dynamics simulation

Especially from the nanoscience perspective, the outstanding challenges of MD simulation are related to the Holy Grail of *multiscale modeling*, i.e. to the quest of bridging together timescales from femtoseconds to minutes and length scales from sub-nanometers to millimeters. For short time and length scales, this quest is coupled to the progress in QMC and other methods (see above). For long time and length scales, the coupling is to kinetic equations and stochastic methods, and eventually to continuum equations. The multiscale challenge includes the ability to do systematic coarse-graining so as to eliminate irrelevant degrees of freedom, and yet to retain an accurate and realistic description in the remaining degrees of freedom.

Examples of technique-related challenges for MD simulation include the following:

Hybrid quantum-classical techniques. The central idea, relevant especially for biologically relevant nanosystems, is to apply full quantum description (Schrödinger equation) only for a small, most crucial part of the system, and to treat the rest classically. The challenge is to develop efficient and reliable methods for the embedding of the quantum part to the environment, described either in terms of Newtonian MD equations or as a more featureless continuum.

Mesoscale modeling. For a typical process simulation it is often necessary to reach length scales in the micrometer range, beyond the limits of what can be achieved by standard MD simulations. Methods for coarse-graining have to be developed. This means lumping together groups of atoms and molecules as "effective particles" with "renormalized" interactions. The challenge is to systematically derive those interactions. Methods such as Lattice Boltzmann Techniques or Dissipative Particle Dynamics are vig-

orously pursued to provide downward and upward links to mesoscale modeling.

Rare event simulations. An often recurring problem in MD simulations is how to pass through the bottleneck caused by rare events in the dynamical evolution. In many cases, free-energy barriers exist and lead to very long simulation times for the relevant phenomena. The indentification of such barriers is far from trivial: the barrier corresponds to an ensemble of transition states in a very high-dimensional space. Computational techniques and algorithms are being developed to locate the transition-state ensemble and thus to guide the system through the rare-event bottleneck.

Solvent/environment effects. *Ab initio* MD enables the study of environmental effects (solvents, ligands etc.) on a specific chemical reaction. *Ab initio* MD is however time-consuming and it is difficult to reach simulation times needed to observe a chemical reaction *in silico*. However, such a simulation can deliver the free energy of a given reaction coordinate and thus open a way to uncover the reaction kinetics. The challenge is develop systematic methods to modeling chemical kinetics.

A remarkable convergence in methods and approaches is taking place in the nanoscience community, where scientists with different backgrounds now strongly interact. Roughly speaking, three streams seem to converge, also in their arsenals for simulation. Biophysics/biochemistry at the molecular level meets computational chemistry and computational physics/materials science. Quantum physics and chemistry meets Newtonian dynamics and statistical thermodynamics, equilibrium ensembles meet nonequilibrium kinetics. Molecular dynamics simulation is at the epicenter of this convergence, and will continue to play a very important role in computational and theoretical nanoscience.

Chapter 11

Electric field calculations in scanning probe microscopy: Generalized image charge method

E. Sahagún, G. M. Sacha, L. S. Froufe-Pérez and J. J. Sáenz
Moving Light & Electrons (MOLE) group,
Departamento de Física de la Materia Condensada,
Universidad Autónoma de Madrid, E-28049 Madrid, Spain

Abstract. We present a method to calculate the electrostatic field between a metallic tip of arbitrary shape and a metallic or insulating sample. The basic idea is to replace the electrodes by a set of "image" charges. These charges are adjusted in order to fit the boundary conditions on the surfaces. This generalized image-charge method is particularly useful for modelling electrostatic fields, forces and capacitances when different length scales are included in the simulation. The versatility of the method allowed for an analysis of the electrostatic problem as a function of the tip apex geometry. The method can be easily extended to analyze multilayered and anisotropic samples.

11.1 Introduction

The long range nature of electrostatic interactions makes them specially suitable to perform noncontact scanning probe microscopy (SPM) imaging of both conducting and insulating materials. By applying a voltage between a force microscope tip and a sample, electrostatic force microscopy (EFM) and related techniques have been used to study capacitance [1], surface potential [2], charge or dopant distribution [3], topography and dielectric properties of metallic and insulating [2, 4] surfaces, the dielectric response of single molecules [5, 6] and to deposit and image localized charges on insulators [7]. In analogy with the

magnetic force microscope [8], EFM has been used to image the domain structure of ferroelectric crystals [9]. Polarization forces have also been used to perform electrostatic spectroscopy [10], to imaging weakly bound materials and liquids [11] and to induce the formation of water nanobridges [12].

As in other SPM techniques, the interpretation of the EFM images is not always evident [13]. Since EFM is a nonlocal technique due to the long range nature of the electrostatic interaction, the detailed shape and dimensions of the tip must then be taken into account for a precise calculation of both force and capacitance [14, 15, 16, 17].

In a different context, the electric field generated by sharp tips is the most critical parameter governing electron field emission and closely related phenomena. Field electron and ion microscopes [18, 19], field emission electron guns [20], thin-film field emission cathodes [21] and many devices based on vacuum field emission are found throughout the sciences. The advances in engineering of ultrasharp tips at atomic scale [22, 23] and, more recently, the development of field emission based flat panel displays [24] and nano-electromechanical single electron transistors [25] generated renewed interest in this field. The remarkable properties of the electron beam generated from these nanometer-sized sources depend critically on the electric field around the tip apex [26, 27, 28, 29, 30].

Additional motivation for our work arises from the combination of field emission devices with scanning tunneling microscope (STM) technology [31, 32, 33, 34, 35]. *Near field emission STM* [36] provides a direct, non invasive approach for investigating at nanometer scale. An early example of such approach was the Young's "topografiner" [37]. When a STM is operated in the near field emission regime (like a topografiner), the electric field strength at the emitter surface determines the I versus V characteristics through the Fowler-Nordheim equation [19, 38]. Operating the instrument in the constant current mode implies an approximatelly constant field at the emitter surface. By solving the Laplace's equation we can then calculate the field strength everywhere between the tip and the sample, and thus determine the relationship between the tip-sample distance S and the emitter voltage V [37].

Most of the predictions of the electric field shape and amplitude in the tip-sample gap used so far are based on simple analytical models or sophisticated numerical calculations. Moreover, theory an experiments on field emission systems have been mainly focused in the far field regime, i.e. when a tip cathode is placed at a macroscopic distance from the sample (anode) surface. The main problems of the simulations are related to the different scales of a typical SPM setup. While the tip-cantilever system size is of the order of micrometers, the tip radius can be of the order of few nanometers and the tip-sample distance can range from macroscopic to sub-nanometer distances.

The aim of this work is to introduce the Generalized Image-Charge Method (GICM) [39, 40, 41] as a very useful technique to overcome the scale problem in SPM geometries. The GICM is optimized to calculate the electrostatic field between a metallic tip of arbitrary (axial symmetric) shape and a sample surface for any tip-surface distance. The basic idea is to replace the electrodes by a set of "image" charges. These charges are adjusted by means of a standard least-squares method in order to fit the boundary conditions on the surfaces. The method is much simpler and faster than standard methods used to solve the Laplace equation. It is also rather flexible and can be applied to many different problems ranging from the design and characterization of field emission guns, the study of field effects in STM experiments

[36, 42] and, in general, modelling electrostatic interactions in SPM geometries.

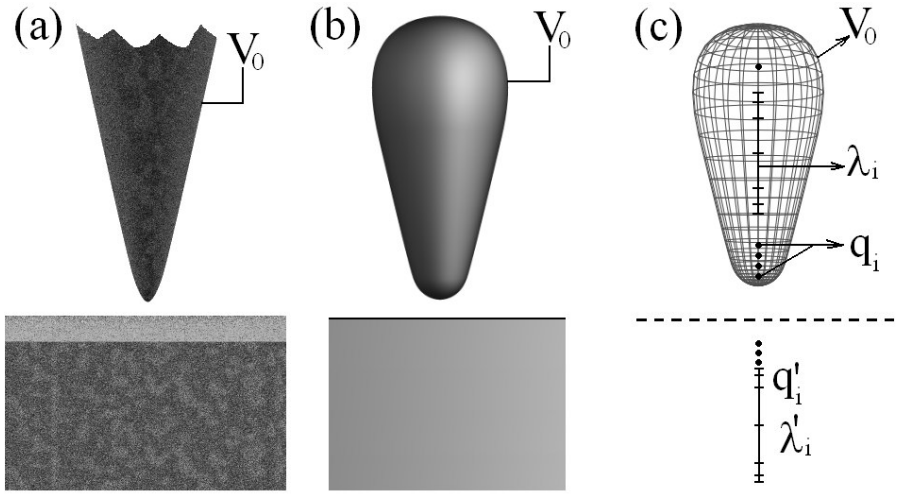


Figure 11.1: Graphic scheme of the GICM basis. A biased tip in presence of a substrate (a) is consider to be a revolution solid (b) and is substituted by a set of charges (q_i, λ_i) and its images (q_i', λ_i') (c) which give rise to an equipotential surface equal in geometry and value to that of the tip.

In section 11.2, we present the Generalized Image Charge Method and discuss the details of the field calculation. The original method [39] was restricted to work with smooth tips but probably the most important restrictions are connected with the substrates. They should be homogeneous and flat. But despite of this restrictions, many generalizations have being made [41]. Some of them are presented in this work. In particular, we will discuss an extension of the GICM to model electrostatic fields for dielectric and anisotropic thin films. As a particular application, in section 11.3 we present some examples and a number of different results of interest in Electrostatic Force Microscopy. The concluding remarks are given in section 11.4.

11.2 Generalized Image-Charge Method (GICM)

The calculation of the electric field generated by very sharp tips is far from be a trivial problem. One of the main difficulties in the computation is to handle accurately the large change in geometrical scale between the tip radius and the tip to sample spacing. The situation is even worse when the tip has not a smooth shape, for example due to the presence of small nanometer-scale protrusions.

In the past decades many methods for calculating the electrostatic field in field emission systems have been developed [20]. Among these, effective charge-density methods [43, 44, 45] are specially appropriate for the calculation of capacitances or the computation of trajectories in electrostatic focusing [43, 44, 45, 29]. With a clever choice of the charge distribution on the surface it is even possible to simulate the electric field at atomic scale [28]. Several techniques based on finite elements have also been developed to simulate electrostatic force in the context of AFM [46].

From a general point of view, the charge density methods emphasize the superposition principle

and the self-consistency of the boundary conditions [45], e.g. the charges on the conductor are calculated by requiring that the potential be a constant on each conductor surface, in contrast with the finite-difference/element approach in which the potential at each point is obtained as the average of the potentials at the adjacent points. Following the same basic ideas, we will introduce the GICM as a method specially suitable to calculate the electrostatic field between a metallic tip of arbitrary shape and a sample surface for any tip-surface distance. The aim of this study is to solve the electrostatic problem presented in figure (11.1a), this is, an SPM biased probe above a certain substrate. The GICM solves this problem in a very intuitive way which not only gives extremely accurate results but implies a very low computational effort.

Making use of the classical electrostatic image method, we replace the electrodes by a set of "image" charges. The central idea is to substitute the whole tip by a set of charges (see figure 11.1). A biased tip can be considered roughly as a revolution solid with a well defined electrostatic potential V_0 over its surface. If we find a system of discrete charges whose potential were the same as the one of the tip's surface, both systems would be equivalent. So knowing the value and position of this charges which "simulate" tip equipotential surfaces the problem is reduced to the calculation of the potential generated by a charge over a substrate. And here is where the classical image-charges method plays its role. It allows to solve in a very easy way the potential due to a charge in presence of flat interfaces.

For simplicity, let us discuss in some detail the case of a sharp tip with axial symmetry in front of a flat homogeneous semi-infinite insulating surface characterized by a dielectric permitivity ϵ. First we place, along the symmetry axis ($\rho = 0$), a set of M segments λ and N discrete charges q in front of the sample surface ($z = 0$). The electrostatic potential of a point charge located at $z = z_n$ is simply given by the textbook image-charge result:

$$\frac{q_n}{\epsilon_0} G(\rho, z; z_n) = \frac{q_n}{\epsilon_0} \begin{cases} (G_0(\rho, z; z_n) - \beta G_0(\rho, z; -z_n)) & \text{for } z > 0 \\ (1-\beta) G_0(\rho, z; z_n) & \text{for } z < 0 \end{cases} \tag{11.1}$$

where $\beta = (\epsilon - 1)/(\epsilon + 1)$ and

$$G_0(\rho, z; z_n) = \frac{1}{4\pi\sqrt{\rho^2 + (z - z_n)^2}}. \tag{11.2}$$

Notice that G_0 and G are the electrostatic Green functions in free space (i.e. just the Coulomb potential) and in the presence of the sample, respectively.

The corresponding potential for linear charge density, uniformly distributed along a segment ($z_m - L_m/2 < z < z_m + L_m/2$), can then be written as:

$$\begin{aligned} \frac{\lambda_m}{\epsilon_0} G^{(L)}(\rho, z; z_m) &= \frac{\lambda_m}{\epsilon_0} \int_{z_m - L_m/2}^{z_m + L_m/2} G(\rho, z; z') dz' \\ &= \begin{cases} \left(G_0^{(L)}(\rho, z; z_m) - \beta G_0^{(L)}(\rho, z; -z_m) \right) & \text{for } z > 0 \\ (1-\beta) G_0^{(L)}(\rho, z; z_m) & \text{for } z < 0 \end{cases} \end{aligned} \tag{11.3}$$

with

$$G_0^{(L)}(\rho, z; z_m) = \frac{1}{4\pi} \ln \left\{ \frac{(z - z_m + L_m) + \sqrt{\rho^2 + (z - z_m + L_m)^2}}{(z - z_m - L_m) + \sqrt{\rho^2 + (z - z_m - L_m)^2}} \right\}. \tag{11.4}$$

The total electrostatic potential at a point, (ρ, z) $(z > 0)$, generated by our distribution of charges would then be

$$V(\rho, z) = \sum_{n=1}^{N} \frac{q_n}{\epsilon_0} G(\rho, z; z_n) + \sum_{m=1}^{M} \frac{\lambda_m}{\epsilon_0} G^{(L)}(\rho, z; z_m) \tag{11.5}$$

The position z_n and strength q_n of each point charge and the length, $2L_m$, the position of the center, z_m, and the charge density, λ_m, of each segment are the unknowns of the problem. **These parameters are adjusted in order to fit the boundary condition on the tip surface**, $V = V_0$ (notice that the boundary condition on the flat surface is already included in Eq. 11.5 through the contribution of the image charges).

Once we have $\{q_n\}$ and $\{\lambda_m\}$, we can obtain the capacitance of the tip-sample system from the expression $C = Q/V_0$. The total charge Q is obtained from the sum of the charges inside the tip, i.e.

$$C = \frac{1}{V_0} \left\{ \sum_n q_n + \sum_m (2\lambda_m L_m) \right\}. \tag{11.6}$$

The force and force gradient can then be calculated from the coulomb interaction or from the derivatives of the capacitance.

11.2.1 Green's Function and multilayered structures

As we have seen, the basic idea of the GICM is to solve the electrostatic problem fulfilling boundary and interfacial conditions by placing image charges. If the system is simple enough, one can place this images using ones intuition. As we will see, the method can be extended and systematized through the use of Green's functions.

Let us again consider a point charge over a general substrate composed by different layers with dielectric permitivity ϵ_i and thickness W_i but still having axial symmetry (see Fig. 11.2). The total electrostatic potential above the sample (i.e. for $z > 0$) is proportional to the total Green's function that can be expanded in cylindrical coordinates as [48]

$$G(\rho, z; z_n) = G_0(\rho, z; z_n) + \frac{1}{4\pi} \int_0^\infty f_1(k) J_0(\rho k) e^{-kz} dk \tag{11.7}$$

where G_0 is the free space Green's function (Eq. 11.2),

$$G_0(\rho, z; z_n) = \frac{1}{4\pi\sqrt{\rho^2 + (z - z_n)^2}} = \frac{1}{4\pi} \int_0^\infty J_0(\rho k) e^{-k|z - z_n|} dk \tag{11.8}$$

and $f_1(k, z_n)$ is a function that depends on the sample. For the homogenous dielectric sample discussed

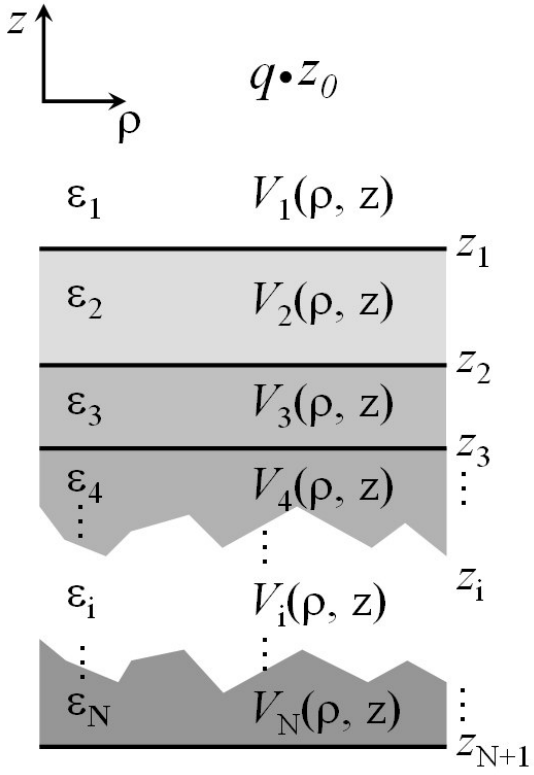


Figure 11.2: Graphic scheme and coordinate system of the problem. The media under the charge can be in principle made of an indefinitely number of layers with different dielectric permitivities.

above, it is simply given by $f_1(k, z_n) = -\beta e^{-kz_n}$.

Within each dielectric layer i (see FIG. 11.2) the Green function can be written as

$$G_i(\rho, z; z_n) = \frac{1}{4\pi}\left(\int_0^\infty f_i(k, z_n) J_0(\rho k) e^{-kz} dk + \int_0^\infty g_i(k, z_n) J_0(\rho k) e^{kz} dk\right) \tag{11.9}$$

Notice that the corresponding Green functions for a linear charge distribution is again given by

$$G_i^{(L)}(\rho, z; z_m) = \int_{z_m - L/2}^{z_m + L/2} G_i(\rho, z; z') dz' \tag{11.10}$$

The different functions $\{f_i(k, z_n), g_i(k, z_n)\}$ can be obtained from the matching conditions at each interface:

$$G = 0 \quad \text{at infinite or on the grounding electrode} \tag{11.11}$$

$$G_i = G_{i+1} \quad \text{at the interfaces} \tag{11.12}$$

$$\epsilon_i \frac{\partial G_i}{\partial z} = \epsilon_{i+1} \frac{\partial G_{i+1}}{\partial z} \quad \text{at the interfaces} \tag{11.13}$$

The complete set of equations obtained from all the matching conditions can be solved numerically by

using, for example, a simple transfer matrix technique.

For the case of a thin film dielectric sample (of thickness h and dielectric permitivity ϵ) on a grounded metallic electrode (see Fig. 11.4) it is easy to find a closed expression for $f_1(k)$ in Eq. 11.7:

$$f_1(k) = -\frac{\epsilon - \tanh(kh)}{\epsilon + \tanh(kh)} e^{-kz_n} \tag{11.14}$$

In summary with this formalism it is possible to calculate in a very systematic way the Green function of mulilayered systems for charges and segments of charge.

11.2.2 Fitting the boundary conditions on the tip surface

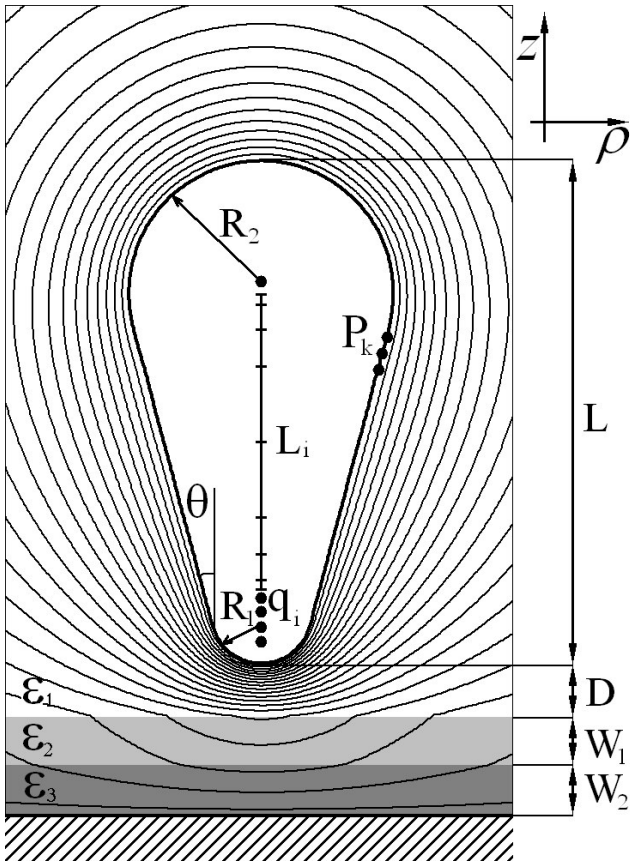


Figure 11.3: Complete scheme of the tip-substrate setup for a sample made of two layers. The equipotential surfaces are computed assuming $\epsilon_1 = 1$, $\epsilon_2 = 4$ and $\epsilon_3 = 2$.

To describe the geometry of the tip we take a sample of P points, $\vec{r}_j = (\rho_j, z_j)$ $(j = 1, P >> N, M)$, on the tip surface where the potential must be constant $V = V_0$. The best values for L_m,z_m and q_i can

then be obtained after minimization of

$$\chi^2 = \sum_{j=1}^{P} \{V(\rho_j, z_j) - V_0\}^2 \tag{11.15}$$

Including the lengths and positions of the charges in the minimization process leads to a non-linear minimization problem which is poorly convergent. This can be solved by using a nonlinear least squares minimization routine [47]. However, the initial guess of the parameters is somehow fussy, sometimes unstable and slows down the calculation. After minimization, for a given tip geometry, the optimized positions and lengths of the effective charges are almost independent on the tip-surface distance [39]. Since the potential is a linear function of the charges, after a first non linear minimization (or after an appropriate guess of positions and lengths) the problem becomes a simple linear least-squares fitting. In this case, the potential can be obtained with a negligible computational effort. An appropriate choice of charge positions and lengths is critical for an optimum performance of the GICM method [41].

For simplicity, and as an example, we will consider the geometry of the general tip shown in figure 11.3. The shape of this tip can be described with a truncated cone and two spheres of radius R_1 and R_2. In practice, in order to get an equipotential having the spherical shape of the tip near the apex a good choice is to place a set of point charges at the positions of the image charges inside a metallic sphere over a plane surface [41]. To define the body of the tip we use a set of segments which are shorter near the apex and much longer in the center of the tip (see Fig. 11.3), being their increment geometrical.

Another important issue in the GICM is the choice of a set of sampling points on the tip surface. Although the tip apex gives the most important contribution to the electrostatic force, the macroscopic contribution must be also taken into account. The number of points must be balanced along the tip surface so that we have a higher density in those places with lower curvature radius [41].

After the minimization process, the equipotential surfaces generated by the set of charges fits almost perfectly the original surface. Of course, the final result comes from a least square minimization so the voltage in all the sample points P_k is not exactly V_0 (notice that the number of sample points is much larger than the number of unknowns). In practice, the potential on any of the sample points deviates less than 1% from V_0.

11.2.3 Anisotropic substrates.

The problem of a tip over an anisotropic substrate can be easily solved under the formalism presented here. Among other different applications, the accurate modelling of electrostatic fields in these systems play an important role in the understanding piezoelectric and dielectric properties of thin films, including ferroelectric properties, piezoelectric coupling coefficients, and domain structures, that can deviate significantly from the corresponding bulk values [49]. The GICM should also find important applications in the optimization of piezoresponse force microscopy (PFM)[50] and its spectroscopic variants.

Let us consider an anisotropic media characterized by a dielectric tensor whose non-zero components

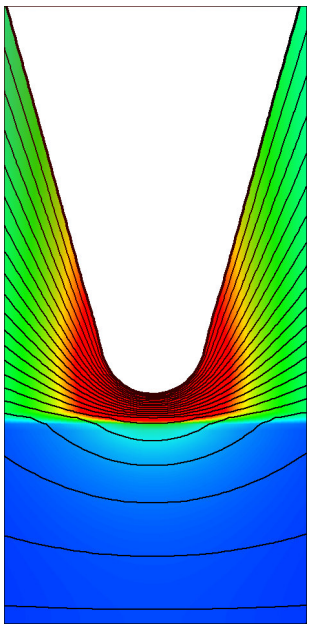


Figure 11.4: Equipotential map with the modulus of the electric field superimposed. ($D = 0.5R_1$, $W = 4R_1$ and $\epsilon = 6$).

are given by

$$\epsilon_{xx} = \epsilon_{yy} \equiv \epsilon_\rho \quad ; \quad \epsilon_{zz} = \epsilon_z. \tag{11.16}$$

In this case the calculation of the electrostatic potential can be done from a simple generalization of the appropriate Green's functions. For example, the Green's function in an infinite homogeneous medium with this anisotropy is given by [51]

$$G_0(\rho, z; z_n) = \frac{1}{4\pi\sqrt{\epsilon_\rho \epsilon_z}} \left\{ \rho^2 + \frac{\epsilon_\rho}{\epsilon_z}(z - z_n)^2 \right\}^{-1/2} \tag{11.17}$$

which, in analogy with Eq. 11.2, can be rewritten as

$$G_0(\rho, z; z_n) = \frac{1}{4\pi\sqrt{\epsilon_\rho \epsilon_z}} \int_0^\infty J_0(\rho k) e^{-k|z-z_n|\sqrt{\frac{\epsilon_\rho}{\epsilon_z}}} dk \tag{11.18}$$

For a multilayered system made of such anisotropic materials, equations 11.7, corresponding to the potential of a point charge over the anisotropic sample remains the same while, within each dielectric layer i, the Green function (Eq. 11.9) becomes

$$G_i(\rho, z; z_n) = \frac{1}{4\pi} \left(\int_0^\infty f_i(k, z_n) J_0(\rho k) e^{-kz\sqrt{\frac{\epsilon_\rho}{\epsilon_z}}} dk + \int_0^\infty g_i(k, z_n) J_0(\rho k) e^{kz\sqrt{\frac{\epsilon_\rho}{\epsilon_z}}} dk \right) \tag{11.19}$$

The different functions $\{f_i(k, z_n), g_i(k, z_n)\}$ can be obtained from the matching conditions at each inter-

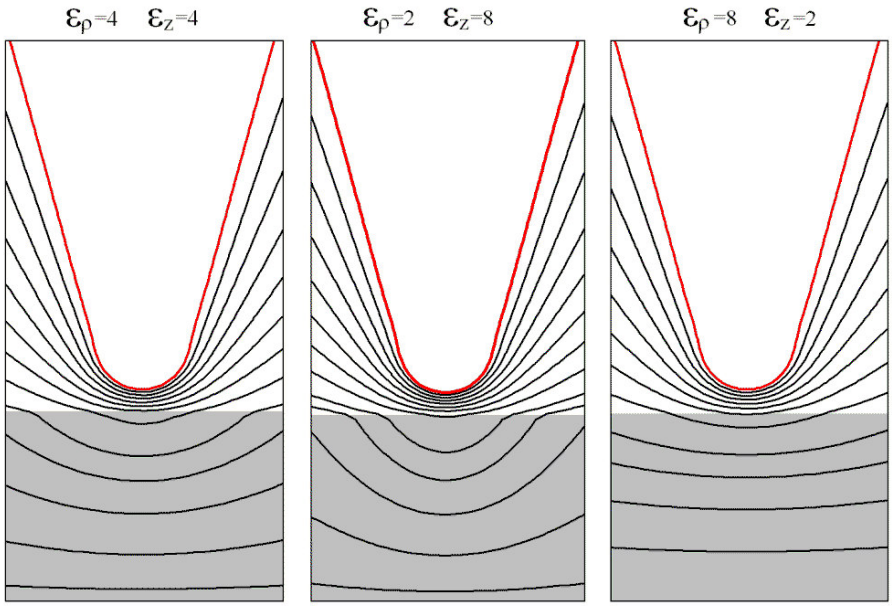


Figure 11.5: Equipotential surfaces for an anisotropic thin film. (for $R = 1$, $L = 1000R$, $D = 0.5R$, $\theta = 15$ and substrate thickness $W = 4R$)

face:

$$G = 0 \quad \text{at infinite or on the grounding electrode} \tag{11.20}$$

$$G_i = G_{i+1} \quad \text{at the interfaces} \tag{11.21}$$

$$\epsilon_{i_z} \frac{\partial G_i}{\partial z} = \epsilon_{i+1_z} \frac{\partial G_{i+1}}{\partial z} \quad \text{at the interfaces} \tag{11.22}$$

For the case of a thin film made of anisotropic material on a grounded metallic electrode it is easy to find a closed expression for $f_1(k)$ in Eq. 11.7:

$$f_1(k) = -\frac{\sqrt{\epsilon_\rho \epsilon_z} - \tanh(kh\sqrt{\frac{\epsilon_\rho}{\epsilon_z}})}{\sqrt{\epsilon_\rho \epsilon_z} + \tanh(kh\sqrt{\frac{\epsilon_\rho}{\epsilon_z}})} e^{-kz_n} \tag{11.23}$$

For the simple case of a metallic tip in front of a semi-infinite anisotropic media the electric fields around the tip (and the corresponding tip-capacitance) corresponds to those generated by an effective homogeneous isotropic media with an effective dielectric constant $\epsilon = \sqrt{\epsilon_\rho \epsilon_z}$ (see Fig.11.5). However, inside the substrate the potential is deformed along the z coordinate.

So finally, materials with anisotropy can be easily modelled with the GICM. The explicit calculations involve the same effort as before. This can be specifically important in modelling fields in PFM [50].

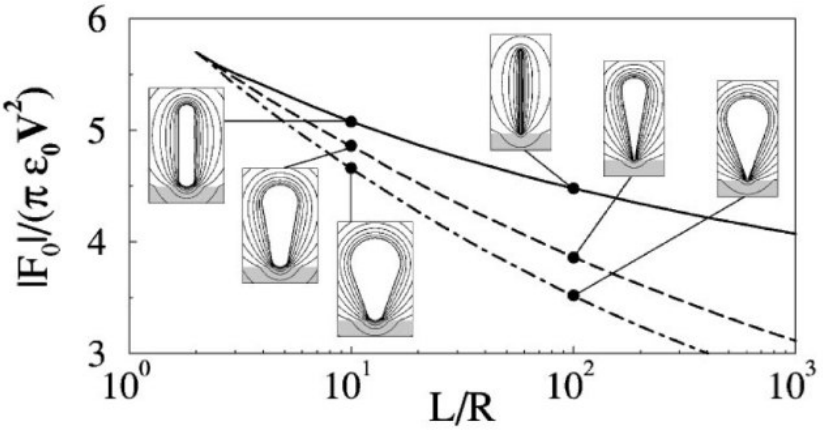


Figure 11.6: Absolute value of the contact force $|F_0|$ between a metallic tip and a plane dielectric sample ($\epsilon = 5.4$) vs the ratio between tip length and radius L/R for different tip angles $\theta = 0^o, 10^o, 20^o$ ($\theta = 0^o$ corresponds to a nanotube-like tip). [After Ref. [15]]

11.3 Electrostatic Force Microscopy

The analysis of electrostatic measurements based on Force Microscopy involves a lot of different elements, most of them at the macroscopic scale. The long range nature of the electrostatic force implies that all of them should be taken into account for a realistic simulation. The GICM has been used to analyze the influence of the macroscopic shape of the tip [15] as well as of the cantilever [17] for both metallic and dielectric samples.

The analysis of electrostatic forces showed that for metallic samples the force law, except for a constant background, only depends on the tip radius of curvature [15]. As a matter of fact, it is possible to determine the actual tip-radius by measuring the electrostatic force-distance curves on a metallic surface [17]. In contrast, for dielectric samples the forces depend on the overall geometry of the tip. Interestingly, it was found [15] that the contact adhesion force does not depend on the tip size and is bound by a simple expression which only depends on the applied bias and the sample dielectric permitivity (see Fig. 11.6).

The effect of a metallic cantilever can be included in the simulation by another series of image charges. Although the contribution of the cantilever can be neglected for metallic samples, calculations based on GICM showed that it can be very important for dielectric films [16]. By using appropriate image-charge ideas, it is possible to analyze the electrostatic problem of small metallic clusters or nanotubes absorved in dielectric samples. Modelling the electrostatic force with the GICM it was possible to predict *quantitatively* the electrostatic signal of metallic objects. A very strong agreement has been obtained between the theory and experimental results of connected and isolated single wall nanotubes [6](see Fig. 11.7).

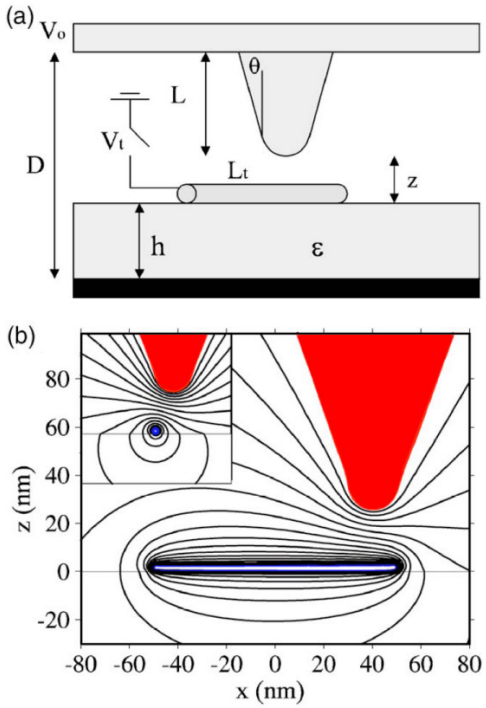


Figure 11.7: Scheme of a typical electrostatic force microscope. (b) Equipotential distribution for a standard tip-sample system using the GICM method. A metallic nanowire has been included in the simulation. Inset shows a 90^o rotation of the main image. [After Ref. [6]]

11.4 Conclusions

We have presented a method able to calculate the electrostatics of a biased tip probe in SPM setups. The main advantage of this method is the low computational cost that implies the inclusion of different orders of magnitude in length scales characteristic of SPM. This allowed us to calculate electrostatic potential and fields, capacitance and force versus distance curves from large tip-sample distances down to contact for different tip geometries. The method can be easily extended to deal with thin films, multilayered and anisotropic samples which could be very useful for local measurements of thin film polarization and thickness.

Bibliography

[1] C.D. Bugg and P.J. King, J. Phys. E **21**, 147 (1988); C.C. Williams, W.P. Hough and A. Rishton, Appl. Phys. Let. **55**, 203 (1989).

[2] Y. Martin, D. W. Abraham and H.K. Wickramasinghe, Appl. Phys. Lett. **52**, 1103 (1988); M. Nonnenmacher, M. O'Boyle and H.K. Wickramasinghe, Ultramicroscopy **42-44**, 268 (1992); T. Hochwitz, H.K. Henning, C. Leveg, C. Daghlian and J. Slinkman, J. Vac Sci. Technol B **14**, 457 (1996); P. Girard, A. N. Titkov, A. Ramonda, V. P. Evtickhiev, and V. P. Ulin, Appl. Surf. Sci. 201, 1 (2002); A. Gil, J. Colchero, J. Gomez-Herrero, and A. M. Baro, Nanotechnology 14, 332 (2003).

[3] C.C. Williams, J. Slinkman, W.P. Hough and H.K. Wickramasinghe, Appl. Phys. Lett. **55**, 1662 (1989); R.C. Barrett and C.F. Quate, Ultramicroscopy **42-44**, 262 (1992).

[4] R. Erlandsson, G.M. McClelland, C.M. Mate and S. Chiang, J. Vac. Sci. Technol A, **6**, 266 (1988); S. Lányi, J. Török and P. Rehurek, J. Vac. Sci. Technol B **14**, 892 (1996).

[5] C. Gómez-Navarro, A. Gil, M. Alvarez, P. J. de Pablo, F. Moreno-Herrero, I. Horcas, R. Fernández-Sánchez, J. Colchero, J. Gómez-Herrero, and A. M. Baro, Nanotechnology **13**, 314 (2002)

[6] G. M. Sacha, C. Gómez-Navarro, J.J. Sáenz and J. Gómez-Herrero Appl. Phys. Lett. **89**, 173122 (2006).

[7] J.E. Stern, B.D. Terris, H.J. Mamin and D. Rugar, Appl. Phys. Lett. **53**, 2717 (1998); B.D. Terris, J.E. Stern, D. Rugar and H.J. Mamin, Phys. Rev. Lett. **63**, 2669 (1989); C. Schonenberger and S. Alvarado, Phys. Rev. Lett. **65**, 3162 (1990).

[8] Y. Martin and H.K. Wickramasinghe, Appl. Phys. Lett. **50**, 1455 (1987); J.J. Sáenz, N. García, P. Grütter, E. Meyer, H. Heinzelmann, R. Wiesendanger, L. Rosenthaler, H.R. Hidber and H.-J. Güntherodt, J. Appl. Phys. **62**, 4293 (1987).

[9] F. Saurenbach and B. Terris, Appl. Phys. Lett. **56**, 1703 (1990); R. Lüthi, H. Haefke, K.-P. Meyer, E. Meyer, L. Howald and H.-J. Güntherodt, J. Appl. Phys. **74** , 7461 (1992); T. Hidaka, T. Maruyama, M. Saitoh, N. Mikoshiba, M. Shimizu, T. Shiosaki, L.A. Wills, R. Hiskes, S.D. Dicarolis and J. Amano, Appl. Phys. Lett. **68**, 2358 (1996); H. Bluhm, A. Wadas, R. Wiesendanger, K.-P. Meyer and L. Szcześniak, Phys. Rev. B **55**, 4 (1997).

[10] D. Gekhtman, Z.B. Zhang and G. Dresselhaus, Phys. Rev. Lett. **82**, 3887 (1999).

[11] J. Hu, X.D. Xiao, D.F. Ogletree and M. Salmeron, Science **268**, 267 (1995); J. Hu, X.D. Xiao and M. Salmeron, Appl. Phys. Lett. **67**, 476 (1995); A. Verdaguer, G. M. Sacha, H. Bluhm, and M. Salmeron, Chem. Rev. **106**, 1478 (2006).

[12] S. Gómez-Moivas, J. J. Sáenz, M. Calleja, and R. García, Phys. Rev. Lett. 91, 056101 (2003); G. M. Sacha, A. Verdaguer, and M. Salmeron, J. Phys. Chem. B 110, 14873 (2006).

[13] S. Gómez-Moñivas, J.J. Sáenz, R. Carminati and J.J. Greffet, Appl. Phys. Lett. **76**, 2955 (2000).

[14] H.W. Hao, A.M. Baró and J.J. Sáenz, J. Vac. Sci. Technol B **9**, 1323 (1991); S. Watanabe, K. Hane, T. Ohye, M. Ito and T. Goto, J. Vac. Sci. Technol. B **11**, 1774 (1993); S. Belaidi, P. Girard and G. Leveque, J. Appl. Phys. **81**, 1023 (1997); C. Girard, C. Joachim, C. Chavy and P. Sautet, Surf. Sci. **282**, 400 (1993)

[15] S. Gómez-Moñivas, L. S. Froufe, A. J. Caamaño, and J. J. Sáenz, Appl. Phys. Lett. **79**, 4048 (2001).

[16] G.M. Sacha and J.J. Sáenz, Appl. Phys. Lett. **85**, 2610 (2004)

[17] G. M. Sacha, A. Verdaguer, J. Martínez, J. J. Sáenz, D. F. Ogletree, and M. Salmeron, Appl. Phys. Lett. **86**, 123101 (2005).

[18] E. W. Müller and T. T. Tsong, *Field Ion Microscopy* (Elsevier Inc, NY, 1969).

[19] R. Gomer, *Field Emission and Field Ionization* (Harvard University Press, Cambridge, MA, 1961).

[20] E. Kasper, in *Adv. in Optical and Electron Microscopy*, vol. 8, p. 207, ed. by R. Barber and V. E. Cosslett, (Academic Press 1982) and references therein.

[21] C. A. Spindt, I. Brodie, L. Humphrey and E. R. Westerberg, J. Appl. Phys. **47**, 5284 (1976).

[22] H.-W. Fink, IBM J. Res. Develop. **30**, 460 (1986); Phys. Scr. **38**, 260 (1988).

[23] Vu Thien Binh, J. Microsc. **152**, 355 (1988); Vu Thien Binh and N. García, J. Phys. I (France) **1**, 605 (1991); Vu Thien Binh, S. T. Purcell, N. García and J. Doglioni, Phys. Rev. Lett **69**, 2527 (1992).

[24] W. Zhu, *Vacuum Micro-Electronics* (Wiley, New York, 2001).

[25] D. V. Scheible, C. Weiss, J. P. Kotthaus and R. H. Blick, Phys. Rev. Lett. **93**, 186801 (2004).

[26] J. J. Sáenz, N. García, Vu Thien Binh and H. de Raedt, in *"Scanning Tunneling Microscopy (STM) and Related Methods"*, ed. by R. J. Behm, N. García and H. Rohrer, NATO ASI Series E: Applied Sciences, Vol. 184, p. 215 (Kluwer, Dordrecht 1990); J. Phys. C (France) **8**, 73 (1991).

[27] P. A. Serena, L. Escapa, J. J. Sáenz, N. García and H. Rohrer, J. of Microscopy **152**, 43 (1988).

[28] D. Atlan, G. Gardet, Vu Thien Binh, N. García and J. J. Sáenz, Ultramicroscopy, **42**, 154 (1992).

[29] M. R. Scheinfein, W. Qian and J. C. H. Spence, J. Appl. Phys. **73**, 2057 (1993); W. Qian, M. R. Scheinfein and J. C. H. Spence, J. Appl. Phys. **73** 7041 (1993).

[30] M. L. Yu, B. W. Hussey, H.-S. Kim and T. H. P. Chang, J. Vac. Sci. Technol. B **12**, 3431 (1994).

[31] *Nanosources and Manipulations of Atoms under High Fields and Temperatures: Applications*, NATO ASI Series E: Applied Sciences, Vol. 235 edited by Vu Thien Binh, N. García and K. Dransfeld (Kluwer, Dordrecht, The Netherlands, 1993).

[32] *Atomic and Nanometer-Scale Modification of Materials: Fundamentals and Applications* NATO ASI Series E: Applied Sciences, Vol. 239 edited by P. Avouris (Kluwer, Dordrecht, The Netherlands, 1993).

[33] W. Stocker, H.-W. Fink and R. Morin, Ultramicroscopy **31**, 379 (1989); H.-W. Fink, W. Stocker and H. Schmid, Phys. Rev. Lett. **65**, 1204 (1990).

[34] Vu Thien Binh, V. Semet and N. García, Appl. Phys. Lett. **65**, 2493 (1994).

[35] L. P. Murray, U. Staufer, E. Bassous, D. P. Kern and T. H. P. Chang, J. Vac. Sci. Technol. B **9**, 2955 (1991).

[36] J. J. Sáenz and R. García, Appl. Phys. Lett. **65**, 3022 (1994).

[37] R. Young, J. Ward and F. Scire, Rev. Sci. Inst. **43**, 999 (1972).

[38] J. He, N. M. Miskovsky and P. H. Cutler, Appl Phys. Lett **59**, 1 (1991).

[39] G. Mesa, E. Dobado-Fuentes and J. J. Sáenz, J. Appl. Phys. **79**, 39 (1996).

[40] G. Mesa and J. J. Sáenz, Appl. Phys. Lett. **69**, 1169 (1996); S. Gómez-Moñivas, L. S. Froufe, R. Carminati, J. J. Greffet, and J. J. Sáenz, Nanotechnology **12**, 496 (2001).

[41] G. M. Sacha, E. Sahagún and J. J. Sáenz J. Appl. Phys. **101**, 024310 (2007).

[42] L. J. Whitman, J. A. Stroscio, R. A. Dragoset and R. J. Celotta, Science **251**, 1206 (1991).

[43] D. R. Cruise, J. Appl. Phys. **34**, 3477 (1963).

[44] A. Renau, F. H. Read and J. N. H. Brunt, J. Phys. E: Sci. Instrum. **15**, 347 (1982).

[45] G. Martínez and M. Sancho, Am. J. Phys. **51**, 170 (1983).

[46] S. L. Carnie, D. Y. C. Chan, and J. Stankovich, J. Colloid Interface Sci. **165**, 116 (1994); Z. H. Quiao, Z. L. Li, and T. Tang, Proc. Symp. Appl. Math. **24**, 252 (2006).

[47] See for example, W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes. The Art of Scientific Computing*, Chap. 15, (Cambridge University Press, 1992).

[48] R. W. Smythe, *Static and dynamic electricity* (McGraw-Hill, New York, 1968).

[49] A. N. Morozovska, S. V. Svechnikov, E. A. Eliseev and S. V. Kalinin, Phys. Rev. B **76**, 054123 (2007) and references therein.

[50] S. V. Kalinin, A. Rar, and S. Jesse, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 2226 (2006); A. Gruverman and A. Kholkin, Rep. Prog. Phys. 69, 2443 (2006).

[51] L. D. Landau, E. M. Lifshitz, L. P. Pitaevskii, *Electrodynamics of Continuous Media* (Pergamon Press, Oxford, 1984).

Conclusion

We have presented a selection of topics that are currently relevant in the field of multi-scale modeling for nanoscale devices. Bridges still need to be built between some of the approaches, in order to create a seamless hierarchy capable of handling the whole spectrum from complex circuits down to single molecules. Even at the circuit level, a need is developing for approaches that are completely new with to respect to those used for classical devices, due to the unavoidable interactions between nanoscale devices and interconnects and to the possibility of introducing device concepts and architectures that go beyond the three-terminal device paradigm.

On the other hand, modeling tools developed for nanostructures, including atomistic simulation schemes, are acquiring importance also for the analysis of downscaled CMOS devices, which are now rapidly approaching the few nanometer size and start exhibiting size quantization effects and single-electron charging phenomena.

Therefore the long term objective should consist in the creation of a suite of tools capable of handling all types of nanoscale devices and circuits (independent of the operating principle, quantum or classical) and to make their usage accessible also to nonspecialists of the single fields. We believe that, in order to reach such a target, a well structured effort is needed, aimed at coordinating and addressing the efforts of the many groups active in nanoscale modeling and at avoiding unnecessary duplications that have often occurred in the past, with several groups developing complex codes with very similar purposes.

Overall, a mature hierarchy of simulation tools for nanoelectronics, accessible to academia and industry, will be instrumental in the development of a host of applications for the properties of nanoscale structures that have been uncovered in the last decades. Availability of quantitatively reliable modeling tools represents a necessary step to unleash the full power offered by nanotechnology and to trigger innovative ideas for the implementation of new device concepts.

Edited in collaboration with

http://www.nanoict.org http://www.m4nano.com http://www.cea.fr

Cover picture

Silicon nanowire with rough surface and superimposed charge density

Courtesy: Aurélien Lherbier and V_sim project (http://www-drfmc.cea.fr/sp2m/L_Sim/V_Sim/)