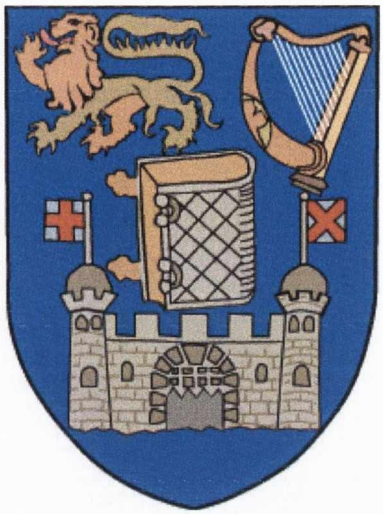


Computational methods for electron transport and their application in nanodevices

By

Ivan Rungger

A thesis submitted for the degree of
Doctor of Philosophy

School of Physics
Trinity College Dublin

October 2008

Declaration

I, Ivan Rungger, hereby declare that this dissertation has not been submitted as an exercise for a degree at this or any other University.

It comprises work performed entirely by myself during the course of my Ph.D. studies at Trinity College Dublin. I was involved in a number of collaborations, and where it is appropriate my collaborators are acknowledged for their contributions.

A copy of this thesis may be lended or copied by the Trinity College Library upon request by a third party provided it spans single copies made for study purposes only, subject to normal conditions of acknowledgement.

Abstract

The present thesis deals with the development of theoretical and computational tools for the first principles study of ground state and electronic transport properties of nanoscale devices and the application of these tools to systems of current interest. The ground state properties are studied within density functional theory (DFT) using the SIESTA code, whereas the transport properties are investigated using the non-equilibrium Green's functions (NEGF) formalism implemented in the SMEAGOL code. This is itself is based on SIESTA. We present our implementation a new algorithm for the calculation of the self-energies for quasi one-dimensional systems. The main advantage of this method is that all the singularities in the computation are avoided wherever possible, so that it is very stable and accurate. We also present a formalism for the inclusion of bound states in the calculation of the non-equilibrium charge density within the NEGF method, which we also use to treat systems with very weakly coupled states. Based on this formalism an adaptive energy-mesh scheme for the integration over energy of the density of states and transmission is implemented.

We first use ground state DFT to study the magneto-structural phase transitions of MnAs. The magnetic interactions are calculated by mapping DFT total energies onto a Heisenberg model, and the Curie temperature is obtained within a mean field approximation. We explain the anomalous features in the phase diagram and in the temperature dependence of the susceptibility of MnAs with the strong dependence of the magnetic exchange interactions on the structural parameters.

We then move on to the investigation of electronic transport through nano-devices. The I-V characteristics of a molecular junction, consisting of a Mn_{12} single molecule magnet weakly coupled to two gold surfaces, are calculated for two different spin configurations of the Mn_{12} center. The I-V curves for the two states are found to be similar, although there are small quantitative differences which might be detectable in experiments.

Finally a detailed study of the electronic transport properties of Fe/MgO/Fe(100) tunnel junctions is presented. It is shown that for such systems at finite voltage the correct occupation of the bound states forming at the Fe/MgO interfaces is crucial in

order to obtain a physically meaningful solution. Two features are found to determine the I-V characteristics: 1) interface states, and 2) Fe band-edges of the states with high transmission. For a junction with a barrier thickness of 4 MgO monolayers two bias regions are identified. The low bias region is characterized by a large contribution to the current from sharp resonant interface states, resulting in a non-monotonic behavior of the TMR as function of bias voltage, whereas the TMR decreases continuously with bias in the high bias regime. If the contribution to the current from the sharp interface resonances is removed, the TMR decreases monotonically with bias for all voltages. We then perform a detailed study of the dependence of the zero bias transport properties on the thickness of the MgO barrier. We further show that oxygen vacancies in the MgO barrier can drastically reduce the TMR, due to a depolarization of the density of states at the vacancy site. We also investigate double MgO barrier junctions, separated by a thin Fe layer, and find evidence in the I-V curves of quantum well states formed in the middle Fe layer. In these junctions the decay of the TMR with bias is slower than for a single barrier junction, which we explain by the relative shift in energy of the density of states in the Fe electrodes with bias.

Acknowledgments

During the course of my postgraduate studies I received a lot of help and support from a large number of people, without whom none of this research would have been possible.

First of all, I would like to thank my supervisor Prof. Stefano Sanvito. When I joined his group he had only two other students and one postdoc, based in the arches underneath the railway. Now, his group is one of the largest in the School of Physics, and he has established himself as one of the leaders in the field of computational materials science and molecular electronics. I have benefited greatly from his advice and teaching during my time in his research group.

Secondly, I would like to thank our external collaborators. Prof. Fernando Iikawa, from the Instituto de Fisica Gleb Wataghin UNICAMP, Brazil, pointed us to the puzzling phase diagram of MnAs, and provided valuable experimental data for the understanding of the lattice distortion effects on the phase transitions of MnAs. For the calculations on the Fe/MgO tunnel junctions our collaborator Oleg Mryasov greatly helped in the understanding of the physics of such systems. Jeronimo Peralta Ramos and his supervisor A. M. Llois pointed us to the fascinating system of the double barrier junctions, and greatly helped us in the understanding of the physical properties. I would also like to thank Plamen Stamenov, a former postgraduate student and now postdoc here in TCD, for the help in understanding the experimental properties of Fe/MgO tunnel junctions.

As I mentioned above, Stefano's group has grown large, with a current total of ten postgraduate students and three postdocs, as well as former postgraduate students and two former postdocs. However, there are a few former and current group members whose assistance was vital to the work performed in this thesis. Alex Rocha was Stefano's very first student, and he did the bulk of the work of developing SMEAGOL, in collaboration with groups in Lancaster (England), and Oviedo (Spain). He has since graduated and is now a researcher in his native Brazil. He provided a lot of assistance during the three and a half years in which we were both in Stefano's group, and we developed some new additions to the SMEAGOL code together. Chaitanya Das

Pemmaraju, a former student in the group, and now a postdoc, was responsible for implementing ASIC in SIESTA, Cormac Toher, a former student and now a postdoc in Germany, investigated the effects of the ASIC and of the derivative discontinuity on the electronic transport. Maria Stamenova, who is now also finishing her Ph.D., studies the time-dependent transport. The discussions about physics with them were always very fruitful. I would also like to acknowledge the postdoc Nadjib Baadji for the discussions about transport in tunnel junctions, and the postgraduate student Andrea Droghetti for the discussions about defects in MgO. Miguel Afonso Oliveira was a postdoc in our group for the best part of three years, during which time he served as our system administrator, along with Tom Archer. They were both of great help in solving my computer problems. I have also had many useful discussions with Tom about density functional theory.

Other members of our group include Ruairi Hanafin, Akinlolu Akande, Durga Sankar Kusanakurthi, and Nuala Caffrey. Former members of our group who have not yet been mentioned are Nicola Jean and William Lee, who still return to TCD on occasion.

It would not be possible to undertake research such as this without funding. The Science Foundation of Ireland (Grants SFI02/IN1/I175 and No. SFI07/ RFP/PHY235) paid my fees and maintenance, and also provided support for attending conferences. The calculations presented in this thesis were performed in the Trinity Centre for High Performance Computing (TCHPC) and the Irish Centre for High End Computing. The staff from TCHPC, particularly Jimmy Tang, Bob Crosbie, Dermot Frost, Geoff Bradley and Jose Redojo also provided technical assistance both on their cluster and visualization suite, as well as on our own internal network.

Members from other groups who we have shared office space with include Andy Wall and Dave Kirwan (who were also both in my class during my undergraduate degree), Claudia Gomes Rocha, Andrew Rowan, and Eireann Cosgriff.

Finally, I thank my wife Anita, my two kids Juna and Moritz, and my parents Ida and Paul for their support throughout these years.

Contents

Chapter 1

General Introduction

The discovery of the giant magnetoresistance (GMR) in 1988 by the groups of P. Grünberg [1] and A. Fert [2] triggered large interest in devices combining electronic and magnetic degrees of freedom. Over the last two decades there has been a rapid increase of the activity in this area of research, in which spin-polarized electronic transport phenomena in metals, semiconductors and molecules are studied. This is now commonly referred to as spin electronics, or also spintronics [3, 4, 5, 6, 7]. Due to its rather broad definition, the field of spintronics covers a large area of research [8].

The prototype spintronics devices are the spin-valves. A typical spin-valve consists of two electrodes that are ferromagnetic metals, separated by a usually thin spacer, as schematically illustrated in Fig. 1.1. The resistance of such a device is sensitive to the mutual orientation of the magnetizations of the electrodes. It is usually small when the magnetizations of the contacts are parallel (P) with respect to each other [Fig. 1.1(a)], but it rises when the two magnetizations are antiparallel (AP) [Fig. 1.1(b)]. The magnetization of one of the ferromagnets is usually pinned, while that of the other is free to rotate, so that it aligns to an externally applied field. A reversal of the magnetic field therefore results in a large change in the resistance of the device. The general dependence of the resistance of a device on the magnetic field is called magnetoresistance (MR). The term giant magnetoresistance was chosen by Fert's group in Ref. [2] since their measured MR was much larger then that of other known systems. A GMR spin-valve is therefore an efficient detector of the orientation of a magnetic field. Read heads for magnetic data storage devices based on this idea are now used in most commercial applications.

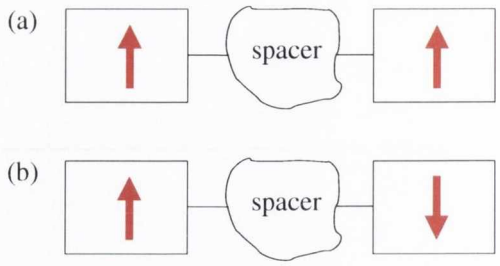


Figure 1.1: Schematic representation of a spin-valve device with parallel alignment of the magnetic moments of the electrodes (a), and with antiparallel alignment (b). The arrows indicate the direction of the magnetizations of the ferromagnetic electrodes.

A measure for the sensitivity of a spin-valve is the MR ratio. For a fixed applied bias voltage it is defined as

$$\mathrm{MR} = \frac{I_{\mathrm{P}} - I_{\mathrm{AP}}}{I_{\mathrm{AP}}}, \tag{1.1}$$

where I_{P} is the current for P alignment, and I_{AP} is the current for the AP. This definition is sometimes referred to as the "optimistic MR", since the values become very large if $I_{\mathrm{AP}} \ll I_{\mathrm{P}}$.

The GMR is found in layered metallic structures, and the mechanisms leading to the GMR effect are well studied. Extensive reviews are given in Refs. [9, 10, 11, 12]. The basic models for the GMR are based on the assumptions that the spins are in a collinear state, so that in each of the ferromagnets one can define the majority spins and the minority spins, and that there is no spin-flip scattering in the junction ("two spin-fluid approximation"). The transmission through a spin-valve then depends on the degree of matching of the Fermi surfaces of the different layers. Since this is usually better for P alignment than for AP alignment, a large MR ratio is found.

Tunnel junctions

Although in the GMR spin-valves the spacer consists of layered metallic materials, over the years many different types of spacers have been used, such as insulators [13, 14, 12, 15, 16], semiconductors [5, 6] or semi-metals. Currently the most relevant devices for applications are the magnetic tunnel junctions, where the spacer is insulating. Whereas in GMR junctions the current measurements can be performed both in the plane of the junction and out of plane, in tunnel junctions the resistance is always measured across the barrier. In practical applications of tunnel junctions, such as sensors or hard-disk drives read heads, there are three important quantities. The first is the zero-bias tunneling magnetoresistance ratio (TMR), the second is the resistance area product of the junction, and the third is the voltage, $V_{1/2}$, at which the TMR

decreases to half its value at zero bias. In order to maximize the signal in a device it is ideal if the zero-bias TMR and $V_{1/2}$ are as large as possible, and if the resistance area product is low. In the first generation of tunnel junctions usually amorphous Al_2O_3 was used as spacer [13]. These junctions however never reached room temperature TMR in excess of 70% [17]. In the new generation of TMR devices crystalline spacers such as MgO are used [15, 16], which can reach TMR values as large as 604% at room temperature and 1144% at 5 K [18, 19]. Such crystalline junctions are not only interesting for device applications, but also from a fundamental point of view. Although the basic mechanism behind the high TMR is known [20, 21, 15, 16], there are still large discrepancies between the theoretical predicted TMR values for such junctions and the experimentally measured values. An understanding of the detailed bias dependence of the current and of the TMR is still lacking, especially with respect to the quantitative influence of the interface structure, of the role of defects in the MgO, of disorder, of inelastic effects, and of the electrodes band structure. At present this is an active area of research and in this work we investigate some of these aspects.

Molecular spintronics

The advances in the field of organic and molecular electronics also opened the possibility to use single molecules or layers of organic materials as spacers. MR ratios up to 40% at a temperature of 11 K have been measured for spacers composed of organic monolayers [22, 23]. Also quasi one-dimensional systems such as carbon nanotubes have been used as spacers [24, 25].

A completely different type of spin-valve can be realized by using non-magnetic electrodes, separated by a magnetic spacer. Such devices could for example be used as memory cells, if the spacer can be stabilized in different magnetic states, and if these can be inferred by an electric readout such as a resistance measurement. In the field of molecular spintronics the single molecule magnets are a potential candidate for such a device [26, 27]. These consist of a central magnetic core of transition metals and ligand ions, which is stabilized by attaching organic functional groups. Magnetic hysteresis has been measured in these molecules at temperatures of the order of a few K, which has been attributed to the intra-molecular magnetic anisotropy [27]. The most investigated of the single molecule magnets are the Mn_{12} based molecules, since they have a large ground state magnetic moment of 20 μ_B and a high magnetic anisotropy [27]. Current vs. voltage curves for individual Mn_{12} based molecules have been obtained using scanning tunneling spectroscopy [28, 29, 30]. It is however still an open question whether the magnetic state of the molecule can be determined

by a readout of the current. This requires the conductance to change significantly for different magnetic states. In this work we address this question with *ab initio* methods.

Spin-injection

The use of spin-valve devices is mainly restricted to spin and magnetic field detectors. The integration of spin and electron transport phenomena can however also be used in different applications. One of the ideas is to develop spin-controlled logic devices [31]. A prerequisite for such devices is the ability to inject spin-polarized currents into non-magnetic semiconductors from ferromagnetic carrier sources. For metal-semiconductor contacts however the spin-injection in the diffusive transport regime is rather low due to the resistance mismatch between the two materials [32]. One way to overcome this problem is to use diluted magnetic semiconductors (DMS) like $In_{1-x}Mn_xAs$ and $Ga_{1-x}Mn_xAs$ [33, 34] as spin injectors [35, 36]. The main drawback in the use of these materials in technological applications, however, is their rather low Curie temperature, which to date does not exceed 200 K [37, 38]. The study of possible candidates for room temperature ferromagnetic DMS is currently an active area of research [39, 40, 41, 42, 43]. Another way to overcome the resistance mismatch problem is to insert a thin insulating layer between the metal and the semiconductor. In this case high spin-injection rates have been measured using spin-filtering tunneling barriers such as MgO [44, 45] or Al_2O_3 [46].

It is however also possible to inject spins directly from the metal to the semiconductor if the material and interface parameters are optimized. Recently a very high injection efficiency of 50% has been obtained using the half metallic Heusler alloy Co_2FeSi as spin injector [47]. Another candidate for spin injection is the ferromagnetic metal MnAs, for which a spin injection efficiency of 6% was reported [48]. Bulk MnAs presents a first-order phase transition at 40^oC, where it changes from ferromagnetic/hexagonal to paramagnetic/orthorhombic [49]. This magneto-structural phase transition has important implications for technological applications. While it is a major drawback for the use of MnAs as spin injecting material at room temperature, the magneto-elastic effects related to the phase transition are useful for transducers [50], and the magneto-caloric properties are interesting for developing refrigeration devices [51]. In this work we therefore investigate the origin of the phase transitions of MnAs using *ab initio* methods, in order to verify existing phenomenological models, and to present suggestions for possible improvements in the device design based on specific strain engeneering [52, 53].

Modeling of electron transport in nano-devices

Transport through nanoscale objects can not be described by ordinary "circuit analysis" based on Ohm-Kirchhoff's laws. This is because the mean free path of the electrons is usually comparable with the size of the device, so that quantum transport theory has to be applied. Reviews about models of classical- and quantum-transport and their applications in calculations are given in Refs. [54, 55, 56, 57]. The applicability of the different models depends on various system specific length scales [54, 55]. One of these is the elastic mean free path, which determines how far an electron can travel without change of momentum. If the device dimensions are smaller than the elastic mean free path, the momentum of the electrons is approximately conserved, and the transport is denoted as "ballistic". Another important quantity is the phase relaxation length, which is the average distance over which an electron can travel before its initial phase is destroyed. For devices smaller than the phase relaxation length there is quantum interference between the different possible paths an electron can take across a device, and the transport is denoted as coherent. If the device dimensions are larger than the phase relaxation length, then there is no long-distance interference, and the transport is diffusive. In crystalline magnetic tunnel junctions for example the barrier thickness is so small that the transport is largely phase coherent, whereas for amorphous Al_2O_3 junctions the phase coherence is largely lost due to dynamic scattering [16]. The spin diffusion length is another important length scale, and corresponds to the distance over which an electron keeps its spin memory. This is usually much longer than the phase relaxation length. Also the screening length is an important factor in electronic transport, since it determines after which distance the potential of a localized charge is screened by the surrounding electrons. This distance is short in metals, but becomes very long in semiconductors and insulators [55]. Effects of impurities are therefore felt at a long distance in semiconductors, but only a rather short distance in metals. Another important quantity for spintronics devices is the spin polarization of a device. We note however that different definitions of the polarization are relevant depending whether the electron transport is in the diffusive or in the ballistic regime [58, 55]. In this work we always assume that the transport is in the ballistic regime, which implies that we neglect all contributions from inelastic effects. We will however discuss the expected contributions from inelastic effects where necessary.

Quantum transport can be calculated in its time dependent evolution [59] or in the steady state [54], which is the approach taken in this work. For the case of non-interacting single-particle Hamiltonians most methods are based on the Landauer-

Büttiker scattering formalism [60], which connects transport properties directly to the electronic structure. One of these methods is the non-equilibrium Green's functions (NEGF) formalism [54, 61, 62, 63], which allows also to calculate the charge density at an applied bias voltage. In principle also inelastic processes can be included in the NEGF formalism [54]. Alternative methods are based on matching wave-functions [64, 65].

A detailed description of the electronic structure of a device is crucial for making quantitative transport calculations. To date the most used method for computing electronic, magnetic and structural properties of materials from first principles (i.e. without using parameters fitted to experiments) is density functional theory (DFT) [66, 67]. A number of numerical implementations exist to calculate the electron transport properties of materials based on DFT [61, 62, 63, 64, 65]. In this work we use the *ab initio* electron transport code SMEAGOL [61, 68], which was developed in a joint effort between S. Sanvito and A. R. Rocha in Trinity College Dublin, J. Ferrer and V. M. Garcia-Suarez in the Universidad de Oviedo, and C. J. Lambert and S. Bailey in the University of Lancaster. In SMEAGOL the DFT Hamiltonian is obtained through an interface with the DFT code SIESTA [69]. SIESTA uses a very efficient localized atomic orbital basis set, which makes possible to treat systems with a large number of atoms in the unit cell. Moreover it generates a self-consistent tight-binding-like Hamiltonian, which can be readily used in the transport technique. In this work we present some of the extensions included in the SMEAGOL code, which increase its stability and enable the calculation of the current vs. voltage characteristics for a new set of physical systems, such as tunneling junctions and systems with weakly coupled states.

1.1 Dissertation Layout

The main focus of this work is the development of theoretical and computational tools for the *ab initio* study of ground state and electronic transport properties of materials, and the application of these tools to systems of current interest. The ground state properties are studied within density functional theory (DFT) using the SIESTA code, whereas the transport properties are investigated using the non-equilibrium Green's functions (NEGF) formalism implemented in SMEAGOL, which itself is based on SIESTA. We present our implementation in the SMEAGOL code of an improved algorithm for the calculation of the self-energies for quasi one-dimensional systems, and of a formalism for the inclusion of completely localized states in the NEGF method. These tools are used to determine current vs. voltage (I-V) curves for electronic transport through Mn_{12} based magnetic molecules and across Fe/MgO(100) single and double barrier magnetic tunnel junctions. A study of the magneto-structural properties of MnAs is presented, where the magnetic interactions are calculated by mapping DFT total energies onto a Heisenberg model.

In chapter 2 we introduce the many-electron problem and give an overview of DFT and its implementation in the SIESTA code. We present the Hohenberg-Kohn and the Kohn-Sham theorems, and give and overview of the existing exchange-correlation functionals.

In chapter 3 we investigate the magneto-structural phase transitions of MnAs between the ferromagnetic hexagonal $B8_1$ structure and the paramagnetic and antiferromagnetic orthorhombic B31 structure. We calculate the magnetic exchange coupling parameters by mapping the DFT total energies for different configurations of the Mn magnetic moments onto a Heisenberg Hamiltonian. This allows us to extract the Curie temperature within a mean field approximation. Based on our *ab initio* calculations we give a model for description of the experimental magneto-structural phase diagram of MnAs, and compare the results to experiments and to the various phenomenological models based on the original idea of Bean and Rodbell [49].

Chapter 4 presents the NEGF formalism used for the quantum transport calculations. This formalism allows to calculate the charge density of a nano-device attached to metallic electrodes with different chemical potentials. The current flowing through the device can be calculated once its charge density is obtained self-consistently. The particular implementation of the NEGF concepts in the *ab initio* transport code SMEAGOL is discussed. A brief discussion of some of the limits of the use of the NEGF formalism in conjunction with DFT is also given.

In chapter 5 a stable and accurate method for the calculation of the self-energies

(SEs) of quasi one-dimensional systems and its implementation in SMEAGOL are outlined. The main advantage of the method is that all the singularities in the computation are avoided where possible, and that it involves no inversions of ill-conditioned matrices. The origin of localized surface states of semi-infinite systems is discussed, and their effect on the accuracy analyzed.

In chapter 6 a formalism for the inclusion of both very weakly coupled and bound states (BSs) in the calculation of the non-equilibrium charge density is outlined. The non-equilibrium occupation of the BSs is set based on the assumption of an effective coupling of these states to the leads. We also present an algorithm to generate an adaptive energy-mesh for the integration over energy of the density of states. In the second part of the chapter the developed methods are applied for the calculation of the transport through weakly coupled molecules. One of these systems consists of a Mn_{12} based magnetic molecule, which is very weakly coupled to gold electrodes through benzene groups. The main question we answer is whether it is possible to distinguish between two magnetic states of the molecule by performing an I-V measurement.

In chapter 7 we study the spin-dependent transport properties of Fe/MgO/Fe(100) tunnel junctions. The I-V curves for a junction of 4 MgO monolayers thickness are calculated and the resulting TMR-V curves presented. The results are discussed using a simple model, based on the relative positions in energy of interface states and band edges of high-transmission states. We also discuss the importance of a correct inclusion of the BSs forming at the Fe/MgO interfaces in the NEGF formalism at finite bias. We then perform a detailed study of the dependence of the zero bias transport properties on the thickness of the MgO barrier, and also investigate the effects of oxygen vacancies (V_O) in the MgO barrier on the I-V and TMR-V curves. In the last part of this chapter we study the transport properties of Fe/MgO/Fe/MgO/Fe(100) junctions, which have a double MgO barrier, separated by a thin Fe layer.

Finally, towards the end of the dissertation we present the general conclusions, and point to some questions related to the presented work that require further investigations. We also present possible future developments of the SMEAGOL transport code.

Chapter 2

Density functional theory and the SIESTA code

In this chapter an overview of the basic concepts of density functional theory (DFT) is given, followed by the description of the practical implementation of the DFT scheme for electronic structure calculations in the SIESTA code. A more detailed description of DFT can be found in a number of books on the subject [70, 71, 72, 73].

First the full many-body description of the ground state for a system of electrons and nuclei is given within the Born-Oppenheimer approximation. Then the basic DFT theorems of Hohenberg-Kohn and Kohn-Sham are presented, where the complex many-body electron system is replaced by an equivalent system of non-interacting particles moving in an effective potential. The various contributions to this effective Hamiltonian are discussed, and special attention is given to the approximations for the exchange-correlation potential.

In the second part the *ab initio* DFT code SIESTA is introduced. This is the basic computational tool used in this work. In SIESTA a linear combination of atomic orbitals is used as basis set, and the nuclei and core shell electrons are replaced by pseudopotentials. The mapping of the real space coordinates representation to the matrix representation used in SIESTA is presented, where the fundamental quantity is the density matrix.

2.1 Density functional theory

2.1.1 The many-electron system

The basic goal of electronic structure calculations is to describe the quantum-mechanical ground state properties of a system composed of N_e electrons and N_n nuclei. If the kinetic energy of the electrons is not too high, then the dynamics of such a system are governed by the time-dependent non-relativistic Schrödinger equation. Since the

mass of the electrons m_e is much smaller than those of the nuclei, the electrons dynamics happens at a time-scale much faster than that of the nuclei. Therefore the the Born-Oppenheimer approximation is commonly used [74, 72], in which the nuclei are assumed to be classical particles generating a static time-independent potential in which the electrons move, and where only the electrons are described quantum-mechanically. Since the aim is to describe the time-independent ground state properties, and the external potential generated by the nuclei is assumed to be constant in time, the system of N_e electrons can be described by the time-independent Schrödinger equation

$$\hat{H}\ \Psi = E\ \Psi. \tag{2.1}$$

Here $\Psi = \Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_e}, s_1, s_2, \ldots, s_{N_e})$ is the many-electron wave function (WF), $\hat{H}$ is the Hamiltonian operator and E is the energy. The vector $\mathbf{r}_i$ represents the coordinates of electron with index i and s_i is its spin. $\hat{H}$ depends on the coordinates of the nuclei and electrons, so that in general

$$\hat{H} = \hat{H}(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{N_n}; \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_e}, s_1, s_2, \ldots, s_{N_e}), \tag{2.2}$$

where the vector $\mathbf{R}_I$ represents the coordinates of nucleus with index I. The explicit form of $\hat{H}$ is given by

$$\hat{H} = \hat{T} + \hat{V}_{ne} + \hat{V}_{ee}, \tag{2.3}$$

where

$$\hat{T} = \sum_{i=1}^{N_e} -\frac{1}{2}\nabla_{\mathbf{r}_i}^2, \tag{2.4}$$

$$\hat{V}_{ne} = \sum_{i=1}^{N_e} v(\mathbf{r}_i), \tag{2.5}$$

$$\hat{V}_{ee} = \sum_{i<j}^{N_e} \frac{1}{r_{ij}}. \tag{2.6}$$

$\hat{T}$ is the kinetic energy operator of the electrons, $\hat{V}_{ne}$ is the electron-nucleus potential, and $\hat{V}_{ee}$ is the electron-electron interaction potential. Here $v(\mathbf{r}_i) = \sum_{I=1}^{N_n} 1/|\mathbf{r}_i - \mathbf{R}_I|$ and $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$. We note that the equations are given in atomic units throughout this work ($\hbar = m_e = e = 1$).

For a given solution of Eq. (2.1) the total energy W of the system of nuclei and electrons is

$$W = E + V_{nn}, \tag{2.7}$$

where the nucleus-nucleus energy $V_{\rm nn}$

$$V_{\rm nn} = \sum_{I<J} \frac{Z_I Z_J}{R_{IJ}} \tag{2.8}$$

has been added to the electron energy E. Here Z_I is the charge of nucleus I and $R_{IJ} = |\mathbf{R}_I - \mathbf{R}_J|$. $V_{\rm nn}$ is a constant for a given position of the nuclei, its value is however important for determining the ground state structure. Furthermore its derivative with respect to the atomic positions is needed for the calculation of the forces.

The physical interpretation of the many-electron fermionic WF Ψ is that the quantity

$$|\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_{\rm e}}, s_1, s_2, \ldots, s_{N_{\rm e}})|^2 \tag{2.9}$$

is the probability to find the electrons at the positions $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_{\rm e}}$ and with spins $s_1, s_2, \ldots, s_{N_{\rm e}}$. It is therefore normalized so that

$$N_{\rm e} = \sum_{s_1, s_2, \ldots, s_{N_{\rm e}}} \int |\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_{\rm e}}, s_1, s_2, \ldots, s_{N_{\rm e}})|^2 \, d\mathbf{r}_1 d\mathbf{r}_2 \ldots d\mathbf{r}_{N_{\rm e}}. \tag{2.10}$$

The probability to find an electron at a position $\mathbf{r}$ with spin s is given by

$$\rho(\mathbf{r}, s) = \sum_{s_2, \ldots, s_{N_{\rm e}}} \int |\Psi(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_{N_{\rm e}}, s, s_2, \ldots, s_{N_{\rm e}})|^2 \, d\mathbf{r}_2 \ldots d\mathbf{r}_{N_{\rm e}}. \tag{2.11}$$

This quantity is called the electron density. It is normalized to

$$N_{\rm e} = \sum_s \int \rho(\mathbf{r}, s) \, d\mathbf{r}. \tag{2.12}$$

The expectation value of an observable A corresponding to an operator $\hat{A}$ is then given by

$$A[\Psi] = \langle \hat{A} \rangle = \langle \Psi \, | \hat{A} | \, \Psi \rangle \,, \tag{2.13}$$

with

$$\begin{aligned} \langle \Psi | \hat{A} | \Psi \rangle \;\; = \;\; & \sum_{s_1, s_2, \ldots, s_{N_{\rm e}}} \int \Psi^*(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_{\rm e}}, s_1, s_2, \ldots, s_{N_{\rm e}}) \\ & \hat{A} \; \Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_{\rm e}}, s_1, s_2, \ldots, s_{N_{\rm e}}) d\mathbf{r}_1 d\mathbf{r}_2 \ldots d\mathbf{r}_{N_{\rm e}}. \end{aligned} \tag{2.14}$$

The notation $A[\Psi]$ explicitly indicates that the expectation value is a functional of the wave function Ψ.

The total energy E of the system corresponds to the expectation value of the Hamiltonian H

$$E[\Psi] = \langle \Psi | \hat{H} | \Psi \rangle, \tag{2.15}$$

where we have again explicitly denoted that E is a functional of Ψ. The total energy can be split up as

$$E[\Psi] = T[\Psi] + V_{\mathrm{ne}}[\Psi] + V_{\mathrm{ee}}[\Psi], \tag{2.16}$$

with the kinetic energy

$$T[\Psi] = \langle \Psi | \hat{T} | \Psi \rangle, \tag{2.17}$$

the electron-nucleus potential energy

$$V_{\mathrm{ne}}[\Psi] = \langle \Psi | \hat{V}_{\mathrm{ne}} | \Psi \rangle, \tag{2.18}$$

and the electron-electron potential energy

$$V_{\mathrm{ee}}[\Psi] = \langle \Psi | \hat{V}_{\mathrm{ee}} | \Psi \rangle. \tag{2.19}$$

In principle one can calculate Ψ and then all the ground state properties of the system. The main problem of the approach presented in this section however is that it is almost impossible to calculate the many-body WF for a large N_{e}. "Large" in this case means more than a few electrons. Different methods have been developed to obtain approximate solutions. In the Hartree-Fock method the WF is expanded as a Slater determinant of single-particle WFs (see Eq. 2.35), so that the many-electron Hamiltonian can be rewritten as a sum of single-particle Hamiltonians [70, 75]. The main problem of the Hartree-Fock approach is however that all the correlations between electrons are neglected, which for example causes major errors in calculations for metallic materials. In the "Configuration-Interaction" (CI) formalism the Hartree-Fock scheme is extended by expanding the WF over a set of different Slater determinants [76]. By using more determinants the correlations can be described accurately and highly accurate results can be obtained. The drawback however is that the computational cost becomes prohibitive for large systems, since the scaling with system size is typically of the order of N_e^6 or even higher. In the following sections the basic concepts of density functional theory (DFT) are outlined, which is another approach to make the many-electron problem tractable. DFT allows to work with a single-particle Hamiltonian like the Hartree-Fock method, and can therefore be used for large systems. The advantage over Hartree-Fock is that it also takes into account electron correlations, although this is done only in an approximate way.

2.1.2 Hohenberg-Kohn theorems

In the following two sections we assume that the particles are spinless, so that $\rho = \rho(\mathbf{r})$. The generalization to spin particles is briefly outlined in Sec. 2.1.4. If E_0 is the

lowest eigenvalue of Eq. (2.1) then for an arbitrary Ψ the following inequality holds

$$E_0 \leq E[\Psi]. \tag{2.20}$$

Minimization of $E[\Psi]$ with respect to all allowed N_e-electron Ψ therefore gives E_0

$$E_0 = \min_{\Psi} E[\Psi]. \tag{2.21}$$

If we write the electron-nucleus potential $\hat{V}_{ne}$ as a general external potential $V(\mathbf{r})$, then the equations above show that the ground state wave function, and consequently all the ground state properties, are uniquely determined by $V(\mathbf{r})$ and by the number of electrons N_e. Therefore also $\rho(\mathbf{r})$ of the ground state is uniquely determined by $V(\mathbf{r})$ and N_e.

The first theorem of Hohenberg and Kohn states that also the opposite is true, namely that "the external potential $V(\mathbf{r})$ is determined, within a trivial additive constant, by the ground state electron density $\rho(\mathbf{r})$" [70, 66]. Since $\rho(\mathbf{r})$ determines the potential $V(\mathbf{r})$ it implicitly also determines all the electronic properties of the system in the ground state. All quantities are therefore functionals of $\rho(\mathbf{r})$, so that we can write the kinetic energy as $T[\rho]$, the electron-nucleus potential energy as $V_{ne}[\rho]$ and the electron-electron potential energy as $V_{ee}[\rho]$. The total energy of the system $E_V[\rho]$ for a given potential $V(\mathbf{r})$ therefore is

$$E_V[\rho] = T[\rho] + V_{ne}[\rho] + V_{ee}[\rho]. \tag{2.22}$$

The big difference to the definitions given in the previous section is that now these quantities depend only on ρ, which is a function of a single variable $\mathbf{r}$, instead of being a functional of the full many-body WF which itself is a function of the N_e variables $\mathbf{r}_i$. Therefore the problem written in this way has been simplified by drastically reducing the degrees of freedom. However the explicit functional dependence of the observables on $\rho(\mathbf{r})$ is not known, and therefore rewriting them as functionals of ρ instead of Ψ is of no practical advantage at this point.

The second Hohenberg-Kohn theorem is analogous to the variational principle for WFs [Eq. (2.21)], where instead of the WF the electron density is used. It states that "for a trial density $\tilde{\rho}(\mathbf{r})$, such that $\tilde{\rho}(\mathbf{r}) \geq 0$ and $\int \tilde{\rho}(\mathbf{r})d\mathbf{r} = N_e$,

$$E_0 \leq E_V[\tilde{\rho}], \tag{2.23}$$

where $E_V[\tilde{\rho}]$ is the energy functional of Eq. (2.22)" [70, 66]. This shows that in principle one can obtain the ground state charge density and all the ground state properties of the system by finding the $\tilde{\rho}$ that minimizes $E_V[\tilde{\rho}]$. In practice however

this is difficult to solve, mainly due to the many constraints the trial charge density has to fulfill [70]. The Kohn-Sham scheme presented in the next chapter is a practical way to overcome these problems and to obtain $\rho(\mathbf{r})$.

2.1.3 Kohn-Sham equations

In order to be able to use above concepts in practical calculations Kohn and Sham have derived a procedure based on the solution of an equivalent non-interacting particle system, where the exact calculation of the kinetic energy and all the many-body effects are included in the so called exchange-correlation potential [67]. First the energy functional of Eq. (2.22) is rewritten in an equivalent form as

$$E_V[\rho] = T_s[\rho] + V_{ne}[\rho] + J[\rho] + E_{xc}[\rho], \tag{2.24}$$

where

$$E_{xc}[\rho] = T[\rho] - T_s[\rho] + V_{ee}[\rho] - J[\rho] \tag{2.25}$$

is the exchange-correlation (XC) energy,

$$J[\rho] = \frac{1}{2}\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \tag{2.26}$$

is the classical Coulomb repulsion energy for a charge density $\rho(\mathbf{r})$,

$$V_{ne}[\rho] = \int \rho(\mathbf{r})V(\mathbf{r})d\mathbf{r}, \tag{2.27}$$

and $T_s[\rho]$ is the kinetic energy of a reference system of non-interacting particles, which will now be introduced. The idea is that $T_s[\rho]$ is a good approximation for $T[\rho]$, so that $T_s[\rho] - T[\rho]$ is small. $E_{xc}[\rho]$ contains all the information about the many-body and non-classical contributions to the energy, and so it contains all the parts that are difficult to calculate exactly. The non-interacting reference system is defined by the following single-particle Schrödinger-like equation [70]

$$\hat{H}_{KS}\ \Psi_i(\mathbf{r}) = \epsilon_i\ \Psi_i(\mathbf{r}), \tag{2.28}$$

where the Kohn-Sham (KS) Hamiltonian $\hat{H}_{KS}$ is

$$\hat{H}_{KS} = -\frac{1}{2}\nabla^2 + V_{eff}(\mathbf{r}). \tag{2.29}$$

Here the single-particle kinetic energy operator $\hat{T}$ is

$$\hat{T} = -\frac{1}{2}\nabla^2. \tag{2.30}$$

The effective KS potential $V_{\text{eff}}(\mathbf{r})$ is defined by

$$\begin{aligned} V_{\text{eff}}(\mathbf{r}) &= \frac{\delta(E_V[\rho] - T_s[\rho])}{\delta\rho(\mathbf{r})} \\ &= \frac{\delta(V_{\text{ne}}[\rho] + J[\rho] + E_{\text{xc}}[\rho])}{\delta\rho(\mathbf{r})}, \end{aligned} \tag{2.31}$$

where the definition of $E_V[\rho]$ in Eq. (2.24) has been used. By taking the functional derivative $V_{\text{eff}}(\mathbf{r})$ writes

$$V_{\text{eff}}(\mathbf{r}) = V(\mathbf{r}) + V_{\text{H}}(\mathbf{r}) + V_{\text{xc}}(\mathbf{r}), \tag{2.32}$$

where

$$V_{\text{H}}(\mathbf{r}) = \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, d\mathbf{r}' \tag{2.33}$$

is the electrostatic Hartree potential (also called Coulomb potential) and

$$V_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}[\rho]}{\delta\rho(\mathbf{r})} \tag{2.34}$$

is the exchange-correlation potential.

The fermionic many-body ground state WF for such a non-interacting system can by constructed as Slater determinant of the matrix of the N_e lowest energy single-particle states

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_e}) = \begin{vmatrix} \Psi_1(\mathbf{r}_1) & \Psi_1(\mathbf{r}_2) & \ldots & \Psi_1(\mathbf{r}_{N_e}) \\ \Psi_2(\mathbf{r}_1) & \Psi_2(\mathbf{r}_2) & \ldots & \Psi_2(\mathbf{r}_{N_e}) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_{N_e}(\mathbf{r}_1) & \Psi_{N_e}(\mathbf{r}_2) & \ldots & \Psi_{N_e}(\mathbf{r}_{N_e}) \end{vmatrix}. \tag{2.35}$$

Using the orthogonality relation for the single-particle WFs $\int \Psi_i^*(\mathbf{r})\Psi_j(\mathbf{r})d\mathbf{r} = \delta_{ij}$ the ground state electron density [Eq. (2.11)] for such a many-body WF is

$$\rho(\mathbf{r}) = \sum_{i=1}^{N_e} \Psi_i^*(\mathbf{r})\Psi_i(\mathbf{r}), \tag{2.36}$$

which is equivalent to populating the N_e lowest lying single-particle levels. It is possible to generalize this to non-integer occupation [77, 70], so that the density becomes

$$\rho(\mathbf{r}) = \sum_{i=1} n_i\ \Psi_i^*(\mathbf{r})\Psi_i(\mathbf{r}), \tag{2.37}$$

where the sum now goes over all WFs and $n_i \in [0, 1]$ is the occupation number of the eigenstate with index i and energy ϵ_i. The correct set of $\{n_i\}$ is the one that minimized the total energy [70] with the constraint that

$$N_e = \sum_i n_i. \tag{2.38}$$

In practice however usually only the lowest lying states are occupied.

The kinetic energy of this non-interacting system with density $\rho(\mathbf{r})$ is

$$T_{\mathrm{s}} = \sum_{i=1} n_i \int \Psi_i^*(\mathbf{r})(-\frac{1}{2}\nabla^2)\Psi_i(\mathbf{r})d\mathbf{r}. \tag{2.39}$$

Although the functional dependence of T_{s} on $\rho(\mathbf{r})$ is unknown, this equation shows that it is possible to calculate T_{s} indirectly as $T_{\mathrm{s}}[\{\Psi_i\}]$ by first solving the KS equation [Eq. (2.28)] to obtain the set of ground state single-particle WFs $\{\Psi_i\}$. The possibility to calculate an approximate value for the kinetic energy is one of the main advantages of the KS scheme in practical applications. The other reason for the success of the KS scheme is that a number of approximations for the exchange-correlation potential have been developed and that these work well for a variety of materials, even if not in all.

We conclude this section with some notes about the physical meaning of the KS eigenvalues. The sum of the eigenvalues ϵ_i

$$\sum_{i=1}^{N_{\mathrm{e}}} \epsilon_i = T_{\mathrm{s}}[\rho] + \int V_{\mathrm{eff}}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \tag{2.40}$$

is not equal to the total energy E_V from Eq. (2.24). This instead can be rewritten as

$$E_V = \sum_{i=1}^{N_{\mathrm{e}}} \epsilon_i - \frac{1}{2}\int \frac{\rho(\mathbf{r})\rho(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' + E_{\mathrm{xc}}[\rho] - \int V_{\mathrm{xc}}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r}. \tag{2.41}$$

The physical meaning of the eigenvalues ϵ_i is that each eigenvalue corresponds to the derivative of the total energy with respect to the occupation number n_i of the state

$$\epsilon_i = \frac{dE_{\mathrm{V}}}{dn_i}. \tag{2.42}$$

This fundamental relation is usually referred to as *Janak theorem*, and is valid for any n_i ranging between 0 and 1 [77, 70]. For molecules it is therefore possible to obtain the electron affinity A, defined as the total energy difference for a molecule in its neutral state (N_0 electrons) and in the singly reduced stated ($N_0 + 1$ electrons)

$$A = E_{\mathrm{V}}(N_0) - E_{\mathrm{V}}(N_0+1) = -\int_0^1 \epsilon_{\mathrm{LUMO}}(n)\, dn, \tag{2.43}$$

where LUMO denotes the lowest unoccupied molecular orbital. The larger the affinity the more favorable it is for the molecule to bind an extra electron. In the same way the ionization potential I, which is the energy difference upon removal of an electron, can be calculated as

$$I = E_{\mathrm{V}}(N_0-1) - E_{\mathrm{V}}(N_0) = -\int_0^1 \epsilon_{\mathrm{HOMO}}(n)\, dn, \tag{2.44}$$

where HOMO denotes the highest occupied molecular orbital. The larger the ionization potential the more energy is required to remove an electron from the molecule. It can be shown that $\epsilon_{\mathrm{HOMO}}(n)$ should be a constant for any value of $n > 0$, and that in general $\epsilon_{\mathrm{HOMO}}(n) > \epsilon_{\mathrm{HOMO}}(0)$ for any $n > 0$ [78, 79, 80]. This discontinuity of $\epsilon_{\mathrm{HOMO}}(n)$ for an infinitesimal n is known as the "derivative discontinuity" (referring to the fact that ϵ_{HOMO} is a derivative of the total energy with occupation number). The ionization potential therefore becomes

$$I = -\epsilon_{\mathrm{HOMO}}, \tag{2.45}$$

where by definition the occupation of the HOMO is assumed to be non-zero. We note however that this relation is only valid if one uses a XC functional that correctly describe the derivative discontinuity. For approximate functionals that do not have this property $-\epsilon_{\mathrm{HOMO}}$ is only a rough approximation for I.

2.1.4 Spin-polarized DFT

The generalization of DFT to a collinear spin-polarized system is achieved by introducing separately a majority electron density $\rho^{\uparrow}(\mathbf{r})$ and a minority electron density $\rho^{\downarrow}(\mathbf{r})$ [81, 70]. The total electron density then is

$$\rho(\mathbf{r}) = \rho^{\uparrow}(\mathbf{r}) + \rho^{\downarrow}(\mathbf{r}), \tag{2.46}$$

and the magnetization density is defined as the difference

$$m(\mathbf{r}) = \rho^{\uparrow}(\mathbf{r}) - \rho^{\downarrow}(\mathbf{r}). \tag{2.47}$$

Majority and minority spins satisfy two independent KS equations

$$\begin{aligned} \hat{H}^{\uparrow}_{\mathrm{KS}}\ \Psi^{\uparrow}_i &= \epsilon^{\uparrow}_i\ \Psi^{\uparrow}_i, \\ \hat{H}^{\downarrow}_{\mathrm{KS}}\ \Psi^{\downarrow}_i &= \epsilon^{\downarrow}_i\ \Psi^{\downarrow}_i, \end{aligned} \tag{2.48}$$

where $\hat{H}^{\sigma}_{\mathrm{KS}} = \frac{\delta E_{\mathrm{KS}}}{\delta \rho^{\sigma}(\mathbf{r})}$. σ indicates the spin and is either $\uparrow$ or $\downarrow$. All the terms of the KS Hamiltonian operators $\hat{H}^{\uparrow}_{\mathrm{KS}}$ and $\hat{H}^{\downarrow}_{\mathrm{KS}}$ are identical and depend only on $\rho(\mathbf{r})$ except for the XC potential, for which in general $V^{\uparrow}_{\mathrm{xc}}(\mathbf{r}) \neq V^{\downarrow}_{\mathrm{xc}}(\mathbf{r})$. The reason is that the XC energy is a functional of both $\rho^{\uparrow}(\mathbf{r})$ and $\rho^{\downarrow}(\mathbf{r})$ ($E_{\mathrm{xc}} = E_{\mathrm{xc}}[\rho^{\uparrow}, \rho^{\downarrow}]$), so that $V^{\uparrow}_{\mathrm{xc}}(\mathbf{r}) = \frac{\delta E_{\mathrm{xc}}[\rho^{\uparrow}, \rho^{\downarrow}]}{\delta \rho^{\uparrow}(\mathbf{r})}$ and $V^{\downarrow}_{\mathrm{xc}}(\mathbf{r}) = \frac{\delta E_{\mathrm{xc}}[\rho^{\uparrow}, \rho^{\downarrow}]}{\delta \rho^{\downarrow}(\mathbf{r})}$ are different. The ground state electron density is obtained by occupying the N_{e} lowest energy levels choosing from both spins

$$\rho^{\sigma}(\mathbf{r}) = \sum_{i=1}^{N^{\sigma}} \Psi^{\sigma *}_i(\mathbf{r}) \Psi^{\sigma}_i(\mathbf{r}). \tag{2.49}$$

where $N_e = N^{\uparrow} + N^{\downarrow}$.

The general procedure for a non-collinear spin-configuration is similar, however in that case 4 electron densities are needed and the KS equation becomes a matrix-vector equation where the 2×2 KS Hamiltonian acts on spinor wave functions [82, 72].

2.1.5 Exchange-correlation functionals

In principle the Kohn-Sham formalism is exact. However this is based on the assumption that the exchange-correlation potential $V_{xc}[\rho](\mathbf{r})$ can be calculated exactly, which in practice is not possible. There are however different approximations for $V_{xc}[\rho](\mathbf{r})$ that work well for a set of systems.

Since $V_{xc}(\mathbf{r})$ is a functional of $\rho(\mathbf{r})$, in general the value of V_{xc} at a point $\mathbf{r}$ depends on the value of the function $\rho(\mathbf{r}')$ over the whole space spanned by $\mathbf{r}'$. The most basic approximation is the local density approximation (LDA), where it is assumed that V_{xc} at point $\mathbf{r}$ is a function of the density at that point only, so that $V_{xc}(\mathbf{r}) = V_{xc,LDA}(\rho(\mathbf{r}))$ [67, 70]. The function $V_{xc,LDA}(\rho(\mathbf{r}))$ is chosen in such a way to be the exact solution for a uniform electron density, and is therefore a good approximation for electron densities that vary slowly over space [67]. Although this condition is rarely fulfilled in real systems, the LDA works well for many systems, mainly because it fulfills all relevant sum rules and has many correct formal features [71]. The generalization of the LDA for spin-polarized systems is the local spin density approximation (LSDA), where $V^{\sigma}_{xc}(\mathbf{r}) = V^{\sigma}_{xc,LSDA}(\rho^{\uparrow}(\mathbf{r}), \rho^{\downarrow}(\mathbf{r}))$ [81].

The so called generalized gradient approximation (GGA) is an expansion of the LDA, where V_{xc} depends also on the gradient of $\rho^{\sigma}(\mathbf{r})$, so that [71, 72]

$$V^{\sigma}_{xc}(\mathbf{r}) = V^{\sigma}_{xc,GGA}(\rho^{\uparrow}(\mathbf{r}), \rho^{\downarrow}(\mathbf{r}), \nabla\rho^{\uparrow}(\mathbf{r}), \nabla\rho^{\downarrow}(\mathbf{r})). \tag{2.50}$$

There are different flavors of GGAs that are built in such a way as to satisfy different criteria [83, 84, 85]. In this work the GGA as parametrized by Perdew, Burke and Ernzerhof (PBE) is used [85].

Both the LDA and GGA are local approximations in the sense that $V^{\sigma}_{xc}(\mathbf{r})$ depends only on the properties of the electron density at that single point $\mathbf{r}$. Therefore they usually both give similar results, so that for systems where LDA gives a fundamentally wrong result also GGA will not improve the results much. The failures of the LDA and GGA are mainly found for materials where the electrons tend to be localized and "strongly interacting", which is for example the case of many transition-metal oxides [72]. One way to solve this problem is to use non-local, orbital-dependent exchange-correlation functionals [71, 73, 72].

A computationally inexpensive approach of this type to correct some of the LDA failures, typically used for transition-metal oxides, is the "LDA+U" approximation [86, 87]. In this case the LDA XC energy is replaced by a Hubbard-U type orbital-dependent energy for some orbitals, usually the d and f orbitals, with the effect that occupied orbitals are shifted to lower energies, whereas unoccupied are shifted to higher energies. The method is not fully *ab initio* since it relies on the parameters U and J, which are usually chosen in such a way as to reproduce some reference properties the material.

One of the fundamental problems of LDA and GGA is the so-called "self-interaction" [88, 70, 72], which is caused by the fact that in the Hartree potential [Eq. (2.33)] the integral is performed over the total $\rho(\mathbf{r})$, which includes the spurious electrostatic interaction of each particle with itself. A similar self energy contribution is also included in the LDA and GGA XC potential. For example, in the single electron case the Coulomb potential acting on a particle should be zero. In the LDA and GGA however the particle will feel the potential generated by itself. In the Hartree-Fock approach the self-interaction in V_{H} is exactly canceled by the exchange term [72], so that the total potential is self-interaction free. The problem in LDA and GGA is that the exchange potential is only approximated and therefore this cancellation is not complete, so that part of the unphysical self-interaction remains. One way to overcome this problem is the use of an exact functional for the exchange part in the XC potential, so that by construction the self-interaction is canceled [89, 71, 72]. The correlation part of the potential has to be added separately, and it is important that this does not reintroduce a spurious self-interaction. The problem of such an approach is that it is difficult to construct the corresponding correlation, which is especially important in transition metals, moreover the method is computationally very expensive. Another possibility is to explicitly add a "self-interaction correction" (SIC) to the XC functional. Different functionals have been developed to this aim, most recent works are however based on the ideas of Perdew and Zunger [88]. The main drawback of these methods however is that they are still computationally much more expensive than LDA or GGA. In this work we use the "atomic-SIC" (ASIC) approximation, described in Ref. [90] and based on the concepts in Refs. [91, 92]. It is a computationally inexpensive approximate method which has been shown to give good results for a variety of systems. For metal-oxides it improves the band gap usually bringing it close to the experimental value. Similarly the ionization potential of molecules is improved compared to LDA [90]. Those improvements are important in electronic transport calculations through molecules [93]. We note however that neither the LDA, nor the GGA or the ASIC correctly describe the "derivative dis-

continuity". We also note that for metallic systems the ASIC often gives unphysical results. In the ASIC an occupation dependent correction is added to the LDA XC potential, with the effect that occupied states are shifted down to lower energies, whereas unoccupied levels are not affected. The physical reason for this is that due to the self-interaction error the LDA XC energy is too high only for occupied states.

In our calculations we use either LDA, GGA or LDA+ASIC, depending on which XC functional is better suited to the studied materials.

2.2 The SIESTA code

There exists a variety of different *ab initio* computer codes that implement the DFT method in order to calculate the electronic structure of materials [94, 95, 96, 97, 98, 99, 100, 101, 69]. The differences between the various code generally lie in the basis set over which the Kohn-Sham WFs are expanded, and in the approximations used, where usually the tradeoff lies between accuracy and speed. In this work the SIESTA code is used [69]. SIESTA employs pseudopotentials to describe the nuclei and the core electrons, and the WF is expanded as a linear combination of non-orthogonal atomic orbitals. In this section a brief overview of SIESTA is given, a more detailed description can be found in Refs. [69, 102, 103, 104, 105]. The main advantages of SIESTA for the purposes of this work are that it is a very efficient code and therefore makes it possible to calculate systems with a rather large number of atoms (of the order of a few thousands). Furthermore the Hamiltonian is written in a tight-binding like form, which makes it easy to integrate the electronic transport formalism developed in tight-binding schemes (see chapter 4). Another big advantage is that the source code is open, so that it can be modified depending on the user's needs.

2.2.1 Pseudopotentials

In order to reduce the number of states in the calculation it is of advantage to split the electrons of an atom into core and valence electrons. The valence electrons are the ones occupying the outer shells that are responsible for the chemical bonding, so that the electronic properties of these states depends strongly on their chemical environment. The core electrons on the other hand are the ones in the fully occupied inner shells. These do not participate to the chemical bonding, and their electronic properties are only slightly affected by the chemical environment, so that the potential generated by the core electrons can be assumed to be independent from the environment. It is

therefore possible to remove the core electrons from the calculation and only consider the valence electrons. The nucleus and the core electrons are then replaced by an effective potential. This effective potential however is singular at the position of the nucleus, and to describe it properly a very fine mesh would be needed. However since the electronic properties of molecules and solids are determined by the charge density in the bonding region, it is not necessary to describe the charge density close to the nucleus accurately. It is therefore possible to replace the effective potential by the so-called pseudopotential V_{PS}, which avoids the singularity at the nucleus, but reproduces the right potential at larger distances. In the same way the corresponding pseudo-wave functions differ from the atomic wave functions close to the nucleus, but they are identical in the bonding region. An atom where the nucleus and the core electrons are replaced by a pseudo-potential is called a pseudo-atom.

In SIESTA Trullier-Martins norm-conserving pseudopotentials are used [106, 107]. The effective single-particle atomic potential $V_{\mathrm{eff,cv}}$ acting on the valence electrons is equal to the sum of the effective potential generated by the nucleus and the core electrons $V_{\mathrm{eff,c}}$ and the one generated by the other valence electrons alone V_{v}, so that $V_{\mathrm{eff,cv}} = V_{\mathrm{eff,c}} + V_{\mathrm{v}}$. It is assumed that $V_{\mathrm{eff,cv}}$ is spherical, so that $V_{\mathrm{eff,cv}} = V_{\mathrm{eff,cv}}(r)$, where r is the distance from the nucleus. The valence WFs can therefore be expanded as a product of a radial part $R_{nl}(r)$ with the spherical harmonics $Y_{lm}(\theta, \phi)$ that only depend on the angles θ and ϕ. The radial atomic-WF $R_{nl}(r)$ are then constructed by self-consistently solving the radial KS equation

$$\left[-\frac{1}{2}\frac{d^2}{dr^2} + \frac{l(l+1)}{2r^2} + V_{\mathrm{eff,c}}(r) + V_{\mathrm{v}}(r)\right] rR_{nl}(r) = \epsilon_{nl} R_{nl}(r), \tag{2.51}$$

where the corresponding equation for the pseudo-atom is

$$\left[-\frac{1}{2}\frac{d^2}{dr^2} + \frac{l(l+1)}{2r^2} + V_{\mathrm{PS},l}(r) + V_{\mathrm{v}}(r)\right] rR_{\mathrm{PS},l}(r) = \epsilon_l R_{\mathrm{PS},l}(r). \tag{2.52}$$

Here $V_{\mathrm{PS},l}(r)$ and $R_{\mathrm{PS},l}(r)$ are the l-dependent pseudo-potential and pseudo-WF respectively. We note that there is no dependence on n, but, since for each quantum number l a different equation is solved, the pseudopotential will be different for each l. $V_{\mathrm{PS},l}(r)$ is constructed from Eq. (2.52) in a unique way by requiring the following four criteria to be satisfied [106]:

1. For a given principal quantum number n the atomic wave functions have $n-1$ nodes. The corresponding pseudo-wave-functions however have no nodes, this means that all information about the n quantum number is lost.

2. The pseudo-wave-function is identical to the atomic wave function beyond a critical radius r_{c}.

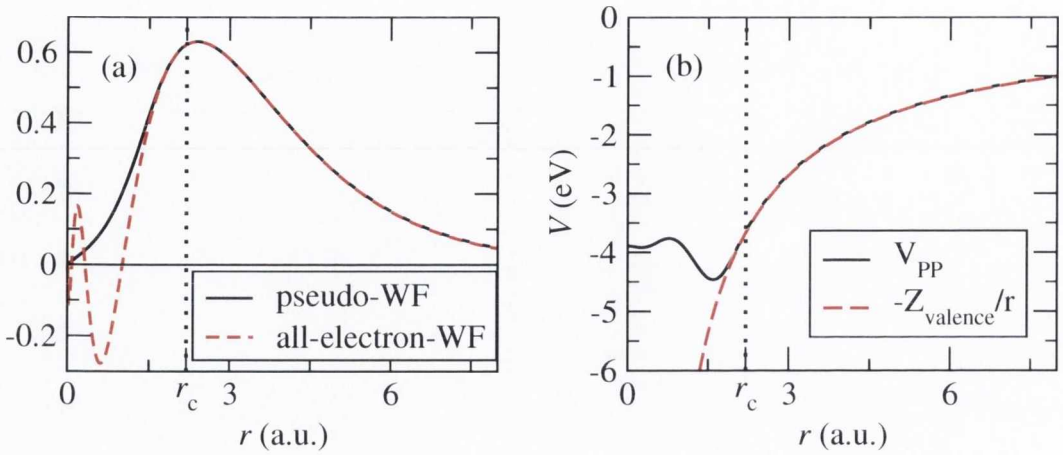


Figure 2.1: (a) Pseudo-WF and all-electron-WF for the Fe-4*s* shell and (b) the corresponding pseudopotential.

3. The charge inside a sphere of radius r_c is the same for both atomic- and pseudo-wave-function.

4. The eigenvalues calculated for both the atomic- and pseudo-wave-function are equal.

In Fig. 2.1(a) the all-electron radial WF is shown together with the corresponding pseudo-WF for the Fe-4*s* shell. In Fig. 2.1(b) the corresponding pseudopotential is shown. The pseudopotential converges to the unscreened potential $-Z/r$ for $r > r_c$, where Z is the number of valence electrons.

For certain elements that have a core charge with a rather large extension it is necessary to add non-linear core corrections [108]. In this case the value of part of the core charge is stored together with the pseudopotential and taken into account for the calculation of the exchange-correlation potential for the pseudo-atom.

Since $\hat{V}_{\mathrm{PS}}$ is different for each l, it can be written as $\hat{V}_{\mathrm{PS}}(r) = \sum_l V_{\mathrm{PS},l}(r)\hat{P}_l$, where $\hat{P}_l$ is a projector onto the l component of the WF. The long tail $-Z/r$ part of the pseudopotential is identical for all l, whereas the l-dependent part has only finite range equal to r_c. It is therefore possible to split the pseudopotential into a long range l-independent V_{local} and short range l-dependent part

$$\hat{V}_{\mathrm{PS}}(r) = V_{\mathrm{local}}(r) + \sum_{lm} \delta V(r)\hat{P}_l \tag{2.53}$$

where $\delta V(r) = \delta V_{\mathrm{PS},l}(r) - V_{\mathrm{local}}(r)$ is short ranged [69, 106]. Furthermore the l-dependent part is approximated into the fully non-local form $\hat{V}_{\mathrm{KB}}$ as proposed by Kleinman and Bylander (KB) [109, 106, 107, 69]. The result is that the pseudopotential operator is of finite range and has the form $\hat{V}_{\mathrm{PS}}(r) = V_{\mathrm{local}}(r) + \hat{V}_{\mathrm{KB}}$.

2.2.2 Basis set

One of the goals of the SIESTA program is to achieve linear scaling in the computational effort with system size. This requires finite range interactions and the corresponding sparsity of all the relevant matrices. The sparsity is achieved by expanding the WF as a linear combination of atomic orbitals (LCAO) of the type described in Ref. [110]. These basis orbitals have a finite radius, and they are strictly zero beyond that. For an atom I located at position $\mathbf{R}_I$ the basis functions are

$$\Phi_{Ilmn}(\mathbf{r}) = \Phi_{Iln}(r_I) Y_{lm}(\frac{\mathbf{r}_I}{r_I}), \tag{2.54}$$

where $\mathbf{r}_I = \mathbf{r} - \mathbf{R}_I$, $r_I = |\mathbf{r}_I|$, l and m are the angular momentum quantum numbers, and the index n indicates the fact that for each l, m an arbitrary number of nodeless radial functions can be used. These are called higher ζ-orbitals [69]. In principle the shape of the basis orbitals $\Phi_{Iln}(r)$ is arbitrary, however in SIESTA these are usually calculated in such a way that they correspond to the solution of the pseudo-atom within a spherical box. In order to confine them inside a certain cutoff radius r_c an infinite hard wall potential is imposed at r_c. The smaller r_c, the more the basis functions differ from the atomic WF. In practice the first ζ-orbital is constructed by solving the radial KS equation for the pseudo-atom [Eq. (2.52)] at a slightly higher energy

$$\left[-\frac{1}{2r}\frac{d^2}{dr^2}r + \frac{l(l+1)}{2r^2} + V_{\mathrm{PS},l}(r) + V_v(r)\right] \Phi_l(r) = (\epsilon_l + \delta\epsilon_l)\, \Phi_l(r), \tag{2.55}$$

with $\Phi_l(r_c) = 0$. The energy shift $\delta\epsilon_l$ can be chosen in such a way that the first node of $\Phi_l(r)$ occurs at $r = r_c$. Higher ζ-orbitals are constructed starting from the first ζ-orbital using the split-norm scheme [111]. The basis can further be enriched by including so-called polarization orbitals, which are obtained by polarizing the basis orbitals with an electric field [69].

2.2.3 Electron Hamiltonian and density matrices

For a set of pseudo-atoms the non-spin-polarized KS Hamiltonian operator $\hat{H}$ (we omit the index "KS" to explicitly label the Kohn-Sham Hamiltonian) as defined in Eq. (2.29) is

$$\hat{H} = \hat{T} + \sum_I V_{\mathrm{local},I}(\mathbf{r}) + \sum_I \hat{V}_{\mathrm{KB},I} + V_{\mathrm{H}}(\mathbf{r}) + V_{\mathrm{xc}}(\mathbf{r}), \tag{2.56}$$

where $\hat{T}$ is the kinetic energy operator [Eq. (2.30)], $V_{\mathrm{H}}(\mathbf{r})$ is the Hartree electrostatic potential [Eq. (2.33)] and $V_{\mathrm{xc}}(\mathbf{r})$ is the exchange-correlation potential. $V_{\mathrm{local},I}(\mathbf{r})$ is the

local part of the pseudopotential of atom I and $\hat{V}_{\mathrm{KB},I}$ is the corresponding Kleinman-Bylander projector [69]. $V_{\mathrm{local},I}(\mathbf{r})$ is a long range potential decaying as $-Z_I/r_I$, where Z_I is the valence charge of pseudo-atom I. It is however possible to obtain a short range potential by screening $V_{\mathrm{local},I}(\mathbf{r})$ with the Coulomb potential generated by the atomic electron valence density $\rho_I(\mathbf{r})$, since $\int \rho_I(\mathbf{r})d\mathbf{r} = Z_I$ and therefore the total charge of such a combined system is zero. $\rho_I(\mathbf{r})$ is obtained by populating the levels of the pseudo-atom according to the free-standing atomic occupation with Z_I electrons. By Adding the Coulomb screening potential $V_{\mathrm{H}}[\rho_I](\mathbf{r})$ generated by $\rho_I(\mathbf{r})$ to $V_{\mathrm{local},I}(\mathbf{r})$ the neutral atom potential $V_{\mathrm{NA},I}(\mathbf{r}) = V_{\mathrm{local},I}(\mathbf{r}) + V_{\mathrm{H}}[\rho_I](\mathbf{r})$ is obtained. We note that $V_{\mathrm{NA},I}(\mathbf{r})$ is zero if $r_I \geq r_{\mathrm{c,max}}$, where $r_{\mathrm{c,max}}$ is the largest of the cutoff radii $r_{\mathrm{c},l}$ for all l of atom I. If we define $\delta\rho(\mathbf{r}) = \rho(\mathbf{r}) - \sum_I \rho_I(\mathbf{r})$, so that $\int \delta\rho(\mathbf{r})d\mathbf{r} = 0$, the total Hamiltonian can be rewritten as

$$\hat{H} = \hat{T} + \sum_I \hat{V}_{\mathrm{KB},I} + \sum_I V_{\mathrm{NA},I}(\mathbf{r}) + \delta V_{\mathrm{H}}(\mathbf{r}) + V_{\mathrm{xc}}(\mathbf{r}), \tag{2.57}$$

where $\delta V_{\mathrm{H}}(\mathbf{r}) = V_{\mathrm{H}}[\delta\rho](\mathbf{r})$ is the Coulomb potential generated by the charge $\delta\rho(\mathbf{r})$. All the potentials in the Hamiltonian are now short-range.

The KS equation [Eq. (2.28)] then is

$$\hat{H}\ \Psi(\mathbf{r}) = E\ \Psi(\mathbf{r}). \tag{2.58}$$

The WF $\Psi(\mathbf{r})$ is expanded over the localized basis orbitals described in Sec. 2.2.2

$$\Psi(\mathbf{r}) = \sum_\nu \psi_\nu \Phi_\nu(\mathbf{r}), \tag{2.59}$$

where the set of indices $\{Ilmn\}$ has been collected into the general index ν. In Eq. (2.59) the sum runs over all N_{u} orbitals in the simulation cell. The KS equation becomes

$$\sum_\nu \hat{H}(\mathbf{r})\Phi_\nu(\mathbf{r})\psi_\nu = E\sum_\nu \Phi_\nu(\mathbf{r})\psi_\nu. \tag{2.60}$$

Multiplying this equation from the left by $\Phi^*_\mu(\mathbf{r})$ and integrating over $\mathbf{r}$ we obtain a matrix equation

$$\sum_\nu H_{\mu,\nu}\ \psi_\nu = E\sum_\nu S_{\mu,\nu}\ \psi_\nu. \tag{2.61}$$

Here we have introduced the overlap matrix

$$S_{\mu,\nu} = \langle\Phi_\mu|\Phi_\nu\rangle = \int \Phi^*_\mu(\mathbf{r})\Phi_\nu(\mathbf{r})d\mathbf{r}, \tag{2.62}$$

and the Hamiltonian matrix

$$H_{\mu\nu} = \langle\Phi_\mu|\hat{H}|\Phi_\nu\rangle = \int \Phi^*_\mu(\mathbf{r})\hat{H}(\mathbf{r})\Phi_\nu(\mathbf{r})d\mathbf{r}. \tag{2.63}$$

The dimension of the matrices S and H is equal to the number of basis orbitals in the simulation cell N_u. For a collinear spin-polarized calculation the Hamiltonian will be different for up and down spin

$$H^{\uparrow}_{\mu\nu} = \langle\Phi_\mu|\hat{H}^{\uparrow}|\Phi_\nu\rangle, \tag{2.64}$$

$$H^{\downarrow}_{\mu\nu} = \langle\Phi_\mu|\hat{H}^{\downarrow}|\Phi_\nu\rangle, \tag{2.65}$$

where the difference between $\hat{H}^{\uparrow}$ and $\hat{H}^{\downarrow}$ lies in the exchange-correlation potential (see Sec. 2.1.4). In the remaining part of this section we will describe an unpolarized system, the generalization to a spin-polarized system is straight-forward.

The normalization of the WFs in real space $\int |\Psi_i(\mathbf{r})|^2 d\mathbf{r} = 1$ leads to the normalization of the expansion coefficients $\psi^\dagger S\psi = 1$. By using Eq. (2.36) the charge density in real space then is

$$\rho(\mathbf{r}) = \sum_{i=1}^{N_\mathrm{e}} |\Psi_i(\mathbf{r})|^2, \tag{2.66}$$

where the sum goes over the N_e lowest lying states. The numerical stability and convergence can be improved by partially occupying also some of the higher lying levels, in which case the charge density becomes

$$\rho(\mathbf{r}) = \sum_i n_i |\Psi_i(\mathbf{r})|^2, \tag{2.67}$$

where the sum now goes over all N_u WFs and $n_i \in [0, 1]$ is the occupation number of the eigenstate with index i and energy E_i. Within SIESTA n_i is given by a Fermi-Dirac distribution $n_i = f(E_i)$

$$f(E_i) = \frac{1}{1 + e^{\frac{E_i - E_\mathrm{F}}{k_\mathrm{B}T}}}, \tag{2.68}$$

where k_B is the Boltzmann constant, T is the electronic temperature and E_F is the Fermi energy.

Then the charge density is expanded as function of the basis orbitals and becomes

$$\rho(\mathbf{r}) = \sum_{\mu\nu} \rho_{\mu\nu} \Phi^*_\nu(\mathbf{r}) \Phi_\mu(\mathbf{r}), \tag{2.69}$$

where we have introduced the density matrix

$$\rho_{\mu\nu} = \sum_i n_i \psi_{i\mu} \psi^*_{i\nu}. \tag{2.70}$$

This is a Hermitian $N_\mathrm{u} \times N_\mathrm{u}$ matrix. Since within the LDA and GGA approximations for the exchange-correlation potential the real space Hamiltonian is a functional of

$\rho(x)$, in matrix representation it is a function of the density matrix $H = H(\rho)$. Since the integral over the simulation cell of the charge density is equal to the number of valence electrons N_e it is possible to express N_e as function of the density matrix

$$N_e = \int \rho(\mathbf{r}) d\mathbf{r} = \sum_{\mu\nu} \rho_{\mu\nu} S_{\nu\mu} = \mathrm{Tr}[\rho S]. \tag{2.71}$$

Using the normalization of the WF this simplifies to

$$N_e = \sum_i n_i. \tag{2.72}$$

For a given set of eigenvalues E_i this equation implicitly determines the Fermi energy used in Eq. (2.70) to obtain ρ.

Since H depends on ρ, and in turn ρ depends on H, the solution has to be self-consistent. Starting from a trial density matrix ρ^{in} the corresponding Hamiltonian $H(\rho^{\mathrm{in}})$ is constructed, and from this the density matrix $\rho^{\mathrm{out}} = \rho(H(\rho^{\mathrm{in}}))$ is obtained. If ρ^{in} is the self-consistent solution then $\rho^{\mathrm{out}} = \rho^{\mathrm{in}}$. In a practical calculation the criterion for self-consistency is $||\rho^{\mathrm{out}} - \rho^{\mathrm{in}}|| < \epsilon_{\mathrm{SC}}$, where $||\ldots||$ is a matrix norm, usually the maximum norm, and ϵ_{SC} is a small number setting the degree of convergence. In SIESTA the self-consistent solution is found by updating ρ^{in} using part of the output density matrix ρ^{out} until the self-consistent criterion is fulfilled [112].

2.2.4 Brillouin zone sampling

An infinite periodic system is obtained by specifying a unit cell by its three lattice vectors together with the internal positions of the atoms inside the unit cell. The crystal is then constructed by stacking cells separated by the lattice vectors in all directions, and the charge density can be constructed by applying Bloch theorem. In this section we denote orbitals inside a reference unit cell by unprimed indices (e.g. μ, ν, ...), and those in a general cell in the full crystal by primed indices (e.g. μ', ν', ...). Sums over unprimed orbital indices therefore run over the N_u orbitals in the unit cell, whereas sums over primed indices run over all the orbitals in the whole infinite crystal. The notation $\nu' \equiv \nu$ indicates that $\Phi_{\nu'}(\mathbf{r})$ and $\Phi_{\nu}(\mathbf{r})$ are equivalent orbitals, related by a lattice vector translation, so that orbital ν is the equivalent orbital inside the reference unit cell to the orbital ν' in an arbitrary cell.

For a given $\mathbf{k}$-vector $\mathbf{k} = (k_x, k_y, k_z)$ in the Brillouin zone (BZ) the WF is expanded as

$$\psi_{\mathbf{k}\mu'} = e^{i\mathbf{k}\bar{\mathbf{R}}_{\mu'}} \psi_{\mathbf{k}\mu} \ , \qquad \mu' \equiv \mu. \tag{2.73}$$

Here $\bar{\mathbf{R}}_{\mu'} = \mathbf{R}_{\mu'} - \mathbf{R}_{\mu}$ is the lattice vector translation between the equivalent orbitals μ and μ'. For each $\mathbf{k}$-point the matrix KS equation [Eq. (2.61)] for the infinite system can be reduced to a matrix equation that involves only the orbitals of the unit cell

$$\sum_{\nu} H_{\mathbf{k}\mu\nu}\ \psi_{\mathbf{k}\nu} = E \sum_{\nu} S_{\mathbf{k}\mu\nu}\ \psi_{\mathbf{k}\nu}, \tag{2.74}$$

where the Hermitian $N_{\mathrm{u}} \times N_{\mathrm{u}}$ matrices $H_{\mathbf{k}}$ and $S_{\mathbf{k}}$ are given by

$$H_{\mathbf{k}\mu\nu} = \sum_{\nu' \equiv \nu} H_{\mu\nu'} e^{i\mathbf{k}\bar{\mathbf{R}}_{\nu'}}, \tag{2.75}$$

$$S_{\mathbf{k}\mu\nu} = \sum_{\nu' \equiv \nu} S_{\mu\nu'} e^{i\mathbf{k}\bar{\mathbf{R}}_{\nu'}}. \tag{2.76}$$

Although in principle these sums run over all the unit cells of the infinite crystal, the number of non-zero terms in the sum is finite since the basis orbitals have strictly finite range. This is reflected in the structure of H and S. The WF with energy E_i in real space is then given by

$$\Psi_{\mathbf{k}i}(\mathbf{r}) = \sum_{\mu'} \psi_{\mathbf{k}i\mu'} \Phi_{\mu'}(\mathbf{r}), \tag{2.77}$$

and also here since only a finite number of orbitals $\Phi_{\mu'}(\mathbf{r})$ is non-zero at a given point $\mathbf{r}$, this sum effectively only goes over this finite number of orbitals and not over the whole crystal. The corresponding charge density is

$$\rho(\mathbf{r}) = \frac{1}{\Omega_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \sum_{i} n_{\mathbf{k}i} |\Psi_{\mathbf{k}i}(\mathbf{r})|^2 d\mathbf{k} = \sum_{\mu'\nu'} \rho_{\mu'\nu'} \Phi^*_{\nu'}(\mathbf{r}) \Phi_{\mu'}(\mathbf{r}). \tag{2.78}$$

The $\mathbf{k}$-space integral goes over the BZ and $\Omega_{\mathrm{BZ}} = \int_{\mathrm{BZ}} d\mathbf{k}$ is the volume of the BZ. The density matrix now is

$$\rho_{\mu'\nu'} = \frac{1}{\Omega_{\mathrm{BZ}}} \int_{\mathrm{BZ}} e^{i\mathbf{k}(\bar{\mathbf{R}}_{\mu'} - \bar{\mathbf{R}}_{\nu'})} \rho_{\mathbf{k}\mu\nu}\ d\mathbf{k}, \qquad \mu' \equiv \mu, \nu' \equiv \nu \tag{2.79}$$

with the $N_{\mathrm{u}} \times N_{\mathrm{u}}$ Hermitian matrix $\rho_{\mathbf{k}\mu\nu}$ defined as

$$\rho_{\mathbf{k}\mu\nu} = \sum_{i} n_{\mathbf{k}i}\ \psi_{\mathbf{k}i\mu} \psi^*_{\mathbf{k}i\nu}. \tag{2.80}$$

If $\rho_{\mathbf{k}\mu\nu}$ is known for all $\mathbf{k}$-points it is possible to construct $\rho_{\mu'\nu'}$ for all μ' and ν'. With Eq. (2.71) the total number of particles in the unit cell becomes

$$N_{\mathrm{e}} = \int_{\mathrm{UC}} d\mathbf{r} \rho(\mathbf{r}) = \sum_{\mu\nu'} \rho_{\mu\nu'} S_{\nu'\mu} = \frac{1}{\Omega_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \mathrm{Tr}[\rho_{\mathbf{k}} S_{\mathbf{k}}]\ d\mathbf{k}, \tag{2.81}$$

where the real space integral over $\mathbf{r}$ spans the unit cell (UC). Using the normalization of the WFs $\int_{\mathrm{UC}} |\Psi_{\mathbf{k}i}(\mathbf{r})|^2 d\mathbf{r} = 1$, which leads to

$$\psi_{\mathbf{k}i}^{\dagger} S_{\mathbf{k}} \psi_{\mathbf{k}i} = 1, \tag{2.82}$$

this becomes

$$N_{\mathrm{e}} = \frac{1}{\Omega_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \sum_i n_{\mathbf{k}i} \, d\mathbf{k}. \tag{2.83}$$

This equation again implicitly sets the Fermi energy. For periodic systems therefore an independent Kohn-Sham equation is solved for each $\mathbf{k}$-point to obtain the set of eigenvalues and WFs. Using all the eigenvalues of all $\mathbf{k}$-points the Fermi energy is determined by Eq. (2.83), which then allows one to obtain the density matrix via Eqs. (2.79) and (2.80).

We conclude this section by providing the expression for the density of states (DOS). Starting from the total number of states N_{u} in the unit cell

$$N_{\mathrm{u}} = \frac{1}{\Omega_{\mathrm{BZ}}} \int_{-\infty}^{\infty} dE \int_{\mathrm{BZ}} \sum_i \delta(E - E_{\mathbf{k}i}) \, d\mathbf{k} \tag{2.84}$$

the DOS $\mathcal{N}(E)$ is defined as

$$\mathcal{N}(E) = \frac{1}{\Omega_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \sum_i \delta(E - E_{\mathbf{k}i}) \, d\mathbf{k}, \tag{2.85}$$

so that $N_{\mathrm{u}} = \int_{-\infty}^{\infty} \mathcal{N}(E) \, dE$. The total DOS can be split up in the contributions from the single orbitals, the so called projected density of states (PDOS), in the following way

$$\mathcal{N}(E) = \sum_{\mu}^{N_{\mathrm{u}}} \mathcal{N}_{\mu}(E), \tag{2.86}$$

with the PDOS for orbital μ, $\mathcal{N}_{\mu}$, defined as

$$\mathcal{N}_{\mu}(E) = \frac{1}{\Omega_{\mathrm{BZ}}} \int_{\mathrm{BZ}} \sum_i^{N_{\mathrm{u}}} \delta(E - E_{\mathbf{k}i}) \sum_{\nu}^{N_{\mathrm{u}}} \psi_{\mathbf{k}i\nu}^{*} S_{\nu\mu} \psi_{\mathbf{k}i\mu} \, d\mathbf{k}. \tag{2.87}$$

The PDOS is a measure for the contribution of the single orbitals to the total DOS.

2.2.5 Total energy

The Kohn-Sham total energy per unit cell is

$$E_{\mathrm{KS}} = Tr[\rho H] - \frac{1}{2} \int_{\mathrm{UC}} V_{\mathrm{H}}(\mathbf{r}) \rho(\mathbf{r}) d\mathbf{r} + E_{\mathrm{xc}}[\rho] - \int_{\mathrm{UC}} V_{\mathrm{xc}}(\mathbf{r}) d\mathbf{r} + \sum_{I<J} \frac{Z_I Z_J}{R_{IJ}}. \tag{2.88}$$

Once the self-consistent charge density is obtained forces and stresses can be calculated by taking the derivative of the total energy with respect to the atomic positions and lattice vectors [72, 69].

2.3 Conclusions

In this chapter, starting from a discussion of the many-body system, we have introduced the DFT formalism and also the fundamental DFT quantities such as the electron density, the Kohn-Sham Hamiltonian and the corresponding equations. In the second part the practical implementation of the DFT scheme contained in the SIESTA code has been outlined. It has been shown how functions and operators expressed in a real space representation (i.e. described as functions of $\mathbf{r}$) are converted into a tight-binding-like matrix representation by projecting them onto a localized basis set. In this matrix representation the fundamental quantities are now the density matrix, the Kohn-Sham Hamiltonian matrix and the overlap matrix. Thus the Kohn-Sham equations become matrix equations. This density matrix formalism is the framework in which the electronic transport theory is presented in chapters 4, 5 and 6.

Chapter 3

Magneto-structural properties of MnAs

MnAs is an extremely promising material for magneto-electronics, since it can grow epitaxially on GaAs [113, 114] and Si [115] forming clean and atomically sharp interfaces [115]. MnAs/GaAs heterojunctions have been extensively studied experimentally [113, 116, 117, 118], and spin injection from MnAs into GaAs has been demonstrated [48]. Bulk MnAs presents a first-order phase transition at $\sim 40^{o}$C, where it changes from ferromagnetic/hexagonal to paramagnetic/orthorhombic [49]. This magneto-structural phase transition has important implications for technological applications. While it is a major drawback for its use as spin injecting material, the magneto-elastic effects are useful for transducers [50], and their magneto-caloric properties are interesting for developing refrigeration devices [51]. The aim of this work is to use *ab initio* density functional theory (DFT) to develop an understanding of the phase transitions of MnAs, which can be compared with experiments and with existing phenomenological models.

This chapter is organized as follows. First a review of the experimental properties of MnAs is presented, and a brief description of the existing phenomenological models is given. Then the results of our *ab initio* calculations are presented and compared to experiments and phenomenological models. The magnetic and structural properties of MnAs are studied by mapping total energies onto a Heisenberg model. The exchange coupling constants are calculated for different distorted unit cells, and the Curie temperature and its dependence on the lattice parameters are evaluated in the mean field approximation. We also predict the ground state volume and lattice structure for the paramagnetic state, this allows us to formulate a model for the second phase transition between two paramagnetic states at 400 K. Finally a simple model for the susceptibility as function of temperature is given, and a semi-quantitative description of the phase diagram of MnAs will emerge.

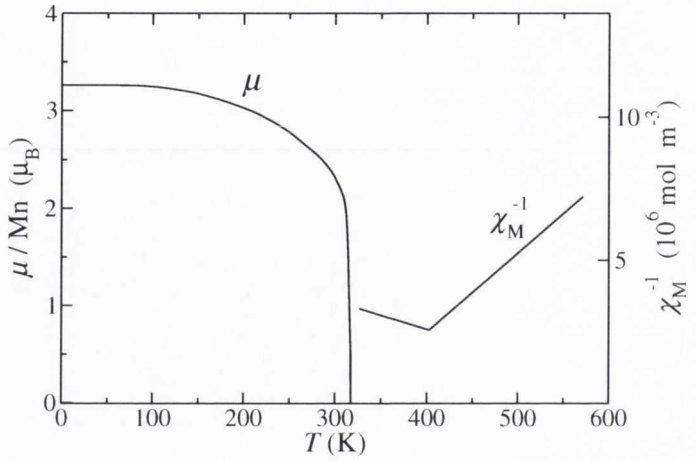


Figure 3.1: Magnetization per Mn atom μ as a function of temperature for ferromagnetic MnAs below 318 K, and inverse susceptibility χ_M^{-1} for paramagnetic MnAs above 318 K (schematically after Ref. [125]).

While most of the chapter deals with bulk properties, in the last section we present a systematic experimental and theoretical study of the first-order phase transition of MnAs thin films epitaxially grown on GaAs under biaxial tensile stress. The experiments were carried out by the group of Prof. F. Iikawa, while we performed the theoretical calculations. When MnAs is grown on GaAs the phase transition temperature depends on the the growth direction [113, 116, 119, 120]. Here we show that this can be mainly attributed to the induced strain. Our results give direct information on the dependence of the phase transition temperature of MnAs films on the lattice parameters. We demonstrate that an increase of the lattice constant in the hexagonal plane raises the phase transition temperature (T_p), while an increase of the perpendicular lattice constant lowers T_p. Our results indicate that the phase transitions of MnAs can indeed be explained by *ab initio* calculations.

3.1 Experimental properties

MnAs is a ferromagnetic metal at low temperature but it becomes paramagnetic at $T_p = 318$ K, when the magnetic moment abruptly vanishes (Fig. 3.1), the resistivity increases discontinuously [121], the volume is reduced by 2.1%, and the lattice structure changes from the hexagonal $B8_1$ (NiAs-type) to the orthorhombic B31 (MnP-type) [122, 49, 123, 124, 125]. A latent heat of 7490 J/kg is associated with this transition [49]. Hysteresis is present with a critical temperature of 307 K upon cooling and of 318 K upon heating [124]. All these properties clearly indicate a first-order phase transition.

Above T_p the distortion reduces continuously, until it vanishes at $T_t \approx 398$ K [122] where the crystal structure reverts back to the $B8_1$. There is no latent heat, but only a discontinuity in the heat capacity of the material, i.e. this phase transition is of second order. For temperatures between T_p and T_t the paramagnetic susceptibility χ has an anomalous behavior. It increases with increasing temperature until it reaches a maximum at T_t. Above T_t it decreases and has a Curie-Weiss behavior (Fig. 3.1). Moreover at T_t there is a lambda point in the specific heat [126]. Application of a magnetic field transforms the B31 structure back to the $B8_1$ above a critical field [127, 128, 50, 129, 121]. Tab. 3.1 gives a short summary of the described properties of MnAs.

	$0 < T < T_p$	$T_p < T < T_t$	$T_t < T$
Crystal structure	Hexagonal $B8_1$	Orthorhombic B31	Hexagonal $B8_1$
Magnetic order	Ferromagnetic	Paramagnetic	Paramagnetic
Magnetic moment	$3.4\mu_B$	$--$	$--$
Susceptibility χ	$--$	$\partial\chi/\partial T > 0$	$\partial\chi/\partial T < 0$

Table 3.1: Some properties of MnAs at zero pressure [124].

Fig. 3.2 shows the phase diagram. If pressure is applied T_p is lowered while T_t increases. Above the critical pressure of 4.6 kbar the ferromagnetic $B8_1$ structure becomes unstable, and the material remains in the B31 structure for all temperatures below T_t. At high pressures and low temperatures different types of ordered magnetic structures are found, with a reduced saturation magnetic moment compared to the zero pressure ferromagnetic phase. Canted spin structures, similar to the helimagnetic structures of MnP [130], are found at a pressure of 4.75 kbar below 210 K, with a local magnetic moment of about 3 μ_B. The different saturation magnetic moments for different pressures therefore correspond to different types of canted magnetic alignments. A hysteresis region lies between the ferromagnetic and the canted regions, where both the $B8_1$ and the B31 structures can be stabilized.

When the magnetic order breaks down and the system becomes paramagnetic, MnAs maintains the B31 structure for all pressures. As the temperature is further increased the structure of the cell continuously changes back towards the $B8_1$, until at T_t it has again the $B8_1$ structure, with $\partial T_t/\partial P > 0$, where P is the pressure.

The magnetocrystalline anisotropy is quite strong in MnAs, with the c-axis being the hard axis, so that the moments prefer to lie in the hexagonal plane [123]. Measurements on the magnetoelastic coupling [131] indicate that the coupling is stronger

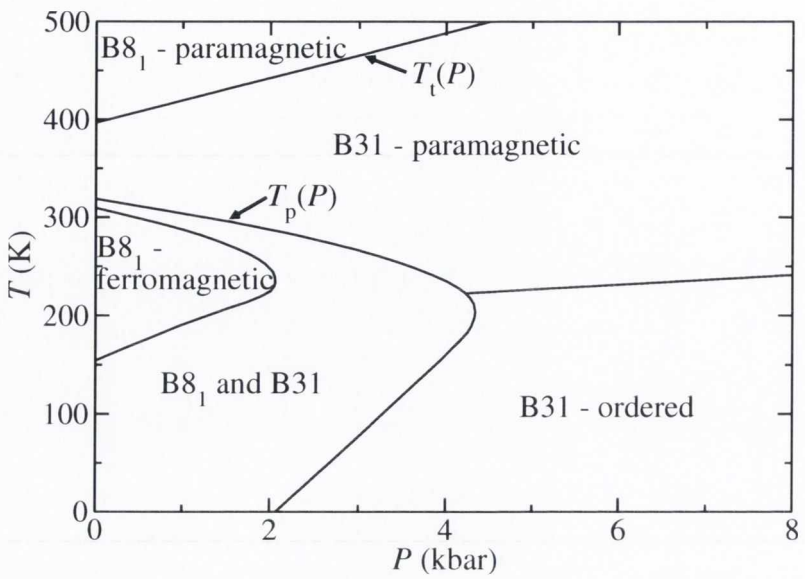


Figure 3.2: Temperature (T) versus pressure (P) phase diagram of MnAs (adapted from Ref. [125]), indicating also T_p and T_t as functions of pressure.

in the hexagonal plane than perpendicular to it.

Fig. 3.3 shows the unit cells of MnAs in the hexagonal $B8_1$ and in the orthorhombic B31 crystal structures and it defines the unit cell vectors **a**, **b** ($\mathbf{b}_h$ for the $B8_1$ structure) and **c**. The $B8_1$ structure consists of stacked hexagonal layers of Mn and As atoms, and the unit cell contains two Mn and two As atoms. The B31 structure is obtained by slightly distorting the $B8_1$, it has twice the volume of the $B8_1$ due to symmetry lowering and contains four Mn and four As atoms. The lattice is nearly hexagonal and the atoms are moved out of the hexagonal symmetry points along the **b** and **c** directions (figure 3.3(b)).

The Mn atoms are mainly displaced in the hexagonal plane along the **b** direction forming chains (Fig. 3.4), while the As atoms are displaced along the **c** axis. In each unit cell one of the planar As atoms is moved upwards and the other downwards with respect to the original position in the $B8_1$ structure, so as to keep the Mn-As distance nearly constant. The displacement u of the Mn atoms in the hexagonal plane lies between 0 and 0.05 b ($b = |\mathbf{b}|$), depending on the temperature and pressure, while the displacement v of the As atoms along the c-axis is between 0 and 0.05 c ($c = |\mathbf{c}|$). The $B8_1$ structure is a special case of the B31 structure, where $b = \sqrt{3}\ a$ and $u = v = 0$.

Therefore we choose the unit cell vectors in such a way that **a** and **c** have the same direction for both the $B8_1$ and the B31 structures. In contrast the directions of the vectors $\mathbf{b}_h$ ($|\mathbf{b}_h| = |\mathbf{a}| = a$) for the hexagonal cell and **b** for the orthorhombic cell are different.

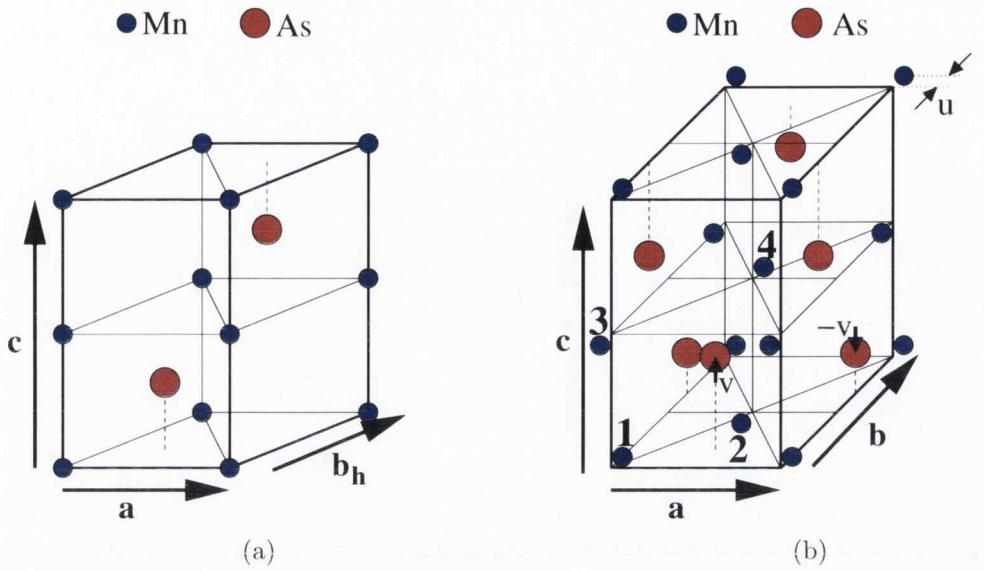


Figure 3.3: (a) $B8_1$ unit cell containing two Mn and two As atoms, (b) B31 unit cell containing four Mn and four As atoms.

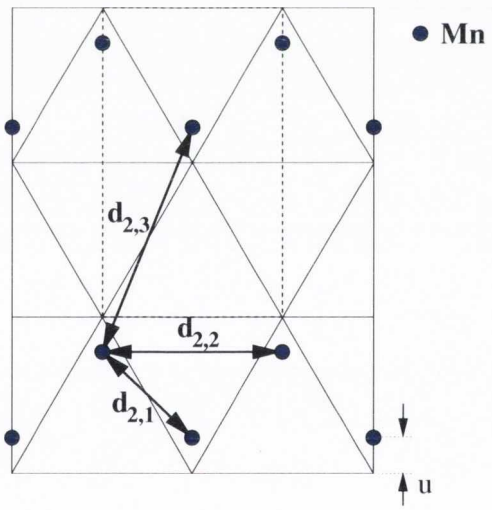


Figure 3.4: Two-dimensional representation of one layer of Mn atoms in the B31 structure. d_{ij} represent the various Mn-Mn distances. The first index i=2 indicates that all the Mn are second nearest-neighbor to each other in the $B8_1$ structure. The second index j =1, 2, 3 labels the three distances arising from the B31 distortion.

The lattice parameters of MnAs as function of temperature and magnetic field have been measured in several works [132, 122, 133, 50]. Fig. 3.5 shows the change of the lattice parameters as a function of temperature for zero pressure [122]. The lattice parameters increase with temperature due to normal thermal expansion. However the in plane lattice parameter a decreases when the temperatures get near T_p, where it jumps from 3.717 Å to the lower value of 3.673 Å. The perpendicular lattice parameter

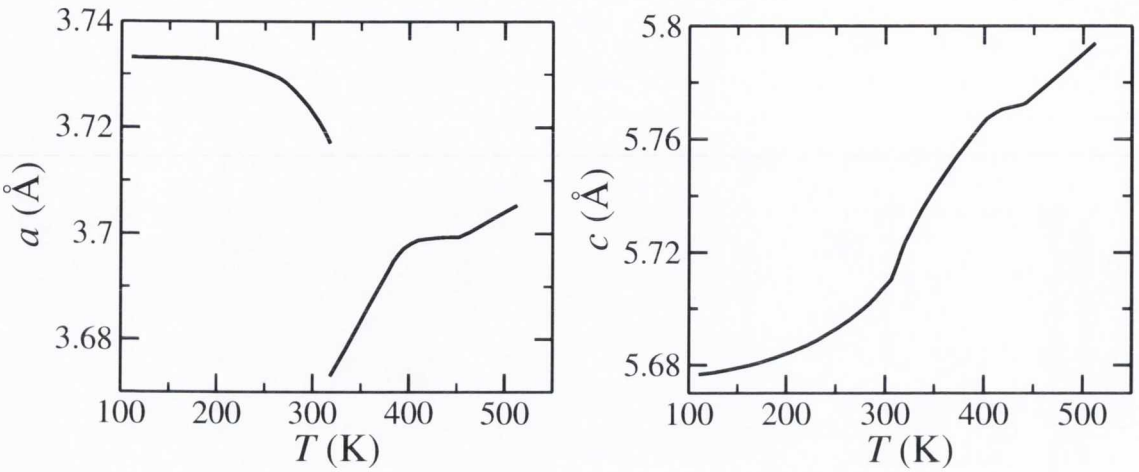


Figure 3.5: Lattice parameters a and c (see figure 3.3) as a function of temperature (adapted from Ref. [122]).

c always increases continuously with temperature. At T_t there is an inflection in the slope of the lattice parameter as a function of temperature, and at about T_s = 450 K the slope changes discontinuously. This discontinuous change of the slope depends on various experimental details, and it is found at about 410 K in other measurements[133].

The exact temperature at which the distortion disappears is somewhat uncertain, and fluctuations may play a role for small distortions. The given temperature for the disappearing of the distortion corresponds to $T_t = 398$ K [122, 133], however measurements for small distortions are difficult and such temperature can only be inferred. As pointed out in Ref. [134] the distortion should appear at temperatures slightly above T_t. Throughout this work we assume that the disappearing of the distortion occurs at T_s, which is the temperature where the thermal expansion coefficient of MnAs changes abruptly.

3.2 Review of existing models

In early theories it was believed that the first order phase transition at T_p was between a ferromagnetic and an antiferromagnetic state. Kittel proposed a model where the distortion from the $B8_1$ to the B31 structure produces the change in sign of one of the exchange coupling constants, giving rise to antiferromagnetic order [135]. However there was no experimental evidence of this antiferromagnetic state and experiments demonstrated that the transition is instead to a paramagnetic state [133].

Therefore Bean and Rodbell (BR) proposed a modification of Kittel's theory, where the exchange interactions are ferromagnetic for both structures, but they are

much weaker in the B31 phase [136, 49]. The main idea is that the exchange interaction decays strongly as the volume decreases. In order to simplify the model, the analysis was based on the extrapolated Curie temperature T_C only, and not on the details of the magnetic interaction at the atomic level. Furthermore the model neglected the anisotropic change of the crystal structure, and assumed that it is possible to describe the change from the $B8_1$ to the B31 structure by a change in volume only. The dependence of T_C on the volume V was described by the following equation

$$T_C = T_0 \left[1 + \beta \frac{(V - V_0)}{V_0} \right] , \tag{3.1}$$

where T_0 is the Curie temperature at the volume V_0, which is the volume that the system would have in the hypothetical absence of exchange interaction. In this context this corresponds to the volume of the B31 structure above T_p. β is a parameter and is determined by fitting the model to experiments. Within the BR model one can show that for certain values of β a first order phase transition between a ferromagnetic and paramagnetic state occurs with a simultaneous change of the volume. However the model does not explain the second order phase transition at T_t. Further improvements such as the introduction of a term proportional to $(V - V_0)^2$ do not change the main results [137].

Later Goodenough made an attempt to explain the anomalous behavior of the susceptibility and the second order phase transition at T_t by extending the BR model and assuming that the local magnetic moment on the Mn atoms depends strongly on the volume [125, 124]. Here the Mn atoms are assumed to be in a high spin state in the $B8_1$ structure and in a low spin state in the B31 structure [125, 124]. However measurements of the local magnetic moment show that the change in the magnetic moment at T_p is very small, therefore the Goodenough model is not applicable to this phase transition. This is probably due to the fact that the MnAs unit cell volume in any crystalline structure is too big to justify a high-spin to low-spin transition. In fact in $MnAs_{1-x}P_x$ the change from high-spin to low-spin state is observed [122, 138, 139], however the low spin state is found for unit cell volumes smaller than 120 Å^3. In contrast the unit cell of MnAs has always a volume of around 130 Å^3, and it always remains in the high-spin state.

In 1982 Kato et al. extended the BR model by taking into account not only the change in volume, but also the change of the crystal structure [140]. Furthermore instead of just using T_C for describing the magnitude of the exchange interactions, they consider exchange coupling constants up to second nearest neighbors. The obtained results are similar to those of Bean and Rodbell. The second order phase transition at T_t is explained by assuming that locally the structure above T_t is still the B31,

but that the distortions from the $B8_1$ are randomly distributed. This assumption is probably not valid, since neutron diffraction experiments clearly indicate that the structure is a regular $B8_1$ above T_t [133].

A Landau-type phenomenological model, where the free energy Φ is expanded as a function of order parameters, is given in Refs. [134, 141, 142] and references therein. In Ref. [134] just two order parameters are used, the relative magnetization σ ($\sigma = 1$ for a ferromagnetic state, $\sigma = 0$ for a paramagnetic state) and the orthorhombic distortion d. The following equation is used for the free energy:

$$\begin{aligned}\Phi(d,\sigma;T,H) = \Phi_0 + c_1\left[T - T_0\left(1-\delta_1\, d^2\right)\right]\sigma^2 + c_2\,\sigma^4 + \\ + c_3\sigma^6 + c_4\left(T - T_D\right)d^2 + c_5 d^4 - M_0\sigma H\left(1-\delta_2 d^2\right),\end{aligned} \tag{3.2}$$

where $c_1, ..., c_5$ and δ_1, δ_2 are expansion coefficients to be fitted to experiment, H is an external magnetic field, M_0 is the saturation magnetization, T_0 is the extrapolated Curie temperate of the low temperature phase, and Φ_0 is a constant. The distortion d plays the same role as the relative change in volume $(V - V_0)/V_0$ of the BR model (Eq. (3.1)). In Eq. (3.1) T_C depends linearly on the change in volume, whereas now it depends quadratically on the distortion d, $T_C = T_0(1 - \delta_1\, d^2)$. This is the correct expansion of T_C, since the linear term in d disappears due to symmetry ($+d$ and $-d$ correspond to the same distortion). The terms in σ^2, σ^4 and σ^6 appear also in the BR model and lead to the first order transition at T_p. The second order phase transition at $T_t = T_D$ is generated by the d^2 and d^4 terms in the expansion.

Also the variation of the magnetic moment with the distortion is contained in the model, although the authors find that the corresponding coefficient δ_2 is essentially zero. For T_0 the Curie temperature extrapolated from the high temperature susceptibility above T_t was considered ($T_0 = 285$ K), and for T_D the temperature at which the inverse susceptibility has a minimum ($T_D = 394$ K). By construction this model yields the correct thermodynamic behavior, and also predicts the increase of the susceptibility between T_p and T_t by means of the reduction of the distortion with increasing temperature.

Variations of this model [141, 142] give similar results. In Ref. [142] the full T-P-H phase diagram of MnAs is explained with a Landau-type expansion, where more order parameters are used. However also in this case a term equivalent to $(T - T_D)d^2$ of Eq. (3.2) is used in order to obtain the phase transition at T_D. Despite the fact that Landau-type expansions give very good agreement between theory and experiments when the right parameters are used, they do not provide insights into the origin of the terms of the expansion, especially of the $(T - T_D)d^2$ term. In Ref. [143] a basic justification of such term is given from first principles within a spin fluctuation theory

constructed from a Hubbard Hamiltonian. It is shown that for a given volume the minimum of the free energy can lie at $d = 0$ for a ferromagnetic state or at $d \neq 0$ for a paramagnetic state.

In 1986-87 Motizuki and Katoh used spin fluctuation theory in order to explain the anomalous behavior of the susceptibility between T_p and T_t [144, 145]. A Hubbard Hamiltonian was used, with model density of states obtained from first principle calculations. They could qualitatively show that the susceptibility increases when going from the B31 structure to the $B8_1$ structure, again mainly due to the fact that T_C increases with increasing temperature.

More recently various tight binding [146, 147] as well as first principles [148, 149, 150, 151, 152, 153, 154, 155, 156] calculations have been performed for MnAs in the $B8_1$ structure. The results generally agree and compare well with the experiments. Only few studies on MnAs in the B31 structure are known to the authors [157, 158, 159]. In Ref. [157] a description of the paramagnetic state of the B31 structure is given by assuming that it coincides with zero local magnetic moment of the Mn atoms. This in fact describes a low spin-state for Mn and not paramagnetism, which rather corresponds to constant magnetic moments randomly oriented by spin fluctuations. Ref. [159] is a later publication and largely confirms our findings. In Ref. [160] it is argued that the actual ground state is only approximately ferromagnetic, and that there might be a slight canting of the magnetic moments.

In the present chapter we investigate the magnetic interaction across the various phase transitions of MnAs. An explanation of the magnetostructural properties in terms of first-principles calculations is given, thereby illustrating the origin and providing a justification of the parameters used by the different models.

3.3 DFT calculations

The first principles calculations within DFT are performed using SIESTA [69]. The generalized gradient approximation (GGA) as parametrized in Ref. [85] is used for the exchange correlation potential, since it has been shown to give good structural properties for hexagonal MnAs [150, 151]. In the valence we consider $4s, 4p$ and $3d$ orbitals for Mn, and $4s$, $4p$ and $4d$ for As. For both Mn and As double ζ polarized local orbitals are used for the s and p angular momenta, whereas for the d orbitals double ζ is used. The number of k-points in the Brillouin zone is specified by a grid cutoff of 20 Å. This corresponds to a $11\times11\times8$ mesh for the $B8_1$ unit cell, giving approximately 1000 k-points in the full Brillouin zone. For the B31 unit cell such cutoff yields a $8\times11\times7$ mesh. The real space mesh cutoff, which determines the

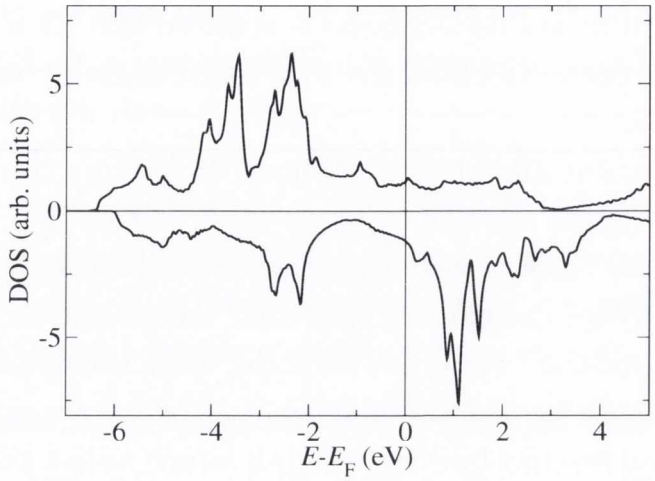


Figure 3.6: Density of states for ferromagnetic MnAs in the $B8_1$ structure.

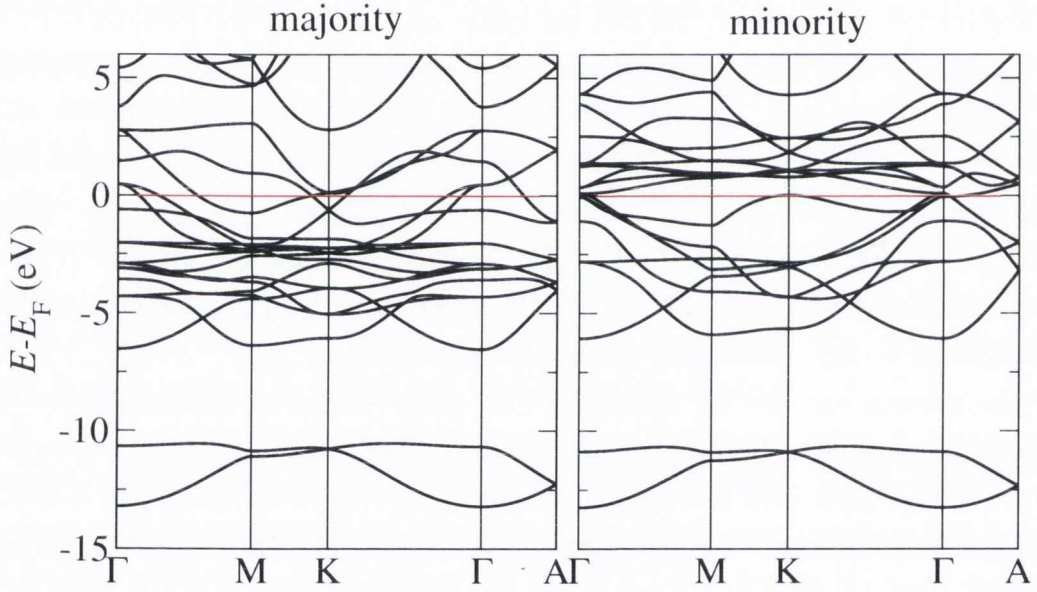


Figure 3.7: Band structure for ferromagnetic MnAs in the $B8_1$ structure

density of the real space grid, is 300 Ry.

After full relaxation of the unit cell to a pressure below 0.1 kbar, and of the atomic positions to forces smaller than 0.01 eV/Å, the B31 unit cell in the ferromagnetic configuration relaxes to a $B8_1$ structure with $a = 3.72$ Å and $c = 5.58$ Å. The experimental values at room temperature are $a = 3.724$ Å and $c = 5.706$ Å. Therefore the relative error is below 1% for a and -2% for c. The lattice parameters at 110 K can be extracted from figure 3.5 and are $a = 3.733$ Å and $c = 5.677$ Å. This demonstrates that GGA reproduces rather well the zero temperature ground state.

For a fixed c/a ratio of 1.54, which is close to the experimental value at the first order phase transition, the energy is minimized for $a = 3.695$ Å, which compares

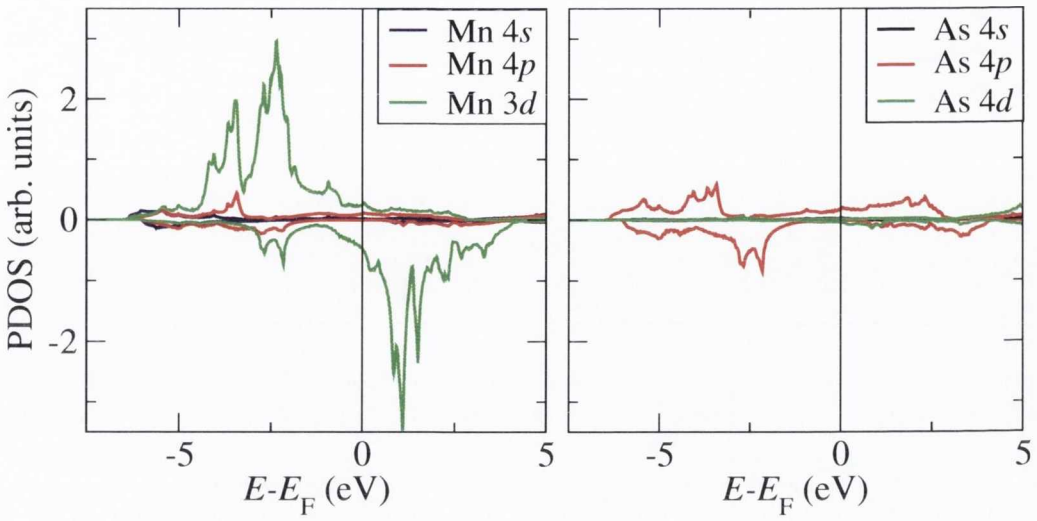


Figure 3.8: Projected density of states for ferromagnetic MnAs in the $B8_1$ structure.

well with the results of other *ab initio* calculations [148, 150, 151]. The magnetic moment per Mn atom is 3.4 μ_B, and compares well with the measured value of 3.4 μ_B [124]. Also the density of states (DOS) and the band structure are similar to previous calculations (Figs. 3.6 and 3.7), and in approximate agreement with experimental data [155, 161]. The projected density of states (PDOS, Eq. 2.87) is shown in Fig. 3.8. It can be seen that MnAs is a strong ferromagnet, i.e. one of the two Mn d spin sub-bands is completely filled. There is a strong hybridization between the Mn-3d and As-4p states. Around the Fermi level there is almost no contribution from the s and p states in the minority spin. We calculated the DOS also using the LDA instead of the GGA, and we note that also in that case the DOS is similar, the peaks of the d states are however shifted to higher energies by about 0.3 eV for the majority spin, and to lower energies by about 0.2 eV for the minority.

The unit cell of the B31 structure contains 4 Mn atoms, allowing for 3 possible independent antiferromagnetic configurations of the local moments of the Mn atoms. The different antiferromagnetic states are $++--$, $+-+-$ and $+--+$. As a matter of notation $++--$ means that the atoms 1 and 2 in the unit cell have opposite magnetic moment than that of atoms 3 and 4. The indices of the Mn atoms in the unit cell are defined in Fig. 3.3(b). A cell relaxation is performed for those three antiferromagnetic configurations. Tables 3.2 and 3.3 list the calculated relaxed structures together with the total energies per Mn atom as compared to the ferromagnetic ground state energy ($E - E_{FM}$). The structure remains of the $B8_1$ type if the local moments are ferromagnetically aligned in the hexagonal plane, whereas it changes to the B31 type if the moments are antiferromagnetically aligned in the

	a(Å)	b(Å)	c(Å)	V(Å^3)	u/b	v/c	$d_{2,1}$(Å)	$d_{2,3}$(Å)
$++++$	3.72	6.47	5.58	134.27	0.00	0.00	3.73	3.73
$++--$	3.56	6.18	5.81	127.93	0.00	0.00	3.56	3.57
$+-+-$	3.55	6.24	5.62	124.54	0.05	0.05	3.10	4.10
$+--+$	3.62	6.29	5.70	129.83	0.04	0.04	3.12	4.17

Table 3.2: Relaxed lattice parameters for different spin configurations.

plane, with u (v) of the order of 0.05 b (0.05 c). There is also a slight displacement of the Mn atoms along **c** of at most 0.01 c, and of the As atoms along **b** of at most 0.01 b. Generally it can be observed that the in-plane lattice parameters contract and the c-parameter expands for the antiferromagnetic states, resulting in a net reduction of the volume V. The calculated lattice parameters are similar to those given in Ref. [157], although the absolute value of the magnetic moments is different. This is probably due to the fact that we use the Mulliken population analysis to determine the local magnetic moment, while in Ref. [157] it is obtained by integrating the magnetization density over a sphere centered on the Mn atoms.

The total energy for the ferromagnetic alignment is the lowest, although the $+-+-$ configuration is higher by only 17 meV/Mn. This indicates that the system should evolve to one of the antiferromagnetic states when put under pressure, since those have a much smaller volume but only a slightly higher energy.

	$\mu_{\mathrm{Mn}}(\mu_{\mathrm{B}})$	$\mu_{\mathrm{As}}(\mu_{\mathrm{B}})$	$E - E_{\mathrm{FM}}$(meV)
$++++$	3.43	-0.24	0
$++--$	3.10	0.00	62
$+-+-$	3.01	-0.08	17
$+--+$	3.33	-0.03	35

Table 3.3: Local magnetic moment of the Mn and As atoms, and total energies per Mn atom for different spin configurations.

Tab. 3.2 gives also the distances between a given Mn atom and its first three nearest neighbor Mn atoms in the hexagonal plane $d_{2,1}$, $d_{2,2}$ and $d_{2,3}$ (see Fig. 3.4). Note that $d_{2,2} = a$ and it is not given explicitly. While these distances are all equal in the hexagonal case, they differ of as much as 1 Å in the B31 structure. Large changes in the distance between the Mn atoms are possible since the nearest neighbor Mn-Mn separation in MnAs is well above the inter-atomic distance 2.61 Å of bulk Mn [162], which can be regarded as the minimal possible distance between Mn atoms. The distance between nearest neighbor Mn and As atoms lies between 2.46 Å and 2.62

Å for all the different configurations, and it therefore changes much less than the Mn-Mn distance.

The local magnetic moment on the Mn (μ_{Mn}) and As (μ_{As}) atoms, calculated using the atomic Mulliken population [163], is also given in Tab. 3.3. The local moment on the Mn atoms ranges between 3.43 μ_{B} for the ferromagnetic configuration to 3.01 μ_{B} for the $+-+-$ configuration. This reduction in the local moment is mainly due to the decrease of the cell volume, and the consequent increase of the hybridization between the Mn-d and As-p orbitals.

In summary these calculations show that the distortion to the B31 structure is caused by an antiferromagnetic alignment of the local magnetic moments in the hexagonal plane.

3.4 Fit to Heisenberg energy

In order to extract the various exchange parameters, calculations are performed for three different B31 supercells using different local magnetic configurations. These supercells contain 8 Mn atoms and are obtained respectively by doubling the B31 unit cell along the **a** lattice vector (supercell 1), along **b** (supercell 2) and along **c** (supercell 3). The calculated total energies are then fitted to a model Heisenberg energy

$$E_{\mathbf{s}_1,\mathbf{s}_2,\ldots} = E_0 - \frac{1}{2}\sum_{i,j} \mathbf{s}_j \mathbf{s}_j \; J_{ij}, \tag{3.3}$$

where $\mathbf{s}_i$ is the magnetic moment of the i-th Mn ion, and the Js are coupling parameters. Here we neglect 4-moment coupling constants J_{ijkl} and the small induced magnetic moment over the As atoms. E_0 is a constant that can be associated to the energy of the paramagnetic phase. In fact if all the local magnetic moments are randomly aligned the contribution coming from $\sum_{i,j} \mathbf{s}_j \mathbf{s}_j \; J_{ij}$ vanishes. In Appendix A the approximations involved in deriving Eq. (3.3) are presented.

In mean field theory the Curie temperature T_{C} for classical Heisenberg $|\mathbf{s}_i| = 1$ exchanged magnetic moments is

$$k_{\mathrm{B}} T_{\mathrm{C}} = \sum_j J_{0j}/3 = J_0/3, \tag{3.4}$$

where k_{B} is the Boltzmann constant [164, 165], The quantity $J_0 = \sum_j J_{0,j}$ is the sum of the exchange coupling constants of a given magnetic moment with all the other moments. In the following sections Eq. (3.4) is always used to extract Curie temperatures, although it is well-known that the mean field approximation overestimates T_{C} [166, 165].

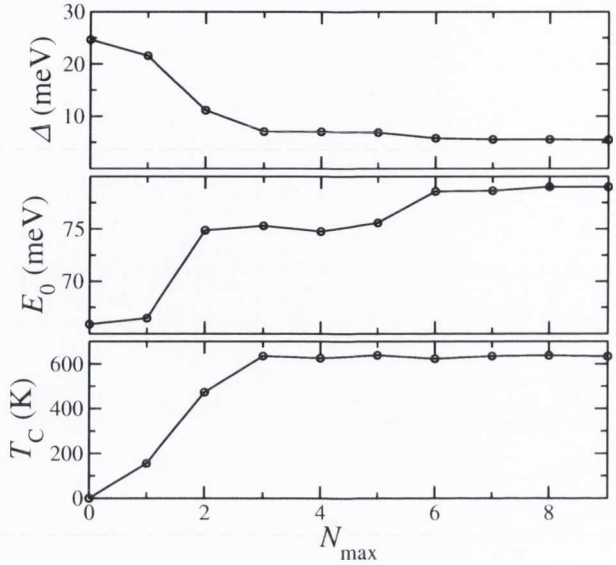


Figure 3.9: Variation of the various exchange quantities as function of the number of coupling coefficients $N_{\max}$ included in the fit. (a) Standard deviation Δ of the energies resulting from Eq. (3.3) as compared to the calculated DFT energies per Mn atom. (b) E_0 per Mn atom (Eq. (3.3)), where the zero of energy is chosen as the energy of the ferromagnetic state. (c) Mean field Curie temperature T_C.

Calculations are performed for all the independent spin configurations of the supercell 1, and for a randomly chosen subset of those of supercells 2 and 3. 35 different configurations of the magnetic moments are used in total. The energies are then fitted by a least-mean-square fit to the coupling parameters of Eq. (3.3). Since the system is metallic with the d-orbitals having finite density of states at the Fermi level, the magnetic interaction is expected to have a long range character. For the chosen supercells it is possible to extract coupling constants up to the ninth nearest neighbor. The lattice parameters used are approximately those for ferromagnetic MnAs in the $B8_1$ structure at the phase transition temperature $T_p = 318$ K (a=3.71 Å, c/a=1.54).

We carefully tested the convergence of our results with the range of the Heisenberg exchange interaction. Fig. 3.9 shows the standard deviation Δ of the energies resulting from Eq. (3.3) as compared to the calculated DFT energies per Mn atom, the value of E_0 per Mn atom, and the mean field Curie temperature T_C as a function of the number of coupling coefficients $N_{\max}$ included in the fit. The standard deviation Δ decays monotonically, remains roughly constant for $N_{\max} \geq 3$, and then reaches a minimum value of around 5 meV for $N_{\max} = 9$. This can be considered as the error resulting from neglecting high moment coupling constants. The value of E_0 changes less over the whole range, being something like an average of the energies of the different magnetic configurations.

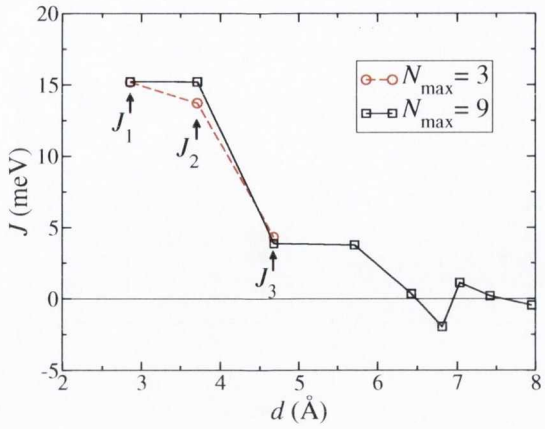


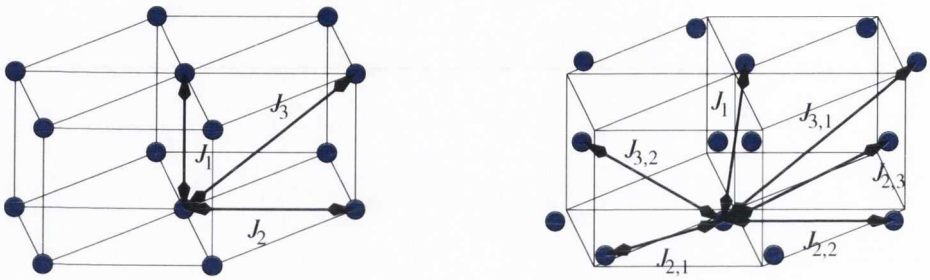
Figure 3.10: Exchange coupling parameters for $N_{\max} = 3$ and $N_{\max} = 9$ as a function of the distance between the magnetic moments.

T_C reaches a constant value of approximately 633 K for $N_{\max} \geq 3$. This indicates that the main contribution arises from the first three nearest neighbor coupling constants. The experimental value of T_C for the low temperature phase lies somewhere between $T_p = 318$ K and $T_t = 400$ K. This means that our mean field T_C overestimates the experimental one by a factor between 1.6 and 2.

Fig. 3.10 shows the calculated exchange coupling constants as a function of the distance for two different fits counting respectively 3rd and 9th nearest neighbor coupling. The first three exchange constants J_1, J_2 and J_3 (see Fig. 3.11(a)) remain nearly unchanged when going from 3rd to 9th nearest neighbor coupling. Interestingly the coupling parameters are positive and therefore ferromagnetic up to $d \sim 6.5$ Å (fifth neighbor interaction). In what follows we consider only coupling parameters up to third nearest neighbors, as they give the main contribution to the magnetic properties of the material.

3.4.1 $B8_1$ to B31 distortion at T_p

The $B8_1$ to B31 structure phase transition at T_p is investigated by calculating the Heisenberg coupling constants for different distorted cells. We start from the $B8_1$ with the experimental lattice parameters near T_p ($a = 3.71$ Å, $b = \sqrt{3}\ a$, $c = 1.54\ a$, $u = v = 0$) and distort the cell linearly to the B31 structure. The amount of distortion d is given in percent, where d =0% stands for the lattice parameters of the ferromagnetic $B8_1$ cell just below T_p, and $d = 100\%$ for the paramagnetic B31 cell above T_p ($a = 3.676$ Å, $b = 1.01\sqrt{3}\ a$, $c = 1.556\ a$, $u = 2.71\ 10^{-2}\ b$, $v = 2.45\ 10^{-2}\ c$) [133]. Calculations are done for distortions between 0% and 220%. Note that the volume decreases with increasing distortion. For these calculations only the supercells 1 and 2



(a) (b)

Figure 3.11: Schematic representation of the atomic positions of the Mn atoms together with the exchange constants for the $B8_1$ (a), and the B31 (b) structures.

are used with a total of 26 different spin configurations. The standard deviation of the fit is approximately constant for all the distortions and is of the order of 5 meV/Mn.

Fig. 3.11(a) shows the Mn atoms of the $B8_1$ structure coupled by first (J_1), second (J_2) and third (J_3) nearest neighbor interaction. In the distorted B31 structure the three coupling constants J_1, J_2 and J_3 are split into six different constants due to symmetry loss. While there is still only one J_1 coupling, the in-plane J_2 splits into three different coupling constants $J_{2,1}$, $J_{2,2}$ and $J_{2,3}$, corresponding to different distances between the Mn atoms in the hexagonal plane (see figure 3.4). Moreover also the third nearest neighbor coupling J_3 splits into three different constants, although two of them are between Mn atoms separated by approximately the same distance at T_p, and so they are assumed to be identical. Hence J_3 effectively splits only into $J_{3,1}$ and $J_{3,2}$.

Fig. 3.12 shows the calculated values for the exchange parameters as a function of the distortion. For 0% distortion the values of $J_{2,1}$, $J_{2,2}$ and $J_{2,3}$ are approximately equal reflecting the hexagonal symmetry. The values of $J_{3,1}$ and $J_{3,2}$ also should be identical although they differ by about 2 meV (note that in the fit we do not force the $B8_1$ symmetry when determining the Js for the undistorted structure, i.e. they are fitted without symmetry constraints). This can be assumed to be the error over the fit. Additional control fits were also performed for different subsets of the 26 spin configurations. The variation over the Js was of 20%, whereas the variation of J_0 was always smaller than 6%.

The value of J_1 remains approximately constant for all the distortions, reflecting the fact that the distance between the quasi-hexagonal layers remains roughly constant. In contrast the in-plane Js change and eventually become antiferromagnetic. In particular the coupling becomes strongly antiferromagnetic for $J_{2,1}$, i.e. for those

Mn atoms that get closer in the hexagonal plane under distortion. Also $J_{2,3}$, which couples the Mn atoms increasing their separation, is reduced and becomes antiferromagnetic for large distortions. Finally the coupling parameters $J_{3,1}$ and $J_{3,2}$ have only minor changes, with $J_{3,1}$ becoming weakly antiferromagnetic for large distortions.

The evolution of the coupling constants with the distortion indicates why for the $+-+-$ and $+--+$ spin configurations (see tables 3.2 and 3.3), where the spins are antiferromagnetically aligned in the hexagonal planes, the lowest energy is found for the B31 structure. The relaxed structure for both spin configurations is similar to a distortion of about 200%. At this distortion the in-plane coupling constants $J_{2,1}$ and $J_{2,3}$ become antiferromagnetic, resulting in a reduction of the total energy as compared to the $B8_1$ structure for those spin configurations. In contrast for the $++++$ and $++--$ configurations, that have a ferromagnetic alignment of the moments in the hexagonal plane, the $B8_1$ structure is stable, since for that structure $J_{2,1}$ and $J_{2,3}$ are positive.

Fig. 3.13(a) shows the relative change of the mean field Curie temperature $T_C(d)/T_C(0)$ for the ferromagnetic state. T_C decreases monotonically with increasing distortion. For 100% distortion (B31 structure at T_p) $T_C(100\%)/T_C(0) = 0.67$, demonstrating that when the phase transition from the $B8_1$ to the B31 structure occurs, the system in the B31 cell is already paramagnetic with very little magnetic order. The experimental Curie temperature T_C^{exp} for the hexagonal cell at $T = T_p$ is not known, since the structure changes.

Fig. 3.13(b) shows the total energy per B31 unit cell as a function of the distortion in the ferromagnetic (FM) and in the $+--+$ antiferromagnetic configurations (AF). This latter is the antiferromagnetic configuration giving the lowest total energy at its minimum among all the ones calculated along the considered distortion. The figure also shows the value of E_0, the energy of the paramagnetic state [see Eq. (3.3)]. The zero in the energy scale is the energy of the ferromagnetic state for $d = 0\%$.

The ferromagnetic state has its energy minimum for $d = 0\%$, and increases parabolically for increasing distortion. This means that the $B8_1$ structure is the one with lowest energy in the ferromagnetic state. In contrast the competing antiferromagnetic configuration has a minimum for about 180% distortion, where the energy is lower than the ferromagnetic phase. The crossing between the two energy curves occurs at about $d \approx 142\%$. The ground state of the system is therefore expected to be ferromagnetic for small distortions, and to become antiferromagnetic for very large distortions. In the intermediate region the ground state can be expected to be some canted spin structure.

E_0 has a very flat minimum for distorted cells, reflecting the fact that the total

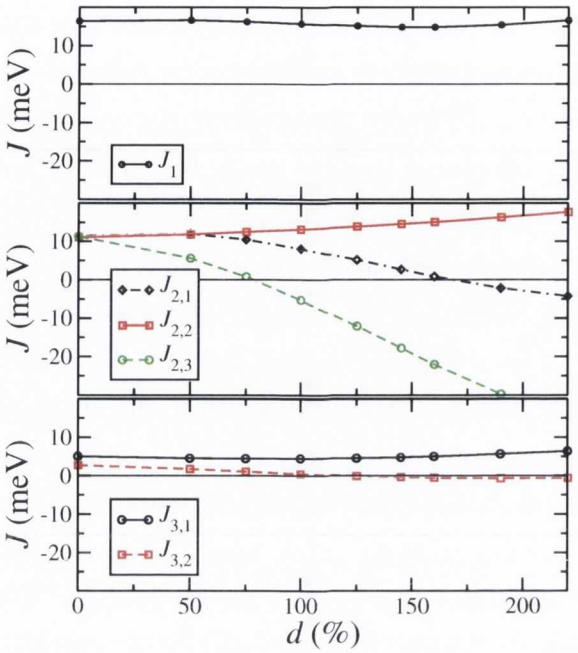


Figure 3.12: Evolution of the exchange coupling constants when distorting the unit cell linearly from the $B8_1$ structure to the B31 structure. $d = 0\%$ represents the $B8_1$ structure at $T_p = 318$ K, $d = 100\%$ represents the B31 structure at T_p. A positive (negative) value of J means ferromagnetic (antiferromagnetic) coupling.

energy increases for the ferromagnetic state, but decreases on average for the antiferromagnetic states. The minimum is found to be at about $d \approx 100\%$, which corresponds indeed to the lattice parameters of the paramagnetic state above the phase transition. This therefore indicates that a structural change from the $B8_1$ to the B31 structure occurs when there is a transition from the ferromagnetic to the paramagnetic state. Our calculations therefore correctly predict the structural transition at T_p.

3.4.2 B31 to $B8_1$ distortion at T_t

For a temperature in between T_p and T_t the MnAs crystal structure continuously changes from B31 back to $B8_1$. As mentioned in Sec. 3.1 the phase transition temperature T_t is usually identified as the temperature where the susceptibility and the specific heat have a maximum. This is at about 398 K. However the distortion should disappear at slightly higher temperatures as pointed out in Ref. [134]. Therefore, since the exact temperature for this second order structural phase transition is not known exactly, we introduce an operative definition and assume that the distortion disappears at a temperature T_s, at which the slope of the in-plane lattice constant as function of temperature $a(T)$ changes discontinuously (see Fig. 3.5). According to

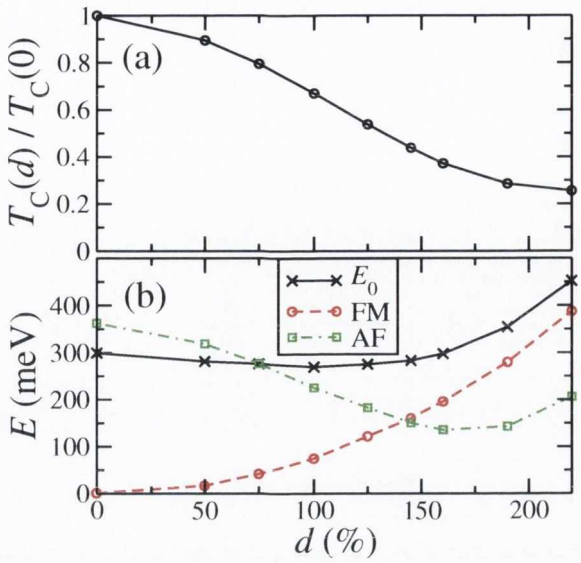


Figure 3.13: (a) Relative change of the mean field Curie temperature $[T_C(d) - T_C(0)]/T_C(0)$ for the ferromagnetic state. (b) Total energy for one B31 unit cell for the ferromagnetic configuration (FM), for the $+--+$ antiferromagnetic configuration (AF), together with E_0 (equation (3.3)), as a function of the distortion d. $d = 0\%$ and $d = 100\%$ represent respectively the $B8_1$ and the B31 phase at $T_p \approx 318$ K.

Fig. 3.5 the lattice constant at T_s is $a(T_s) = \tilde{a} \approx 3.699$ Å, and $(\partial a/\partial T)_{T_s} \approx 0$. At T_t the same lattice parameter is $a(T_t) \approx 3.697$ Å, so that the difference in $\tilde{a}$ is very small.

We argue that the main reason for the second order phase transition at high temperatures is related to the lattice thermal expansion. The idea is that upon volume expansion, the ground state of the paramagnetic phase moves towards the hexagonal structure. We verify this hypothesis by calculating the minimum of E_0 ($E_0^{\min}$) along a distortion of the cell transforming $B8_1$ to B31. In the calculation the volume of the cell is kept constant and we repeat the calculation for different volumes. This allows us to evaluate both $E_0^{\min}$ and the corresponding distortion as a function of the volume. Since for $T > T_p$ MnAs is always paramagnetic, the minimum of E_0 corresponds to the stable distortion d_0 at a given volume. In practice the change in volume can be described simply by the change in the planar lattice constant a, since both b/a and c/a do not deviate much from their value at T_p. Thus we always consider $b = \sqrt{3}\ a$ and $c = 1.556\ a$ and the phase transition is investigated as a function of a only.

The equilibrium distortion d_0 as a function of a is presented in figure 3.14 (filled circles). Indeed the distortion decreases with volume and it disappears for a between $a = 3.695$ Å and $a = 3.71$ Å. Moreover we find 100% distortion for $a \sim 3.66$ Å. These

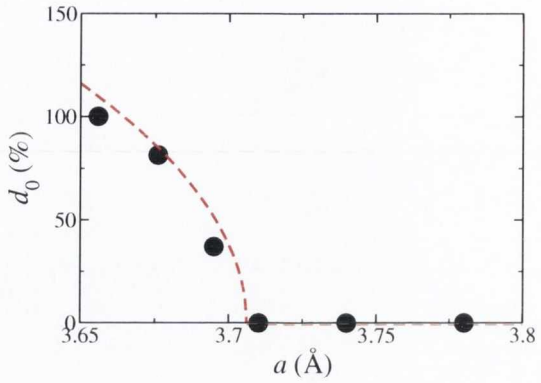


Figure 3.14: Distortion d_0 for the minimum of the paramagnetic ground state as a function of the lattice constant a. The dots are calculated values, the dashed line is a fit with equation (3.5).

values agree rather well with the experimental ones, where the distortion disappears at about $a(T_s) = 3.699$ Å, and 100% distortion is found at $a(T_p) = 3.673$ Å.

In order to interpret these results consider that the distortion is symmetric for $\pm d$ $[E_0(d) = E_0(-d)]$, and therefore E_0 can be expanded in even powers of the distortion $E_0(d) = r_0 + r_1\ d^2 + r_2\ d^4$. Here r_i are parameters to fit to the DFT calculations. In particular note that r_0 corresponds to the energy of the paramagnetic phase when the crystal is undistorted, i.e. it has the hexagonal structure. In this way the minimum of the $E_0(d)$ curve is obtained for $d_0 = \sqrt{-\frac{r_1}{2r_2}}$ if $r_1 < 0$, and for $d_0 = 0$ for $r_1 \geq 0$. We now define $\tilde{a}$ as the lattice constant where $r_1 = 0$. For small distortions the parameters r_1 and r_2 can then be further expanded around $\tilde{a}$ as $r_1 = r_{1,1}(a - \tilde{a})$ and $r_2 = r_{2,0} + r_{2,1}\ (a - \tilde{a}) + r_{2,2}\ (a - \tilde{a})^2$. In order to obtain the leading terms in this expansion, we calculate the value of r_0, r_1 and r_2 for different lattice constants (a=3.656, 3.676, 3.695, 3.71, 3.74 and 3.78 Å). We then first calculate $\tilde{a}$ and $r_{1,1}$ by mapping r_1 onto the expansion $r_1(a) = r_{1,1}(a - \tilde{a})$, and with the obtained value of $\tilde{a}$ we can then in the same way also calculate the expansion coefficients of r_2. The so calculated values for the leading terms are $\tilde{a} = 3.706$ Å, $r_{1,1} = 62.4\ 10^{-3}$ meV/Å and $r_{2,0} = 1.29\ 10^{-7}$ meV (d is given in percent). The equilibrium distortion d_0 up to first order in a is then

$$d_0(a) = \gamma \sqrt{\frac{\tilde{a} - a}{\tilde{a}}}\ \Theta(\tilde{a} - a), \qquad \gamma = \sqrt{\tilde{a} \frac{r_{1,1}}{2\ r_{2,0}}}, \tag{3.5}$$

where $\Theta(x)$ is the Heaviside function. With the values of $\tilde{a}$, $r_{1,1}$ and $r_{2,0}$ given above $\gamma = 947$ is obtained. The resulting distortion is presented in Fig. 3.14 (dashed line).

Importantly, if we use Eq. (3.5) to fit the experimentally determined distortions at $a(T_s) = 3.699$Å ($d_0 = 100\%$) and $a(T_p) = 3.673$Å ($d_0 = 0$), we obtain $\gamma = 1184$ and $\tilde{a} = 3.699$, both in good agreement with our calculated values. This suggests that the

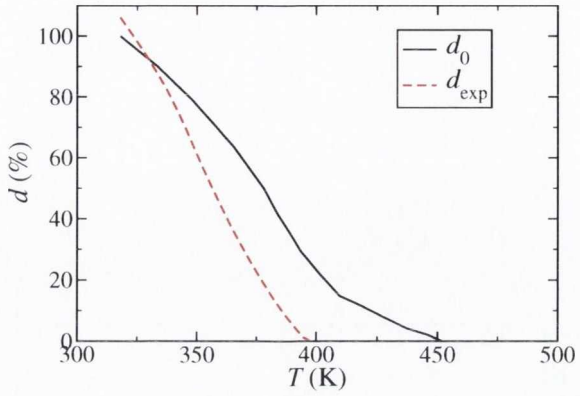


Figure 3.15: Solid line: Distortion d_0 for the minimum of the paramagnetic ground state as a function of temperature calculated with equation (3.5) (with $\gamma = 1184$ and $\tilde{a} = 3.699$ Å) using $a(T)$ taken from Fig. 3.5. Dashed line: experimental distortion extracted from $v(T)$ of Fig. 3 in Ref. [122].

main effects of the distortion to the B31 structure arise from the atomic displacement from the symmetry positions, and that small changes of the ratio of the lattice vectors, neglected in our calculations, play only a secondary role. Using the values $\gamma = 1184$ and $\tilde{a} = 3.699$ Å of the two parameters the evolution of the distortion as a function of temperature T can be obtained by inserting the data for $a(T)$ from Fig. 3.5 in Eq. (3.5). The result is shown in Fig. 3.15 (full line). The figure also shows the experimental distortion d_{exp} obtained from $v(T)$ extracted from Fig. 3 of Ref. [122] (dashed line). The main difference between the two curves is that the distortion d_{exp} becomes zero at 398 K, whereas in our results this happens only at 450 K. This is due to our choice of $T_s = 450$ K, which by definition sets the temperature where the distortion disappears. Close to the phase transition temperature fluctuations play an important role, so that for very small distortions close to the phase transition the description may not be valid.

By using the computed values of r_0, r_1 and r_2 the minimum of E_0 [$E_0^{\min}$=$E_0(d = d_0)$] is calculated as a function of the volume of the unit cell, and it is shown in Fig. 3.16 together with r_0. Recalling that r_0 is the energy of the paramagnetic hexagonal phase, it also can be expanded as function of the lattice constant $r_0 = \epsilon_0 + \epsilon_1 \, (a - a_0)^2$, where a_0 is the equilibrium lattice constant of the hexagonal phase. This, combined with Eq. (3.5), gives an expression for the energy minimum as a function of the lattice constant a

$$\begin{aligned} E_0^{\min}(a) &= \epsilon_0 + \epsilon_1 \, (a - a_0)^2 - \epsilon_2 \, (a - \tilde{a})^2 \, \Theta(\tilde{a} - a), \\ \epsilon_2 &= \frac{r_{1,1}^2}{4 r_{2,0}} = 7546 \, \frac{\text{meV}}{\text{Å}^2}, \end{aligned} \tag{3.6}$$

where ϵ_0, ϵ_1 and a_0 are to be fitted from the calculations of r_0 (Fig. 3.16). The fitted

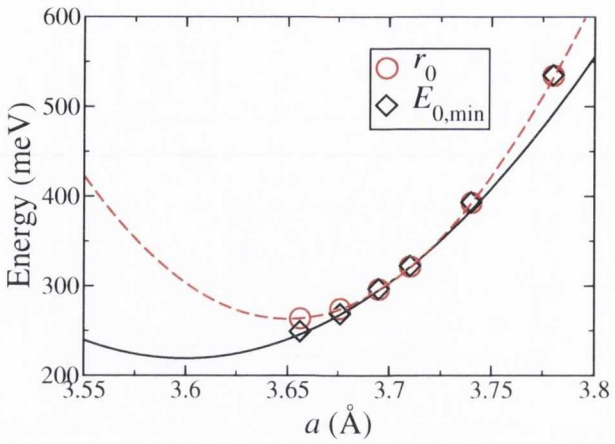


Figure 3.16: Energy expansion coefficient r_0 and $E_{0,\min}$ as function of the lattice constant. The dashed line shows $\epsilon_0 + \epsilon_1 \, (a - a_0)^2$, the solid line shows $\epsilon_0 + \epsilon_1 \, (a - a_0)^2 - \epsilon_2 \, (a - \tilde{a})^2$ (see Eq. (3.6)).

values are $\epsilon_0 = 264$ meV, $\epsilon_1 = 15935$ meV/Å^2 and $a_0 = 3.65$ Å. From equation (3.6) the energy minimum is easily found

$$a_{\min} = a_0 \left(1 - \frac{\epsilon_2}{\epsilon_1 - \epsilon_2} \frac{\tilde{a} - a_0}{a_0} \right) , \tag{3.7}$$

and by using the calculated parameters we estimate $a_{\min} = 3.60$ Å. Since $a_{\min} < a_0$, we derive the important result that the distortion allows the volume to be further reduced as compared to the hexagonal phase. Furthermore the curvature of the energy as a function of a is

$$\frac{\partial^2 E_0^{\min}(a)}{\partial a^2} = 2\epsilon_1 - 2\epsilon_2 \, \Theta \left(\tilde{a} - a \right) , \tag{3.8}$$

which is also reduced by a factor $2\epsilon_2$ when the structure is distorted.

The effect of the thermal expansion on the lattice parameter can now be modeled as a temperature dependent position of a_0, $a_0 = a_0(T)$. The change of the lattice constant with temperature for the distorted phase can therefore be written as

$$\frac{\partial a_{\min}}{\partial T} = \frac{\epsilon_1}{\epsilon_1 - \epsilon_2} \frac{\partial a_0}{\partial T} \tag{3.9}$$

Since ϵ_2 is smaller than ϵ_1 the lattice expands with temperature faster for the distorted phase than for the undistorted phase. If the calculated values for ϵ_1 and ϵ_2 are used the ratio $\epsilon_1/(\epsilon_1 - \epsilon_2)$ is found to be 1.90, which agrees well with the value of 2.25 extracted from Fig. 3.5. Near the phase transition, i.e. where $d \approx 0$, phononic effects due to the different curvatures of the energy and fluctuations should be considered. It is especially interesting that the change of the lattice constant with the temperature goes to zero near the phase transition temperature.

In the same way as E_0 also T_C can be expanded as a function of the lattice constant and of the distortion

$$T_C(a,d) = T_C(\tilde{a}) \left[1 + K_v \frac{a-\tilde{a}}{\tilde{a}} - K_d\, d^2\right], \tag{3.10}$$

where K_v and K_d are parameters, and $T_C(\tilde{a})$ is the Curie temperature for the cell with lattice parameter $\tilde{a}$ and where the atoms are in the hexagonal positions. Note that this equation is a generalization of the Bean-Rodbell model equation [Eq. (3.1)] and of the Landau expansion in Eq. (3.2), since the change in volume and the distortion are now included as independent quantities. The parameters are fitted by a least mean squares fit to the calculated values of T_C obtained for six lattice constants ranging between 3.656 Å and 3.78 Å and for different distortions. We obtain $T_C(\tilde{a})$=573 K, K_v = 6.80 and K_d = 2.62 10^{-5}. At d = 100% the relative change of the Curie temperature is $T_C(\tilde{a}, d = 100\%)/T_C(\tilde{a})$ = 0.74 and corresponds roughly to the value of Fig. 3.13(a). In this case it is slightly larger due to the fact that the volume is kept constant, whereas for the calculations of Fig. 3.13 it shrinks with increasing the distortion.

Next we calculate the dependence of the magnetic moment on the distortion and on the unit cell volume. The dependence is again expanded to lowest order in a and d

$$\mu(a) = \mu(\tilde{a}) \left[1 + \alpha_\mu \frac{a-\tilde{a}}{\tilde{a}} + \alpha_{\mu,d}\, d^2\right] . \tag{3.11}$$

We now have different ways of extracting the magnetic moment of the Mn atoms from our DFT calculations. One possibility is to take the total moment of the cell for the ferromagnetic spin configuration and divide it by the number of Mn atoms. In this way however the small induced moments of the As atoms are subtracted from the moment on the Mn. A second possibility is to take the average Mulliken spin population for the Mn atoms. The advantage of this method is that also antiferromagnetic configurations can be used to determine the average moment, and the induced moments of the As atoms are accounted for. The drawback however is that Mulliken populations are somewhat arbitrary as they depend on the basis set.

By setting $\tilde{a}$ to 3.699 Å, the values obtained using the cell moment are $\mu(\tilde{a})$ = 3.28 μ_B, α_μ=3.28 and $\alpha_{\mu,d}$=−1.15 10^{-6}. Similarly from the average Mulliken population over all the magnetic configurations we obtain $\mu(\tilde{a})$=3.42 μ_B, α_μ=3.48, and $\alpha_{\mu,d} \approx 0$. These results are rather similar to each other. With α_μ=3.28 and $\alpha_{\mu,d}$ = −1.15 10^{-6} a reduction of the lattice constant of 1%, as approximately found at T_p, results in a reduction of the magnetic moment of about 3%, and a distortion of

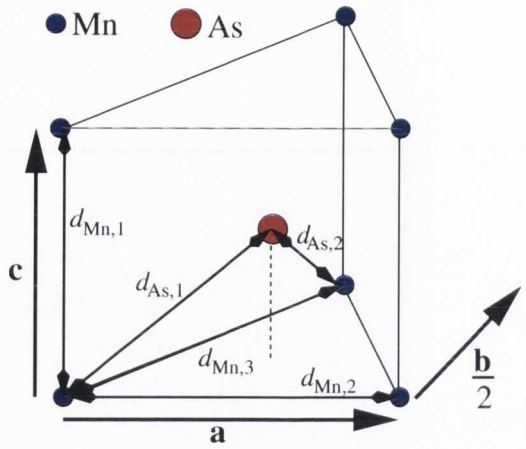


Figure 3.17: Diagram of the positions of Mn and As atoms in orthorhombic MnAs.

$d = 100\%$ results in a reduction of μ of about 1%. Both effects therefore reduce the magnetic moment and are of the same order of magnitude.

3.5 Small distortions of the $B8_1$ structure

In this section the dependence of T_C on the individual lattice parameters and on the distances between the atoms is investigated for the $B8_1$ structure. Our approach is to distort the cell orthorhombically but to leave the atoms in their high symmetry positions. Apart from a general understanding of the phase diagram of MnAs this analysis is useful for predicting the behavior of MnAs when grown on a substrate. For instance when grown on GaAs(001) the substrate induces strain in MnAs, and the unit cell is slightly orthorhombically distorted [113]. This distortion does not correspond to the orthorhombic B31 structure, since the atoms do not move out of the high symmetry positions. Moreover different growth orientations are possible, and the Curie temperature varies accordingly [119]. In the second part of this section we present experimental results, obtained by our collaborators, showing that the phase transition temperature T_p changes when strain is applied to the MnAs film, and compare the data with our model.

When the cell is orthorhombically distorted the first three nearest neighbor coupling constants split into five different constants, corresponding to the ones of the B31 cell (Fig. 3.11), and with the only exception that now $J_{2,1} = J_{2,3}$. The number of total energy calculations used for the fit of the coupling parameters is 16. The change of T_C for each different distortion can be expressed as a function of the change

of each single Mn-Mn and Mn-As distance in the unit cell. This gives

$$\frac{\delta T_C}{T_C} = \frac{\delta J_0}{J_0} = \sum_{\mu=1}^{5} K_\mu N_\mu \frac{\delta d_\mu}{d_\mu}, \tag{3.12}$$

where the sum goes over all five independent distances in the orthorhombic unit cell as defined in Fig. 3.17. The dependence on the angles between the atoms is neglected. N_μ are the multiplicities of each distance d_μ within one unit cell, and have the values $N_{\mathrm{Mn},1} = 4$, $N_{\mathrm{Mn},2} = 4$, $N_{\mathrm{Mn},3} = 8$, $N_{\mathrm{As},1} = 4$, $N_{\mathrm{As},2} = 8$. For the evaluation of the K_μ coefficients, 21 different distortions are considered, including changes of volume, changes of the ratio of the different axes and different displacements of the As atoms. For all the distortions the orthorhombic symmetry however was preserved. The best fit gives

$$\begin{aligned} K_{\mathrm{Mn},1} &= 6.6,\ K_{\mathrm{Mn},2} = 6.2,\ K_{\mathrm{Mn},3} = 5.9, \\ K_{\mathrm{As},1} &= -7.5,\ K_{\mathrm{As},2} = -7.5. \end{aligned} \tag{3.13}$$

The values of $K_{\mathrm{Mn},1}$ and $K_{\mathrm{Mn},2}$ are almost identical, as one should expect from the symmetry and similarly for $K_{\mathrm{As},1}$ and $K_{\mathrm{As},2}$.

Eq. (3.12) describes the fact that the change of T_C is the result of an interplay between the change of the Mn-Mn and Mn-As distances. The calculated K_μ show that, while an increase in the distance between Mn atoms increases T_C, an increase of the Mn-As distance decreases it. However note that the two distances can not be changed independently, thus the net change in T_C depends on the details of the distortion.

J_0 can also be expanded over the orthorhombic lattice parameters

$$\frac{\delta J_0}{J_0} = \sum_{i=1}^{3} K_i \frac{\delta a_i}{a_i}, \text{ with } K_i = \sum_{\mu=1}^{5} K_\mu N_\mu \frac{a_i}{d_\mu} \frac{\partial d_\mu}{\partial a_i}, \tag{3.14}$$

where $a_1 = a$, $a_2 = b$ and $a_3 = c$. The change of the distances between Mn and As atoms is not exactly known for the orthorhombic cell. However it is easy to show that to first order the position of the As atom in the cell does not influence J_0, since up to first order $\delta d_{\mathrm{As},2} = 2\delta d_{\mathrm{As},1}$ when moving the As atom inside the cell. Therefore the As atoms can be assumed to remain in the high symmetry position. Assuming now $K_2 = K_3$ and $K_4 = K_5$ (as imposed by symmetry), the general form of the K_i is

$$\begin{aligned} K_a = K_b &= 6K_2 + \frac{96K_4}{16 + 3\,\tilde{c}^2}, \\ K_c &= 4K_1 + \frac{36K_4\,\tilde{c}^2}{16 + 3\,\tilde{c}^2}, \end{aligned} \tag{3.15}$$

where $\tilde{c} = c/a$. Using the average between K_2 and K_3 and $\tilde{c} = 1.533$ this gives

$$K_a = K_b = 4.9, \quad K_c = -1.0 \; . \tag{3.16}$$

Our results clearly show that stretching the unit cell along the basal plane raises T_{p} (since $K_a > 0$), while stretching along the c-axis lowers T_{p} ($K_c < 0$). An increase of the volume without distorting the cell results in an increase of the ferromagnetic exchange interactions and therefore of T_{C}, since K_a is positive and larger in magnitude than K_c. If the cell changes only its volume the expansion corresponds to the one of Eq. (3.10) with a factor $K_v = 2K_a + K_c = 8.8$. We note that this value is somewhat different from the value of $K_v = 6.8$ given in Sec. 3.4.2. This is due to the different type of analysis performed in this section, which starts from the dependence of T_{C} on the distances between the single atoms. The difference is however small and can be considered a measure for the error in our results.

3.5.1 Comparison with experiment

Here we present a direct comparison of our theoretical predictions with experimental results for the magneto-structural phase transition of MnAs films grown on GaAs. Those films present a non-abrupt phase transition with the coexistence [167] of the two phases in the form of periodically alternating stripes [168, 169] for a large temperature range (~20 °C) [167, 168, 169, 170]. As a result of this phase coexistence, a considerable fraction of the volume of the MnAs epitaxial films is usually in the paramagnetic phase at ~30 °C, which is a strong limitation for applications on room temperature spintronic devices. The growth of MnAs films on different crystal orientations has been suggested as an alternative that can provide higher phase-transition temperatures [170]. The detailed mechanism that associates the crystal distortion (lattice parameter variation) with the phase transition temperature is, however, still unclear,

In order to get more insight into the MnAs phase transition, the experimental group of Prof. F. Iikawa has designed an x-ray diffraction experiment to investigate in detail the dependence of the phase transition temperature on the crystal distortions. Their experimental results, part of which are presented in this section, show a clear variation of the phase transition temperature of MnAs films due to the lattice distortions imposed by the substrate and by the application of an external biaxial stress. They observed that an increase of the unit cell volume does not necessarily result in an increase of the transition temperature, contrary to the predictions of the Bean-Rodbell model [49].

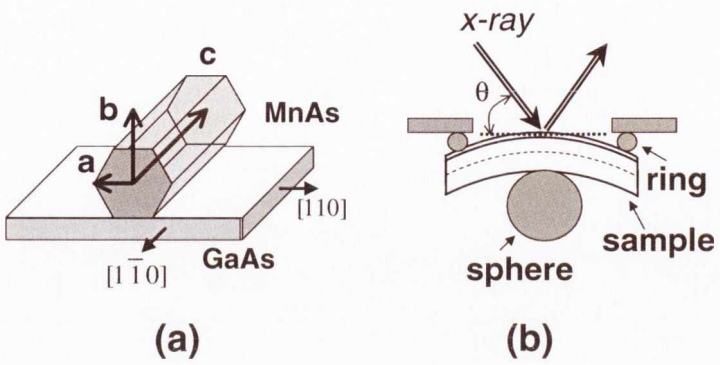


Figure 3.18: (a) Schematic diagram of a unit cell of the A-orientation MnAs/GaAs(001). The lattice parameters **a**, **b** and **c** are defined in the orthorhombic structure notation (see Fig. 3.3(b)). (b) Diagram of a biaxial stress ring cell. The X-ray beam was focused at the center of the ring.

The MnAs films were epitaxially grown on GaAs(001) by molecular beam epitaxy under growth conditions that yield A-oriented films [171], as shown in the Fig. 3.18(a). The orthorhombic unit cell vectors as defined in Fig. 3.3(b) are used. This is more appropriate for MnAs film on GaAs, since the films are deformed by the substrate and lose the hexagonal symmetry to an orthorhombic one even in the ferromagnetic phase. In the following the ferromagnetic phase for both the hexagonal as well as for the slightly distorted structure is denoted as α-phase, whereas the B31 phase is denoted as β-phase. The results presented here were obtained for a 130 nm thick MnAs film. External biaxial strain was applied using a stress ring cell [172] assembled in a temperature controlled continuous N_2 gas flow cryostat. The sample is mounted between a ring and a sphere [Fig. 3.18(b)] and biaxial strain is achieved by moving the sphere towards the ring [172]. The film (much thinner than the substrate), which is located on the upper side of the sample, undergoes a symmetric biaxial tensile strain at the center of the ring. The strain is approximately uniform within ~ 1 mm of diameter from the center, which is of the order of the X-ray beam spot.

At intermediate temperatures around the phase transition temperature X-ray spectra indicate the coexistence of the two phases. Measurements for cooling cycles gave similar results to the ones for the heating cycle, with a small shift of the phase transition to lower temperatures, resulting in a thermal hysteresis. This was also observed in previous works [170]. Fig. 3.19(a) shows the experimentally measured volume fraction F_α *vs.* temperature of the α-MnAs phase for a 130 nm thick MnAs film, with and without an external strain. The fraction F_α was obtained from the ratio of the integrated intensity of the α phase X-ray diffraction peak relative to the sum of the total intensity of the peaks for both α- and β-phases. F_α presents

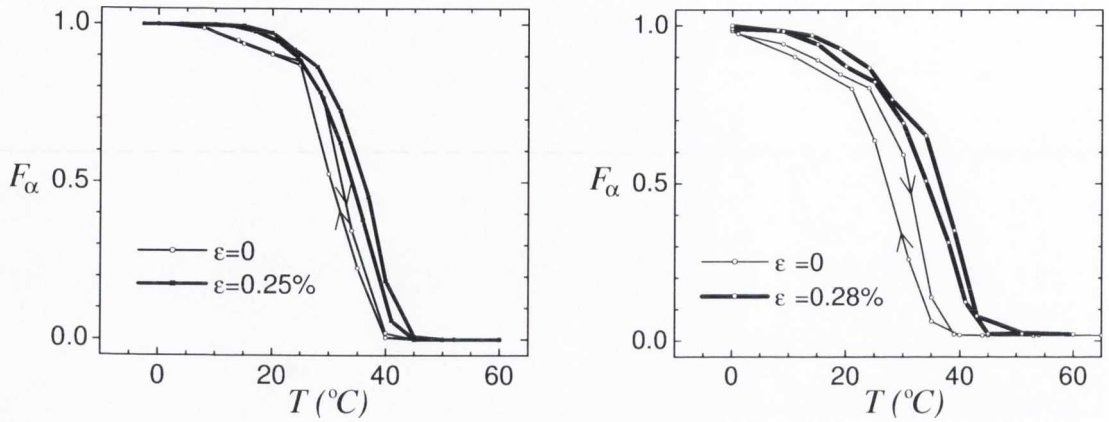


Figure 3.19: Temperature dependence of the volume fraction of α-phase F_α of the 130 nm MnAs film (a) and the 160 nm MnAs film (b), with and without induced strain.

a non-abrupt profile with a thermal hysteresis, as discussed before. Note that the phase transition of the films starts at temperatures smaller than that of the transition temperature of the bulk (40°C). When an external tensile strain is applied, the curves show a clear shift to higher temperatures. Similar results were observed for a 160nm MnAs/GaAs film [Fig. 3.19(b)]. Due to the phase coexistence across a large range of temperatures the phase transition temperature is not well defined for the film. Therefore we choose to define a characteristic transition temperature, T_p, as the temperature where F_α is 0.5.

	MnAs Bulk	film $\epsilon_\parallel$=0	film $\epsilon_\parallel$=0.25%
a (Å)	3.724	3.710 (-0.37%)	3.719 (-0.13%)
b (Å)	6.450	6.440 (-0.15%)	6.434 (-0.25%)
c (Å)	5.706	5.764 (+1.0%)	5.778 (+1.27%)
V (Å^3)	137.06	137.72 (+0.48%)	138.26 (+0.88%)
$\frac{\delta T_P}{T_P}$ (%)	0.0	-3.2	-1.6
$\frac{\delta J_0}{J_0}$ (%)	0.0	-3.0	-2.4

Table 3.4: Lattice parameters, unit cell volume (V) and relative temperature variation $\frac{\delta T_P}{T_P}$ of the α-MnAs phase at T = 20 °C for bulk [132] and for the MnAs film with a strain in the film plane of $\epsilon_\parallel = 0$ and for $\epsilon_\parallel = 0.25\%$ [120] ($b = \sqrt{3}\ a$ at the perfect hexagonal structure). In parenthesis the lattice parameter variation relative to bulk is shown. $\frac{\delta J_0}{J_0}$ are our theoretical data.

In Tab. 3.4 the experimentally measured room-temperature lattice parameters of the MnAs film, as-grown and under external strain, are presented together with those of bulk MnAs taken from Ref. [132]. The lattice parameters of the as-grown film were obtained from specular and grazing incident diffraction (GID) measurements.

Comparing the lattice parameters between bulk and as-grown film, we observe that the MnAs film is stretched along the c direction and is shrunk in the hexagonal basal plane. Similar results were observed in a previous work [170]. The resulting unit cell volume in the film is 0.48% larger than that in the bulk and presents a smaller T_p. When the film is under external tensile strain, the unit cell volume shows a further increase (becoming 0.88% larger than the bulk). In this case, however, the increase of the cell volume is accompanied by an increase of T_p. Bean and Rodbell [49] have predicted an enhancement of T_p upon increasing the MnAs unit cell volume. Our results show, however, that T_p is not a simple function of the cell volume and it does depend on the lattice parameters in a more complex way.

The parameter $J_0 = \sum_j J_{0j}$, which in the mean field approximation is proportional to the Curie temperature, is a measure of the stability of the ferromagnetic phase [49, 140, 173]. The larger the gain in magnetic energy, the higher should be T_p. As a first approximation, we assume that the change of T_p is proportional to the equivalent change of J_0: $\frac{\delta T_p}{T_p} = \frac{\delta J_0}{J_0}$. Tab. 3.4 gives the calculated values of $\frac{\delta J_0}{J_0}$ for the film, with and without an external stain, and its relative change with respect to the bulk. We observe that the calculated values of $\frac{\delta J_0}{J_0}$ are in very good agreement with the corresponding experimental values of $\frac{\delta T_p}{T_p}$ for the as-grown sample. The agreement is not equally good for the film under external biaxial strain. In this case the in-plane parameters were calculated from the out-of-plane one using the elastic tensor, which carries a significant experimental uncertainty. We remark in particular that the theory correctly predicts the direction of the transition temperature changes due to the film distortions: a decrease of T_p from bulk to the as-grown film and a partial recovery of T_p when a biaxial tensile strain is applied to the film.

We may now interpret the experimental results. The strain obtained by growing a MnAs thin film with the A-orientation on GaAs results in the stretching of the c axis and the contraction of both the a and b axes. All these distortions contribute to reduce T_p. When an external biaxial stress is applied along the film plane, the lattice parameters a and c increase while b decreases. The net result is a slight enhancement of T_p, reflecting the balance of the K parameters and the relative large value of K_a. Our calculations also explain why a hydrostatic pressure lowers the phase transition temperature of bulk MnAs, as observed in previous works [49]. Under hydrostatic pressure, the a and b parameters suffer a stronger contraction than c due to the anisotropy of the MnAs elastic constant tensor [174]. This effect, combined with the fact that $|K_a|$ and $|K_b|$ are bigger than $|K_c|$, results in a decrease of T_p.

Based on our results we propose that higher T_p can be achieved by growing MnAs in such a way as to maximize the lattice parameters of the basal plane and minimize

the lattice parameter in the perpendicular direction. The possibility of moving across the phase transition temperature by applying an external biaxial pressure opens a new avenue to mechanically driven magneto-devices. One can envision composite materials comprising MnAs and piezoelectric compounds, where the ferromagnetism can be mechanically switched on and off by an external stress.

3.6 Curie temperature and susceptibility

It is now possible to analyze two peculiar properties of MnAs. The first is the anomalous behavior of the susceptibility χ as a function of temperature between T_{p} and T_{t}. The second is the fact that although the Curie temperature for ferromagnetic MnAs has to be larger than T_{p}, the T_{C} extrapolated from the susceptibility above T_{t} is only 285 K [124]. In this section both these features are explained using the dependence of the susceptibility on the Curie temperature [$\chi = \chi(T_{\mathrm{C}})$], and the strong dependence of the Curie temperature on the lattice parameters found in the previous sections [$T_{\mathrm{C}} = T_{\mathrm{C}}(a, b, c, d)$]. $T_{\mathrm{C}}(T)$ and $\chi(T)$ are therefore determined using the experimentally measured temperature dependence of the lattice vectors $a(T), b(T)$ and $c(T)$ and of the distortion $d(T)$. This analysis also provides a tool for extracting the parameters K_v and K_d from experimental data.

By generalizing Eqs. (3.10) and (3.14) the Curie temperature $T_{\mathrm{C}}(a, b, c, d)$ can be written as

$$T_{\mathrm{C}}(a, b, c, d) = T_{\mathrm{C},0}\left[1 + K_a\left(\frac{a-\tilde{a}}{\tilde{a}} + \frac{b-\tilde{b}}{\tilde{b}}\right) + K_c\,\frac{c-\tilde{c}}{\tilde{c}} - K_d\,d^2\right], \tag{3.17}$$

where we use the fact that $K_a = K_b$ and take the values for K_a and K_c from Eq. (3.16), while $K_d = 2.62\ 10^{-5}$ and $T_{\mathrm{C},0} = T_{\mathrm{C}}(\tilde{a}) = 573$ K are those calculated in Sec. 3.4.2. The reference lattice parameters are chosen to be the lattice vectors at T_{s} ($\tilde{a} = 3.699$ Å, $\tilde{b} = \sqrt{3}\,\tilde{a}$, $\tilde{c} = 1.56\,\tilde{a}$).

Similarly to the Curie temperature also the susceptibility is calculated in the mean field approximation. This is justified for $T \gg T_{\mathrm{C}}$, a condition which is satisfied for paramagnetic MnAs. The molar susceptibility χ_{M} is then given by

$$\begin{aligned} \chi_{\mathrm{M}}^{-1} &= \frac{1}{C_0}\,(T - T_{\mathrm{C}})\,, \\ C_0 &= \frac{N_{\mathrm{A}}\mu_{\mathrm{B}}^2 g^2}{3k_{\mathrm{B}}}\,s\,(s+1)\,. \end{aligned} \tag{3.18}$$

N_{A} is the Avogadro's number, $g \approx 2$ is the Lande's factor for the free electron spin, k_{B} is the Boltzmann constant and s is the atomic total spin. Note that the susceptibility

has an additional temperature dependence since T_C and s depend on the temperature through the lattice distortion. However in what follows we neglect the dependence of s on the lattice parameters so that C_0 is constant over all temperature. An analysis performed by relaxing this approximation gives similar results.

Similarly to Sec. 3.4.2 the model is further simplified by assuming that b/a and c/a are constant above T_p. As indicated in Eq. (3.5) d is a function of the lattice constant, $d(a) = \gamma\sqrt{(\tilde{a}-a)/\tilde{a}}\ \Theta(\tilde{a}-a)$. Moreover we have shown that the experimental distortion as function of the lattice constant is well reproduced when $\gamma = 1184$. Therefore it is now possible to express T_C and χ_M^{-1} as a function of the lattice constant a only

$$\chi_M^{-1}(a) = \frac{1}{C_0}\left[T - T_{C,0}\left(1 + \left(K_v + K_d\,\gamma^2\Theta\left(\tilde{a}-a\right)\right)\frac{a-\tilde{a}}{\tilde{a}}\right)\right], \tag{3.19}$$

where $K_v = 2K_a + K_c = 8.8$. This equation shows that if the lattice expands strongly with temperature χ_M^{-1} decreases.

For temperatures between T_p and about 390 K, as well as above T_s, a increases approximately linearly with temperature (see Fig. 3.5) and can therefore be written as

$$a(T) = a(T_0)\left[1 + \alpha\,\frac{T-T_0}{T_0}\right], \tag{3.20}$$

where the experimental values for the coefficients are $T_{0,+} = T_s = 452$ K, $a(T_s) = \tilde{a} = 3.699$ Å and $\alpha_+ = 0.0126$ for temperatures above T_s (the index "+" denotes the high temperature region above T_s), and $T_{0,-} = T_p = 318$ K, $a(T_p) = 3.673$ and $\alpha_- = 0.0284$ for temperatures between T_p and about 390 K (the index "−" denotes the intermediate temperature region). By inserting equation (3.20) into Eq. (3.19) we obtain for the high temperature region above T_s

$$\chi_M^{-1}(T) = \frac{1}{C_{\text{eff}}}(T - T_{C,\text{eff}}), \tag{3.21}$$

with

$$\begin{aligned} C_{\text{eff}} &= \frac{1}{1 - K_v\,\alpha_+\frac{T_{C,0}}{T_s}}\,C_0, \\ T_{C,\text{eff}} &= \frac{1 + K_v\alpha_+}{1 - K_v\alpha_+\frac{T_{C,0}}{T_s}}\,T_{C,0}. \end{aligned} \tag{3.22}$$

$T_{C,\text{eff}}$ and C_{eff} are the experimentally accessible quantities for the high temperature susceptibility, and due to the expansion of the lattice they are different from $T_{C,0}$ and C_0. The experimentally measured values are $T_{C,\text{eff}} = 285$ K and $C_{\text{eff}} = 3.12\ 10^{-5}$ m^3

K [125], which corresponds to an effective magnetic moment of 3.57 μ_B. From C_{eff} and $T_{\text{C,eff}}$ the values of $T_{\text{C},0} = T_{\text{C}}(\tilde{a})$ and C_0 can now be obtained

$$\begin{aligned} C_0 &= \frac{1 - K_v \alpha_+}{1 - K_v \alpha_+ \left(1 - \frac{T_{\text{C,eff}}}{T_s}\right)} \, C_{\text{eff}}, \\ T_{\text{C},0} &= \frac{1}{1 - K_v \alpha_+ \left(1 - \frac{T_{\text{C,eff}}}{T_s}\right)} \, T_{\text{C,eff}}. \end{aligned} \tag{3.23}$$

All the variables on the right hand side of Eq. (3.23) can be obtained from experiment except K_v. For small K_v the difference between $T_{\text{C},0}$ and $T_{\text{C,eff}}$ is proportional to K_v. Since T_s is larger than the experimental value of $T_{\text{C,eff}}$ the effect of the thermal expansion of the hexagonal structure is a reduction of the slope of the inverse susceptibility as a function of temperature, as well as a reduction of the extrapolated Curie temperature as compared to the real Curie temperature.

In the region where linear expansion holds the slope of the inverse susceptibility above T_t is

$$\frac{\partial \chi^{-1}_{\text{M},+}}{\partial T} = \frac{1}{C_0} \left(1 - K_v \, \alpha_+ \frac{T_{\text{C},0}}{T_t}\right), \tag{3.24}$$

whereas for in the intermediate temperature region above T_p it is

$$\frac{\partial \chi^{-1}_{\text{M},-}}{\partial T} = \frac{1}{C_0} \left(1 - \left(K_v + K_d \, \gamma^2\right) \, \alpha_- \frac{a_0}{\tilde{a}} \frac{T_{\text{C},0}}{T_{0,-}}\right). \tag{3.25}$$

In both regions there is a reduction of the slope due to the expansion of the lattice. However the reduction is much larger for $\chi_{\text{M},-}$ than for $\chi_{\text{M},+}$, since there is the additional term proportional to K_d due to the distortion, and also $\alpha_- > \alpha_+$. As a rough approximation it can be assumed that $(a_0 \, T_{\text{C},0})/(\tilde{a} \, T_{0,-}) \approx 1$, so that $\frac{\partial \chi^{-1}_{\text{M},-}}{\partial T}$ becomes negative for

$$\left(K_v + K_d \, \gamma^2\right) \alpha_- > 1 \, . \tag{3.26}$$

The values of α_- and γ are determined experimentally and describe how the structure changes with temperature, whereas K_v and K_d describe how T_{C} varies for distorted cells. By using our calculated values for K_v and K_d we obtain $(K_v + K_d \, \gamma^2) \, \alpha_- = 1.29$. This is indeed larger than one. Therefore we do predict a negative slope for the inverse susceptibility in the intermediate temperature region. With the value for $\gamma = 947$ estimated in Sec. 3.4.2 we obtain $(K_v + K_d \, \gamma^2) \, \alpha_- = 0.92$, which is still smaller than 1. This is due to the underestimation of the distortion as compared to experiments by using $\gamma = 947$. With $\gamma = 1186$ the reduction of the distortion with increasing lattice constant is large enough to obtain a negative slope of the inverse susceptibility,

whereas $\gamma = 947$ underestimates the reduction of the distortion, and so the slope of the inverse susceptibility is reduced but not to the point to make it negative.

Finally we extract the values for K_v and K_d from the experimental behavior of the Curie temperature. Since the ratio between K_a and K_c can not be obtained from the thermal properties of MnAs, it is therefore assumed that $K_c/K_a = -1/4.9 \approx -0.2$ is fixed and corresponds to our calculated value. For the hexagonal cell ($d = 0$) Eq. (3.17) reads

$$T_C(a,c) = T_{C,0}\left[1 + K_v\left(k_a\frac{a-\tilde{a}}{\tilde{a}} + k_c\,\frac{c-\tilde{c}}{\tilde{c}}\right)\right], \tag{3.27}$$

where $k_a = 2/(2 + K_c/K_a) = 1.11$ and $k_c = (K_c/K_a)/(2 + K_c/K_a) = -0.11$. This has to be valid for all temperatures where the cell is hexagonal, therefore it can not be assumed that the ratio between c and a is constant since it changes abruptly from 1.533 to 1.556 at T_p. By using the expression for $T_{C,0}$ from Eq. (3.23) we obtain K_v

$$K_v = \frac{\frac{T_C(a,c)}{T_{C,\mathrm{eff}}} - 1}{\alpha_+\left(\frac{T_C(a,c)}{T_{C,\mathrm{eff}}} - \frac{T_C(a,c)}{T_s}\right) + k_a\,\frac{a-\tilde{a}}{\tilde{a}} + k_c\,\frac{c-\tilde{c}}{\tilde{c}}}. \tag{3.28}$$

As reference Curie temperature the extrapolated value to room temperature is used, which can be estimated to be about $T_C(a = 3.724$ Å, $c = 1.533a)$=360 K. By inserting the experimental values for the parameters on the right hand side of Eq. (3.28) we obtain $K_v = 18.1$. This is about twice as big as our predicted value. The disagreement may partly be due to the fact that the ratio between K_c and K_a has been fixed for our calculated value. By inserting this value for K_v in Eq. (3.23) we obtain $T_{C,0} = 311$ K, the Curie temperature for the lattice parameters at $T = T_s$.

In order to extract K_d from experiments we use the relative change in the slope of the inverse susceptibility around T_t. This is, according to the Eqs. (3.24) and (3.25),

$$\frac{\frac{\partial\chi^{-1}_{M,+}}{\partial T} - \frac{\partial\chi^{-1}_{M,-}}{\partial T}}{\frac{\partial\chi^{-1}_{M,+}}{\partial T}} = 1 - \frac{1 - (K_v + K_d\,\gamma^2)\;\alpha_-\frac{a_0}{\tilde{a}}\frac{T_{C,0}}{T_0}}{1 - K_v\;\alpha_+\frac{T_{C,0}}{T_t}}, \tag{3.29}$$

and increases with increasing K_v and K_d. All the variables in this equation can be derived from the experimental measurements, except K_v and K_d. Experimentally different values are found for the relative change of the slope [left-hand side of Eq. (3.29)] [134, 175, 176, 177]. These are all of the order of 1.44. By using this value for the relative change of the slope and the previously calculated value $K_v = 18.1$, K_d is found to be 1.78 10^{-5}. This value agrees approximately with our predicted value of 2.62 10^{-5}.

In conclusion Tab. 3.5 summarizes the parameters calculated in this work, by comparing our *ab initio* results obtained from the DFT calculations and the Heisenberg model, with the results obtained by fitting to the experimental data. In general

the value obtained from the *ab initio* calculations roughly agree with the best fit to experiments, although the DFT results underestimate K_v and overestimate K_d.

	$T_{C,0}$ (K)	K_d (10^{-5})	K_v	K_a	K_c	γ	$\tilde{a}$ (Å)
DFT	579	2.62	8.8	4.9	-1.0	947	3.706
FIT	311	1.78	18.1	10.1	-2.1	1184	3.699

Table 3.5: Main parameters used in the description of the phase diagram of MnAs. We compared results obtained from *ab initio* calculations and Heisenberg model (DFT), with those of the best fit of the experimental properties (FIT).

3.7 Phase diagram

It is now possible to draw a qualitative description of the phase diagram of MnAs (see Fig. 3.2). Since the magnetic interactions are strongly dependent on the volume, the effects of the thermal expansion of the lattice are important. First the state of MnAs at zero pressure is considered. For $P = 0$ and temperatures below 155 K both ferromagnetic and canted spin structures can be stabilized. The ground state is the ferromagnetic $B8_1$ structure, however also the B31 structure is stable, because at these low temperatures the thermal energy is not large enough to induce a phase transition from a canted spin structure to a ferromagnetic one. As the temperature increases above 155 K the thermal fluctuation may induce the phase transition, and therefore only the ferromagnetic state survives.

As the temperature is further increased and the system becomes paramagnetic, the B31 structure becomes the stable one, as described in Sec. 3.4.1 [Fig. 3.13(b)]. MnAs then remains paramagnetic for all temperatures, but because of the thermal expansion the orthogonal distortion decreases, until it vanishes when a critical volume is reached (see Fig. 3.14). Then the volume becomes large enough to stabilize the $B8_1$ structure even for a paramagnetic state, leading to the second structural phase transition at T_t.

As pressure is applied the volume is reduced and thus the ferromagnetic exchange interaction decreases. Therefore T_p decreases with pressure, while T_t increases. At a temperature of approximately 230 K the ferromagnetic state is stable up to a pressure of 2 kbar. For high pressures above 2.5-4 kbar the volume is small enough that a canted spin structure becomes the ground state. The temperature at which this state

becomes paramagnetic is approximately 230 K. It increases slightly with pressure, reflecting the fact that the antiferromagnetic interaction increases as the volume is reduced.

3.8 Conclusions

We have investigated, by means of *ab initio* electronic structure calculations, the magneto-structural properties of MnAs. The stable structure for the ferromagnetic state is found to be the $B8_1$ structure. However if antiferromagnetic alignment in the hexagonal plane is imposed the B31 structure becomes more stable. By fitting the DFT total energies of different magnetic configurations to a Heisenberg type energy it is shown that the main contributions to the physical properties originate from the exchange coupling parameters up to 3rd nearest neighbor. The Curie temperature was calculated in the mean field approximation, with values approximately twice as large as the experimental ones.

The main assumption of the phenomenological model of Bean and Rodbell [49] that the ferromagnetic exchange coupling parameters increase when the volume is increased has been confirmed [Eq. (3.16)] using this analysis. However it has been shown that the exchange interactions depend not only on the volume, but that the orthogonal distortion to the B31 structure plays an important role. For the experimentally observed distortions some of the in-plane exchange coupling coefficients become antiferromagnetic. This is the reason for the stability of the B31 structure for those configurations of the magnetic moments that have an antiferromagnetic component in the hexagonal plane. Different canted spin structures are expected to minimize the energy for different distortions, since there are both positive and negative exchange coefficients depending on the amount of distortion of the B31 structure.

Furthermore it has been shown that for paramagnetic states the B31 structure is stable at small volumes, while the $B8_1$ structure is stable above a critical lattice constant of about 3.7 Å. This explains the second order phase transition at T_t, since at that temperature the lattice constant crosses this critical value.

The Curie temperature has been expanded as a function of the lattice vectors and of the amount of distortion. An increase in the volume leads to an enhancement of the Curie temperature, while an increase of the distortion leads to a reduction. With these results in hand the increase of the susceptibility between T_p and T_t has been explained as the result of the increase of the Curie temperature due to the change of the structure from the B31 to the $B8_1$ and to the increase of the volume. By using the experimental variation of the lattice parameters with rising temperature

the susceptibility is indeed found to increase between T_p and T_t.

A fit of the dependence of the Curie temperature on the lattice parameters to best reproduce the experimental behavior is also given. The calculated values agree within a factor two with the values obtained from *ab initio* calculations. Our results are in agreement with the various phenomenological models based on the Bean Rodbell idea. In our work the various parameters used in those models for the magneto-structural properties of MnAs have been derived from first principles and therefore validated.

Chapter 4

Non equilibrium Green's functions

In this chapter the main concepts of electronic transport calculations within the non equilibrium Green's functions (NEGF) formalism are presented. These are used in this work to obtain the steady state electronic transport properties of nano-devices. The NEGF formalism allows one to describe the out of equilibrium state of a quantum system, which consists of a scattering region, also denoted as extended molecule (EM), attached to semi-infinite electron reservoirs, also called leads or electrodes. Each of these reservoirs is assumed to be in local equilibrium, so that the states in the leads are occupied up to the local Fermi energy. The state of the scattering region is then determined by the rate of in- and outflow of electrons from the leads. The NEGF formalism allows the calculation of three central properties of quantum mechanical systems out of equilibrium: first it allows one to obtain the spectrum of states in the scattering region for a given Hamiltonian; secondly it makes possible to calculate the occupation of these states in the scattering region when the leads are out of equilibrium; thirdly one can calculate the current through the device for given chemical potentials of the leads and charge density of the scattering region. The NEGF formalism is nowadays widely used for the calculation of the electronic transport properties of nanoscale devices, overviews of its application can be found in Refs. [55, 178]).

There exist a number of publications where the NEGF equations are presented and their physical meaning is explained [54, 179, 180, 181, 55, 182], these give a good understanding of the dependence of the properties on the different device parameters. A formal derivation of the fundamental equations, based on perturbation theory within second quantization and a time-dependent approach, can be found in the literature [183, 184, 185, 186, 187, 188, 189, 190]. As an alternative, in Refs. [190, 191, 192, 193] the formalism is derived once again for the steady state without using a time dependent formulation. This derivation is valid in the case of elastic coherent transport and a single particle Hamiltonian, which is always the case for

the systems considered here. In this chapter a derivation of the NEGF formalism is presented based on this approach, following the concepts of Refs. [54, 194, 190, 193].

First some general features of Green's functions in electronic structure calculations are presented. Then the Green's function for the particular system consisting of a scattering region attached to semi-infinite leads is introduced. The spectral function and the density of states for such a system are presented in a Green's function formulation, followed by the equations for the occupation of the states in the scattering region when the leads are out of equilibrium. The way to obtain the current through the scattering region is outlined, and the transmission coefficient is introduced. Then the equations are expanded to the case where the leads are two-dimensional (2D) surfaces. Finally some notes on the implementation of the NEGF formalism in the *ab initio* electronic transport code SMEAGOL are given [61, 68]. SMEAGOL is the code used in this work.

4.1 Green's functions: basic properties

A description of the use of Green's functions in solid state physics can be found in a number of books [194, 54, 183, 182], here we just present the general properties of Green's functions relevant to the NEGF formalism. We assume that the system is described by the single-particle Hamiltonian H of Eq. (2.61)

$$H\ \psi_n = E_n\ S\ \psi_n, \tag{4.1}$$

where the N_u-dimensional vectors ψ_n are the single-particle wave functions with eigenvalue E_n, and S is the overlap matrix. Since both H and S are Hermitian the E_n are real. The ψ_n are normalized in such a way that the orthogonality relation becomes $\psi_n^\dagger S\psi_m = \delta_{nm}$. The corresponding completeness relation is then $\sum_n \psi_n\psi_n^\dagger S = \mathbb{1}_{N_\mathrm{u}}$, where $\mathbb{1}_{N_\mathrm{u}}$ is the $N_\mathrm{u} \times N_\mathrm{u}$ identity matrix. The retarded Green's function G is then defined as [194, 54]

$$G(E) = [(E + i\delta)S - H]^{-1}, \tag{4.2}$$

with $\delta \to 0^+$. The advanced Green's function G^A is defined in the same way with $\delta \to 0^-$, so that for real energies E we have $G^\mathrm{A} = G^\dagger$. The Hamiltonian can be written in its spectral representation as

$$H = \sum_{n=1}^{N_\mathrm{u}} E_n\ S\psi_n\psi_n^\dagger S, \tag{4.3}$$

so that the corresponding retarded GF is

$$G(E) = \sum_{n=1}^{N_\mathrm{u}} \frac{1}{E + i\delta - E_n}\psi_n\psi_n^\dagger. \tag{4.4}$$

We note that if H and S are both real, then also the ψ_n can be made real, so that $G(E)$ is symmetric and $G^{\mathrm{A}}(E) = G^*(E)$. The spectral function is defined as

$$A(E) = i\left[G(E) - G^\dagger(E)\right], \tag{4.5}$$

and it can be seen as a generalized density of states [54]. In fact, by using the spectral representation of $G(E)$ it can be rewritten as

$$\begin{aligned} A(E) &= i\sum_{n=1}^{N_{\mathrm{u}}} \frac{1}{E+i\delta-E_n}\psi_n\psi_n^\dagger - \frac{1}{E-i\delta-E_n}\psi_n\psi_n^\dagger & (4.6)\\ &= 2\sum_{n=1}^{N_{\mathrm{u}}} \frac{\delta}{(E-E_n)^2+\delta^2}\psi_n\psi_n^\dagger, & (4.7) \end{aligned}$$

and in the limit $\delta \to 0^+$ this becomes

$$A(E) = 2\pi\sum_{n=1}^{N_{\mathrm{u}}} \delta(E-E_n)\ \psi_n\psi_n^\dagger. \tag{4.8}$$

With this result, and by using the normalization of the ψ_n, the density of states $\mathcal{N} = \sum_{n=1}^{N_{\mathrm{u}}} \delta(E-E_n)$ can be written as

$$\mathcal{N}(E) = \frac{1}{2\pi}\mathrm{Tr}[A(E)S]. \tag{4.9}$$

As shown in Sec. 2.2.4 the total DOS can be split up in the contributions from the single orbitals, the so called projected density of states (PDOS), in the following way

$$\mathcal{N}(E) = \sum_{\mu}^{N_{\mathrm{u}}} \mathcal{N}_\mu(E), \tag{4.10}$$

with the PDOS for orbital μ, $\mathcal{N}_\mu$, defined as

$$\mathcal{N}_\mu(E) = \frac{1}{2\pi}\sum_{\nu}^{N_{\mathrm{u}}} A_{\mu\nu}(E)S_{\nu\mu}. \tag{4.11}$$

Eqs. (4.8) and (4.9) show that the spectral function, and therefore also the GF, contains all the information about the states of the system. We also note that $A(E)$ is a positive semidefinite matrix, since all its eigenvalues are greater than or equal to zero [see Eq. (4.8)]. The density matrix [Eq. (2.70)]

$$\rho = \sum_{n=1}^{N_{\mathrm{u}}} n_n\psi_n\psi_n^\dagger. \tag{4.12}$$

can be equally expressed as function of $A(E)$. If it is assumed that the system is in thermal equilibrium with the environment, for a system of Fermions one obtains

$n_n = f(E_n)$, where $f(E_n)$ is the Fermi-Dirac distribution calculated at the energy E_n, given in Eq. (2.68). At finite temperature the Fermi energy in the Fermi-Dirac distribution is replaced by the chemical potential [70], however we assume that $T \approx 0$, so that the two quantities are identical. The use of a finite temperature in our calculations has only the role of smearing out the region of occupied states in order to improve convergence for metals. Moreover in the calculation of the current (Sec. 4.4) it makes the results less sensitive to noise.

Using Eq. (4.8) the density matrix can then be written as

$$\rho = \frac{1}{2\pi} \int dE \; f(E) \; A(E), \tag{4.13}$$

where the energy integral is over the entire real energy axis (from $-\infty$ to ∞). The Fermi energy is defined implicitly by the number of particles in the system N_e via Eq. (2.72). At equilibrium the knowledge of the spectral function therefore uniquely determines ρ and consequently all the ground state properties of the system. We note that due to the assumption of a thermal equilibrium with a reference system ρ is implicitly a time averaged quantity. For a system out of equilibrium the same argument can be applied, the only difference being that the wave functions ψ_n are split up into separate sets, in local equilibrium with only one of the reservoirs, with different local Fermi energy. We note that out of equilibrium the functional relation between a given ρ and its Hamiltonian is not known, since the KS Hamiltonian is constructed assuming that the system is in the ground state. However we assume that the DFT Hamiltonian [Eq. (2.56)] is valid also out of equilibrium.

4.2 Green's function for the scattering region

The first step towards a non equilibrium transport theory is to split up the system into separate subsystems. The scattering region, also called the "extended molecule" (EM), is the central part and interacts with all the other subsystems, each one corresponding to a separate lead. It is assumed that the leads do not interact directly with each other, but they are connected through the EM. The idea then is that each lead is in local thermal equilibrium and has a corresponding chemical potential. The leads are further split up into a set of left and right leads, where the difference is that the matrix elements representing orbitals of the left leads have indices that are all smaller than the ones of the EM (i.e. they are on the "left" side of the matrix), whereas for the right leads they are all larger (i.e. they are on the "right" side of the matrix). This subdivision is not necessary, but it somewhat helps in splitting up a real physical system. To keep the notation simple, in the following we restrict our attention to a

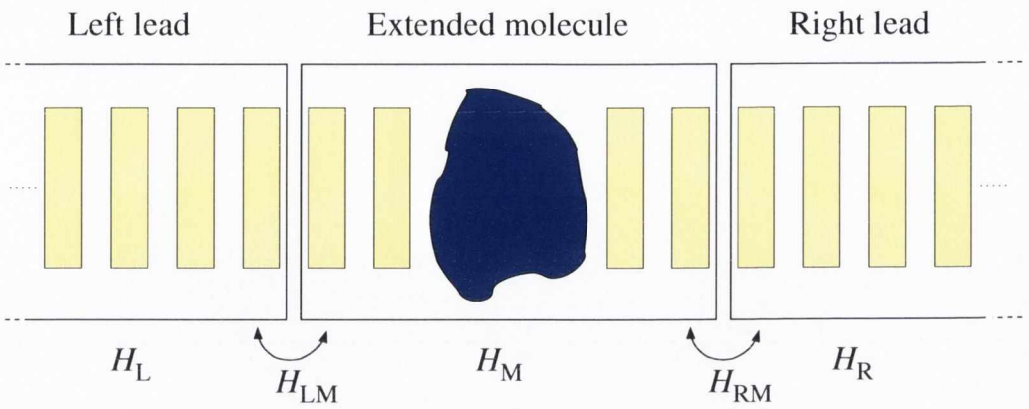


Figure 4.1: Schematic representation of the two-terminal device comprising the left lead, the extended molecule, and the right lead.

system with just one left and one right lead. The fundamental equations are however also introduced for an arbitrary number of leads. A schematic representation of such a two-terminal system is shown in Fig. 4.1. This is also the system used in most of the calculations in this work.

The Hamiltonian H of the two-terminal device has the structure

$$H = \begin{pmatrix} H_{\mathrm{L}} & H_{\mathrm{LM}} & 0 \\ H_{\mathrm{ML}} & H_{\mathrm{M}} & H_{\mathrm{MR}} \\ 0 & H_{\mathrm{RM}} & H_{\mathrm{R}} \end{pmatrix}, \tag{4.14}$$

where H_{L} and H_{R} are the Hamiltonians of the isolated left and right leads respectively, H_{M} is the Hamiltonian of the isolated EM. H_{LM} is the hopping Hamiltonian between the left lead and the EM, and H_{MR} is the same for the right lead (see Fig. 4.1). Since H is Hermitian we have $H_{\mathrm{ML}} = H_{\mathrm{LM}}^{\dagger}$ and $H_{\mathrm{RM}} = H_{\mathrm{MR}}^{\dagger}$. The zero matrices represent the fact that there is no direct interaction between the left and the right lead. The overlap matrix S has the same structure as H

$$S = \begin{pmatrix} S_{\mathrm{L}} & S_{\mathrm{LM}} & 0 \\ S_{\mathrm{ML}} & S_{\mathrm{M}} & S_{\mathrm{MR}} \\ 0 & S_{\mathrm{RM}} & S_{\mathrm{R}} \end{pmatrix}. \tag{4.15}$$

In order to simplify the notation we define the K matrix as $K = H - (E + i\delta)S$, which in block form reads

$$K = \begin{pmatrix} K_{\mathrm{L}} & K_{\mathrm{LM}} & 0 \\ K_{\mathrm{ML}} & K_{\mathrm{M}} & K_{\mathrm{MR}} \\ 0 & K_{\mathrm{RM}} & K_{\mathrm{R}} \end{pmatrix}. \tag{4.16}$$

We note that $K \neq K^{\dagger}$ for all $\delta \neq 0$, however as we take the limit $\delta \to 0^{+}$ the matrix becomes Hermitian, so that $K = K^{\dagger}$ for $\delta = 0$.

By using the general result for the inverse of such a block matrix given in Appendix B [Eqs. (B.2) and (B.3)] the retarded Green's function, $G = [(E + i\delta)S - H]^{-1} = -K^{-1}$, for the whole system can be written as

$$G = \begin{pmatrix} G_{\mathrm{L}} + T_{\mathrm{LM}} G_{\mathrm{M}} T_{\mathrm{ML}} & T_{\mathrm{LM}} G_{\mathrm{M}} & T_{\mathrm{LM}} G_{\mathrm{M}} T_{\mathrm{MR}} \\ G_{\mathrm{M}} T_{\mathrm{ML}} & G_{\mathrm{M}} & G_{\mathrm{M}} T_{\mathrm{MR}} \\ T_{\mathrm{RM}} G_{\mathrm{M}} T_{\mathrm{ML}} & T_{\mathrm{RM}} G_{\mathrm{M}} & G_{\mathrm{R}} + T_{\mathrm{RM}} G_{\mathrm{M}} T_{\mathrm{MR}} \end{pmatrix}. \tag{4.17}$$

Here

$$G_{\mathrm{L}} = -K_{\mathrm{L}}^{-1}, \tag{4.18}$$

$$G_{\mathrm{R}} = -K_{\mathrm{R}}^{-1} \tag{4.19}$$

are the GFs of the isolated left and right lead respectively, and

$$T_{\mathrm{LM}} = G_{\mathrm{L}}\, K_{\mathrm{LM}}, \tag{4.20}$$

$$T_{\mathrm{ML}} = K_{\mathrm{ML}}\, G_{\mathrm{L}}, \tag{4.21}$$

$$T_{\mathrm{RM}} = G_{\mathrm{R}}\, K_{\mathrm{RM}}, \tag{4.22}$$

$$T_{\mathrm{MR}} = K_{\mathrm{MR}}\, G_{\mathrm{R}}. \tag{4.23}$$

These T-Matrices are generally called transfer matrices (see also Sec. 5.1.2). The GF of the extended molecule G_{M} is given by

$$G_{\mathrm{M}} = \left(-K_{\mathrm{M}} - \Sigma_{\mathrm{L}} - \Sigma_{\mathrm{R}}\right)^{-1}. \tag{4.24}$$

Here the left self-energy Σ_{L} and the right self-energy Σ_{R} have been introduced [54]. These are defined as

$$\Sigma_{\mathrm{L}} = K_{\mathrm{ML}}\, G_{\mathrm{L}}\, K_{\mathrm{LM}}, \tag{4.25}$$

$$\Sigma_{\mathrm{R}} = K_{\mathrm{MR}}\, G_{\mathrm{R}}\, K_{\mathrm{RM}}. \tag{4.26}$$

They are non-Hermitian matrices and determine how the states in the EM are modified by the interaction with the leads. The Hermitian part causes a shift of the EM's eigenvalue spectrum, and the antihermitian part leads to a broadening of the energy levels in the EM [54]. This will be discussed in more detail below and in chapter 6. The algorithm for constructing the self-energies for semi-infinite, periodic leads is presented in chapter 5. For an arbitrary number of leads, N_{leads}, Eq. (4.24) for G_{M} becomes

$$G_{\mathrm{M}} = \left(-K_{\mathrm{M}} - \sum_{n=1}^{N_{\mathrm{leads}}} \Sigma_n\right)^{-1}, \tag{4.27}$$

where Σ_n is the self-energy for the n-th lead. If we define the effective Hamiltonian of the EM as

$$H_{\mathrm{M,eff}} = H_{\mathrm{M}} + \sum_{n=1}^{N_{\mathrm{leads}}} \Sigma_n, \tag{4.28}$$

then the G_{M} can be written as

$$G_{\mathrm{M}} = \left[(E + i\delta)S_{\mathrm{M}} - H_{\mathrm{M,eff}}\right]^{-1}. \tag{4.29}$$

The GF for the EM can therefore be viewed like the one for an isolated system, with the only difference that H_{M} has to be replaced by the effective Hamiltonian $H_{\mathrm{M,eff}}$. This takes into account the effects of the leads onto the EM.

For a two-terminal device the full GF G, and with it the properties of the whole system, can therefore be determined by calculating G_{L}, G_{R} and G_{M}. Once G is known the density matrix of the whole system can be calculated via the spectral function [Eq. (4.13)]. Therefore the GF formalism is a way of solving the Kohn-Sham problem without calculating the eigenvalues of H. Just as in a standard DFT calculation however the density matrix and the Hamiltonian H have to be calculated self-consistently. We now assume that all the parts of H, except H_{M}, are independent of the charge density in the scattering region ρ_{M}. This assumption is valid if the EM is chosen large enough, so that changes to the charge density in the center of it do not affect the Hamiltonian at its boundaries. Equivalently the requirement is that all charge fluctuations in the center of the EM are screened once its boundaries are reached. For metallic systems with a short screening-length this requirement is fulfilled in a few atomic layers, for semiconducting systems however the screening-length is very large, so that also the size of the EM must be increased accordingly [55]. In all our calculations we therefore use metals at the boundary of the EM in order to keep the system size small. The main advantage resulting from this assumption is that the self-energies need to be calculated only once at the beginning of the calculation. The matrices ρ_{M} and H_{M} need to be calculated self-consistently for the given set of self-energies. The GF scheme therefore allows us to calculate the properties of a system for a given set of boundary conditions, set by the self-energies.

In order to obtain ρ_{M} for the ground state [see Eq. (4.13)], we first need to evaluate the spectral function for the EM

$$A_{\mathrm{M}} = i(G_{\mathrm{M}} - G_{\mathrm{M}}^{\dagger}). \tag{4.30}$$

A_{M} can also be rewritten in terms of self-energies as

$$\begin{aligned} A_{\mathrm{M}} &= i\, G_{\mathrm{M}}(G_{\mathrm{M}}^{\dagger -1} - G_{\mathrm{M}}^{-1})G_{\mathrm{M}}^{\dagger} \\ &= i\, G_{\mathrm{M}}[-K_{\mathrm{M}}^{\dagger} - \Sigma_{\mathrm{L}}^{\dagger} - \Sigma_{\mathrm{R}}^{\dagger} + K_{\mathrm{M}} + \Sigma_{\mathrm{L}} + \Sigma_{\mathrm{R}}]G_{\mathrm{M}}^{\dagger} \\ &= i\, G_{\mathrm{M}}(\Sigma_{\mathrm{L}} - \Sigma_{\mathrm{L}}^{\dagger} + \Sigma_{\mathrm{R}} - \Sigma_{\mathrm{R}}^{\dagger})G_{\mathrm{M}}^{\dagger} + 2\delta\, G_{\mathrm{M}} S_{\mathrm{M}} G_{\mathrm{M}}^{\dagger}, \end{aligned} \tag{4.31}$$

where we have used Eq. (4.24) and the fact that for real energies E, and Hermitian matrices H_{M} and S_{M}, we have $K_{\mathrm{M}} - K_{\mathrm{M}}^{\dagger} = -2i\delta S_{\mathrm{M}}$. We now introduce the left and right Γ-matrices [54]

$$\Gamma_{\mathrm{L}} = i\left(\Sigma_{\mathrm{L}} - \Sigma_{\mathrm{L}}^{\dagger}\right), \tag{4.32}$$

$$\Gamma_{\mathrm{R}} = i\left(\Sigma_{\mathrm{R}} - \Sigma_{\mathrm{R}}^{\dagger}\right), \tag{4.33}$$

which are proportional to the antihermitian part of the self-energies. Γ_{L} and Γ_{R} are Hermitian, positive semidefinite matrices. With these definitions in hands A_{M} can be written as

$$A_{\mathrm{M}} = G_{\mathrm{M}}(\Gamma_{\mathrm{L}} + \Gamma_{\mathrm{R}})G_{\mathrm{M}}^{\dagger} + 2\delta\, G_{\mathrm{M}} S_{\mathrm{M}} G_{\mathrm{M}}^{\dagger}. \tag{4.34}$$

Following an analogous procedure it can be shown that the positions of G_{M} and $G_{\mathrm{M}}^{\dagger}$ in Eq. (4.34) can also be interchanged

$$A_{\mathrm{M}} = G_{\mathrm{M}}^{\dagger}(\Gamma_{\mathrm{L}} + \Gamma_{\mathrm{R}})G_{\mathrm{M}} + 2\delta\, G_{\mathrm{M}}^{\dagger} S_{\mathrm{M}} G_{\mathrm{M}}. \tag{4.35}$$

For an arbitrary number of leads Eq. (4.34) can be generalized to

$$A_{\mathrm{M}} = G_{\mathrm{M}} \left(\sum_{n=1}^{N_{\mathrm{leads}}} \Gamma_n \right) G_{\mathrm{M}}^{\dagger} + 2\delta\, G_{\mathrm{M}} S_{\mathrm{M}} G_{\mathrm{M}}^{\dagger}. \tag{4.36}$$

A_{M} is therefore proportional to the sum of the Γ-matrices of all the leads. In Secs. 4.3 and 4.4 it will be shown that these describe the rate of in- and out-flow of electrons from/to the leads, and can therefore be regarded as "coupling" matrices to the leads.

The next assumption is that the leads are very large, so that they can be approximated as semi-infinite objects, i.e. each lead has one end joining the scattering region, whereas the other end runs to infinity. In this way the density of states of the leads forms a continuum. As a consequence of the coupling to the leads, for most systems also the spectrum of the EM becomes continuous, which implies that there are no poles in G_{M}. This means that the limit $\delta = 0$ can be taken in the matrices K_{M}, K_{LM} and K_{RM}. Moreover, the second term in Eq. (4.34) vanishes, since this only contributes when there are poles in G_{M} (see chapter 6). Only if there are localized states (also denoted as bound states) in the EM that are not coupled to the continuous spectrum of the leads, then there will be poles in G_{M}. In that case the limit

$\delta = 0$ can not be taken in K_{M}, and the second term in Eq. (4.34) will not vanish. This will be discussed in detail in chapter 6. In the remaining part of this chapter we assume that there are no bound states in the EM.

Another consequence of the infinite size of the leads is that the Fermi energy is now set by the leads, which also act as electron reservoirs. Whereas in a normal DFT calculation the number of electrons in the system is fixed and determines the Fermi energy via Eq. (2.72) [Eq. (2.83) for periodic systems], now the opposite is true: the Fermi energy of the leads determines the number of electrons in the EM.

4.3 Out of equilibrium occupation

The GF formalism described up to this point is just an alternative way of solving the ground state KS problem for a given set of boundary conditions, set by the leads self-energies and by their common Fermi energy. If the EM joining the leads is removed, then each lead is independent and has its own Fermi energy. If the EM is then added back again the system is out of equilibrium, and this will cause a flow of electrons trying to equilibrate the system to one common Fermi energy. However due to the infinite size of the leads it will take an infinite time to reach equilibrium. The system will rather reach a non equilibrium steady state with a constant flow of electrons through the EM. In this section we derive the NEGF equations for the charge distribution in the EM at steady state. The basic idea is that the states in the leads are occupied up to a local Fermi level, and the occupation of the states in the EM is then obtained by the difference between the inflowing and the outflowing electrons. First the spectral function is rewritten in a form that maps each of the states of the system as originating from one particular lead. The occupation the states is then determined by the chemical potential of the lead they originate from. We note that, in contrast to to ground state DFT, in this approach there is no energy minimization principle that gives us the correct occupation, i.e. the variational principle does not hold.

4.3.1 Wave functions

The spectral function contains information about all the wave functions in the system at a given energy E. Here we split up these wave functions into those originating from each of the leads. Since the leads are semi-infinite, they have a continuous spectrum, so that for each energy E there is a finite number $N_{\mathrm{L}}(E)$ of states $\varphi_{\mathrm{L},n}$ in the left

lead with

$$H_{\mathrm{L}}\,\varphi_{\mathrm{L},n} = E\,S_{\mathrm{L}}\,\varphi_{\mathrm{L},n} \qquad n \in \{1,\ldots,N_{\mathrm{L}}(E)\}. \tag{4.37}$$

In the same way the right lead has $N_{\mathrm{R}}(E)$ states $\varphi_{\mathrm{R},n}$ that satisfy

$$H_{\mathrm{R}}\,\varphi_{\mathrm{R},n} = E\,S_{\mathrm{R}}\,\varphi_{\mathrm{R},n} \qquad n \in \{1,\ldots,N_{\mathrm{R}}(E)\}. \tag{4.38}$$

We note that $\varphi_{\mathrm{L},n}$ and $\varphi_{\mathrm{R},n}$ are functions of E. When the isolated leads join the EM, the wave functions will however be different due to their interaction with the EM and the other leads (indirectly). The wave function for the entire system ψ satisfies the KS Schrödinger-like equation

$$(H - ES)\,\psi = 0. \tag{4.39}$$

As a matter of notation we subdivide ψ into a three components vector

$$\psi = \begin{pmatrix} \psi^{\mathrm{L}} \\ \psi^{\mathrm{M}} \\ \psi^{\mathrm{R}} \end{pmatrix}, \tag{4.40}$$

where ψ^{L} (ψ^{R}) is the part of the WF extending over the left (right) lead, and ψ^{M} is the part extending over the EM. The wave functions are now split up into the ones originating from the left lead ψ_{L}, and the one originating from the right lead ψ_{R}. We write ψ_{L} as [194, 193]

$$\psi_{\mathrm{L}} = \begin{pmatrix} \varphi_{\mathrm{L}} \\ 0 \\ 0 \end{pmatrix} + \psi_{\mathrm{L}}^{\Delta}, \tag{4.41}$$

where $\psi_{\mathrm{L}}^{\Delta}$ has the dimension of the full infinite system (EM plus leads), and it corresponds to the change in WF due to the presence of the EM. In an analogous way ψ_{R} is written as

$$\psi_{\mathrm{R}} = \begin{pmatrix} 0 \\ 0 \\ \varphi_{\mathrm{R}} \end{pmatrix} + \psi_{\mathrm{R}}^{\Delta}. \tag{4.42}$$

This is just a way to rewrite the solutions, however the use of the retarded GF will allow us to associate the WFs as originating from a given lead.

The $\psi_{\mathrm{L},n}^{\Delta}$ then satisfy the equation

$$(H - ES)\,\psi_{\mathrm{L},n}^{\Delta} = \begin{pmatrix} 0 \\ -K_{\mathrm{ML}}\,\varphi_{\mathrm{L},n} \\ 0 \end{pmatrix}, \tag{4.43}$$

where we have used Eq. (4.37) and the definition of K. There are two sets of solutions to this equation [194]. The first set is obtained by multiplying Eq. (4.43) with the

retarded GF from the left

$$\psi^{\Delta+}_{\mathrm{L},n} = G \begin{pmatrix} 0 \\ K_{\mathrm{ML}}\ \varphi_{\mathrm{L},n} \\ 0 \end{pmatrix}, \tag{4.44}$$

and the second set is obtained by multiplying it with the advanced GF

$$\psi^{\Delta-}_{\mathrm{L},n} = G^{\dagger} \begin{pmatrix} 0 \\ K_{\mathrm{ML}}\ \varphi_{\mathrm{L},n} \\ 0 \end{pmatrix}. \tag{4.45}$$

The physical difference between the two solutions is that the first describes waves that flow from the left lead into the EM, and the second describes waves that flow from the EM into the left lead [194]. This will be proved in Sec. 4.4, where we will show that the probability current associated to the $\{\psi^{\Delta+}_{\mathrm{L},n}\}$ is positive (i.e. electrons flow from the left lead into the EM), while the one associated to the $\{\psi^{\Delta-}_{\mathrm{L},n}\}$ is negative. If there is no right lead attached to the scattering region these solutions are identical. Since the goal is to distinguish the solutions originating from a given lead, we use the first set of solutions, so that we set $\psi^{\Delta}_{\mathrm{L},n} = \psi^{\Delta+}_{\mathrm{L},n}$ [194]. By using the solution for G of Eq. (4.17), we then obtain for the WF of the entire system

$$\psi_{\mathrm{L},n} = \begin{pmatrix} \mathbb{1}_{\mathrm{L}} + T_{\mathrm{LM}}\ G_{\mathrm{M}}\ K_{\mathrm{ML}} \\ G_{\mathrm{M}}\ K_{\mathrm{ML}} \\ T_{\mathrm{RM}}\ G_{\mathrm{M}}\ K_{\mathrm{ML}} \end{pmatrix} \varphi_{\mathrm{L},n}, \tag{4.46}$$

where $\mathbb{1}_{\mathrm{L}}$ is a unity matrix with the dimensions of the left lead (it is therefore infinite dimensional). In the same way we obtain the wave functions originating in the right lead

$$\psi_{\mathrm{R},n} = \begin{pmatrix} T_{\mathrm{LM}} G_{\mathrm{M}}\ K_{\mathrm{MR}} \\ G_{\mathrm{M}}\ K_{\mathrm{MR}} \\ \mathbb{1}_{\mathrm{R}} + T_{\mathrm{RM}}\ G_{\mathrm{M}}\ K_{\mathrm{MR}} \end{pmatrix} \varphi_{\mathrm{R},n}, \tag{4.47}$$

with $\mathbb{1}_{\mathrm{R}}$ being the unity matrix with the dimension of the right lead.

4.3.2 Spectral function

Since the spectrum of the leads is continuous their spectral function [Eq. (4.8)] can be written as

$$A^{0}_{\mathrm{L}} = i\left(G_{\mathrm{L}} - G^{\dagger}_{\mathrm{L}}\right) = 2\pi \sum_{n=1}^{N_{\mathrm{L}}(E)} \mathcal{N}_{\mathrm{L},n}(E)\ \varphi_{\mathrm{L},n}\varphi^{\dagger}_{\mathrm{L},n}, \tag{4.48}$$

$$A^{0}_{\mathrm{R}} = i\left(G_{\mathrm{R}} - G^{\dagger}_{\mathrm{R}}\right) = 2\pi \sum_{n=1}^{N_{\mathrm{R}}(E)} \mathcal{N}_{\mathrm{R},n}(E)\ \varphi_{\mathrm{R},n}\varphi^{\dagger}_{\mathrm{R},n}. \tag{4.49}$$

Here $\mathcal{N}_{\mathrm{L},n}(E)$ and $\mathcal{N}_{\mathrm{R},n}(E)$ are the density of states for the individual states $\varphi_{\{\mathrm{L/R}\},n}$ of the isolated left and right leads respectively. We now verify that the combined sets $\{\psi_{\mathrm{L},n}\}$ and $\{\psi_{\mathrm{R},n}\}$ contain all the possible solutions for a given energy E by using them to explicitly calculate the spectral function. If the two sets contain all the possible solutions, the SF can be written as

$$A = A_{\mathrm{L}} + A_{\mathrm{R}}, \tag{4.50}$$

where

$$A_{\mathrm{L}} = 2\pi \sum_{n=1}^{N_{\mathrm{L}}(E)} \mathcal{N}_{\mathrm{L},n}(E)\ \psi_{\mathrm{L},n}\psi_{\mathrm{L},n}^{\dagger}, \tag{4.51}$$

$$A_{\mathrm{R}} = 2\pi \sum_{n=1}^{N_{\mathrm{R}}(E)} \mathcal{N}_{\mathrm{L},n}(E)\ \psi_{\mathrm{R},n}\psi_{\mathrm{R},n}^{\dagger}. \tag{4.52}$$

Using Eq. (4.46) A_{L} reads

$$\begin{aligned} A_{\mathrm{L}} &= \begin{pmatrix} \mathbb{1}_{\mathrm{L}} + T_{\mathrm{LM}}G_{\mathrm{M}}K_{\mathrm{ML}} \\ G_{\mathrm{M}}K_{\mathrm{ML}} \\ T_{\mathrm{RM}}G_{\mathrm{M}}K_{\mathrm{ML}} \end{pmatrix} \left(2\pi \sum_{n=1}^{N_{\mathrm{L}}(E)} \mathcal{N}_{\mathrm{L},n}(E)\varphi_{\mathrm{L},n}\varphi_{\mathrm{L},n}^{\dagger} \right) \times \\ &\quad \left(\mathbb{1}_{\mathrm{L}} + K_{\mathrm{ML}}^{\dagger}G_{\mathrm{M}}^{\dagger}T_{\mathrm{LM}}^{\dagger} \ ,\ K_{\mathrm{ML}}^{\dagger}G_{\mathrm{M}}^{\dagger} \ ,\ K_{\mathrm{ML}}^{\dagger}G_{\mathrm{M}}^{\dagger}T_{\mathrm{RM}} \right) \\ &= i \begin{pmatrix} \mathbb{1}_{\mathrm{L}} + T_{\mathrm{LM}}G_{\mathrm{M}}K_{\mathrm{ML}} \\ G_{\mathrm{M}}K_{\mathrm{ML}} \\ T_{\mathrm{RM}}G_{\mathrm{M}}K_{\mathrm{ML}} \end{pmatrix} (G_{\mathrm{L}} - G_{\mathrm{L}}^{\dagger}) \times \\ &\quad \left(\mathbb{1}_{\mathrm{L}} + K_{\mathrm{ML}}^{\dagger}G_{\mathrm{M}}^{\dagger}T_{\mathrm{LM}}^{\dagger} \ ,\ K_{\mathrm{ML}}^{\dagger}G_{\mathrm{M}}^{\dagger} \ ,\ K_{\mathrm{ML}}^{\dagger}G_{\mathrm{M}}^{\dagger}T_{\mathrm{RM}} \right) \end{aligned} \tag{4.53}$$

In the same way we get for A_{R}

$$\begin{aligned} A_{\mathrm{R}} &= i \begin{pmatrix} T_{\mathrm{LM}}G_{\mathrm{M}}K_{\mathrm{ML}} \\ G_{\mathrm{M}}K_{\mathrm{ML}} \\ \mathbb{1}_{\mathrm{R}} + T_{\mathrm{RM}}G_{\mathrm{M}}K_{\mathrm{ML}} \end{pmatrix} (G_{\mathrm{R}} - G_{\mathrm{R}})^{\dagger} \times \\ &\quad \left(K_{\mathrm{ML}}^{\dagger}G_{\mathrm{M}}^{\dagger}T_{\mathrm{LM}}^{\dagger} \ ,\ K_{\mathrm{ML}}^{\dagger}G_{\mathrm{M}}^{\dagger} \ ,\ \mathbb{1}_{\mathrm{R}} + K_{\mathrm{ML}}^{\dagger}G_{\mathrm{M}}^{\dagger}T_{\mathrm{RM}} \right). \end{aligned} \tag{4.54}$$

It is now straight forward to verify that $A_{\mathrm{L}} + A_{\mathrm{R}} = i(G - G^{\dagger})$, with G given in Eq. (4.17). This confirms that $\{\psi_{\mathrm{L},n}\}$ and $\{\psi_{\mathrm{R},n}\}$ contain, for a given energy E, all the states of the entire system. We note again that this result is only valid if there are no localized states in the EM, so that the second term proportional to δ in Eq. (4.34) is zero.

With Eq. (4.53) and Eq. (4.54) the SF of the EM can be split up as

$$A_{\mathrm{M}} = A_{\mathrm{ML}} + A_{\mathrm{MR}}, \tag{4.55}$$

where

$$A_{\mathrm{ML}} = G_{\mathrm{M}}\Gamma_{\mathrm{L}}G_{\mathrm{M}}^{\dagger} \tag{4.56}$$

is the part of the SF generated by states originating from the left lead. These states describe a flow of particles from the left lead into the EM and further into the right lead. In the same way

$$A_{\mathrm{MR}} = G_{\mathrm{M}}\Gamma_{\mathrm{R}}G_{\mathrm{M}}^{\dagger} \tag{4.57}$$

is the part of the SF generated by states originating from the right lead.

4.3.3 Density matrix

In the same way as the SF also the density matrix of the EM can be split up into the components originating from the single leads

$$\rho_{\mathrm{M}} = \rho_{\mathrm{ML}} + \rho_{\mathrm{MR}}. \tag{4.58}$$

The individual parts are

$$\rho_{\mathrm{ML}} = \int dE \sum_{n=1}^{N_{\mathrm{L}}(E)} p_{\mathrm{L},n}(E)\, \mathcal{N}_{\mathrm{L},n}(E)\, \psi_{\mathrm{L},n}^{\mathrm{M}}\psi_{\mathrm{L},n}^{\mathrm{M}\dagger}, \tag{4.59}$$

$$\rho_{\mathrm{MR}} = \int dE \sum_{n=1}^{N_{\mathrm{R}}(E)} p_{\mathrm{R},n}(E)\, \mathcal{N}_{\mathrm{R},n}(E)\, \psi_{\mathrm{R},n}^{\mathrm{M}}\psi_{\mathrm{R},n}^{\mathrm{M}\dagger}, \tag{4.60}$$

with $\psi_{\{\mathrm{L/R}\},n}^{\mathrm{M}}$ being the part of the WF extending over the EM as defined in Eq. (4.40). The occupation number $p_{\{\mathrm{L/R}\},n}(E)$ lies between 0 and 1 and determines the occupation of each state. We now assume that the left and the right leads are in local thermal equilibrium, and therefore have a local Fermi energy $E_{\mathrm{F},\{\mathrm{L/R}\}}$. Then each of the states $\psi_{\{\mathrm{L/R}\}}(E)$ has a probability of being occupied given by the Fermi distribution $f_{\{\mathrm{L/R}\}}(E)$ of the lead it originates from. Here $f_{\mathrm{L}}(E)$ [$f_{\mathrm{R}}(E)$] is the Fermi-Dirac distribution with the Fermi energy equal to $E_{\mathrm{F,L}}$ ($E_{\mathrm{F,R}}$). Since the $\{\psi_{\mathrm{L},n}\}$ describe states originating from the left lead, we have $p_{\mathrm{L},n}(E) = f_{\mathrm{L}}(E)$, and analogously we have $p_{\mathrm{R},n}(E) = f_{\mathrm{R}}(E)$. This is one of the central approximations in the NEGF formalism [54]. It also allows the introduction of a bias voltage V_{b}, which is defined as the difference of the Fermi energies of the two leads divided by the electron charge e, so that $eV_{\mathrm{b}} = E_{\mathrm{F,L}} - E_{\mathrm{F,R}}$. Since the energy eigenvalue spectrum of the entire system is only defined up to a constant, we can set the energy scale of the system by using the Fermi level of the system at equilibrium, E_{F}, as reference and setting $E_{\mathrm{F,L}} = E_{\mathrm{F}} + eV_{\mathrm{b}}/2$ and $E_{\mathrm{F,R}} = E_{\mathrm{F}} - eV_{\mathrm{b}}/2$. This is the convention used throughout

this work. We can therefore write

$$\rho_{\mathrm{ML}} = \int dE\ f_{\mathrm{L}}(E) \left(\sum_{n=1}^{N_{\mathrm{L}}(E)} \mathcal{N}_{\mathrm{L},n}(E)\ \psi^{\mathrm{M}}_{\mathrm{L},n} \psi^{\mathrm{M}\dagger}_{\mathrm{L},n} \right)$$
$$= \frac{1}{2\pi} \int dE\ f_{\mathrm{L}}(E)\ A_{\mathrm{ML}}, \tag{4.61}$$

and in the same way

$$\rho_{\mathrm{MR}} = \frac{1}{2\pi} \int dE\ f_{\mathrm{R}}(E)\ A_{\mathrm{MR}}. \tag{4.62}$$

We now introduce the lesser GF for the EM, $G^{<}_{\mathrm{M}}$, [54, 183, 194], which for such a system is defined by

$$G^{<}_{\mathrm{M}} = i\ G_{\mathrm{M}} \left[f_{\mathrm{L}}(E)\Gamma_{\mathrm{L}} + f_{\mathrm{R}}(E)\Gamma_{\mathrm{R}} \right] G^{\dagger}_{\mathrm{M}}. \tag{4.63}$$

By using this definition, and Eqs. (4.58), (4.61) and (4.62), the density matrix of the EM becomes

$$\rho_{\mathrm{M}} = \frac{1}{2\pi i} \int dE\ G^{<}_{\mathrm{M}}(E). \tag{4.64}$$

This is the central equation of the NEGF formalism [54], and allows one to obtain the charge density of the EM attached to leads also out of equilibrium. If all the leads have the same Fermi energy, then we recover the equilibrium result of Eq. (4.13).

The same procedure can be repeated for an arbitrary number of leads, in which case the total lesser GF becomes

$$G^{<}_{\mathrm{M}} = i\ G_{\mathrm{M}} \left(\sum_{n}^{N_{\mathrm{leads}}} f_n(E)\Gamma_n \right) G^{\dagger}_{\mathrm{M}}. \tag{4.65}$$

4.3.4 Equilibrium and non-equilibrium charge density

Eq. (4.65) for the lesser GF can be rewritten in a form that is more suited for numerical computation, as described in Refs. [61, 62, 63]. We first choose one of the leads to be the reference lead, in this case lead with index 1, and add and remove the same term to Eq. (4.65)

$$G^{<}_{\mathrm{M}} = i\ G_{\mathrm{M}} \left(\sum_{n=1}^{N_{\mathrm{leads}}} f_n(E)\Gamma_n + \sum_{n=2}^{N_{\mathrm{leads}}} f_1(E)\Gamma_n - \sum_{n=2}^{N_{\mathrm{leads}}} f_1(E)\Gamma_n \right) G^{\dagger}_{\mathrm{M}}. \tag{4.66}$$

We can rearrange the sums to

$$G^{<}_{\mathrm{M}} = i\ f_1(E) G_{\mathrm{M}} \left(\sum_{n=1}^{N_{\mathrm{leads}}} \Gamma_n \right) G^{\dagger}_{\mathrm{M}} + \sum_{n=2}^{N_{\mathrm{leads}}} i \left[f_n(E) - f_1(E) \right] G_{\mathrm{M}} \Gamma_n G^{\dagger}_{\mathrm{M}}. \tag{4.67}$$

The first term is proportional to the spectral function [Eq. (4.36)] times the Fermi-Dirac distribution, and corresponds to the lesser GF for a system at equilibrium with the Fermi energy equal to the one of the first lead $E_{\mathrm{F},1}$. This is therefore called the equilibrium part $G^{<}_{\mathrm{M,eq},1}$ of the lesser GF

$$G^{<}_{\mathrm{M,eq},1} = -f_1(E)\left(G_{\mathrm{M}} - G^{\dagger}_{\mathrm{M}}\right). \tag{4.68}$$

The index 1 reflects the fact that lead 1 is the reference lead. The remaining part describes the change to the lesser GF caused by the system being out of equilibrium, and is therefore called the non-equilibrium part of the lesser GF $G^{<}_{\mathrm{M,neq},1}$

$$G^{<}_{\mathrm{M,neq},1} = \sum_{n=2}^{N_{\mathrm{leads}}} i\left[f_n(E) - f_1(E)\right] G_{\mathrm{M}} \Gamma_n G^{\dagger}_{\mathrm{M}}. \tag{4.69}$$

The total $G^{<}_{\mathrm{M}}$ is equal the sum of the two parts

$$G^{<}_{\mathrm{M}} = G^{<}_{\mathrm{M,eq},1} + G^{<}_{\mathrm{M,neq},1}. \tag{4.70}$$

Assuming that lead 1 is the one with the lowest chemical potential, then the set of Eqs. (4.68) through (4.70) show that all the states in the system are occupied up to the chemical potential of lead 1. No state is occupied above the highest chemical potential, and for energies in between the occupation is given by the non-equilibrium charge density and depends on the strength of the coupling to each of the leads and its Fermi energy.

We could equally have used any of the other leads as reference lead. In order to avoid having to chose a reference lead, we can define the equilibrium lesser GF as the average

$$G^{<}_{\mathrm{M,eq}} = -\left(\frac{1}{N_{\mathrm{leads}}} \sum_{n=1}^{N_{\mathrm{leads}}} f_n(E)\right)\left(G_{\mathrm{M}} - G^{\dagger}_{\mathrm{M}}\right), \tag{4.71}$$

and the corresponding non-equilibrium lesser GF $G^{<}_{\mathrm{M,neq}}$ as

$$G^{<}_{\mathrm{M,neq}} = \frac{1}{N_{\mathrm{leads}}} \sum_{n=1}^{N_{\mathrm{leads}}} \sum_{m<n} i\left[f_n(E) - f_m(E)\right] G_{\mathrm{M}} \left(\Gamma_n - \Gamma_m\right) G^{\dagger}_{\mathrm{M}}. \tag{4.72}$$

so that $G^{<}_{\mathrm{M}} = G^{<}_{\mathrm{M,eq}} + G^{<}_{\mathrm{M,neq}}$.

The main advantage of splitting up $G^{<}_{\mathrm{M}}$ into $G^{<}_{\mathrm{M,eq}}$ and $G^{<}_{\mathrm{M,neq}}$ is that $G^{<}_{\mathrm{M,eq}}$ can be analytically extended to complex energies E, and that it has no poles for $\mathrm{Im}(E) > 0$ [195, 61, 62]. In evaluating the energy integral of Eq. (4.64) to obtain ρ it is therefore possible to replace the integral over the real energy axis by one over the positive complex plane, where the $G^{<}_{\mathrm{M,eq}}$ is much smoother and less integration points are

needed as a consequence (see also chapter 6). The integral over the non-equilibrium part $G^{<}_{\mathrm{M,neq}}$ has to be calculated on the real energy axis, however the integration range only extends over the bias window, where the Fermi-Dirac distributions of the leads are different.

4.4 Transmission and Current

4.4.1 Current of a wave function

In the same way as the density matrix also the current can be constructed as a sum over the contributions from the single WFs. In this section the current of a single WF is calculated, and in the next section the total current is then evaluated summing up the contributions from all WFs. Since the WFs ψ are normalized in such a way that $\psi^{\dagger} S \psi$, the electron charge in the EM, q_{M}, is

$$q_{\mathrm{M}} = \psi^{\mathrm{L}\dagger} S_{\mathrm{LM}} \psi^{\mathrm{M}} + \psi^{\mathrm{M}\dagger} S_{\mathrm{M}} \psi^{\mathrm{M}} + \psi^{\mathrm{R}\dagger} S_{\mathrm{RM}} \psi^{\mathrm{M}}, \tag{4.73}$$

where the $\psi^{\{\mathrm{L/R/M}\}}$ are defined in Eq. (4.40). If the overlap terms proportional to S_{LM} and S_{RM} are neglected this becomes

$$q_{\mathrm{M}} = \psi^{\mathrm{M}\dagger} S_{\mathrm{M}} \, \psi^{\mathrm{M}}. \tag{4.74}$$

In order to obtain the current we need to define the time-dependent WF. We now assume that the time dependence of a KS single particle eigenstate with energy E

$$H \, \psi = E \, S \, \psi \tag{4.75}$$

is determined by a time-dependent Schrödinger-like equation

$$H \, \psi_t = i\hbar S \frac{\partial \psi_t}{\partial t}, \tag{4.76}$$

where the index t in ψ_t indicates that this is the explicit time-dependent WF. The solution for the time-dependent WF therefore is $\psi_t = e^{-iEt/\hbar}\psi$. The time derivative of the occupation of the EM for such an eigenstate vanishes:

$$\frac{\partial q_{\mathrm{M}}}{\partial t} = \frac{\partial \psi_t^{\mathrm{M}\dagger} S_{\mathrm{M}} \psi_t^{\mathrm{M}}}{\partial t} = 0. \tag{4.77}$$

This can be interpreted as a sort of continuity equation [55]: the change in charge is zero, because the inflowing current from one lead is equal to the outflowing current

through the other lead. These two currents can be obtained by explicitly taking the time-derivative of q_{M}. The time derivative of ψ_t^{M} is

$$e^{iEt/\hbar}\, S_{\mathrm{M}} \frac{\partial \psi_t^{\mathrm{M}}}{\partial t} = -\frac{i}{\hbar} E\, S_{\mathrm{M}}\, \psi^{\mathrm{M}} \tag{4.78}$$

$$= -\frac{i}{\hbar}\left(K_{\mathrm{ML}}\, \psi^{\mathrm{L}} + H_{\mathrm{M}}\, \psi^{\mathrm{M}} + K_{\mathrm{MR}}\, \psi^{\mathrm{R}}\right), \tag{4.79}$$

where we have used Eq. (4.75) together with the definition of H and S [Eqs. (4.14) and (4.15)]. The time derivative of the occupation of the EM then is

$$\begin{aligned} \frac{\partial q_{\mathrm{M}}}{\partial t} &= \frac{\partial \psi_t^{\mathrm{M}\dagger}}{\partial t}\, S_{\mathrm{M}}\, \psi_t^{\mathrm{M}} + \psi_t^{\mathrm{M}\dagger}\, S_{\mathrm{M}}\, \frac{\partial \psi_t^{\mathrm{M}}}{\partial t} \\ &= \frac{i}{\hbar}\left(\psi^{\mathrm{L}\dagger}\, K_{\mathrm{LM}}\, \psi^{\mathrm{M}} - \psi^{\mathrm{M}\dagger}\, K_{\mathrm{ML}}\, \psi^{\mathrm{L}}\right) \\ &\quad + \frac{i}{\hbar}\left(\psi^{\mathrm{R}\dagger}\, K_{\mathrm{RM}}\, \psi^{\mathrm{M}} - \psi^{\mathrm{M}\dagger}\, K_{\mathrm{MR}}\, \psi^{\mathrm{R}}\right). \end{aligned} \tag{4.80}$$

If we interpret this as a continuity equation, as done in Ref. [55], then we can identify the change in charge as being equal to the sum of the total current flowing in from the left lead I^{L} and the total current flowing in from the right lead I^{R}:

$$\frac{\partial q_{\mathrm{M}}}{\partial t} = I^{\mathrm{L}} + I^{\mathrm{R}} = 0, \tag{4.81}$$

so that $I^{\mathrm{L}} = -I^{\mathrm{R}}$. The two currents of the single WFs are [55]

$$I^{\mathrm{L}} = \frac{i}{\hbar}\left(\psi^{\mathrm{L}\dagger}\, K_{\mathrm{LM}}\, \psi^{\mathrm{M}} - \psi^{\mathrm{M}\dagger}\, K_{\mathrm{ML}}\, \psi^{\mathrm{L}}\right), \tag{4.82}$$

$$I^{\mathrm{R}} = \frac{i}{\hbar}\left(\psi^{\mathrm{R}\dagger}\, K_{\mathrm{RM}}\, \psi^{\mathrm{M}} - \psi^{\mathrm{M}\dagger}\, K_{\mathrm{MR}}\, \psi^{\mathrm{R}}\right). \tag{4.83}$$

These equations can also be derived more formally starting from the current density operator [192]. For an arbitrary number of leads with integer index n this equation can be generalized to

$$I^n = \frac{i}{\hbar}\left(\psi^{n\dagger} K_{n\mathrm{M}}\, \psi^{\mathrm{M}} - \psi^{\mathrm{M}\dagger} K_{Mn}\, \psi^n\right), \tag{4.84}$$

where I^n is the probability current from lead n into the EM, ψ^n is the part of the WF extending over lead n, and $K_{\{n\mathrm{M}/\mathrm{M}n\}}$ are the corresponding hopping matrices. The continuity equation then becomes

$$\sum_n I^n = 0. \tag{4.85}$$

4.4.2 Transmission and Current within the NEGF

The total current from the left lead into the EM I^{L} is equal to the sum of all the contributions from the single WFs

$$\begin{aligned} I^{\mathrm{L}} &= \int dE \sum_{n=1}^{N_{\mathrm{L}}(E)} f_{\mathrm{L}}(E)\, \mathcal{N}_{\mathrm{L},n}(E)\, I^{\mathrm{L}}_{\mathrm{L},n} \\ &\quad + \int dE \sum_{n=1}^{N_{\mathrm{R}}(E)} f_{\mathrm{R}}(E)\, \mathcal{N}_{\mathrm{R},n}(E)\, I^{\mathrm{L}}_{\mathrm{R},n}, \end{aligned} \tag{4.86}$$

where, by using Eq. (4.82) the current $I^{\mathrm{L}}_{\mathrm{L},n}$ due to a single WF originating in the left lead $\psi_{\mathrm{L},n}$ is

$$I^{\mathrm{L}}_{\mathrm{L},n} = \frac{i}{\hbar}\left(\psi^{\mathrm{L}\,\dagger}_{\mathrm{L},n}\, K_{\mathrm{LM}}\, \psi^{\mathrm{M}}_{\mathrm{L},n} - \psi^{\mathrm{M}\,\dagger}_{\mathrm{L},n}\, K_{\mathrm{ML}}\, \psi^{\mathrm{L}}_{\mathrm{L},n}\right), \tag{4.87}$$

and the current $I^{\mathrm{L}}_{\mathrm{R},n}$ due to a WF originating in the right lead $\psi_{\mathrm{R},n}$ is

$$I^{\mathrm{L}}_{\mathrm{R},n} = \frac{i}{\hbar}\left(\psi^{\mathrm{L}\,\dagger}_{\mathrm{R},n}\, K_{\mathrm{LM}}\, \psi^{\mathrm{M}}_{\mathrm{R},n} - \psi^{\mathrm{M}\,\dagger}_{\mathrm{R},n}\, K_{\mathrm{ML}}\, \psi^{\mathrm{L}}_{\mathrm{R},n}\right). \tag{4.88}$$

By inserting the explicit expressions for $\psi_{\mathrm{L},n}$ and $\psi_{\mathrm{R},n}$ [see Eqs. (4.46) and (4.47)], and after some algebraic manipulations, we obtain

$$I^{\mathrm{L}}_{\mathrm{L},n} = \frac{1}{\hbar}\varphi^{\dagger}_{\mathrm{L},n} K_{\mathrm{LM}} G^{\dagger}_{\mathrm{M}} \Gamma_{\mathrm{R}}\, G_{\mathrm{M}} K_{\mathrm{ML}}\, \varphi_{\mathrm{L},n}, \tag{4.89}$$

$$I^{\mathrm{L}}_{\mathrm{R},n} = -\frac{1}{\hbar}\varphi^{\dagger}_{\mathrm{R},n} K_{\mathrm{RM}} G^{\dagger}_{\mathrm{M}} \Gamma_{\mathrm{L}}\, G_{\mathrm{M}} K_{\mathrm{MR}}\, \varphi_{\mathrm{R},n}. \tag{4.90}$$

Here we have used the definitions of G_{M} [Eq. (4.24)], Γ_{L} [Eq. (4.32)], Γ_{R} [Eq. (4.33)], and also the fact that $K_{\mathrm{M}} = K^{\dagger}_{\mathrm{M}}$ and $K_{\mathrm{M\{L/R\}}} = K^{\dagger}_{\mathrm{\{L/R\}M}}$. Since Γ_{R} is positive semidefinite, and using the fact that for any semidefinite matrix M also the matrix $UMU^{\dagger}$ is semidefinite for a arbitrary matrix U [196], it can be seen that $I^{\mathrm{L}}_{\mathrm{L},n} \geq 0$. Therefore it is further proved that the states $\{\psi_{\mathrm{L},n}\}$ describe electrons flowing from the left lead into the EM (and then into the right lead). This was one of the assumptions used in Sec. 4.3.1 to motivate why the retarded GF was chosen when calculating $\psi^{\Delta}_{\mathrm{L},n}$ [Eq. (4.44)], and has now been proved. Had we chosen $\psi^{\Delta+}_{\mathrm{L},n}$ [Eq. (4.45)], obtained by using the advanced GF, then the electron flow would be from the EM into the left lead, so that such states would not originate in the left lead. In the same way it can be shown that $I^{\mathrm{L}}_{\mathrm{R},n} \leq 0$ [due to the negative sign in Eq. (4.90)], so that the $\{\psi_{\mathrm{R},n}\}$ describe electrons flowing from the EM into the left lead, which is again consistent with causality.

The total currents due to the states originating in the left ($I^{\mathrm{L}}_{\mathrm{L}}$) and right ($I^{\mathrm{L}}_{\mathrm{R}}$) lead

are

$$I_{\mathrm{L}}^{\mathrm{L}} = \int dE \sum_{n=1}^{N_{\mathrm{L}}(E)} f_{\mathrm{L}}(E)\, \mathcal{N}_{\mathrm{L},n}(E)\; I_{\mathrm{L},n}^{\mathrm{L}}, \tag{4.91}$$

$$I_{\mathrm{R}}^{\mathrm{L}} = \int dE \sum_{n=1}^{N_{\mathrm{R}}(E)} f_{\mathrm{R}}(E)\, \mathcal{N}_{\mathrm{R},n}(E)\; I_{\mathrm{R},n}^{\mathrm{L}}, \tag{4.92}$$

so that $I^{\mathrm{L}} = I_{\mathrm{L}}^{\mathrm{L}} + I_{\mathrm{R}}^{\mathrm{L}}$. By using the result of Eq. (4.89) we obtain for $I_{\mathrm{L}}^{\mathrm{L}}$

$$I_{\mathrm{L}}^{\mathrm{L}} = \frac{1}{\hbar} \int dE f_{\mathrm{L}}(E) \sum_{n=1}^{N_{\mathrm{L}}(E)} \mathcal{N}_{\mathrm{L},n}(E) \left(\varphi_{\mathrm{L},n}^{\dagger} K_{\mathrm{LM}} G_{\mathrm{M}}^{\dagger} \Gamma_{\mathrm{R}}\; G_{\mathrm{M}} K_{\mathrm{ML}}\; \varphi_{\mathrm{L},n} \right).$$

Since the current for each WF is larger than or at least equal to zero, also $I_{\mathrm{L}}^{\mathrm{L}} \geq 0$. The quantity in brackets is just a number, so that it can be rewritten as a trace. This allows us to rearrange the matrices to

$$I_{\mathrm{L}}^{\mathrm{L}} = \frac{1}{\hbar} \int dE\; f_{\mathrm{L}}(E)\; \mathrm{Tr} \left[K_{\mathrm{ML}} \left(\sum_{n=1}^{N_{\mathrm{L}}(E)} \mathcal{N}_{\mathrm{L},n}(E) \varphi_{\mathrm{L},n} \varphi_{\mathrm{L},n}^{\dagger} \right) K_{\mathrm{LM}}\; G_{\mathrm{M}}^{\dagger} \Gamma_{\mathrm{R}}\; G_{\mathrm{M}} \right].$$

By using Eqs. (4.48) and (4.32) this becomes

$$I_{\mathrm{L}}^{\mathrm{L}} = \frac{1}{h} \int dE\; f_{\mathrm{L}}(E)\; \mathrm{Tr} \left(\Gamma_{\mathrm{L}} G_{\mathrm{M}}^{\dagger} \Gamma_{\mathrm{R}} G_{\mathrm{M}} \right).$$

Here $h = 2\pi\hbar$ is Planck's constant. In an analogous way the total current from the left lead into the EM, carried by the states originating from the right lead, is obtained

$$I_{\mathrm{R}}^{\mathrm{L}} = -\frac{1}{h} \int dE\; f_{\mathrm{R}}(E)\; \mathrm{Tr} \left(\Gamma_{\mathrm{L}} G_{\mathrm{M}}^{\dagger} \Gamma_{\mathrm{R}} G_{\mathrm{M}} \right). \tag{4.93}$$

and also here we note that $I_{\mathrm{R}}^{\mathrm{L}} \leq 0$. We can now define the transmission coefficient T [54, 60, 197] as

$$T = \mathrm{Tr} \left(\Gamma_{\mathrm{L}} G_{\mathrm{M}}^{\dagger} \Gamma_{\mathrm{R}} G_{\mathrm{M}} \right), \tag{4.94}$$

with $T \geq 0$. The total probability current then is

$$I^{\mathrm{L}} = \frac{1}{h} \int dE\; T(E)\; \left[f_{\mathrm{L}}(E) - f_{\mathrm{R}}(E) \right]. \tag{4.95}$$

This is the well-known Landauer-Büttiker result for the current through a two-terminal device [54]. We note that I^{L} depends only on the difference between the Fermi energies of the leads, which by definition is equal to the bias voltage eV_{b}. The equation shows that only the states in the bias window, i.e. those lying in the energy range between $E_{\mathrm{F,L}}$ and $E_{\mathrm{F,R}}$, contribute to the current with an amplitude proportional to $T(E)$. We note that if the density of states in any of the leads is zero at a

given energy, the corresponding Γ-matrix will be zero, and therefore also the current will vanish. Since for a two-terminal device $I^{\mathrm{L}} = -I^{\mathrm{R}}$, we can define the two-terminal current through the EM I as being equal to I^{L}, so that

$$I = \frac{1}{h} \int dE\ T(E)\ [f_{\mathrm{L}}(E) - f_{\mathrm{R}}(E)]\,. \tag{4.96}$$

We note that this is the probability current, to obtain the electron current, I_{e}, it has to be multiplied with the charge of the electron e:

$$I_{\mathrm{e}} = \frac{e}{h} \int dE\ T(E)\ [f_{\mathrm{L}}(E) - f_{\mathrm{R}}(E)]\,. \tag{4.97}$$

If $T(E)$ is constant then $I_{\mathrm{e}} = \frac{e^2}{h} T V_{\mathrm{b}}$. The factor e^2/h is known as the "quantum of conductance" (usually a factor 2 is added for spin), and corresponds to the conductance of a system in the case where $T(E) = 1$. It also shows that I_{e} has always the same sign as V_{b}. In the spin-polarized case the transmission for majority spins $[T^{\uparrow}(E)]$ and the one for minority spins $[T^{\downarrow}(E)]$ are independent, and have to be evaluated separately for each spin. The total transmission then is $T(E) = T^{\uparrow}(E) + T^{\downarrow}(E)$. The total current is $I_e = I_e^{\uparrow} + I_e^{\downarrow}$, where $I^{\uparrow}$ is the current carried by the majority spins, and $I^{\downarrow}$ is the one carried by minority spins. These are given by

$$I_{\mathrm{e}}^{\sigma} = \frac{e}{h} \int dE\ T^{\sigma}(E)\ [f_{\mathrm{L}}(E) - f_{\mathrm{R}}(E)]\,, \tag{4.98}$$

where $\sigma = \uparrow, \downarrow$ is the spin index. We note that in chapters 6 and 7, where we present transport results for different systems, we omit the index "b" for the bias voltage, so that the bias voltage is simply referred to as V, and also for the electron current we omit the label "e". Moreover current and voltage are always given in SI units, and not in atomic units.

For an arbitrary number of leads the inflowing current through lead n is

$$I^n = \frac{1}{h} \int dE \sum_{m \neq n} T_{nm}(E)\ [f_n(E) - f_m(E)]\,, \tag{4.99}$$

where

$$T_{nm} = \mathrm{Tr}\left(\Gamma_n G_{\mathrm{M}}^{\dagger} \Gamma_m G_{\mathrm{M}}\right). \tag{4.100}$$

In this case it is not possible to define a single current through the EM, only the current through each of the leads has a physical meaning. This means that the conductivity is a tensor of order N_{leads}.

4.5 Brillouin zone sampling

The description presented so far is valid for general leads of arbitrary width. For two-dimensional (2D) surfaces the system is infinite in the plane perpendicular to the transport (PPT). If the system is periodic in the PPT we can apply Bloch's theorem in the 2D-plane. In the following we choose the coordinate system in such a way that the x- and y-axes lie in the PPT, so that the transport direction is along the z-axis (this is also the convention used in the SMEAGOL transport code [198]). The general scheme for the Bloch state expansion is analogous to the one for three-dimensional (3D) crystals described in Sec. 2.2.4, with the only difference that along z the system is not periodic. Therefore, whereas in Sec. 2.2.4 we denoted the lattice vectors as some generic 3D-vector $\bar{\mathbf{R}}$ [Eq. (2.73)], for clarity in this section we explicitly define two crystal vectors $\mathbf{R}_1$ and $\mathbf{R}_2$ that span the x-y plane. The full three-dimensional system then is obtained by stacking these two-dimensional layers along the z direction, with no periodicity assumed along the stacking direction.

We label the WF coefficients in such a way that $\psi_{m_1 m_2 \alpha}$ corresponds to the WF coefficient for the localized orbital α (at an arbitrary position along z), in the unit cell at position $m_1\mathbf{R}_1 + m_2\mathbf{R}_2$ in the 2D plane (m_1 and m_2 are integers). The matrix KS equation for this system then reads

$$\sum_{m_1' m_2' \alpha'} H_{m_1 m_2 \alpha, m_1' m_2' \alpha'}\ \psi_{m_1' m_2' \alpha'} = E \sum_{m_1' m_2' \alpha'} S_{m_1 m_2 \alpha, m_1' m_2' \alpha'}\ \psi_{m_1' m_2' \alpha'}. \tag{4.101}$$

We can now apply Bloch's theorem to the wave function in the plane

$$\psi_{\mathbf{k}, m_1 m_2 \alpha} = e^{i\mathbf{k}(m_1\mathbf{R}_1 + m_2\mathbf{R}_2)}\ \psi_{\mathbf{k},\alpha}, \tag{4.102}$$

where $\mathbf{k}$ is a 2D wave-vector with $\mathbf{k}.\hat{\mathbf{z}} = 0$ ($\hat{\mathbf{z}}$ is the unit vector along the z-axis), which lies in the 2D BZ defined by $\mathbf{R}_1$ and $\mathbf{R}_2$. For a given $\mathbf{k}$-point the KS equation then becomes

$$\sum_{\alpha'} H_{\mathbf{k},\alpha\alpha'}\ \psi_{\mathbf{k},\alpha'} = E \sum_{\alpha'} S_{\mathbf{k},\alpha\alpha'}\ \psi_{\mathbf{k},\alpha'}, \tag{4.103}$$

where the effective matrices are defined as

$$H_{\mathbf{k},\alpha\alpha'} = \sum_{m_1, m_2} H_{00\alpha, m_1 m_2 \alpha'}\ e^{i\mathbf{k}(m_1\mathbf{R}_1 + m_2\mathbf{R}_2)} \tag{4.104}$$

$$S_{\mathbf{k},\alpha\alpha'} = \sum_{m_1, m_2} S_{00\alpha, m_1 m_2 \alpha'}\ e^{i\mathbf{k}(m_1\mathbf{R}_1 + m_2\mathbf{R}_2)}. \tag{4.105}$$

For each $\mathbf{k}$-point the effective Hamiltonian and overlap matrices are different, so that the system can be viewed as a collection of $\mathbf{k}$-dependent quasi-1D systems.

It is now again possible to set up a two-terminal device, where $H_{\mathbf{k}}$ and $S_{\mathbf{k}}$ have the same shape as the corresponding matrices for a 1D system [Eqs. (4.14) and (4.15)]. The only difference is that due to the geometry in this case it is not possible to have more than two leads. Therefore all the results obtained in the previous sections of this chapter are valid also in this case for each **k**-point independently. For each **k**-point there is a corresponding lesser Green's function $G^{<}_{\mathrm{M},\mathbf{k}}(E)$ and a density matrix $\rho_{\mathbf{k}}$ that satisfy

$$\rho_{\mathbf{k}} = \frac{1}{2\pi i} \int dE \; G^{<}_{\mathrm{M},\mathbf{k}}(E). \tag{4.106}$$

Analogously to Eq. (2.79) the total density matrix then is

$$\rho_{m_1 m_2, m_1' m_2'} = \frac{1}{\Omega_{\mathrm{BZ}}} \int_{BZ} d\mathbf{k} \; \rho_{\mathbf{k}} \; e^{i\mathbf{k}\left[(m_1 - m_1')\mathbf{R}_1 + (m_2 - m_2')\mathbf{R}_2\right]}, \tag{4.107}$$

where the integral runs over the 2D BZ.

Following the same procedure as for the 1D case for each **k**-point we obtain the corresponding **k**-dependent transmission coefficient $T_{\mathbf{k}}(E)$. The total transmission coefficient $T(E)$ is obtained by integrating $T_{\mathbf{k}}(E)$ over the 2D BZ

$$T(E) = \frac{1}{\Omega_{\mathrm{BZ}}} \int_{\mathrm{BZ}} d\mathbf{k} \; T_{\mathbf{k}}(E). \tag{4.108}$$

The current is then calculated by using Eq. (4.97). In the spin-polarized case the transmission for majority spins $[T^{\uparrow}(E)]$ and the one for minority spins $[T^{\downarrow}(E)]$ are independent, and have to be evaluated separately for each spin

$$T^{\sigma}(E) = \frac{1}{\Omega_{\mathrm{BZ}}} \int_{\mathrm{BZ}} d\mathbf{k} \; T^{\sigma}_{\mathbf{k}}(E) \tag{4.109}$$

where $\sigma = \uparrow, \downarrow$ is the spin index, and $T^{\sigma}_{\mathbf{k}}(E)$ is the **k**-dependent transmission for spin σ. The total transmission then is $T(E) = T^{\uparrow}(E) + T^{\downarrow}(E)$.

4.6 The SMEAGOL code

The concepts introduced in this chapter apply to any system that can be described by a set of localized orbitals of finite range, and the formalism is structurally equivalent to a a tight-binding scheme. The only input needed for the NEGF formalism is the Hamiltonian of the system in a matrix form. In a tight-binding approach the elements of the Hamiltonian are parameters, whereas here the Hamiltonian is constructed using the KS scheme. The NEGF scheme described in this chapter can be interfaced with any DFT code that generates a Hamiltonian in tight-binding-like form. In the SMEAGOL [61, 68] transport code the NEGF formalism has been interfaced with

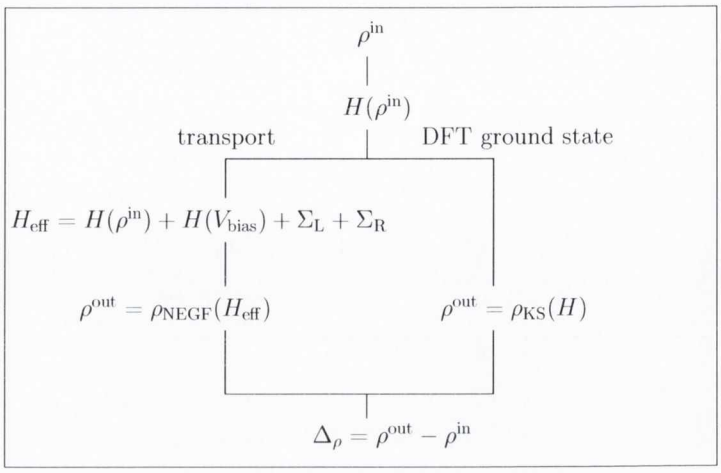


Figure 4.2: Schematic diagram of the interface between SIESTA and SMEAGOL. The left-hand side describes the path followed by a SMEAGOL calculation, while the right-hand side corresponds to the path followed by a SIESTA calculation.

the *ab-initio* DFT code SIESTA [69], which is described in Sec. 2.2. SMEAGOL is a joint project between groups in Trinity College Dublin, Lancaster University and Universidad de Oviedo. The code is distributed free of charge for academic purposes under the "SMEAGOL academic license". A detailed description of the SMEAGOL code is given in Refs. [61, 178], and a practical guide to the use of SMEAGOL is described in detail in the "User Guide" that is distributed together with the code (also found in the appendix of Ref. [178]). One of the advantages of using the SIESTA code for the DFT part is that it constructs the Hamiltonian very efficiently even for large systems. In fact in SIESTA this operation scales linearly with system size. Moreover the Hamiltonian and overlap matrices are constructed in the required tight-binding like form. Other transport codes, that are based on the same concepts, are described in Refs. [62, 63, 199].

Fig. 4.2 shows schematically how the NEGF transport formalism of SMEAGOL is interfaced to SIESTA. The calculation starts by setting up a trial initial density matrix ρ^{in}. With this density matrix the Kohn-Sham Hamiltonian $H(\rho^{\text{in}})$ is constructed in SIESTA. In a normal SIESTA calculation the output charge density ρ^{out} is equal to the ground state KS charge density $\rho^{\text{out}} = \rho_{\text{KS}}(H)$, which is obtained by diagonalizing $H(\rho^{\text{in}})$ and occupying the levels up to the Fermi energy (see Sec. 2.2.3). In a SMEAGOL transport calculation however ρ^{out} is obtained with the NEGF formalism via Eq. (4.64), so that it is equal to the NEGF charge density $\rho_{\text{NEGF}}(H)$. In this case two terms are added to H, the first is the bias-dependent part $H(V_{\text{b}})$, and the second part are the self-energies. The way these are calculated in SMEAGOL is described in chapter 5. The Hamiltonian and overlap matrices needed to construct the

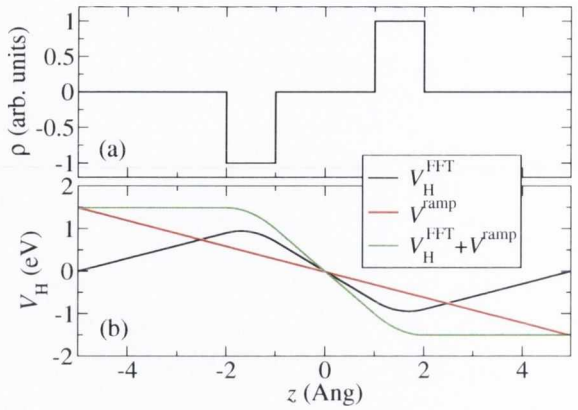


Figure 4.3: (a) Model charge density ρ for a 1D system, and (b) corresponding Hartree potential V_{H} for different boundary conditions.

self-energies are obtained from a previous run for an infinite bulk system corresponding to the leads. Just like in a normal DFT calculation this procedure is iterated until self-consistency is achieved, so that $\Delta_\rho = \rho^{\mathrm{out}} - \rho^{\mathrm{in}}$ is smaller than some chosen tolerance.

The additional potential $H(V_{\mathrm{b}})$ is chosen in such a way that it gives the correct boundary conditions for the electrostatic potential for a finite bias calculation at convergence. The basic principle is illustrated in Fig. 4.3 for a model 1D system. Out of equilibrium the potential drop in a device is generated by a dipole along the transport direction (the z-axis), generated by the difference in the chemical potentials and the resistance of the junction. This dipole charge is schematically shown in Fig. 4.3(a). In order to obtain the electrostatic potential $V_{\mathrm{H}}(z)$ the Poisson equation is solved for this charge density $\rho(z)$. In SIESTA this is done by using a fast Fourier transform (FFT) algorithm. This however implies that periodic boundary conditions are assumed in all directions, so that the potential is identical to the left and to the right end of the unit cell. For the 1D charge distribution of Fig. 4.3(a) the resulting Hartree potential $V_{\mathrm{H}}^{\mathrm{FFT}}(z)$ is the black curve in 4.3(b). It can be seen that indeed $V_{\mathrm{H}}^{\mathrm{FFT}}(z_{\mathrm{l}}) = V_{\mathrm{H}}^{\mathrm{FFT}}(z_{\mathrm{r}})$, where z_{l} is the position of the left end of the unit cell, and z_{r} corresponds to the right end. For a transport calculation however the necessary boundary condition is $V_{\mathrm{H}}(z_{\mathrm{l}}) = V_{\mathrm{H}}(z_{\mathrm{r}}) + eV_{\mathrm{b}}$ [green curve in Fig. 4.3(b)]. Since the Hartree potential is only defined up to a constant and a term linear in z, this boundary condition can be uniquely obtained by adding the term $V^{\mathrm{ramp}}(z) = eV_{\mathrm{b}}\left(z - \frac{z_{\mathrm{l}}+z_{\mathrm{r}}}{2}\right) / (z_{\mathrm{r}} - z_{\mathrm{l}})$ [red curve in Fig. 4.3(b)] to the potential calculated with the FFT. By adding $V^{\mathrm{ramp}}(z)$ to the SIESTA Hartree potential the correct boundary conditions for the electrostatic potential are obtained in SMEAGOL.

We conclude this section with a note on how the integral of Eq. (4.64) is calculated.

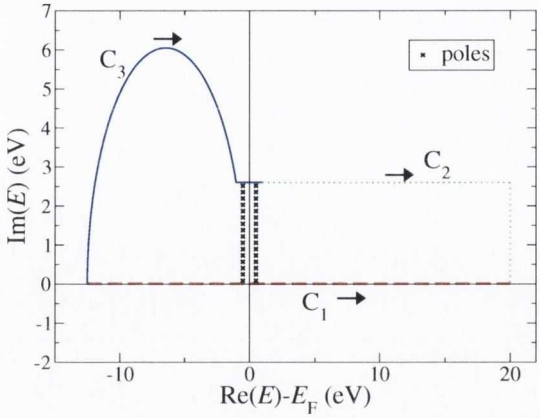


Figure 4.4: Path of the contour integral used in SMEAGOL.

In order to reduce the computational effort the lesser GF, $G^<_{\mathrm{M}}$, is split up into its equilibrium part $G^<_{\mathrm{M,eq}}$ [Eq. (4.71)] and its out of equilibrium part $G^<_{\mathrm{M,neq}}$ [Eq. (4.72)]. The integral over $G^<_{\mathrm{M,neq}}$ is non-zero only in the bias window, which is the energy range where $f_{\mathrm{L}}(E) \neq f_{\mathrm{R}}(E)$. We note that even if this energy range is rather small, the number of integration points can be very large, especially if $G^<_{\mathrm{M,neq}}$ has sharp peaks (see chapter 6). The upper bound of the integral of $G^<_{\mathrm{M,eq}}$ can also be bound by the Fermi-Dirac distribution with the largest Fermi energy, however the lower bound has to be lower than the energy of the state with the lowest energy in the EM. The integration range is therefore rather large, typically of the order of a few tens of eV. Moreover also $G^<_{\mathrm{M,eq}}$ can have sharp peaks for real energies, especially if there are localized states, so that a very fine integration mesh would be needed. However since $G^<_{\mathrm{M,eq}}$ can be analytically continued for positive complex energies (see Sec. 4.3.4), it is possible to perform the integral as a contour integral along the path indicated in Fig. 4.4. The integral along C_1, which is the one to evaluate, is equal to the sum of the ones on C_2 and on C_3, plus the contribution coming from the poles enclosed in the path of the complex integral. The upper bound of C_3 (corresponding to the lower bound of C_2) is chosen in such a way that all the Fermi-Dirac distributions are approximately zero above that energy. The integral over C_2 can therefore be assumed to be zero. Since for energies with positive imaginary part $G^<_{\mathrm{M,eq}}$ has no poles, the only poles inside the contour are the ones of the Fermi-Dirac distributions (Matsubara frequencies [200]). These are found where $1 + e^x = 0$, with $x = (E - E_{\mathrm{F},\{L/R\}})/T$, which is the case for $x = i(2n + 1)\pi$ $(n \in \mathbb{N})$. The poles are therefore found for $\mathrm{Re}(E) = E_{\mathrm{F},\{L/R\}}$ and $\mathrm{Im}(E) = (2n + 1)\pi T$. The distance in energy between two poles is $2\pi T$, so that the number of poles inside the contour is proportional to T. For $T = 0K$ it is not possible to perform this contour integral, since an infinite number of

poles would have to be added. In SMEAGOL the integral over C_1 is replaced with the one over C_3, adding also the contributions from the poles. By doing this the number of integration points is drastically reduced, usually less than about 100 points are needed for this numerical integration. If we use room temperature for T, then the number of poles used is usually of the order of a few tens.

4.7 Limits of the NEGF formalism

We conclude this chapter by listing some of the limits of the NEGF formalism used in conjunction with DFT, as described in this section. Although this scheme works well for many systems, there are many cases where it leads to erroneous predictions. We note that in this chapter we have only outlined the NEGF concepts for elastic transport. Inelastic effects, caused for example by electron-phonon or electron-photon scattering, can however in principle be included in the NEGF formalism in a mean-field type approach [54].

First we note that it is assumed that the single-particle KS-eigenvalues correspond to the energy levels for the real interacting-electron system (an assumption which is usually made also in ground state DFT calculations). Furthermore it is assumed that the KS Hamiltonian is also valid out of equilibrium, although this Hamiltonian is derived only for the ground state. An extensive discussion of this topic can be found in Ref. [201]. There it is shown that as a first approximation the KS Hamiltonian can be used for transport calculations, but that by doing this a few contributions to the electron transport are neglected.

If one still insists in using DFT as a single particle theory, the problem of the correct position of the KS-eigenvalues must be faced. Just as in ground state DFT, for many systems the LDA or GGA exchange correlation (XC) potentials do not give the correct results. The transport and charging properties critically depend on the position of the energy eigenvalues. However if local XC-potentials are used these are often at a wrong energy, and this for example causes major errors in calculations of transport across molecules [93]. Moreover, since LDA and GGA do not correctly reproduce the derivative discontinuity (see Secs. 2.1.3 and 2.1.5), weakly coupled states can be fractionally charged, whereas the charge should always be integer for such systems (or at least very close to integer) [202]. The use of more advanced XC-functionals however should solve these problems, and the possibility of the use of these for transport problems is a currently active area of research. The ASIC XC potential (Sec. 2.1.5) does usually improve the alignment of the energy levels, however it still fails to describe the derivative discontinuity [93].

Another limit is the fact that the NEGF uses a single particle picture to describe the electrons. For this reason it can not correctly describe transport through very weakly coupled states, like quantum dots in the Coulomb blockade regime, even if an appropriate XC potential was used. In this cases many-body effects, that are not included in the NEGF formalism described here, may play an important role. The transport properties of these systems are usually calculated using a master-equation approach [203, 204, 205, 206, 207].

In many systems localized states are found in the EM, that are not coupled to the leads. The occupation of such states in the EM is not defined within the NEGF formalism. In chapter 6 we set their occupation "by hand" by using a physically meaningful occupation. However this is not possible for all systems. Time-dependent calculations show that in this case it is indeed not always possible to obtain a steady state current [187, 208]. In Ref. [209] it is shown that for some systems the NEGF predicts multiple solutions at a given bias voltage, but within a time dependent approach only one of these is found to be stable.

From a practical point of view another source of uncertainty, independent of the NEGF, is the fact that the experimental atomic positions of typical nano-scale devices are usually not exactly known. Since the transport properties critically depend on these, the theoretical predictions can deviate drastically from experiment if the atomic positions are not the same.

4.8 Conclusions

In this chapter the main concepts of the NEGF formalism for elastic transport have been outlined. The NEGF formalism can be used to describe the quantum-mechanical transport properties of nano-devices at finite applied bias potential. The main advantage of this formalism is that it allows to determine the bias-dependent occupation of the states in the EM, which is not possible using standard ground-state DFT. The current can then be calculated by integrating the transmission coefficient over the bias window. We have described the particular implementation of the NEGF formalism in the *ab initio* transport code SMEAGOL, which is used for the calculations in chapters 5, 6 and 7. Finally we have listed some of the problems and limitations of the NEGF formalism used in conjunction with DFT.

Chapter 5

Self-energies for semi-infinite periodic quasi-1D systems

In chapter 4 the leads self-energies (SEs) are introduced [Eqs. (4.25) and (4.26)]. These play a central role within the NEGF formalism, since they determine the effect of the semi-infinite leads on the scattering region. In order to calculate the SEs, the GFs for the semi-infinite system, defined in Eqs. (4.18) and (4.19), are needed. It is now assumed that the coupling Hamiltonian matrices $H_{\mathrm{LM/RM}}$ (also generally denoted as hopping matrices) have a strictly finite range, so that only a finite number of atomic layers in the leads are coupled directly to the EM. In this case only a limited number of matrix elements of the leads GF are needed in order to obtain the SE via Eqs. (4.25) and (4.26). This part of the GF is called the surface Green's function (SGF), since the needed matrix elements are those of the terminal surface layer of the isolated lead (detached from the EM). In this chapter a complete prescription for the numerical calculation of the SGF and SE of semi-infinite, periodic, quasi-one-dimensional (quasi-1D) systems is presented. Moreover a detailed error analysis on the SE, as obtained with the presented algorithm, is performed. This work extends previous results, generating a robust algorithm to be used in conjunction with *ab initio* electronic structure methods.

Generally the SGFs can be obtained either with recursive methods [210, 211, 212, 213] or by using a semi-analytic formula [214, 63, 215, 216, 197, 65]. Recursive methods are affected by poor convergence for some critical systems, typically when the Hamiltonian for the leads is rather sparse. Semi-analytical methods instead bypass those problems by construction, however major difficulties arise if the hopping matrices are singular or, more generally, ill conditioned. Unfortunately the condition of the Hamiltonian is set by the electronic structure of the leads and by the unit cell used, and thus it is largely not controllable. In the original version of SMEAGOL the algorithm described in Ref. [214] was used to construct the SGF. This method

is however undefined for singular hopping matrices. By using a generalized singular value decomposition (GSVD) the singularities in the leads Hamiltonian could be partially eliminated [68], however due to the inherent instability of the used algorithm the number of failures in the calculation of the SGF is still very large. Here we present an improved semi-analytical method that overcomes these limitations and thus represents a robust algorithm for quantum transport based on *ab initio* DFT Hamiltonians. This extended scheme is now implemented in the SMEAGOL code.

In the first part of this chapter the extended algorithm for the calculation of the SE is presented. First the construction of the Green's function of an infinite 1D system as derived in Ref. [214] is recast into a more general form based on the notion of a complex group velocity. Then we present an extension of such method to the calculation of the SGF and SE that is defined also for the case of singular hopping matrices. This largely improves the numerical accuracy. However we find that even such an improved scheme sometimes fails if the hopping matrices are close to being singular. We overcome this limitation by performing a transformation of the hopping matrix that reduces its condition number κ, defined as the ratio between its largest to its smallest singular value [217, 218]. This transformation limits the maximum absolute value of the imaginary part of the Bloch wave vectors, increasing both accuracy and stability. Two approaches are presented: the first is based on a singular value decomposition (SVD), and the second consists in adding a random noise matrix of small amplitude to the Hamiltonian. A SVD transformation has been previously employed for example to calculate the complex band-structure of long molecules [219]. By using the SVD the size of the Hamiltonian matrix can be reduced, so that the computation becomes considerably faster for large systems.

In the second part of this chapter we present three examples of calculations performed with the new implementation. A detailed error analysis of the scheme is performed. The highest accuracy is obtained if no inversion of the usually ill conditioned hopping matrix is involved. We compare the results to the ones obtained by using the original method of Ref. [214], finding a considerable improvement. Although the algorithm appears very robust, our detailed error analysis reveals that for a given system the accuracy is lost at some specific energies. This is caused by the divergence of one of the SE eigenvalues. The physical origin of this behavior lies in the presence of surface states that are very weakly coupled to the semi-infinite leads. These surface states appear whenever at a given energy the set of Bloch functions (with both real and imaginary wave vectors) for the infinite quasi-1D system is linearly dependent. In the simplest case this corresponds to two Bloch functions being equal inside the unit cell. A small imaginary part is added to the energy in a small

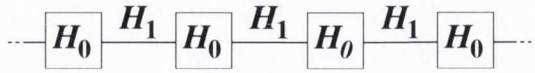


Figure 5.1: Schematic representation of the system with onsite Hamiltonian H_0 and hopping H_1. The overlap matrix has the same structure.

energy range around the surface state. It is shown that this has little effect on the transport properties in the high transmission regime, whereas for low transmission it has a substantial influence on the results. Crucially only a very small imaginary part is used, and moreover this is added only around the energy of the surface state, so that the error can be carefully controlled.

5.1 Retarded Green's function for an infinite system

Following the scheme introduced in Ref. [214] the construction of the retarded Green's function for an infinite quasi-1D system is now recalled. This is the starting point for the calculation of the SGF. It is assumed that the interaction is of finite range, so that the Hamiltonian elements between two distant orbitals is strictly zero. This assumption is always fulfilled if the localized basis set of SIESTA is used (see Sec. 2.2.2). The size of the unit cell can then be chosen to guarantee interaction only to the first nearest neighboring unit cells. The total Hamiltonian of the system $H_{zz'}$ (the integers z and z' label the unit cells along the 1D stacking direction) can then be written as

$$H_{zz'} = H_0\ \delta_{zz'} + H_1\ \delta_{z,z'-1} + H_{-1}\ \delta_{z,z'+1}, \tag{5.1}$$

where H_0, H_1 and H_{-1} are $N \times N$ matrices, with N being the number of orbitals comprised in the unit cell (see Fig. 5.1). If time-reversal symmetry holds then $H_0 = H_0^\dagger$, and $H_{-1} = H_1^\dagger$. We further assume that the overlap matrix $S_{zz'}$ has the same structure and range of the Hamiltonian

$$S_{zz'} = S_0\ \delta_{zz'} + S_1\ \delta_{z,z'-1} + S_{-1}\ \delta_{z,z'+1}, \tag{5.2}$$

where S_0, S_1 and S_{-1} are again $N \times N$ matrices with the same meaning of their Hamiltonian counterparts. In SMEAGOL the matrices $H_0, H_1, H_{-1}, S_0, S_1$ and S_{-1} are obtained from the converged solution of a SIESTA ground state calculation.

5.1.1 Bloch states expansion

The solutions of the Hamiltonian equation for the associated infinite periodic 1D system $\sum_{z'} H_{zz'}\ \psi_{z'} = E\ \sum_{z'} S_{zz'}\ \psi_{z'}$ are Bloch functions $\psi_z = e^{ikz}\phi$, where ψ_z and

ϕ are N-dimensional vectors and k is the wave vector, which in general is a complex number. For a given real or complex energy E there are $2N$ solutions with wave vectors k_n and corresponding wave functions ϕ_n. Each of them satisfies

$$\begin{aligned}\left(H_0 + H_1 e^{ik_n} + H_{-1} e^{-ik_n}\right) \phi_n \quad &= \\ E\left(S_0 + S_1 e^{ik_n} + \quad S_{-1} \quad e^{-ik_n}\right) \phi_n.& \end{aligned} \tag{5.3}$$

We now define $K_\alpha = H_\alpha - ES_\alpha$, $(\alpha = -1, 0, 1)$. In contrast to the definition of Sec. 4.2 here we do not add a positive infinitesimal imaginary part δ to the energy, so that for real energies $K_\alpha = K_\alpha^\dagger$. The above equation can then be rewritten as

$$\left(K_0 + K_1 e^{ik_n} + K_{-1} e^{-ik_n}\right) \phi_{\mathrm{R},n} = 0, \tag{5.4}$$

where the additional index R denotes explicitly that the solution is a right eigenvector. The corresponding left eigenvector $\phi_{\mathrm{L},n}$ satisfies the associated equation

$$\phi_{\mathrm{L},n}^\dagger \left(K_0 + K_1 e^{ik_n} + K_{-1} e^{-ik_n}\right) = 0. \tag{5.5}$$

For real energies we have $\phi_{\mathrm{L},n} = \phi_{\mathrm{L}}(k_n) = \phi_{\mathrm{R}}(k_n^*)$, so that in the case of real k_n (propagating states) left and right eigenvectors are equal. For complex k_n left and right eigenvectors are different, describing left- and right-decaying states.

The sets $\{k_n\}$, $\{\phi_{\mathrm{R},n}\}$ and $\{\phi_{\mathrm{L},n}\}$ that satisfy Eqs. (5.4) and (5.5) at a given energy can be found by solving a quadratic eigenvalue problem [220, 221] of the form

$$\begin{pmatrix} -K_0 & -K_{-1} \\ \mathbb{1}_N & \mathbb{0}_N \end{pmatrix} \Phi_{\mathrm{R},n} = e^{ik_n} \begin{pmatrix} K_1 & \mathbb{0}_N \\ \mathbb{0}_N & \mathbb{1}_N \end{pmatrix} \Phi_{\mathrm{R},n} \tag{5.6}$$

for the right eigenvectors, and

$$\Phi_{\mathrm{L},n}^\dagger \begin{pmatrix} -K_0 & -K_{-1} \\ \mathbb{1}_N & \mathbb{0}_N \end{pmatrix} = e^{ik_n} \Phi_{\mathrm{L},n}^\dagger \begin{pmatrix} K_1 & \mathbb{0}_N \\ \mathbb{0}_N & \mathbb{1}_N \end{pmatrix} \tag{5.7}$$

for the left eigenvectors. Here $\mathbb{1}_N$ is the $N \times N$ unit matrix, $\mathbb{0}_N$ is the $N \times N$ zero matrix (a general $i \times j$ zero matrix is denoted as $\mathbb{0}_{i,j}$), and

$$\Phi_{\mathrm{R},n} = \begin{pmatrix} e^{i\frac{k_n}{2}} \\ e^{-i\frac{k_n}{2}} \end{pmatrix} \frac{\phi_{\mathrm{R},n}}{\sqrt{v_n}}, \tag{5.8}$$

$$\Phi_{\mathrm{L},n}^\dagger = \frac{i\phi_{\mathrm{L},n}^\dagger}{\sqrt{v_n}} \begin{pmatrix} e^{i\frac{k_n}{2}} & , -e^{-i\frac{k_n}{2}} K_{-1} \end{pmatrix}. \tag{5.9}$$

The normalization constant is the square root of the complex group velocity $v_n = \partial E / \partial k_n$ $(\hbar = 1)$ equal to

$$v_n = \frac{i}{l_n} \phi_{\mathrm{L},n}^\dagger \left(K_1 e^{ik_n} - e^{-ik_n} K_{-1}\right) \phi_{\mathrm{R},n}, \tag{5.10}$$

$$l_n = \phi_{\mathrm{L},n}^\dagger \left(S_0 + S_1 e^{ik_n} + S_{-1} e^{-ik_n}\right) \phi_{\mathrm{R},n}. \tag{5.11}$$

In the following we assume that the eigenvectors $\phi_{\mathrm{R},n}$ and $\phi_{\mathrm{L},n}$ are always normalized to give $l_n = 1$. For real energies $v(k_n^*) = v_n^*$, so that the group velocity is real if the imaginary part of k_n is zero. Note that, at variance with Ref. [214], Eqs. (5.6) and (5.7) avoid the inversion of K_1, so that they eliminate a possible source of singularities in the calculation of k_n, $\phi_{\mathrm{R},n}$ and $\phi_{\mathrm{L},n}$.

The full sets of left $\{\Phi_{\mathrm{L},n}\}$ and right eigenvectors $\{\Phi_{\mathrm{R},n}\}$ form a complete and orthogonal basis. The orthogonality relation is

$$\Phi_{\mathrm{L},n}^\dagger \begin{pmatrix} K_1 & \mathbb{0}_N \\ \mathbb{0}_N & \mathbb{1}_N \end{pmatrix} \Phi_{\mathrm{R},m} = c_n \delta_{nm}, \tag{5.12}$$

where c_n is a constant. This leads to

$$i\phi_{\mathrm{L},n}^\dagger \left(K_1 e^{ik_n} - e^{-ik_m} K_{-1}\right) \phi_{\mathrm{R},m} = v_n \; c_n \; \delta_{nm}. \tag{5.13}$$

For $n = m$ this equation is only satisfied if $c_n = 1$, in which case it corresponds to the definition of v_n. With the chosen normalization the basis is therefore orthonormal. The corresponding completeness relation then reads

$$\sum_{n=1}^{2N} \Phi_{\mathrm{R},n} \Phi_{\mathrm{L},n}^\dagger \begin{pmatrix} K_1 & \mathbb{0}_N \\ \mathbb{0}_N & \mathbb{1}_N \end{pmatrix} = \begin{pmatrix} \mathbb{1}_N & \mathbb{0}_N \\ \mathbb{0}_N & \mathbb{1}_N \end{pmatrix}, \tag{5.14}$$

and provides the three following useful relations

$$\sum_{n=1}^{2N} \frac{\phi_{\mathrm{R},n} \phi_{\mathrm{L},n}^\dagger}{v_n} = \mathbb{0}_N, \tag{5.15}$$

$$K_1 \sum_{n=1}^{2N} i e^{ik_n} \frac{\phi_{\mathrm{R},n} \phi_{\mathrm{L},n}^\dagger}{v_n} = \mathbb{1}_N, \tag{5.16}$$

$$K_{-1} \sum_{n=1}^{2N} -i e^{-ik_n} \frac{\phi_{\mathrm{R},n} \phi_{\mathrm{L},n}^\dagger}{v_n} = \mathbb{1}_N. \tag{5.17}$$

Note that in Eqs. (5.14-5.17) the sums run over all $2N$ solutions. If $K_1 = K_{-1}^\dagger$ and $K_0 = K_0^\dagger$ Eqs. (5.16) and (5.17) are equivalent.

5.1.2 Green's function

The retarded Green's function $g_{zz'}$ of the 1D system is defined by means of Green's equation [Eq. (4.2)]

$$\sum_{z'} g_{zz'} \left[(E + i\delta) \, S_{z'z''} - H_{z'z''}\right] = \delta_{zz''}, \tag{5.18}$$

with $\delta \to 0^+$ real. In what follows we present and expand, by using left and right Bloch functions, the solution to Eq. (5.18) given in Ref. [214] only in terms of the right eigenvectors ϕ_{R}. First we divide the $2N$ $\phi_{\mathrm{R},n}$ vectors into N right-going states, with either $\mathrm{Im}(k_n) > 0$ (right decaying) or $\mathrm{Im}(k_n) = 0$ and $v_n > 0$ (right propagating), and N left-going states, with either $\mathrm{Im}(k_n) < 0$ (left decaying) or $\mathrm{Im}(k_n) = 0$ and $v_n < 0$ (left propagating). As a matter of notation in order to distinguish left- from right-going states, in what follows we indicate the right-going states with k, ϕ and v, and the left-going states with a bar over these quantities, i.e. $\bar{k}$, $\bar{\phi}$ and $\bar{v}$.

As in Ref. [214] we introduce the duals $\tilde{\phi}_{\mathrm{R},n}$ of the right-going states $\phi_{\mathrm{R},n}$ defined by $\tilde{\phi}^{\dagger}_{\mathrm{R},n}\phi_{\mathrm{R},m} = \delta_{nm}$, and the duals $\tilde{\bar{\phi}}_{\mathrm{R},n}$ of the left-going states $\bar{\phi}_{\mathrm{R},n}$ defined by $\tilde{\bar{\phi}}^{\dagger}_{\mathrm{R},n}\bar{\phi}_{\mathrm{R},m} = \delta_{nm}$. If we define the matrices Q and $\bar{Q}$ as

$$\begin{aligned} Q &= \begin{pmatrix} \phi_{\mathrm{R},1} & \phi_{\mathrm{R},2} & \cdots & \phi_{\mathrm{R},N} \end{pmatrix}, \\ \bar{Q} &= \begin{pmatrix} \bar{\phi}_{\mathrm{R},1} & \bar{\phi}_{\mathrm{R},2} & \cdots & \bar{\phi}_{\mathrm{R},N} \end{pmatrix}, \end{aligned} \tag{5.19}$$

then the duals can be obtained by simple inversion:

$$\begin{aligned} \begin{pmatrix} \tilde{\phi}_{\mathrm{R},1} & \tilde{\phi}_{\mathrm{R},2} & \cdots & \tilde{\phi}_{\mathrm{R},N} \end{pmatrix} &= \left(Q^{-1}\right)^{\dagger}, \\ \begin{pmatrix} \tilde{\bar{\phi}}_{\mathrm{R},1} & \tilde{\bar{\phi}}_{\mathrm{R},2} & \cdots & \tilde{\bar{\phi}}_{\mathrm{R},N} \end{pmatrix} &= \left(\bar{Q}^{-1}\right)^{\dagger}. \end{aligned} \tag{5.20}$$

The inversions in Eqs. (5.20) are usually well defined, unless Q and $\bar{Q}$ do not have full rank. We will return on this aspect in Sec. 5.5, for the moment we assume that the duals can always be constructed.

The retarded GF calculated in Ref. [214] is then

$$g_{zz'} = \begin{cases} \sum_{n=1}^{N} \phi_{\mathrm{R},n} e^{ik_n(z-z')} \tilde{\phi}^{\dagger}_{\mathrm{R},n} \; \mathcal{V}^{-1} & z \geq z' \\ \sum_{n=1}^{N} \bar{\phi}_{\mathrm{R},n} e^{i\bar{k}_n(z-z')} \tilde{\bar{\phi}}^{\dagger}_{\mathrm{R},n} \; \mathcal{V}^{-1} & z \leq z', \end{cases} \tag{5.21}$$

with the matrix $\mathcal{V} = g_{zz}^{-1} = g_{00}^{-1}$ given by

$$\mathcal{V} = K_{-1} \left(\sum_{n=1}^{N} e^{-ik_n} \phi_{\mathrm{R},n} \tilde{\phi}^{\dagger}_{\mathrm{R},n} - \sum_{n=1}^{N} e^{-i\bar{k}_n} \bar{\phi}_{\mathrm{R},n} \tilde{\bar{\phi}}^{\dagger}_{\mathrm{R},n} \right). \tag{5.22}$$

We now introduce the right transfer matrices T_{R} and $\bar{T}_{\mathrm{R}}$,

$$T_{\mathrm{R}} = \sum_{n=1}^{N} \phi_{\mathrm{R},n} e^{ik_n} \tilde{\phi}^{\dagger}_{\mathrm{R},n}, \tag{5.23}$$

$$\bar{T}_{\mathrm{R}} = \sum_{n=1}^{N} \bar{\phi}_{\mathrm{R},n} e^{-i\bar{k}_n} \tilde{\bar{\phi}}^{\dagger}_{\mathrm{R},n}. \tag{5.24}$$

These are equivalent to the bulk transfer matrices introduced in Refs. [212, 222, 213] in the context of the recursive Green's function approach. Note that both T_{R} and $\bar{T}_{\mathrm{R}}$

have eigenvalues with complex modulus ≤ 1. For an integer z the following relations hold

$$\begin{aligned} (T_{\mathrm{R}})^{z} &= \sum_{n=1}^{N} \phi_{\mathrm{R},n} e^{ik_n z} \tilde{\phi}^{\dagger}_{\mathrm{R},n}, \\ \left(\bar{T}_{\mathrm{R}}\right)^{z} &= \sum_{n=1}^{N} \bar{\phi}_{\mathrm{R},n} e^{-i\bar{k}_n z} \tilde{\bar{\phi}}^{\dagger}_{\mathrm{R},n}, \end{aligned} \tag{5.25}$$

which allow us to write the Green's function of Eq. (5.21) as

$$g_{zz'} = \begin{cases} (T_{\mathrm{R}})^{z-z'} g_{00} & z \geq z' \\ \left(\bar{T}_{\mathrm{R}}\right)^{z'-z} g_{00} & z \leq z' \end{cases} . \tag{5.26}$$

In the same way $\mathcal{V}$ is rewritten as

$$\mathcal{V} = g_{00}^{-1} = K_{-1} \left(T_{\mathrm{R}}^{-1} - \bar{T}_{\mathrm{R}}\right) . \tag{5.27}$$

Note that although the matrices T_{R} and $\bar{T}_{\mathrm{R}}$ are in general well defined, the inverse of these matrices is not. In fact, if K_1 and K_{-1} are singular, there are some k_n with $\mathrm{Im}(k_n) \to \infty$, so that $e^{ik_n} = 0$ (see Sec. 5.3.1). In this case T_{R} does not have full rank and is therefore singular. The same argument holds for $\bar{T}_{\mathrm{R}}$. Eq. (5.27) can therefore be used only if the matrices K_1 and K_{-1} are not singular.

A possible way for overcoming such limitation is by using an equivalent form for the Green's function based on the left and right eigenvectors. The starting point is the relation (5.15) that will allow us to find the connection between the duals and the left eigenvectors. Eq. (5.15) contains a sum over both left- and right-going states. By moving the contribution of the left-going states to the right side of the equation, we obtain $\sum_{n=1}^{N} \frac{\phi_{\mathrm{R},n}\phi^{+}_{\mathrm{L},n}}{iv_n} = -\sum_{n=1}^{N} \frac{\bar{\phi}_{\mathrm{R},n}\bar{\phi}^{+}_{\mathrm{L},n}}{i\bar{v}_n} = B$, where we have introduced the auxiliary matrix B. By multiplying B from the left with either $\tilde{\phi}^{\dagger}_{\mathrm{R}}$ or $\tilde{\bar{\phi}}^{\dagger}_{\mathrm{R}}$ we obtain respectively $\tilde{\phi}^{\dagger}_{\mathrm{R},n} = \frac{1}{iv_n}\phi^{\dagger}_{\mathrm{L},n}B^{-1}$ and $\tilde{\bar{\phi}}^{\dagger}_{\mathrm{R},n} = \frac{1}{-i\bar{v}_n}\bar{\phi}^{\dagger}_{\mathrm{L},n}B^{-1}$. The matrix B is determined by inserting these relations into Eq. (5.22) and by using the identity (5.17). The result is $B = g_{00}$. The relation between the dual basis and the left eigenvectors is therefore

$$\tilde{\phi}^{\dagger}_{\mathrm{R},n} = \frac{1}{iv_n}\phi^{\dagger}_{\mathrm{L},n}g_{00}^{-1} \quad , \quad \tilde{\bar{\phi}}^{\dagger}_{\mathrm{R},n} = \frac{1}{-i\bar{v}_n}\bar{\phi}^{\dagger}_{\mathrm{L},n}g_{00}^{-1}. \tag{5.28}$$

This result allows us to rewrite the Green's function of Eq. (5.21) in a shorter form

$$g_{zz'} = \begin{cases} \sum_{n=1}^{N} \frac{1}{iv_n}\phi_{\mathrm{R},n}e^{ik_n(z-z')}\phi^{\dagger}_{\mathrm{L},n} & z \geq z' \\ \sum_{n=1}^{N} \frac{1}{-i\bar{v}_n}\bar{\phi}_{\mathrm{R},n}e^{i\bar{k}_n(z-z')}\bar{\phi}^{\dagger}_{\mathrm{L},n} & z \leq z'. \end{cases} \tag{5.29}$$

This result represents a generalization to complex energies and to systems breaking time-reversal symmetry of the solution given in Refs. [223, 224] for Hermitian

Hamiltonians, real energy and an orthogonal tight binding model. This derivation shows that the Green's function can be equivalently expressed by using the right eigenvectors and their duals [Eq. (5.21)], or both the right and left eigenvectors [Eq. (5.29)]. It is thus possible to move from one representation to the other through Eq. (5.28) that relates the duals to the left eigenvectors. One can then decide which representation to use, depending on the specific problem investigated. We note that Eq. (5.29) has the benefit that g_{00} can be calculated also in the case where the two matrices K_1 and K_{-1} are singular. For those k_n where $\mathrm{Im}(k_n) \to \infty$ the group velocity becomes $v_n = i\ \phi^{\dagger}_{\mathrm{L},n} K_0 \phi_{\mathrm{R},n}$ and is therefore well defined [$\bar{v}_n = -i\ \bar{\phi}^{\dagger}_{\mathrm{L},n} K_0 \bar{\phi}_{\mathrm{R},n}$ for $\mathrm{Im}(\bar{k}_n) \to -\infty$].

As a matter of completeness we show that a representation entirely based on the left Bloch functions and their duals $\tilde{\phi}_{\mathrm{L},n}$ and $\tilde{\bar{\phi}}_{\mathrm{L},n}$ is also possible. By multiplying Eq. (5.29) respectively by $\tilde{\phi}_{\mathrm{L},n}$ and $\tilde{\bar{\phi}}_{\mathrm{L},n}$ from the right we obtain the two relations

$$\tilde{\phi}_{\mathrm{L},n} = \frac{1}{iv_n} g_{00}^{-1} \phi_{\mathrm{R},n} \quad , \quad \tilde{\bar{\phi}}_{\mathrm{L},n} = \frac{1}{-i\bar{v}_n} g_{00}^{-1} \bar{\phi}_{\mathrm{R},n}. \tag{5.30}$$

The left transfer matrices T_{L} and $\bar{T}_{\mathrm{L}}$ are now defined as

$$T_{\mathrm{L}} = \sum_{n=1}^{N} \tilde{\phi}_{\mathrm{L},n} e^{ik_n} \phi^{\dagger}_{\mathrm{L},n} \tag{5.31}$$

$$\bar{T}_{\mathrm{L}} = \sum_{n=1}^{N} \tilde{\bar{\phi}}_{\mathrm{L},n} e^{-i\bar{k}_n} \bar{\phi}^{\dagger}_{\mathrm{L},n}, \tag{5.32}$$

and the Green's function of Eq. (5.29) can be rewritten as

$$g_{zz'} = \begin{cases} g_{00}\ (T_{\mathrm{L}})^{z-z'} & z \geq z' \\ g_{00}\ (\bar{T}_{\mathrm{L}})^{z'-z} & z \leq z' \end{cases}. \tag{5.33}$$

The structure of Eq. (5.33) is the same as that of Eq. (5.26), with the difference that now g_{00} is multiplied to the left of the transfer matrix. Finally we extend Eq. (5.27) and present four equivalent relations for the inverse of g_{00}

$$\begin{aligned} g_{00}^{-1} &= K_{-1} \left(T_{\mathrm{R}}^{-1} - \bar{T}_{\mathrm{R}}\right) = K_1 \left(\bar{T}_{\mathrm{R}}^{-1} - T_{\mathrm{R}}\right) \\ &= \left(T_{\mathrm{L}}^{-1} - \bar{T}_{\mathrm{L}}\right) K_{-1} = \left(\bar{T}_{\mathrm{L}}^{-1} - T_{\mathrm{L}}\right) K_1. \end{aligned} \tag{5.34}$$

The second of these relations can be shown by multiplying Eq. (5.16) by g_{00}^{-1} from the right and then by using Eq. (5.28). In the same way the third and fourth equations can be obtained by multiplying equations (5.16) and (5.17) by g_{00}^{-1} from the left.

In the following we will use mostly the quantities expressed in terms of the right eigenvectors only, however the same conclusions can be derived using the left eigenvectors.

5.1.3 Density of states

As an example of the use of the Green's function in the form of Eq. (5.29) we determine the spectral function A and the DOS, $\mathcal{N}$, of the infinite quasi-1D system. The spectral function [Eq. (4.5)] for such a system is

$$A_{zz'} = i\left[g - g^\dagger\right]_{zz'} = i\left[g_{zz'} - (g_{z'z})^\dagger\right]. \tag{5.35}$$

Using Eq. (4.9) the DOS projected on the unit cell z, $\mathcal{N}_z$, then is

$$\mathcal{N}_z = \frac{1}{2\pi}\mathrm{Tr}\left[\sum_{z'} A_{zz'}S_{z'z}\right]. \tag{5.36}$$

By using Eq. (5.2) this becomes

$$\mathcal{N}_z = \frac{1}{2\pi}\mathrm{Tr}\left[A_{zz}S_0 + A_{z,z-1}S_1 + A_{z,z+1}S_{-1}\right]. \tag{5.37}$$

In general the main contribution originates from the first term in the sum, which can be interpreted as the onsite DOS $\tilde{\mathcal{N}}_z$

$$\tilde{\mathcal{N}}_z = \frac{1}{2\pi}\mathrm{Tr}\left[A_{zz}S_0\right]. \tag{5.38}$$

We now calculate A and $\mathcal{N}$ for real E, so that $K_{-1} = K_1^\dagger$ and $K_0 = K_0^\dagger$. In this case for $\mathrm{Im}(k_n) = 0$ we have $\phi_{\mathrm{L},n} = \phi_{\mathrm{R},n}$, whereas if $\mathrm{Im}(k_n) \neq 0$ then $\phi_{\mathrm{L},n} = \phi_{\mathrm{L}}(k_n) = \bar{\phi}_{\mathrm{R}}(k_n^*)$. In the same way for $\mathrm{Im}(\bar{k}_n) = 0$ we have $\bar{\phi}_{\mathrm{L},n} = \bar{\phi}_{\mathrm{R},n}$, whereas if $\mathrm{Im}(\bar{k}_n) \neq 0$ then $\bar{\phi}_{\mathrm{L},n} = \bar{\phi}_{\mathrm{L}}(\bar{k}_n) = \phi_{\mathrm{R}}(\bar{k}_n^*)$. Therefore for each right decaying state with $\mathrm{Im}(k_n) > 0$ there is a left decaying state with $\bar{k}_n = k_n^*$ and $v(\bar{k}_n)^* = v(k_n)$. By using these relations when inserting the Green's function of Eq. (5.29) in the definition of $A_{zz'}$, the contribution from all the decaying states cancels out. The only remaining contributions come from the propagating states, also denoted as open channels. For these $k_n^* = k_n$, $\bar{k}_n = -k_n$ and $v(\bar{k}_n) = -v(k_n)$. With these constraints, and by using Eq. (5.29), the spectral function becomes

$$A_{zz'} = \sum_n^{N_{\mathrm{open}}} \frac{e^{ik_n(z-z')}}{v_n}\phi_{\mathrm{R},n}\phi_{\mathrm{R},n}^\dagger + \frac{e^{-ik_n(z-z')}}{v_n}\bar{\phi}_{\mathrm{R},n}\bar{\phi}_{\mathrm{R},n}^\dagger, \tag{5.39}$$

where N_{open} is the number of open channels (number of Bloch functions at a given energy with real positive k vector). If there are no open channels $A_{zz'} = 0$ and the Green's function is Hermitian. Finally, by using Eqs. (5.37) and (5.39), and the fact that the eigenvectors are normalized to give $l_n = 1$ [see Eq. (5.11)], the DOS at the site $z = 0$ is simply

$$\mathcal{N}_0 = \frac{1}{\pi}\sum_n^{N_{\mathrm{open}}} \frac{1}{v_n}. \tag{5.40}$$

This is the well known result for the DOS of infinite periodic 1D systems [60].

5.2 Surface Green's function and self-energy

The retarded Green's function g_S for a quasi-periodic system, where the left- and right-hand sides are separated at the position $z = 0$ (the left-hand side part extends from $z = -\infty$ to $z = -1$, and the right-hand side part from $z = 1$ to $z = \infty$, with no coupling between the cells located at $z = -1$ and $z = 1$), can be constructed from the Green's function g for the infinite chain as demonstrated in Ref. [214]:

$$g_{S,zz'} = g_{zz'} - g_{z0}\, g_{00}^{-1}\, g_{0z'}. \tag{5.41}$$

For $z, z' < 0$ therefore $g_{S,zz'}$ corresponds to the GF for the isolated left lead G_L [defined in Eq. (4.18)], and for $z, z' > 0$ it corresponds to the GF for the isolated right lead G_R [defined in Eq. (4.19)]. We then define the left SGF as $g_L = g_{S,-1,-1}$, and the right SGF as $g_R = g_{S,11}$, so that these correspond to the GF for the semi-infinite system at the surface layer.

It is now assumed that the coupling matrix to the left lead H_{LM} [see Eq. (4.14)] has the form

$$H_{LM} = \begin{pmatrix} \vdots \\ \mathbb{0}_N \\ \mathbb{0}_N \\ H_1 \end{pmatrix}, \tag{5.42}$$

and analogously that the one to the right lead is

$$H_{RM} = \begin{pmatrix} H_{-1} \\ \mathbb{0}_N \\ \mathbb{0}_N \\ \vdots \end{pmatrix}. \tag{5.43}$$

The overlap matrices $S_{\{L/R\}M}$ are assumed to have an analogous form. This implies that there is always one leads unit cell at each end of the EM unit cell (in general left and right-hand side lead can however be different). When setting up a SMEAGOL calculation such a constraint needs to be fulfilled. This can always be achieved by adding a layer of leads on each side of the EM. With this structure of $H_{\{L/R\}M}$ the SEs, defined in Eqs. (4.25) and (4.26), become:

$$\Sigma_L = K_{-1} g_L K_1, \tag{5.44}$$

$$\Sigma_R = K_1\, g_R K_{-1}. \tag{5.45}$$

In this form the SEs are an inherent property of the leads, since all the quantities on the right-hand side of the equation are themselves properties of the leads. These SEs are therefore independent of the EM they are joined to.

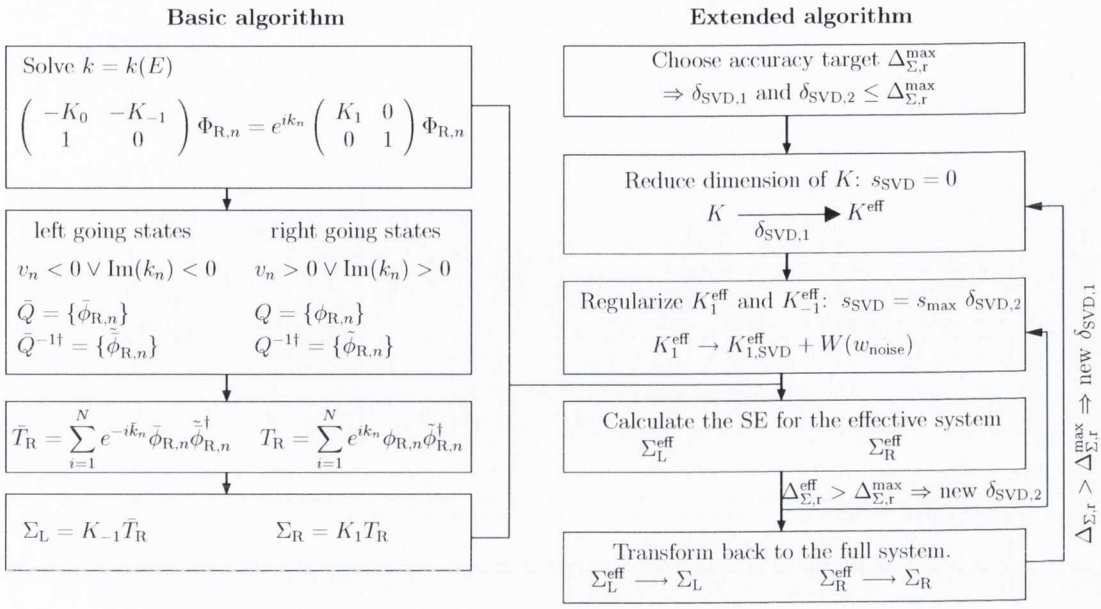


Figure 5.2: Schematic diagram of the basic algorithm described in Sec. 5.2 and of the extended algorithm described in Sec. 5.3.

The explicit form of the SGF can be obtained by using Eqs. (5.41) and (5.26)

$$\begin{aligned} g_{\mathrm{L}} &= \left(I_N - \bar{T}_{\mathrm{R}} T_{\mathrm{R}}\right) g_{00}, \\ g_{\mathrm{R}} &= \left(I_N - T_{\mathrm{R}} \bar{T}_{\mathrm{R}}\right) g_{00}. \end{aligned} \tag{5.46}$$

This corresponds to the form derived in Ref. [214]. Such a result can be simplified by using the relations in Eq. (5.34) for g_{00} to

$$\begin{aligned} g_{\mathrm{L}} &= \bar{T}_{\mathrm{R}}\ K_1^{-1}, \\ g_{\mathrm{R}} &= T_{\mathrm{R}}\ K_{-1}^{-1}. \end{aligned} \tag{5.47}$$

These equations unfortunately are only defined if K_1 and K_{-1} are not singular. The same problem however does not affect the left and right SE, since they simply become

$$\Sigma_{\mathrm{L}} = K_{-1} \bar{T}_{\mathrm{R}}, \tag{5.48}$$

$$\Sigma_{\mathrm{R}} = K_1\ T_{\mathrm{R}}. \tag{5.49}$$

In complete analogy the same expressions obtained by using the left transfer matrices are $\Sigma_{\mathrm{L}} = T_{\mathrm{L}} K_1$ and $\Sigma_{\mathrm{R}} = \bar{T}_{\mathrm{L}} K_{-1}$. This result is equivalent to those obtained in Refs. [212, 213, 216, 215, 197, 65] and derived with different approaches, demonstrating the equivalence of those to our semi-analytical formula. Since NEGF-based transport codes simply require Σ_{L} and Σ_{R}, our scheme allows the calculations of system with arbitrarily complicated electronic structure. A schematic tree diagram describing the steps involved in obtaining the SE is shown in Fig. 5.2 ("basic algorithm").

Eqs. (5.48) and (5.49) demonstrate that the SE can be calculated directly, without explicitly calculating g_{L} and g_{R}. In situations where also the SGFs are needed, these can be obtained by using the relations

$$g_{\mathrm{L}} = -\left[K_0 + \Sigma_{\mathrm{L}}\right]^{-1}, \tag{5.50}$$

$$g_{\mathrm{R}} = -\left[K_0 + \Sigma_{\mathrm{R}}\right]^{-1}. \tag{5.51}$$

This can be derived by adding one layer to the left and one to the right surfaces respectively [210]. In Appendix C.1 we show that the SEs calculated with Eqs. (5.48) and (5.49) indeed fulfill the above equations. Moreover with the use of Eqs. (5.47) and (5.50) we can now regularize Eq. (5.27) also for the case where T_{R} is singular by writing it as

$$g_{00}^{-1} = -K_0 - \Sigma_{\mathrm{L}} - \Sigma_{\mathrm{R}}. \tag{5.52}$$

We have therefore a scheme where the SEs are identified as the principal quantities, whereas the SGF and g_{00} are derived from these.

When we compare the method of Ref. [214] with the equations derived above, we notice that now it is not necessary to calculate the matrix g_{00} and its inverse using Eq. (5.27) in order to obtain the SE. Since this is not defined in the case of singular K_1 and K_{-1}, we expect the method presented here to be more stable and accurate. Also the problems caused close to band edges by the Van Hove singularities in g_{00} are avoided, since these singularities do not appear in the SGF and SE. Moreover the method in Ref. [214] relies on the calculation of the SGF in order to obtain the SE, whereas here the SGF is not needed. As we will show in Sec. 5.5, close to surface states the error in the SGF is much larger than the one for the SE, so that we also expect a large improvement in the accuracy for energies where surface states are found.

5.3 Reducing the condition number of K_1 and K_{-1}

The accuracy with which the SEs are calculated depends on the accuracy involved in solving Eq. (5.6), a quadratic eigenvalue problem extensively studied in the past [220, 221]. However most solution methods have problems if K_1 or K_{-1} are close to being singular, or more generally if their condition number κ is large. In this case some of the complex eigenvalues tend to infinity and others to zero at the same time, and this results in a loss of accuracy in numerical computations. When calculating T_{R} ($\bar{T}_{\mathrm{R}}$) however the contributions from the states with $\mathrm{Im}(k_n) \to \infty$ [$\mathrm{Im}(\bar{k}_n) \to -\infty$] are vanishingly small. It is therefore useful to limit the range of the eigenvalues $\{e^{ik_n}\}$

in such a way that the important eigenstates with small $|\mathrm{Im}(k_n)|$ and $|\mathrm{Im}(\bar{k}_n)|$ can be calculated accurately, while losing precision for the less important eigenstates with large $|\mathrm{Im}(k_n)|$ and $|\mathrm{Im}(\bar{k}_n)|$. In this section we show how this can be achieved by decreasing $\kappa(K_1)$ and $\kappa(K_{-1})$. Here we assume that $K_1 = K_{-1}^\dagger$, which is the case for real E, so that $\kappa(K_1) = \kappa(K_{-1})$. Minor modifications are needed for the general case (see Appendix C.2).

In order to obtain $\kappa(K_1)$ first a SVD of the matrix is performed:

$$K_1 = USV^\dagger. \tag{5.53}$$

Here, U and V are unitary matrices, and S is a diagonal matrix, whose diagonal elements s_n are the singular values. These are real and positive, and ordered so that $s_{n+1} \leq s_n$. If $s_{\max}$ is the largest singular value and $s_{\min}$ the smallest one, then the condition number is defined as $\kappa(K_1) = s_{\max}/s_{\min}$, with K_1 singular if $s_{\min}$ is zero $[\kappa(K_1) \to \infty]$.

We now replace S by an approximate S_{SVD}, whose diagonal elements $s_{\mathrm{SVD},n}$ are

$$s_{\mathrm{SVD},n} = \begin{cases} s_n & s_n \geq s_{\max}\ \delta_{\mathrm{SVD}} \\ s_{\mathrm{SVD}} & s_n < s_{\max}\ \delta_{\mathrm{SVD}} \end{cases}, \tag{5.54}$$

and accordingly K_1 with $K_{1,\mathrm{SVD}} = US_{\mathrm{SVD}}V^\dagger$. The tolerance parameter δ_{SVD} is a real positive number that determines the condition number of $K_{1,\mathrm{SVD}}$.

We now present two possible choices for s_{SVD}. The first is to set $s_{\mathrm{SVD}} = 0$, resulting in $K_{1,\mathrm{SVD}}$ being singular. We can then perform a unitary transformation in order to eliminate the degrees of freedom associated to $s_{\mathrm{SVD},n} = 0$, and obtain an effective K_1 matrix (K_1^{eff}) with reduced size for which $\kappa(K_1^{\mathrm{eff}}) \leq \delta_{\mathrm{SVD}}^{-1}$. The second possibility is to set $s_{\mathrm{SVD}} = s_{\max}\ \delta_{\mathrm{SVD}}$, so that by definition we have $\kappa(K_{1,\mathrm{SVD}}) \leq \delta_{\mathrm{SVD}}^{-1}$. The accuracy obtained with both strategies is similar, the advantage of using the first however is that the size of the matrices is reduced, so that for big systems the computation is much faster. In our implementation we use both methods together, first we reduce the size of the system by setting $s_{\mathrm{SVD}} = 0$, and then, if necessary, we further reduce the condition number for the effective system by limiting the smallest singular value.

5.3.1 Reduction of system size

Here we set all the M singular values s_n smaller than $s_{\max}\delta_{\mathrm{SVD}}$ to zero, so that there are $N_{\mathrm{eff}} = N - M$ singular values s_n with $s_n \geq s_{\max}\delta_{\mathrm{SVD}}$. The transformations needed in order to obtain the right SE are now presented (the procedure for the left SE is analogous). We apply the unitary transformation $K'_{zz'} = U^\dagger K_{zz'} U$, $\phi'_{\mathrm{R},n} = U^\dagger \phi_{\mathrm{R},n}$,

and we define $K'_1 = U^\dagger K_{1,\mathrm{SVD}} U$, $K'_{-1} = U^\dagger K_{-1,\mathrm{SVD}} U$, $K'_0 = U^\dagger K_0 U$. Since M singular values of $K_{1,\mathrm{SVD}}$ are zero the transformed matrices have the structure

$$\begin{aligned} K'_1 &= \begin{pmatrix} K_{1,\mathrm{c}} & K_{1,\mathrm{u}} \\ \mathbb{0}_{M,N_{\mathrm{eff}}} & \mathbb{0}_{M,M} \end{pmatrix}, \quad K'_0 = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \\ K'_{-1} &= \begin{pmatrix} K_{-1,\mathrm{c}} & \mathbb{0}_{N_{\mathrm{eff}},M} \\ K_{-1,\mathrm{u}} & \mathbb{0}_{M,M} \end{pmatrix}, \quad \phi'_{\mathrm{R},n} = \begin{pmatrix} \phi_{\mathrm{c},n} \\ \phi_{\mathrm{u},n} \end{pmatrix}, \end{aligned} \tag{5.55}$$

where the dimensions of the new matrices are: $N_{\mathrm{eff}} \times N_{\mathrm{eff}}$ for $K_{1,\mathrm{c}}$, $K_{-1,\mathrm{c}}$ and A, $N_{\mathrm{eff}} \times M$ for $K_{1,\mathrm{u}}$ and B, $M \times N_{\mathrm{eff}}$ for $K_{-1,\mathrm{u}}$ and C, and $M \times M$ for D. Finally $\phi_{\mathrm{c},n}$ is a column vector of dimension N_{eff}, and $\phi_{\mathrm{u},n}$ is of dimension M. The transformed form of Eq. (5.4) is

$$\left(K'_0 + K'_1 e^{ik_n} + K'_{-1} e^{-ik_n}\right) \phi'_{\mathrm{R},n} = 0. \tag{5.56}$$

Due to the structure of K'_{-1} there are M solutions to this equation with $e^{ik_n} = 0$ and $\phi_{\mathrm{c},n} = 0$. We therefore split up the right-going states into those with finite $e^{ik_n} \neq 0$ and those with $e^{ik_n} = 0$. For the first set, from Eq. (5.56), we obtain

$$\phi_{\mathrm{u},n} = F_n \phi_{\mathrm{c},n}, \tag{5.57}$$

with

$$F_n = -D^{-1}\left(K_{-1,\mathrm{u}} e^{-ik_n} + C\right). \tag{5.58}$$

The $\phi_{\mathrm{c},n}$ are then solutions of an effective system with reduced size

$$\left(K_0^{\mathrm{eff}} + K_1^{\mathrm{eff}} e^{ik_n} + K_{-1}^{\mathrm{eff}} e^{-ik_n}\right) \phi_{\mathrm{c},n} = 0, \tag{5.59}$$

where the effective matrices are

$$\begin{aligned} K_1^{\mathrm{eff}} &= K_{1,\mathrm{c}} - K_{1,\mathrm{u}} D^{-1} C, \\ K_{-1}^{\mathrm{eff}} &= K_{-1,\mathrm{c}} - B D^{-1} K_{-1,\mathrm{u}}, \\ K_0^{\mathrm{eff}} &= A - B D^{-1} C - K_{1,\mathrm{u}} D^{-1} K_{-1,\mathrm{u}}. \end{aligned} \tag{5.60}$$

We can now solve the quadratic eigenvalue problem [Eq. (5.6)] for this effective system to get the set of N_{eff} eigenvectors $Q_{\mathrm{c}} = \begin{pmatrix} \phi_{\mathrm{c},1} & \phi_{\mathrm{c},2} & \dots & \phi_{\mathrm{c},N_{\mathrm{eff}}} \end{pmatrix}$ and eigenvalues $\{e^{ik_n}\}$ for the right-going states. The M eigenvectors of the second set of solutions with $e^{ik_n} = 0$ are given by $\phi_{\mathrm{c},n} = 0$ with a general $\phi_{\mathrm{u},n}$. The set of eigenvectors of the full K' matrix therefore is

$$Q = \begin{pmatrix} Q_{\mathrm{c}} & \mathbb{0}_{N_{\mathrm{eff}},M} \\ Q_{\mathrm{u}} & Q_0 \end{pmatrix}, \tag{5.61}$$

with $Q_{\mathrm{u}} = \begin{pmatrix} F_1\phi_{\mathrm{c},1} & F_2\phi_{\mathrm{c},2} & \dots & F_{N_{\mathrm{eff}}}\phi_{\mathrm{c},N_{\mathrm{eff}}} \end{pmatrix}$, and Q_0 is a general matrix of solution vectors for the states with $e^{ik_n} = 0$. From this we obtain the set of duals

$$Q^{-1} = \begin{pmatrix} Q_{\mathrm{c}}^{-1} & \mathbb{0}_{N_{\mathrm{eff}},M} \\ -Q_0^{-1} Q_{\mathrm{u}} Q_{\mathrm{c}}^{-1} & Q_0^{-1} \end{pmatrix}. \tag{5.62}$$

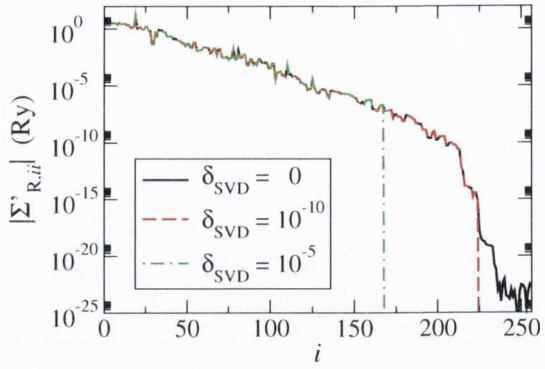


Figure 5.3: Absolute value $|\Sigma'_{\mathrm{R},ii}|$ of the diagonal elements of the transformed right SE for different values of δ_{SVD}.

Using these results we can now calculate the transfer matrix T'_{R} of the transformed system

$$T'_{\mathrm{R}} = \sum_{n=1}^{N_{\mathrm{eff}}} e^{ik_n} \begin{pmatrix} \phi_{\mathrm{c},n}\tilde{\phi}^{\dagger}_{\mathrm{c},n} & \mathbb{0}_{N_{\mathrm{eff}},M} \\ F_n\phi_{\mathrm{c},n}\tilde{\phi}^{\dagger}_{\mathrm{c},n} & \mathbb{0}_{M,M} \end{pmatrix}, \tag{5.63}$$

where we have also used the fact that $e^{ik_n} = 0$ for the second set of solutions. We note that setting the M smallest singular values s_n to zero causes the last M columns of T'_{R} to be zero too. Moreover the explicit calculation of Q_0 is not needed in order to obtain T'_{R}. From this and Eq. (5.49) we obtain the right SE

$$\Sigma'_{\mathrm{R}} = \begin{pmatrix} \Sigma^{\mathrm{eff}}_{\mathrm{R}} - K_{1,\mathrm{u}}D^{-1}K_{-1,\mathrm{u}} & \mathbb{0}_{N_{\mathrm{eff}},M} \\ \mathbb{0}_{M,N_{\mathrm{eff}}} & \mathbb{0}_{M,M} \end{pmatrix}, \tag{5.64}$$

where

$$\Sigma^{\mathrm{eff}}_{\mathrm{R}} = K_1^{\mathrm{eff}} \sum_{n=1}^{N^{\mathrm{eff}}} e^{ik_n}\phi_{\mathrm{c},n}\tilde{\phi}^{\dagger}_{\mathrm{c},n} \tag{5.65}$$

is the SE of the effective system.

The structure of Σ'_R shows that by applying this unitary transformation we have ordered the elements of the SE by absolute size, moving those columns (rows) with the smallest values to the right (bottom). By setting the smallest singular values of K_1 to zero those columns and rows of the SE with small values have also been set to zero. This is illustrated in Fig. 5.3, where the absolute value of the diagonal elements of the transformed self-energy $|\Sigma'_{\mathrm{R},ii}|$ is shown for a (8,0) zigzag carbon nanotube at the Fermi energy E_{F} (see Sec. 5.4 for a detailed description of the system). The $|\Sigma'_{\mathrm{R},ii}|$ are basically identical for different δ_{SVD} up to $i = N_{\mathrm{eff}}$, and indeed by increasing the value of δ_{SVD} more diagonal elements of Σ'_R are set to zero. We note that N_{eff} is of similar size as N in Fig. 5.3, since the system is rather short along z and a small basis set is used (i.e. N is small). For large systems and rich basis sets the ratio

$N_{\rm eff}/N$ will decrease. The physical interpretation of the zero columns and rows in the SE is that the M states with $k_n \to \infty$ decay infinitely fast, so that the interaction of those states is limited to the site they are localized at. Finally the SE of the original system can be obtained by applying the inverse unitary transformation

$$\Sigma_{\rm R} = U\Sigma'_{\rm R}U^\dagger, \tag{5.66}$$

and in contrast to $\Sigma'_{\rm R}$ the matrix $\Sigma_{\rm R}$ is a dense $N \times N$ matrix.

Note that in order to obtain the left SE we perform the unitary transformation $K'_{zz'} = V^\dagger K_{zz'}V$, $\phi'_{{\rm R},n} = V\phi'_{{\rm R},n}$, and then follow an analogous procedure. In this case however instead of the right-going states the left-going ones are used.

5.3.2 Limiting the smallest singular value

We can limit the lower bound of the singular values s_n by setting $s_{\rm SVD} = s_{\rm max}\delta_{\rm SVD}$ in Eq. (5.54). In this case the approximated K matrix is obtained by replacing K_1 with $K_{1,{\rm SVD}}$. The error introduced is now of the order of $s_{\rm max}\ \delta_{\rm SVD}$. Ideally $s_{\rm max}\ \delta_{\rm SVD}$ should be of the order of the machine numerical precision, so that the error is minimal. However sometimes increasing $s_{\rm max}\ \delta_{\rm SVD}$ beyond that value improves the results, therefore $\delta_{\rm SVD}$ is left as a parameter to adjust depending on the material system investigated. This will be discussed extensively in the next section.

A simpler but equally effective possibility for limiting the smallest singular value of a matrix is that of adding a small random perturbation [225, 217]. Thus another strategy for reducing the condition number of K_1 is that of replacing K_1 with $K_{1,{\rm noise}} = K_1 + W(w_{\rm noise})$, where $W(w_{\rm noise})$ is a matrix whose elements are random complex numbers with an average absolute value $|W_{ij}| \sim w_{\rm noise}$. In particular we choose the $|W_{ij}|$ in such a way that both ${\rm Re}(W_{ij})$ and ${\rm Im}(W_{ij})$ are random numbers in the range $[-w_{\rm noise}, w_{\rm noise}]$. We find that if $w_{\rm noise} = s_{\rm max}\ \delta_{\rm SVD}$ the addition of noise usually gives results as accurate as those obtained with the SVD procedure, but the calculation is faster since instead of performing a SVD we just perform a sum of the matrices.

In Fig. 5.2 we present our final extended algorithm as it has been implemented in SMEAGOL. This now includes the following regularization procedure of K_1. First the size of K_1, and hence of the whole problem, is reduced by using the scheme described in Sec. 5.3.1, with a tolerance parameter $\delta_{\rm SVD} = \delta_{{\rm SVD},1}$. This generates an effective matrix $K_1^{\rm eff}$ whose condition number $\kappa(K_1^{\rm eff})$ is reduced by adding a small noise matrix $W(w_{\rm noise})$. Such a step is extremely fast and enhances considerably the numerical stability of the calculation. In most cases the SE for the effective system

can then be calculated and no further regularization steps are needed. However, in some cases the calculation of the SE still fails. This, for example, happens when the solution of Eq. (5.6) for the effective system fails, or else when the calculated number of left-going states erroneously differs from the number of right-going states. In these critical situations we further decrease $\kappa(K_1^{\text{eff}})$ by limiting the smallest singular value of K_1^{eff} as described in Sec. 5.3.2 with a tolerance parameter $\delta_{\text{SVD}} = \delta_{\text{SVD},2}$. SMEAGOL automatically adjusts $\delta_{\text{SVD},1}$, $\delta_{\text{SVD},2}$ and w_{noise} within a given range until the SE is calculated. In our test calculations for a number of different systems we found no situation where such a scheme has failed. In contrast when the standard algorithm of Ref. [214] is employed the number of failures was considerable. Note that our extended algorithm can also be used in conjunction with recursive methods for evaluating the SE [210, 211, 212, 213]. Also in this case it will decrease the computing time for large systems due to the reduced size of the effective K matrix.

5.4 Error analysis

When recursive algorithms are used the accuracy of the SE is automatically known as it coincides with the convergence criterion. Poor convergence is found when the error can not be reduced below a given tolerance. Direct methods, as the one presented here, are in principle error free in the sense that when the solution is found, this is in principle exact. For this reason the numerical errors arising from semi-analytical schemes usually are not estimated. In this section we perform this estimate and present a detailed error analysis for three different material systems.

In order to estimate the numerical accuracy we use the recursive relations of Eqs. (5.50) and (5.51), written as

$$\begin{aligned} \Sigma_{\text{L}}^{\text{out}} &= -K_{-1}\left[K_0 + \Sigma_{\text{L}}^{\text{in}}\right]^{-1} K_1 \\ \Sigma_{\text{R}}^{\text{out}} &= -K_1\ \left[K_0 + \Sigma_{\text{R}}^{\text{in}}\right]^{-1} K_{-1}, \end{aligned} \tag{5.67}$$

where $\Sigma_{\{\text{L/R}\}}^{\text{in}}$ are calculated with our extended algorithm, and $\Sigma_{\{\text{L/R}\}}^{\text{out}}$ are obtained by evaluating the right-hand side term of the above equations. When the solution is exact then $\Sigma_{\text{L}}^{\text{out}} = \Sigma_{\text{L}}^{\text{in}}$ and $\Sigma_{\text{R}}^{\text{out}} = \Sigma_{\text{R}}^{\text{in}}$. Therefore we can define a measure of the error Δ_Σ as

$$\Delta_\Sigma = \left|\left|\Sigma_{\{\text{L/R}\}}^{\text{out}} - \Sigma_{\{\text{L/R}\}}^{\text{in}}\right|\right|_{\text{max}}, \tag{5.68}$$

where $||\ldots||_{\text{max}}$ stands for the maximum norm [218], the corresponding relative error is $\Delta_{\Sigma,\text{r}} = \Delta_\Sigma / \left|\left|\Sigma_{\{\text{L/R}\}}\right|\right|_{\text{max}}$. The accuracy criterion used in the extended algorithm is the following. We first set $\delta_{\text{SVD},1}$, w_{noise} and eventually $\delta_{\text{SVD},2}$ and compute $\Delta_{\Sigma,\text{r}}$. This should be lower than a target accuracy $\Delta_{\Sigma,\text{r}}^{\text{max}}$. If this is not the case then the SE

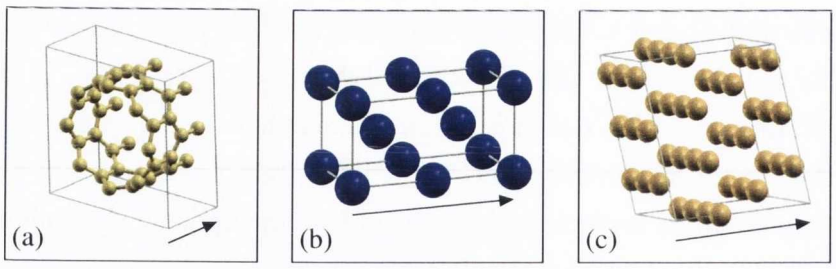


Figure 5.4: Unit cells of the three systems investigated in this work: (a) (8,0) zigzag carbon nanotube, (b) bcc Fe oriented along the (100) direction, and (c) fcc Au oriented along the (111) direction. The black arrow indicates the direction of the stacking z, i.e. the direction of the transport.

are recalculated with a different set of tolerance parameters, until $\Delta_{\Sigma,\mathrm{r}}$ reaches the desired accuracy. If this condition is never achieved the final SE is the one with the smallest $\Delta_{\Sigma,\mathrm{r}}$.

We now calculate the SE for different variations of the method, chosen in order to highlight the problems arising from ill conditioned K_1 and K_{-1} and to show the difference between the basic method of Ref. [214] and the extensions presented here. There are two main differences between the two methods. The first is that here we solve Eq. (5.6) without inverting K_1, whereas in Ref. [214] K_1^{-1} is used to solve the inverse band-structure relation $k = k(E)$. Clearly this second choice is less accurate if K_1 is close to singular. However it is much faster computationally, so that it might be of advantage for big systems. The second difference is that here it is not necessary to calculate g_{00} via Eq. (5.27), so that one does not need to invert T_R and $\bar{T}_\mathrm{R}$.

In order to investigate the effect of these two aspects independently, we have calculated the SE using the following four methods. In method 1 we use the algorithm presented in this work. In particular we use Eq. (5.6) to solve the quadratic eigenvalue problem and Eqs. (5.48) and (5.49) to obtain the SE [for the right SE we actually use a different form of Eq. (5.6); see Appendix C.3]. Method 2 is essentially the same, with the only difference that instead of solving Eq. (5.6) we use the eigenvalue method of Ref. [214]. In method 3 we solve Eq. (5.6), but we use Eq. (5.46) to calculate the SGF, from which the SEs are obtained via Eqs. (5.44) and (5.45), and g_{00} is calculated from Eq. (5.27). Finally method 4 is the algorithm of Ref. [214].

In order to obtain a statistically significant average of the errors, we plot a histogram of the calculated errors for both Σ_L and Σ_R for a large energy range. Here we use the absolute error, since it can readily be compared to the energy scale of the problem. Note that although the relative error might be small, the absolute error can be very large if $\left|\left|\Sigma_{\{\mathrm{L/R}\}}\right|\right|_{\max} \gg 1$ Ry. Furthermore, in order to keep the analysis

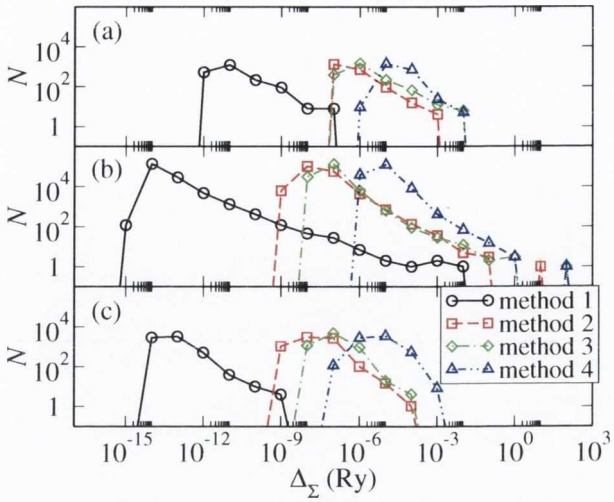


Figure 5.5: Histogram of the errors in the calculation of the self-energy Δ_Σ for three different systems. (a) (8,0) zigzag carbon nanotube, (b) bcc Fe, and (c) fcc Au. N is the number of times a given error Δ_Σ occurs (not normalized).

simple, in all the calculations of this section we do not reduce the system size nor do we add noise ($w_{\text{noise}} = 0$). We regularize K_1 and K_{-1} by using $s_{\text{SVD}} = s_{\max}\, \delta_{\text{SVD}}$ in Eq. (5.54). Since the error depends on the chosen δ_{SVD}, here we calculate Δ_Σ for a set of δ_{SVD} in the range $[0, 10^{-23}, 10^{-22}, \ldots, 10^{-4}, 10^{-3}]$. We then present the smallest Δ_Σ found for δ_{SVD} taken in that range. This is the smallest possible error achievable with a given method and allows us to extract informations on the range of optimal SVD values for a given method.

As first example a (8,0) zigzag carbon nanotube [226] is presented [the unit cell is shown in Fig. 5.4(a)]. The length of the periodic unit cell is 4.26 Å along the nanotube, with 32 carbon atoms in the unit cell. The LDA approximation (no spin-polarization) is used for the exchange correlation potential. We consider $2s$ and $2p$ orbitals for carbon with double ζ and a cutoff radius r_c for the first ζ of $r_c = 5$ bohr. Higher ζ are constructed with the split-norm scheme with a split-norm of 15% [69]. The real-space mesh cutoff is 200 Ry. The matrices H_0, H_1, S_0 and S_1 are extracted from a ground- state SIESTA DFT calculation for an infinite periodic nanotube (see Sec. 4.6). We calculate the SE for the semi-infinite nanotube at 1024 energy points in a range of ±5 eV around the Fermi energy.

Fig. 5.5(a) shows the histogram of the errors in the SE, where N is the number of times a given error Δ_Σ appears. In general the figure shows that for this system the average error increases when going from method 1 to method 2 and method 3, and finally to method 4. The error obtained with method 1 is on average about 6 orders of magnitude smaller than the one obtained with method 4. The main reason behind

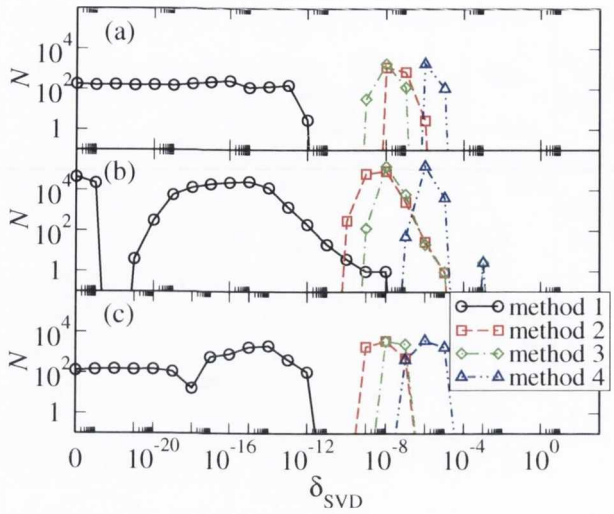


Figure 5.6: Histogram of δ_{SVD} giving the smallest error in the self-energy. (a) (8,0) zigzag carbon nanotube, (b) bcc Fe, and (c) fcc Au. N is the number of times a given δ_{SVD} generates the smallest error (not normalized).

this dramatically improved accuracy is that method 1 does not involve any steps where a singular K_1 leads to divergences. Method 4 on the other hand is strongly dependent on the condition number of K_1, since it necessitates to invert K_1 and T_{R} (or $\bar{T}_{\mathrm{R}}$). Methods 2 and 3 are on average about one order of magnitude more precise than method 4. Since they both still involve one of the two inversions, the difference is however not large.

Fig. 5.6(a) shows the histogram of the optimum δ_{SVD} used for the calculations of the SE. Here we plot the number of times N a particular δ_{SVD} has given the smallest error in the set of calculations. A larger optimal value for δ_{SVD} indicates a stronger dependence of the computational scheme on $\kappa(K_1)$. For method 1 the range of used δ_{SVD} is smaller than 10^{-12}. If we force δ_{SVD} to be zero we get almost the same accuracy as the one shown in Fig. 5.5(a), which confirms that the accuracy of for method 1 depends little on $\kappa(K_1)$ for this system. However also for this method there is a set of energies (a few percent of the total number) where the solution of Eq. (5.6) fails if δ_{SVD} is too small. The optimal δ_{SVD} for the other methods is orders of magnitude larger than that of method 1, and it is never smaller than 10^{-9}. The absolute error induced by replacing K_1 by $K_{1,\mathrm{SVD}}$ is of the order of $\delta_{\mathrm{SVD}}\, s_{\max}$. Usually $s_{\max}$ is of the order of 1 Ry, so that the error is of the order of δ_{SVD} Ry. Therefore since in methods 2–4 a large value of δ_{SVD} is needed in order to improve $\kappa(K_{1,\mathrm{SVD}})$, also the resulting error is large.

The second example is bcc Fe [Fig. 5.4(b)], oriented along the (100) direction. The lattice parameters are the same as in Ref. [20]. There are four Fe atoms in the

unit cell. We apply periodic boundary conditions in the direction perpendicular to the stacking, so that these correspond to four Fe planes. The length of the cell along the stacking direction is 5.732 Å. A double ζ s (r_c=5.6 bohr), single ζ p (r_c=5.6 bohr) and single ζ d (r_c=5.2 bohr) basis is used. The real-space mesh cutoff is 600 Ry, and the DFT calculation is converged for 7x7 k-points in the Brillouin zone orthogonal to the stacking. The SEs have been calculated for the converged DFT calculation at 32 different energies in a range of ± 1 eV around the Fermi energy, and for 10,000 k-points in the 2D Brillouin zone perpendicular to the stacking direction. For each k-point there is a different set of matrices K_0, K_1 and K_{-1}, so that for each k-point there is a different SE. The histogram for the error of the calculated self-energy Δ_Σ is shown in Fig. 5.5(b), and the histogram for the optimal δ_{SVD} in Fig. 5.6(b). The general behavior is similar to the one found for the carbon nanotube. We note that, although for the vast majority of the calculations the error in the SE is small, there is a long tail in the histograms of Fig. 5.5(b) indicating the presence of a small number of large errors. This is present for all the methods, with maximum errors of 10^{-2} Ry for method 1, and 100 Ry for method 4. Closer inspection shows that the reason for the increase of the error for certain energies and k-points is caused by a divergence in $\left|\left|\Sigma_{\{\mathrm{L,R}\}}\right|\right|_{\mathrm{max}}$. This will be illustrated in more detail in the next section.

Finally we consider fcc Au [Fig. 5.4(c)], with the stacking along the (111) direction. The unit cell consists of three planes of nine gold atoms each. These are the typical leads used for the calculations of the transmission properties of molecules attached to gold [227, 228, 202, 93]. We use double ζ s (r_c=6.0 bohr) and single ζ d (r_c=5.5 bohr) and four k-points in the Brillouin zone perpendicular to the stacking. The mesh cutoff is 400 Ry. The SEs have been calculated for 418 energy points, from about 15 eV below to about 10 eV above the Fermi energy. The general behavior [Figs. 5.5(c) and 5.6(c)] is again similar to that of the previous examples. Also here the error for method 1 is about 6 orders of magnitude smaller than that of method 4, with methods 2 and 3 giving some marginal improvement.

Our results show that the scheme presented here in general allows the calculation of the SE with high accuracy. The main advantage of method 1 is rooted in the possibility of using a much smaller δ_{SVD}. For big systems sometimes one might prefer to use method 2, since it is considerably faster than method 1 and gives the second best accuracy. In this case we first calculate the SE with method 2 and check the error. Only for those energy points where the error is above some maximum value (of the order of 10^{-5} Ry, for example) the calculation is repeated with method 1 to improve the accuracy. Finally the results show that for all methods the SVD transformation of K_1 is necessary, although for method 1 it is needed only a fraction

of the times. For big systems, in particular if the unit cell is elongated along the stacking direction, or if a rich basis set is used, $\kappa(K_1)$ will generally increase as there will be some singular values of K_1 going to zero. In these cases also method 1 will require a SVD transformation for most energies. The range of δ_{SVD} should however be similar to the one shown in Fig. 5.6, so that also the error in the SE should be of the same order of magnitude. We also note that in order to keep the analysis simpler here we have not used the reduction of system size described in Sec. 5.3.1, for such large systems it is however crucial in order to decrease the computational effort and regularize K_1 at the same time.

5.5 Surface states

The center of the error distribution for method 1 (Fig. 5.5) is located at small Δ_Σ, usually smaller than 10^{-11} Ry. However the histogram has also a tail reaching up to very large errors. These are found only at some critical energies, as demonstrated in Fig. 5.7(a), where we show Δ_Σ for the carbon nanotube calculated over 1024 energy points, in a range of 2 eV around the Fermi energy. The average error is of the order of 10^{-12} Ry, but at energies around -0.8 eV and -0.34 eV the error drastically increases. Indeed a finer energy mesh at these points suggests a divergence. The origin of the large errors at particular energies can be investigated by looking at the eigenvalues $g_{\text{L},i}$ of the SGF g_{L}. In Fig. 5.7(b) the largest and the smallest absolute values for the eigenvalues, respectively $g_{\text{L,max}}$ and $g_{\text{L,min}}$, are plotted as a function of energy ($g_{\text{L,min}} \leq |g_{\text{L},i}| \leq g_{\text{L,max}}$). It can be seen that $g_{\text{L,max}}$ diverges close to the energies where the error increases, i.e. we can associate large errors in g_{L} with a divergence in its spectrum. Since Σ_{L} is calculated from Eq. (5.48), the only possible origin for the divergence is in the norm of some of the $\tilde{\bar{\phi}}_{\text{R},n}$. As these are obtained by inverting the matrix $\bar{Q} = \begin{pmatrix} \bar{\phi}_{\text{R},1} & \bar{\phi}_{\text{R},2} & \dots & \bar{\phi}_{\text{R},N} \end{pmatrix}$ [Eq. (5.19)], one deduces that the set of vectors $\{\bar{\phi}_{\text{R},n}\}$ is not linearly independent. For these energies $\kappa(\bar{Q}) \to \infty$. We therefore can simply check the magnitude of $\kappa(\bar{Q})$ to determine whether there is a divergence of the SE close to a particular energy.

Physically the divergence of the SE translates into the presence of a surface state [212, 210] at that particular energy. Consider the spectral representation of g_{L}

$$g_{\text{L}}(E) = \sum_{n=1}^{N} \frac{1}{E + i\delta - E_n} \psi_n \tilde{\psi}_n^\dagger, \tag{5.69}$$

where E_n are the eigenvalues, ψ_n are the right eigenvectors of the effective surface Hamiltonian matrix $H_0 - \Sigma_{\text{L}}$ with overlap S_0, and $\tilde{\psi}_n$ are the left eigenvectors of the

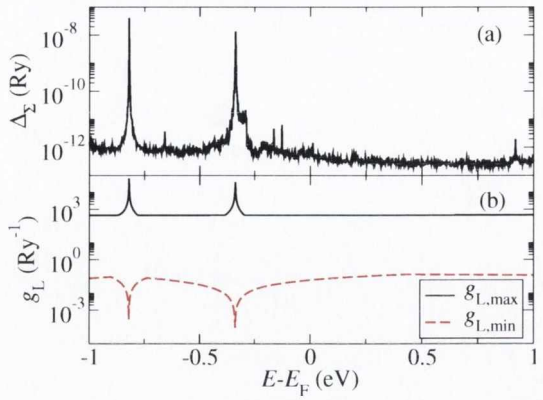


Figure 5.7: The error analysis for the carbon nanotube of Fig. 5.4: (a) absolute error Δ_Σ of the self-energy as a function of the energy E, (b) maximum ($g_{\mathrm{L,max}}$) and minimum ($g_{\mathrm{L,min}}$) eigenvalues of g_{L}.

same Hamiltonian. A localized surface state is found when there is a real eigenvalue $E_n(E)$ at $E_n(E) = E$ [or more generally if $\mathrm{Im}(E_n(E))$ is very small].

From the recursive relation (5.50) one can deduce that for an infinite eigenvalue there is also a corresponding vanishing eigenvalue. Therefore in Fig. 5.7(b) for energies where $g_{\mathrm{L,max}} \to \infty$ we have also $g_{\mathrm{L,min}} \to 0$. Close to the singularity we can therefore expand the two eigenvalues as $g_{\mathrm{L,max}} \propto \frac{1}{E+i\delta-E_n}$ and $g_{\mathrm{L,min}} \propto E + i\delta - E_n$. For $E = E_n$ the largest eigenvalue in Eq. (5.69) is then equal to δ^{-1}, and the smallest is equal to δ. To avoid divergence therefore the magnitude of the g_{L} eigenvalues can be bounded to a finite value δ^{-1} by introducing a small imaginary part to the energy for energies in the vicinity of a surface state.

Another possibility for limiting the size of $g_{\mathrm{L,max}}$ is to bound the singular values of $\bar{Q}$ from below in the same way as it is done for K_1 (Sec. 5.3.2). This essentially imposes the $\bar{\phi}_{\mathrm{R},n}$ to be linearly independent from each other. However, with this scheme it is not possible to conserve the Green's function causality, so that the SGF might have eigenvalues lying on the positive imaginary axis. Moreover we loose control over the accuracy of the computed SGF and SE. Both these problems are avoided when using a finite δ.

We now investigate the DOS and transport properties of a system when the finite imaginary part δ (broadening) is added to the energy. We consider as an example the carbon nanotube of Fig. 5.4(a). In Fig. 5.8(a) the onsite surface DOS $\tilde{\mathcal{N}}_0$, as defined in Eq. (5.38), is shown for $\delta = 0$ Ry, $\delta = 10^{-6}$ Ry, $\delta = 10^{-5}$ Ry and $\delta = 10^{-4}$ Ry. For $\delta = 0$ the surface DOS vanishes for energies between -0.42 eV and +0.39 eV, indicating the presence of a gap around the Fermi energy. Note that there are no Van Hove singularities in $\tilde{\mathcal{N}}_0$, since we never divide by the group velocity when calculating

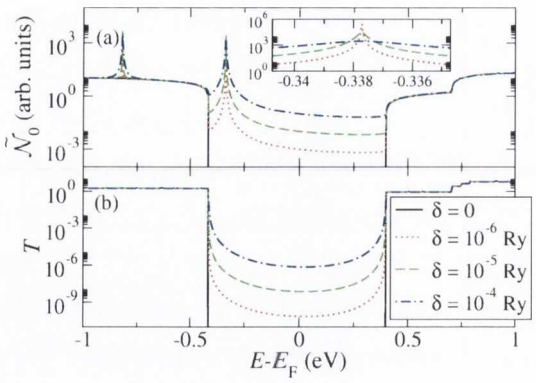


Figure 5.8: Density of states and transmission coefficient for the carbon nanotube of Fig. 5.4. (a) Density of states of the surface layer $\tilde{\mathcal{N}}_0$ as a function of energy E, calculated for different broadenings δ. The inset is a zoom at energies around -0.34 eV. (b) Transmission coefficient T for different values of δ.

the SGF. For finite δ and energies away from the band gap, the DOS is essentially identical to that calculated for $\delta = 0$, however inside the gap $\tilde{\mathcal{N}}_0$ does not vanish but saturates to a small value proportional to δ^{-1}. Moreover whereas the surface states are not visible for $\delta = 0$, they appear in the DOS for finite δ, and their full width at half maximum (FWHM) equals 2δ.

We then move to the transport by calculating the transmission coefficient $T(E)$ [Eq. (4.94)] for a carbon nanotube attached to semi-infinite leads made from an identical carbon nanotube. Since this is a periodic system $T(E)$ must equal the number of open channels, so that it can only have integer values. This is indeed the case for $\delta = 0$ [Fig. 5.8(b)]. For finite δs the transmission coefficient is only approximately an integer, especially inside the energy gap region, where the finite surface DOS introduced by δ leads to a non-zero transmission. The transmission in the gap is proportional to δ^2 (note that the scale is logarithmic), since on both sides of the scattering region the artificial surface DOS is proportional to δ. In this region of small transmission therefore the results might change by orders of magnitude depending on the value of δ. For all values of δ however we find no contributions to the transmission coming from the surface state, indicating that these do not carry current. These results show that adding a finite value δ to the energy has little effect on the actual transmission if this is large. However when the transmission is small, as in the case of tunnel junctions, the finite δ introduces an additional contribution to the conduction that might arbitrarily affect the results. It is thus imperative for those systems to identify surface states and use the imaginary δ only in a narrow energy interval around them.

Finally we can give an estimate of the relative accuracy $\Delta_{\Sigma,\mathrm{r}}(\delta) = \Delta_\Sigma / \left|\left|\Sigma\right|\right|_{\max}$ at

the energy corresponding to the surface state. As discussed before the origin of the error is the inversion of $\bar{Q}$, needed in order to calculate the duals. The relative error introduced by the inversion of $\bar{Q}$ is proportional to $\kappa(\bar{Q})$ [229, 230, 225, 217, 218]. Close to a surface state the smallest singular value is of the order of δ, so that $\kappa(\bar{Q}) \propto \delta^{-1}$. As this is the dominant source of error in the calculation of the SE close to a surface state, we can approximate the relative error as

$$\Delta^{\text{in}}_{\Sigma,\text{r}} = c_1 \, \delta^{-1}, \tag{5.70}$$

where c_1 is a constant that depends on the machine precision and on the details of the algorithm. The label "in" explicitly indicates that this is the error in the SE calculated with the extended algorithm [$\Sigma^{\text{in}}_{\{\text{L,R}\}}$ in Eq. (5.67)]. The absolute error $\Delta^{\text{in}}_{\Sigma}$ is equal to the relative error times $||\Sigma||_{\text{max}}$, which is itself proportional to δ^{-1}, so that we get $\Delta^{\text{in}}_{\Sigma} \propto \delta^{-2}$.

When using Eq. (5.68) to estimate the error in the SE we introduce an additional error due to the inversion involved in implicitly obtaining g_{L}. The largest singular value of g_{L} is proportional to δ^{-1}, and the smallest one is proportional to δ, so that the relative error introduced by the inversion is proportional to $\kappa(g_{\text{L}}^{-1}) = \kappa(g_{\text{L}}) \propto \delta^{-2}$. For small δ we can therefore write for the error in $\Sigma^{\text{out}}_{\text{L}}$

$$\Delta^{\text{out}}_{\Sigma,\text{r}} = c_2 \, \delta^{-2}, \tag{5.71}$$

where c_2 is again a constant. Since the two errors are random and uncorrelated, the total estimated error can be approximated by adding the contributions from the two inversions

$$\Delta^2_{\Sigma,\text{r}} \approx (\Delta^{\text{in}}_{\Sigma,\text{r}})^2 + (\Delta^{\text{out}}_{\Sigma,\text{r}})^2. \tag{5.72}$$

$\Delta_{\Sigma,\text{r}}$ is therefore a good estimate for the true error $\Delta^{\text{in}}_{\Sigma,\text{r}}$ if $\Delta^{\text{out}}_{\Sigma,\text{r}}$ is small. Close to surface states however $\Delta^{\text{out}}_{\Sigma,\text{r}} \gg \Delta^{\text{in}}_{\Sigma,\text{r}}$, so that $\Delta_{\Sigma,\text{r}}$ largely overestimates the true error.

To verify these estimates numerically we present a scheme for calculating $\Delta^{\text{in}}_{\Sigma,\text{r}}$ and $\Delta^{\text{out}}_{\Sigma,\text{r}}$ independently. For each SE we perform a second calculation where we add a small amount of noise to the input matrices K_0, K_1, and K_{-1}, so that we obtain the self-energy $\Sigma_{\text{L,noise}}$ for a slightly perturbed system. The noise is added as a random relative perturbation of each element of the matrices. As we decrease the magnitude of the noise the difference between Σ_{L} and $\Sigma_{\text{L,noise}}$ is reduced until it becomes constant for noise smaller than a critical value. In this range of minimum noise even if the difference in the input matrices decreases, the difference in the output matrices is constant, it therefore corresponds to the error in the calculation. As one might expect we find that this critical value of noise is of the same order of magnitude as the numerical accuracy used (approximately 10^{-15} in our calculations). We can therefore obtain

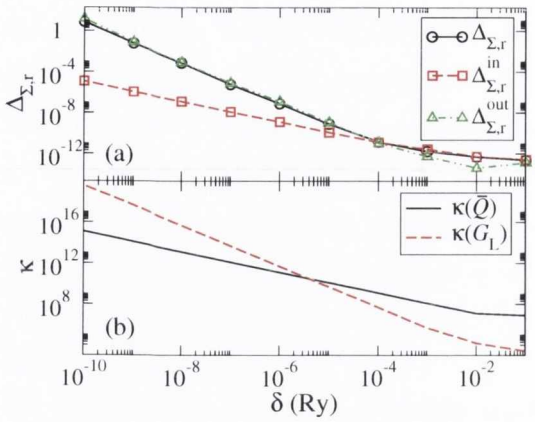


Figure 5.9: (a) Relative error of the self-energy $\Delta_{\Sigma,r}$ ($\Delta_{\Sigma,r}^{in}$ represents the true error), (b) condition number κ of $\bar{Q}$ and g_L, as a function of the broadening δ for the carbon nanotube of Fig. 5.4 calculated at the surface state energy.

$\Delta_{\Sigma,r}^{in} = \left|\left|\Sigma_L^{in} - \Sigma_{L,noise}^{in}\right|\right|_{max} / \left|\left|\Sigma_L^{in}\right|\right|_{max}$ and $\Delta_{\Sigma,r}^{out} = \left|\left|\Sigma_L^{out} - \Sigma_{L,noise}^{out}\right|\right|_{max} / \left|\left|\Sigma_L^{out}\right|\right|_{max}$, with the magnitude of the noise equal to the critical value.

We have calculated the maximum error for a set of 128 energy points located within 10^{-11} Ry around the energy of the surface state at -0.34 eV for different values of δ. The result is shown in Fig. 5.9(a). Indeed for small δ $\Delta_{\Sigma,r}^{in}$ follows Eq. (5.70) with $c_1 \approx 10^{-15}$ Ry, $\Delta_{\Sigma,r}^{out}$ follows Eq. (5.71) with $c_2 \approx 10^{-19}$ Ry^2, and $(\Delta_{\Sigma,r})^2 \approx (\Delta_{\Sigma,r}^{in})^2 + (\Delta_{\Sigma,r}^{out})^2$. In Fig. 5.9(b) the condition numbers $\kappa(\bar{Q})$ and $\kappa(g_L)$ are shown, confirming $\kappa(\bar{Q}) \propto \delta^{-1}$ and $\kappa(g_L) \propto \delta^{-2}$. This demonstrates that close to surface states $\Delta_{\Sigma,r}$ is mainly caused by the calculation of g_L. Thus $\Delta_{\Sigma,r}$ largely overestimates the real error $\Delta_{\Sigma,r}^{in}$, which even for $\delta = 10^{-10}$ Ry has an acceptable size of $\Delta_{\Sigma,r}^{in} \approx 10^{-5}$.

Since c_1 and c_2 are generally system dependent, in practical calculations we use a value of δ ranging between 10^{-7} Ry and 10^{-6} Ry for energies in the vicinity of surface states, mainly in order to limit the absolute error. Moreover δ is added in an energy range corresponding approximately to the FWHM of the imaginary part of $(E - E_n + i\delta)^{-1}$, which is equal to 2δ. Although this range is only of the order of $10^{-7} - 10^{-6}$ Ry, in practical calculations where both energy and k-point sampling are fine, the number of times when this prescription is applied can be rather large (see Fig. 5.5).

The above analysis confirms that close to surface states also direct semi-analytical methods have the same accuracy problems of recursive methods. This fact is usually ignored in the literature [214, 63, 215, 197], where it is assumed that the accuracy is constant for a given algorithm. Here we show that the accuracy of a method is solely determined by the value of c_1, which, as indicated in Sec. 5.4, can vary over many

orders of magnitude. Our analysis also shows that methods requiring the explicit calculation of g_{L} from its inverse are much less accurate close to surface states than those calculating Σ_{L} directly.

5.6 Conclusions

By extending the scheme proposed in Ref. [214], we have presented a different but equivalent form for calculating the Green's functions of an infinite quasi-1D system, as well as the SGF and SE for the semi-infinite system. We have then constructed an extended algorithm containing also the necessary steps to regularize the ill conditioned hopping matrices. This is found to be crucial in order to obtain a numerically stable algorithm. By applying a unitary transformation based on a SVD we remove the rapidly decaying states and calculate the SE for an effective system with reduced size. We further decrease the condition number of the hopping matrices by adding a small random perturbation and by limiting the smallest singular value.

We have performed a detailed error analysis on the numerical calculation of the SE, showing that if the algorithm does not involve an inversion of the hopping matrices K_1 (or K_{-1}) high accuracy is obtained. We also find that the error is not constant as a function of energy. It is shown that an increase in accuracy is needed especially close to energies where the SE and SGF diverge, which corresponds to the presence of surface states in the semi-infinite system. At these energies we improve the accuracy by adding a small imaginary part to the energy. We have shown that this procedure affects the transport properties little in the high transmission limit. However, for low transmission this adds some spurious surface density of states contributing significantly to the total transmission. The transport can therefore be strongly affected, so that the imaginary part should be added only in a small energy range around the poles and it should be as small as possible.

The algorithm is highly numerically stable and extremely accurate. Most importantly errors and accuracy can be closely monitored. We believe that this is an ideal algorithm to be used with *ab initio* transport schemes, where the condition of the Hamiltonian and its sparsity is controlled by the convergence of the electronic structure and cannot be fixed *a priori*. In all the calculations presented in chapters 6, 7 the extended algorithm is therefore used to obtain the SEs.

Chapter 6

Weakly coupled and bound states in electronic transport

The spectral function of the EM [Eq. (4.34)] contains a term, which is proportional to the small imaginary part added to the energy, δ. In chapter 4, where the NEGF formalism was introduced, we neglected that term in the calculation of the density matrix. This was justified by the assumption that all the states in the EM are coupled to the leads. However, if there are localized states in the EM, that are not coupled to the propagating states in the leads, this term can not be neglected, and gives a finite contribution even in the limit $\delta \to 0^+$. These states are called bound states (BSs). General properties of BSs localized at the interface between two materials are discussed in Refs. [194, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243]. It is shown that BSs appear if there are no open channels in the leads at their energy, or along special points in the 2D BZ perpendicular to the stacking direction.

In the first part of this chapter the modifications to the NEGF equations, needed to account for the BSs at finite bias, are presented. The fundamental difficulty in dealing with BSs is that their non-equilibrium occupation is undefined within the NEGF formalism. Although this problem is well known, in standard NEGF based algorithms it is usually assumed that there are no BSs in the system, at least not at energies inside the bias window [62, 63]. In Ref. [244] the BSs for carbon nanotube systems are calculated, and a way to occupy them at finite bias is proposed, based on a symmetry decomposition of the charge density, and on a shift of the energy-integration limits. This solution is however highly system specific, since the shift of the integration limits is probably not applicable when the number of BSs increases. The results in the reference show the importance of a correct occupation of the BSs at finite bias for the calculation of the capacitance. In [245] it is proposed to set the occupation of the BSs based on the geometric position. This is achieved indirectly, the atomic position is inferred by using an atomic weighting factor, which depends

on the total coupling of an atom to each of the leads. This choice is rather arbitrary, since the total coupling of an atom to the leads depends on its geometric position, but also on many other factors.

In this work we start from the assumption that in real systems there are interactions outside the ones contained in our single-particle DFT Hamiltonian, and that these lead to an effective coupling to the leads [246, 247, 248, 249, 243]. By adding this effective coupling to the formalism, we can set the occupation of the BSs. For many systems the physically meaningful effective coupling can be determined, although this is certainly not possible for all systems. We perform an eigenstate expansion of the effective Hamiltonian H_{eff} [Eq. (4.28)] and GF, which gives us a way to find and characterize the BSs. For a true BS the imaginary part of the corresponding eigenvalue of H_{eff} has to be zero. We introduce the concept of generalized BS (GBS), for which we only require that the complex modulus of the imaginary part is smaller than some maximum value. This leads to a common description for both BSs and weakly coupled states. An effective coupling parameter to the each of the leads for the eigenstates of H_{eff} is derived. This allows us to approximate the contribution of the GBSs to the density matrix and transmission with a simple model.

For systems with very weakly coupled states, an extremely fine energy mesh is needed to numerically evaluate the integrals over energy of the non-equilibrium part of the lesser GF [Eq. (4.72)]. This is required in order to obtain the density matrix, and of the transmission coefficient [Eq. (4.97)], and therefore the current. By using a finite value for δ, and by calculating the GBSs, we introduce a BSs correction scheme, which allows us to evaluate these integrals using a rather coarse energy mesh. In this way the computational cost for such systems is drastically reduced. As an alternative way, and based on the same concepts, a method to obtain an adaptive energy mesh for the integration of the non-equilibrium part of the electron density is presented. This method is also used for the calculation of the energy-dependent transmission coefficient.

These concepts are then applied to a set of example systems. First a simple tight-binding model is presented, that shows in which typical situations BSs are formed, and allows us to analyze their properties. Then we present two sets of calculations of systems with very weakly coupled states, where both the BSs correction scheme and the adaptive energy grid algorithm are applied. The first example is a calculation for a small test system consisting of a C_3 molecule placed in a parallel-plate gold capacitor. The second is for a rather large system, where a Mn_{12} molecule is attached to two gold leads. In chapter 7 the concepts will be applied to Fe/MgO/Fe(100) single- and double-barrier tunnel junctions. We will also show that it is essential to correctly

include the BSs in the calculation, in order to obtain a physically meaningful result.

6.1 Bound states in the NEGF formalism

The spectral function A_{M} for the EM has been derived in Eq. (4.34) and reads

$$A_{\mathrm{M}}(E) = G_{\mathrm{M}} \left(\Gamma_{\mathrm{L}} + \Gamma_{\mathrm{R}}\right) G_{\mathrm{M}}^{+} + \delta \; G_{\mathrm{M}} S_{\mathrm{M}} G_{\mathrm{M}}^{\dagger}. \tag{6.1}$$

If there are no poles in G_{M}, then the term proportional to δ vanishes in the limit of $\delta \rightarrow 0^{+}$. This was assumed in all the derivations in chapter 4, and therefore the term was neglected. In the general case however G_{M} has poles at specific energies. Each pole corresponds to a state that is localized in the EM, with no coupling to the propagating states of the leads. In these cases we can not neglect the second term, which describes exactly these localized states. We usually refer to these localized states also as bound states (BSs). We can split up $A_{\mathrm{M}}(E)$ into the part $A_{\mathrm{CS}}(E)$, originating from states that are coupled to the propagating states of the leads

$$A_{\mathrm{CS}}(E) \;\; = \;\; G_{\mathrm{M}} \left(\Gamma_{\mathrm{L}} + \Gamma_{\mathrm{R}}\right) G_{\mathrm{M}}^{+}, \tag{6.2}$$

and the part $A_{\mathrm{BS}}(E)$ originating from the bound states

$$A_{\mathrm{BS}}(E) \;\; = \;\; 2\delta \; G_{\mathrm{M}} S_{\mathrm{M}} G_{\mathrm{M}}^{\dagger}, \tag{6.3}$$

so that

$$A_{\mathrm{M}}(E) = A_{\mathrm{CS}}(E) + A_{\mathrm{BS}}(E). \tag{6.4}$$

We can define the onsite DOS, $\mathcal{N}_{\mathrm{M}}(E)$, of the EM, in which the overlap terms to the leads are neglected, as

$$\mathcal{N}_{\mathrm{M}}(E) = \frac{1}{2\pi} \mathrm{Tr}[A_{\mathrm{M}} S_{\mathrm{M}}]. \tag{6.5}$$

This can equally be split up into

$$\mathcal{N}_{\mathrm{CS}}(E) \;\; = \;\; \frac{1}{2\pi} \mathrm{Tr}[A_{\mathrm{CS}} S_{\mathrm{M}}], \tag{6.6}$$

$$\mathcal{N}_{\mathrm{BS}}(E) \;\; = \;\; \frac{1}{2\pi} \mathrm{Tr}[A_{\mathrm{BS}} S_{\mathrm{M}}], \tag{6.7}$$

so that

$$\mathcal{N}_{\mathrm{M}}(E) = \mathcal{N}_{\mathrm{CS}}(E) + \mathcal{N}_{\mathrm{BS}}(E). \tag{6.8}$$

Whereas $\mathcal{N}_{\mathrm{CS}}(E)$ is a smooth function of energy, $\mathcal{N}_{\mathrm{BS}}(E)$ corresponds to a set of δ-functions, one for each of the poles in $G_{\mathrm{M}}(E)$. The maximum number of electrons in the states connected to the leads q_{CS} is

$$q_{\mathrm{CS}} = \int_{-\infty}^{\infty} \mathcal{N}_{\mathrm{CS}}(E) \; dE, \tag{6.9}$$

and the maximum number of electrons that can be associated to BSs q_{BS} is

$$q_{\mathrm{BS}} = \int_{-\infty}^{\infty} \mathcal{N}_{\mathrm{BS}}(E)\, dE, \tag{6.10}$$

so that the total maximum number of electrons q_{M} that can populate the EM is $q_{\mathrm{M}} = q_{\mathrm{CS}} + q_{\mathrm{BS}}$, and it is approximately equal to the number of orbitals in the scattering region N.

Within the NEGF formalism the total electron density ρ_{M} of the EM for a two-terminal device, as derived in Eq. (4.64), is

$$\rho_{\mathrm{M}} = \frac{1}{2\pi i} \int_{-\infty}^{\infty} G^{<}_{\mathrm{M}}(E) dE, \tag{6.11}$$

with the lesser GF, $G^{<}_{\mathrm{M}}(E)$, [Eq. (4.63)] given by

$$G^{<}_{\mathrm{M}}(E) = i\, G_{\mathrm{M}} \left(\Gamma_{\mathrm{L}} f_{\mathrm{L}} + \Gamma_{\mathrm{R}} f_{\mathrm{R}}\right) G^{\dagger}_{\mathrm{M}}. \tag{6.12}$$

If all the states are assumed to be occupied, which is obtained by setting $f_{\mathrm{L}}(E) = f_{\mathrm{R}}(E) = 1$, then $G^{<}_{\mathrm{M}}(E)$ should be equal to $iA_{\mathrm{M}}(E)$. As can be seen comparing Eq. (6.12) with Eq. (6.1) for $f_{\mathrm{L}}(E) = f_{\mathrm{R}}(E) = 1$, in $G^{<}_{\mathrm{M}}(E)$ the term proportional to δ is missing. The reason is that in the derivation of Eq. (4.64) we assumed that all the states in the EM are connected to the propagating states in the leads, and that G_{M} has no poles.

Whereas it is straightforward to include the term describing the BSs to the SF via Eq. (6.1), this is not possible for $G^{<}_{\mathrm{M}}(E)$, since the occupation of the BSs in the EM is not defined within the NEGF formalism. The reason is that the occupation of a state within NEGF is based on the difference between in-flux from the leads and out-flux into the leads. For a BS however both in- and out-flux are zero, and this causes the occupation to be undefined. Within a time-dependent approach the occupation depends on the system history [194, 187, 208, 250, 251]. For non-interacting particles the occupation is constant for all times, and remains equal to the initial occupation. However, due to interactions that are not included in the DFT Hamiltonian (e.g. electron-electron interactions beyond mean field theory, electron-phonon coupling, electron-photon coupling, electron magnon coupling), in many systems also the occupation of BSs will change over time, until a steady state occupation is reached. In the next sections we present a way to set the occupation of BSs, based on the assumption of the existence of these additional interactions. We note that in a practical calculation it is not possible to distinguish between a true BS, with exactly zero coupling, and a very weakly coupled state. Therefore both have to be treated similarly.

6.2 Occupation of bound states

In a DFT ground state calculation all the states, and therefore also the BSs, are occupied up to the Fermi energy. When comparing to experiment it is implicitly assumed that the ground state is reached after thermal equilibration with the environment. The processes leading to the ground state are however outside the standard DFT theory. The thermal relaxation can be caused by electron-electron interactions, electron-phonon scattering, electron-magnon scattering [246, 243], electron-photon scattering, and other inelastic effects. Within the NEGF formalism, even at zero bias the occupation of the BSs is undefined. However by splitting the lesser GF into its equilibrium and non-equilibrium contribution in form of Eqs. (4.71) and (4.72), one implicitly assumes that all the BSs are occupied up to the Fermi energy of the leads. The underlying assumption is the same as for ground state DFT, namely that although they are not coupled to the leads, due to thermal relaxation with the environment their occupation is well defined. The interactions are again outside the DFT Hamiltonian. We can now introduce a generic relaxation time τ for a BS, that is inversely proportional to its coupling to the current/voltage electrodes [246]. The longer the relaxation time, the weaker are the interactions of the BS with the lead. In a system consisting of an EM coupled to leads, we can split up the relaxation time as

$$\frac{1}{\tau} = \frac{1}{\tau_{\mathrm{EM}}} + \frac{1}{\tau_{\mathrm{L}}} + \frac{1}{\tau_{\mathrm{R}}}, \tag{6.13}$$

where τ_{L} (τ_{R}) is the relaxation time caused by the additional interactions with the left (right) lead, and τ_{EM} is the relaxation time due to the additional processes within the EM itself. τ_{L} and τ_{R} can for example have a finite value, if the structure of the leads deviates slightly from the perfect crystalline structure assumed when constructing the self-energies. For energies, where the DOS for the perfectly crystalline leads is exactly zero, it might be finite for the real, imperfect structure. In this case, whereas the coupling to the leads is exactly zero for the perfectly crystalline leads, it might therefore be finite for the real leads. In the same way a finite value for τ_{EM} can also be caused by a slight difference from the atomic structure from the one assumed in the calculations. In addition to the structural imperfections, there are always inelastic effects that further reduce the relaxation times.

All the scattering processes included in τ_{EM} eventually lead to an electron propagating into either the left or the right electrode. We can therefore include these into effective leads relaxation times $\tau^{\mathrm{eff}}_{\{\mathrm{L/R}\}}$ and set $\tau_{\mathrm{EM}} = \infty$ at the same time. The total

τ then is

$$\frac{1}{\tau} = \frac{1}{\tau_{\mathrm{L}}^{\mathrm{eff}}} + \frac{1}{\tau_{\mathrm{R}}^{\mathrm{eff}}}. \tag{6.14}$$

The meaning of this equation is that for a BS there is an effective coupling to each of the two leads, caused by interactions outside the DFT Hamiltonian. Since we are just interested in the steady state non-equilibrium occupation, the absolute value of the relaxation time is not important, and the occupation is given by the ratio between $\tau_{\mathrm{L}}^{\mathrm{eff}}$ and $\tau_{\mathrm{R}}^{\mathrm{eff}}$. The stronger a state is thermally coupled to one lead with respect to the other, the closer its occupation follows the Fermi energy of that lead. Since the effective relaxation times to the leads are outside the NEGF and DFT formalism, they have to be set by hand as boundary conditions, motivated by the physical properties of the system. An example system where this can be done is a parallel plates capacitor at finite bias. If there is a bound state somewhere close to the left surface, then the physically meaningful constraint for the relaxation time is that $\tau_{\mathrm{L}}^{\mathrm{eff}} \ll \tau_{\mathrm{R}}^{\mathrm{eff}}$, which results in state being occupied up to the left Fermi energy. For a state on the right-hand side we have $\tau_{\mathrm{L}}^{\mathrm{eff}} \gg \tau_{\mathrm{R}}^{\mathrm{eff}}$, meaning that it is occupied up to the right lead's Fermi energy.

These arguments, based on the relaxation time concept, are now included in the NEGF in a formal way. In this formalism we introduce the effective coupling to the leads, that has the same role as the inverse relaxation time. The starting point is the definition for the retarded GF G_{M} for the EM [Eq. (4.2)]

$$G_{\mathrm{M}}(E) = \left[(E + i\delta)S_{\mathrm{M}} - H_{\mathrm{M}} - \Sigma_{\mathrm{L}} - \Sigma_{\mathrm{R}}\right]^{-1}. \tag{6.15}$$

Instead of taking the limit $\delta \to 0^{+}$, we now use a finite positive value for δ, although very small ($\delta \ll 1$). The use of a finite δ corresponds to the physical assumption that all the states have a finite lifetime. We can therefore define an additional self-energy

$$\Sigma^{\delta} = -i\delta\ S_{\mathrm{M}}, \tag{6.16}$$

that takes into account all the effects that lead to the finite lifetime, and that are not included in the effective Hamiltonian. As discussed above, the origin of Σ^{δ} can for example be inelastic effects, but also slight deviations from the assumed structure. The retarded GF can then be written as

$$G_{\mathrm{M}}(E) \;=\; \left[ES_{\mathrm{M}} - H_{\mathrm{M}} - \Sigma_{\mathrm{L}} - \Sigma_{\mathrm{R}} - \Sigma^{\delta}\right]^{-1}. \tag{6.17}$$

We note that the Hermitian part of Σ^{δ} is zero, so that it does not lead to a shift of the eigenvalue spectrum, and its only effect is to broaden the levels. The coupling matrix Γ^{δ} associated to Σ^{δ} is

$$\Gamma^{\delta} = i\left(\Sigma^{\delta} - \Sigma^{\delta\dagger}\right) = 2\delta S_{\mathrm{M}}. \tag{6.18}$$

so that $\Sigma^\delta = -\frac{i}{2}\Gamma^\delta$. The SF for the BSs [Eq. (6.3)] can then be written as

$$A_{\mathrm{BS}}(E) = G_{\mathrm{M}}\,\Gamma^\delta\,G_{\mathrm{M}}^\dagger, \tag{6.19}$$

where we note that now $A_{\mathrm{BS}}(E)$ depends on the used finite value for δ.

In order to determine the occupation of the BSs in the EM, we now split up the total additional coupling matrix Γ^δ into the coupling to the left lead $\Gamma^\delta_{\mathrm{L}}(E)$, and the coupling to the right lead $\Gamma^\delta_{\mathrm{R}}(E)$

$$\Gamma^\delta = \Gamma^\delta_{\mathrm{L}}(E) + \Gamma^\delta_{\mathrm{R}}(E). \tag{6.20}$$

Γ^δ has to be split up in such a way as to reproduce the effective coupling to the leads expected from the physical constraints. In general, the matrices $\Gamma^\delta_{\{\mathrm{L/R}\}}(E)$ are both position and energy dependent. The relative value of $\Gamma^\delta_{\mathrm{L}}$ and $\Gamma^\delta_{\mathrm{R}}$ at an energy E will then set the occupation of the BS. Since $\Gamma^\delta_{\{L/R\}}$ are coupling matrices, they have to be Hermitian, positive semidefinite. We now present a possible way to construct such matrices. First we note that since the overlap matrix S_{M} is a positive definite matrix, it can be decomposed as

$$S_{\mathrm{M}} = S_{\mathrm{M}}^{\frac{1}{2}} S_{\mathrm{M}}^{\frac{1}{2}}, \tag{6.21}$$

where $S_{\mathrm{M}}^{\frac{1}{2}}$ is a unique, positive definite, Hermitian matrix [196]. It can be seen as the square root of S_{M}, since the eigenvalues of $S_{\mathrm{M}}^{\frac{1}{2}}$ are equal to the square root of those of S_{M}, and the eigenvectors are identical. We can then write $\Gamma^\delta_{\{L/R\}}$ as

$$\begin{aligned} \Gamma^\delta_{\mathrm{L}} &= 2\delta\; S_{\mathrm{M}}^{\frac{1}{2}} \alpha(E) S_{\mathrm{M}}^{\frac{1}{2}}, \\ \Gamma^\delta_{\mathrm{R}} &= 2\delta\; S_{\mathrm{M}}^{\frac{1}{2}} \left[\mathbb{1}_N - \alpha\,(E)\right] S_{\mathrm{M}}^{\frac{1}{2}}. \end{aligned} \tag{6.22}$$

where the newly introduced, energy-dependent matrix $\alpha(E)$ is positive semidefinite and Hermitian, with eigenvalues banded between 0 and 1. It is straightforward to verify that with this definition both $\Gamma^\delta_{\{L/R\}}$ are indeed positive semidefinite and Hermitian. If it is assumed that S_{M} is close to the unity matrix ($S_{\mathrm{M}} \approx \mathbb{1}_N$), then we can write $\Gamma^\delta_{\{L/R\}}$ in an alternative, approximate way as

$$\begin{aligned} \Gamma^\delta_{\mathrm{L}} &= 2\delta\{\alpha(E), S_{\mathrm{M}}\}, \\ \Gamma^\delta_{\mathrm{R}} &= 2\delta\{\mathbb{1}_N - \alpha(E), S_{\mathrm{M}}\}, \end{aligned} \tag{6.23}$$

where $\{A, B\} = (AB + B^\dagger A^\dagger)/2$ for two general square matrices A and B. We note that with this definition $\Gamma^\delta_{\{L/R\}}$ are still Hermitian, but only approximately positive semidefinite. If $S_{\mathrm{M}} = \mathbb{1}_N$, then Eqs. (6.23) and (6.22) are equivalent.

We now define $G_{\mathrm{M}}^{\mathrm{BS}<}(E)$ to be the contribution originating from this additional coupling to the lesser GF

$$G_{\mathrm{M}}^{\mathrm{BS}<}(E) = G_{\mathrm{M}} \left[\Gamma_{\mathrm{L}}^{\delta} f_{\mathrm{L}}(E) + \Gamma_{\mathrm{R}}^{\delta} f_{\mathrm{R}}(E)\right] G_{\mathrm{M}}^{\dagger}. \tag{6.24}$$

The total lesser GF can then be written as

$$G_{\mathrm{M}}^{<}(E) = iG_{\mathrm{M}} \left(\Gamma_{\mathrm{L}}^{\mathrm{eff}} f_{\mathrm{L}} + \Gamma_{\mathrm{R}}^{\mathrm{eff}} f_{\mathrm{R}}\right) G_{\mathrm{M}}^{\dagger}, \tag{6.25}$$

where we have introduced the total effective coupling matrices $\Gamma_{\mathrm{L}}^{\mathrm{eff}}$ and $\Gamma_{\mathrm{R}}^{\mathrm{eff}}$, given by

$$\begin{aligned} \Gamma_{\mathrm{L}}^{\mathrm{eff}} &= \Gamma_{\mathrm{L}} + \Gamma_{\mathrm{L}}^{\delta}, \\ \Gamma_{\mathrm{R}}^{\mathrm{eff}} &= \Gamma_{\mathrm{R}} + \Gamma_{\mathrm{R}}^{\delta}. \end{aligned} \tag{6.26}$$

These contain both the NEGF part of the coupling and also the additional finite-lifetime coupling. With the lesser GF given by Eq. (6.25), the occupation of the BSs is well defined, and determined by the choice of the matrix $\alpha(E)$. As discussed above, α has to be chosen in such a way as to reproduce the expected effective coupling due to the additional interactions not described by the effective Hamiltonian. We can again take the parallel plate capacitor with a BS on the surfaces as example. The BS on the left surface has a small effective coupling to the left lead, whereas the coupling is zero to the right lead. The opposite is the case for the BS located on the right-hand side surface. This physically meaningful effective coupling can be achieved by setting the matrix α to be a diagonal matrix, with diagonal elements equal to 1 for the matrix elements corresponding to an orbital on the left part of the EM, and 0 for the ones corresponding to orbitals of the right part:

$$\alpha = \begin{pmatrix} \mathbb{1}_{N_{\mathrm{EL}}} & \mathbb{0}_{N_{\mathrm{EL}},N_{\mathrm{ER}}} \\ \mathbb{0}_{N_{\mathrm{ER}},N_{\mathrm{EL}}} & \mathbb{0}_{N_{\mathrm{ER}}} \end{pmatrix}. \tag{6.27}$$

Here we have assumed that the indices of the orbitals in the EM are ordered from left to right; N_{EL} (N_{ER}) is the number of orbitals in the left-hand (right-hand) part of the capacitor. Therefore for such a system we can set the occupation of the BSs depending on their position in space. This is also reflected by the fact that α does not depend on E. Such a choice of α makes the BSs on the left (right) surface occupied up to $E_{\mathrm{F,L}}$ ($E_{\mathrm{F,R}}$). We will use the same approach to set the occupation of BSs appearing at the interfaces of Fe/MgO/Fe(100) tunnel junctions (chapter 7).

Analogously to the procedure highlighted in Sec. 4.3.4 we can split the lesser GF into

$$G_{\mathrm{M}}^{<}(E) = G_{\mathrm{eq}}^{<}(E) + G_{\mathrm{neq}}^{<}(E) + G_{\mathrm{neq}}^{\mathrm{BS}<}(E), \tag{6.28}$$

with the equilibrium part

$$G^{<}_{\mathrm{eq}}(E) = -\frac{f_{\mathrm{L}} + f_{\mathrm{R}}}{2}\left(G_{\mathrm{M}} - G^{\dagger}_{\mathrm{M}}\right), \tag{6.29}$$

the non-equilibrium part

$$G^{<}_{\mathrm{neq}}(E) = i\frac{f_{\mathrm{L}} - f_{\mathrm{R}}}{2}\, G(\Gamma_{\mathrm{L}} - \Gamma_{\mathrm{R}})G^{\dagger}_{\mathrm{M}}, \tag{6.30}$$

and the non-equilibrium part describing the BSs, given by

$$G^{\mathrm{BS}<}_{\mathrm{neq}}(E) = i\,\frac{f_{\mathrm{L}} - f_{\mathrm{R}}}{2}\, G_{\mathrm{M}}(\Gamma^{\delta}_{\mathrm{L}} - \Gamma^{\delta}_{\mathrm{R}})G^{\dagger}_{\mathrm{M}}. \tag{6.31}$$

Neglecting $G^{\mathrm{BS}<}_{\mathrm{neq}}$ is equivalent to setting $\alpha = \frac{1}{2}\,\mathbb{1}_N$, so that $\Gamma^{\delta}_{\mathrm{L}} = \Gamma^{\delta}_{\mathrm{R}}$. This results in the BSs with energies lying in the bias window being half filled. As described in Sec. 4.3.4, the division between the equilibrium and the non-equilibrium part is not uniquely defined. If $G^{\mathrm{BS}<}_{\mathrm{neq}}$ is neglected, different definitions lead to a different implicit occupation of the BSs. In a self-consistent calculation this generates completely different electrostatic potentials, and consequently different self-consistent solutions. If BSs are present in the bias window, and $G^{\mathrm{BS}<}_{\mathrm{neq}}$ is not considered, the results are therefore largely random.

The density matrix can be split up in an analogous way as

$$\rho_{\mathrm{M}} = \rho_{\mathrm{eq}} + \rho_{\mathrm{neq}} + \rho^{\mathrm{BS}}_{\mathrm{neq}}, \tag{6.32}$$

with

$$\rho_{\mathrm{eq}}(E) = \frac{1}{2\pi i}\int dE\; G^{<}_{\mathrm{eq}}(E), \tag{6.33}$$

$$\rho_{\mathrm{neq}}(E) = \frac{1}{2\pi i}\int dE\; G^{<}_{\mathrm{neq}}(E), \tag{6.34}$$

$$\rho^{\mathrm{BS}}_{\mathrm{neq}}(E) = \frac{1}{2\pi i}\int dE\; G^{\mathrm{BS}<}_{\mathrm{neq}}(E). \tag{6.35}$$

We conclude this section by introducing a different, but equivalent way to set the effective coupling, which is more appropriate for certain systems. Instead of defining an energy-dependent $\alpha(E)$ we can define a set of N_α energy-independent matrices α_n, each of which is positive semidefinite, and with eigenvalues ranging between 0 and 1. The set $\{\alpha_n\}$ has to satisfy

$$\sum_{n=1}^{N_\alpha} \alpha_n = \mathbb{1}_N. \tag{6.36}$$

We can then introduce a corresponding set of N_α fictitious leads, each of them with its own Fermi energy $E_{\mathrm{F},n}$, and effective coupling

$$\Gamma^{\delta}_{n} = \delta\; S^{\frac{1}{2}}_{\mathrm{M}}\; \alpha_n\; S^{\frac{1}{2}}_{\mathrm{M}}. \tag{6.37}$$

The contribution of the BSs to the lesser GF then is

$$G_{\mathrm{M}}^{\mathrm{BS}<}(E) = G_{\mathrm{M}} \sum_{n=1}^{N_\alpha} \Gamma_n^\delta f_n(E)\, G_{\mathrm{M}}^\dagger, \tag{6.38}$$

with $f_n(E)$ being the Fermi-Dirac distribution with Fermi energy $E_{\mathrm{F},n}$. It is possible to determine the matrix $\alpha(E)$ giving the same $G^{\mathrm{BS}<}(E)$ by equating Eqs. (6.24) and (6.38), which gives

$$\alpha(E) = \sum_{n=1}^{N_\alpha} \alpha_n \frac{f_n(E) - f_{\mathrm{R}}(E)}{f_{\mathrm{L}}(E) - f_{\mathrm{R}}(E)}. \tag{6.39}$$

This shows that both approaches are equivalent. We will use Eq. (6.39) to set the occupation of the BSs in Sec. 7.7, where double-barrier Fe/MgO/Fe/MgO/Fe(100) tunnel junctions are investigated.

6.3 Generalized bound states

For a two-terminal device the effective Hamiltonian of the EM, defined in Eq. (4.28), is

$$H_{\mathrm{eff}} = H_{\mathrm{M}} - \Sigma_{\mathrm{L}} - \Sigma_{\mathrm{R}}, \tag{6.40}$$

and $G_{\mathrm{M}} = [(E + i\delta)S_{\mathrm{M}} - H_{\mathrm{eff}}]^{-1}$. In the same way as described in Sec. 4.1 we can now write the effective Hamiltonian and the GF by using their spectral representation, the major difference however is that, in contrast to the DFT Hamiltonian, the matrix H_{eff} is not Hermitian. As a consequence the eigenvalues of H_{eff} are complex valued, and the right and left eigenvectors differ from each other. For a given complex eigenvalue ϵ_n the right eigenvectors ψ_n satisfy the equation

$$H_{\mathrm{eff}}\psi_n = \epsilon_n S_{\mathrm{M}}\psi_n, \tag{6.41}$$

whereas the left eigenvectors $\tilde{\psi}_n$ for the same eigenvalue fulfill

$$\tilde{\psi}_n^\dagger H_{\mathrm{eff}} = \tilde{\psi}_n^\dagger \epsilon_n S_{\mathrm{M}}. \tag{6.42}$$

We note that the sets $\{\epsilon_n\}$, $\{\psi_n\}$, and $\{\tilde{\psi}_n\}$ are all energy dependent, since H_{eff} itself is energy dependent. Moreover, due to the imposed causality the eigenvalues ϵ_n have a negative imaginary part. The eigenvectors are normalized in such a way that $\tilde{\psi}_n^\dagger S_{\mathrm{M}}\psi_m = \delta_{nm}$ and $\tilde{\psi}_n^\dagger S_{\mathrm{M}}\tilde{\psi}_n = 1$. H_{eff} can then be written in spectral form as

$$H_{\mathrm{eff}} \;=\; \sum_{n=1}^{N} \epsilon_n S_{\mathrm{M}}\, \psi_n \tilde{\psi}_n^+\, S_{\mathrm{M}}, \tag{6.43}$$

and the retarded GF is

$$G_{\mathrm{M}} = \sum_{n=1}^{N} \frac{1}{E + i\delta - \epsilon_n} \psi_n \tilde{\psi}_n^{+}. \tag{6.44}$$

A BS is found if there is a pole in G_{M} for $\delta \to 0^{+}$ [194], which is the case at a real energy $E_{\mathrm{BS},\mu}$ if

$$\mathrm{Re}\left[\epsilon_n(E_{\mathrm{BS},\mu})\right] = E_{\mathrm{BS},\mu}, \tag{6.45}$$

and

$$\mathrm{Im}[\epsilon_n(E_{\mathrm{BS},\mu})] = 0. \tag{6.46}$$

The BS index μ runs from 1 to N_{BS}, where N_{BS} is the number of BSs, which corresponds to the number of poles of G_{M}. We define $\psi_{\mu}^{\mathrm{BS}} = \psi_n(E_{\mathrm{BS},\mu})$ and $\tilde{\psi}_{\mu}^{\mathrm{BS}} = \tilde{\psi}_n(E_{\mathrm{BS},\mu})$ as the wave functions evaluated at the BS energy.

This concept can be generalized to include also weakly coupled states, by defining a maximum value η^{W} for the imaginary part of the eigenvalues. We define the generalized bound states (GBSs) as those states, where at a real energy E_{μ}^{W}

$$\mathrm{Re}\left[\epsilon_n(E_{\mu}^{\mathrm{W}})\right] = E_{\mu}^{\mathrm{W}}, \tag{6.47}$$

and

$$-\mathrm{Im}[\epsilon_n(E_{\mu}^{\mathrm{W}})] \leq \eta^{\mathrm{W}}, \tag{6.48}$$

are fulfilled at the same time. As a matter of notation, we denote the variables relating to BSs and GBSs with Greek indices. The GBSs therefore include both true BSs and also those weakly coupled states, for which the broadening in the energy is smaller than η^{W}. The notion of GBSs is important for numerical computations, where the imaginary part of ϵ_n is never exactly zero, so that Eq. (6.46) is not applicable. By choosing the value of η^{W} it is therefore possible to set a lower limit to the imaginary part of the eigenvalues that is still considered numerically accurate. We note that the in general η^{W} can also be chosen to be large, in which case the GBSs represent a set of weakly coupled states, where the coupling is lower than some tolerance, which is set by η^{W}. The total number of GBSs in the EM N^{W} depends on the choice of η^{W}, and it can have any value. We define

$$\epsilon_{\mu}^{\mathrm{W}} = \epsilon_n(E_{\mu}^{\mathrm{W}}) \tag{6.49}$$

as the energy eigenvalue of the GBS with index μ, and analogously $\psi_{\mu}^{\mathrm{W}} = \psi_n(E_{\mu}^{\mathrm{W}})$ and $\tilde{\psi}_{\mu}^{\mathrm{W}} = \tilde{\psi}_n(E_{\mu}^{\mathrm{W}})$ as the wave function evaluated at the energy of the GBS.

If we use a finite value for δ in constructing G_{M}, then all those states have to be considered BSs, which contribute significantly only to A_{BS}, and not to A_{CS}. This is

the case for all states with $-\mathrm{Im}(\epsilon_n) \ll \delta$. On the other hand, we consider those states as coupled to the leads, whose contribution to A_{BS} is negligible. This is the case for those states with $-\mathrm{Im}(\epsilon_n) \gg \delta$. Those states however, where $-\mathrm{Im}(\epsilon_n)$ is of the order of δ, contribute to both A_{BS} and A_{CS}, and have therefore to be considered partly as BSs, and partly as coupled states. If we choose η^{W} about one order of magnitude larger than δ, then the resulting set of GBSs includes all the states that contribute significantly to A_{BS}. Those GBSs, with $-\mathrm{Im}(\epsilon_n)$ of the order of δ or larger, contribute also to A_{CS}.

The GBSs can be found in a graphical way, as schematically shown in Fig. 6.1. First the real part of all the eigenvalues $\{\epsilon_n\}$ is plotted as function of E. These are shown as continuous black curves in Fig. 6.1(a). The next step is to remove all those eigenvalues with large imaginary part, i.e. where $\mathrm{Im}[\epsilon_n(E)] > \eta^{\mathrm{W}}$. We assume that this is the case for two of the eigenvalue curves. The remaining two eigenvalues with $\mathrm{Im}[\epsilon_n(E)] \leq \eta^{\mathrm{W}}$ are plotted in 6.1(b) as black curves. For these the second criterion for the GBSs [Eq. (6.48)] is always automatically fulfilled. The remaining criterion [Eq. (6.47)] is then also fulfilled at those energies, where the black curves cut the dashed red curve, which is simply a plot of E as function of E itself. In this particular case we find two GBSs, marked with blue circles. The energy of the GBS E^{W} is then simply the value of E at the intersection of the two lines (marked by a circle). We use this scheme also in our numerical computations. The eigenvalues $\{\epsilon_n\}$ are first calculated on an energy mesh, and the GBSs are then found by interpolating the eigenvalues curves between the mesh points. This simple method works rather well, and allows one to reliably find the GBSs. This is shown in the examples presented at the end of this chapter. However, if the eigenvalues show highly non-linear behavior as functions of energy, the method might not find all the GBSs.

Using the definitions of Γ_{L} [Eq. (4.32)] and Γ_{R} [Eq. (4.33)], together with the definition of H_{eff} [Eq. (4.28)], we can write the following identity:

$$\Gamma_{\mathrm{L}} + \Gamma_{\mathrm{R}} = i\left(H_{\mathrm{eff}} - H_{\mathrm{eff}}^{\dagger}\right). \tag{6.50}$$

If we multiply this equation from the left by $\tilde{\psi}_n^{\dagger}$, and from the right with $\tilde{\psi}_m$, we obtain

$$\tilde{\psi}_n^{\dagger}\left(\Gamma_{\mathrm{L}} + \Gamma_{\mathrm{R}}\right)\tilde{\psi}_m \;=\; i(\epsilon_n - \epsilon_m^{*})\tilde{\psi}_n^{\dagger} S_{\mathrm{M}} \tilde{\psi}_m \tag{6.51}$$

We can now define left and right coupling coefficients $\gamma_{\mathrm{L},n}^{E}$ and $\gamma_{\mathrm{R},n}^{E}$ for a general state with index n as

$$\begin{aligned} \gamma_{\mathrm{L},n}^{E} &= \tilde{\psi}_n^{\dagger}\Gamma_{\mathrm{L}}\tilde{\psi}_n, \\ \gamma_{\mathrm{R},n}^{E} &= \tilde{\psi}_n^{\dagger}\Gamma_{\mathrm{R}}\tilde{\psi}_n. \end{aligned} \tag{6.52}$$

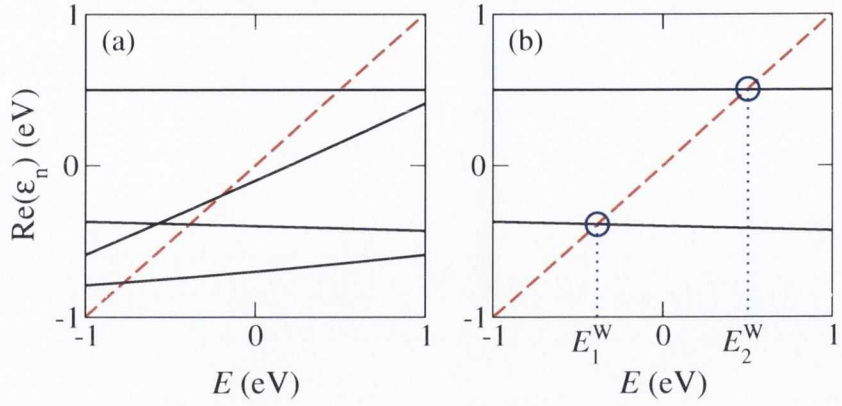


Figure 6.1: Schematic representation of the graphical method for the location of the GBSs; the real part of the eigenvalues $\epsilon_n(E)$ of H_{eff} is plotted as function of energy E for all eigenvalues (a), and for the subset with $\text{Im}[\epsilon_n(E)] \leq E^{\text{W}}$ (b). The circles indicate the energy of the GBSs.

Here the superscript "E" denotes that these coupling coefficients are energy dependent. Since $\Gamma_{\{L/R\}}$ are positive semidefinite Hermitian, the quantities $\gamma^E_{\{L/R\},n}$ are real and positive. By setting $n = m$ in Eq. (6.51), and by using the normalization of the $\tilde{\psi}_n$, we then obtain the following relation for the imaginary part of the eigenvalue

$$\text{Im}(\epsilon_n) = -\frac{1}{2}\left(\gamma^E_{\text{L},n} + \gamma^E_{\text{R},n}\right). \tag{6.53}$$

This means that at each energy, the imaginary part of the eigenvalues can be split into the components due to the coupling with the individual leads. For the generic μ-th GBS we finally define

$$\begin{aligned} \gamma_{\text{L},\mu} &= \tilde{\psi}^{\text{W}\dagger}_\mu \, \Gamma_{\text{L}}(E^{\text{W}}_\mu) \, \tilde{\psi}^{\text{W}}_\mu, \\ \gamma_{\text{R},\mu} &= \tilde{\psi}^{\text{W}\dagger}_\mu \, \Gamma_{\text{R}}(E^{\text{W}}_\mu) \, \tilde{\psi}^{\text{W}}_\mu, \end{aligned} \tag{6.54}$$

as the coupling coefficients at the energy of the GBS E^{W}_μ, so that

$$\text{Im}(\epsilon^{\text{W}}_\mu) = -\frac{1}{2}\left(\gamma_{\text{L},\mu} + \gamma_{\text{R},\mu}\right). \tag{6.55}$$

This relation shows that the imaginary part of the energy, that corresponds to the broadening of a weakly coupled level, is a sum of the part associated to the coupling to the left lead $-\gamma_{\text{L},\mu}/2$, and the part associated to the coupling to the right lead $-\gamma_{\text{R},\mu}/2$. These coupling coefficients have the same meaning as the ones used in tight-binding models [54, 202, 194]. It is important to note that $\gamma_{\{L/R\},\mu}$ are in general functions of the applied bias voltage. As we will show with the example of a weakly coupled magnetic molecule in Sec. 6.10, these couplings can change drastically as a function of bias, reflecting the fact that the nature of the states changes.

6.4 Bound states contribution to the Density matrix

Using the spectral representation of the GF, we obtain for the non-equilibrium part of the lesser GF describing the GBSs [Eq. (6.31)]

$$G^{\mathrm{BS}<}_{\mathrm{neq}} = i\frac{f_{\mathrm{L}} - f_{\mathrm{R}}}{2} \sum_{n,m=1}^{N} \frac{1}{E + i\delta - \epsilon_n} \frac{1}{E - i\delta - \epsilon_m^*} \psi_n \tilde{\psi}_n^\dagger \left(\Gamma_{\mathrm{L}}^\delta - \Gamma_{\mathrm{R}}^\delta\right) \tilde{\psi}_m \psi_m^\dagger. \tag{6.56}$$

The contribution to the charge density $\rho^{\mathrm{BS}}_{\mathrm{neq}}$, given in Eq. (6.35), then is

$$\rho^{\mathrm{BS}}_{\mathrm{neq}} = \sum_{n,m=1}^{N} \int_{-\infty}^{\infty} dE\, \frac{f_{\mathrm{L}} - f_{\mathrm{R}}}{4\pi} \frac{1}{E + i\delta - \epsilon_n} \frac{1}{E - i\delta - \epsilon_m^*} \psi_n \tilde{\psi}_n^\dagger \left(\Gamma_{\mathrm{L}}^\delta - \Gamma_{\mathrm{R}}^\delta\right) \tilde{\psi}_m \psi_m^\dagger \tag{6.57}$$

Since the matrices $\Gamma^\delta_{\{L/R\}}$ are proportional to δ, when we choose η^{W} somewhat larger than δ, the significant contributions to the integral come only from the GBSs that satisfy the criteria (6.47) and (6.48). The value of η^{W} thus determines how many states are important for the integral, in practical computations we typically choose $\eta^{\mathrm{W}} = 10\ \delta$. The sum over n and m in Eq. (6.57) can then be replaced by a sum over the GBSs only

$$\rho^{\mathrm{BS}}_{\mathrm{neq}} = \sum_{m,n\ \in\ \mathrm{GBS}} \int_{-\infty}^{\infty} dE\, \frac{f_{\mathrm{L}} - f_{\mathrm{R}}}{4\pi} \frac{1}{E + i\delta - \epsilon_n} \frac{1}{E - i\delta - \epsilon_m^*} \psi_n \tilde{\psi}_n^\dagger \left(\Gamma_{\mathrm{L}}^\delta - \Gamma_{\mathrm{R}}^\delta\right) \tilde{\psi}_m \psi_m^\dagger, \tag{6.58}$$

where "$m, n \in$ GBS" indicates that the sum runs only over those states m and n that have a GBS at some E. For each GBS with index μ the main contribution to the integral comes from energies within a range of δ around E^{W}_μ. If we now further assume that the weakly coupled states are not degenerate, or more generally that the spacing between the states is much larger than δ, then the integral is only significantly different from zero for $m = n$. With these assumptions we can also take the wave function dependent part, as well as the Fermi-Dirac distributions, out of the integral, with the functions evaluated at $E = E^{\mathrm{W}}_\mu$. $\rho^{\mathrm{BS}}_{\mathrm{neq}}$ can therefore be approximated as

$$\rho^{\mathrm{BS}}_{\mathrm{neq}} = \sum_{\mu}^{N^{\mathrm{W}}} \frac{f_{\mathrm{L}}(E^{\mathrm{W}}_\mu) - f_{\mathrm{R}}(E^{\mathrm{W}}_\mu)}{4\pi} \psi^{\mathrm{W}}_\mu \tilde{\psi}^{\mathrm{W}\dagger}_\mu \left(\Gamma_{\mathrm{L}}^\delta - \Gamma_{\mathrm{R}}^\delta\right) \tilde{\psi}^{\mathrm{W}}_\mu \psi^{\mathrm{W}\dagger}_\mu \int_{-\infty}^{\infty} dE \frac{1}{E + i\delta - \epsilon_n} \frac{1}{E - i\delta - \epsilon_n^*}\bigg|_{n\equiv\mu}, \tag{6.59}$$

where "$n \equiv \mu$" indicates that n is the index of the eigenvalue including the GBS with index μ, so that $\epsilon_\mu^{\mathrm{W}} = \epsilon_n(E_\mu^{\mathrm{W}})$ [see Eq. (6.49)]. Assuming that $\mathrm{Re}(\epsilon_n)$ varies linearly in an energy range of the order of a few δ around E_μ^{W}, and that $\mathrm{Im}(\epsilon_n)$ is constant over the same range, the integral can be evaluated, so that $\rho_{\mathrm{neq}}^{\mathrm{BS}}$ is

$$\begin{aligned}\rho_{\mathrm{neq}}^{\mathrm{BS}} &= \sum_{\mu}^{N^{\mathrm{W}}} \frac{f_{\mathrm{L}}(E_\mu^{\mathrm{W}}) - f_{\mathrm{R}}(E_\mu^{\mathrm{W}})}{2}\, \psi_\mu^{\mathrm{W}} \tilde{\psi}_\mu^{\mathrm{W}\dagger} \left(\Gamma_{\mathrm{L}}^{\delta} - \Gamma_{\mathrm{R}}^{\delta}\right) \tilde{\psi}_\mu^{\mathrm{W}} \psi_\mu^{\mathrm{W}\dagger} \\ &\quad \frac{1}{-2\ \mathrm{Im}(\epsilon_\mu^{\mathrm{W}}) + 2\delta}\ \frac{1}{|1-\beta_\mu|}. \end{aligned} \tag{6.60}$$

Here the factor

$$\beta_\mu = \left.\frac{\partial \mathrm{Re}(\epsilon_n)}{\partial E}\right|_{n\equiv\mu, E=E_\mu^{\mathrm{W}}} \tag{6.61}$$

is the derivative of the real part of the eigenvalues with respect to the energy, evaluated at $E = E_\mu^{\mathrm{W}}$. If $\mathrm{Re}(\epsilon_n)$ varies significantly around $E = E_\mu^{\mathrm{W}}$, then the factor $1/|1-\beta_\mu|$ takes into account the fact that the number of electrons, which can be put into a GBS, is different from 1. This will be illustrated with the tight-binding example in Sec. 6.8. β_μ can be either calculated approximately, by interpolating the slope of $\epsilon_n(E)$ between two energy points, or else it can be determined semi-analytically from

$$\begin{aligned}\frac{\partial \epsilon_n}{\partial E} &= \frac{\partial \tilde{\psi}_n^\dagger H_{\mathrm{eff}} \psi_n}{\partial E} \\ &= \tilde{\psi}_n^\dagger \frac{\partial H_{\mathrm{eff}}}{\partial E} \psi_n \qquad (6.62) \\ &= \tilde{\psi}_n^\dagger \left(\frac{\partial \Sigma_{\mathrm{L}}}{\partial E} + \frac{\partial \Sigma_{\mathrm{R}}}{\partial E}\right) \psi_n. \qquad (6.63)\end{aligned}$$

Here we have used the definition of H_{eff} [Eq. (4.28)] and the normalization $\tilde{\psi}_n^\dagger S_{\mathrm{M}} \psi_n = 1$. In Appendix C.4 it is shown how the derivative of the SE with respect to energy can be obtained. If we now also use the definitions in Eq. (6.54), together with relation (6.55), $\rho_{\mathrm{neq}}^{\mathrm{BS}}$ becomes

$$\rho_{\mathrm{neq}}^{\mathrm{BS}} = \sum_{\mu}^{N^{\mathrm{W}}} \frac{f_{\mathrm{L}}(E_\mu^{\mathrm{W}}) - f_{\mathrm{R}}(E_\mu^{\mathrm{W}})}{2}\, \frac{\tilde{\psi}_\mu^{\mathrm{W}\dagger} \left(\Gamma_{\mathrm{L}}^{\delta} - \Gamma_{\mathrm{R}}^{\delta}\right) \tilde{\psi}_\mu^{\mathrm{W}}}{\gamma_{\mathrm{L},\mu} + \gamma_{\mathrm{R},\mu} + 2\delta} \frac{1}{|1-\beta_\mu|} \psi_\mu^{\mathrm{W}} \psi_\mu^{\mathrm{W}\dagger}. \tag{6.64}$$

We now define p_μ as

$$p_\mu = \frac{\tilde{\psi}_\mu^{\mathrm{W}\dagger} \left(\Gamma_{\mathrm{L}}^{\delta} - \Gamma_{\mathrm{R}}^{\delta}\right) \tilde{\psi}_\mu^{\mathrm{W}}}{2\delta} = 2\tilde{\psi}_\mu^{\mathrm{W}\dagger} S_{\mathrm{M}}^{\frac{1}{2}} \alpha(E_\mu^{\mathrm{W}}) S_{\mathrm{M}}^{\frac{1}{2}} \tilde{\psi}_\mu^{\mathrm{W}} - 1, \tag{6.65}$$

where we have also used the definitions of $\Gamma_{\{\mathrm{L/R}\}}^{\delta}$ [Eq. (6.22)]. p_μ is a real number ranging between -1 and 1, and it is a measure of the relative coupling of a GBS to

the left and right leads. If all the GBSs have coupling only to the right lead, then $\alpha = \mathbb{0}_N$, and $p_\mu = -1$. On the other hand, if the GBSs have effective coupling only to the left lead, then $\alpha = \mathbb{1}_N$, and $p_\mu = 1$. If the coupling of a GBS is much stronger to the left (right) lead, then $p_\mu \approx 1$ ($p_\mu \approx -1$). Finally for symmetric coupling $p_\mu \approx 0$. Instead of choosing the matrix α, we can alternatively set the occupation of each individual BS by settings the value of p_μ. $\rho_{\mathrm{neq}}^{\mathrm{BS}}$ can then be rewritten as

$$\rho_{\mathrm{neq}}^{\mathrm{BS}} = \sum_{\mu}^{N^{\mathrm{W}}} p_\mu \frac{f_{\mathrm{L}}(E_\mu^{\mathrm{W}}) - f_{\mathrm{R}}(E_\mu^{\mathrm{W}})}{2} \frac{2\delta}{\gamma_{\mathrm{L},\mu} + \gamma_{\mathrm{R},\mu} + 2\delta} \frac{1}{|1-\beta_\mu|} \psi_\mu^{\mathrm{W}} \psi_\mu^{\mathrm{W}\dagger}. \quad (6.66)$$

For true BSs we have to take the limit $\delta \to 0^+$, in which case, by following an analogous procedure as that used for the density matrix, we obtain for the BSs contribution to the SF

$$A_{\mathrm{BS}} = \sum_{\mu=1}^{N_{\mathrm{BS}}} \delta(E - E_{\mathrm{BS},\mu}) \frac{2\pi}{|1-\beta_\mu|} \psi_\mu^{\mathrm{BS}} \psi_\mu^{\mathrm{BS}\dagger}, \quad (6.67)$$

and for the BSs contribution to the non-equilibrium part of the lesser GF

$$G_{\mathrm{neq}}^{\mathrm{BS},<} = i(f_{\mathrm{L}} - f_{\mathrm{R}}) \sum_{\mu=1}^{N_{\mathrm{BS}}} \delta(E - E_{\mathrm{BS},\mu})\, p_\mu \frac{\pi}{|1-\beta_\mu|} \psi_\mu^{\mathrm{BS}} \psi_\mu^{\mathrm{BS}\dagger}. \quad (6.68)$$

It is straightforward to verify that $\rho_{\mathrm{neq}}^{\mathrm{BS}}$ obtained by integrating this equation is identical to Eq. (6.66) for $\delta = 0$, if we use the fact that for true BSs $\gamma_{\mathrm{L},\mu} = \gamma_{\mathrm{R},\mu} = 0$.

6.5 Self-consistent occupation for generalized bound states

For true BSs the relative coupling number p_μ has to be set in order to reproduce the physically meaningful occupation, as described in Sec. 6.2. This can be done by choosing an appropriate α matrix. For the parallel plate capacitor with the BS on the surface, the choice of α described in Sec. 6.2 would simply result in $p_\mu = 1$ for the BS located on the left surface, and $p_\mu = -1$ for the BS located on the right surface. In conjunction with Eq. (6.66), this shows again that the BSs on the left (right) surface follow the left (right) Fermi energy. If the GBS is weakly coupled, then its occupation at a given bias is in principle defined. We can therefore choose whether to set p_μ using the matrix α as for the true BSs, or whether to choose p_μ in such a way, that the resulting occupation of the state corresponds to the one for $\delta = 0$. In this section we describe how to set p_μ in the second case.

For one weakly coupled GBS with index μ, the main contribution $\rho^{\mathrm{BS}}_{\mathrm{neq},\mu}$ to the total $\rho^{\mathrm{BS}}_{\mathrm{neq}}$, given in Eq. (6.66), is

$$\rho^{\mathrm{BS}}_{\mathrm{neq},\mu} = p_\mu \frac{f_{\mathrm{L}}(E^{\mathrm{W}}_\mu) - f_{\mathrm{R}}(E^{\mathrm{W}}_\mu)}{2} \frac{2\delta}{\gamma_{\mathrm{L},\mu} + \gamma_{\mathrm{R},\mu} + 2\delta} \frac{1}{|1-\beta_\mu|} \psi^{\mathrm{W}}_\mu \psi^{\mathrm{W}\dagger}_\mu . \tag{6.69}$$

Since it is a weakly coupled state, its contribution $\rho_{\mathrm{neq},\mu}$ to the total ρ_{neq} [Eq. (6.34)] can be calculated analogously to $\rho^{\mathrm{BS}}_{\mathrm{neq},\mu}$, and the result is

$$\rho_{\mathrm{neq},\mu} = \frac{f_{\mathrm{L}}(E^{\mathrm{W}}_\mu) - f_{\mathrm{R}}(E^{\mathrm{W}}_\mu)}{2} \frac{\gamma_{\mathrm{L},\mu} - \gamma_{\mathrm{R},\mu}}{\gamma_{\mathrm{L},\mu} + \gamma_{\mathrm{R},\mu} + 2\delta} \frac{1}{|1-\beta_\mu|} \psi^{\mathrm{W}}_\mu \psi^{\mathrm{W}\dagger}_\mu . \tag{6.70}$$

Due to the fact that the state is weakly coupled, $\rho^{\mathrm{BS}}_{\mathrm{neq},\mu} = 0$ in the limit $\delta \to 0^+$. We can therefore determine p_μ by requiring that the total contribution of the GBSs to the density matrix at a finite δ is identical to the one for $\delta = 0$, which is the case if

$$\rho_{\mathrm{neq},\mu}|_{\delta\to 0^+} = \rho_{\mathrm{neq},\mu} + \rho^{\mathrm{BS}}_{\mathrm{neq},\mu}|_{\delta\neq 0} . \tag{6.71}$$

With Eqs. (6.69) and (6.70) this condition becomes

$$\frac{\gamma_{\mathrm{L},\mu} - \gamma_{\mathrm{R},\mu}}{\gamma_{\mathrm{L},\mu} + \gamma_{\mathrm{R},\mu}} = \frac{\gamma_{\mathrm{L},\mu} - \gamma_{\mathrm{R},\mu}}{\gamma_{\mathrm{L},\mu} + \gamma_{\mathrm{R},\mu} + 2\delta} + p_\mu \frac{2\delta}{\gamma_{\mathrm{L},\mu} + \gamma_{\mathrm{R},\mu} + 2\delta} . \tag{6.72}$$

If we solve this equation for p_μ, we obtain

$$p_\mu = \frac{\gamma_{\mathrm{L},\mu} - \gamma_{\mathrm{R},\mu}}{\gamma_{\mathrm{L},\mu} + \gamma_{\mathrm{R},\mu}} . \tag{6.73}$$

We can therefore obtain the correct NEGF $\delta = 0$ result even using a finite value for δ, if we add the BSs correction term $\rho^{\mathrm{BS}}_{\mathrm{neq},\mu}$ and set p_μ to the value in Eq. 6.73. We call this method the bound states correction scheme (BSCS). For very weakly coupled states, setting $\delta = 0$ would require a very fine energy integration mesh for the numerical evaluation of $\rho_{\mathrm{neq},\mu}$. However, by using a finite value of δ, the integration mesh needs only to resolve peaks with a full width at half maximum (FWHM) of the order of δ. If δ is chosen much larger than the true width of the peak for $\delta = 0$, then the number of integration points can be drastically reduced, which results in a much faster computation. The drawback however is that the GBSs and their couplings have to be determined, however for $\delta \gg \gamma_{\mathrm{L},\mu} + \gamma_{\mathrm{R},\mu}$ the extra work is negligible compared to the gain in reducing the number of integration points.

In an analogous way to the non equilibrium part of the charge density [Eq. (6.70)], we can also calculate the contribution of a weakly coupled state $\rho_{\mathrm{eq},\mu}$ to the equilibrium part ρ_{eq} [Eq. (6.33)], and the result is

$$\rho_{\mathrm{eq},\mu} = \frac{f_{\mathrm{L}}(E^{\mathrm{W}}_\mu) + f_{\mathrm{R}}(E^{\mathrm{W}}_\mu)}{2} \frac{\gamma_{\mathrm{L},\mu} + \gamma_{\mathrm{R},\mu}}{\gamma_{\mathrm{L},\mu} + \gamma_{\mathrm{R},\mu} + 2\delta} \frac{1}{|1-\beta_\mu|} \psi^{\mathrm{W}}_\mu \psi^{\mathrm{W}\dagger}_\mu . \tag{6.74}$$

By adding $\rho_{\mathrm{eq},\mu}$ and $\rho_{\mathrm{neq},\mu}$ for $\delta = 0$, we obtain the occupation $\bar{p}_\mu$ of a molecular state as

$$\begin{aligned} \bar{p}_\mu &= \frac{f_{\mathrm{L}}(E_\mu^{\mathrm{W}}) + f_{\mathrm{R}}(E_\mu^{\mathrm{W}})}{2} + \frac{f_{\mathrm{L}}(E_\mu^{\mathrm{W}}) - f_{\mathrm{R}}(E_\mu^{\mathrm{W}})}{2} p_\mu \\ &= \frac{f_{\mathrm{L}}(E_\mu^{\mathrm{W}})}{2}(1 + p_\mu) + \frac{f_{\mathrm{R}}(E_\mu^{\mathrm{W}})}{2}(1 - p_\mu). \end{aligned} \tag{6.75}$$

To conclude this section we note that the standard NEGF scheme is not applicable in cases with very small and symmetric coupling, like quantum dots [206, 207]. Usually a different method, based on rate equations, is used to determine the occupation for these states [203, 204, 205, 206, 207]. Even in those cases however we can use the GBSs to separate out the weakly coupled states from the strongly coupled ones in a NEGF calculation. By determining the coupling coefficients [Eq. (6.54)] their occupation can then be determined using the rate equations formalism. The strongly coupled states on the other hand can be occupied with the NEGF formalism.

6.6 Mesh refinement algorithm

Using the BSCS, described in the previous section, it is possible to integrate the lesser GF using a rather coarse mesh, even when very weakly coupled states are present. Another approach is to use an adaptive energy grid instead of a grid with fixed mesh-point density. This grid should be fine at the energies of the weakly coupled states, and coarse in the energy regions where the DOS is rather smooth. Different approaches have been studied in Ref. [252] in order to generate such a mesh. Here we present an alternative method based on the concepts described in the previous sections. The main idea is to start the determination of the energy mesh by adding a rather large value of δ to the energy, so that all the possible peaks in the DOS due to weakly coupled states are broadened and they can be located with a rather coarse energy mesh. The mesh is then refined in an energy range around these broad peaks, and the size of δ is reduced. This is repeated until δ is smaller than the FWHM of the sharpest peak in the DOS.

Unfortunately, the contribution from the weakly coupled states to the total DOS is usually overshadowed by the much larger contribution from the strongly coupled states. In typical calculations of molecules attached to metallic leads, the major contribution to the DOS comes from the bulk-like states in the metallic electrodes. The peaks in the DOS due to weakly coupled states are often barely visible, and therefore difficult to locate. The main idea now is that instead of using the total DOS $\mathcal{N}_{\mathrm{M}}(E)$ to locate the peaks, we only use the contribution to the DOS generated by the

GBSs $\mathcal{N}_{\rm BS}(E)$, evaluated at a finite value of δ. For a very large δ the main contribution to the total DOS $\mathcal{N}_{\rm M}(E) = \mathcal{N}_{\rm CS}(E) + \mathcal{N}_{\rm BS}(E)$ [Eq. (6.8)] comes from $\mathcal{N}_{\rm BS}(E)$ [Eq. (6.7)], whereas for small δ the largest part of the DOS is usually contained in the contribution coming from the coupled states $\mathcal{N}_{\rm CS}(E)$ [Eq. (6.6)]. As discussed in Sec. 6.1, $\mathcal{N}_{\rm BS}(E)$ contributes in the limit $\delta \to 0^+$ only if true BSs are present, for finite δ however it contributes also due to the presence of weakly coupled states.

$\mathcal{N}_{\rm BS}(E)$ is now written in an approximate form, that allows to determine the width and height of possible peaks as function of δ. Using Eqs. (6.7), (6.19), together with the spectral representation of $G_{\rm M}$ [Eq. (6.44)], $\mathcal{N}_{\rm BS}(E)$ can be written as

$$\mathcal{N}_{\rm BS}(E) = \frac{1}{\pi} \sum_{n,m=1}^{N} \delta \frac{1}{E + i\delta - \epsilon_n} \frac{1}{E - i\delta - \epsilon_m^*} \tilde{\psi}_n^\dagger S_{\rm M} \tilde{\psi}_m \psi_m^\dagger S_{\rm M} \psi_n. \quad (6.76)$$

As described in Sec. 6.3, for a given value of δ we can choose $\eta^{\rm W}$ in such a way, that the main contribution to $\mathcal{N}_{\rm BS}(E)$ is due to the resulting set of $N^{\rm W}$ GBSs. This can be achieved by setting $\eta^{\rm W}$ about one order of magnitude larger than δ. As an approximation, we can then replace the sum over all states in Eq. (6.76) by the sum over the GBSs only. In this section we assume that the $\{\epsilon_n\}$ are constant over energy, at least in the vicinity of the energy of the GBSs. If we further neglect overlap terms, it is enough to consider the terms with $n = m$. Eq. (6.76) can then be approximated to

$$\mathcal{N}_{\rm BS}(E) \approx \frac{1}{\pi} \sum_{\mu=1}^{N^{\rm W}} \delta \frac{1}{E + i\delta - \epsilon_\mu^{\rm W}} \frac{1}{E - i\delta - \epsilon_\mu^{\rm W*}} \tilde{\psi}_\mu^{\rm W\dagger} S_{\rm M} \tilde{\psi}_\mu^{\rm W} \; \psi_\mu^{\rm W\dagger} S_{\rm M} \psi_\mu^{\rm W}. \quad (6.77)$$

The wave functions are normalized in such a way that $\tilde{\psi}_\mu^{\rm W\dagger} S_{\rm M} \tilde{\psi}_\mu^{\rm W} = 1$. The term $\psi_\mu^{\rm W\dagger} S_{\rm M} \psi_\mu^{\rm W}$ in contrast can have any value. However here we assume that also $\psi_\mu^{\rm W\dagger} S_{\rm M} \psi_\mu^{\rm W} = 1$. This is usually a good approximation for weakly coupled states, for which $\tilde{\psi}_\mu^{\rm W} \approx \psi_\mu^{\rm W}$. The approximate relation for $\mathcal{N}_{\rm BS}(E)$ then is

$$\mathcal{N}_{\rm BS}(E) = \frac{1}{\pi} \sum_{\mu=1}^{N^{\rm W}} \frac{\delta}{\left[E - {\rm Re}(\epsilon_\mu^{\rm W})\right]^2 + \left[\delta - {\rm Im}(\epsilon_\mu^{\rm W})\right]^2}. \quad (6.78)$$

We note that since ${\rm Im}(\epsilon_\mu^{\rm W}) \leq 0$, then $\delta - {\rm Im}(\epsilon_\mu^{\rm W}) > 0$. $\mathcal{N}_{\rm BS}(E)$ therefore is simply a sum of Lorentzian functions, one for each of the GBSs. The contribution from a single GBS is

$$\mathcal{N}_{{\rm BS},\mu}(E) = \frac{1}{\pi} \frac{\delta}{\left[E - {\rm Re}(\epsilon_\mu^{\rm W})\right]^2 + \left[\delta - {\rm Im}(\epsilon_\mu^{\rm W})\right]^2}, \quad (6.79)$$

so that $\mathcal{N}_{\mathrm{BS}}(E) = \sum_{n=1}^{N^{\mathrm{W}}} \mathcal{N}_{\mathrm{BS},\mu}(E)$. The center of the Lorentzian function $\mathcal{N}_{\mathrm{BS},\mu}(E)$ is at $E = \mathrm{Re}(\epsilon_\mu^{\mathrm{W}})$, and the FWHM is equal to $2\left[\delta - \mathrm{Im}(\epsilon_\mu^{\mathrm{W}})\right]$. The minimum FWHM, which is found for those GBSs with $\mathrm{Im}(\epsilon_\mu^{\mathrm{W}}) = 0$, is given by 2δ. In order to resolve reliably all the peaks, the needed energy mesh spacing Δ_E^δ for a given δ has to be of the order of δ. We now introduce the free parameter r_Δ, which sets Δ_E^δ for a given δ by

$$\Delta_E^\delta = r_\Delta \, \delta. \tag{6.80}$$

In our calculations we usually choose r_Δ somewhere in the range between 0.1 and 1, which guarantees that all the peaks in $\mathcal{N}_{\mathrm{BS}}(E)$ are resolved. This feature is the first important part of our adaptive mesh algorithm: no matter how sharp the peaks in the DOS are, by using a finite value of δ they are broadened, so that they can always be resolved with a regular mesh of spacing Δ_E^δ.

In order to obtain an adaptive mesh algorithm we now need to identify in which energy regions the mesh has to be refined. This information can be obtained from the value of $\mathcal{N}_{\mathrm{BS},\mu}(E)$. By definition the energy spacing Δ_E^δ is fine enough, in order to resolve peaks with a FWHM of 2δ. Therefore all the peaks generated by GBSs with $|\mathrm{Im}(\epsilon_\mu^{\mathrm{W}})| \geq \delta$ can be resolved even if $i\delta$ was not added to the energy. For such a GBS with $|\mathrm{Im}(\epsilon_\mu^{\mathrm{W}})| \geq \delta$, and using Eq. (6.79), we have

$$\pi \, \delta \, \mathcal{N}_{\mathrm{BS},\mu}(E) \leq \frac{1}{2} \qquad \text{for} \quad |\mathrm{Im}(\epsilon_\mu^{\mathrm{W}})| \geq \delta. \tag{6.81}$$

On the other hand, for those GBSs with $|\mathrm{Im}(\epsilon_\mu^{\mathrm{W}})| < \delta$, the mesh is not fine enough, and needs to be refined. For those GBSs with $|\mathrm{Im}(\epsilon_\mu^{\mathrm{W}})| < \delta$ we have

$$\pi \, \delta \, \mathcal{N}_{\mathrm{BS},\mu}(E) > \frac{1}{2}. \qquad \text{for} \quad |\mathrm{Im}(\epsilon_\mu^{\mathrm{W}})| < \delta. \tag{6.82}$$

In the same way, from Eq. (6.78) we can see that if the total $\pi\delta\mathcal{N}_{\mathrm{BS}}(E) \leq \frac{1}{2}$, then there is no state with $|\mathrm{Im}(\epsilon_\mu^{\mathrm{W}})| < \delta$ in a region of $\pm\delta$ around this energy. If $\pi\delta\mathcal{N}_{\mathrm{BS}}(E) > \frac{1}{2}$, then there might be a such a state within $\pm\delta$, so that the mesh has to be refined in that region. This allows us to define the following criterion for the mesh refinement: the mesh has to be refined at all those energies where

$$\pi \, \delta \, \mathcal{N}_{\mathrm{BS}}(E) > \frac{1}{2}. \tag{6.83}$$

More generally we can therefore introduce a refinement parameter r_δ, and require the mesh to be refined if

$$\pi \, \delta \, \mathcal{N}_{\mathrm{BS}}(E) > r_\delta. \tag{6.84}$$

In practical calculations we usually use $r_\delta = 1/2$, however the value can also be reduced.

state 1	state 2	state 3
$\mathrm{Re}(\epsilon_1) = -8.0$ eV	$\mathrm{Re}(\epsilon_2) = 0.0$ eV	$\mathrm{Re}(\epsilon_3) = 4.0$ eV
$\mathrm{Im}(\epsilon_1) = -2.0$ eV	$\mathrm{Im}(\epsilon_2) = -0.1$ eV	$\mathrm{Im}(\epsilon_3) = -0.001$ eV

Table 6.1: Parameters used in the 3 state example system.

This allows us to formulate the adaptive mesh algorithm as follows: we start by choosing values for r_Δ, r_δ and an initial value δ_0 for δ. Next we calculate $\mathcal{N}_{\mathrm{BS}}(E)$ on a mesh with spacing $\Delta_E^\delta = r_\Delta \delta$ using Eq. (6.7). For all those energies where $\pi\ \delta\ \mathcal{N}_{\mathrm{BS}}(E) > r_\delta$, we have to refine the mesh in an energy range of $E - \delta$ to $E + \delta$. The refined mesh spacing is given by the choice of a new, reduced value of $\delta = \delta_1$, so that now $\Delta_E^\delta = r_\Delta \delta_1$. This procedure is repeated until for a certain value of δ there are no energies with $\pi\ \delta\ \mathcal{N}_{\mathrm{BS}}(E) > r_\delta$. Once the energy mesh is determined, we can set $\delta = 0$ for all the calculations on that mesh. If BSs are present, then we can define a lowest value δ_{BS} for δ, below which the states are considered BSs (see Sec. 6.3). In these cases the mesh is refined until $\delta < \delta_{\mathrm{BS}}$. The occupation of the BSs has then to be set as described in Sec. 6.2. We note that this recursive algorithm can therefore also be used as an alternative way to locate the BSs.

We now illustrate the algorithm on an example system, in which there are only 3 states, each one is located at a different energy and possessing a different broadening. The values of the used parameters are given in Tab. 6.1. State 1 is a broad state, state 2 has an intermediate energy width, and state 3 is a rather localized state. For such a system we can write $\mathcal{N}_{\mathrm{CS}}(E)$ as

$$\mathcal{N}_{\mathrm{CS}}(E) = \frac{1}{\pi} \sum_{n=1}^{3} \frac{-\mathrm{Im}(\epsilon_n)}{[E - \mathrm{Re}(\epsilon_n)] + [\delta - \mathrm{Im}(\epsilon_n)]^2}, \tag{6.85}$$

$\mathcal{N}_{\mathrm{BS}}(E)$ as

$$\mathcal{N}_{\mathrm{BS}}(E) = \frac{1}{\pi} \sum_{n=1}^{3} \frac{\delta}{[E - \mathrm{Re}(\epsilon_n)] + [\delta - \mathrm{Im}(\epsilon_n)]^2}, \tag{6.86}$$

and the total DOS $\mathcal{N}_{\mathrm{M}}(E)$ as

$$\mathcal{N}_{\mathrm{M}}(E) = \frac{1}{\pi} \sum_{n=1}^{3} \frac{\delta - \mathrm{Im}(\epsilon_n)}{[E - \mathrm{Re}(\epsilon_n)] + [\delta - \mathrm{Im}(\epsilon_n)]^2}. \tag{6.87}$$

In Fig. 6.2 these three DOSs are shown for $\delta = 0.1$ eV. For the strongly coupled state 1 almost all the DOS is contained in $\mathcal{N}_{\mathrm{CS}}(E)$, whereas it is barely visible in

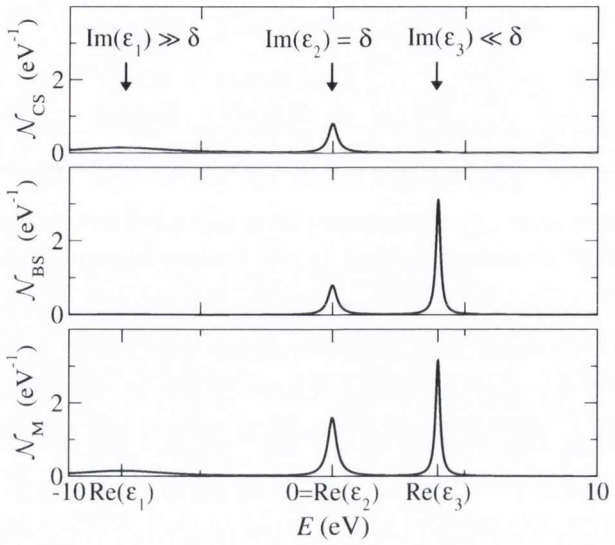


Figure 6.2: Contributions to the DOS for the 3 states example system with the parameters given in Tab. 6.1, for $\delta = 0.1$ eV.

$\mathcal{N}_{\mathrm{BS}}(E)$. For state 2 $\mathrm{Im}(\epsilon_2) = \delta$, so that its contributes with the same amplitude to both $\mathcal{N}_{\mathrm{CS}}(E)$ and $\mathcal{N}_{\mathrm{BS}}(E)$. State 3 finally has has $\mathrm{Im}(\epsilon_3) \ll \delta$, and therefore almost all the DOS is contained in $\mathcal{N}_{\mathrm{BS}}(E)$, whereas it is barely visible in $\mathcal{N}_{\mathrm{CS}}(E)$. This shows that as the absolute value of the imaginary part of the states decreases, the corresponding $\mathcal{N}_{\mathrm{CS}}(E)$ decreases, while $\mathcal{N}_{\mathrm{BS}}(E)$ increases. The FWHM of state 1 is approximately equal to $-2\mathrm{Im}(\epsilon_3)$, whereas for state 3 it is approximately equal to 2δ. This illustrates the broadening of the weakly coupled states induced by δ.

In Fig. 6.3 the adaptive mesh algorithm for this system is illustrated. On the left panel we plot $\pi\delta\mathcal{N}_{\mathrm{BS}}(E)$, which is used in order to detect where to refine the mesh via Eq. (6.84), for different values of δ. The dashed horizontal line corresponds to the chosen value of the refinement parameter $r_\delta = 0.5$. For those energies, where $\pi\delta\mathcal{N}_{\mathrm{BS}}(E) > r_\delta$ (black part of the curve), the mesh needs to be refined. For $\pi\delta\mathcal{N}_{\mathrm{BS}}(E) \leq r_\delta$ (green part of the curve), no refinement of the mesh is needed. The shaded area corresponds to the energy region where for a given δ the mesh has to be refined. For $\delta = 1$ eV no refinement is needed for state 1, however for states 2 and 3 the mesh needs to be refined in the two broad grey areas. For $\delta = 0.1$ eV the mesh is fine enough also for state 2, however for state 3 it still needs to be refined. The same is true for $\delta = 0.01$ eV. Only when δ reaches 0.001 eV, the energy mesh is fine enough also for state 3. At this stage the mesh refinement would stop. If δ is further reduced to 0.0001 eV, no peaks are visible on this scale in $\pi\delta\mathcal{N}_{\mathrm{BS}}(E)$. We note that as long as $\delta > -\mathrm{Im}(\epsilon_n)$ the height of the peaks is approximately constant, but the width of the peaks shrinks progressively as we reduce δ. Once $\delta < -\mathrm{Im}(\epsilon_n)$, then the

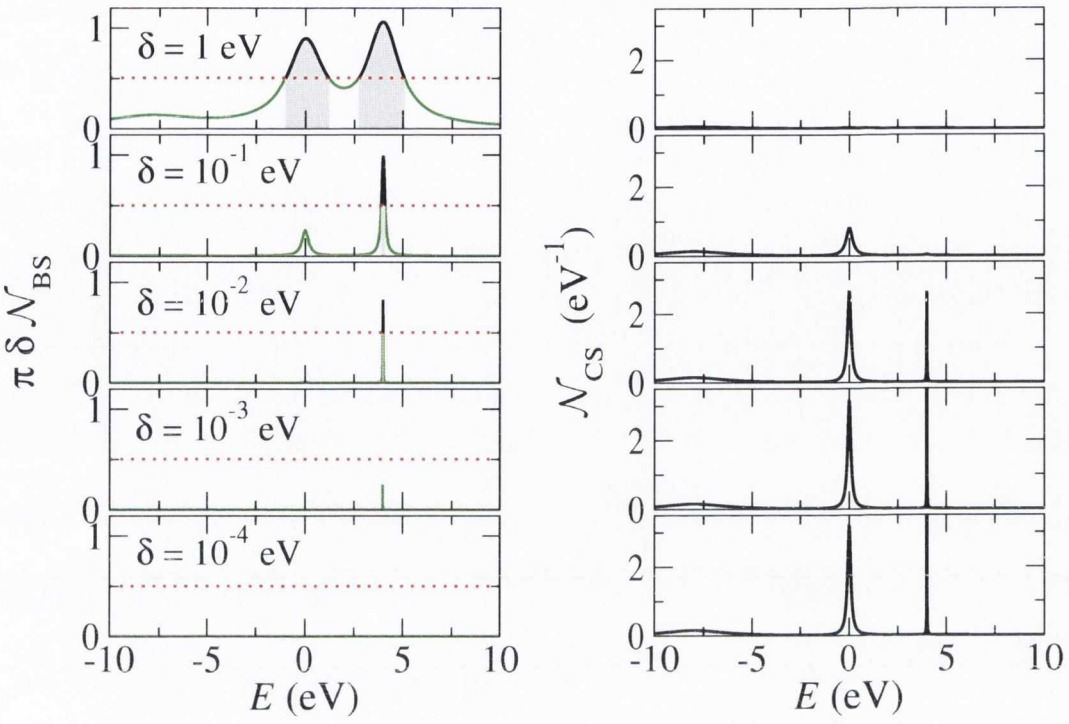


Figure 6.3: Left panel: $\delta\pi\mathcal{N}_{\rm BS}$, used in the criterion for the mesh refinement in Eq. (6.84), as function of energy E, for different values of the added imaginary part δ. Right panel: $\mathcal{N}_{\rm CS}$ as function of energy. The parameters of the 3 states example system are given in Tab. 6.1.

width stays approximately constant, but the height is reduced. In the right panel of Fig. 6.3 $\mathcal{N}_{\rm CS}(E)$ is shown for the same values of δ. It can be seen that reducing δ leads to an increase in $\mathcal{N}_{\rm CS}(E)$, until it is approximately constant once $\delta < 0.001$ eV.

Instead of calculating the full $\mathcal{N}_{\rm BS}(E)$, one can just calculate the dominant element in the sum of Eq. (6.78). The largest contribution comes from that GBS, for which at a given energy E the term $|E - \epsilon_\mu^{\rm W}|$ is minimal, so that $1/|E - \epsilon_\mu^{\rm W}|$ is maximized. We define $\epsilon_{\rm max} = \epsilon_{\mu_{\rm max}}^{\rm W}$, where $\mu_{\rm max}$ is the index of the GBS with the minimum value of $|E - \epsilon_\mu^{\rm W}|$ for a given energy. We can therefore equally use the quantity

$$\bar{\mathcal{N}}_{\rm BS}(E) = \frac{1}{\pi}\frac{\delta}{[E - {\rm Re}(\epsilon_{max})] + [\delta - {\rm Im}(\epsilon_{max})]^2} = \frac{1}{\pi}\frac{\delta}{|E - \epsilon_{max} + i\delta|^2} \quad (6.88)$$

instead of $\mathcal{N}_{\rm BS}(E)$ in the criterion for the mesh refinement [Eq. (6.84)]. If we define $\bar{g}_{\rm min}$ as the eigenvalue with the smallest complex modulus of the matrix $ES_{\rm M} - H_{\rm eff}$, then

$$\bar{\mathcal{N}}_{\rm BS}(E) = \frac{1}{\pi}\frac{\delta}{|\bar{g}_{\rm min} + i\delta|^2} \quad (6.89)$$

The advantage now is that in order to obtain $\bar{g}_{\min}$ it is not necessary to calculate all the eigenvalues of the matrix. There are numerical algorithms, such as the shifted Arnoldi scheme [253, 254], that allow the direct calculation of the eigenvalue with the smallest complex modulus. For large systems the calculation of $\bar{\mathcal{N}}_{\rm BS}(E)$ is therefore much faster than the one of $\mathcal{N}_{\rm BS}(E)$, so that this method should be more efficient. Moreover, in $\mathcal{N}_{\rm BS}(E)$ there is a background contribution coming from the strongly coupled states, with an amplitude proportional to δ, which grows for large systems. This problem is avoided when using $\bar{\mathcal{N}}_{\rm BS}(E)$, since by construction it depends only on one eigenvalue, and is therefore size-independent. For our test systems this method worked rather well, although we note that the recursive algorithms to calculate $\bar{g}_{\min}$ show convergence problems for certain energies.

6.7 Transmission through weakly coupled states

Using the GF in the form of Eq. (6.44), the transmission coefficient [Eq. (4.94)] can be written in its spectral representation as

$$T(E) = \sum_{n,m=1}^{N} \frac{1}{E+i\delta-\epsilon_n} \frac{1}{E-i\delta-\epsilon_m^*} \tilde{\psi}_n^\dagger \Gamma_{\rm L} \tilde{\psi}_m \psi_m^\dagger \Gamma_{\rm R} \psi_n. \tag{6.90}$$

For a system where the transmission is entirely through weakly coupled states, we can use the same approximations introduced in Sec. 6.4 for the evaluation of $\rho_{\rm neq}^{\rm W}$. We set the value of $\eta^{\rm W}$ in such a way, that all the weakly coupled states contributing to the transport satisfy Eq. (6.48), and they are therefore contained in the set of GBSs determined by such $\eta^{\rm W}$. If we apply the same approximations of Sec. 6.4, the resulting form for $T(E)$ is

$$T(E) = \sum_{\mu=1}^{N^{\rm W}} \frac{1}{E+i\delta-\epsilon_\mu^{\rm W}} \frac{1}{E-i\delta-\epsilon_\mu^{\rm W*}} \tilde{\psi}_\mu^{\rm W\dagger} \Gamma_{\rm L} \tilde{\psi}_\mu^{\rm W} \; \psi_\mu^{\rm W\dagger} \Gamma_{\rm R} \psi_\mu^{\rm W}. \tag{6.91}$$

with $\delta \to 0^+$. The finite value for δ leads to the same broadening of the peaks as in the DOS.

By using the definitions of $\gamma_{\{\rm L/R\},\mu}$ [Eq. (6.54)], and assuming that for the weakly coupled states $\psi_\mu^{\rm W} \approx \tilde{\psi}_\mu^{\rm W}$, in the limit $\delta = 0$ we can write the transmission as

$$T(E) = \sum_{\mu=1}^{N^{\rm W}} \frac{\gamma_{{\rm L},\mu} \; \gamma_{{\rm R},\mu}}{\left[E - {\rm Re}(\epsilon_\mu^{\rm W})\right]^2 + \frac{1}{4}\left[\gamma_{{\rm L},\mu} + \gamma_{{\rm R},\mu}\right]^2}. \tag{6.92}$$

This form of the transmission coefficient corresponds to the Bright-Wigner formula, and is often used in tight-binding models for the transmission through molecular

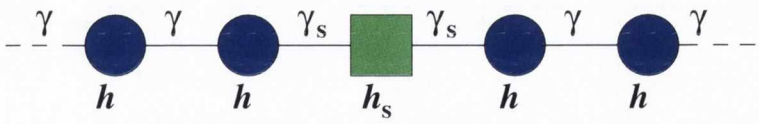


Figure 6.4: Schematic representation of the 1D tight-binding model.

states [54]. For a set of weakly coupled states, the transmission written in the form of Eq. (6.92) generally gives a better understanding of the transport properties than the general NEGF of Eq. (4.94). In particular, one can determine the nature of the molecular states contributing to $T(E)$, as well as the strength of the coupling of these states to the leads.

For one single state the maximum transmission is

$$T\left[\mathrm{Re}(\epsilon_\mu^{\mathrm{W}})\right] = 4\frac{\gamma_{\mathrm{L},\mu}\gamma_{\mathrm{R},\mu}}{(\gamma_{\mathrm{L},\mu}+\gamma_{\mathrm{R},\mu})^2}, \tag{6.93}$$

which is equal to 1 if $\gamma_{\mathrm{L},\mu} = \gamma_{\mathrm{R},\mu}$, and becomes smaller as the difference between $\gamma_{\mathrm{L},\mu}$ and $\gamma_{\mathrm{R},\mu}$ gets larger. The height of a peak in the transmission is therefore a measure of the asymmetry of the coupling to the leads. The FWHM is equal to $\gamma_{\mathrm{L},\mu} + \gamma_{\mathrm{R},\mu} = -2\ \mathrm{Im}(\epsilon_\mu)$. The integral over energy of the transmission gives the saturation current $I_{\mu,max}$ of the state, which is

$$I_{\mu,max} = \frac{2\pi e}{h}\frac{\gamma_{\mathrm{L},\mu}\gamma_{\mathrm{R},\mu}}{\gamma_{\mathrm{L},\mu}+\gamma_{\mathrm{R},\mu}}. \tag{6.94}$$

In Secs. 6.9 and 6.10 we verify the validity of the approximate model for the transmission, and find good agreement between the model and the exact solution.

For a BS the current is obtained by taking the limit $\gamma_{\mathrm{L/R},\mu} \to 0$ in Eq. 6.94, which gives $I_{\mu,max} = 0$. BSs therefore do not carry current. However if an effective coupling of the BSs to the leads is included, the current through these states is finite. This approach is taken in Refs. [246, 247, 248, 249, 243], and it is shown that in cases where the current is very small, the contribution of the BSs to the total current is not negligible. For some of the systems it is shown that the main contribution to the total current comes from the BSs.

6.8 1D-chain tight-binding model

In this section we illustrate the basic properties of the BSs, as presented in Secs. 6.1 and 6.3, using a simple tight-binding model. Such a model is also used chapter 6 of Refs. [194, 65] to illustrate the zero-bias properties of BSs due to impurities. The model system consists of an infinite 1D chain of atoms (Fig. 6.4). We note that in

this section all quantities are given in arbitrary units. The leads are described by one orbital per unit cell, with an onsite Hamiltonian $H_0 = h$ and hopping matrix $H_1 = H_{-1} = \gamma$ ($\gamma \in \mathbb{R}$). The EM consists of a single atom with $H_{\mathrm{M}} = h_{\mathrm{s}}$ and $H_{\mathrm{ML}} = H_{\mathrm{MR}} = \gamma_{\mathrm{s}}$. The orbitals are all assumed to be orthogonal, so that the overlap matrix S_{M} is the identity matrix. The dispersion relation $E(k)$ for the leads then is

$$E = h + 2\gamma \cos k, \tag{6.95}$$

with $k \in (-\pi, \pi]$. The inverse of the dispersion relation $k(E)$ is

$$\cos k = \frac{E - h}{2\gamma}. \tag{6.96}$$

For each E there are two k-points, describing left- and right-going states respectively. If we define $x = \frac{E-h}{2|\gamma|}$, and using the results of chapter 5, the self-energies for this system are

$$\Sigma_{\mathrm{L}}(E) = \Sigma_{\mathrm{R}}(E) = \Delta - i\frac{\Gamma}{2}, \tag{6.97}$$

where

$$\Delta = \frac{\gamma_{\mathrm{s}}^2}{|\gamma|}\left[x - \mathrm{sgn}(x)\sqrt{x^2 - 1}\;\theta(|x| - 1)\right] \tag{6.98}$$

is the real part, and

$$\Gamma = 2\frac{\gamma_{\mathrm{s}}^2}{|\gamma|}\sqrt{1 - x^2}\;\theta(1 - |x|) \tag{6.99}$$

is the imaginary part of the self-energy. The function $\theta(x)$ is the Heaviside step function. Fig. 6.5 shows Δ and Γ as function of E, for the set of parameters $\gamma = 1, h = 0, \gamma_{\mathrm{s}} = 1$. The real part of the self-energy increases linearly for $|x| < 1$, and decays proportional to x^{-2} for large x. The imaginary part is always smaller or equal to zero, and is non-zero only for $|x| < 1$. The GF of the EM then is

$$G_{\mathrm{M}} = \frac{1}{E + i\delta - h_{\mathrm{s}} - \Sigma_{\mathrm{L}} - \Sigma_{\mathrm{R}}}, \tag{6.100}$$

with $\delta \to 0^+$. The energy-dependent eigenvalue ϵ of the effective Hamiltonian $H_{\mathrm{eff}} = h_{\mathrm{s}} + \Sigma_{\mathrm{L}} + \Sigma_{\mathrm{R}}$ is

$$\epsilon = h_{\mathrm{s}} + \Sigma_{\mathrm{L}} + \Sigma_{\mathrm{R}}. \tag{6.101}$$

In the remaining part of this section we set $h = 0$, which corresponds to a shift of the zero of the energy scale to h, and $\gamma = 1$, which sets the energy scale. We evaluate the GF and SF for different values of h_{s} and γ_s, first at zero-bias, and then at finite bias. One of the main results at zero bias is that, except for $\gamma_s = \gamma$ and $h_{\mathrm{s}} = h$, there are always BSs in the system, as also shown in Ref. [194]. At finite bias the BSs appear and disappear, depending on the bias voltage.

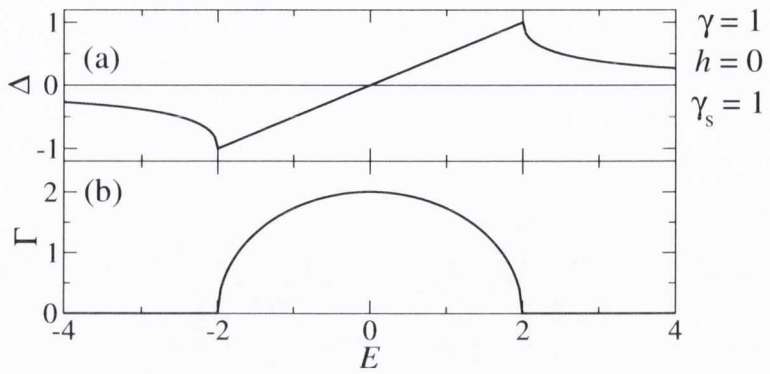


Figure 6.5: Real (Δ) and imaginary (Γ) components of the self-energy for the 1D model system [Eq. (6.97)].

6.8.1 Zero bias

At zero bias we have $\Sigma_\mathrm{L} = \Sigma_\mathrm{R}$, and the GF of the EM becomes

$$G_\mathrm{M} = \begin{cases} \frac{1}{(1-\gamma_\mathrm{s}^2)E - h_\mathrm{s} + i\gamma_\mathrm{s}^2\sqrt{4-E^2}+i\delta} & |E| \leq 2 \\ \frac{1}{(1-\gamma_\mathrm{s}^2)E - h_\mathrm{s} + \mathrm{sgn}(E)\gamma_\mathrm{s}^2\sqrt{E^2-4}+i\delta} & |E| \geq 2 \end{cases} . \tag{6.102}$$

The different behavior for $E \leq 2$ and $E \geq 2$ is caused by the Heaviside step function in the SEs. Using the definition in Eq. (4.30), the SF of the EM then results to

$$A_\mathrm{M} = \begin{cases} \frac{2\gamma_\mathrm{s}^2\sqrt{4-E^2}+2\delta}{\left[(1-\gamma_\mathrm{s}^2)E - h_\mathrm{s}\right]^2 + \left(\gamma_\mathrm{s}^2\sqrt{4-E^2}+\delta\right)^2} & |E| \leq 2 \\ \frac{2\delta}{\left[(1-\gamma_\mathrm{s}^2)E - h_\mathrm{s} + \mathrm{sgn}(E)\gamma_\mathrm{s}^2\sqrt{E^2-4}\right]^2 + \delta^2} & |E| \geq 2 \end{cases} . \tag{6.103}$$

Special case $h_\mathrm{s} = h$ and $\gamma_\mathrm{s} = \gamma$

If $h_\mathrm{s} = h$ and $\gamma_\mathrm{s} = \gamma$ the system corresponds to an infinite periodic chain of atoms. For this special case the SF becomes

$$A_\mathrm{M} = \begin{cases} \frac{2}{\sqrt{4-E^2}+\delta} & |E| \leq 2 \\ \frac{2\delta}{\mathrm{sgn}(E)(E^2-4)+\delta^2} & |E| \geq 2 \end{cases} . \tag{6.104}$$

Taking the limit $\delta \to 0^+$ we get

$$A_\mathrm{M} = \frac{2}{\sqrt{4-E^2}}\,\theta\left(4-E^2\right). \tag{6.105}$$

With Eqs. (6.2) and (6.102) we can calculate $A_\mathrm{CS} = iG(\Gamma_\mathrm{L} + \Gamma_\mathrm{R})G^\dagger$, the result for $\delta \to 0^+$ is identical to the right-hand side of above equation. For A_BS we obtain $A_\mathrm{BS} = 0$, so that there are no BSs in the system and as a result $A_\mathrm{M} = A_\mathrm{CS}$. This is the only case where there are no BSs in the system. By changing the parameters from this special case we can see how bound states appear in general.

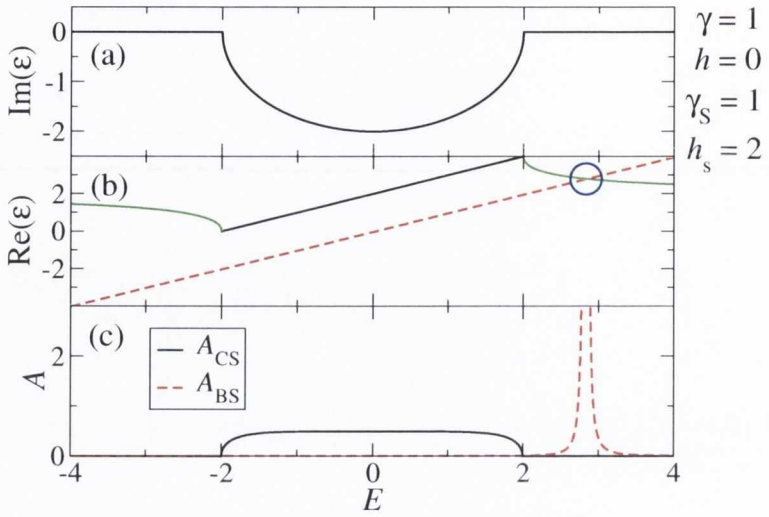


Figure 6.6: Imaginary (a) and real (b) part of the eigenvalue of the effective Hamiltonian ϵ, and (c) contributions to the SF, as function of energy E, at zero bias. The dashed red line in (b) is a plot of E as function of E itself [$f(E) = E$], the blue circle indicates the energy of the BS, and the green part of the curve corresponds to the range of eigenvalues where $\mathrm{Im}(\epsilon) = 0$.

Special case $h_s \neq h$ and $\gamma_s = \gamma$

For $\gamma_s = \gamma$, but $h_s \neq h$, the eigenvalue ϵ of H_{eff} is

$$\epsilon = \begin{cases} h_s + E - i\sqrt{4 - E^2} & |E| \leq 2 \\ h_s + E - \mathrm{sgn}(E)\sqrt{E^2 - 4} & |E| \geq 2 \end{cases} . \tag{6.106}$$

In Fig. 6.6(a) and 6.6(b) the imaginary and real parts of ϵ are shown as function of energy. The general behavior of ϵ is largely determined by the behavior of the self-energies (see Fig. 6.5). $\mathrm{Im}(\epsilon)$ is zero except for $E \in (-2, 2)$. A BS is found at an energy E_{BS} where $\mathrm{Re}\,[\epsilon(E_{\mathrm{BS}})] = E_{\mathrm{BS}}$ and $\mathrm{Im}\,[\epsilon(E_{\mathrm{BS}})] = 0$ [see Eqs. (6.45) and (6.46)]. In Fig. 6.6(b) the BS is found graphically, in the way described in Sec. 6.3, for $h_s = 2$. First those regions of $\mathrm{Re}\,[\epsilon(E)]$, where $\mathrm{Im}\,[\epsilon(E)] = 0$, are marked with a green color. Then all the crossings between the function $f(E) = E$ (dashed red line) and the green part of $\mathrm{Re}\,[\epsilon(E)]$ are determined. These correspond to the energy of the BSs. For the chosen set of parameters a BS is found at $E_{\mathrm{BS}} \approx 2.83$. From this graphical solution one can see that there is always a BS in the system, as long as $h_s \neq h$, because there is always a crossing between the two curves in the green part. If the curves cross in one of the black regions of $\mathrm{Re}\,[\epsilon(E)]$ the first condition [Eq. (6.45)] is fulfilled, but not the second one [Eq. (6.46)], so that these states are coupled to the leads.

By using Eq. (6.103) the SF becomes

$$A_{\mathrm{M}} = \begin{cases} \frac{2\sqrt{4-E^2}+2\delta}{h_{\mathrm{s}}^2+\left(\sqrt{4-E^2}+\delta\right)^2} & |E| \leq 2 \\ \frac{2\delta}{\left[-h_{\mathrm{s}}+\ \mathrm{sgn}(E)\sqrt{E^2-4}\right]^2+\delta^2} & |E| \geq 2 \end{cases}, \tag{6.107}$$

and taking the limit $\delta \to 0^+$ we get

$$A_{\mathrm{M}} = \begin{cases} \frac{2\sqrt{4-E^2}}{h_{\mathrm{s}}^2+(4-E^2)} & |E| \geq 2 \\ \frac{2\pi}{|1-\beta_0|}\delta(E-E_{\mathrm{BS}}), & |E| \geq 2 \end{cases}. \tag{6.108}$$

Here $E_{\mathrm{BS}} = \mathrm{sgn}(h_{\mathrm{s}})\sqrt{4+h_{\mathrm{s}}^2}$ and $\beta_0 = \partial \mathrm{Re}(\epsilon)/\partial E|_{E=\epsilon_{\mathrm{BS}}} = 1 - \sqrt{4+h_{\mathrm{s}}^2}/|h_{\mathrm{s}}|$ [with $\mathrm{Re}(\epsilon)$ taken from Eq. (6.106)]. For any $h_{\mathrm{s}} \neq 0$ therefore there is a BS. Again both A_{CS} and A_{BS} can be evaluated explicitly to be

$$A_{\mathrm{CS}} = \frac{2\sqrt{4-E^2}}{h_{\mathrm{s}}^2+(4-E^2)}\Theta\left(4-E^2\right) \tag{6.109}$$

$$A_{\mathrm{BS}} = \frac{2\pi}{|1-\beta_0|}\delta(E-E_{\mathrm{BS}}), \tag{6.110}$$

and it can indeed be verified that for this simple model $A_{\mathrm{M}} = A_{\mathrm{BS}} + A_{\mathrm{CS}}$. Moreover, the form of A_{BS} corresponds to the general form given in Eq. (6.67), if applied to this model. In Fig. 6.6(c) A_{CS} and A_{BS} are shown. A small finite value of δ is used in order to visualize the δ-function in A_{BS}. A_{CS} is smooth and flat over a wide energy-range, equal to the band-width of the leads, whereas A_{BS} is localized at one particular energy. This is the typical behavior of these two contributions to the total SF. In Fig. 6.7(a) E_{BS} is shown as function of h_{s}. For all values of h_{s} there is a BS, located at an energy outside the bandwidth of the leads. For $h_{\mathrm{s}} = 0$ there is no BS, this is reflected by the discontinuity in the curve at $h_{\mathrm{s}} = h_0$.

We can integrate the SF over energy to obtain the maximum number of electrons q_{CS} that can be associated to A_{CS} [Eq. (6.9)]

$$q_{\mathrm{CS}} = 1 - \frac{|h_{\mathrm{s}}|}{\sqrt{h_{\mathrm{s}}^2+4}}, \tag{6.111}$$

and the maximum number of electrons q_{BS} associated to A_{BS} [Eq. (6.10)]

$$q_{\mathrm{BS}} = \frac{|h_{\mathrm{s}}|}{\sqrt{h_{\mathrm{s}}^2+4}}. \tag{6.112}$$

The total number of electrons that can be placed in the EM is $q_{\mathrm{M}} = q_{\mathrm{CS}} + q_{\mathrm{BS}} = 1$. In Fig. 6.7(b) q_{CS} and q_{BS} are shown as function of h_{s}. q_{CS} is 1 for $h_{\mathrm{s}} = 0$ and decreases monotonically for increasing h_{s}. On the other hand, q_{BS} is 0 for $h_{\mathrm{s}} = 0$, reflecting the

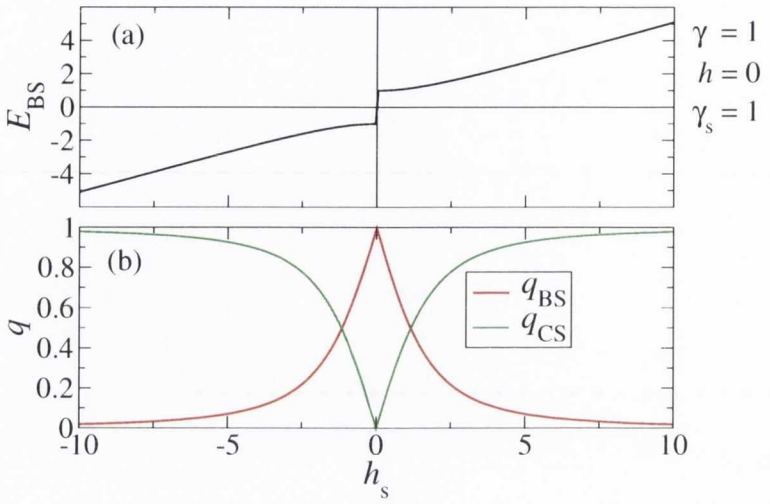


Figure 6.7: (a) Energy of the BS E_{BS} as function of the onsite Hamiltonian of the EM h_{s}, and (b) maximum number of electrons in the coupled states q_{CS} and bound states q_{BS}, as function of energy E.

fact that in this case there is no BS. Then it increases monotonically for increasing h_{s}. The distribution of charge between q_{BS} and q_{CS} is determined by β_0. It is therefore essential to include this factor in the computations, in order to obtain the correct total charge. We note that in Ref. [194] the Van Hove singularities for $h_{\mathrm{s}} = 0$ at $E = \pm 2$ are interpreted as BSs. The reason why we do not include them among the BSs, is that $q_{\mathrm{BS}} = 0$ for this special case.

Special case $\gamma_{\mathrm{s}} = 0$

By using Eq. (6.103), the SF for $\gamma_{\mathrm{s}} = 0$ becomes

$$A_{\mathrm{M}} = \frac{\delta}{(E - h_{\mathrm{s}})^2 + \delta^2}, \tag{6.113}$$

which in the limit $\delta \to 0^+$ is

$$A_{\mathrm{M}} = 2\pi\, \delta(E - h_{\mathrm{s}}). \tag{6.114}$$

This corresponds to a completely decoupled, localized state at energy h_{s}. This can always contain one electron regardless of the value of h_{s}.

6.8.2 Finite bias

A finite bias potential V is applied non-self-consistently by simply shifting the onsite energies of the leads by $+V/2$ on the left-hand side and $-V/2$ on the right-hand side. The GF at finite bias is then given by

$$G(E, V) = \frac{1}{E + i\delta - h_{\mathrm{s}} - \Sigma_{\mathrm{L}}(E - V/2) - \Sigma_{\mathrm{R}}(E + V/2)}. \tag{6.115}$$

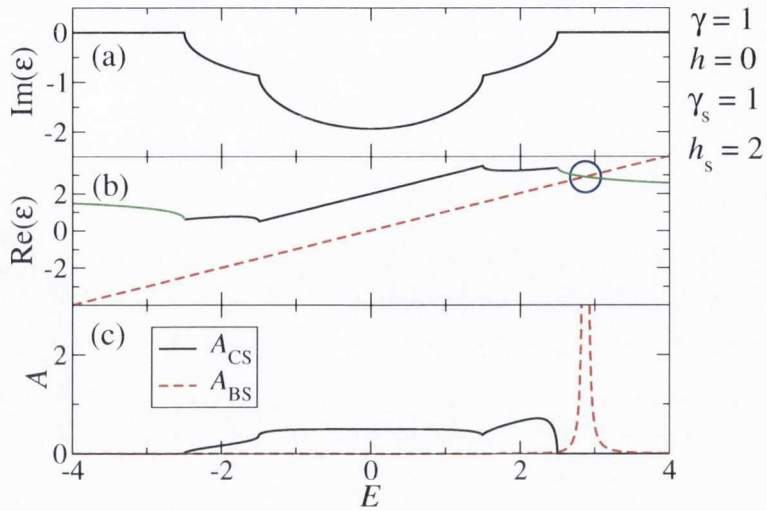


Figure 6.8: Imaginary (a) and real (b) part of the eigenvalue ϵ of the effective Hamiltonian, and (c) the two contributions to the SF, as function of energy E, for an applied bias potential $V = 1$. The dashed red line in (b) is a plot of E as function of E itself $[f(E) = E]$, the blue circle indicates the energy of the BS, and the green part of the curve corresponds to the range of eigenvalues where $\mathrm{Im}(\epsilon) = 0$.

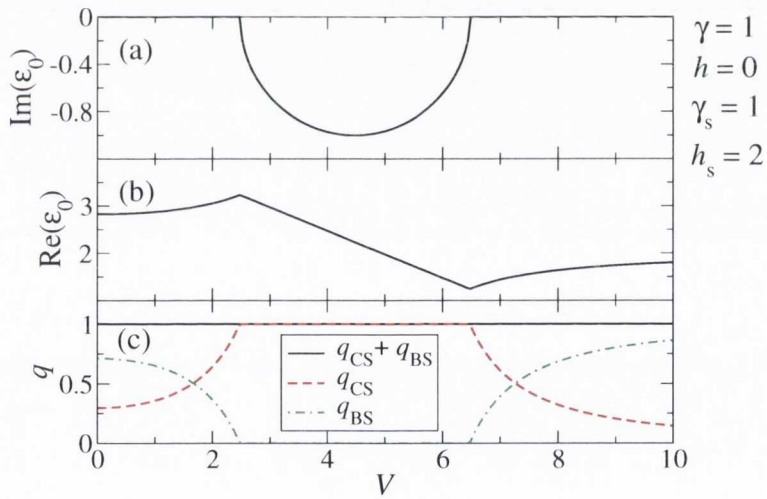


Figure 6.9: Real (a) and imaginary (b) parts of the eigenvalue ϵ_1 of the BS $[\epsilon_1 = \epsilon(E_{\mathrm{BS}})]$, and (c) maximum number of electrons storable respectively in the coupled states q_{CS} and bound states q_{BS}, as function of the applied bias potential V.

We now analyze the special case where $h_s \neq h$ and $\gamma_s = \gamma$, in order to understand how the contribution of the bound states to the total DOS changes with bias. The GF and SGF are computed numerically. In Figs. 6.8(a) and 6.8(b) the imaginary and real part of ϵ are shown for an applied bias of $V = 1$. The energy range where $\mathrm{Im}(\epsilon) \neq 0$ is now increased by V. There is still a BS, but E_{BS} now moves closer to the band-edge of one of the leads. This causes the factor β_0 to increase, so that the number of electrons that can be placed in the BS is reduced. Also the SF [Fig. 6.8(c)]

shows an increase of the range where $A_{\rm CS} \neq 0$.

In Figs. 6.9(a) and 6.9(b) the imaginary and real part of $\epsilon_1 = \epsilon(E_{\rm BS})$ are shown as function of bias voltage. For small and large voltages $\mathrm{Im}(\epsilon_1) = 0$, so that there is a BS. For intermediate bias voltages, between 2.4 and 6.3, $\mathrm{Im}(\epsilon_1) < 0$, and therefore there is no BS. The energy of the bound state $\mathrm{Re}(\epsilon_1)$ changes with the bias, especially for those bias voltages where $\mathrm{Re}(\epsilon_1)$ is close to a band-edge of one of the leads. The number of electrons that fits in the BS $q_{\rm BS}$ [Fig. 6.9(c)] changes continuously with the bias, and it is zero when there is no BS. This result shows that for a finite bias calculation the contribution of the BSs to the density matrix has to be calculated independently at each bias. BSs can appear and disappear, with the amount of charge stored in the BSs changing continuously with bias. We also point out the importance of the prefactor $1/|1-\beta_0|$ in Eq. (6.110), and of the corresponding factor $1/|1-\beta_\mu^{\rm W}|$ in the general case (Sec. 6.4), since it determines how much charge can be placed into the BS. If this prefactor is neglected, then $q_{\rm BS}$ is either 0 if there is no BS, or 1 if there is a BS. This however often leads to a wrong total charge in the EM, which is the case for example if the BS is close to the band-edge of one of the leads.

6.9 C_3 molecule in an Au capacitor

In this section we perform a finite bias calculation of a small test system, consisting of a parallel plate gold capacitor with a weakly coupled C_3 molecule placed in the middle (see Fig. 6.10). The goal is to test our implementation of the BSCS (Sec. 6.5) and of the mesh refinement algorithm (Sec. 6.6) in SMEAGOL. The distance between molecule and Au surface is chosen in such a way that the coupling is very weak. Moreover, it is slightly shifted to the left side, so that although the coupling is weak to both surfaces, it is much stronger to the left than the right electrode. In this way the energy levels of the molecule should mainly follow the Fermi energy of the left lead. Although this is an artificial system, its general features apply to all systems presenting weakly coupled states.

In order to reduce the size of the calculation for the gold electrodes, we include the Au d orbitals in the core. In the context of electronic transport this is a rather good approximation, as long as the bias is not too large [255]. The main reason is that close the the Fermi energy the contribution from the Au d orbitals is negligible. We use a minimal single ζ s basis for the Au atoms, with a cutoff radius of $r_c = 5.5$ bohr. This is a rather short cutoff, however this is only a proof of concept, and the aim is not to reproduce a real physical system. For the carbon we use double ζ s, and double ζ polarized p, with $r_c = 6.5$ bohr for the first ζ for both s and p

Figure 6.10: Unit cell of the parallel plate gold capacitor, with a C_3 molecule between the two plates.

orbitals. An electronic temperature of 300 K and a 2×2 k-points mesh are used in the plane of the capacitor. We perform a spin-polarized calculation, and use the LSDA approximation for the exchange correlation potential. The lattice parameters of the Au(111) surface correspond to those used in Sec. 5.4. The atomic spacings between the carbon atoms in the C_3 molecule are equal to 1.30 Å, which corresponds to our calculated value at equilibrium for the isolated molecule. This value is slightly larger than the experimental distance of 1.277, but agrees rather well with previous calculations Å [256]. The energy gap between the HOMO and the LUMO, obtained from a calculation with the isolated molecule, is equal to 2.22 eV, and compares well to previous GGA calculations [257]. The value obtained using CI however is much larger, equal to 12.4 eV [257].

In Fig. 6.11 the zero bias spin-dependent transmission coefficient is shown. The molecular states can clearly be identified as sharp peaks in the transmission. Due to the interaction with the Au surface, which leads to a very small charging of the molecule, the molecular states are slightly spin-split. The LUMO is spin-split and pinned slightly above E_{F}, the HOMO is at about -2.2 eV below E_{F}. Since an electronic temperature of 300 K is used, the step in the Fermi function is smeared out by about 25 meV, so that both the majority and minority LUMO levels are partially filled.

The GBSs formalism developed in this chapter is used to find and characterize the weakly coupled states. We solve the eigenvalue problem for the effective Hamiltonian [Eq. (6.41)], and find all the GBSs with $-\mathrm{Im}(\epsilon_{\mu}^{\mathrm{W}}) < 10^{-3}$ Ry. In Fig. 6.12 the total transmission is shown on a logarithmic scale. The diamonds on the energy axis correspond to the values of the real part of the energy eigenvalue for all the found GBSs. It can be seen that for each peak in the transmission a GBS is found. This confirms that with the method described in Sec. 6.3 we can indeed find all the weakly coupled states at the right energies. The GBSs found slightly below 4 eV do not correspond to molecular states, but to surface states of the gold. Since these are

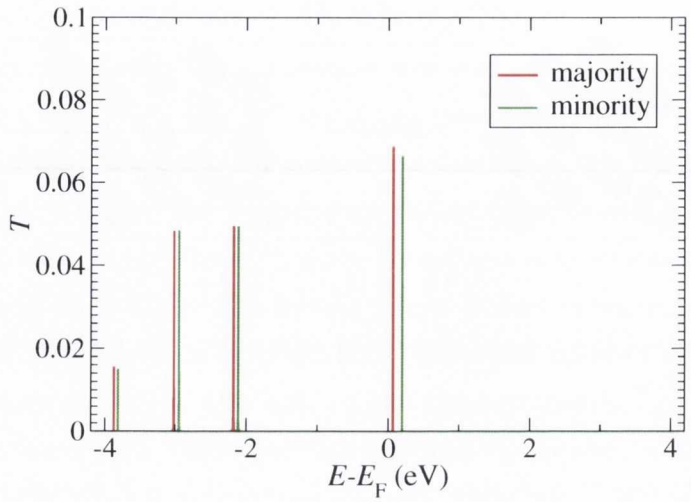


Figure 6.11: Zero bias spin-dependent transmission coefficient T as function of energy.

almost not coupled to the lead on the opposite side, the height of the transmission peaks is very small. In fact for the set of GBSs at about 3.8 eV above E_{F} there is no visible peak in the transmission. A closer analysis shows that these GBSs are only found for one out of the 4 k-points, namely for the Γ-point. The coupling of these states to the leads is however very weak, of the order of the machine precision, and also highly asymmetric, so that the peaks do not show up on the scale used for the figure. These peaks with very small amplitude can be found in the transmission for the Γ point alone at the energies of the GBSs. For each weakly coupled state with index μ the effective coupling to the leads $\gamma_{\mathrm{L},\mu}$ and $\gamma_{\mathrm{R},\mu}$ is extracted using Eq. (6.54). The model transmission coefficient is then calculated using Eq. (6.92), and the result is shown as dashed blue curve in Fig. 6.12. The agreement with the exact transmission is very good, all the peaks have the same position, height and FWHM. Away from the peaks the agreement is not so good, since in this region the approximations made in deriving Eq. (6.92) are not valid. In order to emphasize the agreement between the exact solution and the model, in the inset we show the transmission on a linear scale in an energy range close to the LUMO. The circles indicate the exact transmission, whereas the dashed line is the transmission calculated using the model. The fact that the two curves basically overlap confirms that for such a weakly coupled system the model for the transmission described in Sec. 6.7 is indeed appropriate.

In order to test the validity of the BSCS, presented in Sec. 6.3, we perform a finite bias calculation, where we use $\delta = 10^{-4}$ Ry and apply the BSCS to correctly populate the GBSs. Since for this system the agreement between the model transmission and the exact transmission at zero bias is very good, we expect the BSCS to work well. In order to integrate the non-equilibrium part of the lesser GF over energy we use an

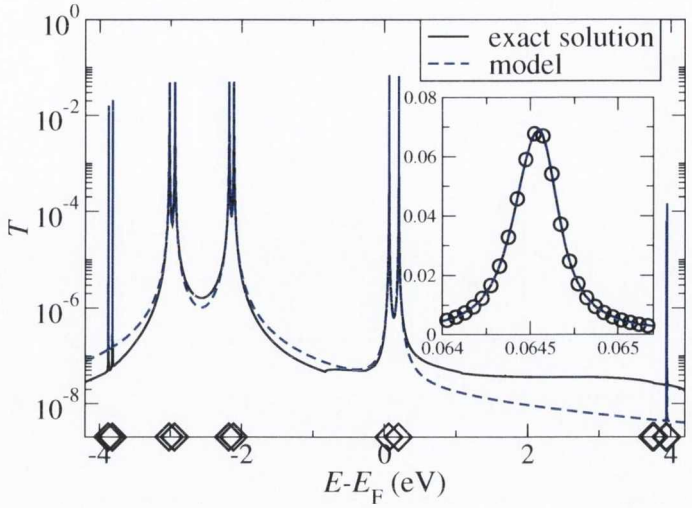


Figure 6.12: Zero bias total transmission coefficient T as function of energy. The black curve is the exact solution, whereas the dashed blue curve is obtained by using the model transmission of Eq. (6.92). The inset shows a zoom over a narrow energy range, where the black circles indicate the exact transmission. The diamonds plotted on the energy axis indicate the energies, at which a GBSs is found.

energy mesh-point spacing of about 8 10^{-5} Ry, so that we can resolve with reasonable accuracy the peaks with a minimal FWHM of $2\delta = 2\ 10^{-4}$ Ry. In Fig. 6.13(a) the obtained value for $\mathrm{Re}(\epsilon_\mu^{\mathrm{W}})$, corresponding to the position of the molecular levels, is shown as function of bias (black curves) for the self-consistent solution. At zero bias the energy scale is chosen in such a way that $E_{\mathrm{F}} = 0$. The dashed red line indicates the position of $E_{\mathrm{F,L}}$ as function of bias, the green one that of $E_{\mathrm{F,R}}$. The states follow almost exactly the left Fermi level, with the LUMO pinned above it for all bias voltages, so that the energy of the levels increases linearly with the bias. The only states that do not follow this trend are the surface states of the gold, located at about 4 eV above E_{F} at zero bias. The surface states located on the left-hand side lead follow $E_{\mathrm{F,L}}$, those located on the right-hand side one follow $E_{\mathrm{F,R}}$, so that their energy decreases with increasing bias. At a voltage of about 2 V the HOMO levels enter the bias window. They can easily enter the bias window, since their charge is approximately 1 even when they are inside the bias window. The reason for this is that, being coupled much stronger to the left lead, the flow of charge from the C_3 into the right lead is much slower than the inflow from the left lead, so that the occupation of the levels is always approximately one. Since charging is negligible, the relative position of the levels in the molecule is constant.

We now compare the results with the ones of a second calculation, where we do not use the BSCS, but we set $\delta = 0$. We then use a very fine mesh of energy points on the

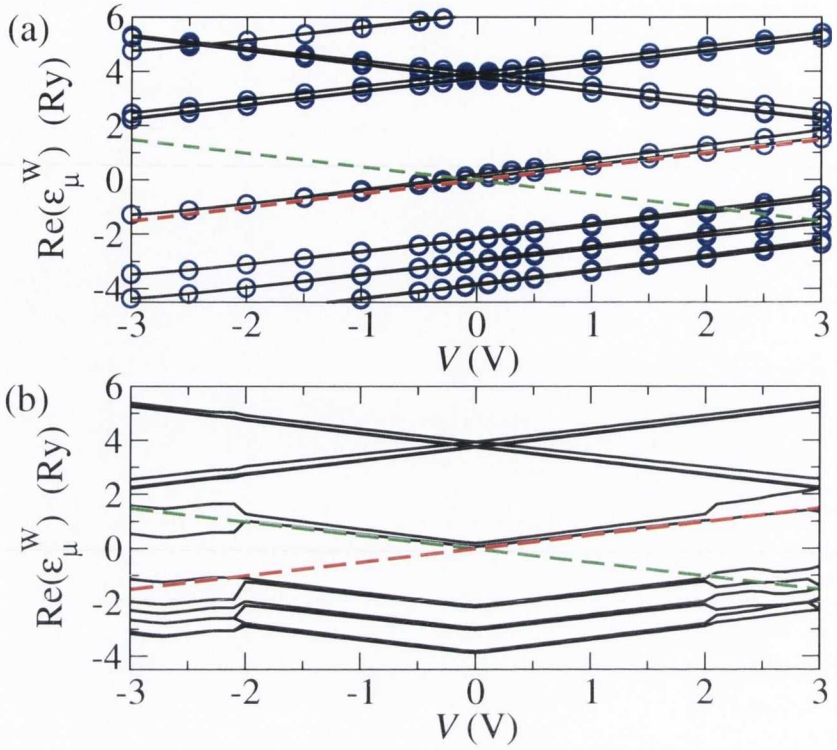


Figure 6.13: Real part of the GBSs eigenvalues $\mathrm{Re}(\epsilon_\mu^{\mathrm{W}})$ (black lines), as function of the voltage V. The red (green) dashed line indicates the position of $E_{\mathrm{F,L}}$ ($E_{\mathrm{F,R}}$). The black lines in (a) are calculated from a self-consistent solution, obtained by using the BSCS, with $\delta = 10^{-4}$ Ry. In (b) the self-consistent solution is obtained without including the BSCS, with $\delta = 10^{-4}$. The blue circles in (a) indicate $\mathrm{Re}(\epsilon_\mu^{\mathrm{W}})$ for a self-consistent solution, obtained with $\delta = 0$.

real axis, with a mesh-point spacing of about 5 10^{-6} Ry. The FWHM of the HOMO and LUMO peaks is about 2 10^{-5} Ry, so that such a fine mesh can resolve the sharp peaks, and therefore calculate the proper occupation exactly. Once self-consistency is achieved, we calculate the energies of the GBSs. These are shown as blue circles in Fig. 6.13(a). They overlap very well with the positions of the GBSs obtained from the self-consistent calculation with a finite δ and the BSCS. This clearly demonstrates that the BSCS proposed in Sec. 6.5 gives the correct solution at finite bias when a finite value for δ is used. We note that for this system the computational time needed when using the BSCS scheme is much shorter than the one for the exact calculation with the high mesh-point density.

As last test the self-consistent solution at finite bias is calculated by using $\delta = 10^{-4}$ Ry, but without correcting for the GBSs. This means that we neglect the term $\rho_{\mathrm{neq}}^{\mathrm{BS}}$ in Eq. (6.32). Again we calculate the positions of the GBSs for the self-consistent solution, and the result is shown in Fig. 6.13(b) (black curves). It can be clearly seen that the position of the levels as function of bias is very different from the correct

one of Fig. 6.13(a). This is due to the fact that the charging of the molecular orbitals is not correctly taken into account. As discussed in Sec. 6.2, neglecting the BSs correction term is equivalent to assuming that the weakly coupled states are equally coupled to both leads. For such a system all the levels avoid entering the bias window, since that would imply change their occupation by about half of an electron. In fact the levels remain outside the bias window at all voltages between -2 V to +2 V. Outside this voltage range, some levels are forced to enter the bias window, since the bias is larger than the HOMO-LUMO gap. This then leads to a rearrangement of the levels at higher voltage, with the effect that the molecule remains in a nearly neutral charging state. Whereas this represents the correct behavior for a symmetrically coupled system, in this case this result is clearly incorrect, since the coupling is constructed to be highly asymmetric. This shows that neglecting the BSs contribution to the density matrix $\rho_{\mathrm{neq}}^{\mathrm{BS}}$ can lead to very unphysical results, also for the I-V. We therefore conclude that for all transport calculations at finite bias it is necessary to check whether there are very weakly coupled states in the system. If such states are present, one can either use a very fine energy mesh and $\delta = 0$, or else use a finite value for δ, with a corresponding coarser mesh, and add the BSs correction term. We also performed the calculation with $\delta = 0$, and using the adaptive mesh algorithm described in Sec. 6.6. The obtained results are basically identical to the ones obtained using a very fine regular mesh, confirming that the mesh refinement algorithm works well in these situations. The computational time is much faster if the adaptive mesh is used.

We conclude this section by noting that the calculation of the GBSs is not only useful to obtain the correct self-consistent solution. It can also provide us several additional informations about the nature of the molecular levels, about their coupling to the leads, and about their occupation. In Fig. 6.14 the evolution with bias of $\mathrm{Im}(\epsilon_\mu^{\mathrm{W}})$, $\gamma_{\mathrm{L},\mu}$, $\gamma_{\mathrm{R},\mu}$ and the occupation $\bar{p}_\mu$ [Eq. (6.75)] is shown for all GBSs within the energy range shown in Fig. 6.13(a). As expected, $\gamma_{\mathrm{L},\mu}$ is much larger than $\gamma_{\mathrm{R},\mu}$ (note the different energy scale for $\gamma_{\mathrm{L},\mu}$ and $\gamma_{\mathrm{R},\mu}$). It can be seen that both $\gamma_{\mathrm{L},\mu}$ and $\gamma_{\mathrm{R},\mu}$, and as a consequence also $\mathrm{Im}(\epsilon_\mu^{\mathrm{W}})$, depend on the bias. The bias-dependence can be caused by a rearrangement of the charge distribution in the molecule for a given level, and also by the fact that for a given energy level the leads self-energy changes with bias. A closer inspection of the results shows that the larger values of $\gamma_{\mathrm{L},\mu}$ can be attributed to the molecular levels, whereas the very small values of $\gamma_{\mathrm{L},\mu}$ can be attributed to the Au surfaces states. For the molecular levels $\gamma_{\mathrm{L},\mu}$ changes approximately linearly with the bias. For almost all the states the occupation $\bar{p}_\mu$ is roughly either 1 or 0. Only the LUMO level has a fractional occupation considerably

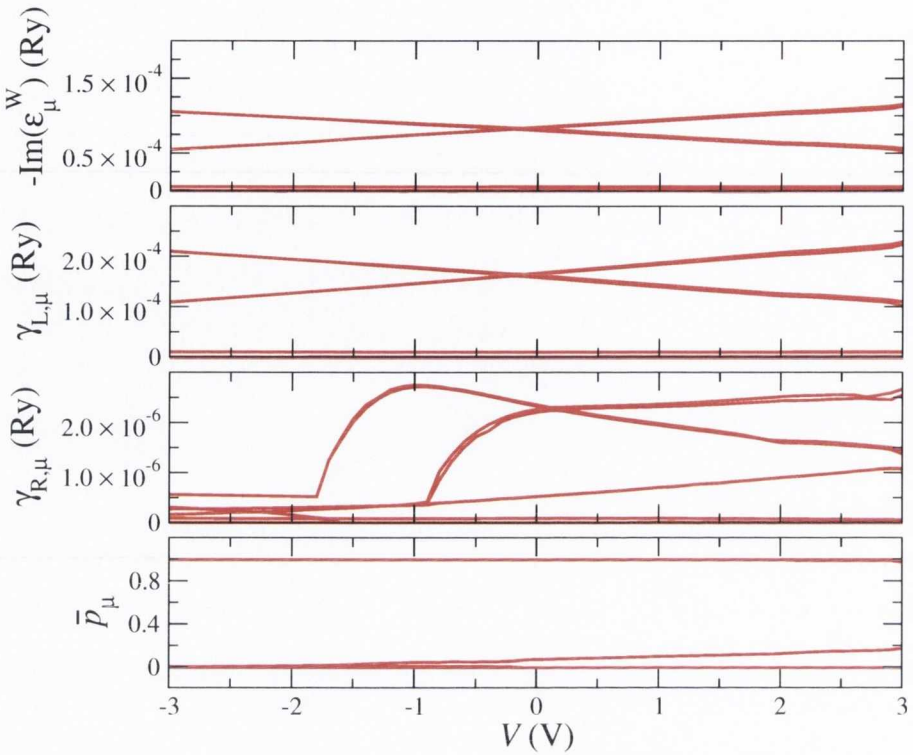


Figure 6.14: Imaginary part of the GBSs eigenvalues $\mathrm{Im}(\epsilon_\mu^{\mathrm{W}})$, their coupling coefficients to the left ($\gamma_{\mathrm{L},\mu}$) and right lead ($\gamma_{\mathrm{R},\mu}$), and their occupation $\bar{p}_\mu$, as function of the applied voltage V.

different from an integer value. Its occupation goes from about $\bar{p}_\mu = 0$ at -3 V, to about $\bar{p}_\mu = 0.2$ at +3 V. The increase of occupation with bias is approximately linear. It is this change of occupation that is responsible for the shift of the energy levels as function of bias. The higher the charge on the molecule, the higher the energy levels.

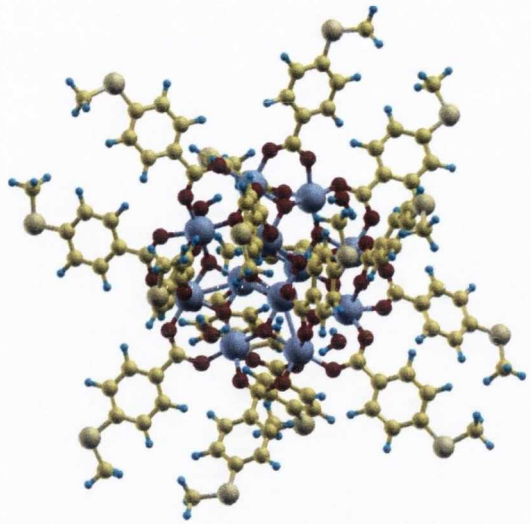


Figure 6.15: Structure of the Mn_{12}-based magnetic molecule.

6.10 Transport through a Mn_{12}-based magnetic molecule

In this section the application of the BSCS to a real physical system of large dimensions is presented. Like the previous example this is a system where the molecular states are very weakly coupled to the leads. In order to avoid the use of a huge number of energy points for the integration of $G^{<}_{\rm neq}$ at finite bias, we use again a finite δ to broaden the levels, and use the BSCS to correctly add the BSs charge. The general analysis is similar to the one of the previous section. In this section we use the adaptive energy mesh to calculate the transmission coefficient, in order to resolve the very sharp peaks. We have also verified that using the adaptive energy mesh with $\delta = 0$ gives the same self-consistent solution as the one obtained with the BSCS for finite δ.

The Mn_{12}-based magnetic molecule considered here is shown in Fig. 6.15. It consists of a core of 12 Mn and 48 O atoms, connected to 12 ligands, consisting of benzene groups [258, 28]. The Mn atoms lie almost in a planar arrangement, with 8 Mn atoms on an outer circle, and 4 Mn atoms on an inner one. The 8 Mn atoms in the outer circle are in a Mn^{3+} state, whereas the inner ones are in a Mn^{4+} state. In the ground state the magnetic moments of the 8 Mn^{3+} ions are all parallel to each other, and antiparallel to those of the 4 Mn^{4+} ions. The total magnetic moment M_s of the ground state is therefore $M_s = 20\ \mu_B$. At very low temperatures these molecules show magnetic hysteresis [28]. Experiments on the transport properties show Coulomb blockade regions in the I-V curves, indicating weak coupling, and

at certain bias voltages a negative differential conductance is found [258]. Scanning tunneling microscope measurements with the molecules on an Au substrate show a conductance gap of about 2 V [29, 30].

The main reason why this molecule is studied, is because it is a candidate for potential applications as single-molecule memory cell. The idea is that the molecule can be switched between different magnetic configurations, and that the readout of the state can be performed by a single current measurement. A very important question is therefore whether the spin state of the molecule can be deduced by an electric measurement. Here we address this question by calculating the current vs. voltage for two different magnetic configurations. The first configuration is the ground state, and the second one is one of the different possible magnetic excited states. These calculations were done in collaboration with C. D. Pemmaraju, who performed the DFT ground state calculations, and also some preliminary calculations on the transport properties [259]. In Ref. [259] the experimental properties of such molecules and previous theoretical calculations are reviewed.

The structure of the molecule attached to the gold electrodes used in our calculations is shown in Fig. 6.16. The molecule is attached with two of the ligands to each of the gold surfaces, the distances between the S end-groups and the Au surface are all chosen to be approximately equal to 2.0 Å. This results in an approximately symmetric coupling of the molecule to both sides. 12 of the benzene groups are not attached to the gold, so that no current can flow through these. In the transport calculations therefore these 12 ligands have been replaced by H atoms, in order to reduce the size of the EM. This is expected not to affect the transport results, especially since careful tests have shown that the energy levels close to the Fermi energy do not change by removing some of the benzene groups [259]. Fig. 6.16 also shows the alignment of the moments on the magnetic ions in the $M_s = 20\ \mu_B$ ground state. A green up (red down) arrow indicates that the ion is parallel to the majority (minority). The experimentally measured lowest excited state is a $M_s = 18\ \mu_B$ state, which is expected to be obtained by a collective excitations of all the ions [260, 258, 261]. This however can not be realized in our DFT calculations, so that we choose a different $M_s = 18\ \mu_B$ excited state, where we simply flip the spin of one of the Mn^{3+} and of the neighboring Mn^{4+} ions. This is one of the possible low energy magnetic excitations. Fig. 6.17 shows the molecule in this excited state, placed between the Au electrodes, as used in the transport calculations.

As done in the previous section, we include the 5-d orbitals of Au into the core, and use a double ζ s basis with a cutoff radius of the first ζ $r_c = 6.9$ bohr for the valence. For the C we use double ζ s ($r_c = 4.2$ bohr) and double ζ p ($r_c = 6.7$ bohr),

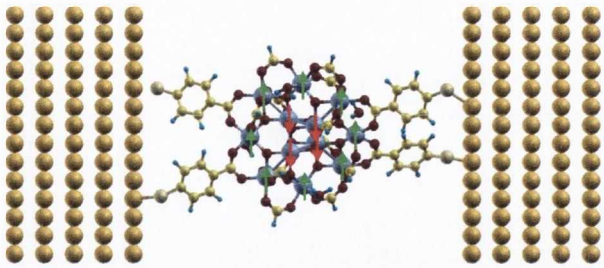


Figure 6.16: Modified Mn_{12}-based magnetic molecule used in the transport calculations, attached to the Au(111) electrodes. The arrows indicate the magnetic alignment of the Mn ions in the ground state configuration with $M_s = 20\ \mu_B$ (green arrows: majority, red arrows: minority).

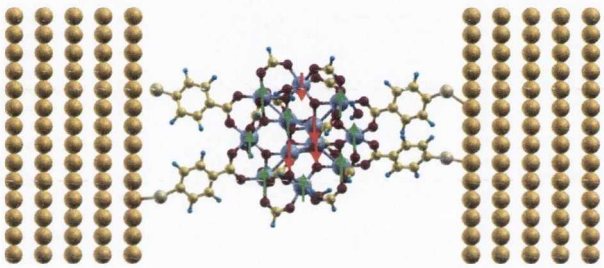


Figure 6.17: Same as Fig. 6.16, where the Mn_{12}-based magnetic molecule has the magnetic configuration corresponding to the excited state with $M_s = 18\ \mu_B$ (green arrows: majority, red arrows: minority).

for S we use double ζ s ($r_c = 5.0$ bohr), double ζ polarized p ($r_c = 6.2$ bohr) and single ζ d ($r_c = 5.2$ bohr), for O we use double ζ s ($r_c = 4.2$ bohr), double ζ polarized p ($r_c = 4.9$ bohr) and single ζ d ($r_c = 3.7$ bohr), for H we use double ζ s ($r_c = 5.5$ bohr), for Mn we use double ζ polarized s ($r_c = 5.9$ bohr), single ζ polarized p ($r_c = 5.8$ bohr) and double ζ d ($r_c = 5.9$ bohr). All the r_c are given for the first ζ only, higher ζ are constructed with the split-norm scheme, with a split-norm of 15% [69]. Due to the large size of the unit cell, no k-points are used in the direction perpendicular to the transport (only the Γ point), and the equivalent real space mesh cutoff is 400 Ry. For the exchange and correlation potential the GGA, as parametrized in Ref. [85], is used.

A detailed analysis of the DFT ground state is given in Ref. [259], here we just briefly present the main results, that are relevant for the transport. Fig. 6.18 shows the PDOS, projected on the two different types of Mn atoms, for the molecule attached to the gold surface, in the $M_s = 20\ \mu_B$ ground state. We note that the states in Fig. 6.18 have been broadened, in order to visualize them, but their real width is much smaller. We denote the four closely spaced levels just below E_F as the HOMO levels, and the closely spaced levels at about 0.3-0.4 eV as the LUMO levels. These HOMO

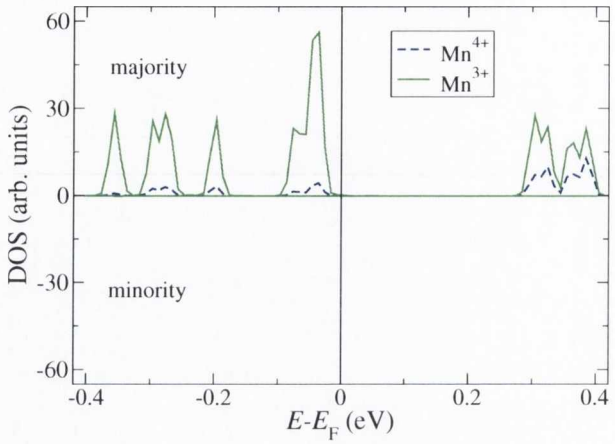


Figure 6.18: Spin-polarized projected density of states onto the Mn ions for the molecule in the $M_s = 20\ \mu_B$ ground state shown in Fig. 6.16. Positive values are for the majority, negative values for the minority.

and LUMO levels are mainly made of Mn and O states. The four closely spaced levels just below E_F can be mainly attributed to the Mn^{3+} ions. At low bias the transport is expected to be dominated by these states. In this ground state configuration there are no states in the minority spin close to E_F, so that the current is expected to be fully polarized at low bias. The GGA HOMO-LUMO gap of about 0.35 eV is rather small compared to the experimental value of about 2 eV [262, 29]. Using the LDA+U approximation for the exchange correlation potential, it is possible to obtain a value close to experiment by tuning the value of the parameters [259, 263]. The character of the HOMO and LUMO levels is however unchanged when compared to the GGA. Moreover, in experiments usually a gate voltage is applied in such a way, to bring the conducting levels close to the E_F of Au. The zero bias GGA result therefore resembles the experimental situation at zero bias, where such a gating voltage is applied. We therefore study the low bias behavior using the GGA approximation.

In Fig. 6.19 the zero bias transmission coefficient is shown. For each of the peaks in the PDOS there is also a corresponding peak in T. The peaks are all very sharp, with a FWHM in the range of 10^{-4} to 10^{-3} eV, indicating that the molecular states are very weakly coupled to the Au surface. The reason for this is that at these energies there are no states in the benzene ligands that couple to the central $Mn_{12}O_{12}$ part of the molecule, so that the wave function decays exponentially from the surface to the molecule. This results in a very small coupling. For all the peaks the maximum value of T is about 1, which means that the coupling is approximately equal to both sides [see Eq. (6.93)]. This is expected, since the molecule is placed in the junction in a symmetric way.

Given the sharpness the peaks, we would need an energy point spacing of about

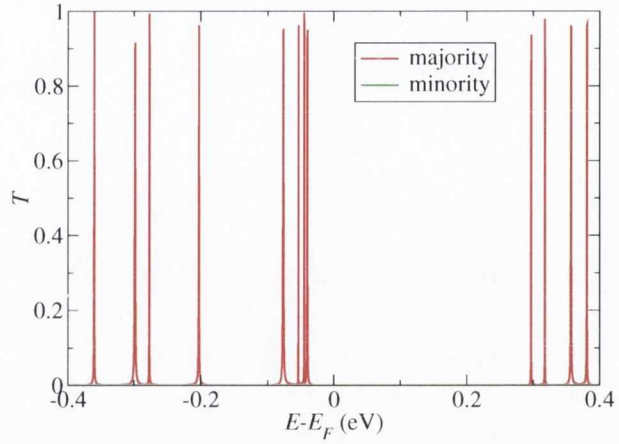


Figure 6.19: Zero-bias transmission coefficient for the $M_s = 20\ \mu_B$ ground state.

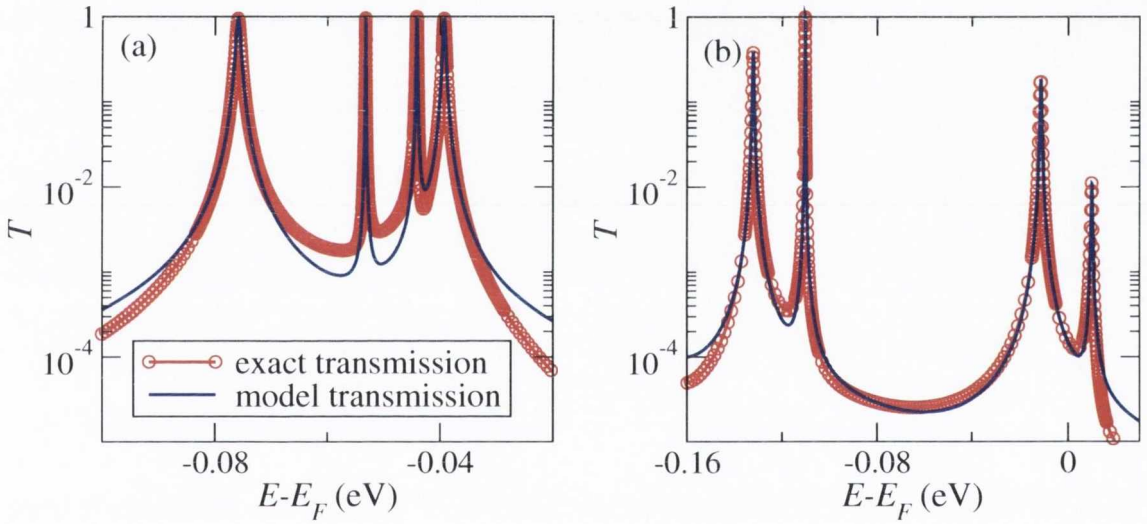


Figure 6.20: Comparison between the exact solution of the $M_s = 20\ \mu_B$ transmission coefficient (red line), and the one obtained using the model BSs transmission [Eq. (6.92)], for two applied bias voltages: (a) zero bias, and (b) $V = 200$ mV. The circles indicate the mesh points used for the exact calculation, determined using the mesh-refinement scheme described in Sec. 6.6.

10^{-6}-10^{-5} Ry in order to integrate properly $G^{<}_{\mathrm{neq}}$ at finite bias. We therefore use the BSCS with $\delta = 10^{-4}$ Ry, in order to limit the number of energy points needed to integrate $G^{<}_{\mathrm{neq}}$. In this way, we calculated the self-consistent solution in a bias range of ± 0.4 V. To verify that all GBSs are found at the correct energies, and also that no spurious GBSs are found, for each bias we carefully checked that for all the found GBSs there is a corresponding peak in the transmission. This is illustrated in Fig. 6.20, where T is shown for the 4 peaks close to E_F for a bias of 0 V (a) and 0.2 V (b). The red line shows T, calculated with the NEGF formalism using Eq. (4.94). In order to resolve the sharp transmission peaks properly, we used the adaptive mesh

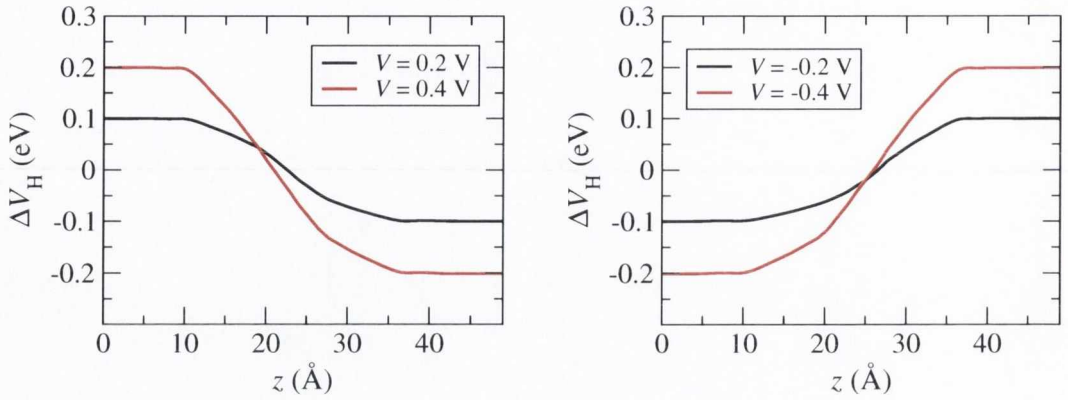


Figure 6.21: Planar average along x and y of the difference ΔV_{H} between the electrostatic potential at a finite bias voltage and the one at zero bias, as function of position z, for different bias voltages.

scheme described in Sec. 6.6. Each red circle in the figure corresponds to a mesh point. It can be seen that the density of points increases close to the peaks. To illustrate the effect of the parameter r_Δ, that sets the density of the grid for a given value of δ, we use $r_\Delta = 0.2$ at zero bias, whereas at $V = 0.2$ V we use $r_\Delta = 0.5$. The energy mesh is therefore denser in the zero bias case. In both cases the peaks are properly resolved, which shows that the smaller value of r_Δ is large enough for this system. We conclude that also in this case the adaptive energy mesh algorithm is suited for the integration the transmission, and consequently also to integrate the DOS and lesser GF. In order to check the correctness of the found GBSs, we plot T, calculated with the approximate model Eq. (6.92) (blue lines). The agreement close to the peaks is almost exact, confirming that the GBSs are found at the correct energies, and also that the calculated coupling to the leads via Eq. (6.54) is correct. The transmission away from the peaks is different, however this is expected, since the model transmission is only valid in the vicinity of the peaks. Due to the very good agreement between the model transmission and the exact one, we conclude that also the BSCS should be suited to reproduce the correct density matrix at finite bias, and therefore work well for this materials system. As further test we calculated the self-consistent solution using the adaptive energy mesh algorithm. The results for the transmission are the same as for the self-consistent BSCS solution, which confirms that both methods work well for this system. We note that at $V = 0.2$ V some of the transmission peaks shrink dramatically (note the logarithmic scale in Fig. 6.20), which is caused by the fact that at this bias, contrary to the zero bias case, the coupling to the leads is highly asymmetric. This important effect will be discussed in more detail further down.

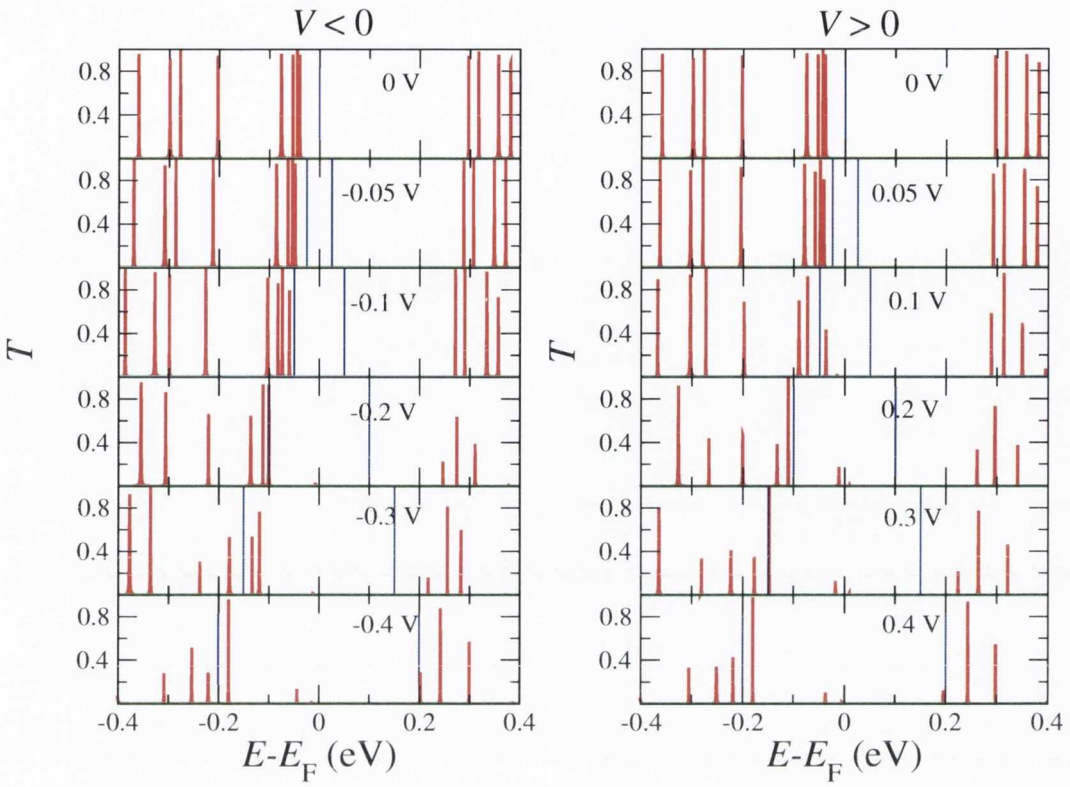


Figure 6.22: Bias dependent transmission coefficient for the $M_s = 20\ \mu_B$ ground state (red: majority, green: minority) on a linear scale; the blue vertical lines indicate the boundaries of the bias window.

We now move to analyzing the finite bias behavior of the system in detail. In Fig. 6.21 we show planar average ΔV_H along x and y of the difference between the self-consistent electrostatic Hartree potential at a finite bias and that at 0-bias along the junction stack. The junction stack corresponds to the direction of the transport, the z-axis in our case, and the planar average is taken along the plane perpendicular to the transport (the x-y plane). We refer to the drop in ΔV_H between the left and right electrode as the "potential drop". This is a very useful quantity for the understanding of the finite-bias transport properties. Fig. 6.21 shows ΔV_H for four different voltages. ΔV_H is flat inside the Au electrodes and decays monotonically across the molecule. For negative bias the potential drop is approximately equal to the mirror image of the one for positive bias. This indicates again that the molecule is placed in a symmetric position, so that the charging of the molecule is similar for positive and negative bias.

In Fig. 6.22 the spin-resolved transmission coefficient is shown for different bias voltages. The blue vertical lines indicate the bias window. We note that since an electronic temperature of 300 K is used, the Fermi function is smoothed out, so that states outside, but close to the bias window, also contribute to the current. The same

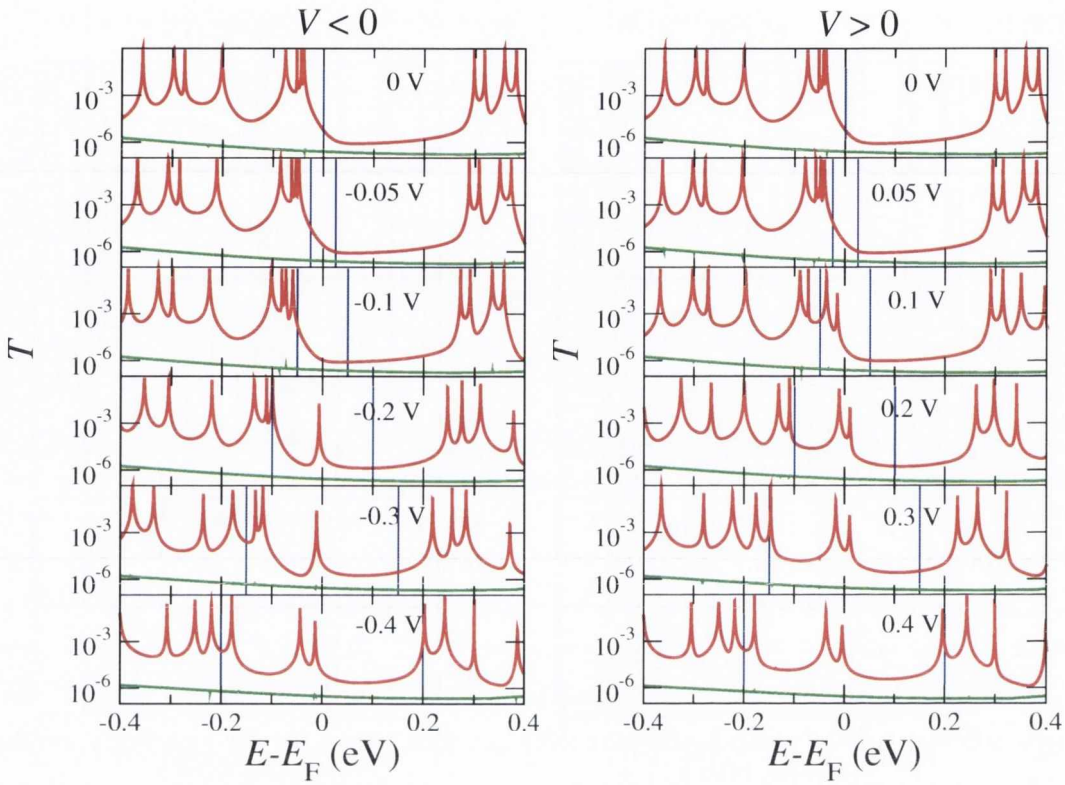


Figure 6.23: Bias dependent transmission coefficient for the $M_s = 20\ \mu_B$ ground state (red: majority, green: minority) on a logarithmic scale; the blue vertical lines indicate the boundaries of the bias window.

transmission is shown again in Fig. 6.22 on a logarithmic scale, which makes all the peaks clearly visible. We first discuss the behavior with positive bias. The HOMO states just below E_F remain outside the bias window up to a bias of 0.05 V, where there is only a small shift of the levels. At 0.1 V two of the levels enter the bias window. The height of these two transmission peaks however is much smaller than 1, especially for the state that moves to the middle of the bias window, which is barely visible on the linear scale. As discussed in Sec. 6.7, this can only be caused by the fact that the coupling of these states to the electrodes is now highly asymmetric. As the bias is further increased, this shrinking if found for all the peaks that enter the bias window. For negative bias we find the same general behavior, also here all the peaks entering the bias window shrink. The difference is that up to -0.1 V the states are all still outside of the bias window, and that at -0.2 V there is just one state inside the bias window, compared to the two states at +0.2 V (see Fig. 6.23). These small differences are caused by the slight asymmetry of the molecular junction.

In order to understand the cause of the shrinking of the peaks, we calculate γ_L and γ_R for all the found GBSs [Eq. (6.54)]. Indeed we find that at zero bias left and right

coupling coefficients are similar, and that they become asymmetric with bias. For positive bias the states entering the bias window are coupled more to the left lead, so that $\gamma_L \gg \gamma_R$. In this way, even if the states are inside the bias window, their charge is approximately conserved. Using Eq. (6.75) it can be seen that for positive bias the occupation for states in the bias window is approximately equal to $\bar{p} = \gamma_L/(\gamma_L + \gamma_R)$, and this is approximately 1 for $\gamma_L \gg \gamma_R$. By changing the coupling the states can therefore easily enter the bias window, whereas for symmetric coupling they would loose half the charge by doing this. In this case they would always be pinned outside the bias window, just like in the example for the C_3 molecule, where the BSs term was neglected [Fig. 6.13(b)]. For negative bias, we find that the states entering the bias window have $\gamma_R \gg \gamma_L$. For negative bias, the occupation of the states in the bias window is given by $\bar{p} = \gamma_R/(\gamma_L + \gamma_R)$, so that also here the change in charge is minimal when the HOMO levels enter the bias window.

The question arises on how the states can change the coupling in such a dramatic way. In order to answer this question, in Fig. 6.24 we plot the real part of the wave function for the 4 GBSs levels close to E_F, for a bias of -0.2 V, 0 V and 0.2 V. For each wave function we also give the value of the real part of the energy eigenvalue E_R with respect to E_F, of the imaginary part E_I, and also $p = (\gamma_L - \gamma_R)/(\gamma_L + \gamma_R)$ [Eq. (6.73)]. As discussed in Sec. 6.5, this term is approximately 1 if $\gamma_L \gg \gamma_R$, 0 if $\gamma_R \approx \gamma_L$, and -1 if $\gamma_L \ll \gamma_R$. It is therefore a measure of the asymmetry of the coupling. For zero bias all 4 wave functions extend evenly over the whole core of the molecule, and in fact here $p \approx 0$ for all levels. The coupling of all the states is therefore approximately the same to both leads. At $V = 0.2$ V the wave functions change drastically. Instead of spreading over the whole molecule, some of the WFs localize on the left or right side. The two peaks with the largest E_R are inside the bias window (see Fig. 6.23) at this bias, for these two peaks $p \approx 1$, as already described for general voltages. The energies of the other two GBSs are outside the bias window. One of these states is localized evenly between left and right side, and in fact here $p \approx 0$. The other one is located on the right side, and for this state $p \approx 1$. The same general relation between the localization of the WF and the couplings is also found for $V = -0.2$ V. The conclusion therefore is, that the change in p is obtained by localizing the WF either on the left or right side of the large molecule. This change of the molecular levels is possible, since the four levels are closely spaced, so that a linear combination of these is formed to obtain new eigenstates at a slightly modified potential. At higher bias more levels get close to the bias window. New linear combinations of states are formed, so that some of the levels can again enter the bias window without significantly changing the charging state. Since all the levels

inside the bias window, which are the ones contributing to the current, have a strongly reduced transmission, the current is strongly suppressed compared to what would be expected from the zero bias transmission.

In Fig. 6.25(a) the resulting $I - V$ is shown. The current is very small, of the order of a few tens of nA, reflecting the fact that the system conducts in the very weak coupling regime. Moreover we note that the current almost completely spin-polarized, since the current in the minority is negligible. For very small positive bias below 50 mV, the current flows mainly through the HOMO states pinned at the Fermi energy. At higher bias however, once the states change coupling and enter the bias window, the current flows mainly through the states in the bias window. Due to their highly asymmetric coupling however they carry only little current, so that the current decreases slightly with increasing bias. Once the bias gets close to 400 mV, the LUMO levels start to enter the bias window, which results in an increase of the current. We note however that the expected experimental value for the HOMO-LUMO gap is much larger than the GGA value of about 0.35 eV. Therefore in experiment the LUMO levels are expected to contribute to the current only at much higher bias voltages. For negative bias the general behavior is similar, which again reflects the fact that the molecule is placed in an approximately symmetric way between the two surfaces. The drop in current at -400 mV is caused by a rearrangement of the molecular levels at that bias.

We now perform the same analysis also for the excited state with $M_s = 18\ \mu_B$ (Fig. 6.17), in order to verify if the spin-state affects the transport properties. In Fig. 6.26 the PDOS of the isolated molecule in the excited state is shown. It is similar to the one for the GS, the difference being that one of the four peaks below the Fermi energy is now a minority state. We can therefore expect also the minority spin to contribute significantly to the current in the excited state. In Fig. 6.27 the resulting bias dependent transmission coefficient is shown on a linear scale, and in Fig. 6.28 it is shown on a logarithmic scale. The transmission is qualitatively similar to the one of the ground state, with the difference being that one of the 4 HOMO peaks is now a minority peak. There is also a second minority peak at about 0.3 eV. At finite bias the general behavior is again that the peaks are shifted out of the bias window up to some voltage, above which the states change the coupling to the Au, so that some of them enter the bias window. However, since there is one less state in the majority spin to re-hybridize the HOMO levels, these re-hybridize in a different way than in the ground state. We note especially that the minority state close to E_F can not form linear combinations with any other states, so that it is always pushed out of the bias window, until the LUMO levels enter. The change of WFs, and of the corresponding

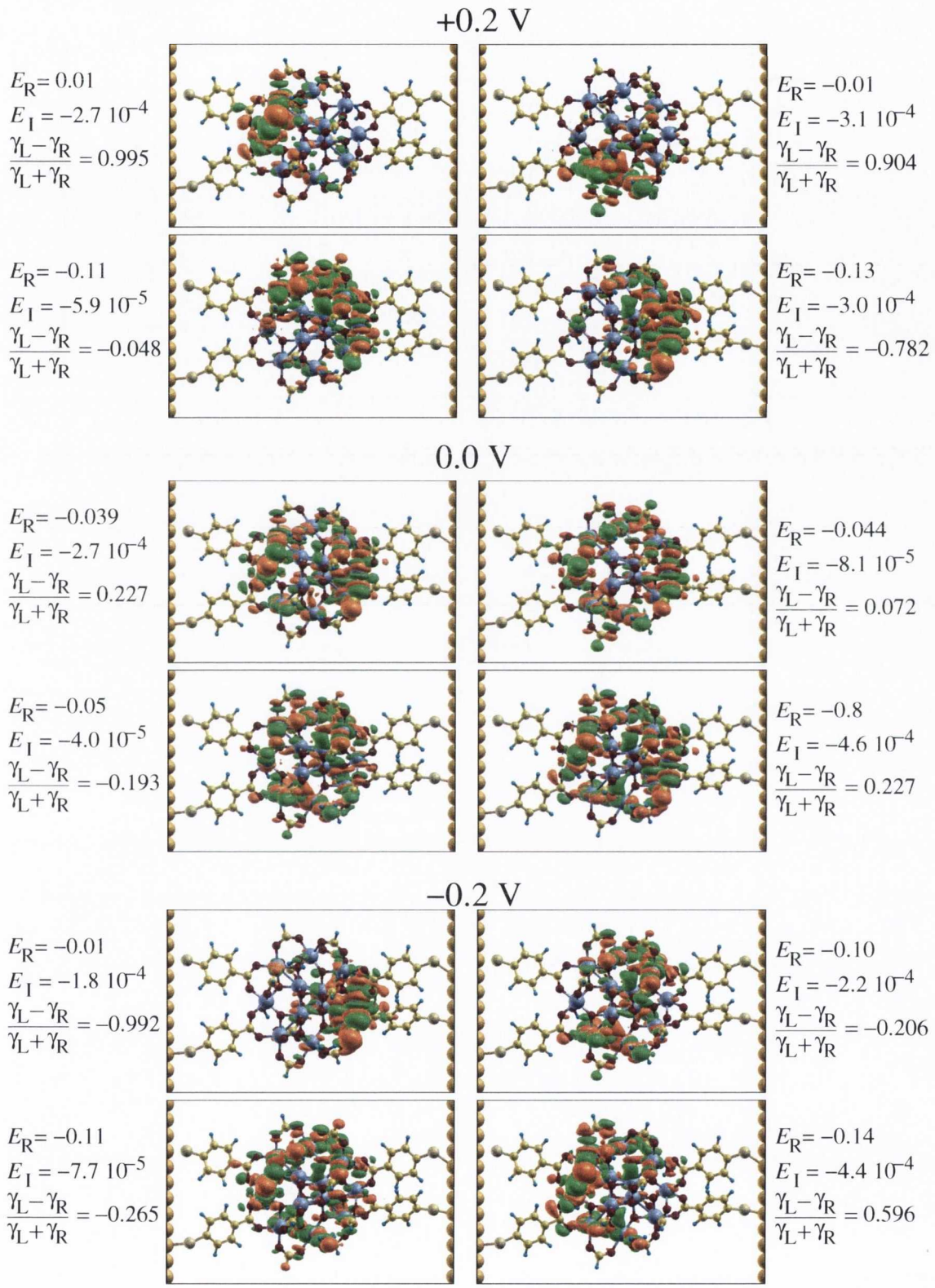


Figure 6.24: Real space plots of the $M_s = 20\ \mu_B$ ground state WFs for the four states closest to the Fermi energy for three different bias voltages. All energies are given in units of eV.

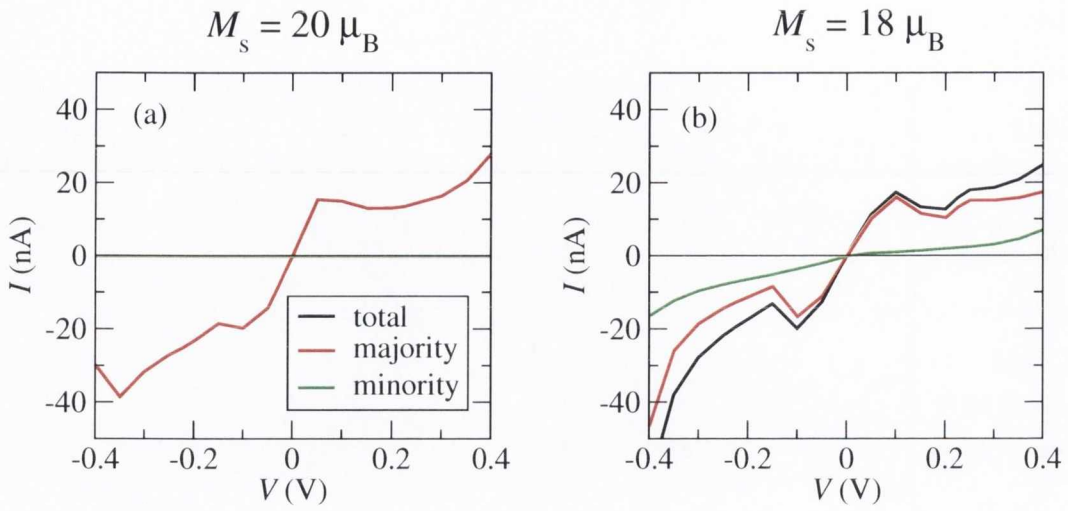


Figure 6.25: Spin-resolved current I versus voltage V curves for the $M_s = 20\ \mu_B$ ground state (a) and the $M_s = 18\ \mu_B$ excited state (b).

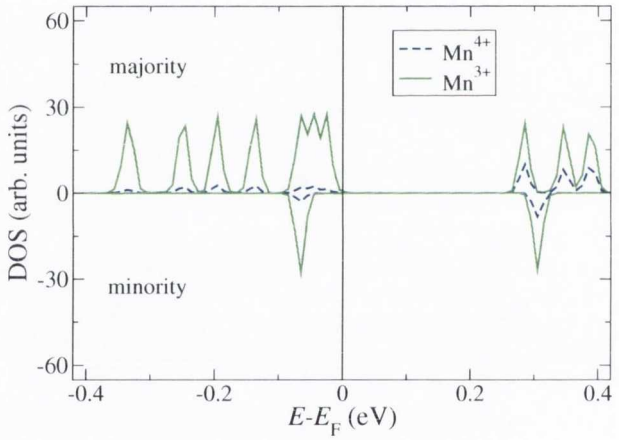


Figure 6.26: Spin-polarized projected density of states onto the Mn ions for the molecule in the $M_s = 18\ \mu_B$ ground state shown in Fig. 6.17. Positive values are for the majority, negative values for the minority.

coupling coefficients, with bias is illustrated in Fig. 6.29. Indeed there is no visible difference in the WF for the minority states for the different bias voltages, which is reflected by the fact that the coupling coefficients are almost constant. The values show that the state is always coupled slightly more to the right lead than to the left one. This is a consequence of the choice of the position of the Mn ions, at which the spins are flipped in the excited state, since the minority WF is localized in that region. The majority states on the other hand change in an analogous way as for the GS.

In Fig. 6.25(b) the resulting spin-polarized $I - V$ curve for the excited state is shown. The general behavior is similar to the one for the ground state, and the to-

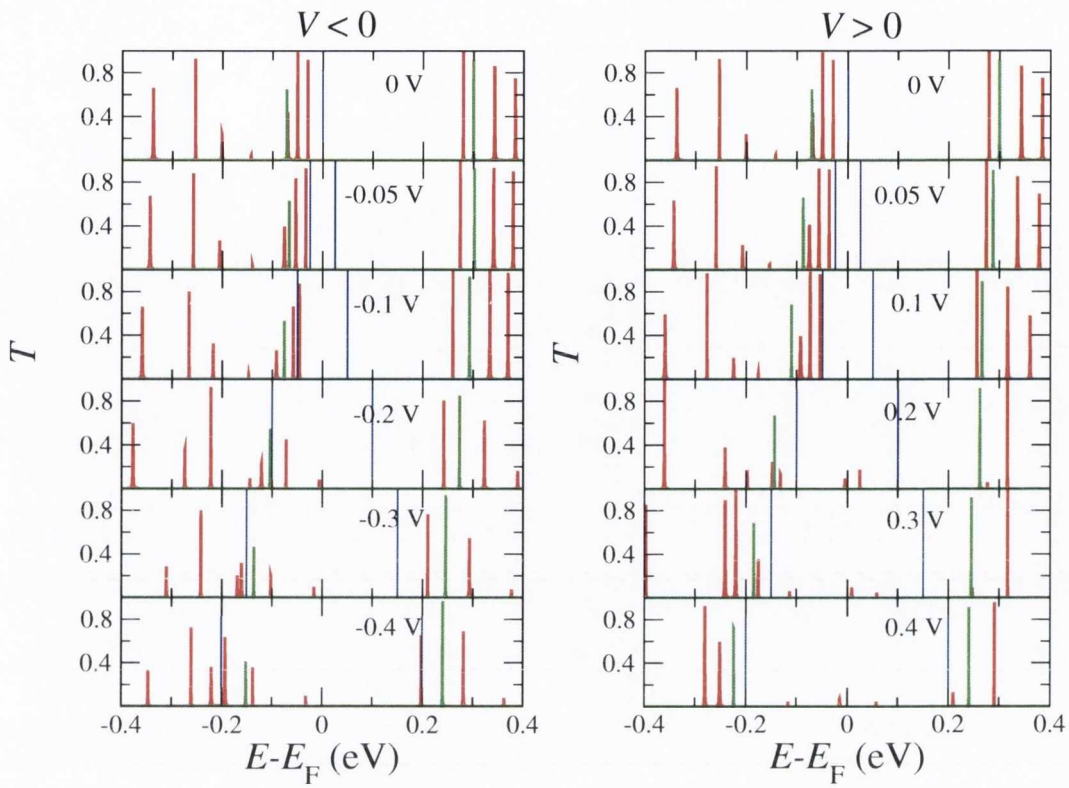


Figure 6.27: Bias dependent transmission coefficient for the $M_s = 18\ \mu_B$ excited state (red: majority, green: minority) on a linear scale; the blue vertical lines indicate the boundaries of the bias window.

tal current is of a similar magnitude for both states. Contrary to the ground state however, there is now also a contribution to the current from the minority spin. Due to the similarity of the results between the two states, we conclude that it is difficult to distinguish the two states from a readout of the current alone. Although the polarization of the current is very different, the total current is similar. Different orientations of the molecule might lead to larger changes of the current than differences in the magnetic state. A change of the spin-state during an $I-V$ measurement might however be detectable.

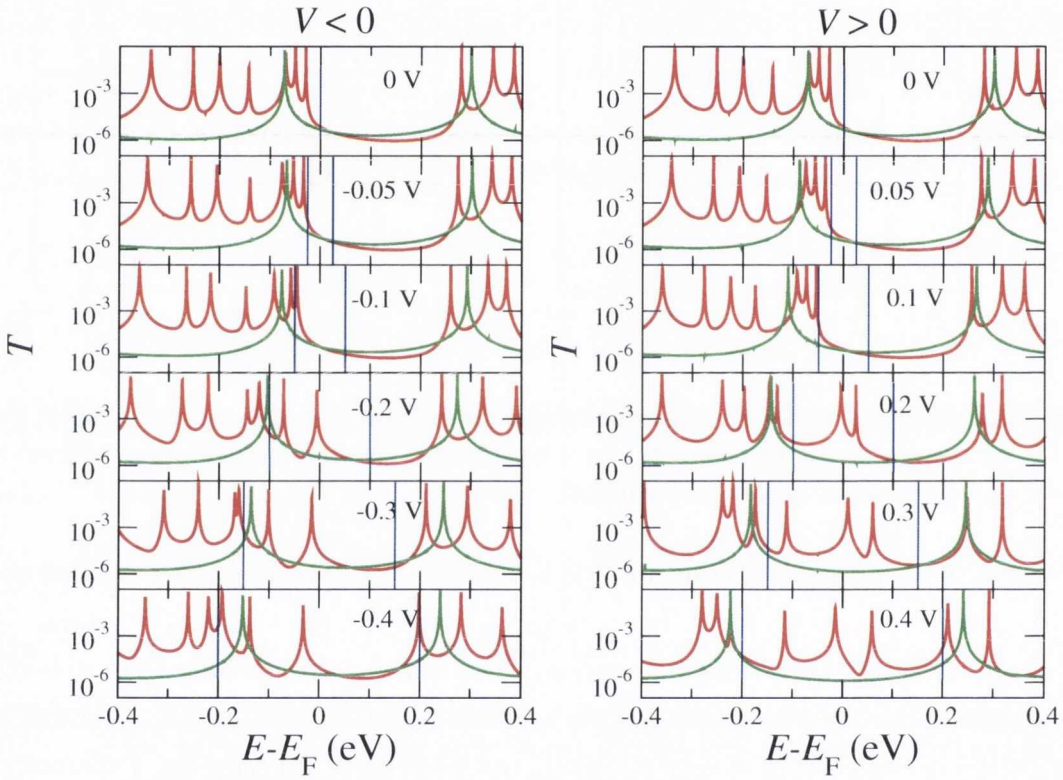


Figure 6.28: Bias dependent transmission coefficient for the $M_s = 18\ \mu_B$ excited state (red: majority, green: minority) on a logarithmic scale; the blue vertical lines indicate the boundaries of the bias window.

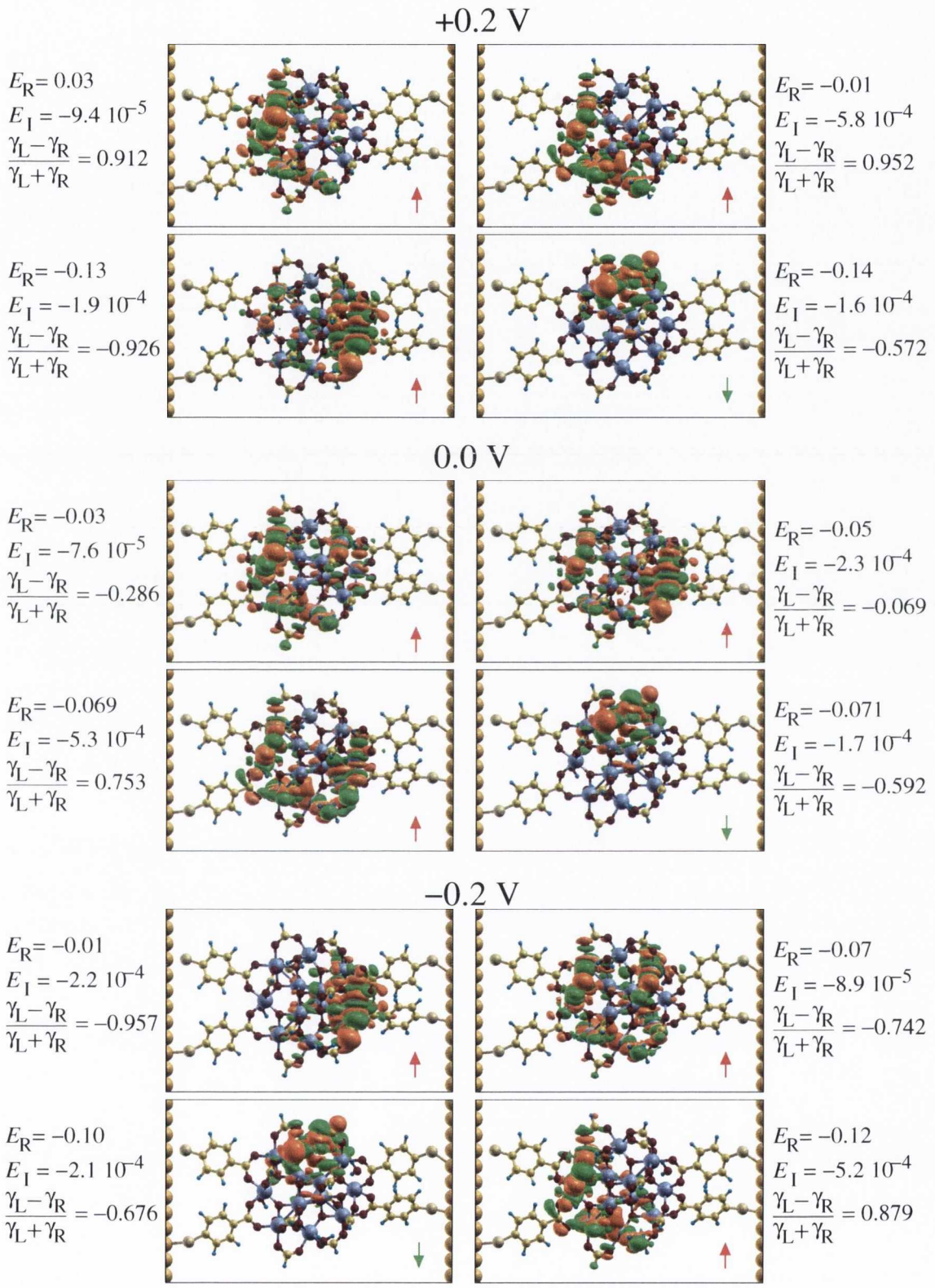


Figure 6.29: Real space plots of the $M_s = 18\ \mu_B$ excited state WFs for the four states closest to the Fermi energy for three different bias voltages. The arrows indicate the spin of the WFs. All energies are given in units of eV.

6.11 Conclusions

A major challenge in NEGF calculations is the treatment of weakly coupled and bound states [62, 63, 243, 246, 242, 244, 245]. While in some cases the standard NEGF formalism might not be suited to treat such states [206, 207, 187, 208, 250, 251], for many materials systems we can formulate a valid description. Apart from the conceptual difficulties, weakly coupled states cause technical problems in the calculation, due to the fact that they cause arbitrarily sharp peaks in the DOS. If a regularly spaced energy mesh is used, the integrations of the lesser GF and of the transmission coefficient over energy become almost intractable. We have presented two possible solutions: the first is the BSCS, based on the concept of generalized bound states, and the second is the adaptive energy mesh. Both methods are found to work well for our test systems, which consisted of a small C_3 molecule, and a large magnetic Mn_{12}-based molecule, attached to gold electrodes. We expect the adaptive energy mesh to be the more robust and flexible method for the integrations over energy. The main advantage of the calculation of the GBSs is, that it gives many additionally informations about the system, such as effective couplings to the leads, and the nature of the states in the EM. One can then use simplified formulae for transmission and DOS. It is also possible to separate out those states from the strongly coupled ones in the calculation of the density matrix. It might therefore be possible to set up a calculation, where the NEGF is used only for the strongly coupled states, whereas for the occupation of the GBSs other formalisms are applied.

For true BSs, for which the coupling to the leads is exactly zero, the situation is more difficult. We have shown how their occupation can be set, based on the assumption that there are interactions outside the ones described by the DFT Hamiltonian. These lead to an effective coupling to the leads, which we have included in the NEGF formalism by the effective coupling matrices. For systems such as capacitors, or tunnel junctions, it is usually possible to determine the effective coupling based on the spatial location of the BSs. This will be illustrated in detail in chapter 7.

We have presented different examples, where weakly coupled and bound states appear. For a simple 1D tight-binding model, we have determined the energy of the BS as function of bias voltage, and also the maximum number of electrons that can be associated to BS. The important result is that it is necessary to calculate the contribution from the BSs independently for each bias voltage, since these can appear and disappear with bias. Most importantly, the charge that can be placed in a BS is a continuous function of bias, and can assume any value between zero and one electrons. In the calculations performed with our implementation of the

BSCS, and of the mesh refinement scheme in SMEAGOL for the small C_3 molecule, we have illustrated how the self-consistent finite bias results critically depend on the correct inclusion of weakly coupled states in the theory. For the magnetic molecule the calculation of the GBSs and of their properties allowed us to understand the peculiar shifting of the energy levels with bias, and to determine the origin of the shrinking of the transmission peaks as they enter the bias window. We found the surprising result, that, as a consequence of the fact that the preferred molecular state is charge neutral, the nature of the molecular states changes drastically for different bias voltages. This leads to a negative differential conductance.

Chapter 7

Electronic transport through Fe/MgO/Fe(100) tunnel junctions

Modern magnetic sensors, such as read heads for hard disk drives, are based on the tunnel magnetoresistance (TMR) effect. This is the drop in resistance of a magnetic tunnel junction (MTJ) formed by two magnetic layers when the mutual alignment of their magnetization vectors changes from antiparallel (AP) to parallel (P). The TMR magnitude is given by

$$\mathrm{TMR} = \frac{I_{\mathrm{P}} - I_{\mathrm{AP}}}{I_{\mathrm{AP}}}, \tag{7.1}$$

with $I_\alpha(V)$ being the current at the voltage V for the α configuration (P or AP). Huge TMR ratios have been achieved in epitaxial, all crystalline Fe/MgO [264] and CoFeB/MgO [265] MTJs, reaching up to 604% at room temperature and 1144% at 5 K [18, 19]. These large values of TMR are largely attributed to the phase coherent and transverse momentum conserving transport. Extensive reviews on the transport properties of crystalline tunnel junctions are given in Refs. [16] and [15].

Tunneling junctions are usually grown in a layered form, either by sputtering [265] or by molecular beam epitaxy (MBE) [264]. The advantage of the MBE growth is that high quality junctions with well defined interfaces can be grown. The crystalline interfaces are atomically sharp over a long range [264]. Sputtering, in conjunction with an annealing step, leads to local crystallization at the interface. The long range order is worse than that obtained with MBE, and also the junction thickness is less well defined. The advantage of the sputtering techniques however is that they are much cheaper. For both methods usually the bottom Fe electrode is grown on a substrate, often an anti-ferromagnetic layer. On top of this a few layers of MgO are deposited, over which the top metallic electrode is then grown. It is generally found that the interface between Fe(100) and MgO(100) is formed in such a way that the O atoms of the first MgO layer grow on top of the Fe interface atoms, with the in-plane unit cell of MgO rotated by 45° with respect to the Fe unit cell [266,

267, 268, 20, 269]. The details of the experimentally determined interface structure vary somewhat for different experiments. In Refs. [270, 266, 267, 268, 271, 272] it is shown that in the first few layers of Fe partial oxidation takes place. Theoretical calculations show that the transport properties can change drastically if such an FeO layer is present [273, 274, 275]. Neutron diffraction experiments suggest that magnetically dead layers can be formed at the interface, especially if grown in O_2 atmosphere [276], which might also be attributable to an FeO layer at the interface. In Refs. [16, 277, 278, 279, 280, 281, 282, 283] however it is claimed that there is no oxidation in the Fe. Theoretical calculations suggest that the formation of a FeO layer can be controlled by varying the O/Mg excess during growth [284]. The recent experimental results of Refs. [285, 286, 271] indicate that both an oxidized and a non-oxidized junction can be grown, depending on the growth conditions. We note also that not all present experimental techniques might be able to detect a single FeO layer at the Fe/MgO interface, as indicated in Ref. [287]. Apart from the partial oxidation of the interface, even in high quality junctions different types of defects can be found, such as O and Mg vacancies [285, 288]. Due to the small lattice mismatch between Fe and MgO of about 3.9 %, lattice dislocations are formed at the interface between MgO and Fe [264]. In order to maximize the TMR, often CoFe is used for the ferromagnetic electrodes. For junctions made by sputtering, Boron is usually added, which allows the growth of non-crystalline electrodes. These are then annealed after the sputtering process, so that crystalline interfaces with the MgO are obtained. In this work we consider ideal junctions, without Fe oxidation. We do however study the influence of defects, in the form of O vacancies in the MgO.

Several theoretical calculations investigating the linear response limit predict very large TMR for such junctions [20, 289, 290, 269, 274, 291]. In these works the current is calculated in the ballistic limit by using the two spin-fluid approximation, where the spin-currents for majority ($\uparrow$) and minority ($\downarrow$) spins do not mix. Inelastic effects and spin-flip events are not considered. Moreover periodic boundary conditions are assumed in the plane perpendicular to the transport, so that Bloch theorem can be applied in the plane. The total transmission is then calculated with Eq. (4.109), where the spin- and $\mathbf{k}$-dependent transmission is integrated over the 2D BZ to give the total spin-dependent transmission. The TMR is found to be governed not only by the spin-polarization of the electrode DOS, but also by the details of the wave functions matching across the barrier. This is analyzed in detail in Ref. [20, 21], where it is shown that the decay of a wave function across the barrier depends mainly on two factors: the first is the symmetry of the wave function, and the second is its $\mathbf{k}$-point in the 2D BZ perpendicular to the transport direction, since it is found

that the decay varies strongly for different **k**-points in the 2D BZ. The slowest decay is found for states with Δ_1 symmetry at the Γ-point, which have no momentum component in the plane perpendicular to the transport. States with Δ_5 symmetry are found to decay faster across the barrier. If the transport direction is along the z axis, and the x and y axes are in the plane perpendicular to it, then the states with Δ_1 symmetry are those that transform like a linear combination of functions with $1, z, 2z^2 - x^2 - y^2$ symmetry [20]. The states with Δ_5 symmetry on the other hand transform like functions with zx and zy symmetry. Fe has a high transmission Δ_1 band at the Γ-point, which is found at energies around the Fermi energy only for the majority ($\uparrow$) spins, whereas for the minority ($\downarrow$) spins no wave function with such a symmetry is found close to the Fermi energy. Since these bands dominate the tunneling current, Fe electrodes separated by a MgO barrier effectively behave as half-metals, and the TMR is expected to be very large [20, 289].

An important, but much less investigated aspect, is the relation between the electronic states and the I-V characteristics in these highly crystalline MTJs. Interface states and details of the Fe band-structure, otherwise washed out by disorder, play an important role in the transport and indeed can be identified through the I-V curves and its derivatives [$G(V) = \mathrm{d}I/\mathrm{d}V$ and $S(V) = \mathrm{d}^2I/\mathrm{d}V^2$]. For instance, high-quality MTJs (2-3 nm MgO thickness) show a pronounced broad peak on the $S(V)$ curve at about 1 V for the AP configuration, and a number of small peaks at lower voltages in the P configuration [292, 293]. Combined with a quantitative theory these measurements can provide a wealth of information, and help the device design.

This chapter is organized as follows. After presenting some basic properties of bulk Fe and MgO, we study the electron transport in an ideal Fe/MgO/Fe(100) tunnel junction with 4 MgO MLs ($\sim$1 nm thick). First the zero bias transport properties are analyzed and compared to the existing literature. We then move to the finite bias properties. The importance of the correct occupation of the BSs in the non-equilibrium case is shown by analyzing the potential drop and charging, and the properties of the BS as function of bias are analyzed in detail. We then present the bias-dependent transmission and I-Vs. The resulting TMR decreases at high bias and eventually even becomes negative above about 1.7 V. We show how the transport properties at finite bias are related to the layer- and symmetry-projected DOS of the Fe electrodes. We conclude the discussion of the 4 MgO MLs junction by verifying the influence on the transport of small changes in the structure of the junction and by calculating the changes due to the use of a self-interaction corrected

exchange-correlation functional in the MgO.

In the following section we then study the dependence of the transmission on the MgO thickness. We show that the results for large barrier thickness are strongly dependent on the used basis set, and relate this to the complex band structure of MgO. We calculate the energy-dependent damping across the MgO for thick barriers, and find that at the Fermi energy the damping for $\uparrow$ spins is only slightly smaller than for the $\downarrow$ spins, so that only a small increase of the TMR with bias is predicted at large thickness. This is in agreement with the results in Refs. [291, 294].

Generally the theoretically obtained values for the TMR for ideal junctions, which are of the order of several thousands of percent [20, 290, 269, 274, 291], are much larger than the ones obtained in experiments, which are never higher than around one thousand percent [264, 265, 18, 19]. This might be attributable to different effects, such as disorder [295, 296, 297, 294], a partial oxidation of the interface Fe layer [273, 274, 275, 291], non-collinear spin-configurations at the Fe interface layer, or to defects in the MgO [298, 294, 299]. Inelastic effects are also likely to play an important role in experiments [292, 246, 300, 15] and usually lead to an additional reduction of the TMR. Also many-body effects might play a role, especially at small bias. A detailed description of the inelastic and many-body effects on the transport is given in Ref. [300]. Since we always work in the ballistic, energy conserving, limit, we do not include these inelastic effects. In this work we investigate the transport properties of junctions with oxygen vacancies in the MgO and find that indeed the TMR decreases drastically for oxygen deficient junctions, especially if the vacancies are close to the interface.

In the final part of this chapter we analyze the transport properties of double barrier junctions. Experiments for such junctions show that these can reduce the decay of the TMR with bias [301]. Moreover evidence of quantum well states in the middle Fe layer is found [302, 303, 304]. In our calculation we find both these effects and analyze their origin.

7.1 Basic properties and system setup

MgO is an insulator and crystallizes in the NaCl structure, with an experimental equilibrium lattice constant of 4.21 Å [305, 306, 307] and a band gap of 7.8 eV [305, 307]. For both the O and Mg atoms, in our calculations the used basis set is double ζ s and p, with cutoff radii for all the first ζs of 6 Å. We have thoroughly tested convergence of the results with respect to the use of richer basis sets. For the equilibrium lattice constant we obtain 4.19 Å for LDA, which matches well the

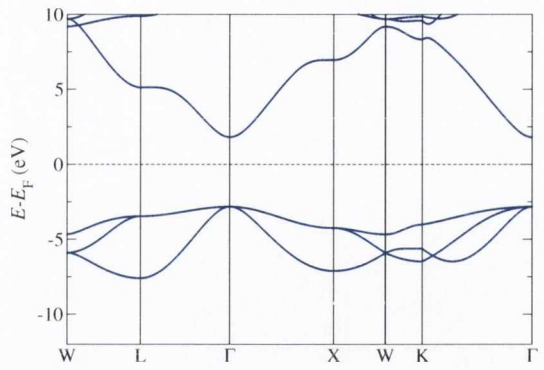


Figure 7.1: MgO band-structure.

experimental value and also previous *ab initio* calculations [308]. For GGA we obtain 4.29 Å, which is somewhat larger than the experimental value. In Fig. 7.1 the LDA band structure is shown for the equilibrium lattice constant. It matches well with previous calculations [309, 305, 308]. We note that the band gap at the Γ point is only 4.64 eV, which is about 3.2 eV smaller than the experimental value. This discrepancy is caused by the self-interaction in the LDA exchange correlation potential (Sec. 2.1.5). In fact, if we add the ASIC (see Sec. 2.1.5) to the LDA exchange correlation potential, then the band gap increases to 6.95 eV, and agrees much better with the experimental value. We note that the ASIC contribution depends on the choice of the so-called α-parameter (see Ref. [90] for the definition). In our ASIC calculations we use an α-parameter of 0.5, which is the default value for semiconducting periodic systems. In Ref. [90] it is shown that the experimental gap for MgO can be obtained by increasing the α-parameter to about 0.65 [90]. Since the LDA band gap, although smaller than the experimental value, is still rather large, we perform most of the calculations using the LDA. In Secs. 7.4.2 and 7.5 we do however compare the results to those obtained using the ASIC. As we will show, the main features of the I-V and TMR-V do not change when the ASIC is introduced.

Fe is a ferromagnetic metal and crystallizes in the bcc structure, with a lattice constant of 2.8665 Å [310, 311] at room temperature. The agreement of the LDA band structure and DOS with the experimental ones is rather good [310, 312]. One of the major failures of the LDA however is that the ferromagnetic state does not correspond to the lowest energy state. Therefore, when the LDA is used, a fcc non-magnetic ground state is predicted, in contrast to experiment [313]. Moreover, the LDA equilibrium lattice constant of about 2.76 Å [313] for ferromagnetic bcc Fe is rather small when compared to experiment. The GGA, as parametrized in Ref. [85], brings the lattice constant very close to the experimental value, and also correctly pre-

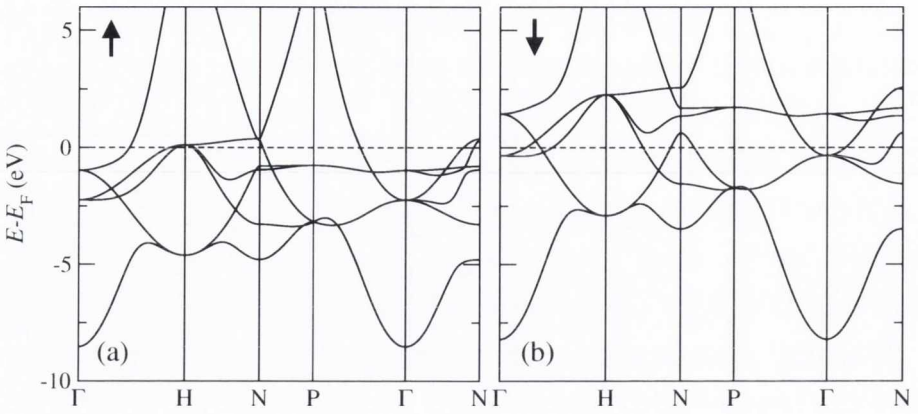


Figure 7.2: Fe band-structure for majority (a) and minority spins (b).

dicts a ferromagnetic ground state [314, 313]. We thoroughly tested the convergence of Fe band structure, DOS and transport properties for different basis sets, with the aim to find a minimal basis set, that still correctly describes these properties. The resulting basis set for Fe is double ζ s (r_c=5.6 bohr), single ζ p (r_c=5.6 bohr) and single ζ d (r_c=5.2 bohr). We note that the first set of transport results, as published in Ref. [315], were calculated using a double ζ (DZ) s (r_c = 7 bohr), single ζ (SZ) p (r_c = 7 bohr), and DZ d (r_c = 5.6 bohr) basis. Since the results are very similar to the ones obtained with the smaller basis set, here we present only the latter. For bulk Fe we obtain a relaxed lattice constant of 2.79 Å for LDA, and 2.88 Å for GGA, which agrees well with other calculations [313]. The LDA band structure is shown in Fig. 7.2. We can identify the parabolic s like bands starting at about -8.5 eV, and the spin-split d bands, ranging from about -4.9 eV to 0.3 eV for the $\uparrow$ spins, and from about -3.5 eV to +2.5 eV for the minority spins. The band structure matches well with other LDA calculations [316], and also to experiments [312]. The calculations presented in the remaining part of this chapter are all performed using the LDA. We did however also repeat some of the calculations using the GGA. We find that the general results are similar for both LDA and GGA, although some features at and around the Fermi energy are different. This is caused by the fact that the band structure for LDA and GGA differs close to the Fermi energy.

A general feature of tunneling junctions is that the lattice vectors of the top and bottom electrodes are slightly different. This is due to the strain induced by the lattice mismatch between the metals and the insulating layer. The strain in the MgO is reduced away from the interface, so that the in-plane lattice vectors of the top Fe electrode adapt to the MgO. In order to make the computations feasible, we need to apply periodic boundary conditions perpendicular to the stacking direction. For this

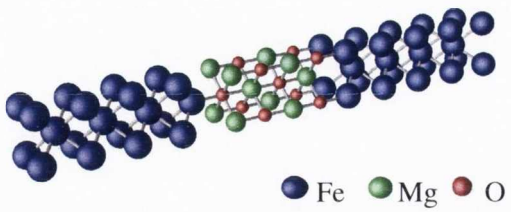


Figure 7.3: Unit cell used for the Fe/MgO/Fe(100) junction with 4 MgO MLs. Periodic boundary conditions are applied perpendicular to the stacking direction.

reason it is necessary to assume that the in-plane lattice vectors of both electrodes are identical. For the same reason we assume that the in-plane MgO lattice adapts perfectly to the Fe electrodes, which means that the used MgO in-plane lattice vectors are $\sqrt{2}$ times the one of Fe [20]. For the Fe electrodes we use the bulk Fe lattice parameters, with a lattice vector of 2.866 Å. The in-plane lattice vector of MgO then results to $\sqrt{2} \times 2.866 \approx 4.05$ Å. In Fig. 7.3 the unit cell for a 4 MgO monolayers (MLs) Fe/MgO/Fe(100) junction is shown, periodic boundary conditions are applied in the plane perpendicular to the stacking direction. In all our calculations we use 8 Fe layers on each side of the MgO in order to converge to bulk. We note that the interface structure obtained experimentally when MgO is grown on Fe can be different from the one obtained when Fe is grown on MgO. In all our calculations however we use a completely symmetric interface on both sides.

For the atomic positions along the stacking direction slightly different values are used in the literature [20, 269, 274, 304]. In Ref. [20] it is assumed that there is no relaxation of the MgO lattice along the stacking, so that the lattice constant along the stacking is also equal to 4.05 Å. The Fe-O distance at the interface used in the calculations is 2.16 Å. We refer to this set of coordinates as the unrelaxed coordinates. In Ref. [269] a relaxation of the atomic positions and lattice vectors along the stacking is performed for 3 MLs of MgO. The result is that due to the in-plane compression, the MgO expands along the stacking direction, in order to keep the volume of the unit cell approximately constant. Another result of the relaxation is that the Mg atoms in the interface layer are slightly displaced towards the Fe substrate with respect to the O atoms. The calculated Fe-O distance is 2.21 Å, and the distance between the O atoms along the stacking is 2.196 Å (compared to 2.025 Å for the unrelaxed coordinates). Based on these results, we construct MgO barriers with an arbitrary number of MLs by using 2.196 Å as the spacing between the MLs. Except for the first interface layer, it is assumed that the Mg and O atoms always have the same z coordinates. We refer to this set of coordinates as the relaxed coordinates. In our calculations we use both the unrelaxed and the relaxed coordinates. We will

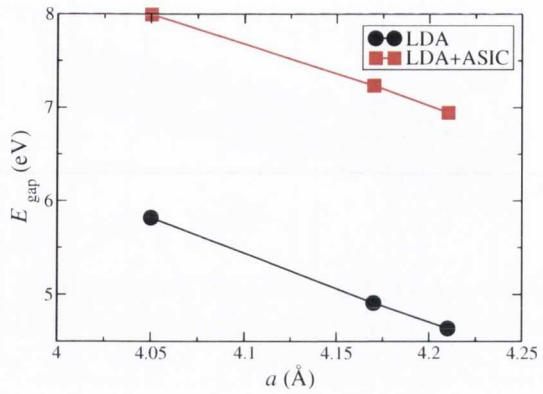


Figure 7.4: MgO band gap E_{gap} as function of the lattice constant a. The experimental lattice constant is 4.21 Å.

show that although the general results are similar for both the sets of coordinates, there are two main differences. The first is that the MgO band gap differs by about 1 eV. This is because the band gap changes sensibly when the lattice constant is reduced, as shown in Fig. 7.4. It increases by about 1 eV when the lattice constant is reduced from the experimental value of 4.21 Å to 4.05 Å, for both LDA and ASIC. The effective lattice constant for the relaxed coordinates is $(4.05^2 \times 4.4)^{1/3} \approx 4.16$ Å, and it is rather close to the equilibrium value. The calculated LDA band gap for the unrelaxed coordinates is 5.8 eV, and that for the relaxed coordinates is 4.8 eV. This difference leads to a different barrier height, so that we expect some quantitative differences in the calculated currents. The second main difference in the results for the relaxed and unrelaxed coordinates is that, due to the different interfaces with the Fe, the coupling of surface states into the MgO varies. At the end of the next section we will show that this leads to a stronger transmission through ISs for the unrelaxed coordinates.

A 7×7 **k**-points mesh is used during the self-consistent cycle to converge the charge density in all the transport calculations. We carefully checked the convergence of the transport properties with respect to this value. Once self-consistency is achieved, a 100×100 **k**-point mesh is used over the full BZ for evaluating the transmission coefficient in a single post-processing step. This finer mesh is necessary in order to resolve sharp resonances. We use a real space mesh cutoff of 600 Ry and an electronic temperature equal to 300 K. Also here we have verified that the results are almost independent on the choice of the temperature, the main difference being that some features are more pronounced at low electronic temperature.

7.2 Zero bias transport properties for a 4 ML junction

In this section we present the zero bias transport properties for a 4 ML junction with relaxed coordinates. At the end of the section we will then compare the results with those obtained for unrelaxed coordinates. In Fig. 7.5 the SIESTA DOS for bulk Fe is shown, together with the PDOS (defined in Sec. 4.1) for the 8th Fe layer (the furthest away from the MgO) and for the interface Fe layer. The agreement between the bulk DOS and the one for the 8th layer is rather good, indicating that the effects of the MgO on the potential are largely screened at this distance. The PDOS for the interface layer is rather different. One of the major differences is the appearance of a peak in the minority PDOS around the Fermi energy (the peak itself is sightly above the Fermi energy). This peak therefore corresponds to a IS, found also in previous calculations for the same junction [20, 290, 269], and also for Fe/vacuum/Fe junctions [65, 295]. To visualize the surface state in real space we have also calculated the local density of states (LDOS) for all the states in an energy window of ± 0.2 eV around E_{F} (see Fig. 7.6). The LDOS is the real space charge density generated by all the eigenvalues lying in a specified energy range. It can be seen that there is an enhancement of the charge at the interfaces between Fe and MgO in the minority spin, whereas for the majority spin no charge is visible on the chosen scale at the interface. We also calculate the band alignment for the junction, and find that the Fermi level of Fe is about 1.8 eV below the conduction band of MgO, and about 3.0 eV above the valence band. In Fig. 7.1, where the MgO band structure is shown, the Fermi level is approximately set according to this band alignment.

We now move to analyzing the properties of the Fe leads relevant to the transport. The leads unit cell is made up of 4 Fe layers, stacked along the (100) direction. By calculating the inverse band structure for the leads, as described in Sec. 5.1, and using Eq. (5.40), we calculate the leads DOS

$$\mathcal{N} = \frac{1}{\pi \Omega_{\mathrm{BZ}}} \int_{\mathrm{BZ}} d\mathbf{k} \sum_{n}^{N_{\mathbf{k},\mathrm{open}}} \frac{1}{v_{\mathbf{k},n}}, \tag{7.2}$$

where the integral goes over the 2D BZ perpendicular to the transport direction. Here Ω_{BZ} is the area of the BZ, $N_{\mathbf{k},\mathrm{open}}$ is the number of open channels at a given $\mathbf{k}$-point, and $v_{\mathbf{k},n}$ is the group velocity for channel n. In Fig. 7.7(a) the DOS per Fe atom is shown. This is equal to $\mathcal{N}/4$ since there are 4 atoms in the unit cell. We further

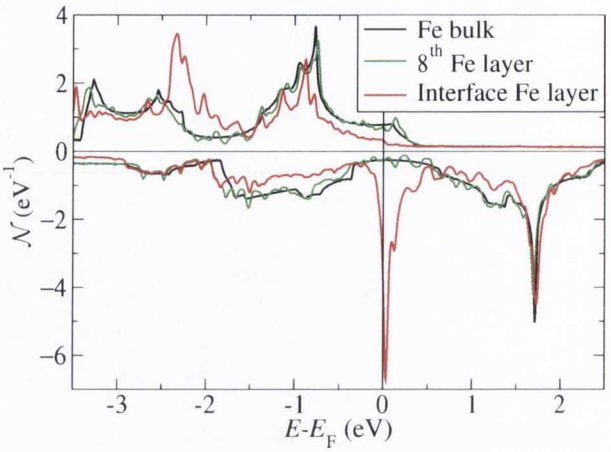


Figure 7.5: Bulk DOS for one Fe atom (black curve), DOS for the Fe atom at a distance of 8 layers from the MgO (green curve), and DOS for the Fe atom at the interface layer (red curve). Positive values represent the ↑ spins, negative values the ↓ spins.

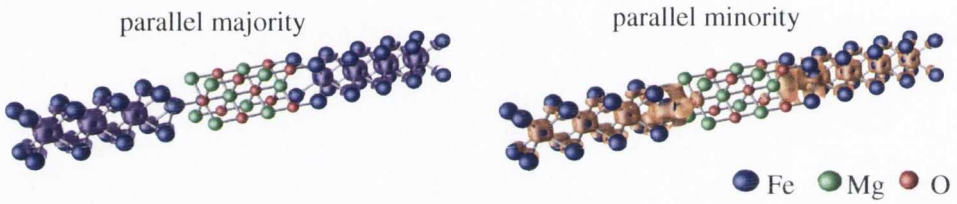


Figure 7.6: Isosurfaces for the charge density associated to the bands in a range of 0.2 eV around the Fermi energy, for P ↑ (left) and ↓ (right).

calculate the DOS times the group velocity,

$$\mathcal{N}v = \frac{1}{\pi\Omega_{\mathrm{BZ}}}\int_{\mathrm{BZ}} d\mathbf{k} \sum_{n}^{N_{\mathbf{k},\mathrm{open}}} \frac{1}{v_{\mathbf{k},n}} v_{\mathbf{k},n} = \frac{1}{\pi}\int_{\mathrm{BZ}} d\mathbf{k}\; N_{\mathbf{k},\mathrm{open}}, \tag{7.3}$$

which is proportional to the average number of open channels [Fig. 7.7(b)]. And finally we calculate the DOS times the square of the group velocity

$$\mathcal{N}v^2 = \frac{1}{\pi\Omega_{\mathrm{BZ}}}\int_{\mathrm{BZ}} d\mathbf{k} \sum_{n}^{N_{\mathbf{k},\mathrm{open}}} v_{\mathbf{k},n}, \tag{7.4}$$

which is shown in Fig. 7.7(c). We note that the quantities $\mathcal{N}v$ and $\mathcal{N}v^2$ are in principle dependent on the choice of the z direction. However for typical $3d$ metals they are almost isotropic. From Fig. 7.7 it can be seen that whereas the contributions from the d states dominate in the DOS, where they are visible as sharp peaks, in $\mathcal{N}v$ the contribution from s and d like bands is of the same order of magnitude. Finally for $\mathcal{N}v^2$ the contribution from the s bands dominates, since they have a larger group velocity compared to the d bands. As described in [58], and also in Sec. 2.2 of Ref.

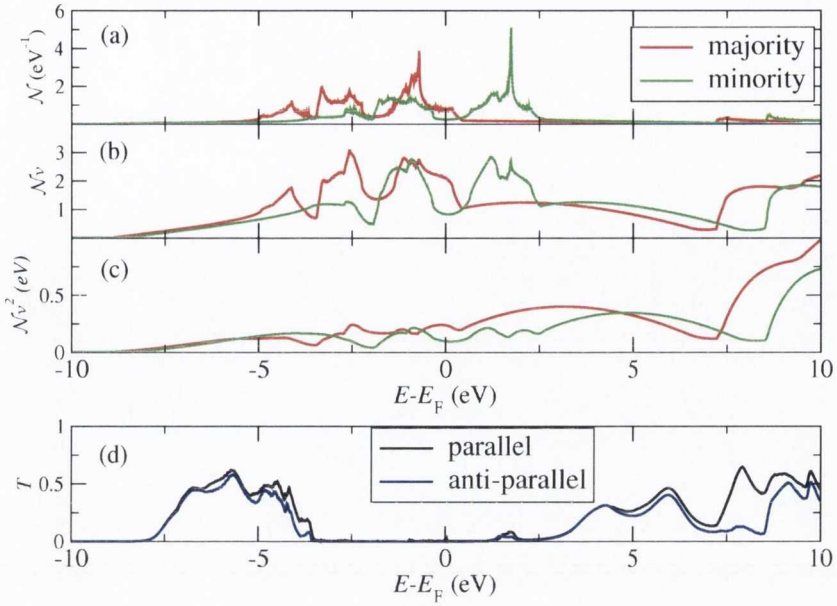


Figure 7.7: Spin-polarized DOS $\mathcal{N}$ for one Fe atom (a), average number of channels per unit cell $\mathcal{N}v$ (b), DOS times the square of the group velocity $\mathcal{N}v^2$ (c), transmission for a 4 MgO MLs junction per unit cell T (d), for the P and AP configurations.

[55], $\mathcal{N}v$ determines the transport properties in the ballistic regime, and $\mathcal{N}v^2$ is the relevant quantity in the diffusive regime. Since we calculate the transport properties in the ballistic regime, the transmission is directly dependent to the number of channels (see also chapter 4). From Fig.7.7(b) we can see that around the Fermi energy both s and d bands can contribute to the transport. In the diffusive regime one can expect the s bands to dominate.

In Fig. 7.7(d) the total transmission coefficient is shown for both P and AP configuration of the magnetic electrodes for the 4 MgO MLs junction. The transmission vanishes on this scale between about -3.5 eV to 1.3 eV around the Fermi energy, due to the band gap in the MgO. In Fig. 7.8 the spin-dependent transmission is shown on a logarithmic scale, together with the average number of channels in the Fe leads $n_c = \mathcal{N}v$. For the P configuration, and for energies in the range of about ± 1 eV around E_F, the transmission for $\uparrow$ spins is much larger than that for the $\downarrow$ spins. Very close to E_F however there is a sharp peak in the minority transmission, which is due to the surface state close to E_F. Below about -1 eV also the $\uparrow$ transmission drops, due to the fact that this is the energy of the band-edge of the majority Δ_1 state, $\Delta_1^{\uparrow}$, at the Γ point (see Fig. 7.27). At this energy we also find a IS in the $\uparrow$ spins, which causes the peak in the transmission. The sharp increase in transmission in the $\downarrow$ spins at about 1 eV to 1.5 eV is due to the fact that at 1.5 eV there is the band-edge of the minority Δ_1 states, $\Delta_1^{\downarrow}$, at the Γ point (see Fig. 7.27). For other

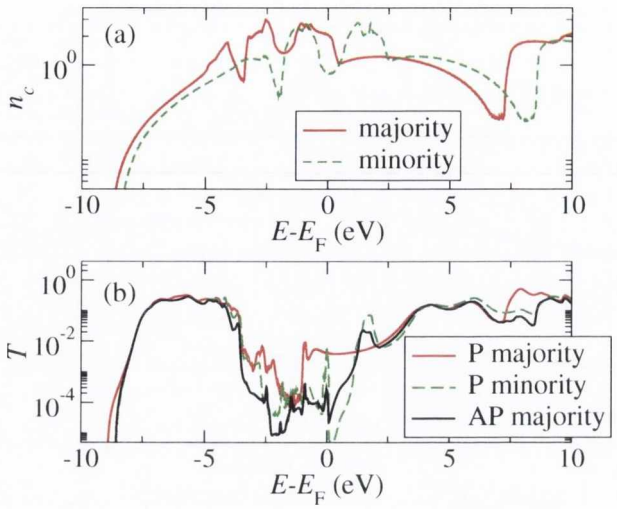


Figure 7.8: Average number of channels per unit cell $n_c = \mathcal{N}v$ for $\uparrow$ and $\downarrow$ (a), and transmission for a 4 MgO MLs junction T (b), for the P $\uparrow$, P $\downarrow$, and and AP $\uparrow$.

energies inside the MgO band gap the transmission varies, following also the change in the number of channels. An increase in n_c usually is translated in an increased transmission. For energies outside of the MgO band gap the transmission is roughly proportional to the number of channels, with the scattering across the MgO bands leading to some variations. As a first approximation, the transmission in the AP configuration can be seen as a convolution of the majority and minority transmission in the parallel one [55, 68]. Around E_F it is much lower than the one for the P configuration. The resulting 0-bias TMR at E_F is about 1780%. We will discuss the energy- and spin-dependent transmission coefficient in more detail in Sec. 7.4, where we will also discuss its dependence on the applied bias voltage.

The general results for the transmission close to E_F and therefore also for the 0-bias TMR agree with other calculations [20, 274, 290, 269, 317]. The main discrepancy is in the magnitude of the IS peaks. Whereas in Ref. [290] for a 4ML junction the P $\downarrow$ transmission is larger at E_F than the P $\uparrow$ transmission, in Refs. [20, 274, 269] the contribution from resonances is negligible in the P configuration. This discrepancy can have different origins. Some of the differences are attributable to the slightly different lattice parameters used in the various calculations, as we will show at the end of this section. One of the problems is also that in many calculations only the value at E_F is given. This depends sensibly on the exact position in energy of the $\downarrow$ surface state. In Sec. 7.5 we will also show that the choice of the basis set can change the amplitude of the resonances, although for a 4 ML junction this effect is not large. In collaboration with my colleague Nadjib Baadji a separate set of calculations was performed for the same system, using a tight-binding LMTO ASA code. Also in this

case we find a large contribution of the ISs to the transmission for the 4 ML junction. One way to reduce the contribution of resonances in the transmission is to add a small imaginary part to the energy when evaluating the GF (see Sec. 7.4). In experiments the resonant transport through such a surface state is probably highly suppressed by interface roughness. We also note that for thicker junctions the contribution from such a surface state is negligible in both the P and AP configuration (Sec. 7.5).

In Fig. 7.9 the **k**-resolved transmission coefficient is plotted in the 2D BZ perpendicular to the transport direction, for different energies around E_{F}. Each of the colored squares extends over the full 2D BZ, and the transmission is shown on a logarithmic scale with a color code. Blue corresponds to regions where the transmission is smaller than 10^{-7}, whereas red corresponds to a transmission of 1. Green and yellow denote intermediate values. The $\uparrow$ transmission [Fig. 7.9(a)] is almost the same for all energies. It has a broad maximum at the Γ-point ($k_x = k_y = 0$), and oscillates towards the border of the BZ. These oscillations are caused by the **k** dependent complex band structure of MgO [20]. The $\downarrow$ transmission on the contrary changes considerably across this energy range. It has an almost constant background pattern with small transmission, but there are also high peaks, caused by the IS. Their height and position in the BZ change strongly as function of energy, reflecting the small energy width of the IS. The shape of the peak for energies below E_{F} is similar to the one given in Ref. [20] at E_{F}. The height at E_{F} is larger in our case. This is probably due to the fact that the IS position relative to E_{F} is slightly different. In fact for energies slightly below E_{F}, the height decreases rapidly. The transmission for the majority spin in the AP configuration is shown in Fig.7.9(c). The figure shows that at each **k**-point the transmission in the AP configuration approximates a convolution between the P $\uparrow$ and P $\downarrow$ transmission.

In Fig. 7.10 the **k** dependent number of channels of the Fe leads is shown for $\uparrow$ and $\downarrow$. The color scale is linear, blue stands for 0, orange for 3, and red for 5. Comparing Figs. 7.9 and 7.10, it can be seen that for those **k**-points, where the number of channels is zero, also the transmission vanishes. The fact that there are no open channels for some **k** points has major implications for the finite bias calculations, since at all those **k**-points bound states appear in the system. This will be discussed in detail in the next section. Although these carry no current in our calculations, in practical experiments they are might contribute to the current, since in experiment the coupling to the Fe electrodes is never exactly zero (see Sec. 6.7) [246]. In Refs. [247, 248, 249, 243] the contribution of such localized states to the current is calculated within the Bardeen perturbation treatment [318, 319], in which such an effective coupling is implicitly assumed. However as shown in Ref. [243], the

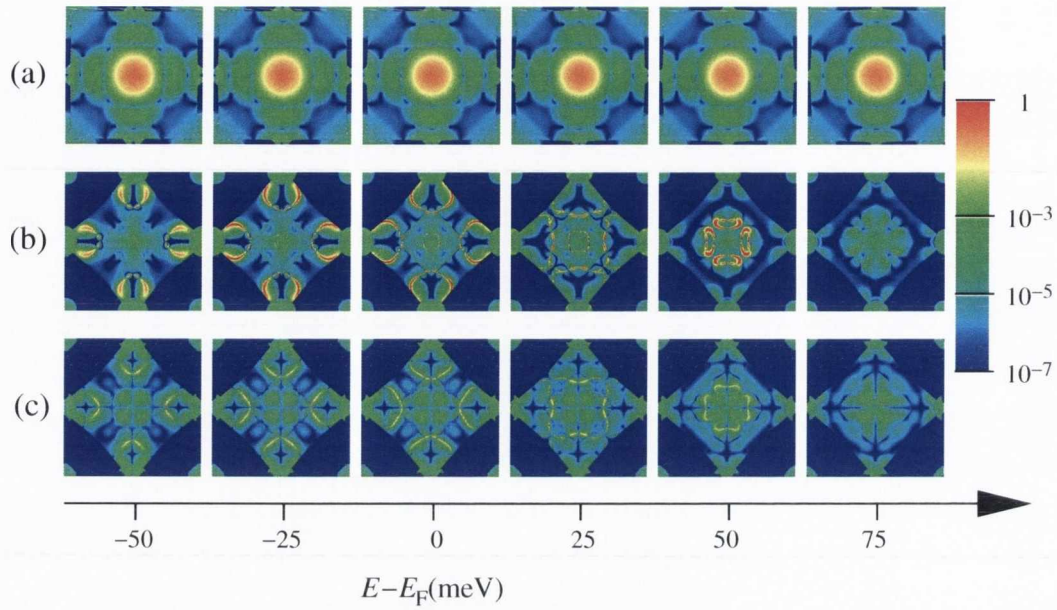


Figure 7.9: **k**-dependent transmission plotted over the 2D BZ for different energies, and for P ↑ in the first row (a), fr P ↓ in the second row (b), and for AP ↑ in the third row. The color code is on a logarithmic scale, red stands for a transmission of 1, and blue stands for a transmission below 10^{-7}. The values are calculated for the relaxed coordinates using the LDA.

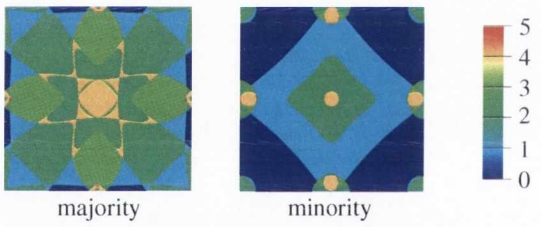


Figure 7.10: **k**-dependent plot over the whole BZ of the number of open channels in the Fe leads at the Fermi energy for ↑ and ↓. The color code is on a linear scale, and ranges from zero (blue) to 5 (red).

additional contribution of localized interface states to the total current is rather small for the Fe/MgO junctions and therefore negligible.

We conclude this section by comparing the zero bias transmission with the one obtained by using the unrelaxed coordinates of Ref. [20]. We perform both LDA and GGA calculations for those coordinates, in order to see whether the change of the exchange-correlation functional has a significant effect on the transport properties. In Fig. 7.11 the spin-dependent transmission coefficient is compared for the three different calculations. Fig. 7.11 (a) is calculated using the relaxed coordinates with the LDA, Fig.7.11 (b) is calculated using the unrelaxed coordinates with the LDA, and Fig.7.11 (c) is calculated using the unrelaxed coordinates with the GGA. The

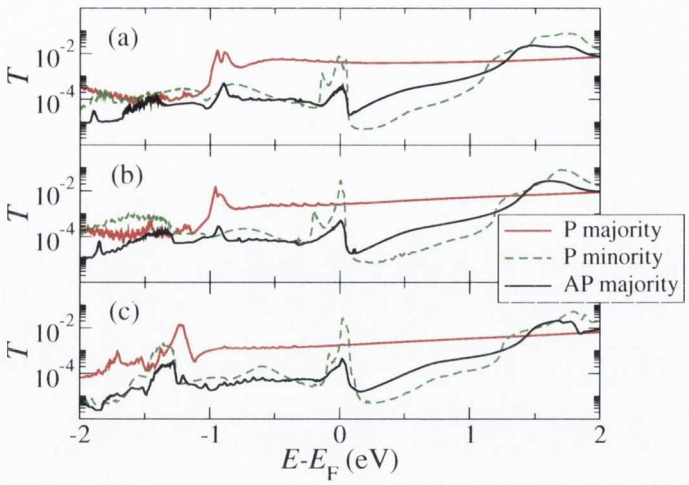


Figure 7.11: Transmission coefficient T calculated for the relaxed coordinates and the LDA (a), for the unrelaxed coordinates and LDA (b), and for the unrelaxed coordinates and GGA (c). The red curves represent the P ↑ transmission, the green curves the P ↓ one, and the black curves the AP ↑ one.

general results are similar for all the three cases. The major difference between the two LDA calculations is that the resonance ↓ peak at E_F is higher for the unrelaxed structure. The same is true for the ↑ peak at about -1 eV. The main difference between LDA and GGA calculations is that some of the peaks shift slightly in energy. This is expected because of the differences in the LDA and GGA Fe band structure. In Figs. 7.12 (LDA) and 7.13 (GGA) the corresponding **k**-resolved transmission is shown for the unrelaxed coordinates. Also here the general behavior is similar to the one of Fig. 7.9, with the difference that the peaks due to the ↓ IS shift slightly in energy and height. For the unrelaxed structure the Fe Fermi energy is 1.8 eV below the MgO conduction band, and 4 eV above the MgO valence band.

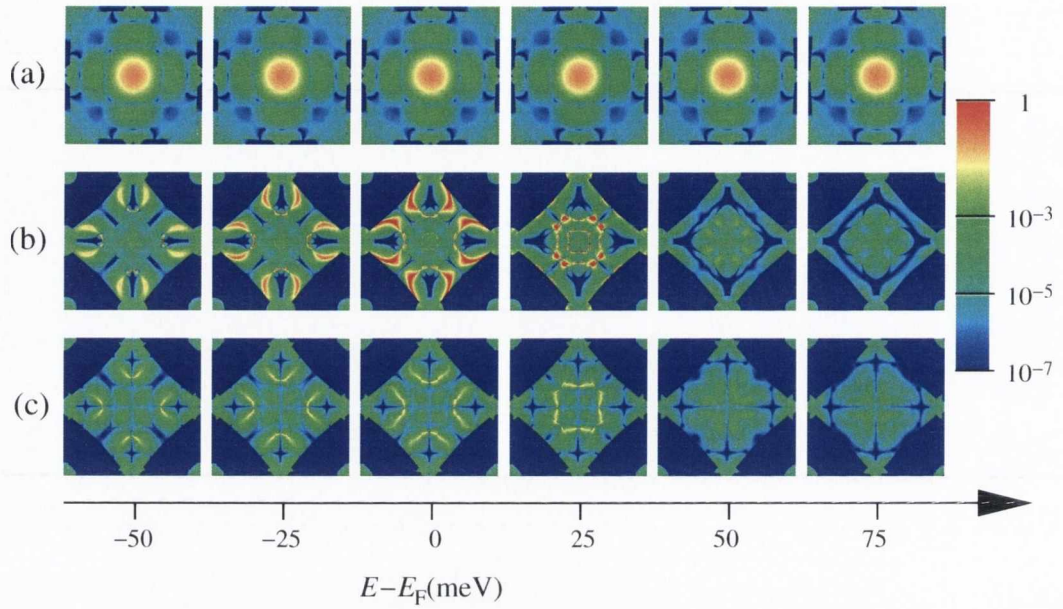


Figure 7.12: **k**-dependent transmission, as shown in Fig. 7.9, but calculated for the unrelaxed coordinates using the LDA.

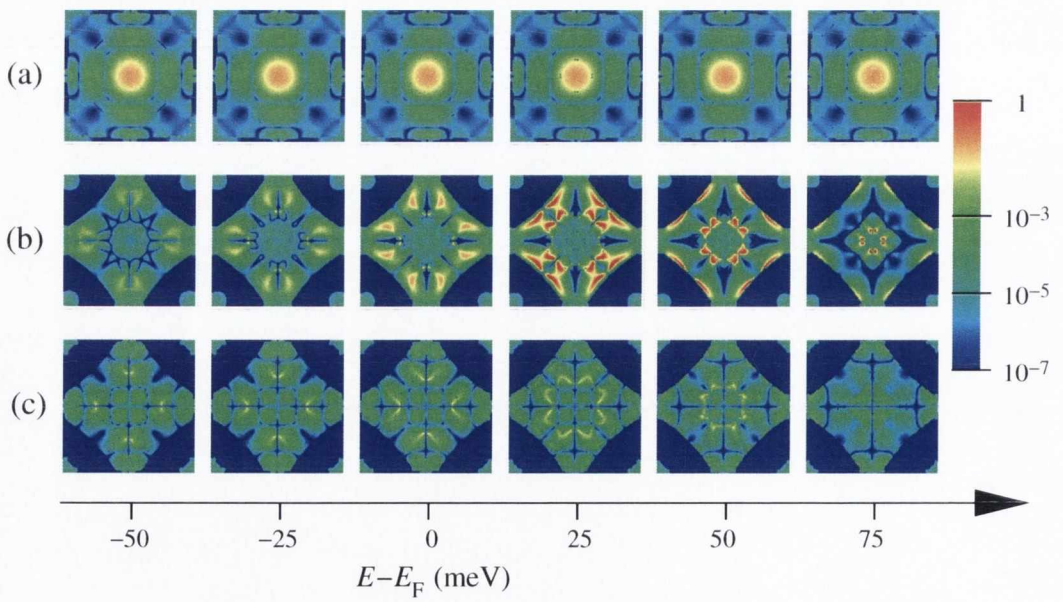


Figure 7.13: **k**-dependent transmission, as shown in Fig. 7.9, but calculated for the unrelaxed coordinates using the GGA.

7.3 Bound states in Fe/MgO/Fe(100) junctions

After having analyzed the main zero bias properties, we now move on to finite bias. In order to obtain the transport properties, we first have to calculate the self-consistent solution at finite bias. This is not straight forward, since bound states (BSs) may appear. In chapter 6 we have outlined in detail the general features of BSs, how to find them, and also how to occupy them for given physical constraints. The reason for the appearance of BSs is that there are **k**-points, for which there are no open channels in the Fe leads (see for example Fig. 7.10), as also described in Refs. [242, 65, 246]. For the majority spin the fraction of the BZ with zero channels at E_{F} is restricted to a small area close to the boundaries of the BZ (Fig. 7.10). Therefore it is reasonable to assume that the contribution from the BSs to the total charge is only a small perturbation. For the minority spin however the blue areas in Fig. 7.10 extend over approximately half the BZ. For all those **k**-points BSs can appear, and the contribution of the BSs to the total charge is therefore expected to be large. In this section we will show that at finite bias it is crucial to correctly take into account the BSs located in the bias window, in order to obtain a physically meaningful potential drop. For all the calculations in this section we use the unrelaxed coordinates of Ref. [20].

In Sec. 6.3 we have shown how BSs can be found by calculating the eigenvalues ϵ_n of the effective Hamiltonian H_{eff} [Eq. (6.41)]. We note that for each **k**-point the effective Hamiltonian is different. A BS is found at all real energies $E_{\mathrm{BS},\mu}$ where the conditions (6.45) and (6.46) are fulfilled, so that there is an n with $\epsilon_n(E_{\mathrm{BS},\mu}) = E_{\mathrm{BS},\mu}$. This implies that $\mathrm{Im}[\epsilon_n(E_{\mathrm{BS},\mu})] = 0$. As already explained, in numerical computations the second condition is usually not fulfilled exactly. We can therefore only calculate the set of generalized bound states (GBSs) with eigenvalues $\epsilon_\mu^{\mathrm{W}} = \epsilon_n(E_\mu^{\mathrm{W}})$ at a real energy E_μ^{W}, for which we only require that $\mathrm{Re}[\epsilon_n(E_\mu^{\mathrm{W}})] = E_\mu^{\mathrm{W}}$ and $\mathrm{Im}[\epsilon_n(E_\mu^{\mathrm{W}})] \leq \eta^{\mathrm{W}}$. The choice of η^{W} determines the set of weakly coupled states. For a true BS we require that η^{W} is of the order of the numerical precision. Here we choose the value of $\eta^{\mathrm{W}} = 10^{-14}$ Ry setting the boundary between weakly coupled and bound states, so that the set of GBSs for which $\eta^{\mathrm{W}} = 10^{-14}$ Ry corresponds to the set of true BSs. This value can also be increased in practical calculations, depending on the numerical accuracy.

We now calculate the BSs for a given **k**-point, using the graphical method described in Fig. 6.1 of Sec. 6.3, with $\eta^{\mathrm{W}} = 10^{-14}$ Ry. The chosen **k**-point is given by $k_x = k_y = 0.58\ \pi/a$. This is one of the **k**-points with vanishing number of channels in the $\downarrow$. Here, a is the Fe lattice constant. The position of this **k**-point is illustrated

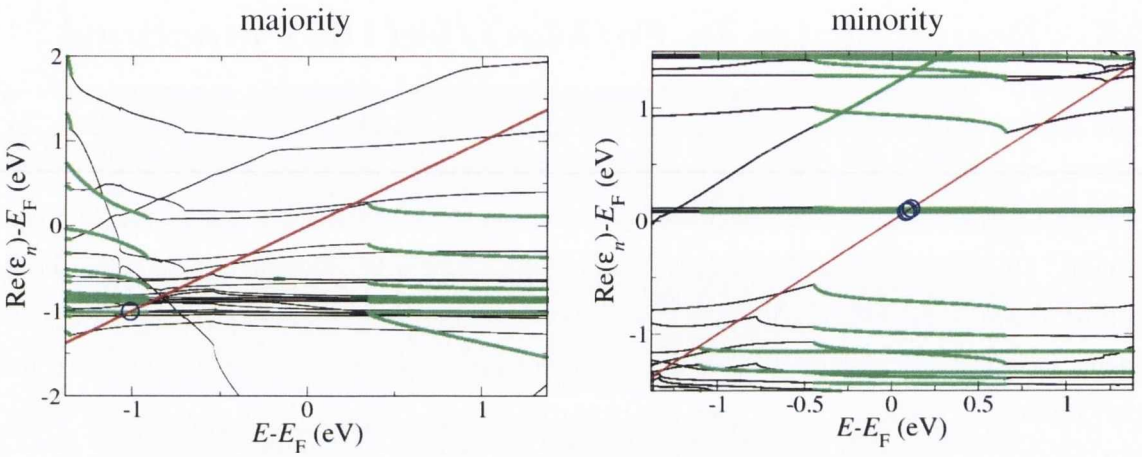


Figure 7.14: Real part of the eigenvalues ϵ_n of H_{eff}, for $\uparrow$ and $\downarrow$, as function of energy E, for $k_x = k_y = 0.58\ \pi/a$, and at zero bias. The green parts of the curves represent the set of eigenvalues with an imaginary part smaller than 10^{-14} Ry. The red line is a plot of the energy as function of the energy itself, the blue circles indicate the energies of the BSs.

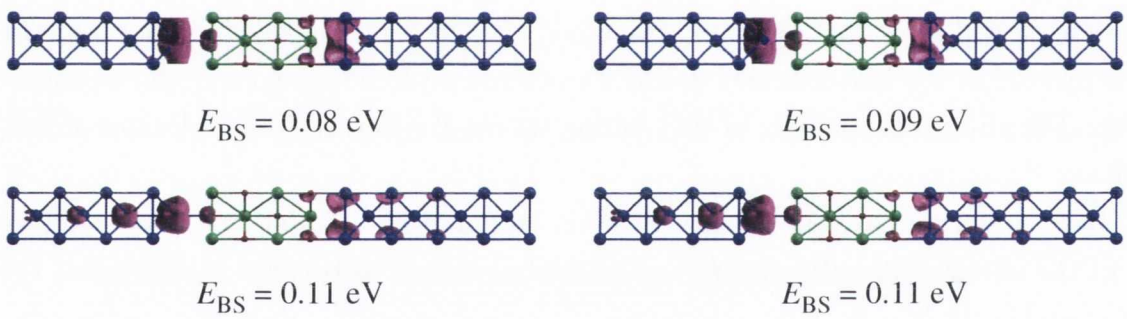


Figure 7.15: Charge density of the 4 $\downarrow$ interface states close to E_{F}, for $k_x = k_y = 0.58\ \pi/a$, and at zero bias.

with a cross in the right part of Fig. 7.20(a), where the **k**-resolved $\downarrow$ number of channels is shown. In Fig. 7.14 we plot the real part of the eigenvalues in a range of 1.5 eV around E_{F}. The black parts of the curves indicate that there $\text{Im}(\epsilon_n) > 10^{-14}$ Ry, whereas the light green parts indicate that there $\text{Im}(\epsilon_n) \leq 10^{-14}$ Ry. The red curve simply plots the energy as function of energy. A BS is found whenever the red curve cuts the green part of the eigenvalues, these crossings are emphasized by the blue circles. Whereas for the majority spins BSs are only found at about 1 eV below E_{F}, for the minority we find four almost degenerate states close to E_{F}. Therefore we demonstrate that the BSs in the minority can enter the bias window already at very small bias. At high bias, of the order of 1 V, it is possible that also BSs in the majority enter the bias window.

We now analyze the properties of the BSs in the $\downarrow$ close to E_{F}. At zero bias we

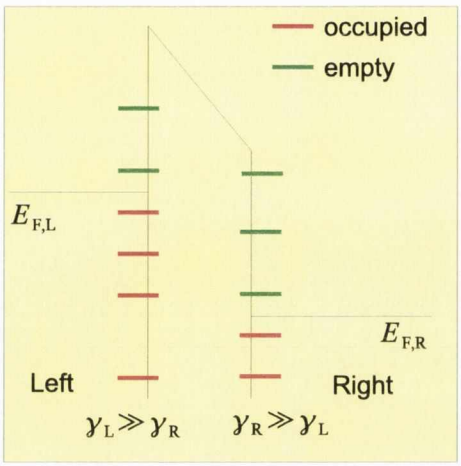


Figure 7.16: Schematic representation of the occupation of the BSs. Red lines represent occupied BSs, whereas green lines represent unoccupied BSs. At the left interface the BSs are occupied up to $E_{F,L}$, and at the right interface they are occupied up to $E_{F,R}$.

find two sets of BSs of different symmetry, and in each of the two sets there are two BSs. The reason for the appearance of two BSs with the same symmetry is that there is always one IS on the left-, and one on the right-hand side of the junction. At zero bias, and for such a symmetric junction, left and right ISs form a pair of bonding and anti-bonding states. This is also described in Ref. [242], and is a general feature of ISs. For the bonding state the total WF consists of an equally weighted linear combination of orbitals located on the left and on the right of the barrier with no phase shift, whereas in the anti-bonding state they have a 180° phase difference. Since the coupling across the MgO is very small, bonding and anti-bonding states are almost degenerate. For such a symmetric junction, and zero bias, the BSs therefore extend over both the left-hand and right-hand part of the junction. This is illustrated in Fig. 7.15, where we plot the probability density for all four BSs. At finite bias however we expect the states to localize either on the left, or on the right side, due to the breaking of the symmetry caused by the electrostatic polarization in the junction. In Sec. 6.10 we found a localization with bias of this type for the weakly coupled states in the Mn_{12} based magnetic molecule. Once the BSs localize on one side of the junction, we can expect that the BSs localized on the left-hand (right-hand) side to have a much stronger effective coupling to the left (right) lead [246]. The way to impose such an effective coupling has been described in Sec. 6.2, and is done by choosing an appropriate α matrix. For the tunneling junctions we simply adapt the solution given in Eq. (6.27) for a capacitor. The resulting α matrix is

$$\alpha = \begin{pmatrix} \mathbb{1}_{N_{EL}} & \mathbb{0}_{N_{EL},N_{ER}} \\ \mathbb{0}_{N_{ER},N_{EL}} & \mathbb{0}_{N_{ER}} \end{pmatrix}. \tag{7.5}$$

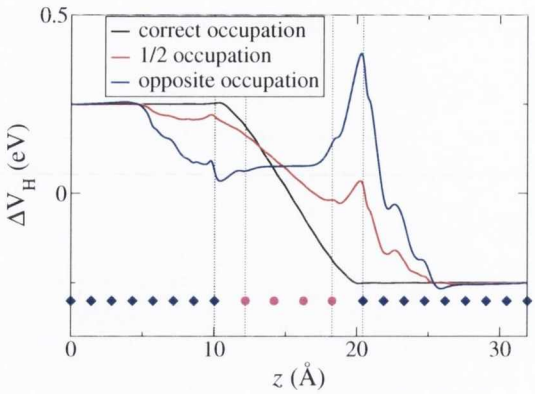


Figure 7.17: Planar average along x and y of the difference ΔV_{H} between the electrostatic potential at a bias voltage of 0.5 V and the one at zero bias, as function of position z, for different occupations of the BSs. The blue diamonds indicate the position of the Fe layers, and the magenta circles indicate the position of the MgO layers.

Here it is assumed that the indices of the orbitals in the EM are ordered from left to right; N_{EL} (N_{ER}) is the number of orbitals in the left (right) Fe electrode, plus the orbitals of the left (right) half of the MgO barrier. We use the approximations given in Eq. (6.23) to calculate the effective coupling matrices. We then calculate the BSs contribution to the non-equilibrium charge density using Eqs. (6.31) and (6.35). The resulting occupation of ISs from this choice of α matrix is schematically illustrated in Fig. 7.16. The left and right local Fermi energies are indicated as flat lines on the left and right side of the MgO barrier. The colored horizontal lines at the interfaces between the metal and the insulator denote ISs. Red lines indicate that the state is occupied, whereas green lines indicate that the state is empty. With this choice of α the result is that the BSs on the left side are occupied up to $E_{F,L}$, and the ones on the right side up to $E_{F,R}$.

In our calculations we use a finite value of δ equal to 10^{-4} Ry. We performed test calculations with a smaller value of 10^{-5} Ry, but the results were basically unchanged. The mesh spacing for the integrals on the real energy axis is always chosen somewhat smaller than δ. In order to verify the correctness of the obtained self-consistent solution we plot the difference ΔV_{H} between the planar average of the self-consistent Hartree potential at a finite bias and that at 0-bias along the junction stack (black curve in Fig. 7.17), for an applied bias of 0.5 V (the quantity ΔV_{H} and the notion of potential drop are introduced in Fig. 6.21 of Sec. 6.10). ΔV_{H} is approximately flat in the Fe electrodes and drops nearly linearly in the MgO. This is indeed the expected behavior, analogous to the result we obtain for junctions without BSs, indicating that the solution is correct. In order to emphasize the importance of the correct

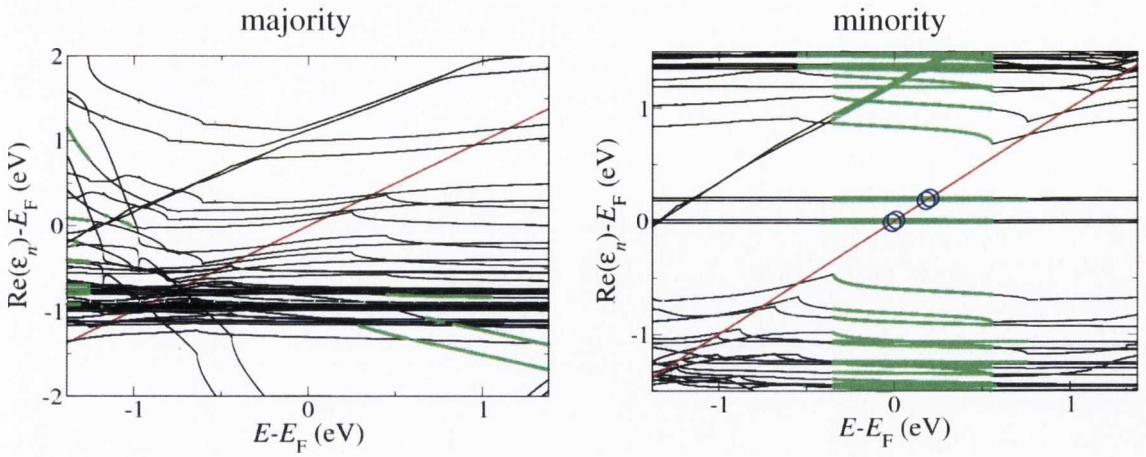


Figure 7.18: Real part of the eigenvalues ϵ_n of H_{eff}, for $\uparrow$ and $\downarrow$, as function of energy E, for $k_x = k_y = 0.58\ \pi/a$, and at an applied voltage of 0.2 V. The green parts of the curves represent the set of eigenvalues with an imaginary part smaller than 10^{-14} Ry. The red line is a plot of the energy as function of the energy itself, the blue circles indicate the energies of the BSs.

occupation of the BSs, we calculated the self-consistent solution using different α matrices. One possible choice is $\alpha = 0.5\mathbb{1}_N$, so that $\rho_{\text{neq}}^{\text{BS}}$ is zero [see Eqs. (6.31) and (6.35)]. This is equivalent to the result one would obtain by neglecting the BSs contribution. The result for such an α is shown as the red line in Fig. 7.17. This result is clearly unphysical, since the potential drops to a large extent also in the metal. If we choose α in such a way to obtain the opposite occupation to the physically meaningful one, so that the ISs on the left (right) side are occupied up to the right (left) quasi Fermi energy. In this case the self-consistent potential drop is the one shown as blue curve in Fig. 7.17. This solution is now even less physical, since the potential drops almost entirely in the metal, while it is approximately flat in the MgO. The transport properties of tunneling junctions at finite bias are to a large extent determined by the relative shift of the interface DOS on both sides of the junction. This shift is mainly determined by the potential drop. A wrong potential drop therefore leads to wrong predictions of the transport properties at finite bias. From the results of Fig. 7.17 it is clear that if the BSs are not correctly occupied, the resulting transport properties are largely unphysical, and can be very different from the correct ones.

We now investigate the properties of the BSs at a finite bias. In Fig. 7.18 the eigenvalues of H_{eff} are plotted as a function of energy E for an applied bias of 0.2 V, and at the same $\mathbf{k}$-point of Fig. 7.14. The colors of the curves are chosen in the same way as for the zero bias case, shown in Fig. 7.14. No BS is found for majority spin, indicating that the BS found at 0 bias is now coupled to one of the leads. The four BSs

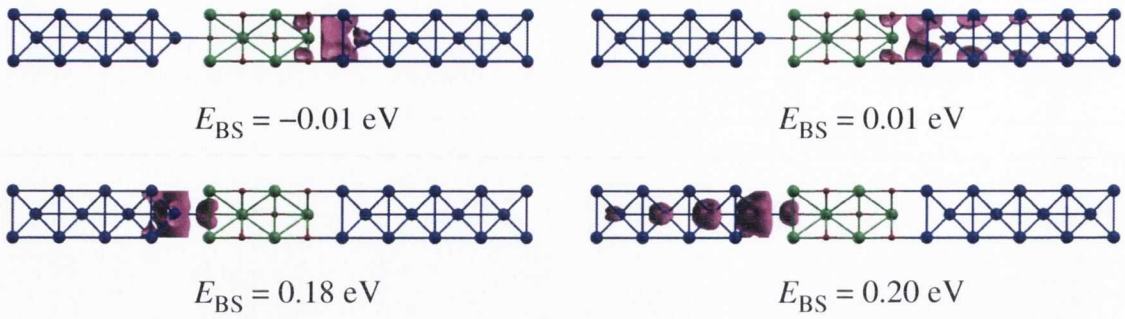


Figure 7.19: Charge density of the 4 ↓ interface states close to E_F, for $k_x = k_y = 0.58\ \pi/a$, and at a bias voltage of 0.2 V.

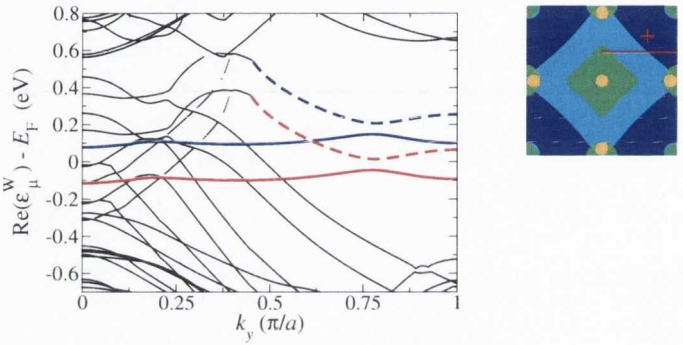


Figure 7.20: $\mathrm{Re}[E^W(\eta^W \to \infty)]$ for ↓ as function of k_y for a fixed $k_x = 0.37\ \pi/a$, at 0.2 V. The red and blue parts of the curves show $\mathrm{Re}[E^W(\eta^W = 10^{-3}\ \mathrm{Ry})]$, and are therefore the eigenvalues corresponding to weakly coupled states. The red (blue) color indicates the weakly coupled states associated with the right (left) lead. The right figure shows the number of ↓ open channels in the Fe leads. The red cross indicates the **k**-point where $k_x = k_y = 0.58\ \pi/a$, the red line is the path in **k**-space along which the left part of the figure is calculated.

in the minority bands are now split in two sets, separated by 0.2 eV, corresponding to the applied bias potential. In Fig. 7.19 the charge density is plotted for the 4 ↓ BSs, for each of them its position on the real energy axis is also given. It can clearly be seen, that the ones at higher energy are located on the left side of the junction, and the ones at lower energy are located on the right side. This confirms our previous assumption, that the BSs localize at one of the two interfaces at finite bias. From the change of the energy of the BSs at finite bias (Fig. 7.19) compared to zero bias (Fig. 7.15), one can see that the BSs on the left side follow closely $E_{F,L}$, so that in this case they are shifted up in energy by 0.1 eV, and the ones on the right side follow closely $E_{F,R}$, so that in this case they are shifted down in energy by 0.1 eV. This is the general behavior expected for highly asymmetrically coupled states.

Up to now the analysis of the BSs was restricted to one **k**-point, the next step is to determine the behavior of the BSs over the full 2D BZ. To this aim we first calculate

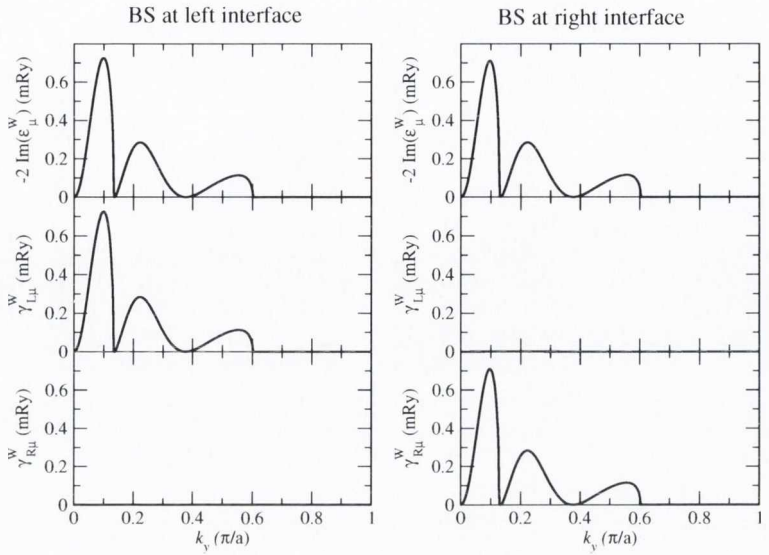


Figure 7.21: Imaginary part of the GBS eigenvalue $\epsilon_\mu^{\mathrm{W}}$ and the corresponding coupling coefficients $\gamma_{\{\mathrm{L/R}\},\mu}$ as function of k_y, for $k_x = 0.37\ \pi/a$, at 0.2 V. In the left (right) column of the figure the chosen set of eigenvalues $\epsilon_\mu^{\mathrm{W}}$ corresponds to the ones, whose real part is shown as solid blue (red) line in Fig. 7.20.

all the GBSs for $\eta^{\mathrm{W}} \to \infty$, so that also all crossings of the black curves in Fig.7.18 are included in the set of GBSs. In Fig. 7.20 all the found values for $E^{\mathrm{W}}(\eta^{\mathrm{W}} \to \infty)$ are plotted as function of k_y, for fixed $k_x = 0.37\ \pi/a$. The corresponding path in the 2D BZ is indicated as red line in the right part of Fig. 7.20, where the $\mathbf{k}$ resolved number of channels is shown over the whole 2D BZ. The discontinuities in the curves are for states with very large imaginary part, which usually have also a very large $\partial E^{\mathrm{W}}/\partial E$, so that they are difficult to find with the interpolation scheme. For smaller imaginary parts however the BSs are found reliably. The colored parts of the curves are the ones where the imaginary part of $\epsilon^{\mathrm{W}} < 10^{-3}$ Ry, and therefore correspond to weakly coupled states. All these states are ISs, since their coupling to the leads is very small. The blue (red) solid and dashed curves are the eigenvalues corresponding to weakly coupled states on the left (right) side of the junction. It can be seen that the blue curves are shifted from the red curves by about 0.2 eV. This again confirms that at self-consistency left- and right-hand side ISs are split in energy by the bias voltage.

In Fig. 7.21 the imaginary part, and the coupling coefficients $\gamma_{\{\mathrm{L/R}\}}$ [Eq. (6.54)] are shown for two sets of weakly coupled states. In the first (second) column of Fig. 7.21 the results are shown for the set of the eigenvalues corresponding to the solid blue (red) curve in Fig. 7.20. In the left column therefore we follow the evolution of the coupling coefficients as function of k_y for the IS on the left side, whereas in

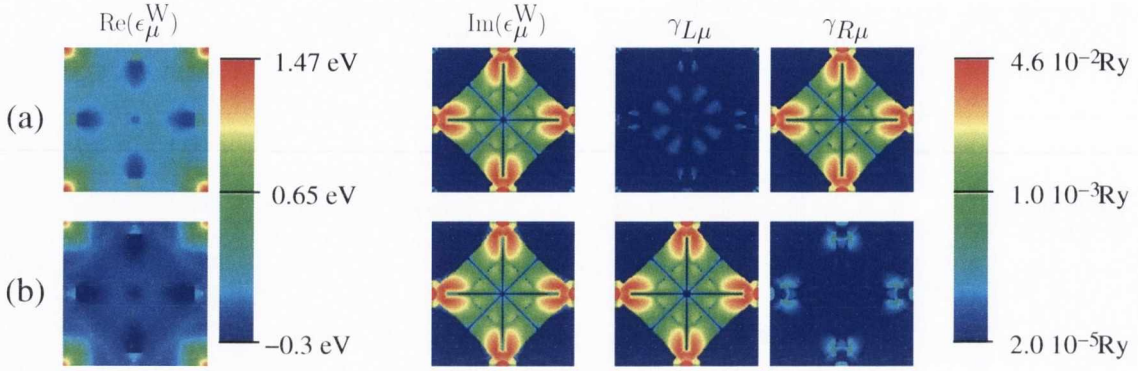


Figure 7.22: Re(ϵ_μ^W), Im(ϵ_μ^W), γ_L and γ_R as function of k_x and k_y, for the weakly coupled states of Fig. 7.21, extended over the whole 2D BZ (see text for details).

the second column the same is shown for the corresponding IS on the right side. For $k_y \gtrsim 0.6\ \pi/a$ the value of Im(ϵ_μ^W) goes to zero, showing that at these **k**-points these IS are BSs. A comparison with the right part of Fig. 7.20 shows that this is the range where the number of channels in the leads is zero. For $k_y \lesssim 0.6\ \pi/a$ there are channels in the leads, and in fact the Im(ϵ_μ^W) is usually different from zero, except at a some special high symmetry points and lines. One of these special lines is at $k_y = 0$, and an other one for example for $k_y = k_x$. For the part of the curves where Im(ϵ_μ^W) $\neq 0$, and for the IS located on the left interface, the value of $\gamma_{L,\mu}^W$ is much larger then the one of $\gamma_{R,\mu}^W$. For the IS located on the right side the opposite is true. This further motivates our choice of effective coupling for such a tunneling junction [Eq. (7.5)].

In order to obtain the evolution of the coupling of the IS to the leads over the full BZ, we calculate the data shown in Figs. 7.20 and 7.21 for the whole BZ. The results are shown in Fig. 7.22, where each small square figure extends over the 2D BZ. The figures on the top (bottom) row are for the IS on the right-hand (left-hand) side of the junction. The first column of figures on the left shows the evolution of the real part of the IS energy, on a linear color scale. The energy of the states is approximately flat over a large part of the BZ. In the region where the dispersion is large, the assignment of the states is not unique, since crossings with other levels occur. Here we follow the states in such a way, that they are the flattest possible, while preserving the symmetry of the states. One can see that the IS located at the left interface always has a higher energy then the one located on the right interface. In the other sub-figures Im(ϵ_μ^W), γ_L, and γ_R are shown. A comparison of the figures for Im(ϵ_μ^W) with Fig. 7.10 for the minority shows that in all those areas of the BZ, where the number of channels is zero, the IS is a BS. Moreover Im(ϵ_μ^W) vanishes along

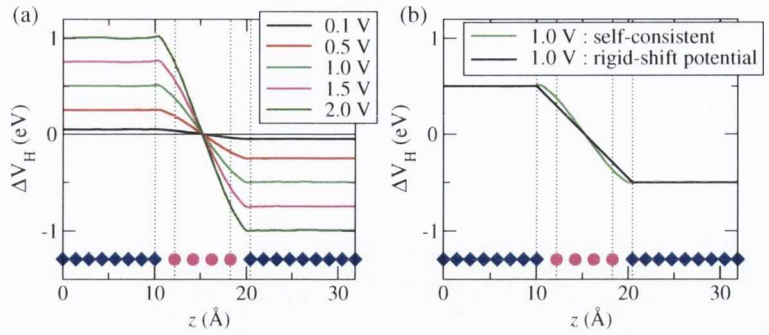


Figure 7.23: (a) Planar average along x and y of the difference ΔV_{H} between the electrostatic potential at a finite bias voltage and the one at zero bias, as function of position z, for different bias voltages. The blue diamonds indicate the position of the Fe layers, and the magenta circles indicate the position of the MgO layers. (b) Comparison of the self-consistent solution for ΔV_{H} (green curve) with the one obtained by applying the rigid shift approximation (black curve) for a bias of 1 V.

the high symmetry line where $k_x = k_y$, and along pars of the lines where $k_x = 0$ and $k_y = 0$. A comparison with the $\mathbf{k}$-dependent transmission (Fig. 7.9) for the $\downarrow$ spin shows that for those regions, where $\mathrm{Im}(\epsilon_\mu^{\mathrm{W}})$ vanishes, also the transmission through ISs is suppressed. This is for example the case along the previously described high symmetry lines. Whereas for the $\uparrow$ spins the modulation of the transmission over the BZ has been ascribed to the MgO complex band structure, the modulation of the transmission through ISs is attributable to both their damping across the MgO barrier and to their coupling to the Fe leads. The figures for γ_{L} and γ_{R} again show that the IS on the left-hand side has a much stronger coupling to the left lead, whereas the one on the right-hand side is coupled stronger to the right lead. The coupling of a IS to the opposite site of the junction is not zero, but is very small due to the MgO barrier. At very high bias this can however change.

In the last part of this section we investigate the potential drop and charging as function of bias. Fig. 7.23(a) shows ΔV_{H} for a set of bias voltages, ranging from 0.1 V to 2.0 V. For all voltages ΔV_{H} is approximately flat in the Fe leads, and drops almost linearly in the MgO. The corresponding planar average of the difference $\Delta\rho$ between the real space charge at finite bias and the one at zero bias is shown in Fig. 7.24. It satisfies the relation $\Delta\rho(z) \propto -d^2\Delta V_{\mathrm{H}}(z)/dz^2$. Since the value of $\Delta\rho$ depends on the bias voltage V, instead of plotting $\Delta\rho$ itself, we plot $\Delta\rho/eV$ (e being the electron charge). We use arbitrary units, although they are the same for all the graphs. For all different voltages the curve of $\Delta\rho/eV$ as function of position z is basically identical. This clearly indicates that the charging in the junction, and especially at the interface layers, is linear in V. The reason for this linear behavior is that the magnitude of the

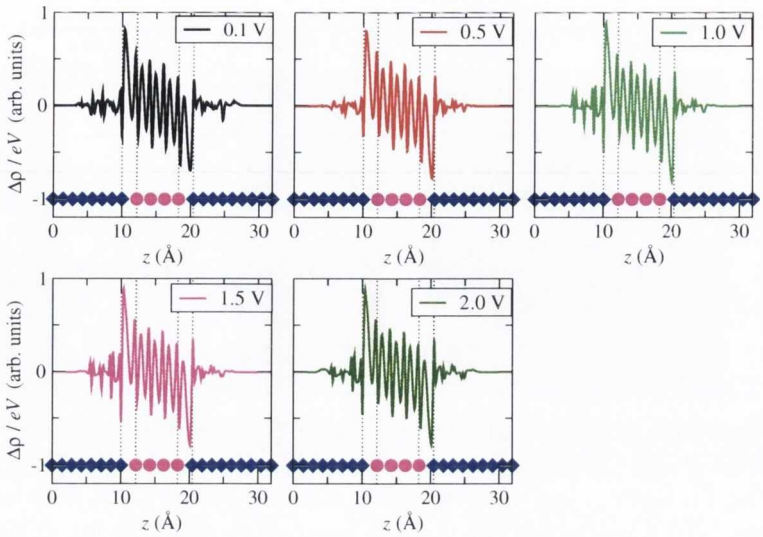


Figure 7.24: Planar average along x and y of the finite bias charging $\Delta\rho$, scaled by the applied bias voltage potential V and the electron charge e, as function of position z. Each sub-figure shows $\Delta\rho/eV$ for a different bias voltage, the color code for the different voltages corresponds to the one in Fig. 7.23. The blue diamonds indicate the position of the Fe layers and the magenta circles indicate the position of the MgO layers.

planar average of the difference in charge is much smaller than the one of the planar average of the charge itself. The ratio of the two quantities is of the order of 10^{-3}. The finite bias therefore leads only to a very small perturbation of the total charge. Even though the relative change in charge is very small, it is large enough to modify the potential by a few eVs. At all voltages we find $\Delta\rho(z)$ increasing linearly with V, and charge accumulating at the extremal layers of the electrodes just before MgO. Inside the MgO $\Delta\rho(z)$ oscillates due to the electric field induced polarization. This is confirmed by a DFT calculation for an isolated MgO slab of the same thickness in an equivalent electric field, which shows analogous oscillations.

Our calculated potential drop justifies the often used rigid shift approximation of the potential for calculating the current at finite bias. In that approximation the finite bias transmission coefficient is calculated using the Hamiltonian calculated self-consistently only at zero bias, plus a rigid shift in the electrodes chemical potential by the applied bias. Across the barrier a linear potential drop is assumed [274]. A comparison of the self-consistent ΔV_{H} with the one applied in the rigid shift approximation is shown in Fig. 7.23(b) for a bias voltage of 1 V, where the black curve corresponds to the rigid shift approximation, and the green curve is the self-consistent solution. The calculation of the self-consistent solution at finite bias is much more difficult to achieve than that at zero bias, since the algorithm is much more involved,

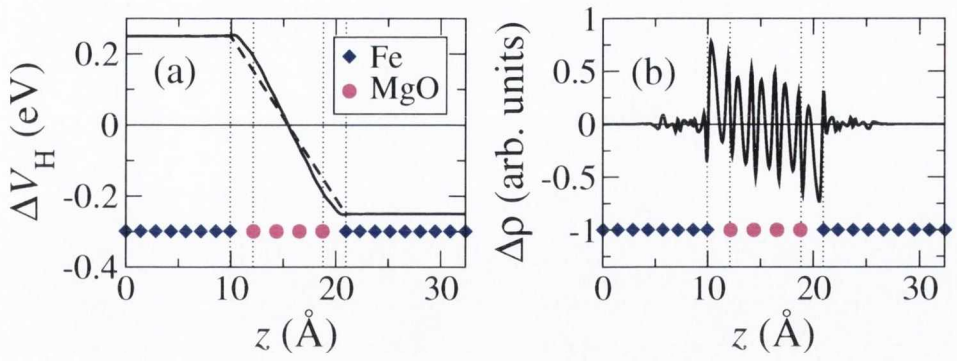


Figure 7.25: (a) Planar average ΔV_{H} of the difference between the Hartree potential at 0.5 V and the one at 0-bias (full line), the dashed line indicates ΔV_{H} applied in the rigid shift approximation. (b) Planar average $\Delta\rho$ of the difference between the charge density at 0.5 V and the one at 0-bias. The diamonds and dots indicate the location of the Fe and MgO layers.

and also since the needed computational resources are much larger. The rigid shift approximation is therefore a way to obtain approximate finite-bias I-V curves in a highly simplified way. The good agreement between the self-consistent potentials and the rigid-shift potential, together with the fact that the charging in the junction is small, indicates that the use of this approximation is well justified. We will further corroborate the validity of this approximation at the end of Sec. 7.4.

7.4 Bias-dependent transmission and I-V curves

In this section we investigate the I-V characteristics of the 4 MgO ML junction. We demonstrate that the features of the I-V originate from the sweeping of the Δ_1 band-edges and of ISs across the bias window. These generic features are emphasized here, as they resemble closely those studied in molecular devices [68] and magnetic point contacts [320]. We use the relaxed coordinates and the LDA.

First, we look at the electrostatic Hartree potential drop across the junction for the relaxed structure. In Fig. 7.25(a) we show the difference ΔV_{H} between the planar average of the self-consistent Hartree potential at a finite bias and that at 0-bias along the junction stack (z-axis). In the same way as for the unrelaxed coordinates, ΔV_{H} is flat in the electrodes and decays linearly in the MgO. The dashed line indicates the potential drop for the rigid shift approximation, described at the end of the previous section. In this section we will also directly compare the I-V obtained with the self-consistent solution with the one obtained using the rigid shift approximation. In Fig. 7.25(b) we also show the difference $\Delta\rho$ between the planar average of the self-consistent charge density at a finite bias and at 0-bias. Also $\Delta\rho$ shows an analogous

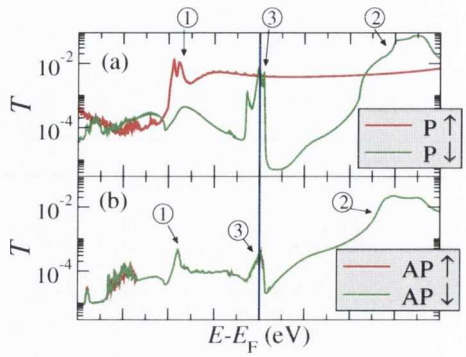


Figure 7.26: Spin dependent transmission coefficient T for P (a) and AP (b) alignment as a function of energy E.

behavior to the one for the unrelaxed structure (Fig. 7.24).

We start our analysis of the transport properties at finite bias by first analyzing in detail the zero-bias transmission in the energy range relevant for a bias of up to 2 V. To this aim in Fig. 7.26 the spin-polarized zero-bias transmission is shown in an energy range of ± 2 eV around E_F. As briefly discussed in Sec. 7.2, three features appear in the $V = 0$ transmission coefficient: ① a sharp increase (note the logarithmic scale of the figure) in transmission at around -1 eV for the $\uparrow$ spins in the P configuration, ② a similar, although smoother increase above +1 eV for the $\downarrow$ spins in the P configuration and for the AP configuration, and ③ a sharp resonance at E_F for the $\downarrow$ spins in the P configuration, which is also present in the AP configuration, although much smaller.

In order to map the features in the transmission to the corresponding electronic states, in Fig. 7.27 we plot the Fe band-structure as function of k_z for $k_x = k_y = 0$ [Fig. 7.27(a)], the P transmission [Fig. 7.27(b)], the average number of open channels for Fe [Fig. 7.27(c)], the Fe bulk DOS [Fig. 7.27(d)], and the Fe PDOS at the interface layer [Fig. 7.27(e)]. A comparison of the transmission with the DOS for the bulk and for the interface layer shows that the $\downarrow$ resonance in the transmission is caused by the $\downarrow$ IS. A comparison of the transmission with the band-structure suggests that the first two features are associated with the Δ_1 band-edges, respectively for the $\uparrow$ and $\downarrow$ spins. However, from Fig.7.27 this assignment is not completely clear. For the $\uparrow$ spins the transmission starts to rise at the $\Delta_1^{\uparrow}$ band-edge, and then it is approximately constant for increasing energy. There is also a sharp resonance in the $\uparrow$ transmission just above the $\Delta_1^{\uparrow}$ band-edge. In contrast for the $\downarrow$ spins the transmission starts to rise abruptly already at about 0.5 eV below the $\Delta_1^{\downarrow}$ band-edge. It then forms a broad peak that extends to about 0.5 eV above the $\Delta_1^{\downarrow}$ band-edge. The peaked structure of the transmission might indicate the presence of ISs.

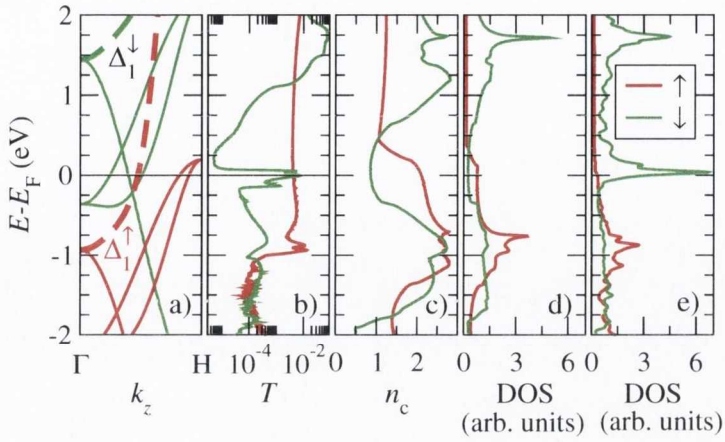


Figure 7.27: (a) Bulk Fe band-structure along the $\Gamma \rightarrow$ H direction (the bands with Δ_1-symmetry are emphasized), (b) T in the P configuration, (c) average number of open channels per **k**-point for bulk Fe, n_c, (d) bulk Fe DOS, (e) interface Fe-layer DOS. Note that the Δ_1 band-edges coincide approximately with a rather sharp increase in the transmission coefficient T.

In order to clarify the exact nature of features ① and ②, in Fig. 7.28 we analyze the transmission and layer-dependent Δ_1 PDOS at the Γ point only ($k_x = k_y = 0$). All the PDOSs are given in arbitrary units, positive values are used for the majority, and negative values for the minority spins. The PDOS within the transport calculation is obtained by using the retarded GF and Eq. 4.11. In order to obtain the layer-dependent Δ_1 PDOS we sum up the PDOS for all the s, p_z and d_{z^2} Fe orbitals in a given Fe layer. In Fig. 7.28(a) the Δ_1 PDOS is shown for bulk Fe. The Van Hove type singularities at -0.93 eV for the $\uparrow$ spins, and at +1.47 eV for the $\downarrow$ spins clearly indicate the band-edges of the states with Δ_1 symmetry in the two spins. Below these two band-edges the Δ_1 PDOS is exactly zero. A comparison with the transmission coefficient [7.28(d)] shows that at all the Van Hove singularities it changes discontinuously. It does not go to zero below the Δ_1 band-edges, since there are also contributions to the transmission coming from states with different symmetry, and from different **k**-points. From the figure one can however see that this contribution is much smaller than the contribution from the Δ_1 states. The Δ_1 PDOS for the 8th layer is similar to the bulk PDOS, the oscillations indicate the effects of the surface on the DOS. The structure of the Δ_1 PDOS for the interface Fe is however different. For the $\uparrow$ spins the Van Hove type peak in the Δ_1 PDOS at -1 eV is still present. In the $\downarrow$ however there is no Van Hove type peak. The band-edge is rather split into two localized peaks. We note that in order to calculate the PDOS we added an imaginary part $\delta = 10^{-3}$ Ry to visualize the two peaks. The peak below the $\Delta_1^\downarrow$ band-edge would otherwise not be visible, since it is a BS. The peaks indicate

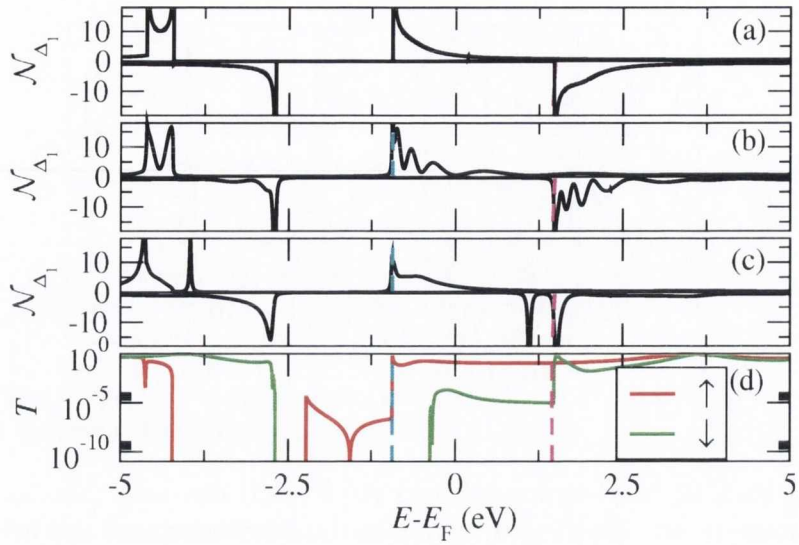


Figure 7.28: DOS $\mathcal{N}_{\Delta_1}$ projected on the states with Δ_1 symmetry, in arbitrary units (positive values for ↑, negative values for ↓), and transmission at the Γ point. (a) Δ_1 DOS for bulk Fe, (b) Δ_1 DOS of the Fe atom placed 8 ML away from the Fe/MgO interface, (c) Δ_1 DOS of the Fe atom at the Fe/MgO interface. (d) transmission for ↑ (red) and ↓ (green). The dashed vertical lines indicate the band-edges of the Δ_1 states at the Γ point for ↑ (cyan) and ↓ (magenta). In order to visualize all the peaks in the DOS, a small imaginary part has been added to the energy.

a IS, the reason for the occurrence of two peaks is that the ISs on the left and right side of the junction form a set of bonding- and anti-bonding states. We note that these ISs can still be regarded as a band-edge feature, since they clearly derive from the bulk $\Delta_1^{\downarrow}$ band-edge.

Whereas at the Γ-point the Δ_1 band-edges lead to steps in the transmission, in the total P transmission these are smoothed out. This is due to the evolution of the Δ_1 layer PDOS for small k_x and k_y. In Fig. 7.29 the bulk Fe Δ_1 PDOS is plotted for $k_x = k_y = 0$ (black curve) and for $k_x = k_y = 0.1\ \pi/a$ (red curve). The Van Hove singularities are found at almost the same energies, however $k_x = k_y = 0.1\ \pi/a$ the Δ_1 PDOS does not drop to zero below the band-edges. Instead it has a small finite value up to about 2 eV below the band-edges, where very sharp Van Hove singularities appear. This can indeed lead to high transmission even below the Γ-point Δ_1 band-edges. In order to verify this we plot the transmission and layer-dependent Δ_1 PDOS at $k_x = k_y = 0.1\ \pi/a$ (Fig. 7.30). The figure shows the same quantities as Fig. 7.28, but for a different **k**-point. The layer-dependent Δ_1 PDOS is similar to the one for the Γ-point. There are however three major differences: the first is that the small finite Δ_1 contribution below the band-edges leads to a smooth decay of the transmission as one moves down in energy, compared to the step-like behavior for

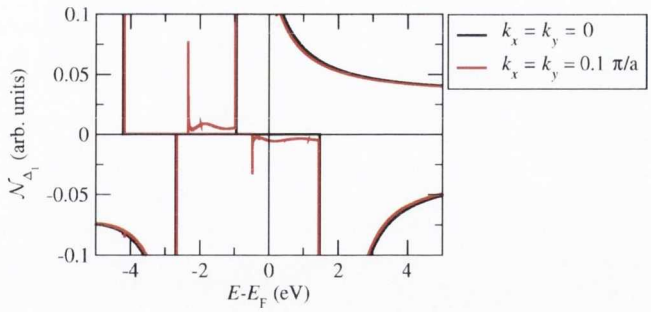


Figure 7.29: DOS $\mathcal{N}_{\Delta_1}$ projected on the states with Δ_1 symmetry for bulk Fe, at two different **k**-points, in arbitrary units (positive values for ↑, negative values for ↓). The black curves are for the DOS at $k_x = k_y = 0$, the red curves are for the DOS at $k_x = k_y \approx 0.1\ \pi/a$.

the Γ-point. The second is that both ↓ ISs at about 1 eV are coupled to the leads, and therefore both contribute to the transmission, whereas for the Γ-point only the part of the peak above the band-edge contributes to the transmission. We note that, in contrast to Fig. 7.28(c), in order to plot the interface Fe PDOS in Fig. 7.30(c) we did not add any imaginary part to the energy, so that the width of all the peaks reflects their coupling to the leads. Adding a finite imaginary part to the energy did not reveal new peaks, demonstrating that at this **k**-point there are no BSs. The third feature is that the position and separation of the two peaks in the ↓ PDOS above one eV changes.

With these informations the origin of the structure of feature ② is explained in the following way. The bulk Fe $\Delta_1^{\downarrow}$ band-edge at about 1.5 eV leads to a surface state, which due to the coupling across the MgO barrier splits in a pair of bonding- and anti-bonding states. This explains the double-peaked structure of feature ②. While for the Γ-point only the part of the IS above the band-edge has finite transmission, because the bulk $\Delta_1^{\downarrow}$ PDOS is exactly zero below the band-edge. For other **k**-points however the $\Delta_1^{\downarrow}$ PDOS has a finite value also below the main Δ_1 band-edge, the energy range over which this finite value extends increases with increasing $|k_x|$ and $|k_y|$. The part of the two ISs PDOS contributing to the transmission therefore increases with increasing $|k_x|$ and $|k_y|$. For this reason both ↓ IS PDOS peaks are visible in the transmission, but the one at lower energy is much smaller in amplitude than the one at higher energy. Also the rather smooth structure of feature ① is explainable in the same way with the **k**-dependent ↑ layer-dependent Δ_1 PDOS. Since the interface $\Delta_1^{\uparrow}$ PDOS is closer to the bulk PDOS, the transmission does not show a pronounced peaked structure, but is rather constant over energy. Only a small and rather sharp resonance is formed.

For the AP configuration we only analyze the ↑ transmission, since by symmetry

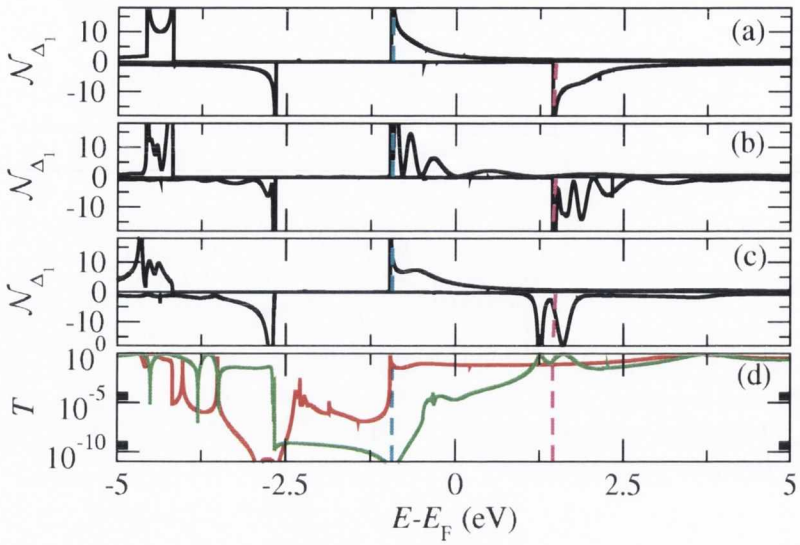


Figure 7.30: DOS $\mathcal{N}_{\Delta_1}$ projected on the states with Δ_1 symmetry, in arbitrary units (positive values for $\uparrow$, negative values for $\downarrow$), and transmission at a **k**-point close to the Γ point ($k_x = k_y \approx 0.1\ \pi/a$). (a) Δ_1 DOS for bulk Fe, (b) Δ_1 DOS of the Fe atom placed 8 ML away from the Fe/MgO interface, (c) Δ_1 DOS of the Fe atom at the Fe/MgO interface. (d) transmission for $\uparrow$ (red) and $\downarrow$ (green). The dashed vertical lines indicate the band-edges of the Δ_1 states at the Γ point for $\uparrow$ (cyan) and $\downarrow$ (magenta). Since all the peaks are coupled to the leads, no imaginary part has been added in order to broaden them.

the antiparallel is identical. In the $\uparrow$ AP configuration the Δ_1 PDOS on the left side of the junction is the one for P $\uparrow$, and the one on the right side is the one for P $\downarrow$. Feature ① at -1 eV is therefore barely visible, because the P $\downarrow$ PDOS is zero at that energy. Only a small resonance is notable at that energy. For increasing energy however the P $\downarrow$ PDOS increases, so that close to the $\Delta_1^{\downarrow}$ band-edge the transmission increases dramatically. The IS in the $\Delta_1^{\downarrow}$ therefore leads to a large peak in the transmission, which is almost of the same size as the corresponding one in the P configuration. Since the IS is only found on one side of the junction, it does not split up into bonding- and anti-bonding states. The IS resonance at E_F (feature ③) is also present in the AP configuration, although drastically reduced in size. The $\downarrow$ IS at E_F has also a Δ_1 component, so that it can in principle couple to the $\Delta_1^{\uparrow}$ states on the other side of the junction. However whereas the P $\uparrow$ transmission is large close to the Γ-point, for the $\downarrow$ ISs it is large at other positions in the BZ. Therefore in the AP configuration the transmission is strongly reduced.

We now move on to analyzing the bias-dependent transmission coefficient. Its behavior is mainly driven by the evolution of the described three main features with bias. In Figs. 7.31 and 7.32 the bias-dependent spin-polarized transmission is shown for P and AP configuration of the electrodes. First we describe the evolution of

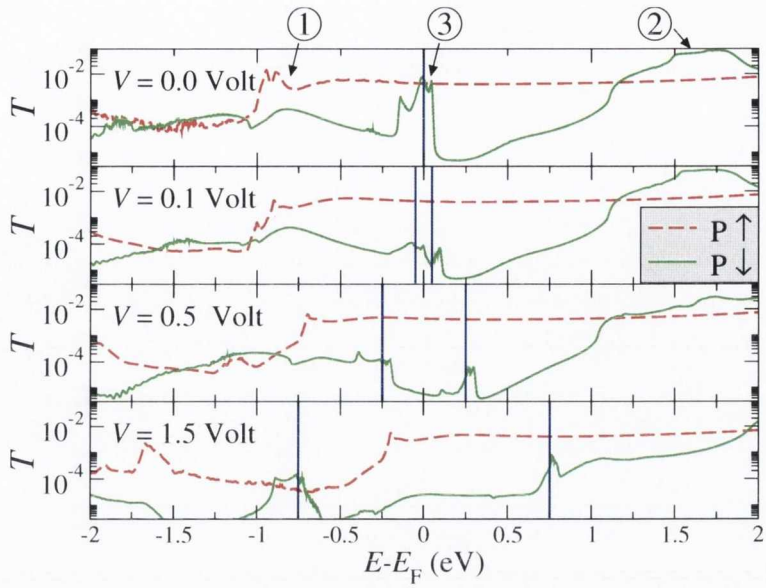


Figure 7.31: Spin dependent transmission coefficient T for the P configuration as a function of energy, E, and for different biases, V. The vertical lines are placed at $E = E_F \pm eV/2$ and enclose the bias window.

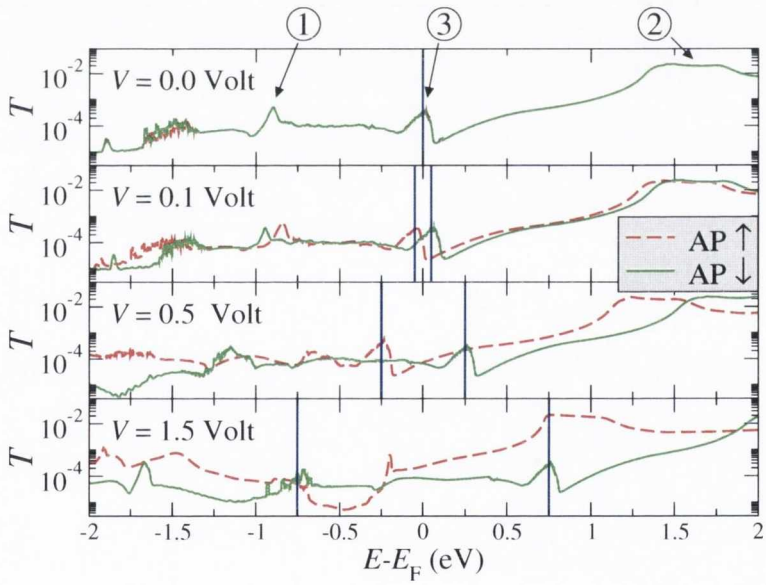


Figure 7.32: Spin dependent transmission coefficient T for the AP configuration as a function of energy, E, and for different biases, V. The vertical lines are placed at $E = E_F \pm eV/2$ and enclose the bias window.

the tree features for P alignment, and then for AP alignment. Since feature ① is mainly determined by the electronic structure of Fe alone, the associated sharp rise in transmission changes position as a function of bias following $\Delta_1^{\uparrow}$ band-edge at higher energy. For positive bias this is the case for the one of the left-hand side lead, whose energy as function of bias is equal to $E_{\Delta_1}^{\uparrow} + eV/2$. For instance the

sharp increase of $T^{\uparrow}(E;V)$ moves to -0.75 eV and -0.25 eV respectively for voltages of 0.5 V and 1.5 V, following $E^{\uparrow}_{\Delta_1} + eV/2$. At a voltage of about 1 V the sharp rise in transmission enters the bias window. As the voltage increases even further, the increase in current is reduced, since the high transmission region in the P $\uparrow$ extends only over part of the bias window. The broad peaks associated to feature ② broaden with increasing bias, but also shrink considerably in size. At $V = 1.5$ V the peaks have almost disappeared. In the P configuration therefore the zero-bias peak in the $\downarrow$ transmission does not contribute substantially to the total current, since once it enters the bias window, its height is much smaller than the transmission amplitude for the $\uparrow$ spins. The evolution of feature ③, the sharp peak in the $\downarrow$ transmission at E_{F}, with bias shows a similar behavior. At finite bias, the zero-bias peak splits into two sub-peaks with a strongly reduced transmission, separated in energy by the bias voltage. The shrinking of the peak is complete at a bias voltage of about 20 meV, for higher voltages the height of the peaks is approximately constant.

The dynamics of features ② and ③ for changing bias are a general property of ISs, spatially localized at the interface between Fe and MgO. This is illustrated schematically in Fig. 7.33. Consider the panels (a) and (b), where two identical ISs are localized on either sides of the tunnel barrier. This is the situation encountered here for the $\downarrow$ spins. The transport is then resonating across the barrier at the IS energy $E_{\mathrm{R}} = E_{\mathrm{L}}$. As shown in Sec. 7.3, in general a IS is coupled more strongly to one of the electrodes, and it will trace closely its quasi Fermi energy. For instance for positive bias and a IS localized on the left-hand (right-hand) side of the junction [Fig. 7.33(b)] we obtain $E_{\mathrm{L}}(V) = E_{\mathrm{L}}(0) + eV/2$ [$E_{\mathrm{R}}(V) = E_{\mathrm{R}}(0) - eV/2$]. This brings the states on either side of the junction out of resonance and generally suppresses the transmission. Thus the peak in $T(E;0)$, originating from a resonating IS across the barrier, will evolve into two smaller peaks separated by an energy eV. Indeed this is the behavior observed in Fig. 7.32. For instance the two peaks centered at E_{F} are separated by 0.5 eV and 1.5 eV for voltages of 0.5 V and 1.5 V, respectively.

A second possible situation is when the ISs on the left-hand and right-hand side of the tunnel junction have a different origin and are placed at different energy. In this case we do not expect zero-bias resonance, however there will be a critical voltage at which the resonant condition is met. In this case we expect the rise of a large peak in the transmission coefficient at a bias $eV = E_{\mathrm{L}}(0) - E_{\mathrm{R}}(0)$. This situation has never been encountered for the symmetric MTJ investigated here, but it is likely to appear in real junctions.

We now analyze the bias-dependent transmission for AP alignment, shown in Fig

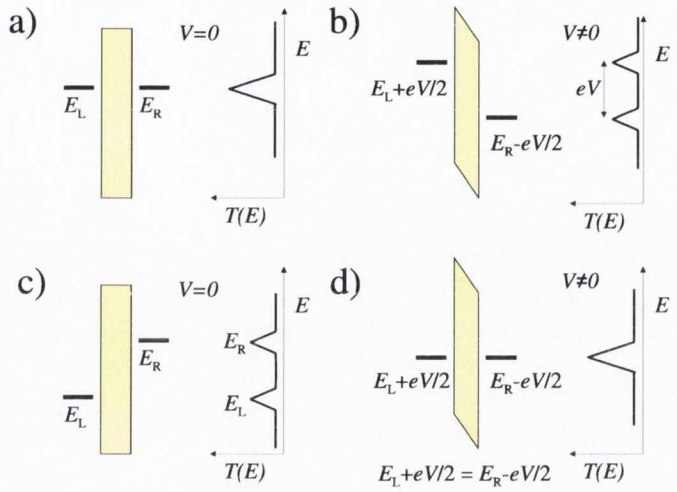


Figure 7.33: Schematic representation of the bias dependence of the transmission through interface state for P alignment. (a) and (b) are for the left (L) and right (R) interface states having the same energy and therefore resonating at V=0; (c) and (d) are for the L and R interface states not resonating at zero-bias.

7.32. In this case only features ② and ③, leading to the peaks at E_F and at about 1.5 eV respectively, are important. With increasing bias both peaks do not shrink in size, they just split up by the applied bias eV. The mechanism is illustrated schematically in Fig. 7.34 for the ↑, the situation for the ↓ is analogous. The only difference between the ↑ transmission and the ↓ transmission is that in the ↑ the peaks shift to lower energy, whereas in the ↓ they shift to higher energies. The spin ↓ IS on the right side of the junction couples to the broad $\Delta_1^{\uparrow}$ band on the left side. Therefore at any bias the transmission at the peak is approximately constant, but the peak shifts according to the energy of the IS. Only when the bias is large enough, that the energy of the IS is below the bottom of the Δ_1 band, the transmission peak will shrink. The result of such a bias-dependence is that the transmission peaks related to the sharp IS at E_F are pinned at the boundaries of the bias window for all voltages, and that the broad ↑ peak in transmission shifts closer to the bias window with increasing bias. This leads to an increase of the AP current, especially once the voltage is large enough that the main part of the peak can enter the bias window. This is in contrast to the case of the P configuration, where this peak never significantly contributes to the current.

We now move to analyzing the I-V characteristics (Figs. 7.35 and 7.36) and the TMR (Fig. 7.37). In Fig. 7.35 the spin-polarized currents are shown for P and AP alignment, and in Fig. 7.36 the total current $I(V)$, the conductance $G(V) = dI/dV$ and $S(V) = d^2I/dV^2$ are plotted, both for P alignment (left column) and for AP alignment (right column). The results are plotted for different values of the imaginary

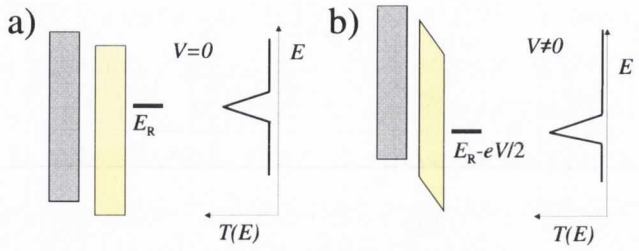


Figure 7.34: Schematic representation of the bias dependence of the transmission through interface states for AP alignment.

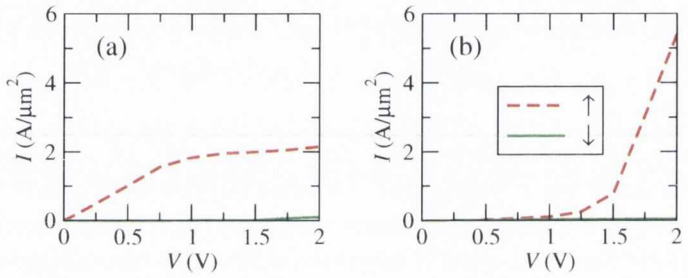


Figure 7.35: Spin-polarized current I as function of voltage V for P (a) and AP (b) alignment of the Fe electrodes.

part δ added to the energy for the evaluation of the transmission coefficient. We first focus on the case where $\delta = 0$ (black curves in Figs. 7.36 and 7.37), which means that no imaginary part is added. We start by analyzing the low bias region ($V < 0.4$ V). The most apparent feature in this bias range is a sharp reduction of the TMR from its zero-bias value followed by a rapid increase which peaks at $V \approx 0.3$ V. The sharp reduction can be associated with a decrease in the P $\downarrow$ spins current originating from the loss of the IS resonant condition at $V \approx 20$ mV, a bias which roughly corresponds to the line-width of the IS. For $V < 20$ mV the P current is shared by the two spin-species, while for $V > 20$ mV the $\uparrow$ component dominates. Such a reduction in the P $\downarrow$ component at the bias corresponding to the resonant condition loss can be clearly observed in the $S(V)$ plot of Fig. 7.36. As the bias further increases the conductance in the P configuration is approximately constant. In contrast the conductance of the AP configuration is slightly reduced for V between 0.1 and 0.3 V. This behavior results in an increase of the TMR with the broad peak at 0.3 V. The high bias region is characterized by a decrease of the conductance in the P configuration at around 1 V and a dramatic increase of the AP current for $V > 1.5$ V. This produces a strong reduction of the TMR with bias and an almost complete suppression for $V > 1.5$ V. In fact at about 1.75 V the TMR becomes negative. Such a voltage range should be put in comparison with the band offset. In our LDA calculations E_F for Fe is

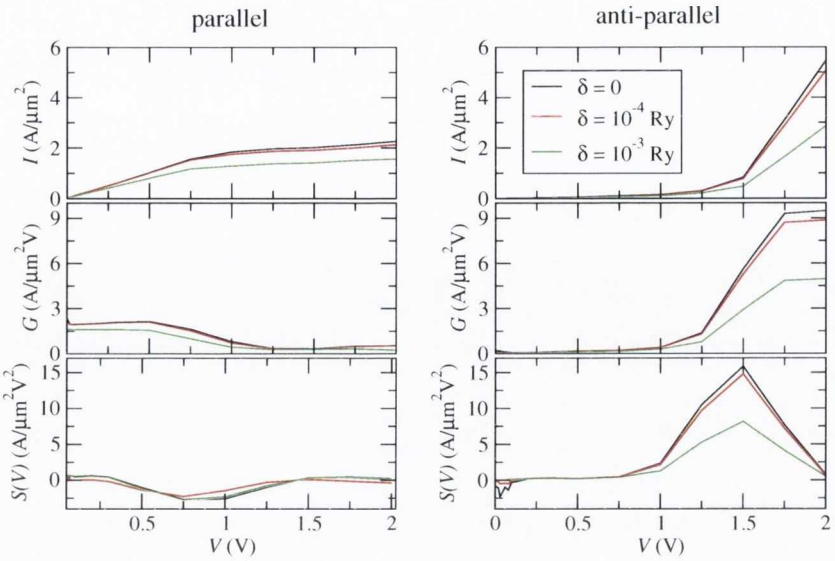


Figure 7.36: Current I, conductance G, and $S = d^2I/dV^2$ as function of voltage V for P (left figures) and AP (right figures) alignment of the Fe electrodes. The curves are plotted for different values of the imaginary part δ added to the energy.

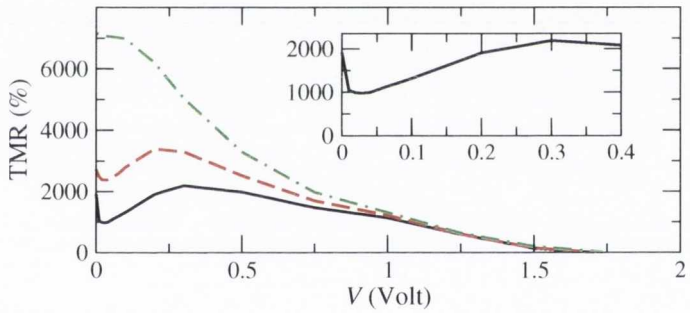


Figure 7.37: TMR as function of voltage V for different values of the imaginary part of the energy δ. The black curve is for $\delta - 0$, the red for $\delta = 10^{-4}$ Ry, and the green for $\delta = 10^{-3}$ Ry. In the inset the TMR is shown in the low-bias region for $\delta = 0$.

positioned ~1.8 eV below the MgO conduction band minimum and ~3.0 eV above the valence band maximum. This means that voltages of the order of 1.5 V are still rather far from those needed for tunneling across a reduced barrier.

The high-bias behavior is dominated by the relative energy shift with bias of the Fe Δ_1^σ states. The origin of the reduction of the conductance in the P configuration for $V > 1$ V is that once the $\Delta_1^\uparrow$ band-edge of the left lead enters the bias window, the high transmission region only extends over part of such window (Fig. 7.32), so that the increase of the current with bias is reduced by a factor of about 2. Note that, as discussed above, the current in the $\downarrow$ is negligible. In the AP configuration the increase of the conductance with increasing bias is due to the fact that the broad

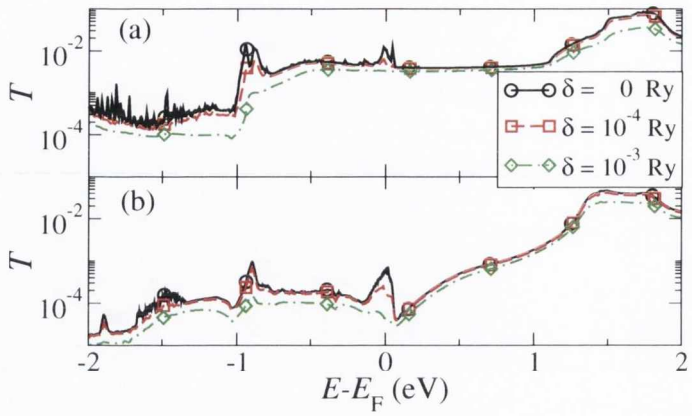


Figure 7.38: 0-bias transmission coefficient for different values of the imaginary part δ added to the energy for the P (a) and the AP (b) configuration.

$\downarrow$ peak in the transmission moves closer to the bias window. Once the $\Delta_1^{\downarrow}$ band-edge moves into the bias window, which is the case at about 1.5 V, there is a drastic increase of the AP current.

It is also interesting to comment on the $S(V)$ plots in Fig. 7.36. For both the magnetic configurations one can observe a peaked structure. This is observed also experimentally [292] and attributed to resonances in the transmission, which could be due to both elastic- and inelastic scattering. In the high-bias region our I-V for the P configuration does not increase as rapidly as that found in typical experiments. One reason for such difference is that the MgO barrier calculated here is very thin, and the current consequently is already large at small bias. In fact our 0-bias transmission coefficient is rather flat around the Fermi energy, for thicker junctions it is expected to rise exponentially for increasing energy, since those energies are closer to the MgO conduction band. Therefore we expect that for larger thicknesses the reduction of the conductance at 1 V is not found. Calculations performed for a 4 MgO ML and for a 8 MgO ML junction for the unrelaxed coordinates confirm this. The I-V for this system is shown in Figs. 7.68 and 7.69, in the case without oxygen vacancy. Another reason for the deviation in the high bias regime can be inelastic effects, and also the presence of defects at the interface and in the MgO. In Sec. 7.6 we investigate the effect of defects in the MgO by calculating the I-Vs for an MgO barrier with an oxygen vacancy (V_O). The most prominent feature in the $S(V)$ curves is the large broad peak at about 1.5 V, found only in the AP configuration, and caused by the broad interface states associated to the $\Delta_1^{\downarrow}$ band edge at the Γ-point. In experiments such a peak in the AP configuration is indeed found at about 1 V, and no such peak is found in the P configuration [292, 293].

We now investigate the effect of generic disorder, motivated by observations [293].

Disorder is modeled at a simple level by adding a small imaginary part δ to the energy when calculating $T(E;V)$, i.e. it corresponds to a uniform level broadening [290]. The effects of such addition on the zero bias transmission are shown in Fig. 7.38, while those on the I-V, $G(V)$, $S(V)$ and TMR are shown in Figs. 7.36 and 7.37. Fig. 7.38 clearly shows that increasing δ results in a gradual suppression of the $T(E;0)$ resonance at E_F, and in general of all the sharp features in the transmission, the remaining features change little with increasing δ. For $\delta = 10^{-4}$ Ry the peak is still clearly visible for P and AP alignment, and only for $\delta = 10^{-3}$ Ry it completely disappears. As a result the TMR at low bias largely increases and the non-monotonic behavior for $V < 0.4$ V is suppressed. This is mainly determined by the reduction in transmission at E_F in the AP configuration. In contrast, the high bias region is barely affected by δ, indicating that these results are not sensitive to such a perturbation. The decay of the TMR with bias is similar to the one obtained in typical experiments for high-quality junctions [264]. The main difference however is that the absolute value is much larger in our calculations. Although our results are indicative of the effects of disorder on the TMR, they should be taken with caution. The broadening δ introduces unstructured disorder, and the transmission of all the spin-channels is equally reduced. In reality one may expect the transmission to either increase or decrease depending on the type of scattering center, which in general will act differently on the different spin-channels. This will in general result in an enhancement of the current in the AP alignment, causing a reduction of the TMR (see also Sec. 7.6). In Ref. [294] this is achieved by adding a random perturbation to the potential. Although this leads to a reduction of the value of the TMR, the mechanisms for the reduction of the TMR with bias should however be the sames as the ones discussed here.

Finally we compare the I-V obtained using the self-consistent solution, with the one obtained using the rigid shift approximation, described at the end of Sec. 7.3. The comparison of the self-consistent potential drop and the one applied in the rigid shift approximation is shown in Fig. 7.25. Since the agreement is rather good, we expect also the I-V to be similar. The resulting I-V curves are shown in Fig. 7.39. Indeed the figure shows that the result of the rigid shift approximation is in very good agreement with the exact self-consistent solution. For these kind of tunneling junctions, where the charging in the barrier is small, we have therefore shown that the rigid shift approximation gives good results.

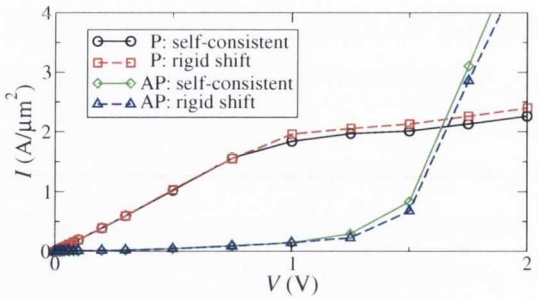


Figure 7.39: Comparison of the P and AP I-Vs obtained with the self-consistent solution, with the one obtained applying the rigid shift potential.

7.4.1 I-V characteristics for unrelaxed coordinates

In order to verify how sensitive the results are to small changes in the atomic structure, we calculate the I-V characteristics for a junction with unrelaxed coordinates, and compare the results to the one for the relaxed coordinates. The bias dependent transmission is shown in Fig. 7.40 for P alignment, and in Fig. 7.41 for AP alignment. The results are very similar to the ones obtained for the relaxed coordinates (Figs. 7.31 and 7.32). There are only some quantitative differences, which we have discussed in Sec. 7.2 for zero bias. The resulting I-V and TMR are shown in Fig. 7.42. Also here the general behavior is similar to the one for the unrelaxed coordinates. There are two main quantitative differences: the reduction in conductance in the P configuration is less pronounced here, and the TMR still shows non-monotonic behavior even for $\delta = 10^{-3}$ Ry. This second difference is due to the fact that the transport through resonances is enhanced for the relaxed coordinates, so that a larger value of δ would be necessary to completely remove their contributions. Generally the comparison shows that a small variation in the structure do not change the main features.

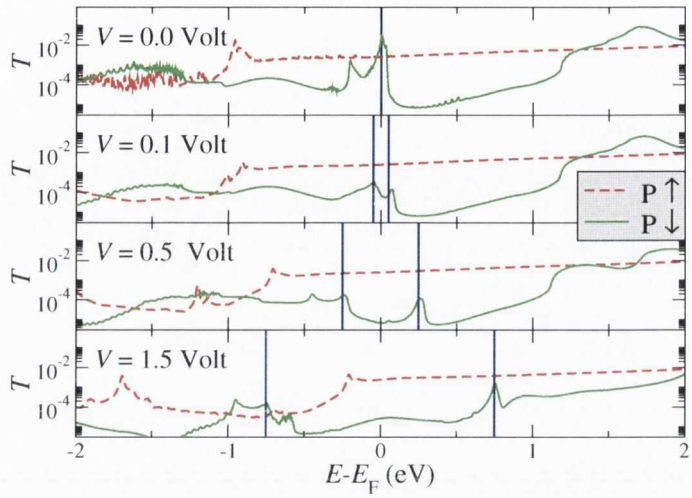


Figure 7.40: Spin dependent transmission coefficient T for the P configuration as a function of energy, E, and for different biases, V. The vertical lines are placed at $E = E_F \pm eV/2$ and enclose the bias window. The transmission is calculated for the unrelaxed coordinates.

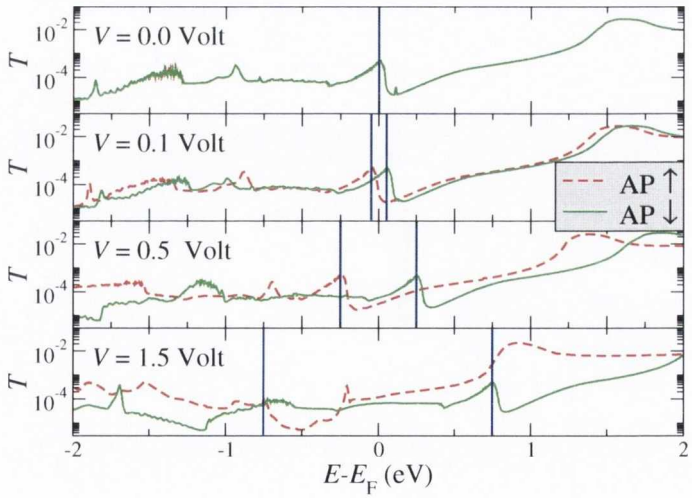


Figure 7.41: Spin dependent transmission coefficient T for the AP configuration as a function of energy, E, and for different biases, V. The vertical lines are placed at $E = E_F \pm eV/2$ and enclose the bias window. The transmission is calculated for the unrelaxed coordinates.

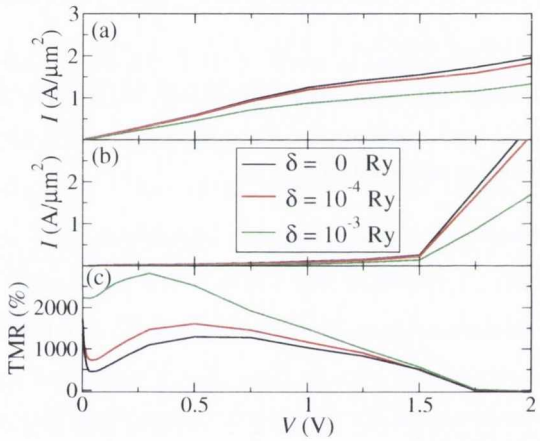


Figure 7.42: Current I for the P (a) and AP (b) configurations, and TMR (c) as function of voltage V, calculated for the unrelaxed coordinates. The results are shown for different values of the imaginary part δ added to the energy.

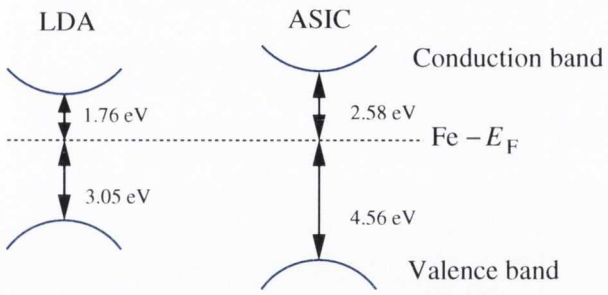


Figure 7.43: Schematic representation of the band-alignment for the relaxed coordinates, and for both LDA and ASIC.

7.4.2 *I-V* characteristics for the ASIC

As described in Sec. 7.2, when the LDA exchange correlation functional is used, the band gap of MgO is underestimated by about 3 eV. The application of the ASIC increases the band gap and brings it very close to the experimental value (Fig. 7.4). One of the major drawbacks of the ASIC however is, that it can usually not correctly describe metals [90], which are usually described well by the LDA. In the Fe/MgO/Fe tunnel junction, we therefore apply the ASIC only on the Mg and O atoms, whereas we use the LDA without the ASIC for the Fe. This is achieved by setting the ASIC correction in the pseudopotential of the Fe atoms to zero. While this is certainly a good approximation far away from the Fe/MgO interface, where both Fe and MgO have bulk-like properties, at the interface it might be less appropriate. However, since also the LDA alone describes MgO rather well, we expect this to be a good approximation. For the calculations in this section we use the relaxed coordinates.

The calculated band alignment between the Fe and MgO is schematically shown in Fig. 7.43. The relative alignment of the bands is similar in both LDA and ASIC. Due to the larger band gap in the ASIC however the absolute distance of both conduction and valence band from E_F increases compared to LDA. We therefore expect a reduction of the transmission around E_F.

In Fig. 7.44 the 0-bias transmission is plotted for both LDA (black dashed lines) and ASIC (red lines), for P majority (a), P minority (b), and AP majority (c). For all the cases the energy range with small transmission increases for ASIC, due to the increased band gap. Generally for all energies in the gap the transmission is also lower than the LDA one. Around E_F the ASIC transmission is lower by approximately a factor three when compared to the LDA one. We note however that the height of all the resonances is almost unchanged when going from LDA to ASIC. As we will show in the next section, where we analyze the complex band structure of MgO, this is due to the fact that in our calculations the transmission through these states is to a large

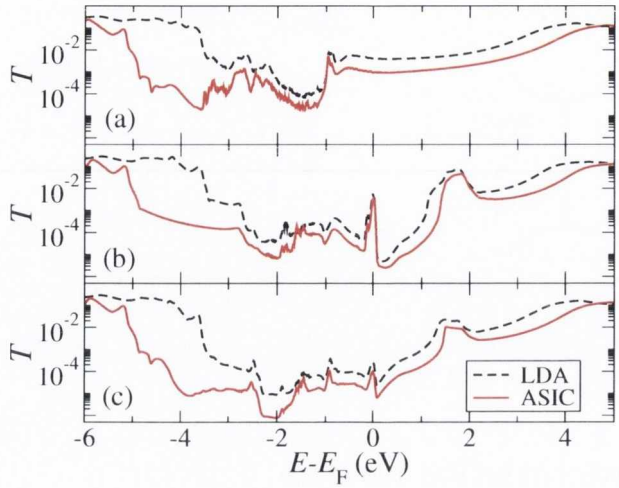


Figure 7.44: Comparison of the zero bias transmission for LDA and ASIC for P ↑ (a), P ↓ (b) and AP ↑ (c).

extent independent of the MgO band gap.

The resulting I-V is shown in Figs. 7.45 (a) and (b) for P and AP alignment respectively, and is similar to the LDA one (Fig. 7.36) for both the P and AP configurations. The main difference is that the ASIC current is about three times smaller than the LDA current at a given bias. Therefore also the TMR [Fig. 7.45 (c)] is of the same magnitude as for the LDA (shown in Fig. 7.37), although slightly reduced. Also here we use a finite value of δ in the calculation of T to gradually reduce the contribution from the ISs. The behavior is again similar to the LDA case. For $\delta = 10^{-3}$ Ry the contribution from ISs is removed, and we find the TMR decreasing monotonically as function of bias voltage [green curve in Fig. 7.45 (c)], with a value at zero bias of about 5000%.

We conclude that qualitatively LDA and ASIC give the same results. The main effect of the ASIC is a reduction of the current, for both P and AP configuration, due to the increased band gap. Since the reduction is of the same order of magnitude for both P and AP alignment, the TMR is almost unchanged.

7.4.3 Comparison with other calculations

We conclude this section by comparing our results with those available in the literature. To our knowledge, the I-V curves for 4 MgO MLs ideal junctions have been calculated non-self-consistently in Refs. [274, 291] within a rigid shift approximation, and self-consistently in Refs. [321, 317] for a 5 MgO MLs junction. Recently the I-Vs for such a junction have also been calculated within a single-band tight-binding model [322] and parameters fitted to the results of Ref. [274]. When comparing the

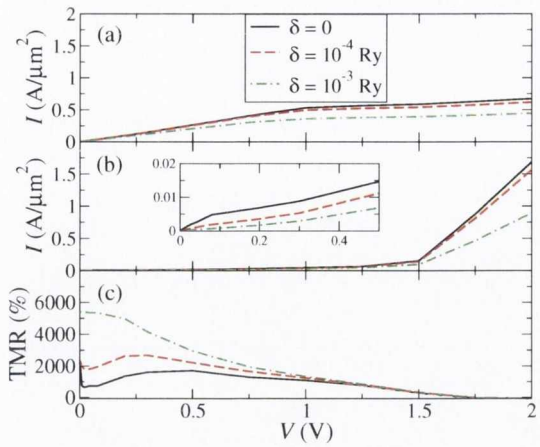


Figure 7.45: Current I for the P (a) and AP (b) configurations, and TMR (c) as function of voltage V, calculated for the relaxed coordinates, using the ASIC. The results are shown for different values of the imaginary part δ added to the energy. The inset in part (b) shows a zoom-in at low bias.

results between this set of calculations, one has to take into account that different interface geometries are used.

Up to about 1 V the results of all the publications are similar, if one takes into account that the contribution of the $\downarrow$ IS close to E_F to the current is very sensitive to the details of the calculation. For example, our low bias results match quite well with those of Ref. [317], for the case where we smear out the IS by adding an imaginary part $\delta = 10^{-3}$ Ry to the energy.

Unfortunately in Refs. [274, 291] no data is given for voltages above 1 V, so that we can not compare the results in the high bias region. For voltages above 1 V our results agree rather well with those of the tight-binding calculation in Ref. [322], but are in contrast with those of the self-consistent calculation of Ref. [317]. The difference is mainly in the I-V for the P configuration, which in Ref. [317] is dominated by the $\downarrow$ spins at high bias, whereas these give only a negligible contribution in our work (see Fig. 7.35). In our analysis of the I-Vs we have discussed in detail why at high bias the $\downarrow$ ISs, associated to the $\Delta_1^{\downarrow}$ band-edge at 1.5 eV above the Fermi energy, can only contribute to the current in the AP configuration, but not in the P configuration (Figs. 7.33 and 7.34). This is a very important point, especially when comparing the results with experiments, where a large peak in the $S(V)$ curves is found at about 1 V only for AP alignment [292, 293]. For the I-V curves calculated in Ref. [317] however a corresponding peak in $S(V)$ should appear also in the P configuration.

The authors of the tight-binding study of Ref. [322] kindly provided us additional, not yet published data, such as the spin-polarized currents for an improved set of parameters, fitted to the results of Ref. [291]. Their spin-polarized I-Vs match

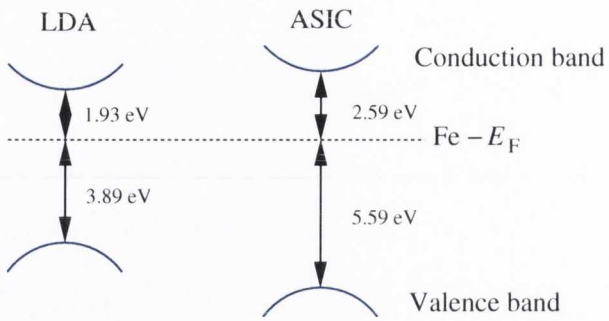


Figure 7.46: Schematic representation of the band alignment for LDA and ASIC (unrelaxed coordinates).

very well with our results and most importantly they do not show significant contribution to the current from the ↓ spins in the P configuration. This confirms our conclusions, that the only possible solution is the one where there is no significant contribution from the ↓ spins at high voltage in the P configuration. At this point we can only speculate on the reasons why in Ref. [317] a fundamentally different result is obtained. The first reason for the difference in the results might be related to the self-consistently obtained potential drop, which depends sensitively on the occupation of the BS (see Fig. 7.17). A second possible reason might be related to the different relaxed coordinates used in the two works, but in our opinion this is less likely, since the zero bias transmission coefficient in Ref. [317] is rather similar to ours.

7.5 Thickness dependence of the transmission

In the previous sections the results for a junction with a barrier thickness of 4 MLs were presented. For MgO tunneling junctions, an exponential increase of the resistance with the thickness of the MgO barrier is found experimentally at large thicknesses [264]. In Refs. [20, 290, 291] the thickness dependence is analyzed theoretically for the Fermi energy. In this section we analyze the dependence of the 0-bias transmission over a broad energy range, of the order of the whole MgO band gap, on the thickness of the junction. We perform calculations for both LDA and ASIC, and use the unrelaxed coordinates. In Fig. 7.46 the band alignment is shown for the unrelaxed coordinates. It is similar to the alignment for the relaxed coordinates (Fig. 7.43), but with the MgO band gap increased by about 1 eV due to the lattice compression (see Sec. 7.1).

If we assume an exponential dependence on the thickness, the spin dependent

k-resolved transmission $T^\sigma(\mathbf{k})$ can be modeled as

$$T^\sigma(\mathbf{k}) = \sum_j t_j^\sigma(\mathbf{k})\ e^{-\alpha_j^\sigma(\mathbf{k}) n_{\mathrm{ML}}}, \tag{7.6}$$

where $\sigma = \uparrow, \downarrow$ is the spin index and n_{ML} is the number of MgO MLs. The sum over j runs over all available states at a given energy, $t_j^\sigma(\mathbf{k})$ is the transmission extrapolated for 0 ML for state j, and $\alpha_j^\sigma(\mathbf{k})$ is its decay coefficient. The total T^σ is then obtained by integrating $T^\sigma(\mathbf{k})$ over the 2D BZ [Eq. (4.108)]. For thin junctions the transmission is mainly determined by the states with the largest coefficient t_j^σ, i.e. by those states that have a high DOS close to the Fe/MgO interface. For thick junctions however only the states with the smallest $\alpha_j^\sigma(\mathbf{k})$ contribute to the transport. In this case we can approximate the total transmission T^σ as

$$T^\sigma = t_0^\sigma\ e^{-\alpha^\sigma n_{\mathrm{ML}}}, \tag{7.7}$$

where the total damping coefficient α^σ is the smallest of the decay coefficients of all the $\alpha_j^\sigma(\mathbf{k})$ over the entire BZ and t_0^σ is the total transmission extrapolated for 0 ML from the high thickness regime. Above equation becomes exact in the limit of infinite thickness. By calculating T^σ for two different thicknesses $n_{\mathrm{ML},1}$ and $n_{\mathrm{ML},2}$ (both of them have to be rather large) we obtain the damping coefficient as

$$\alpha^\sigma = \frac{1}{n_{\mathrm{ML},2} - n_{\mathrm{ML},1}} \ln\left(\frac{T_1^\sigma}{T_2^\sigma}\right), \tag{7.8}$$

where T_1^σ (T_2^σ) is the transmission for $n_{\mathrm{ML},1}$ ($n_{\mathrm{ML},2}$) MgO MLs.

We calculated the 0-bias energy dependent transmission for 4 to 24 MgO MLs, in steps of 4 MLs. The result is shown in Fig. 7.47 (first row) for P majority and minority, and for AP majority. The transmission is similar for all thicknesses outside the MgO band gap (see also Fig. 7.46). Inside the gap it decreases with increasing thickness. We note that the noise in the curves is approximately constant for increasing thickness, which indicates that the accuracy in the results is constant for all the considered thicknesses. When going from 4 to 8 MLs, the relative height of the transmission features in the gap changes, whereas above 8 ML it is approximately constant for all thicknesses. This indicates that for 4 ML we are still in the thin junction regime, whereas for 8 ML we are already in the thick junction regime, where only the states with the smallest decay contribute significantly to the transport.

We note that whereas the transmission decreases rapidly when moving from the band-edge energy towards the middle of the gap, it is almost constant in the middle of the gap, and even increases for the P majority spins. Moreover we note that

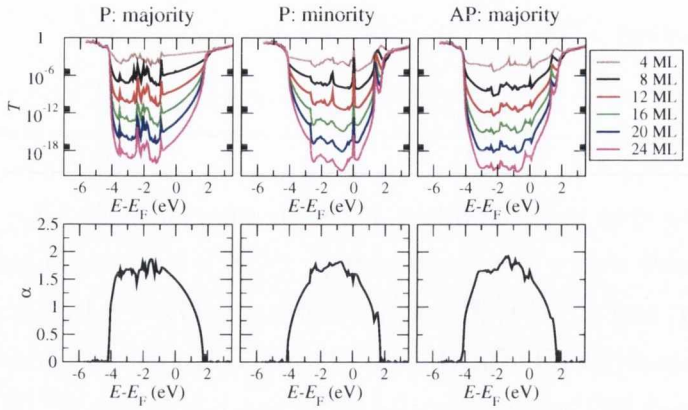


Figure 7.47: Top row of figures: Energy-dependent transmission T for different MgO barrier thicknesses, calculated using the LDA, for P↑, P↓, AP↑; second row of figures: the corresponding damping coefficient α, calculated using $n_{\mathrm{ML},1} = 16$ and $n_{\mathrm{ML},2} = 24$.

the contribution from the ISs in the P minority at E_{F} is visible for all thicknesses. This is in contradiction to the results of Refs. [20, 291], where it is shown that the contribution from such a peak vanishes for large thicknesses. The reason for this difference is that in Refs. [20, 291] it is shown that only **k**-points close to the Γ point contribute to the transmission, whereas in our case, by analyzing the **k**-dependent transmission, we find that even for thick junctions large contributions to the transmission also come from other points in the BZ. We now investigate this difference in the results by calculating the energy-dependent damping coefficient and the complex band structure for MgO.

We calculate α^{σ} with Eq. (7.8), using $n_{\mathrm{ML},1} = 16$ and $n_{\mathrm{ML},2} = 24$. The resulting energy dependent damping is shown in the second row of Fig. 7.47 for P majority, P minority, and AP majority. By using different values for $n_{\mathrm{ML},1}$ and $n_{\mathrm{ML},2}$, the results are analogous, confirming the applicability of the model for the tunneling transmission. α is approximately zero in the energy range outside the MgO gap, and has a semi-circular shape in the gap. The boundaries of the tunneling regime agree rather well with the calculated band alignment (Fig. 7.46). We note that the value of $\alpha(E)$ is similar for P ↑, P ↓ and for AP ↑, especially around E_{F}. This leads to an approximately constant TMR as function of thickness, or at least to a rather slow increase. This agrees with the results of Refs. [291, 294]. If we additionally include a small amount of inelastic processes, and assume that there is always some disorder in the junction, we expect the TMR to be reduced. We can then conclude that for large thicknesses the TMR becomes approximately constant, as also shown in Ref. [294]. This is indeed the situation found in experiment [264].

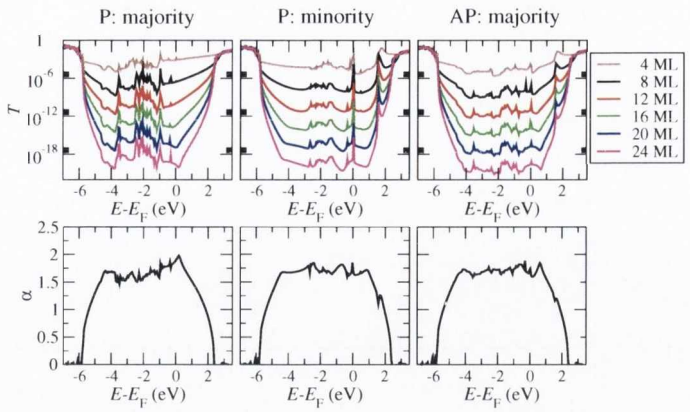


Figure 7.48: Top row of figures: Energy-dependent transmission T for different MgO barrier thicknesses, calculated using the ASIC, for P↑, P↓, AP↑; second row of figures: the corresponding damping coefficient α, calculated using $n_{\mathrm{ML},1} = 16$ and $n_{\mathrm{ML},2} = 24$.

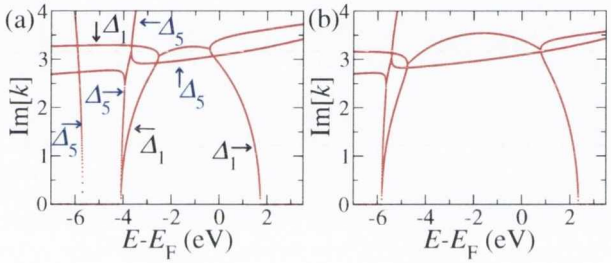


Figure 7.49: MgO complex band structure at the Γ point for LDA (a) and ASIC (b).

From Fig. 2(a) of Ref. [264] the experimental damping coefficient at large thickness and low bias can be extracted to be $\alpha_{\mathrm{exp}} \approx 1.3$, and is identical for both P and AP configuration, leading to an approximately constant TMR for large thicknesses. This is rather similar to our LDA result, which is about 1.5 for all configurations. However, since in the LDA the MgO band gap is underestimated by about 3 eV, we also expect the LDA to underestimate the value for the damping. We therefore calculate the thickness dependent transmission and damping using the ASIC approximation. The result is shown in Fig. 7.48. Apart from an increase of the energy range of the tunneling region, there is no major difference between the LDA and ASIC transmission for all thicknesses. In fact, in the middle of the gap the transmission is even quantitatively similar for LDA and ASIC. The flat transmission region in the middle of the gap is even more pronounced than in LDA. This is clearly reflected by the corresponding damping coefficient, which is semi-circular close to the MgO band-edges, but becomes flat at energies of about 1 eV inside the gap. The damping in this flat region lies in the range of about 1.5 to 1.7, and is similar for P ↑ and ↓, and AP ↑. This value is also close to the LDA result.

In order to find the origin of the flat region in the energy-dependent damping we calculate the **k**-dependent complex band structure (CBS) of MgO. The CBS is generally used to analyze the tunneling process across insulating barriers [323, 20, 324, 219]. This can be obtained solving Eq. (5.6) for an infinite MgO system, stacked along the transport direction z in the same way as in the tunneling junction. The chosen unit cell contains 4 MgO MLs along z. In Fig. 7.49 the so obtained imaginary part of the wave-vector $k = k_z$ is shown as function of energy for both LDA (a) and ASIC (b) at the Γ point ($k_x = k_y = 0$). As a matter of notation we note that in this section the variable k indicates the component along the z direction, k_z, of a general 3D wave-vector. The units of k are chosen in such a way, that $\mathrm{Im}(k)$ corresponds to the decay of the wave function per unit cell (corresponding to the definition in Sec. 5.1.1). For the chosen unit cell it therefore corresponds to the decay across 4 MgO MLs, so that $|\psi_{n_{\mathrm{ML}}+4}| = e^{-k}\,|\psi_{n_{\mathrm{ML}}}|$. The transmission coefficient then is proportional to $e^{-2k\frac{n_{\mathrm{ML}}}{4}} = e^{-\frac{k}{2}n_{\mathrm{ML}}}$, so that the damping coefficient resulting from k is equal to $k/2$. The alignment with respect to E_{F} in Fig. 7.49 is set to match the band alignment in the tunneling junction (Fig. 7.46). To each band in the LDA CBS the corresponding symmetry is added, as defined in Ref. [20]. The CBS partially agrees with the result of Ref. [20], the main difference however is that we find additional bands, which are rather flat as function of energy, and which cross the semi-circular bands. There is one flat band with Δ_1 symmetry and one with Δ_5 symmetry. These hybridize with the bands of the same symmetry from the semi-circular bands. Whereas the semi-circular bands are different for LDA and ASIC, due to the different band gap, the flat bands are very similar. A plot of the CBS for different (k_x, k_y)-points shows that compared to the Γ point the band gap increases, leading to larger imaginary parts of the semi-circular bands. However the rather flat bands change much less for different **k**-points. In the last part of this section we will show that these flat bands are indeed the origin of the flat region that we find in the damping coefficient.

Before comparing the CBS to the damping coefficient in detail, we first investigate the origin of the flat bands. We calculate the CBS for the 1D-chain single orbital tight-binding model of Sec. 6.8. The inverse dispersion relation for such a system is [Eq. (6.96)]

$$\cos k = \frac{E - h}{2\gamma}. \tag{7.9}$$

For $\left|\frac{E-h}{2\gamma}\right| \leq 1$ the value of k is real, whereas for $\left|\frac{E-h}{2\gamma}\right| > 1$ we have $\mathrm{Im}(k) \neq 0$. In Fig. 7.50(a) $\mathrm{Im}(k)$ is plotted as function of energy. The asymptotic behavior for $|E| \to \infty$ is $e^k \approx |E|/2\gamma$, so that $\mathrm{Im}(k)$ increases logarithmically with $|E|$. In this case there is no energy range with a flat inverse dispersion $k(E)$. However, if we include a finite

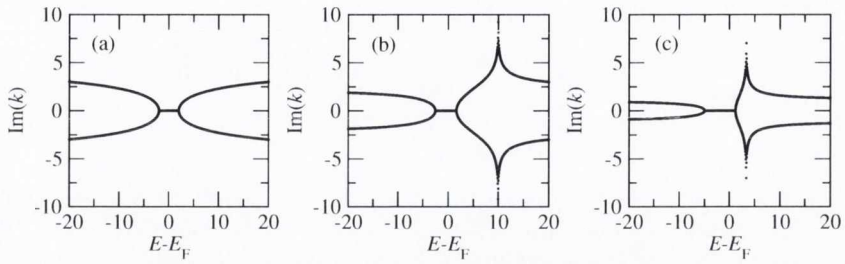


Figure 7.50: Complex band structure for the 1D model (in arbitrary units), for $s = 0$ (a), $s = 0.1$ (b), and $s = 0.3$ (c).

nearest neighbor overlap s, the asymptotic behavior changes drastically. In case of a finite overlap the inverse dispersion relation can be calculated to

$$\cos k = \frac{E - h}{2(\gamma - Es)}. \tag{7.10}$$

This result corresponds to the one obtained in a SMEAGOL calculation, where a non-orthogonal basis set is used. The asymptotic limit for $|E| \to \infty$ now is $\cos(k) \approx -1/2s$, so that k is constant for large $|E|$. If we assume s to be smaller than 0.5, then $\frac{1}{2s} > 1$, so that in the limit $|E| \to \infty$ we have $\mathrm{Im}(k) \neq 0$. Whereas for $s = 0$ $\mathrm{Im}(k)$ increases with increasing energy, for $s \neq 0$ it therefore approaches a constant. This is illustrated in Fig. 7.50, where $\mathrm{Im}(k)$ is plotted for increasing values of s. We also note that at $Es = \gamma$ there is a pole in $k(E)$. This simple model shows that the origin of the flat complex bands lies in the finite overlap between the orbitals.

In order to determine the corresponding asymptotic behavior in a SMEAGOL calculation for MgO, we calculate $\mathrm{Im}(k)$ in a large energy range, of about 100 Ry around E_F. The result is shown in Fig. 7.51(a), and, analogously to the 1D tight-binding model, $\mathrm{Im}(k)$ becomes constant for large $|E|$. We also find poles in $\mathrm{Im}(k)$ for certain energies, the value of $\mathrm{Im}(k)$ at these poles is however limited by using the regularization scheme described in Sec. 5.3. The asymptotic behavior of $\mathrm{Im}(k)$ for large energies can be obtained by taking the limit $|E| \to \infty$ in Eq. (5.3), which then becomes

$$\left(S_0 + S_1 e^{ik_n} + S_{-1} e^{-ik_n}\right) \phi_n = 0, \tag{7.11}$$

where n is the index of state with WF ϕ_n and wave-vector k_n. The resulting ϕ_n and k_n are energy independent. The calculated energy independent values of k_n for the MgO system are shown as red lines in Fig. 7.51(a). It can be seen that for large energies indeed all the $\mathrm{Im}(k)$ (black curves) converge to these asymptotic values (red curves). A zoom at around E_F shows that the lowest lying asymptotic values for $\mathrm{Im}(k)$ are at the origin of the flat bands crossing through the semi-circle around E_F.

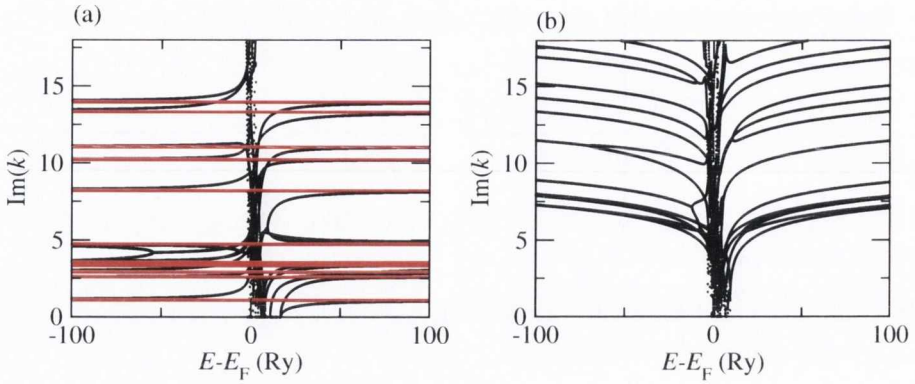


Figure 7.51: (a) MgO complex band structure calculated using the LDA, for a large energy range (black curves). The red curves correspond to the limit calculated using Eq. (7.11). (b) MgO complex band structure calculated setting $S_1 = S_{-1} = 0$.

To confirm that the flat bands are caused by the finite overlap, we calculate the asymptotic solutions for $S_1 = S_{-1} = 0$, which for positive $\mathrm{Im}(k_n)$ are obtained by solving

$$H_{-1}e^{-ik_n}\phi_n = ES_0\phi_n, \tag{7.12}$$

so that $\mathrm{Im}(k_n) \propto \ln(|E|)$ increases with increasing $|E|$. We calculated $\mathrm{Im}(k_n)$ for the MgO system, where we have set $S_1 = S_{-1} = 0$, and obtained the dispersion of Fig. 7.51(b). Instead of reaching constant values, now all the $\mathrm{Im}(k_n)$ increase with increasing $|E|$. A plot of $e^{\mathrm{Im}(k_n)}$ as function of E shows straight lines, indicating a logarithmic dependency. We have therefore confirmed that the origin of the flat bands lies in the finite overlap.

By reducing the overlap between orbitals, we can therefore reduce the absolute value of the elements of S_1 and S_{-1}, so that we expect to shift the flat regions to higher $\mathrm{Im}(k)$. In order to verify this, we have calculated the CBS at the Γ point for 3 different basis sets (in each case we use the same basis for Mg and O): the first is equal to the one used in all our calculations so far [double ζ (DZ), with a cutoff radius $r_c = 6$ bohr], the second basis set is DZ ($r_c = 4$ bohr), and the third is a single ζ (SZ) ($r_c = 4$ bohr). The CBS for the different basis sets, and for both LDA and ASIC, is shown in Fig. 7.52. It can be seen that indeed by reducing the length of r_c from 6 bohr to 4 bohr, the flat bands shift to higher $\mathrm{Im}(k)$. When further moving to a SZ basis, the number of bands is reduced, so that the flat regions move to even higher $\mathrm{Im}(k)$. For the DZ ($r_c = 6$ bohr) basis the flat bands cut through the lowest lying semi-circular state around E_F for both LDA and ASIC, for the DZ ($r_c = 4$ bohr) this is the case only in the ASIC. For the SZ ($r_c = 4$ bohr) the flat bands never cross the low lying semi-circular state. We note that for other **k**-points in the 2D BZ, the flat

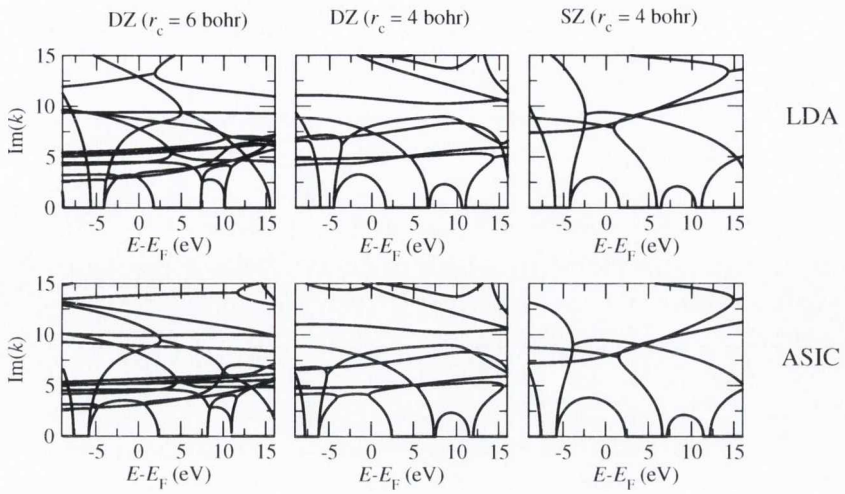


Figure 7.52: MgO complex band structure at the Γ-point, for three different basis sets, and for LDA (top row) and ASIC (bottom row). The figures in the first column are calculated with the DZ ($r_c = 6$ bohr) basis, the ones in the second column with the DZ ($r_c = 4$ bohr) basis, and the ones in the third column with the SZ ($r_c = 4$ bohr) basis.

bands reach to somewhat lower Im(k) also for the SZ ($r_c = 4$ bohr) basis.

We now analyze how these changes in the CBS for different basis sets affect the transport properties for thin and for thick junctions. Since the damping coefficient α is expected to follow the lowest lying Im(k) for each energy, it should vary drastically for the different basis sets. We calculated the thickness dependence of the spin dependent total transmission for the P configuration, and for three different basis sets. In Fig. 7.53 the LDA $T^{\uparrow}$ and $T^{\downarrow}$ are plotted for 4, 16 and 24 ML, together with the corresponding α, obtained comparing the T^{σ} for the 16 ML and 24 ML junctions. The transmission for the DZ ($r_c = 4$ bohr) and the SZ ($r_c = 4$ bohr) bases are similar, and differ strongly from the DZ ($r_c = 6$ bohr) basis for energies lying around the center of the MgO energy-gap. Close to the band-edges all three basis sets give similar results. We note that whereas the P minority resonance peak is visible for all thicknesses for the DZ ($r_c = 6$ bohr) basis, it is filtered out for thick junctions for the other two basis sets. The results for the 4 MgO MLs junction differ less over the basis sets, and the P minority peak is present for all of them. A calculation on a finer energy mesh reveals that its height changes by a factor of not more than about 3 over the different basis sets.

In the last row of Fig. 7.53 the damping α is plotted for both majority and minority, together with the CBS at the Γ point. The values of Im(k) are converted to a corresponding damping coefficient, by scaling them by a factor of 2. In the same way as the transmission coefficient, the damping coefficients for the different bases

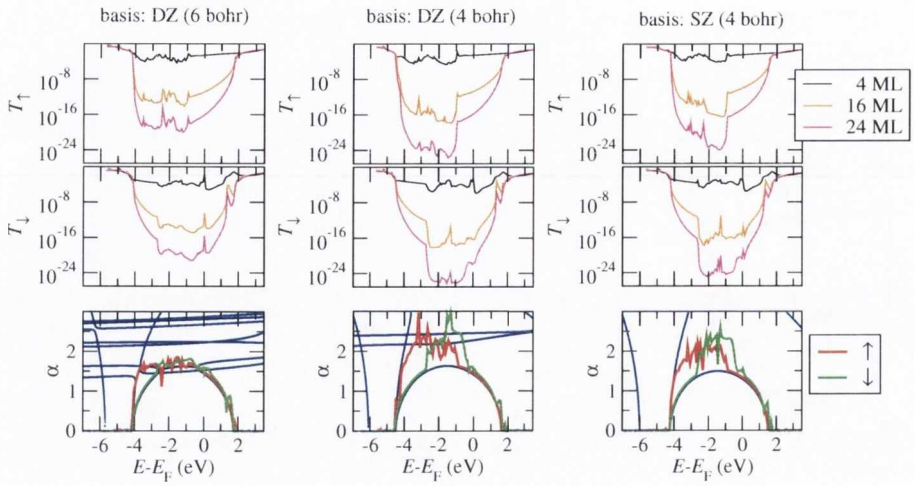


Figure 7.53: Energy-dependent transmission $T^{\uparrow}$ for P↑ and $T^{\downarrow}$ for P↓, for different MgO barrier thicknesses, calculated using the LDA. The results in the first column of figures are calculated using the DZ ($r_c = 6$ bohr) basis, the ones in the second column with the DZ ($r_c = 4$ bohr) basis, and the ones in the third column with the SZ ($r_c = 4$ bohr) basis. In the bottom row of figures the corresponding damping α is shown for P↑ (red) and P↓ (green), calculated using $n_{\mathrm{ML},1} = 16$ and $n_{\mathrm{ML},2} = 24$. The blue lines in the plot of α show the imaginary part of the complex wave-vector, scaled in such a way, that their value corresponds to the predicted damping coefficient.

are all similar for energies close to the band-edges. For energies around the middle of the gap α is similar for the DZ ($r_c = 4$ bohr) and the SZ ($r_c = 4$ bohr) bases, for which it is much larger than the value obtained for the DZ ($r_c = 6$ bohr) basis. The comparison of the transmission with the CBS shows very good agreement for all basis sets. As long as there are states in the Fe electrodes with a Δ_1 symmetry for **k**-points around the Γ point, α follows the lowest Δ_1 band. For energies, where such states are not present in the Fe leads, α rapidly changes to the next smallest complex band. For the P majority spins this transition occurs at about 1 eV below E_F. The discontinuity to the next bands at this energy is visible for the DZ ($r_c = 4$ bohr) basis. For the SZ ($r_c = 4$ bohr) basis a similar discontinuity is found, but there is no band matching α. The reason is that the bands at those values of α do not appear at the Γ point, but at different **k**-points. For the DZ ($r_c = 6$ bohr) basis no such jump of α is visible, which is caused by the fact that the flat bands cut across the top of the semi-circle in the CBS.

The analogous results for the ASIC are shown in Fig. 7.54. The general trend across the different basis sets is the same as for LDA, and there is again a good agreement between CBS and damping. Since for the ASIC however also for the DZ ($r_c = 4$ bohr) basis the flat bands are close to the maximum of the semi-circle, there

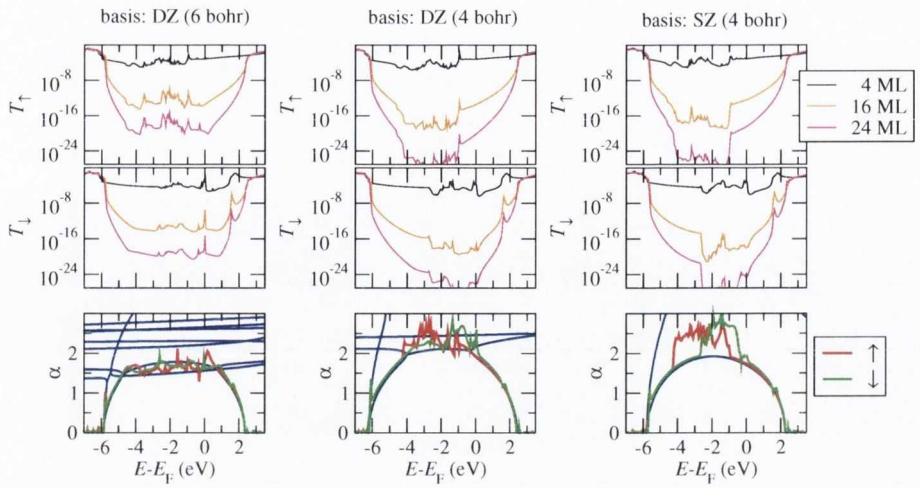


Figure 7.54: The quantities represented are the same as in Fig. 7.53, but calculated with the ASIC.

is a notable difference in α in the middle of the gap compared to the SZ ($r_c = 4$ bohr) basis. The results show that the behavior of the damping coefficient α shown in Figs. 7.47 and 7.48 can be understood in terms of the complex band structure. We find that the flat bands in the CBS contribute to the transmission, and therefore determine the damping at those energies where they have the smallest $\mathrm{Im}(k)$. In Fig. 7.55 we show the **k**-dependent transmission at E_F, obtained using the DZ ($r_c = 6$ bohr) basis, for P ↑ (first row), P ↓ (second row), and AP ↑, evaluated for 4, 16 and 24 MgO MLs. The color code is chosen in such a way, that for each graph the red color corresponds to the maximum transmission. Therefore the red spots indicate in which parts of the BZ the transmission is larger. The blue color is chosen to be 10^{-10} times smaller than this maximum value. In Fig. 7.56 the analogous **k**-dependent transmission is shown for the SZ ($r_c = 4$ bohr) basis. The main difference is that for the DZ ($r_c = 6$ bohr) basis even at large thickness the P ↓ and AP ↑ transmissions have large contributions for **k**-points far away from the Γ-point. In contrast for the SZ ($r_c = 4$ bohr) basis the main contributions to the transmission come only for **k**-points close to the Γ point. The results for the SZ ($r_c = 4$ bohr) basis are in good agreement with those of Ref. [291].

As last comparison between the different basis sets we perform a calculation for a junction with a thickness of 16 and 24 MgO MLs, but where we replace the MgO by vacuum. We do however leave the Mg and O basis orbitals in the vacuum in order to have a finite transmission between the two Fe electrodes. The result for the energy dependent transmission and damping is shown in Fig. 7.57. The transmission is very small for energies below E_F, however for energies above about 4.4 eV the transmission

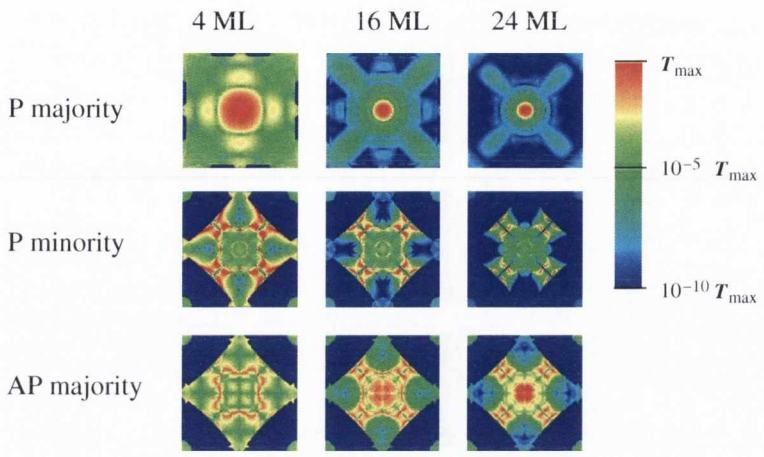


Figure 7.55: **k**-resolved transmission coefficient for P↑, P ↓, and AP ↑, for different MgO thicknesses, using the LDA and the DZ (r_c = 6 bohr) basis. In each figure a different logarithmic color scale is used, where the red color corresponds to the maximum transmission of each figure T_{max}, and the blue color corresponds to 10^{-10} T_{max}.

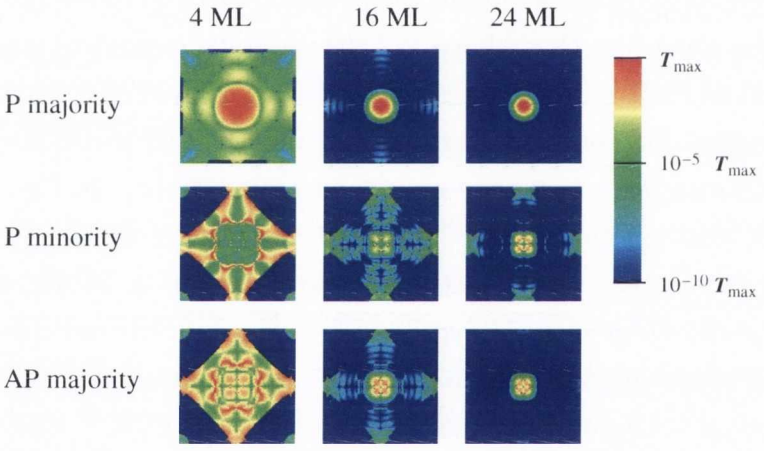


Figure 7.56: The quantities represented are the same as in Fig. 7.55, but calculated using the SZ (r_c = 4 bohr) basis.

is of the order of 1. Using the total potential output from SMEAGOL we calculate the energy of the vacuum to be 4.2 eV (indicated in Fig. 7.57 by the vertical dashed line). For energies higher than the one of the vacuum there is no barrier between the electrodes, and the transmission is therefore high. For such a system we expect the transmission to become smaller as the energy is reduced, for energies below the vacuum level. However for both basis sets we find that the transmission reaches a constant value for very low energies. This is reflected in the damping coefficient, which also reaches a constant value as the energy is lowered. The difference between the two basis sets is that this constant value is already reached at about 3 eV for the DZ (r_c = 6 bohr) basis, whereas it is only reached at about 0 eV for the SZ (r_c = 4

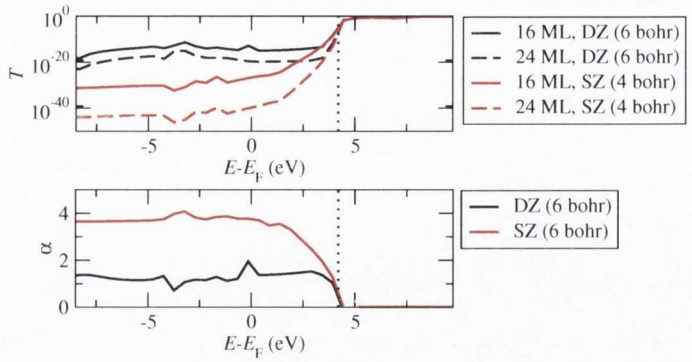


Figure 7.57: P↑ transmission coefficient T for a 16 ML (full lines) and a 24 ML junction (dashed lines), together with the resulting damping coefficient α, for a junction where the MgO has been replaced by vacuum, and where the basis orbitals on the MgO atoms are still included. Two different basis sets are compared, a DZ (r_c =6 bohr) basis (black color), and a SZ (r_c =4 bohr) (red color). The vertical dotted line indicates the value of the vacuum potential energy.

bohr) basis. For very small energies therefore both basis sets produce an unphysical saturation for the damping coefficient. However the energy region, where the damping is correctly described, is larger for the SZ (r_c = 4 bohr) basis. The flat bands in the MgO band structure are the ones responsible for this unphysical saturation of the damping coefficient. We therefore conclude that, in order to describe correctly the tunneling for thick junctions, the basis set has to be chosen in such a way, that no flat bands cross (or come too close) to the semi-circular complex part of the band structure of the band gap. For our calculations with the MgO tunneling junctions therefore we conclude that we can only obtain the correct thickness dependence for both LDA and ASIC with the SZ (r_c = 4 bohr) basis. We note that this is in contrast to the usual view that the quality of the results, obtained using a localized basis set, improves as the basis set gets richer. This can become problematic, since in order to calculate an accurate band-structure for many types of insulators a rather rich basis set has to be used, however the use of such a rich basis set will probably lead to an incorrect description of the damping coefficient at large thicknesses.

In Figs. 7.58 and 7.59 the thickness dependent transmission and damping coefficient are shown for LDA and ASIC respectively, calculated with the SZ (r_c = 4 bohr) basis. The results for the damping coefficient can be expected to be accurate as long as it follows the Δ_1 semi-circular band. Importantly, this is the case for energies around E_F for all configurations. The value in the regions of very high damping might be affected by the flat complex bands also with this basis set.

Around E_F the P ↑ damping coefficient is only slightly smaller than the one for AP ↑. The TMR will therefore increase with thickness, however rather slowly. The

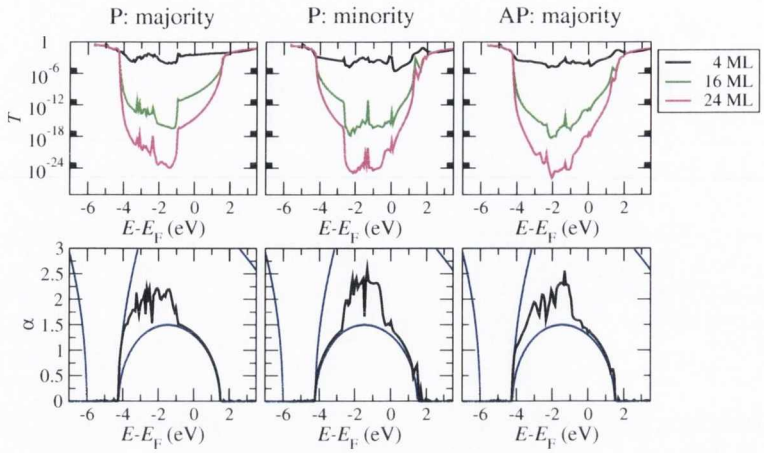


Figure 7.58: In the top row of figures the energy-dependent transmission T for P↑, P↓, and for AP↑ is plotted for different MgO barrier thicknesses. All the results are calculated using the LDA and the SZ (r_c = 4 bohr) basis. In the bottom row of figures the corresponding damping α is shown (black lines), calculated using $n_{ML,1} = 16$ and $n_{ML,2} = 24$. The blue lines in the plot of α show the imaginary part of the complex wave-vector, scaled in such a way, that their value corresponds to the predicted damping coefficient.

main spin-filter effect is already achieved after a few layers, of about 4-12, and is determined by the ratio of the Δ_1 DOS for ↑ and ↓ for **k**-points close to the Γ-point. This contrasts the prediction of Ref. [20], but agrees well with the results of Ref. [291]. In Ref. [20] a different damping coefficient is found. This however is determined by the fact that the transmission in the antiparallel configuration is still dominated by the ISs, which have a different damping. At large thicknesses, only the states close to the Γ point contribute, and therefore the thickness dependence is similar for both majority and minority. The TMR as function of thickness is therefore expected to be constant at large thicknesses, especially if the surface state moves away from E_F.

In Fig. 7.60 finally we show the **k**-dependent transmission along the diagonal of the BZ, so that $k_x = k_y$, for P↑ and P↓. It can be clearly seen that whereas for the 4 MgO MLs junction the main contributions to the P ↓ come from **k**-points over all the range, for the thicker junctions the main contributions are restricted to a small range close to the Γ-point.

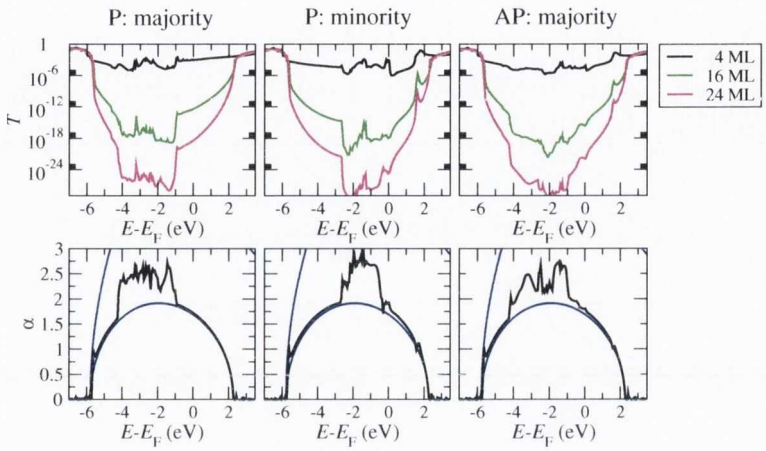


Figure 7.59: The quantities represented are the same as in Fig. 7.58, but calculated with the ASIC.

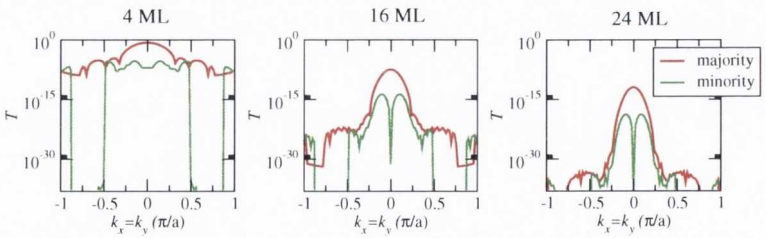


Figure 7.60: LDA spin-polarized transmission coefficient for P↑ (red) and P↓ (green), and for different MgO barrier thicknesses, as function of k_x, along the line in the BZ where $k_x = k_y$. The results are calculated using the SZ ($r_c = 4$ bohr) basis.

7.6 Oxygen vacancies in the MgO barrier

In the calculations presented so far the theoretically predicted TMR is higher than the one measured experimentally, which is never larger than a few hundred percent [264, 265]. This is consistent with other calculations [20, 290, 274, 269]. It has been shown that oxidation of the interface Fe layers can lead to a drastic reduction of the TMR, which can even become negative for asymmetric oxidation of the electrodes [273, 274, 291, 275]. In Ref. [264] it is noted that lattice dislocations are found at the Fe/MgO interface, which can lead to a reduction of the TMR. Another possibility for a reduced value of the TMR is the inclusion of defects in the MgO, and at the Fe/MgO interface. Calculations for a Fe/vacuum/Fe junctions indicate that disorder at the interface can drastically reduce the TMR [295], in agreement with the results of Ref. [275] for a disordered and randomly oxidized Fe/MgO/Fe junction. Experimental results indicate that the density of defects in the MgO depends on the growth conditions [285, 288, 325]. Such a large defect density is found to lead to an effectively reduced MgO band gap [288, 325]. In Ref. [288] it is shown that by annealing the sample the band gap opens to the bulk MgO value, indicating that the density of defects is reduced. Measurements of isolated defects indicate a defect level close to the valence band, which is tentatively attributed to Mg vacancies (V_{Mg}), and a set of levels between the E_F and the conduction band, attributed to oxygen vacancies (V_O). The authors note however that this correspondence is not completely established at this stage. In Ref. [325] a detailed study of the possible defects in MgO, grown on an Ag substrate are presented. They find different possible defects, with energies spread over large part of the MgO band gap. One of the defects they find is V_O, whose energy lies approximately in the middle of the MgO band gap. This agrees well with other theoretical predictions [326, 327, 328]. *Ab initio* calculations with Fe/MgO/Fe junctions in Refs. [329, 299] show that for V_O LDA predicts a defect band centered about 1 eV below the Fe Fermi energy. In Ref. [298] experimental evidence is shown on the decrease of the TMR with V_O, supported by a theoretical model.

In order to investigate the bias dependent influence of defects on the transport, we perform calculations with a V_O in the MgO. As we will show, such a defect lies very close to the Fe Fermi energy, in good agreement with the results of previous calculations [329, 299], and is therefore indicative of all defects that lead to a small additional DOS in the vicinity of E_F. The exact characterization of defects in MgO is a complex task, which should include a relaxation of the lattice for all defects. Moreover, not all defects can be described accurately with DFT. However, previous calculations show that DFT can accurately predict the properties of V_O's [326, 327, 328].

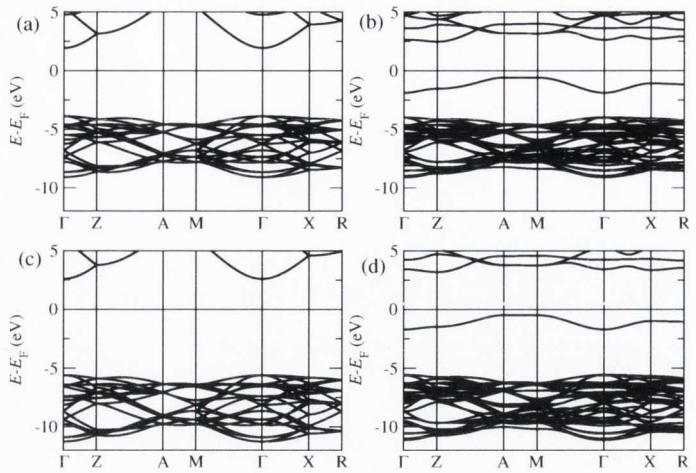


Figure 7.61: LDA MgO band structure without (a) and with V_O (b). Panels (c) and (d) show the corresponding band-structures for ASIC.

In order to keep the size of the calculations tractable, we use a rather high V_O density. We construct a 2×2 supercell in the plane perpendicular to the transport direction. The V_O is then obtained by removing one O atom in one of the MgO MLs. The planar V_O density is therefore 1/4, the total defect density for a 4 MLs junction is 1/16, for a 8 MLs it is 1/32. In all the calculations of this section we do not relax the structure around the defect, and for the Fe/MgO junction we use the unrelaxed coordinates (Ref. [20]). We note that in our calculations the V_O is in a charge neutral state, whereas experimentally the vacancy can exist in different charging states. In Fig. 7.61 the LDA band structure for a defect-free MgO (a) is compared to the one obtained for a MgO with single V_O in a 4 ML unit cell (b). The V_O leads to a defect band lying approximately in the middle of the gap. Importantly, this defect level is not spin-split. The band shows a rather large dispersion, which is due to the high in-plane defect density. This high density also leads to a change of the band-structure at the bottom of the conduction band. Calculations performed for larger cells show that the shape of the bottom of the conduction band is restored, and that the dispersion of the defect band is reduced. In Figs. (c) and (d) the analogous band structures are shown for ASIC. The main difference is that the ASIC increases the band gap. The defect level is still approximately located in the middle of the gap. These results are in good agreement with those of other calculations [327, 328, 299, 329], for both LDA and ASIC. The calculations in Ref. [326] on the other hand predict that the defect levels are much closer to the conduction band. In fact the assignment of the levels close to the conduction band to V_O in Ref. [288] is based on those calculations. However this assignment is probably incorrect, since it is in contradiction to most of the theoretical predictions [327, 328, 299] and to the experimental findings of Ref.

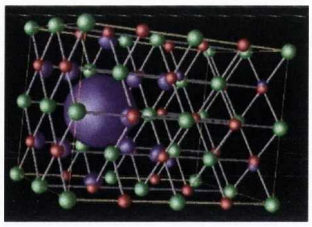


Figure 7.62: Isosurface of the LDOS for the V_O defect band (purple color); red spheres represent O atoms, and green spheres represent Mg atoms.

[285].

In Fig. 7.62 the charge density associated to the defect level is shown (LDOS, see Sec. 7.2). It is located around the V_O site, and has an s-type symmetry. Consequently it is also of Δ_1 symmetry, and can therefore couple well to the high transmission Δ_1 states.

In Fig. 7.61 the bands are aligned with respect to the Fe E_F in correspondence to the calculated band alignment in the Fe/MgO junction (shown schematically in Fig. 7.46). In our calculations the levels associated to V_O's are always below the Fermi energy, and therefore in a charge neutral state. Although the energy difference of the defect band to the conduction band increases in ASIC as compared to LDA, the alignment with respect to E_F is rather similar. Also the dispersion of the defect band is almost the same for LDA and ASIC. Therefore we expect the transport results for such a defect in a tunneling junction to be similar for both LDA and ASIC. We therefore perform all the calculations with the LDA. We performed a test calculation to verify some of the key results, such as the zero-bias transmission coefficient. We always found very similar results for both LDA- and ASIC transmission coefficient in the energy range of a few eV around E_F. There are just minor quantitative differences, of the same type as the ones discussed in Sec. 7.4.2.

We have calculated the transmission properties of a 4ML junction, with a V_O in the 2^{nd} MgO ML from the Fe interface. We also performed calculations for two defective 8 ML junctions, one where the V_O is located in the 2^{nd} ML, and one where it is located in the 4^{th} ML from the Fe/MgO interface. We note that for the transport calculations we use a the DZ ($r_c = 4$ bohr) basis for the MgO, this gives a good LDA damping coefficient, and can also be expected to capture the change in the charge density close to the vacancy. In fact the band-structures shown in Fig. 7.61 are only very weakly dependent on the basis. In Fig. 7.63 the spin-polarized transmission coefficients for P and AP alignment of the Fe electrodes are shown for an ideal 4 ML junction, and for the junction with the V_O in the 2^{nd} ML. In the energy range from about -2 eV to E_F the transmission is enhanced in the defective junction. This range

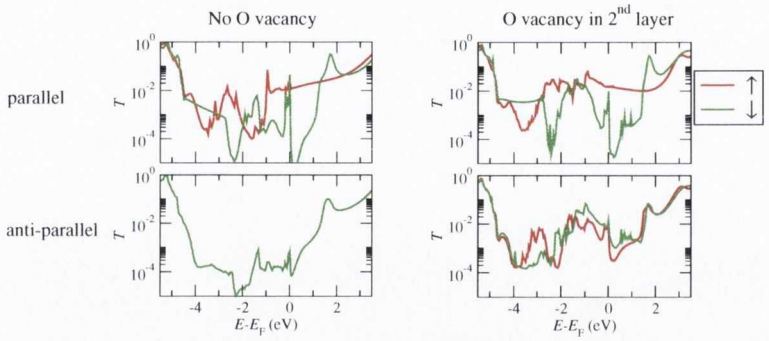


Figure 7.63: Transmission coefficient T for P (top row of figures) and AP alignment (bottom row of figures) of the Fe electrodes, for ↑ (red) and ↓ (green). The results are for a 4 MgO MLs junction, the left column of figures shows the results for an ideal junction, whereas the right column is for a junction with a V_O in the 2nd layer from the interface.

agrees rather well with the band-width of the defect band, confirming that in the Fe/MgO junction the top of the defect level is close to E_F. In the vicinity of E_F the transmission for the P configuration is similar with and without vacancy. At positive energies the transmission with the vacancy is slightly reduced, which is due to the changes in the conduction band, caused by the high density of defects (see Fig. 7.61). In the AP configuration however the transmission around E_F is strongly enhanced for the junction with the vacancy, when compared to the ideal junction. It is in fact of a similar size as the one for the P configuration, indicating a vanishing TMR for such a high defect density.

In Fig. 7.64 the spin-polarized transmission is shown for the 8 ML junction, for the defect-free MgO, for the V_O in the 2nd ML, and for the V_O in the 4th ML. For the P configuration the transmission around E_F is almost unchanged for the three different systems. For the AP configuration however the transmission increases in presence of defects. We note that as the defect moves closer to the interface, the increase in the AP transmission around E_F is enhanced. In fact, for the V_O in the 2nd ML, below E_F the transmission is of a similar size as for the P configuration, analogously to the 4 ML case. In Fig. 7.65 the total transmission for the three 8 ML junctions is shown in a small energy range around E_F, in order to emphasize the change in the transmission for the different position of the vacancy. For junctions, that are highly defective close to the interface, we therefore predict a large reduction of the TMR. The experimentally found vanishing TMR for very thin junctions [264] might therefore be attributable to a large density of defects for such very small thicknesses. Even at larger thicknesses the TMR is limited by the amount of defects close to the interface.

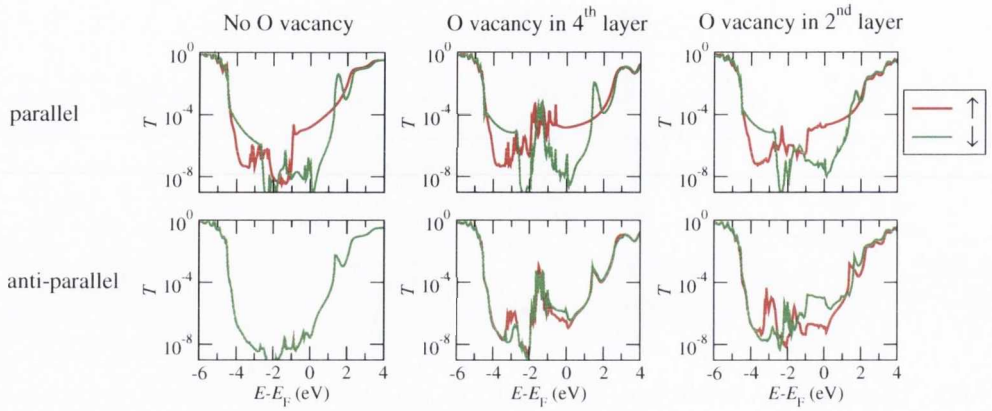


Figure 7.64: Transmission coefficient T for P (top row of figures) and AP alignment (bottom row of figures) of the Fe electrodes, for ↑ (red) and ↓ (green). The results are for a 8 MgO MLs junction with no V_O (leftmost panels), with a V_O in the 4th layer from the interface (middle panels), and with a V_O in the 2nd layer from the interface (rightmost panels).

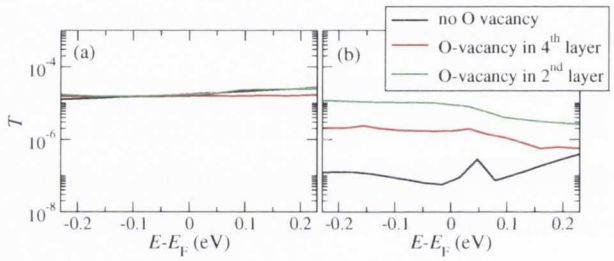


Figure 7.65: Total transmission coefficient T for P (a) and AP alignment (b) of the Fe electrodes. The results are for a 8 MgO MLs junction with no V_O (black curves), with a V_O in the 4th layer from the interface (red curves), and with a V_O in the 2nd layer from the interface (green curves).

We now analyze the origin of the enhanced transmission for the defective junctions in the AP configuration. First we note that the enhancement of the current is much stronger in the minority spin than in the majority spin (Fig. 7.64). The origin of this can be understood from the analysis of the **k**-dependent transmission coefficient, which is plotted in Fig. 7.66 at an energy of -0.2 eV below E_F. This is lower than the lowest energy of the surface state, and lies in the region of high transmission for the AP configuration in the defective junctions. The **k**-dependent transmission is plotted for the ideal junction, and for the 8 ML junction with the V_O in the 2nd ML, for both P and AP configuration. In all the graphs the transmission is dominated by a small area of **k**-points close to the Γ point. For the ideal junction the transmission is highest for the P ↑ states, since in that case there is a large density of the high-transmission Δ_1 like states on both sides of the junction. For the P ↓ the contribution of the Δ_1 like states on both sides of the junction is rather small, and therefore the transmission

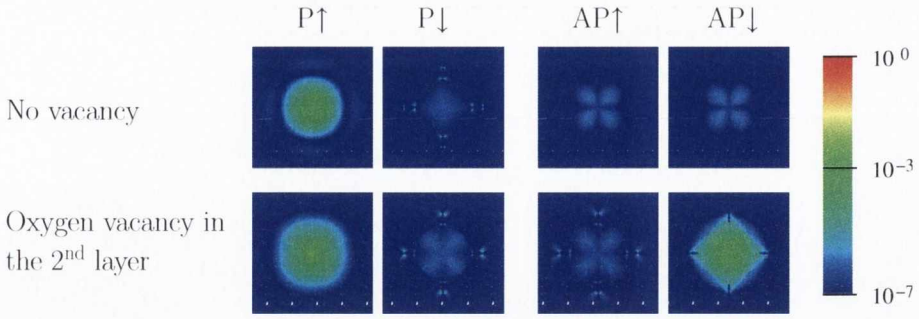


Figure 7.66: **k**-dependent transmission for P↑ (first column), P↓ (second column), AP↑ (third column), and AP↓ (fourth column), at an energy $E - E_F = -0.2$ eV. The first row of figures is for an ideal 8 MgO MLs junction, the second row is for a junction with a V_O in the 2nd MgO layer from the interface.

is reduced when compared to the P ↑. In the AP configuration the transmission is identical for ↑ and ↓, since the ideal junction is completely symmetric. Since on one side there is a high density of Δ_1 states, whereas on the other side there is a low density, the transmission is much smaller than for P ↑, but somewhat larger than for P ↓.

For the junction with the V_O in the 2nd ML, the situation is very different. Since the vacancy is not spin-polarized, the electrons flow through the vacancy states at this energy has approximately the same density of Δ_1 states for both ↑ and ↓. On the left side of the junction, to which the V_O is very close, therefore there is a large Δ_1 DOS for both ↑ and ↓. On the right side, and for P configuration, the ↑ has a much larger contribution from the Δ_1 states than the ↓, so that the P ↑ transmission is similar to the one of the ideal junction, whereas the P ↓ transmission is similar to the one of the ideal junction for AP alignment. For AP alignment of the defective junction the situation is reversed, and the larger transmission comes from the ↓, and is similar in size to the P ↑ transmission. The AP ↑ transmission on the other hand is similar in size to the AP transmission for the ideal junction. We therefore conclude that the enhancement in transmission in the AP configuration in a defective junction is caused by the depolarization of the Δ_1 states at the vacancy site. If this is very close to the Fe/MgO interface, it effectively leads to a depolarization at the interface. If the vacancy lies in the middle of the junction, the effect is less pronounced, since the states need to tunnel to the vacancy site from both interfaces, where the Δ_1 DOS is small. However the TMR is reduced also in this case. Our *ab initio* result agrees qualitatively with the conclusions of Ref. [298].

We now investigate the finite bias behavior of the defective junctions. In Fig. 7.67 the self-consistent potential drop ΔV_H and charging Δ_ρ are shown for different bias

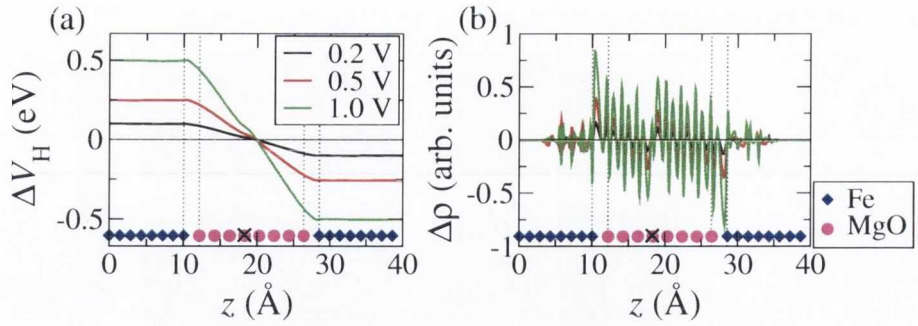


Figure 7.67: (a) Planar average ΔV_H of the difference between the Hartree potential at different bias voltages and the one at 0-bias. (b) Planar average $\Delta\rho$ of the difference between the charge density at 0.5 V and the one at 0-bias for the same bias voltages. The diamonds and dots indicate the location of the Fe and MgO layers. The results are for a 8 MgO MLs junction with a V_O in the 4th layer from the interface (indicated by a dot with a '×').

voltages for the 8 ML junction, with a V_O in the 4th ML, and in the P configuration. The potential drops almost linearly across the 8 MgO layers. At the vacancy site there is a slight deviation from the linearity, however the effect is only small. There is a small dipole forming at the vacancy layer, but the total charge at the vacancy site is constant with bias. The energy levels associated with it can therefore easily enter the bias window. In Sec. 7.4 we have shown that if the potential drops approximately linearly across the insulator, it is possible to apply the rigid shift potential approximation, instead of calculating the fully self-consistent solution at finite bias. We therefore calculated the I-V characteristics using this approximation, in order to reduce the computational effort. For the bias steps shown in Fig. 7.67 we compared the current obtained from the self-consistent calculation with the one obtained applying the rigid shift potential, and the results were almost identical, confirming the applicability of the approximation also for the defective junction.

In Fig. 7.68 the I-V curves and the resulting TMR are shown for the two 4 ML junctions, and in Fig. 7.69 they are shown for the three 8 ML junctions. As expected from the 0-bias results, the main effect of the vacancies is to enhance the AP current. The closer the vacancy is to the Fe/MgO interface, the stronger the enhancement. The V_O's do not lead to a significant change of the P current, although there are some quantitative differences. It can be seen that the TMR is drastically reduced in the junctions containing the V_O's. It basically vanishes if the vacancy is in the 2nd ML. For the defect-free junction we note that the reduction in conductance at about 1 V is visible for the 4 ML junction, but not for the 8 ML junction. This confirms our discussion in Sec. 7.4, where for thicker junctions we predicted a reduction of this change in conductance.

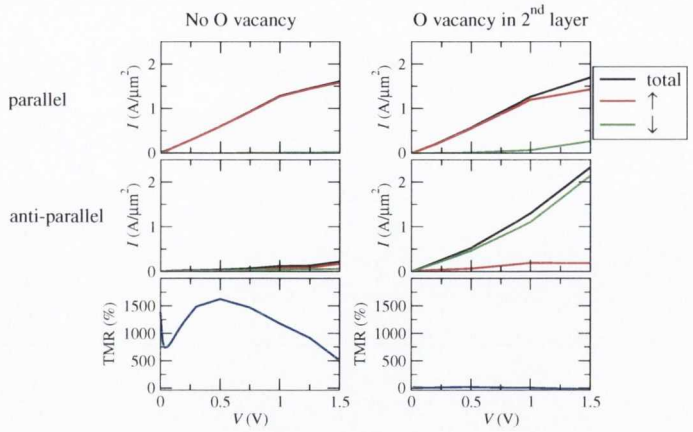


Figure 7.68: Spin-polarized current for P (top row of figures) and for for AP alignment (middle row of figures) of the Fe electrodes, and TMR (bottom row of figures). In the plots for the current the red curves correspond to the ↑ current, the green curves the ↓ current, and the black curves to the total current. The results are for a 4 MgO MLs junction with no V_O (first column of figures), and with a V_O in the 2nd MgO layer from the interface (second column of figures).

For lower defect densities the dispersion of the defect-band is smaller, and so the decrease of the TMR at zero bias might be less, but the TMR will drastically shrink with an applied voltage. The behavior shown here is valid for all the defects that lead to a depolarization of the Δ_1 states around the Fe Fermi energy. In our calculation the defect states are located below E_F, but an analogous behavior is expected for states lying between E_F and the MgO conduction band. Our results also show that the overestimation of the TMR for an ideal junction is caused by an underestimation of the AP current. The AP current can be expected to drastically increase for non-ideal junctions, whereas the change in the P current is expected rather small.

In this section we have only described the effects of V_O's. The number of different possible defects in MgO is however very large. Each type of defect leads to a defect-band at a different energy in the band gap, as shown in Refs. [325, 326, 327, 328]. As an example, in Fig. 7.70 we show the band-structure for MgO with one V_{Mg} and one V_O pair close to each other, calculated using the ASIC. Such a defect is likely to change its charging state when the MgO is attached to Fe. The effects on the transport for such a defect pair are probably even larger than the ones for the isolated V_O, since the band gap lies between two defect bands and is now very small.

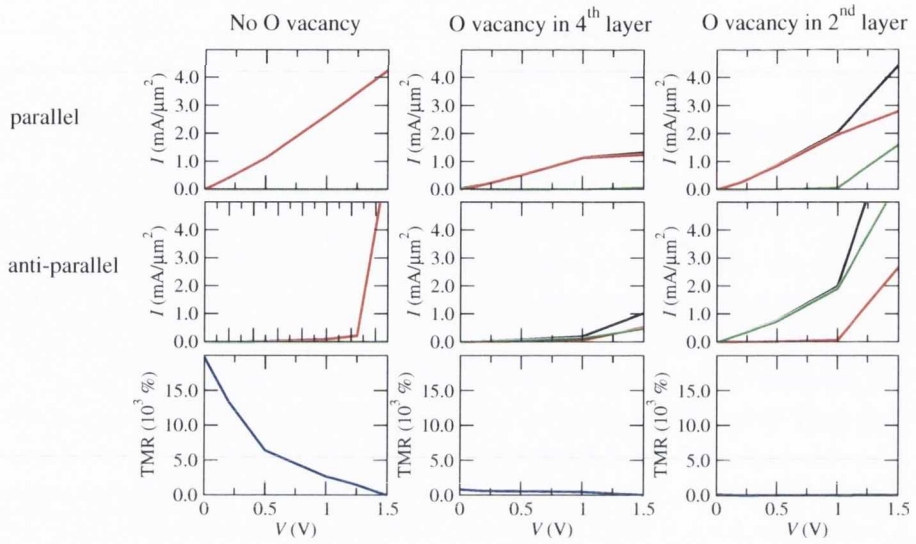


Figure 7.69: Spin-polarized current I for P (top row of figures) and for for AP alignment (middle row of figures) of the Fe electrodes, and TMR (bottom row of figures). In the plots for the current the red curves correspond to the $\uparrow$ current, the green curves the $\downarrow$ current, and the black curves to the total current. The results are for a 8 MgO MLs junction with no V_O (leftmost panels), with a V_O in the 4th layer from the interface (middle panels), and with a V_O in the 2nd layer from the interface (rightmost panels).

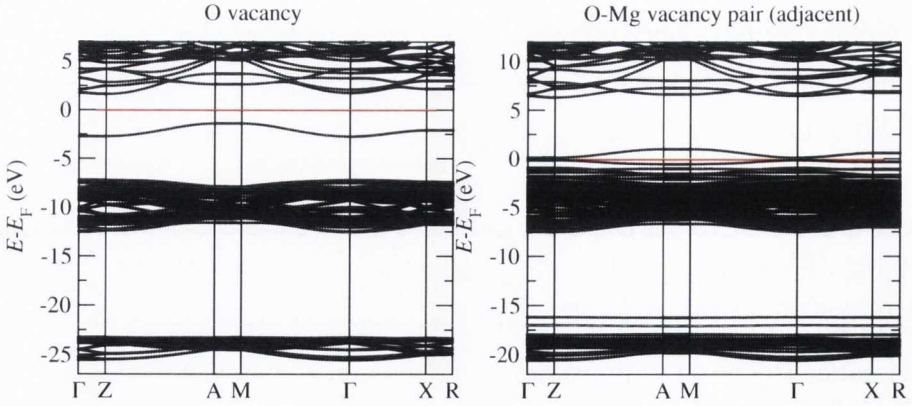


Figure 7.70: MgO band structure, calculated using the ASIC, with a single V_O (left panel), and with a V_{Mg}-V_O pair (right panel).

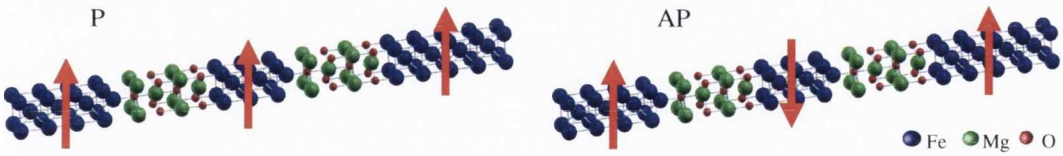


Figure 7.71: Unit cell used for the double MgO barrier junctions, with a schematic representation of the parallel (P) and antiparallel (AP) configurations. The red arrows indicate to the magnetization direction of the electrodes.

7.7 Double barrier junctions

Instead of growing only a single MgO barrier between two ferromagnetic electrodes, it is also possible to grow multiple insulator/metal layers [301, 302]. Here we investigate the transport properties of double MgO barrier junctions. One of the possible advantages of such systems is that the decay of the TMR with bias can be slower in comparison to single barrier junctions [301]. Such a slower decay of the TMR with bias is expected, since the potential drop is shared across the two barriers, and since an additional spin-dependent scattering potential is introduced by the in-between magnetic slab. Moreover, oscillations of the conductance as function of voltage have been measured [302]. These are attributed to the presence of quantum well states, confined in the middle Fe layer [302, 303, 304].

The system setup is illustrated in Fig. 7.71: two 4 MgO MLs thick insulating barriers are separated by 6 MLs of Fe. For the Fe/MgO interfaces we use the relaxed coordinates (see Sec. 7.1). Whereas for single barrier junctions (SBJs) in the antiparallel configuration the magnetization of the two Fe electrodes points in opposite directions, for a double barrier junction (DBJ) the antiparallel orientation is obtained by flipping the magnetization of the middle Fe layer, leaving the magnetization of the two Fe electrodes unchanged (see Fig. 7.71) [301, 302]. We note that in Ref. [302] it is shown that the thickness of the middle Fe layer is not constant over the sample, and that isolated Fe islands are formed, separated by regions with no Fe between the two MgO barriers.

The 0-bias spin-dependent transmission for the P and the AP configuration is shown in Figs. 7.72(a) and (b). In Figs. 7.72 (c) and (d) the same is shown for the SBJ (see Sec. 7.2). The order of magnitude of the transmission around E_F is similar for both the single and the DBJs. This also shows that the transmission through a DBJ with two 4 MgO MLs is much higher than the one through a SBJ with 8 MgO MLs. The main difference to the transmission of a SBJ is the appearance of a large number of peaks in the transmission, for both P and AP, $\uparrow$ and $\downarrow$. Such an oscillatory

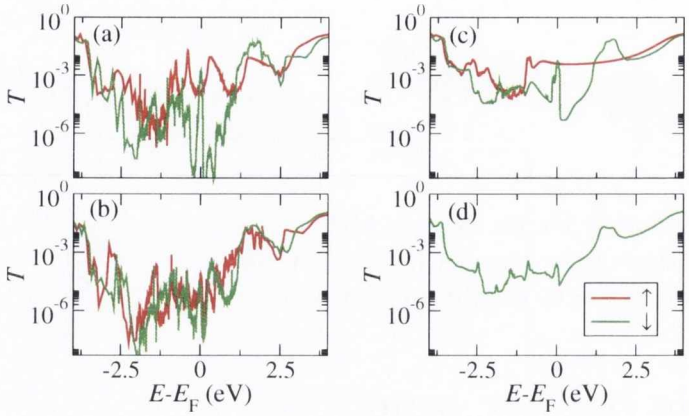


Figure 7.72: Spin dependent transmission coefficient T for the DBJ, for P (a) and AP (b) alignment of the Fe electrodes, and for the SBJ with 4 MgO MLs for P (c) and AP alignment (d).

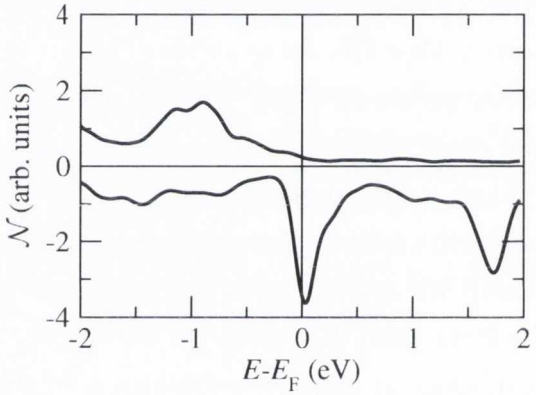


Figure 7.73: Total density of states $\mathcal{N}$ projected onto the interface Fe atom of the middle Fe layer. Positive values are for majority, and negative values for minority spin.

transmission coefficient can indeed lead to the measured conductance oscillations. The oscillation density is larger in the AP configuration, since as a first approximation the transmission in the AP configuration can be seen as a convolution of P majority and minority transmission. It therefore can be expected to include the peaks of both P $\uparrow$ and P $\downarrow$. We note however that the transmission in the AP configuration is highly resonant. This is also visible by the large amount of noise in the transmission coefficient. To obtain a smooth transmission coefficient, the number of **k**-points used in the BZ would have to be increased. If a small imaginary part, of the order of 10^{-4} to 10^{-3} Ry, is added to the energy when calculating the transmission coefficient, the peaks shrink drastically or disappear completely. For the P configuration this effect is less pronounced, although also here the peaks shrink considerably. Therefore effects related to the quantum well states are probably visible only in high-quality epitaxial junctions.

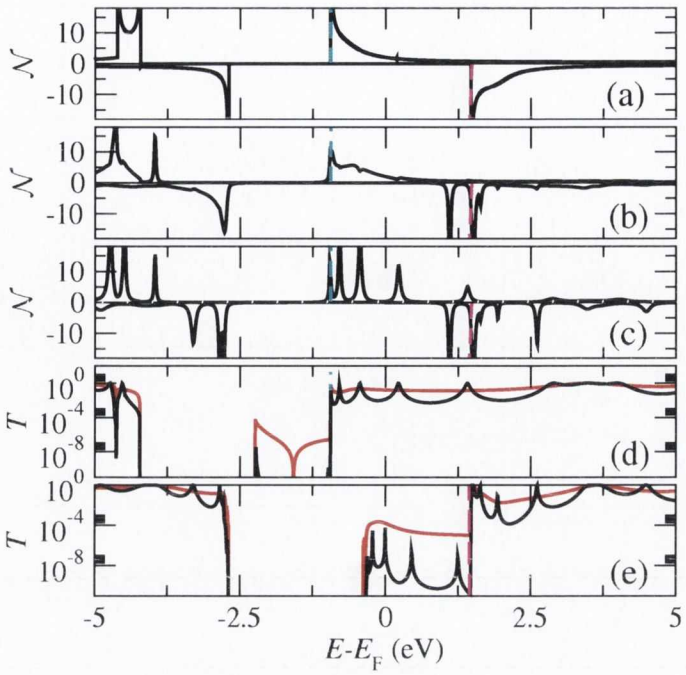


Figure 7.74: DOS $\mathcal{N}_{\Delta_1}$ projected on the states with Δ_1 symmetry, in arbitrary units (positive for $\uparrow$, negative values for $\downarrow$), and transmission at the Γ point. (a) Δ_1 DOS for bulk Fe, (b) Δ_1 DOS of the Fe atom at the interface between the left electrode and the MgO, (c) Δ_1 DOS of the left interface Fe atom of the middle Fe layer, (d) P $\uparrow$ transmission for the DBJ (black curves) and SBJ (red curves), and (e) P $\downarrow$ transmission for the DBJ (black curves) and SBJ (red curves)$\downarrow$ (green). The dashed vertical lines indicate the band-edges of the Δ_1 states at the Γ point for $\uparrow$ (cyan) and $\downarrow$ (magenta). In order to visualize all the peaks in the DOS, a small imaginary part has been added to the energy.

Some of the transmission peaks are also present in the single junction, such as the ones in the P $\downarrow$ located at E_F and at about $E_F+1.5$ eV. As discussed in Sec. 7.4, these two peaks are caused by Fe $\downarrow$ ISs. Other peaks, such as the ones in the P $\uparrow$ around E_F are only found in the DBJ. In order to find their origin, we analyze the DOS of the middle Fe layer, calculated by using the ground state DFT density matrix. In Fig. 7.73 the DOS projected onto the interface Fe atom of the middle Fe layer is shown for P alignment. This PDOS is very similar to the one of the interface Fe atom of a single 4 ML junction (Fig. 7.27). There are no visible peaks that could explain the $\uparrow$ peaks in the transmission coefficient. We therefore analyze the **k**-dependent PDOS for the states with a Δ_1 symmetry, as already done in Sec. 7.4 for the SBJ. In Fig. 7.74 the DOS $\mathcal{N}_{\Delta_1}$ at the Γ point ($k_x = k_y = 0$), projected onto the states with a Δ_1 symmetry, is shown together with the transmission coefficient for the Γ point only. Fig. 7.74(a) shows the DOS for bulk Fe, Fig. 7.74(b) is the DOS projected onto the interface Fe atom of the left electrode. These two DOSs are basically identical to

the corresponding of the SBJ (Fig. 7.28), so that all the related results discussed in Sec. 7.4 are also valid here. Fig. 7.74(c) is the DOS projected onto the left interface atom of the middle Fe layer. Here very distinct peaks appear in the $\uparrow$ around E_F. The position of these peaks is in very good agreement to the position of the peaks in the transmission at the Γ point [black curve in Fig. 7.74(d)]. The peaks in the $\downarrow$ transmission [black curve in Fig. 7.74(e)] around E_F are caused by $\downarrow$ states of a different symmetry, since, as discussed in Sec. 7.4, there is no Δ_1 state in the $\downarrow$ around E_F. These peaks are however very sharp, and their contribution to the current can be expected to be negligible. Since the transmission is dominated by the Δ_1 states at the Γ point, the total transmission [Fig. 7.72](a) shows the peaks at the same energies as found in the transmission for the Γ point only. We therefore conclude that the quantum well states do indeed lead to oscillations in the transmission, and therefore in the conductance. These results are in good agreement with the results of Ref. [303], where a detailed analysis of the Δ_1 DOS is performed for a large number of different thicknesses of the middle layer. Importantly, a transport experiment measures only those quantum well states for $k_x = k_y \approx 0$, and with Δ_1 symmetry. If all quantum well states would contribute equally to the conductance, no oscillations in the transmission would be expected, since the total DOS of the middle Fe layer is rather smooth and flat (Fig. 7.73). With increasing thickness of the middle layer the density of quantum well states increases, so that also the number of oscillations in the transmission coefficient can be expected to increase, eventually leading to a smooth, monotonic curve for very large thickness.

We now move on to the analysis of the finite bias properties. As shown in Sec. 7.3 for the SBJ, it is important to occupy the BSs in a physically meaningful way at all voltages. The BSs at the Fe/MgO interface of the left (right) Fe electrode can be expected to be coupled much more to the left (right) electrode, and therefore they can be occupied with the same approach used for the single junction. For the BSs appearing in the middle Fe layer it is not possible to determine such an energy-independent effective coupling to the electrodes. Their occupation has to be set in such a way, that the physically expected potential drop is obtained at self-consistency. Our junction is completely symmetric, we therefore expect the absolute value of the charging at each of the interfaces to be the same. This leads to an approximately linear potential drop in the MgO, with a flat ΔV_H in the middle Fe layer, its value lying in the middle between ΔV_H of the left- and ΔV_H of the right electrode. We tried different ways to occupy the BSs of the middle layer. The only way we could obtain this expected potential drop was by occupying the BSs up to a local, position-dependent Fermi energy, where the BSs on the left (right) Fe electrode

are occupied up to $E_{F,L}$ ($E_{F,R}$), and the ones of the middle Fe layer are occupied up to $(E_{F,L} + E_{F,R})/2$. This approach is described for a general system at the end of Sec. 6.2. In this particular case we use 3 energy-independent matrices α, defined as follows:

$$\alpha_1 = \begin{pmatrix} \mathbb{1}_{N_{\mathrm{EL}}} & \mathbb{0}_{N_{\mathrm{EL}},N_{\mathrm{ML}}} & \mathbb{0}_{N_{\mathrm{EL}},N_{\mathrm{ER}}} \\ \mathbb{0}_{N_{\mathrm{ML}},N_{\mathrm{EL}}} & \mathbb{0}_{N_{\mathrm{ML}}} & \mathbb{0}_{N_{\mathrm{ML}},N_{\mathrm{ER}}} \\ \mathbb{0}_{N_{\mathrm{ER}},N_{\mathrm{EL}}} & \mathbb{0}_{N_{\mathrm{ER}},N_{\mathrm{ML}}} & \mathbb{0}_{N_{\mathrm{ER}}} \end{pmatrix}, \tag{7.13}$$

with a corresponding local Fermi energy $E_{F,1} = E_{F,L}$. In an analogous way as for the SBJ [Eq. (6.27)], N_{EL} is the number of orbitals of the left electrode, including those of the leftmost two MLs of MgO. N_{ER} is the number of orbitals of the right electrode, including those of the rightmost two MLs of MgO. N_{ML} is the number of orbitals of the middle Fe layer, including those of two MgO MLs on each side. Here it is again assumed, that the orbitals are ordered in such a way, that the orbitals on the left have a lower index with respect to those on the right. The other two α matrices are given by

$$\alpha_2 = \begin{pmatrix} \mathbb{0}_{N_{\mathrm{EL}}} & \mathbb{0}_{N_{\mathrm{EL}},N_{\mathrm{ML}}} & \mathbb{0}_{N_{\mathrm{EL}},N_{\mathrm{ER}}} \\ \mathbb{0}_{N_{\mathrm{ML}},N_{\mathrm{EL}}} & \mathbb{1}_{N_{\mathrm{ML}}} & \mathbb{0}_{N_{\mathrm{ML}},N_{\mathrm{ER}}} \\ \mathbb{0}_{N_{\mathrm{ER}},N_{\mathrm{EL}}} & \mathbb{0}_{N_{\mathrm{ER}},N_{\mathrm{ML}}} & \mathbb{0}_{N_{\mathrm{ER}}} \end{pmatrix}, \tag{7.14}$$

with a corresponding local Fermi energy $E_{F,2} = (E_{F,L} + E_{F,R})/2$, and

$$\alpha_2 = \begin{pmatrix} \mathbb{0}_{N_{\mathrm{EL}}} & \mathbb{0}_{N_{\mathrm{EL}},N_{\mathrm{ML}}} & \mathbb{0}_{N_{\mathrm{EL}},N_{\mathrm{ER}}} \\ \mathbb{0}_{N_{\mathrm{ML}},N_{\mathrm{EL}}} & \mathbb{0}_{N_{\mathrm{ML}}} & \mathbb{0}_{N_{\mathrm{ML}},N_{\mathrm{ER}}} \\ \mathbb{0}_{N_{\mathrm{ER}},N_{\mathrm{EL}}} & \mathbb{0}_{N_{\mathrm{ER}},N_{\mathrm{ML}}} & \mathbb{1}_{N_{\mathrm{ER}}} \end{pmatrix}, \tag{7.15}$$

with a corresponding local Fermi energy $E_{F,3} = E_{F,R}$. We can calculate the energy-dependent effective coupling corresponding to this set of α-matrices, as described at the end of Sec. 6.2. Assuming that $E_{F,L} > E_{F,R}$, we find that the BSs in the middle layer are coupled mainly to the left electrode for energies smaller than $(E_{F,L} + E_{F,R})/2$, whereas for energies larger than $(E_{F,L} + E_{F,R})/2$ the coupling is much stronger to the right electrode. Although we used this effective coupling in all our finite bias calculations, further investigations are needed in order to understand why this effective coupling is the one that gives the physically meaningful result.

The planar average of the potential drop obtained with this set of α matrices at a bias voltage of 0.5 V is shown in Fig. 7.75(a). As required, it is indeed flat in the Fe layers, and drops approximately linearly in the two MgO barriers. The potential drop is equal on both barriers, so that the change of the potential in the middle Fe layers as function of bias is zero. The corresponding planar average of the charging [Fig. 7.75(b)] for each of the MgO barriers is similar to the one for the single junction [Fig. 7.25(b)].

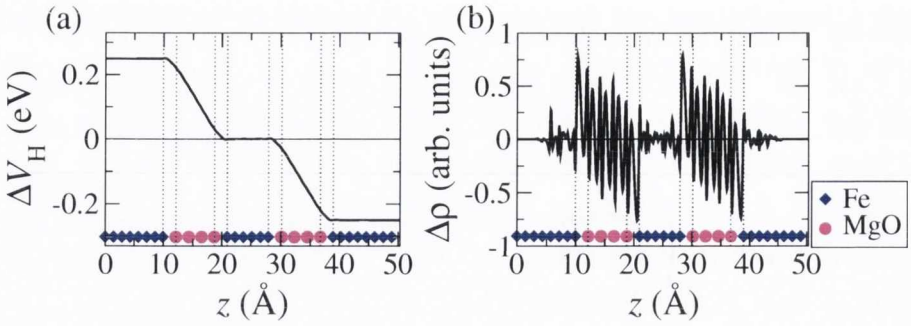


Figure 7.75: (a) Planar average ΔV_{H} of the difference between the Hartree potential at 0.5 V and the one at 0-bias. (b) Planar average $\Delta\rho$ of the difference between the charge density at 0.5 V and the one at 0-bias. The diamonds and dots indicate the location of the Fe and MgO layers.

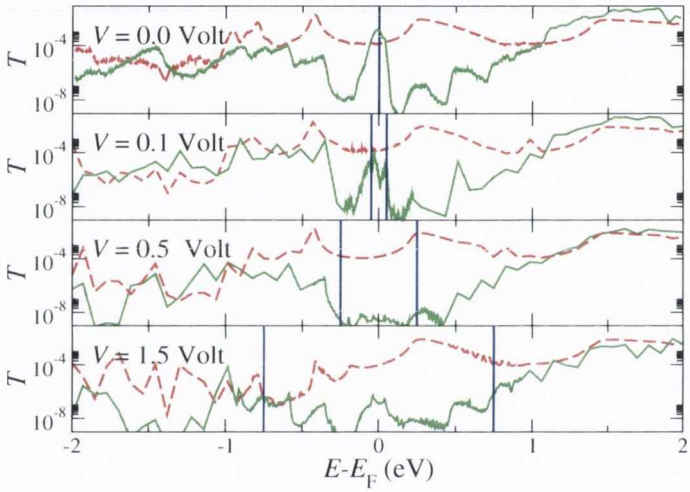


Figure 7.76: Spin dependent transmission coefficient T for the P configuration as a function of energy, E, and for different biases, V. The vertical lines are placed at $E = E_{\mathrm{F}} \pm eV/2$ and enclose the bias window.

The bias dependent transmission coefficient is shown in Fig. 7.76 for the P configuration, and in Fig. 7.77 for the antiparallel configuration. The transmission peaks for the P $\uparrow$ around E_{F} do not move with energy, and also their height is approximately constant. This is due to the fact that the energies of the quantum well states do not change as function of bias, since the potential in the middle layer is itself constant (Fig. 7.75). The transmission of these peaks is however reduced, once the bottom of the Fe $\uparrow$ Δ_1 band of the left electrode moves to higher energies. This is for example the case for the first peak below E_{F} at a voltage of 1.5 V. Such a behavior is analogous to the one found for the SBJ in Sec. 7.4, and it is entirely caused by the relative shift of the Δ_1 band-edges of the electrodes. The peaks that involve states on the left-hand or right-hand side electrodes, such as the $\downarrow$ IS at E_{F}, move as function of

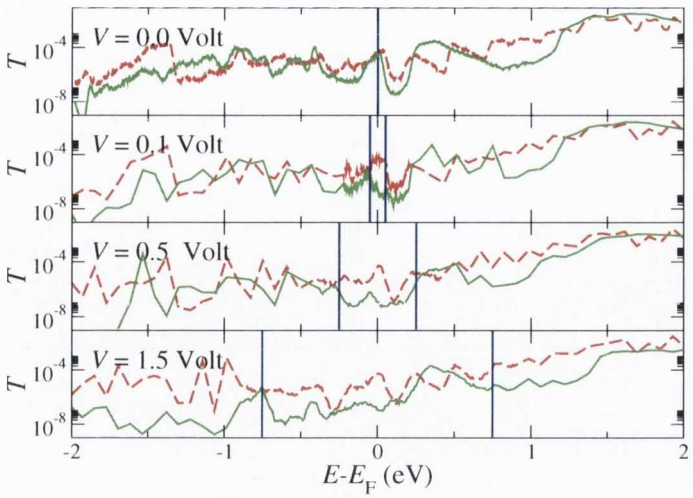


Figure 7.77: Spin dependent transmission coefficient T for the AP configuration as a function of energy, E, and for different biases, V. The vertical lines are placed at $E = E_F \pm eV/2$ and enclose the bias window.

bias. We note that also for the DBJ the $\downarrow$ peak associated to the Δ_1 band-edge does not contribute to the current, since the ISs get out of resonance with applied bias, in the same way as for the SBJ. In the AP transmission the peaks move in an analogous way with applied bias: all those that are associated with quantum well states are in a fixed position, whereas the ones associated with the electrodes move according to the applied potential. For the SBJ we found that the peak in transmission at about 1.5 eV shifts with bias, without changing its height (Fig. 7.32), and therefore leads to a high current when it enters the bias window. For the DBJ the situation is very different, as schematically shown in Fig. 7.78. For AP $\uparrow$ spins [Fig. 7.78(a) and (b)] the peak does not move with bias, since the Δ_1 ISs are located in the middle layer. Therefore the peak will only enter the bias window at a bias that is twice as high as the one for the SBJ, where the peak moves towards the bias window for increasing voltage. For AP $\downarrow$ [Fig. 7.78(c) and (d)] the ISs are located on the electrode Fe layers. These shift with bias accordingly, and therefore get out of resonance for high enough bias. The result is that the $\downarrow$ spins do not contribute to the current also for the AP configuration, which is dominated by the $\uparrow$ spins at high bias. We can therefore expect the conductance in the AP configuration to rise with increasing bias, in the same way as for the SBJ, but the increase will be slower. This should therefore result in a slower decay of the TMR with bias.

The resulting I-V and bias-dependent TMR are shown in Fig. 7.79. Since in the P configuration there is a dip in the transmission around E_F, the low bias current is rather small, but increases rapidly once the bias window reaches the first peak in the $\uparrow$

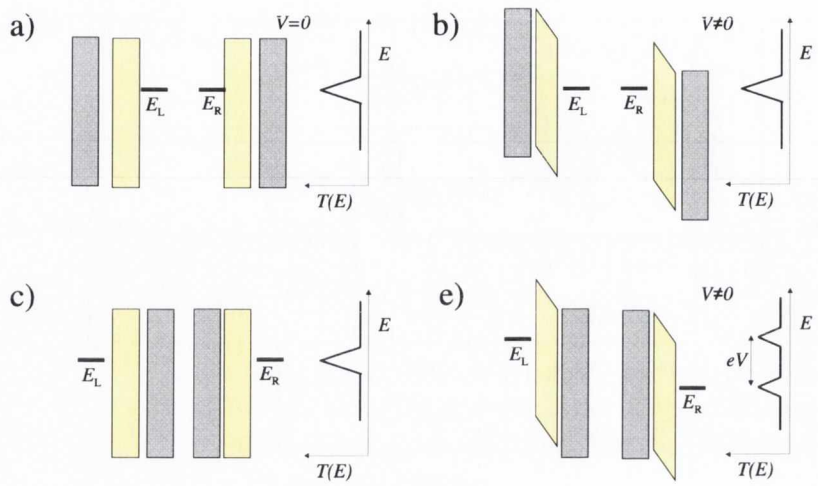


Figure 7.78: Schematic representation of the transmission through ISs for the AP configuration, (a) and (b) are for the ↑ transmission, and (c) and (d) for the ↓ transmission.

transmission at about 0.2 eV, which is the case for applied voltages larger than 0.4 V. The next ↑ peak to enter the bias window is the one at about -0.4 eV, which is the case for a voltage of about 0.8 V. Once both these peaks are inside the bias window, the P current saturates. At higher bias there is even a slight reduction of the current, which is due to the fact that the $\Delta_1^{\uparrow}$ band-edge of the left lead is at a higher energy than the peak at -0.4 eV, so that the conduction through this peak is suppressed. The current in the AP configuration is always much smaller than the one in the P configuration, even at high bias. This is in contrast to the single barrier result (Fig. 7.36), where the AP current rises drastically above about 1.25 V. As already mentioned, the reason for the smaller current in the DBJ with AP configuration is that the broad transmission peak at about 1.5 eV does not contribute significantly to the current. The resulting TMR is therefore large even at high voltages, whereas the TMR for the SBJ drops to zero already at about 1.5 V (Fig. 7.37). These calculations therefore confirm the experimentally measured slower decay of TMR with bias for double junctions [301]. We note that in the experimental junction this is only the case for positive voltage, the authors in Ref. [301] argue that this is due to asymmetric interfaces. The drop of TMR just above zero bias voltage is due to the increase of the AP current with respect to the P one. The subsequent rise of the TMR is mainly due to the increase of conductance in the P configuration, caused itself by the first quantum well state entering the bias window. For thicker middle layers we expect the low bias behavior to get more similar to the one for the single junction, due to the higher density in energy of quantum well states. The low current region can be expected to be reduced for thicker middle layers. We note however that the high bias behavior is independent

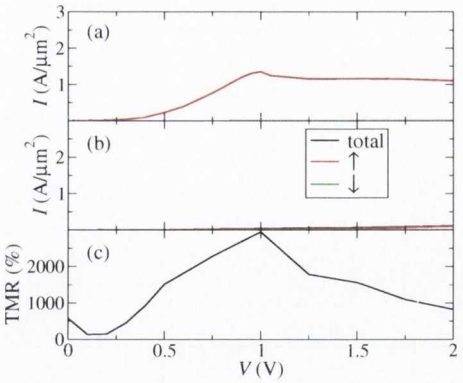


Figure 7.79: Spin-polarized current I for the P (a) and AP (b) configurations, and TMR (c) as function of voltage V for the DBJ.

of the thickness of the Fe middle layer, and is mainly determined by the relative shift of the $\downarrow \Delta_1$ band-edges in the Fe electrodes, and of the connected ISs.

We conclude this section by discussing the bias-dependent conductance. Fig. 7.80 shows the conductance as function of bias voltage. Since we calculated the I-V only for a rather coarse voltage mesh, the conductance is not very accurate, however the main features are still recognizable. For the SBJ the conductance as function of bias is much smoother (Fig. 7.36), since the variations are mainly due to the shift of the Fe Δ_1 band-edges in the electrodes. The P $\uparrow$ conductance is very small at zero bias, and at 0.4 and 0.8 V there are the two peaks caused by the transmission through quantum well states. At high bias the conductance is approximately zero, or even slightly negative. The P $\downarrow$ conductance is much smaller than the $\uparrow$ conductance for all voltages. Also the conductance for the AP configuration is always much smaller than the P $\uparrow$ conductance (note the different scales in Fig. 7.80). In the AP configuration the conductance oscillates visibly as function of voltage. Both the P and AP conductance therefore reflect the presence of quantum well states in form of oscillations. These oscillations have also been found experimentally [301], and attributed to quantum well states. The different positions of the conductance peaks are probably due to the fact that their middle Fe layer is thicker.

Since for the P configuration the transmission in an energy range of about 1 eV around E_F is approximately constant for different bias voltages, we can also calculate the current by simply integrating the 0-bias transmission coefficient for the P $\uparrow$. In Fig. 7.81 we show the so obtained conductance, compared with the one calculated self-consistently for the P configuration for an electronic temperature of 300 K. As expected, the two curves agree very well. A reduction of the electronic temperature in the non-self-consistent calculation reduced the broadening of the peaks, so that they become much sharper. In Ref. [301] such a strong dependence of the width of

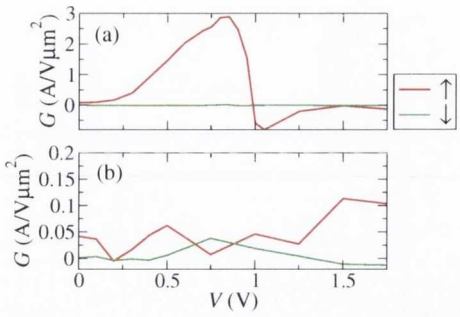


Figure 7.80: Spin-polarized conductance G for the P (a) and AP (b) configurations. Note the different scale for P and AP.

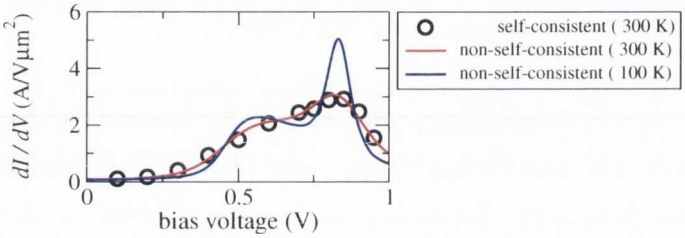


Figure 7.81: Conductance $G = dI/dV$ for the P configuration, for the self-consistent solution at an electronic temperature of 300 K (black circles), and for the non-self-consistent solution at an electronic temperature of 300 K (red curve) and 100 K (black curve).

the peaks on temperature is experimentally observed.

7.8 Conclusions

We investigated the bias dependence of the current and TMR of epitaxial Fe/MgO/Fe tunnel junctions from first principles. We identify two different bias regions, which are affected by two types of electronic resonances: (i) interface resonance states and (ii) band-edges of the states with Δ_1 symmetry. For a 4 MgO MLs junction, at voltages smaller than 20 mV the I-V characteristics and the TMR are dominated by resonant transport through narrow interface states (ISs) in the minority spin-band. In the parallel configuration this contribution is quenched by a voltage comparable to the energy width of the IS, whereas it persists at all voltages in the antiparallel configuration. The low bias region is therefore characterized by a non-monotonic behavior of the TMR.

In the high bias region ($V > 0.4$ V) the TMR decreases monotonically, mainly due to band-edge related ISs. The transport is mainly determined by the relative positions of the $\Delta_1^{\uparrow}$ and $\Delta_1^{\downarrow}$ band-edges in the two Fe electrodes. The increase of the current with bias is much larger in the AP configuration than in the P configuration. This is caused by the $\Delta_1^{\downarrow}$ band-edge moving closer to, and finally entering the bias window. We have also shown how disorder can suppress the transport through ISs, leading to a monotonic decay of the TMR with bias at all voltages, in good agreement to experiment [264]. The value of the TMR however overestimates the experimental one, which is due to the fact that we consider an ideal junction.

We have also investigated the dependence of the transport properties on the thickness of the MgO barrier, and found that the results are sensitive to the used basis set. By analyzing the complex band structure of MgO we found that the reason for this is the finite overlap between the basis orbitals. The basis set with minimal overlap was found to give the best results at large thicknesses.

The effect of defects in the MgO has been studied in the form of oxygen vacancies. In a junction with oxygen vacancies in the MgO the TMR can be drastically reduced compared to the one for an ideal junction. The closer the vacancies are to the interface, the larger is the reduction of the TMR. The origin for this is an increase of the current in the AP configuration, which is caused by a spin-depolarization of the states with Δ_1 symmetry at the vacancy site.

We finally have studied the transport properties of double MgO barrier junctions. In the low bias region the conductance is found to oscillate due to the presence of quantum well states in the middle Fe spacer. This causes a non-monotonic behavior of the TMR at low bias. At high bias the TMR decays monotonically with bias, but the decay is slower than that of the single barrier junction. We explained this by

a simple model, based on the relative positions of the Δ_1 band edges and interface states. The results are in qualitative agreement with experiments [301, 302].

Chapter 8

Conclusions and future work

In summary, we have developed theoretical and computational tools for the calculation of electronic transport within the non-equilibrium Green's functions (NEGF) formalism and applied them to some topics of current interest. An accurate and stable method for the calculation of the self-energies of semi-infinite quasi-one-dimensional systems has been presented, and a formalism for the inclusion of bound states in the NEGF method at finite bias voltage has been derived. These concepts have been implemented in the electronic transport code SMEAGOL [68, 61]. We used the developed tools for finite bias transport calculations through Mn_{12} based magnetic molecules and across Fe/MgO(100) single and double barrier tunnel junctions. We also performed ground state density functional theory (DFT) calculations using the SIESTA code, where we developed a model for the magneto-structural phase transitions of MnAs by mapping the DFT total energies onto a Heisenberg model.

We began in chapter 2 by giving an overview of DFT and its implementation in the SIESTA code. We presented the Hohenberg-Kohn and the Kohn-Sham theorems for the many-electron system, which are at the core of DFT. An overview of existing exchange-correlation functionals was given, and the atomic self-interaction correction (ASIC) was introduced [90]. In the second part of the chapter we presented the *ab initio* DFT code SIESTA.

In chapter 3 we investigated the magneto-structural phase transitions of MnAs between the ferromagnetic hexagonal $B8_1$ structure and the paramagnetic and anti-ferromagnetic orthorhombic B31 structure. We calculated the magnetic exchange coupling parameters by mapping the DFT total energies for different configurations of the Mn magnetic moments onto a Heisenberg Hamiltonian. This allowed us to extract the Curie temperature within a mean field approximation. We found that the $B8_1$ structure favors a ferromagnetic ground state, whereas the B31 structure favors an anti-ferromagnetism ground state. The ferromagnetic Curie Temperature is found to decrease by reducing the volume and by distorting the cell from the $B8_1$ to the

B31 structure. This dependence on the volume confirms the main assumption of the various phenomenological models based on the original idea of Bean and Rodbell [49]. Furthermore it has been shown that for paramagnetic states the B31 structure is stable at small volumes, while the $B8_1$ structure is stable above a critical lattice constant of about 3.7 Å. By using these results we could provide an explanation of the phase diagram and of the anomalous behavior of the susceptibility, which is determined mainly by the changes of the lattice structure and volume as function of temperature. Whereas in our model we predict a transition from the low temperature ferromagnetic state to a paramagnetic state, accompanied by a change in structure from the $B8_1$ to the B31, recent experiments for thin MnAs films indicate a transition to an anti-ferromagnetic phase for temperatures close to the phase transition temperature T_p [330, 331]. One of the possible explanations is that the paramagnetic phase above T_p is not completely random, and that there is some local anti-ferromagnetic order. Further investigations are needed to understand the origin of these experimental findings.

In chapter 4 the main concepts of the NEGF formalism for quantum transport calculations were presented. This formalism allows us to calculate the charge density of nano-devices attached to semi-infinite leads with different chemical potentials. Once the charge density is calculated self-consistently, the electric current flowing through the device can be obtained by integrating the bias-dependent transmission coefficient over the bias window. The implementation of the NEGF concepts in the *ab initio* transport code SMEAGOL has been outlined. Although the NEGF method works well for many systems, we also discussed some of its limitations, especially when used in conjunction with DFT.

An improved method for the calculation of the self-energies (SEs) of quasi one-dimensional systems was presented in chapter 5. By using this method for the calculation of the SEs in SMEAGOL both the accuracy and stability of the code are improved compared to the previous implementation, which is described in Refs. [214, 68]. The main advantage of the improved method is that all the singularities in the computation are avoided wherever possible, and that it involves no inversions of ill-conditioned matrices. Optionally a transformation based on a singular value decomposition (SVD) can be performed, in order to reduce the condition number of the hopping matrices, and also to make the computations faster. We performed a detailed analysis of the accuracy, and found that at some energies the accuracy is very low. Our analysis showed that the loss of accuracy is due to the appearance of localized surface states for the semi-infinite system. At the energy of the surface state one of the eigenvalues of the SE diverges, and this results in a reduction of the accuracy. Since surface states are an inherent physical property of the system, and are independent of the calcula-

tion method, such singularities can not be avoided. We have restored the accuracy at these particular energies by adding a small imaginary part to the energy. We showed that this can cause significant errors in the low transmission regime, whereas for high transmission the errors are negligible.

In chapter 6 a formalism for the inclusion of bound states (BSs) in the calculation of the non-equilibrium charge density was outlined. Their occupation was set based on the assumption of an effective coupling of these states to the leads. This can usually be determined based on physical constraints, such as the geometrical location of a BS inside a junction. We have also introduced the concept of generalized bound states (GBSs), which allowed us to describe both weakly coupled and bound states with a single formalism. The presence of very weakly coupled states is reflected by sharp narrow peaks in the density of states (DOS), which have to be integrated over energy in the out of equilibrium case. For such systems an extremely fine energy mesh is needed, making the calculations intractable if a regularly spaced mesh is used. We have presented two possible solutions: the first is the bound states correction scheme (BSCS), in which the peaks are broadened and the occupation of the GBSs is set separately, and the second is the use of an adaptive energy mesh algorithm. We have applied both methods to an example system with very weak coupling, consisting of a C_3 molecule in the middle of a parallel plate gold capacitor, and found them to work well. We have also applied these methods to study the transport through a Mn_{12} magnetic molecule, which is very weakly coupled to the gold electrodes through benzene groups. We found that at finite bias the set of closely spaced orbitals below the Fermi energy rehybridizes, localizing on either the left- of right-hand side of the molecule. This results in a highly asymmetric coupling of the states, which is reflected by a negative differential conductance at specific voltages. We calculated I-V curves for two different spin states of the molecule. The I-Vs for the two states are found to be similar, although there are small differences, which might be detectable in experiments. We note that one of the drawbacks of our proposed treatment of BSs is that their effective coupling has to be added "by hand". It might however be possible to define a local Fermi energy inside the extended molecule, which determines the occupation of the BSs. This local Fermi energy could be determined by the difference of the potential at finite bias and the one at zero bias. The results obtained for the single and double barrier Fe/MgO tunnel junctions in chapter 7 suggest that this might be a viable option.

Finally, in chapter 7 we studied the spin-dependent transport properties of Fe/MgO/Fe(100) tunnel junctions. We showed that in order to obtain a physically meaningful self-consistently potential at finite bias it is crucial to occupy the BSs

localized at the Fe/MgO interface correctly. For a junction with a barrier thickness of 4 MgO monolayers two features were found to determine the I-V characteristics: 1) interface states, and 2) band-edges of the states with high transmission in the Fe electrodes, which are those of Δ_1 symmetry [20]. The resulting TMR-V curve is non-monotonic below about 0.4 V (low bias region), and decreases monotonically for larger voltages (high bias region), until it eventually becomes negative above 1.7 V. By adding a small imaginary part to the energy we could remove the contribution of the sharp interface states to the conductance, in which case we found the TMR-V curve to decrease monotonically with increasing bias even for small voltages. For thicker junctions, where the contribution from the sharp interface states is filtered out, we found a much higher TMR, which decays monotonically for all bias voltages, in good agreement to experimental data [264]. We also studied junctions with oxygen vacancies (V_O) in the MgO, and found that the presence of such defects drastically reduces the value of the TMR. The decrease of the TMR is caused by a reduction of the spin-polarization of the states with Δ_1 symmetry at the vacancy site, which leads to an increase of the current in the anti-parallel configuration, especially if the V_O is close to the interface. In the last part of this chapter we studied the transport properties of double barrier junctions. We found that the features of the I-V curves are determined by the same two factors that characterize those of single barrier junctions, but also by the quantum well states with Δ_1 symmetry that are formed in the middle Fe layer. These cause oscillations in the conductance, which are also reflected in a non-linear behavior of the TMR-V curve at low bias. At high voltage the TMR decreases with bias, but the decrease is smaller than the one for the single barrier junction. We illustrated the origin of the slower decay using a simple model, based on the evolution of the Δ_1 minority interface states of the two Fe electrodes with bias. Our results are in qualitative agreement with experimental data [302].

Although by using SMEAGOL in its current form we can calculate the transport properties of a number of materials systems, there are many useful improvements that can be added in a future work. The first concerns the underlying DFT part, and consists in the use of improved exchange correlation functionals, which should ideally describe the derivative discontinuity in a proper way [202]. This is expected to be the case for exact-exchange based functionals [332, 78, 333, 334, 335]. The integration of these in SMEAGOL for zero bias, and eventually also for finite bias calculations would improve the description of weakly coupled systems [202]. We note however

that for such systems the single-particle NEGF approach might not be suited, and a rate-equation type approach is more appropriate [203, 204, 205, 206, 207]. A possible way to incorporate such systems with the NEGF formalism in SMEAGOL might be to use a hybrid approach, where the occupation of the states with strong coupling is determined by the standard NEGF equations, whereas the occupation of the weakly coupled states is set independently, for example using rate equations. Such a scheme is similar in spirit to the bound states correction scheme, in which the occupation of the weakly coupled states is already set separately from the standard NEGF part.

Another possible improvement is the inclusion of inelastic effects, which play a major role in some systems [336, 337, 300, 246]. The electron-phonon interaction for example can be treated approximately using the self-consistent Born approximation [338, 183]. In order to simulate unstructured disorder the coherent potential approximation (CPA) can be included in SMEAGOL [273].

Another important extension that would give access to a whole new class of physical systems is the ability to treat large systems with tens to hundred thousands of atoms, such as long organic molecules in solution, or self-assembled organic layers. In order to achieve this goal the scaling of the computational time and memory usage must be linear with increasing system size. Different approaches on how to achieve such a scaling in non-equilibrium calculations are given in the literature [339, 340, 341, 317]. In practical terms the use of a minimal basis set is another important aspect for large scale simulations, and can cut down the computational costs significantly. Possible ways to achieve this are the use of maximally localized Wannier-functions [342] or the use of an adaptive basis set [343]. Ideally the implementation of a code that can treat large systems and the improvement of the underlying DFT exchange correlation functionals should go hand in hand.

Bibliography

[1] G. Binasch, P. Grünberg, F. Saurenbach, and W. Zinn, Phys. Rev. B **39**, 4828 (1989).

[2] M. N. Baibich, J. M. Broto, A. Fert, F. N. Van Dau, F. Petroff, P. Etienne, G. Creuzet, A. Friederich, and J. Chazelas, Phys. Rev. Lett. **61**, 2472 (1988).

[3] G. A. Prinz, Science **282**, 1660 (1998).

[4] S. A. Wolf, D. D. Awschalom, R. A. Buhrman, J. M. Daughton, S. von Molnar, M. Roukes, A. Chtchelkanova, and D. Treger, Science **294**, 1488 (2001).

[5] M. Ziese and M. J. Thornton, eds., *Spin Electronics* (Springer, Berlin/Heidelberg, 2001).

[6] J. Fabian, A. Matos-Abiaguea, C. Ertlera, P. Stano, and I. Zutic, acta physica slovaca **57**, 565 (2007).

[7] C. Chappert, A. Fert, and F. N. Van Dau, Nature Materials **6**, 813 (2007).

[8] A. Fert, Angew. Chem. Int. Ed. **47**, 5956 (2008).

[9] M. A. M. Gijs and G. E. W. Bauer, Adv. in Phys. **46**, 285 (1997).

[10] E. Y. Tsymbal and D. G. Pettifor, Solid State Physics **56**, 113 (2001).

[11] J. Bass and W. P. Pratt Jr., J. Magn. Magn. Mater. **200**, 274 (1999).

[12] S. Maekawa and T. Shinjo, eds., *Spin dependent transport in magnetic nanostructures* (CRC Press, Boca Raton, 2002).

[13] J. S. Moodera, J. Nassar, and G. Mathon, Annual Review of Materials Science **29**, 381 (1999).

[14] P. M. Levy and S. Zhang, Cur.Opin.in Sol.State & Mat.Sci. **4**, 223 (1999).

[15] C. Tiusan, F. Greullet, M. Hehn, F. Montaigne, S. Andrieu, and A. Schuhl, J. Phys.: Condens. Matter **19**, 165201 (2007).

[16] S. Yuasa and D. D. Djayaprawira, J. Phys. D: Appl. Phys. **40**, R337 (2007).

[17] D. Wang, C. Nordman, J. Daughton, Z. Qian, and J. Fink, IEEE Trans. Magn. **40**, 2269 (2004).

[18] S. Ikeda, J. Hayakawa, Y. Ashizawa, Y. M. Lee, K. Miura, H. Hasegawa, M. Tsunoda, F. Matsukura, and H. Ohno, Appl. Phys. Lett. **93**, 082508 (2008).

[19] Y. M. Lee, J. Hayakawa, S. Ikeda, F. Matsukura, and H. Ohno, Appl. Phys. Lett. **90**, 212507 (2007).

[20] W. H. Butler, X.-G. Zhang, T. C. Schulthess, and J. M. MacLaren, Phys. Rev. B **63**, 054416 (2001).

[21] W. H. Butler, Sci. Technol. Adv. Mater. **9**, 014106 (2008).

[22] J. R. Petta, S. K. Slater, and D. C. Ralph, Phys. Rev. Lett. **93**, 136601 (2004).

[23] Z. H. Xiong, D. Wu, Z. V. Vardeny, and J. Shi, Nature **427**, 821 (2004).

[24] K. Tsukagoshi, B. W. Alphenaar, and H. Ago, Nature **401**, 572 (1999).

[25] B. W. Alphenaar, K. Tsukagoshi, and H. Ago, Physica E **6**, 848 (2000).

[26] S. J. Blundell and F. L. Pratt, J. Phys.: Condens. Matter **16**, R771 (2004).

[27] R. Sessoli, H. L. Tsai, A. R. Schake, S. Wang, J. B. Vincent, K. Folting, D. Gatteschi, G. Christou, and D. N. Hendrickson, J. Am. Chem. Soc. **115(5)**, 1804 (1993).

[28] L. Zobbi, M. Mannini, M. Pacchioni, G. Chastanet, G. Bonacchi, C. Zanardi, R. Biagi, U. del Pennino, D. Gatteschi, A. Cornia, et al., Chem. Commun. (Cambridge) p. 1640 (2005).

[29] S. Voss, M. Fonin, U. Rüdiger, M. Burgert, and U. Groth, Appl. Phys. Lett. **90**, 133104 (2007).

[30] S. Fonin, M. Voss, M. Burgert, Y. S. Dedkov, U. Groth, and U. Rüdiger, J. Phys.: Conference Series **100**, 052070 (2008).

[31] I. Zutic, J. Fabian, and S. C. Erwin, J. Phys.: Condens. Matter **19**, 165219 (2007).

[32] G. Schmidt, D. Ferrand, L. W. Molenkamp, A. T. Filip, and B. J. van Wees, Phys. Rev. B **62**, R4790 (2000).

[33] H. Ohno, Science **281**, 951 (1998).

[34] H. Ohno, J. Magn. Magn. Mater. **200**, 110 (1999).

[35] Y. Ohno, D. K. Young, B. Beschoten, F. Matsukura, H. Ohno, and D. D. Awschalom, Nature **402**, 790 (1999).

[36] R. Fiederling, M. Keim, G. Reuscher, W. Ossau, G. Schmidt, A. Waag, and L. W. Molenkamp, Nature **402**, 787 (1999).

[37] T. Jungwirth, K. Y. Wang, J. Masek, K. W. Edmonds, J. König, J. Sinova, M. Polini, N. A. Goncharuk, A. H. MacDonald, M. Sawicki, et al., Phys. Rev. B **72**, 165204 (2005).

[38] M. Wang, R. P. Campion, A. W. Rushforth, K. W. Edmonds, C. T. Foxon, and B. L. Gallagher, cond-mat/0808.1464 (2008).

[39] T. Jungwirth, J. Sinova, J. Masek, J. Kucera, and A. H. MacDonald, Rev. Mod. Phys. **78**, 809 (2006).

[40] C. Liu, F. Yun, and H. Morkoc, Journal of Materials Science: Materials in Electronics **16**, 555 (2005).

[41] J. M. D. Coey, Current Opinion in Solid State and Materials Science **10**, 83 (2006).

[42] S. A. Chambers, Surf. Sci. Rep. **61**, 345 (2006).

[43] C. D. Pemmaraju, R. Hanafin, T. Archer, H. B. Braun, and S. Sanvito, Phys. Rev. B **78**, 054428 (2008).

[44] X. Jiang, R. Wang, R. M. Shelby, R. M. Macfarlane, S. R. Bank, J. S. Harris, and S. S. P. Parkin, Phys. Rev. Lett. **94**, 056601 (2005).

[45] R. Wang, X. Jiang, R. M. Shelby, R. M. Macfarlane, S. S. P. Parkin, S. R. Bank, and J. S. Harris, Applied Physics Letters **86**, 052901 (pages 3) (2005).

[46] B. T. Jonker, G. Kioseoglou, A. T. Hanbicki, C. H. LI, and P. E. Thompson, Nature Physics **3**, 542 (2007).

[47] M. Ramsteiner, O. Brandt, T. Flissikowski, H. T. Grahn, M. Hashimoto, J. Herfort, and H. Kostial, Phys. Rev. B **78**, 121303(R) (2008).

[48] M. Ramsteiner, H. Y. Hao, A. Kawaharazuka, H. J. Zhu, M. Kästner, R. Hey, L. Däweritz, H. T. Grahn, and K. H. Ploog, Phys. Rev. B **66**, 081304(R) (2002).

[49] C. P. Bean and D. S. Rodbell, Phys. Rev. **126**, 104 (1962).

[50] V. A. Chernenko, L. Wee, P. G. McCormick, and R. Street, J. Appl. Phys. **85**, 7833 (1999).

[51] S. Gama, A. A. Coelho, A. de Campos, A. M. Carvalho, F. C. G. Gandra, P. J. von Ranke, and N. A. de Oliveira, Phys. Rev. Lett. **93**, 237202 (2004).

[52] D. H. Mosca, F. Vidal, and V. H. Etgens, Phys. Rev. Lett. **101**, 125503 (2008).

[53] V. Garcia, Y. Sidis, M. Marangolo, F. Vidal, M. Eddrief, P. Bourges, F. Maccherozzi, F. Ott, G. Panaccione, and V. H. Etgens, Phys. Rev. Lett. **99**, 117205 (2007).

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

[55] S. Sanvito, in *Handbook of Computational Nanotechnology* (American Scientific, Stevenson Ranch, CA, 2005), Vol. 5; Also available at cond-mat/0503445.

[56] M. Di Ventra, *Electrical Transport in Nanoscale Systems* (Cambridge University Press, Cambridge, UK, 2008).

[57] P. Weinberger, Phys. Rep. **377**, 281 (2003).

[58] I. I. Mazin, Phys. Rev. Lett. **83**, 1427 (1999).

[59] S. Kurth, G. Stefanucci, C. Almbladh, A. Rubio, and E. Gross, Phys. Rev. B **72**, 035308 (2005).

[60] M. Büttiker, Y. Imry, R. Landauer, and S. Pinhas, Phys. Rev. B **31**, 6207 (1985).

[61] A. Rocha, V. Garcia-Suarez, S. Bailey, C. Lambert, J. Ferrer, and S. Sanvito, Phys. Rev. B **73**, 085414 (2006).

[62] M. Brandbyge, J. Mozos, P. Ordejón, J. Taylor, and K. Stokbro, Phys. Rev. B **65**, 165401 (2002).

[63] H. Taylor, J. Guo and J. Wang, Phys. Rev. B **63**, 245407 (2001).

[64] H. J. Choi and J. Ihm, Phys. Rev. B **59**, 2267 (1999).

[65] P. A. Khomyakov, G. Brocks, V. Karpan, M. Zwierzycki, and P. J. Kelly, Phys. Rev. B **72**, 035450 (2005).

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

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

[68] A. Rocha, V. Garcia-Suarez, S. Bailey, C. Lambert, J. Ferrer, and S. Sanvito, Nature Materials **4**, 335 (2005).

[69] J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejon, and D. Sanchez-Portal, J. Phys.: Condens. Matter **14**, 2745 (2002).

[70] R. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, Oxford, UK, 1989).

[71] C. Fiolhais, F. Nogueira, and M. Marques, *A Primer in Density Functional Theory* (Springer, 2003).

[72] R. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, UK, 2004).

[73] K. Burke, *The abc of dft*, `http://dft.rutgers.edu/kieron/beta/b4.pdf` (2003).

[74] M. Born and J. R. Oppenheimer, Ann. Physik **84**, 457 (1927).

[75] E. Merzbacher, *Quantum Mechanics* (John Wiley & Sons Inc., New York, 1998).

[76] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry* (Macmillan, New York, 1982).

[77] J. Janak, Phys. Rev. B **18**, 7165 (1978).

[78] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Phys. Rev. Lett. **49**, 1691 (1982).

[79] J. Perdew and M. Levy, Phys. Rev. B **56**, 16021 (1997).

[80] C.-G. Zhan, J. A. Nichols, and D. A. Dixon, J. Phys. Chem. **107**, 4184 (2003).

[81] U. Von Barth and L. Hedin, J. Phys. C **5**, 1629 (1972).

[82] J. E. Peralta, G. E. Scuseria, and M. J. Frisch, Phys. Rev. B **75**, 125119 (2007).

[83] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

[84] J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).

[85] J. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[86] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **44**, 943 (1991).

[87] V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, J. Phys: Condens. Matter **9**, 767 (1991).

[88] J. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[89] T. Grabo, T. Kreibich, and E. K. U. Gross, Molecular Engineering **7**, 27 (1997).

[90] C. Pemmaraju, T. Archer, D. Sanchez-Portal, and S. Sanvito, Phys. Rev. B **75**, 045101 (2007).

[91] A. Filippetti and N. Spaldin, Phys. Rev. B **67**, 125109 (2003).

[92] D. Vogel, P. Krüger, and J. Pollmann, Phys. Rev. B **54**, 5495 (1996).

[93] C. Toher and S. Sanvito, Phys. Rev. Lett. **99**, 056801 (2007).

[94] X. Gonze, J.-M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G.-M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, et al., Comput. Mat. Science **25**, 478 (2002).

[95] S. Blügel and G. Bihlmayer, in *Computational Nanoscience: Do It Yourself!*, edited by J. Grotendorst, S. Blügel, and D. Marx (John von Neumann Institute for Computing, Jülich, 2006), pp. 85–129, also available as `http://www.fz-juelich.de/nic-series/volume31/bluegel.pdf`.

[96] *Vienna ab-initio simulation package*, URL `http://cms.mpi.univie.ac.at/vasp/`.

[97] P. Blaha, K. Schwarz, G. Madsen, D. Kvasnicka, and J. Luitz, *Wien2k*, URL `http://www.wien2k.at/`.

[98] *Quantumespresso*, URL `http://www.quantum-espresso.org/`.

[99] *Openmx*, URL `http://www.openmx-square.org/`.

[100] *Gaussian*, URL `http://www.gaussian.com/`.

[101] *Turbomole*, URL `http://www.turbomole.com/`.

[102] P. Ordejon, Comput. Mater. Sci. **12**, 157 (1998).

[103] P. Ordejón, E. Artacho, and J. Soler, Phys. Rev. B **53**, 10441 (1996).

[104] P. Ordejon, D. A. Drabold, M. P. Grumbach, and R. M. Martin, Phys. Rev. B **48**, 14646 (1993).

[105] D. Sanchez-Portal, P. Ordejon, E. Artacho, and J. M. Soler, Int. J. Quantum Chem. **65**, 453 (1997).

[106] N. Troullier and J. Martins, Phys. Rev. B **43**, 1993 (1991).

[107] N. Troullier and J. Martins, Phys. Rev. B **43**, 8861 (1991).

[108] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).

[109] L. Kleinman and D. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[110] O. Sankey and D. Niklewski, Phys. Rev. B **40**, 3979 (1989).

[111] E. Artacho, D. Sanchez-Portal, P. Ordejón, A. García, and J. Soler, Phys. Stat. Sol. b **215**, 809 (1999).

[112] P. Pulay, J. Comput. Chem. **3**, 556 (1982).

[113] A. K. Das, C. Pampuch, A. Ney, T. Hesjedal, L. Däweritz, R. Koch, and K. H. Ploog, Phys. Rev. Lett. **91**, 087203 (2003).

[114] M. Tanaka, J. P. Harbison, M. C. Park, Y. S. Park, T. Shin, and G. M. Rothberg, Appl. Phys. Lett. **65**, 1964 (1994).

[115] K. Akeura, M. Tanaka, M. Ueki, and T. Nishinaga, Appl. Phys. Lett. **67**, 3349 (1995).

[116] K. H. Ploog, Physica E **24**, 101 (2004).

[117] J. H. Song, J. J. Lee, Y. Cui, J. B. Ketterson, and S. Cho, Appl. Phys. Lett. **85**, 4079 (2004).

[118] J. H. Song, Y. Cui, J. J. Lee, Y. Kim, J. B. Ketterson, and S. Cho, J. Appl. Phys. **95**, 7288 (2004).

[119] M. J. S. P. Iikawa, F. Brasil, O. D. D. Couto, C. Adriano, C. Giles, and Däweritz, Appl. Phys. Lett. **85**, 2250 (2004).

[120] F. Iikawa, M. J. S. P. Brasil, C. Adriano, O. D. D. Couto, C. Giles, P. V. Santos, L. Däweritz, I. Rungger, and S. Sanvito, Phys. Rev. Lett. **95**, 077203 (2005).

[121] J. Mira, F. Rivadulla, J. Rivas, A. Fondado, T. Guidi, R. Caciuffo, F. Carsughi, P. G. Radaellil, and J. B. Goodenough, Phys. Rev. Lett. **90**, 097203 (2003).

[122] T. Suzuki and H. Ido, J. Phys. Soc. Jpn. **51**, 3149 (1982).

[123] R. W. De Blois and D. S. Rodbell, Phys. Rev. **130**, 1347 (1963).

[124] J. B. Goodenough and J. A. Kafalas, Phys. Rev. **157**, 389 (1967).

[125] N. Menyuk, J. A. Kafalas, K. Dwight, and J. B. Goodenough, Phys. Rev. **177**, 942 (1969).

[126] F. Groønvold, S. Snildal, and E. F. Westrum Jr., Acta Chem. Scand. **24**, 285 (1970).

[127] A. Zięba, K. Selte, A. Kjekshus, and A. F. Andresen, Acta Chem. Scand. **A32**, 173 (1978).

[128] A. Zięba, Y. Shapira, and S. Foner, Phys. Lett. **91A**, 243 (1982).

[129] F. Ishikawa, K. Koyama, K. Watanabe, and H. Wada, Jpn. J. Appl. Phys. **42**, L918 (2003).

[130] J. B. Forsyth, S. J. Pickart, and P. J. Brown, Proc. Phys. Soc. **88**, 333 (1966).

[131] R. Koch, C. Pampuch, H. Yamaguchi, A. K. Das, A. Ney, L. Däweritz, and K. H. Ploog, Phys. Rev. B **70**, 092406 (2004).

[132] B. T. M. Willis and H. P. Rooksby, Proc. Phys. Soc. (London) **B67**, 290 (1954).

[133] R. H. Wilson and J. S. Kasper, Acta Crystallogr. **17**, 95 (1964).

[134] L. Pytlik and A. Zięba, J. Magn. Magn. Mater. **51**, 199 (1985).

[135] C. Kittel, Phys. Rev. **120**, 335 (1960).

[136] D. S. Rodbell and C. P. Bean, J. Appl. Phys. **33**, 1037 (1962).

[137] K. Bärner, E. A. Zavadskii, and D. Y. Suminov, Phys. Stat. Sol. B **214**, 411 (1999).

[138] S. Haneda, N. Kazama, Y. Yamaguchi, and H. Watanabe, J. Phys. Soc. Jpn. **42**, 31 (1977).

[139] s. Haneda, N. Kazama, Y. Yamaguchi, and H. Watanabe, J. Phys. Soc. Jpn. **42**, 1212 (1977).

[140] T. Kato, K. Nagai, and T. Aisaka, J. Phys. C: Sol. Stat. Phys. **16**, 3183 (1983).

[141] I. M. Vitebskii, V. I. Kamenev, and D. A. Yablonskii, Sov. Phys. Solid State **23**, 121 (1981).

[142] S. K. Asadov, E. A. Zavadskii, V. I. Kamenev, E. P. Stefanovskii, A. L. Sukstanskii, and B. M. Todris, Phys. Solid State **42**, 1696 (2000).

[143] V. I. Val'kov and A. V. Golovchan, Low Temp. Phys. **30**, 711 (2004).

[144] K. Motizuki and K. Katoh, J. Magn. Magn. Mater. **54-57**, 1097 (1986).

[145] K. Motizuki, J. Magn. Magn. Mater. **70**, 1 (1987).

[146] R. Podloucky, J. Magn. Magn. Mater. **43**, 204 (1984).

[147] R. Podloucky, J. Phys. F: Met. Phys. **14**, L145 (1984).

[148] S. Sanvito and N. A. Hill, Phys. Rev. B **62**, 15553 (2000).

[149] P. Ravindran, A. Delin, P. James, B. Johansson, J. M. Wills, R. Ahuja, and O. Eriksson, Phys. Rev. B **59**, 15680 (1999).

[150] A. Continenza, S. Picozzi, W. T. Geng, and A. J. Freeman, Phys. Rev. B **64**, 085204 (2001).

[151] Y.-J. Zhao, W. T. Geng, A. J. Freeman, and B. Delley, Phys. Rev. B **65**, 113202 (2002).

[152] R. de Paiva, J. L. A. Alves, R. A. Nogueira, J. R. Leite, and L. M. R. Scolfaro, Braz. J. Phys. **34**, 568 (2004).

[153] A. Debernardi, M. Peressi, and A. Baldereschi, Comput. Mater. Sci. **27**, 175 (2003).

[154] A. Debernardi, M. Peressi, and A. Baldereschi, Mater. Sci. Eng. C **23**, 1059 (2003).

[155] M. Shirai and Y. Tokioka, J. Electron Spectrosc. Relat. Phenom. **88-91**, 357 (1998).

[156] K. Motizuki, K. Katoh, and A. Yanase, J. Phys. C: Sol. Stat. Phys. **19**, 495 (1986).

[157] M. K. Niranjan, B. R. Sahu, and L. Kleinman, Phys. Rev. B **70**, 180406(R) (2004).

[158] V. I. Valkov and A. V. Golovchan, Low Temp. Phys. **31**, 528 (2005).

[159] L. M. Sandratskii and E. Sasioglu, Phys. Rev. B **74**, 214422 (2006).

[160] L. M. Sandratskii, R. Singer, and E. Sasioglu, Phys. Rev. B **76**, 184406 (2007).

[161] K. Shimada, O. Rader, A. Fujimori, A. Kimura, K. Ono, N. Kamakura, A. Kakizaki, M. Tanaka, and M. Shirai, J. Electron Spectrosc. Relat. Phenom. **88-91**, 207 (1998).

[162] T. Asada and K. Terakura, Phys. Rev. B **47**, 15992 (1993).

[163] R. S. Mulliken, J. Chem. Phys. **23**, 1833 (1955).

[164] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders College Publishing, USA, 1976).

[165] A. I. Liechtensteina, M. I. Katsnelson, V. P. Antropov, and V. A. Gubanov, J. Magn. Magn. Mater. **67**, 65 (1986).

[166] M. Pajda, J. Kudrnovsky, I. Turek, V. Drchal, and P. Bruno, Phys. Rev. B **64**, 174402 (2001).

[167] V. M. Kaganer, B. Jenichen, F. Schippan, W. Braun, L. Däweritz, and K. H. Ploog, Phys. Rev. Lett. **85**, 341 (2000).

[168] T. Plake, M. Ramsteiner, V. M. Kaganer, B. Jenichen, M. Kästner, L. Däweritz, and K. H. Ploog, Appl. Phys. Lett. **80**, 2523 (2002).

[169] M. Kästner, C. Herman, L. Däweritz, and K. H. Ploog, J. Appl. Phys. **92**, 5711 (2002).

[170] F. Iikawa, M. J. S. P. Brasil, O. D. D. Couto, C. Adriano, C. Giles, and L. Däweritz, Appl. Phys. Lett. **85**, 2250 (2004).

[171] F. Schippan, A. Trampert, L. Däweritz, and K. H. Ploog, J. Vac. Sci. Tecnol. **B17**, 1716 (1999).

[172] M. P. F. Godoy, M. K. K. Nakaema, F. Iikawa, W. Carvalho Jr., E. Ribeiro, and A. L. Gobby, Rev. Sci. Instrum. **75**, 1947 (2004).

[173] I. Rungger and S. Sanvito, Phys. Rev. B **74**, 024429 (2006).

[174] M. Dörfler and K. Bärner, phys. stat. sol. (a) **17**, 141 (1973).

[175] C. Guillaud, J. phys. radium **12**, 223 (1951).

[176] Z. S. Basinski, R. O. Kornelsen, and W. B. Pearson, Trans. Ind. Inst. Met. **13**, 143 (1960).

[177] K. Selte, A. Kjekshus, and A. F. Andresen, Acta Chem. Scand. **A28**, 61 (1974).

[178] A. R. Rocha, *Theoretical and computational aspects of electronic transport at the nanoscale* (Ph.D. Thesis, Trinity College Dublin, 2006), also available as `http://www.smeagol.tcd.ie/PAPER/AlexThesis.html`.

[179] S. Datta, Superlattices and Microstructures **28**, 253 (2000).

[180] S. Datta, Electron Devices Meeting, 2002. IEDM '02. Digest. International p. 703 (2002).

[181] F. Zahid, M. Paulsson, and S. Datta, Advanced Semiconductors and Organic Nano-Techniques, edited by H. Morkoc, Academic Press (2003).

[182] A. P. Jauho, *Progress in nonequilibrium Green's functions II* (World Scientific Publishing Co., Singapore, 2003).

[183] H. Haug and A. P. Jauho, Quantum Kinetics in Transport and Optics of Semiconductors, Springer, Berlin (1996).

[184] A. P. Jauho (2006), URL `http://www.nanohub.org/resources/1877/`.

[185] P. C. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959).

[186] G. Stefanucci and C.-O. Almbladh, Phys. Rev. B **69**, 195318 (2004).

[187] G. Stefanucci, Phys. Rev. B **75**, 195115 (2007).

[188] G. Stefanucci, S. Kurth, E. K. U. Gross, and A. Rubio, cond-mat/0607333 (2006).

[189] L. Keldysh, Sov. Phys. JETP **20**, 1018 (1965).

[190] S. Datta, Nanotechnology **15**, S433 (2004).

[191] Y. Xue, S. Datta, and M. A. Ratner, Chem. Phys. **281**, 151 (2002).

[192] M. P. Anantram, M. S. Lundstrom, and D. E. Nikonov, cond-mat/0610247 (2007).

[193] M. Paulsson, cond-mat/0210519v2 (2006).

[194] E. Economou, *Green's Functions in Quantum Physics* (Spinger-Verlag, Berlin, 1995).

[195] A. R. Williams, P. J. Feibelman, and N. D. Lang, Phys. Rev. B **26**, 5433 (1982).

[196] M. Brooks, *The matrix reference manual* (2005), URL `http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/intro.html`.

[197] P. S. Krstić, X.-G. Zhang, and W. H. Butler, Phys. Rev. B **66**, 205319 (2002).

[198] *Smeagolwiki*, URL `http://smeagol.tchpc.tcd.ie/wiki`.

[199] S. Ke, H. Baranger, and W. Yang, Phys. Rev. B **70**, 085410 (2004).

[200] A. M. Zagoskin, *Quantum Theory of Many-Body Systems* (Springer, New York, 1998).

[201] M. Koentopp, K. Burke, and F. Evers, Phys. Rev. B **73**, 121403(R) (2006).

[202] C. Toher, A. Filippetti, S. Sanvito, and K. Burke, Phys. Rev. Lett. **95**, 146402 (2005).

[203] F. Elste and C. Timm, Phys. Rev. B **71**, 155403 (2005).

[204] S. Braig and P. W. Brouwer, Phys. Rev. B **71**, 195324 (2005).

[205] E. Bonet, M. M. Deshmukh, and D. C. Ralph, Phys. Rev. B **65**, 045317 (2002).

[206] B. Muralidharan, A. W. Ghosh, and S. Datta, Molecular Simlulation **32**, 751 (2006).

[207] B. Muralidharan, A. W. Ghosh, and S. Datta, Phys. Rev. B **73**, 155410 (2006).

[208] E. Khosravi, S. Kurth, G. Stefanucci, and E. Gross, Appl. Phys. A **93**, 355 (2008).

[209] C. G. Sanchez, M. Stamenova, S. Sanvito, D. R. Bowler, and A. P. Horsfield, J. Chem. Phys. **124**, 214708 (2006).

[210] B. Wenzien, J. Kudrnovský, V. Drchal, and M. Šob, J. Phys.: Condens. Matter **1**, 9893 (1989).

[211] J. Kudrnovský, I. Turek, V. Drchal, P. Weinberger, N. E. Christensen, and S. K. Bose, Phys. Rev. B **46**, 4222 (1992).

[212] M. P. López Sancho, J. M. López Sancho, and J. Rubio, J. Phys. F: Met. Phys. **14**, 1205 (1984).

[213] M. Buongiorno Nardelli, Phys. Rev. B **60**, 7828 (1999).

[214] S. Sanvito, C. Lambert, J. Jefferson, and A. Bratkovsky, Phys. Rev. B **59**, 11936 (1999).

[215] A. Umerski, Phys. Rev. B **55**, 5266 (1997).

[216] T. Ando, Phys. Rev. B **44**, 8017 (1991).

[217] T. Tao and V. Vu, arXiv:math/0703307v1 (2007).

[218] N. J. Higham, *Accuracy and Stability of Numerical Algorithms* (SIAM, 2002).

[219] J. K. Tomfohr and O. F. Sankey, Phys. Rev. B **65**, 245105 (2002).

[220] F. Tisseur and K. Meerbergen, SIAM Review **43**, 235 (2001).

[221] C.-H. Guo, Linear Algebra Appl. **385**, 391 (2004).

[222] M. P. López Sancho, J. M. López Sancho, and J. Rubio, J. Phys. F: Met. Phys. **15**, 851 (1985).

[223] R. E. Allen, Phys. Rev. B **19**, 917 (1979).

[224] Y.-C. Chang and J. N. Schulman, Phys. Rev. B **25**, 3975 (1982).

[225] A. Sankar, D. A. Spielman, and S.-H. Teng, SIAM J. Matrix Anal. Appl. **28**, 446 (2006).

[226] L. G. Bulusheva, A. V. Okotrub, D. A. Romanov, and D. Tomanek, J. Phys. Chem. A **102**, 975 (1998).

[227] Y. Xue and M. A. Ratner, Phys. Rev. B **68**, 115406 (2003).

[228] Y. Xue and M. A. Ratner, Phys. Rev. B **68**, 115407 (2003).

[229] L. N. Trefethen and D. Bau, *Numerical linear algebra* (SIAM, 1997).

[230] J. J. Du Croz and N. J. Higham, IMA J. Num. Anal. **12**, 1 (1992).

[231] G. F. Koster and J. C. Slater, Phys. Rev. **95**, 1167 (1954).

[232] G. F. Koster and J. C. Slater, Phys. Rev. **96**, 1208 (1954).

[233] M. Di Ventra, C. Berthod, and N. Binggeli, Phys. Rev. B **62**, R10622 (2000).

[234] R. A. Brown, Phys. Rev. **156**, 889 (1967).

[235] J. Callaway, Phys. Rev. **154**, 515 (1967).

[236] J. Bernholc, N. O. Lipari, and S. T. Pantelides, Phys. Rev. B **21**, 3545 (1980).

[237] J. Bernholc and S. T. Pantelides, Phys. Rev. B **18**, 1780 (1978).

[238] G. A. Baraff, J. A. Appelbaum, and D. R. Hamann, Phys. Rev. Lett. **38**, 237 (1977).

[239] W. E. Pickett, S. G. Louie, and M. L. Cohen, Phys. Rev. Lett. **39**, 109 (1977).

[240] W. E. Pickett and M. L. Cohen, Phys. Rev. B **18**, 939 (1978).

[241] J. Pollmann and S. T. Pantelides, Phys. Rev. B **18**, 5524 (1978).

[242] O. Wunnicke, N. Papanikolaou, R. Zeller, P. H. Dederichs, V. Drchal, and J. Kudrnovský, Phys. Rev. B **65**, 064425 (2002).

[243] D. Wortmann, *An Embedding Green Function Approach for Electron Transport through Interfaces* (Ph.D. Thesis, Forschungszentrum Jülich, 2003), also available as `http://darwin.bth.rwth-aachen.de/opus3/volltexte/2003/674/pdf/03_204.pdf`.

[244] P. Pomorski, L. Pastewka, C. Roland, H. Guo, and J. Wang, Phys. Rev. B **69**, 115418 (2004).

[245] R. Li, J. Zhang, S. Hou, Z. Qian, . Shen, X. Zhao, and Z. Xue, Chem. Phys. **336**, 127 (2007).

[246] P. Mavropoulos, M. Lezaic, and S. Blügel, Phys. Rev. B **72**, 174428 (2005).

[247] D. Wortmann, H. Ishida, and S. Blügel, Phys. Rev. B **72**, 235113 (2005).

[248] D. Wortmann, H. Ishida, and T. Ohwaki, Phys. Rev. B **70**, 085409 (2004).

[249] D. Wortmann, H. Ishida, and S. Blügel, Phys. Rev. B **66**, 075113 (2002).

[250] E. Khosravi, G. Stefanucci, S. Kurth, and E. Gross, cond-mat/08030914 (2008).

[251] A. Dhar and D. Sen, Phys. Rev. B **73**, 085119 (2006).

[252] O. Baumgartner, M. Karner, S. Holzer, M. Pourfath, T. Grasser, and H. Kosina, NSTI Nanotech Technical Proceedings **3**, 145 (2007), also available as `http://www.iue.tuwien.ac.at/dl/189/354/656/VISTA-Status-Report-2007-12.pdf`

[253] D. C. Sorensen, in *Parallel Number Alogirthm Proceedings ICASE/LaRC Workshop*, edited by D. Keyes, A. Sameh, and V. Venkatakrishnan (Kluwer Academic Publishers, 1995).

[254] R. B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK Users Guide: Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods* (SIAM, 1998).

[255] C. Toher and S. Sanvito, Phys. Rev. B **77**, 155402 (2008).

[256] A. Van Orden and R. J. Saykally, Chem. Rev. **98**, 2313 (1998).

[257] *Computational chemistry comparison and benchmark database*, URL `http://cccbdb.nist.gov/`.

[258] H. B. Heersche, Z. de Groot, J. A. Folk, H. S. J. van der Zant, C. Romeike, M. R. Wegewijs, L. Zobbi, D. Barreca, E. Tondello, and A. Cornia, Phys. Rev. Lett. **96**, 206801 (2006).

[259] S. C. D. Pemmaraju, *Implementation of Self Interaction Free Density Functional methods and applications to Dilute Magnetic Semiconductors* (Ph.D. Thesis, Trinity College Dublin, 2007).

[260] K. Park, M. R. Pederson, and C. S. Hellberg, Phys. Rev. B **69**, 014416 (2004).

[261] M. R. Wegewijs, C. Romeike, H. Schoeller, and W. Hofstetter, New J. Phys. **9**, 344 (2007).

[262] U. del Pennino, V. De Renzi, R. Biagi, V. Corradini, L. Zobbi, A. Cornia, D. Gatteschi, F. Bondino, E. Magnano, M. Zangrando, et al., Surf. Sci. **600**, 4185 (2006).

[263] D. W. Boukhvalov, A. I. Lichtenstein, V. V. Dobrovitski, M. I. Katsnelson, B. N. Harmon, V. V. Mazurenko, and V. I. Anisimov, Phys. Rev. B **65**, 184435 (2002).

[264] S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, and K. Ando, Nature Materials **3**, 868 (2004).

[265] S. S. P. Parkin, C. Kaiser, A. Panchula, P. M. Rice, B. Hughes, M. Samant, and S.-H. Yang, Nature Materials **3**, 862 (2004).

[266] C. Tusche, H. L. Meyerheim, N. Jedrecy, G. Renaud, A. Ernst, J. Henk, P. Bruno, and J. Kirschner, Phys. Rev. Lett. **95**, 176101 (2005).

[267] H. L. Meyerheim, R. Popescu, N. Jedrecy, M. Vedpathak, M. Sauvage-Simkin, R. Pinchaux, B. Heinrich, and J. Kirschner, Phys. Rev. B **65**, 144433 (2002).

[268] H. L. Meyerheim, R. Popescu, J. Kirschner, N. Jedrecy, M. Sauvage-Simkin, B. Heinrich, and R. Pinchaux, Phys. Rev. Lett. **87**, 076102 (2001).

[269] D. Wortmann, G. Bihlmayer, and S. Blügel, J. Phys.: Condens. Matter **16**, S5819 (2004).

[270] H. Oh, S. B. Lee, J. Seo, H. G. Min, and J.-S. Kim, Appl. Phys. Lett. **82**, 361 (2003).

[271] C. Tusche, H. L. Meyerheim, N. Jedrecy, G. Renaud, and J. Kirschner, Phys. Rev. B **74**, 195422 (2006).

[272] Y. S. Dedkov1, M. Fonin, U. Rüdiger, and G. Güntherodt, Appl. Phys. A **82**, 489 (2006).

[273] X.-G. Zhang, W. H. Butler, and A. Bandyopadhyay, Phys. Rev. B **68**, 092402 (2003).

[274] C. Heiliger, P. Zahn, B. Y. Yavorsky, and I. Mertig, Phys. Rev. B **72**, 180406(R) (2005).

[275] P. Bose, A. Ernst, I. Mertig, and J. Henk, Phys. Rev. B **78**, 092403 (2008).

[276] J.-B. Laloë, A. Ionescu, S. Easton, N.-J. Steinke, T. J. Hayward, H. Kurebayashi, J. A. C. Bland, T. R. Charlton, R. M. Dalgliesh, and S. Langridge, Appl. Phys. Lett. **93**, 012505 (2008).

[277] M. Sicot, S. Andrieu, C. Tiusan, F. Montaigne, and F. Bertran, J. Appl. Phys. **99**, 08D301 (2006).

[278] M. Sicot, S. Andrieu, P. Turban, Y. Fagot-Revurat, H. Cercellier, A. Tagliaferri, C. De Nadai, N. B. Brookes, F. Bertran, and F. Fortuna, Phys. Rev. B **68**, 184406 (2003).

[279] M. Sicot, S. Andrieu, F. Bertran, and F. Fortuna, Phys. Rev. B **72**, 144414 (2005).

[280] P. Luches, S. Benedetti, M. Liberati, F. Boscherini, I. I. Pronin, and S. Valeri, Surf. Sci. **583**, 191 (2005).

[281] P. Luches, P. Torelli, S. Benedetti, E. Ferramola, R. Gotter, and S. Valeri, Surf. Sci. **601**, 3902 (2007).

[282] M. Zajac, K. Freindl, K. Matlak, M. Slezak, T. T. Slezak, N. Spiridis, and J. Korecki, Surf. Sci. **601**, 4305 (2007).

[283] L. Plucinski, Y. Zhao, B. Sinkovic, and E. Vescovo, Phys. Rev. B **75**, 214411 (2007).

[284] B. D. Yu and J.-S. Kim, Phys. Rev. B **73**, 125408 (2006).

[285] M. Müller, F. Matthes, and C. M. Schneider, Europhysics Lett. **80**, 17007 (2007).

[286] F. J. Palomaresa, C. Munuera, C. Martinez Boubeta, and A. Cebollada, J. Appl. Phys. **97**, 036104 (2005).

[287] N. Baadji, *Ab-initio study of magnetism and magneto-transport in hybrid systems (Iron/semi-conductor/Iron)* (Ph.D. Thesis,Institut de Physique et de Chimie des Materiaux de Strasbourg (IPCMS), Louis Pasteur University, 2006).

[288] P. G. Mather, J. C. Read, and R. A. Buhrman, Phys. Rev. B **73**, 205412 (2006).

[289] J. Mathon and A. Umerski, Phys. Rev. B **63**, 220403(R) (2001).

[290] K. D. Belashchenko, J. P. Velev, and E. Y. Tsymbal, Phys. Rev. B **72**, 140404(R) (2005).

[291] C. Heiliger, P. Zahn, B. Y. Yavorsky, and I. Mertig, Phys. Rev. B **77**, 224407 (2008).

[292] Y. Ando, T. Miyakoshi, M. Oogane, T. Miyazaki, H. Kubota, and S. Ando, K. Yuasa, Appl. Phys. Lett. **87**, 142502 (2005).

[293] P. J. Zermatten, G. Gaudin, G. Maris, M. Miron, A. Schuhl, C. Tiusan, and M. Greullet, F. Hehn, Phys. Rev. B **78**, 033301 (2008).

[294] H. Itoh, J. Phys. D: Appl. Phys. **40**, 1228 (2007).

[295] V. M. Xu, P. X. andKarpan, K. Xia, M. Zwierzycki, I. Marushchenko, , and P. J. Kelly, Phys. Rev. B **73**, 180402(R) (2006).

[296] K. Xia, M. Zwierzycki, M. Talanana, P. J. Kelly, and G. E. W. Bauer, Phys. Rev. B **73**, 064420 (2006).

[297] V. Drchal, J. Kudrnovský, P. Bruno, P. H. Dederichs, I. Turek, and P. Weinberger, Phys. Rev. B **65**, 214414 (2002).

[298] G. X. Miao, Y. J. Park, J. S. Moodera, M. Seibt, G. Eilers, and M. Münzenberg, Phys. Rev. Lett. **100**, 246803 (2008).

[299] J. Ozeki, H. Itoh, and J. Inoue, J. Magn. Magn. Mater. **310**, e644 (2007).

[300] P. Stamenov, *Metals, Semimetals and Semiconductors for Spinelectronics* (Ph.D. Thesis, Trinity College Dublin, 2007), also available as `http://www.maths.tcd.ie/ stamenov/download/PhDThesis.pdf`.

[301] T. Nozaki, A. Hirohata, N. Tezuka, S. Sugimoto, and K. Inomata, Appl. Phys. Lett. **86**, 082501 (2005).

[302] N. Nozaki, T. Tezuka and K. Inomata, Phys. Rev. Lett. **96**, 027208 (2006).

[303] Y. Wang, Z.-Y. Lu, X.-G. Zhang, and X. F. Han, Phys. Rev. Lett. **97**, 087210 (2006).

[304] J. Peralta-Ramos, A. M. Llois, I. Rungger, and S. Sanvito, Phys. Rev. B **78**, 024430 (2008).

[305] P. K. de Boer and R. A. de Groot, J. Phys.: Condens. Matter **10**, 10241 (1998).

[306] Y. Fei, Am. Mineral. **272**, 84 (1999).

[307] D. M. Roessler and W. C. Walker, Phys. Rev. **159**, 733 (1967).

[308] H. Baltache, R. Khenata, M. Sahnoun, M. Driz, B. Abbar, and B. Brouhafs, Physica B **344**, 334 (2004).

[309] T. Kotani, Phys. Rev. B **50**, 14816 (1994).

[310] J. Izquierdo, A. Vega, L. C. Balbas, D. Sanchez-Portal, J. Junquera, E. Artacho, J. M. Soler, and P. Ordejon, Phys. Rev. B **61**, 13639 (2000).

[311] *Webelements: the periodic table on the web*, URL `http://www.webelements.com`.

[312] A. M. Turner, A. W. Donoho, and J. L. Erskine, Phys. Rev. B **29**, 2986 (1984).

[313] J.-H. Cho and M. Scheffler, Phys. Rev. B **53**, 10685 (1996).

[314] T. Asada and K. Terakura, Phys. Rev. B **46**, 13599 (1992).

[315] I. Rungger, A. R. Rocha, O. Mryasov, O. Heinonen, and S. Sanvito, J. Magn. Magn. Mater. **316**, 481 (2007).

[316] J. Callaway and C. S. Wang, Phys. Rev. B **16**, 2095 (1977).

[317] D. Waldron, L. Liu, and H. Guo, Nanotechnology **18**, 424026 (2007).

[318] J. Bardeen, Phys. Rev. Lett. **6**, 57 (1961).

[319] J. A. Appelbaum and W. F. Brinkman, Phys. Rev. **186**, 464 (1969).

[320] A. R. Rocha and S. Sanvito, Phys. Rev. B **70**, 094406 (2004).

[321] D. Waldron, V. Timoshevskii, Y. Hu, K. Xia, and H. Guo, Phys. Rev. Lett. **97**, 226802 (2006).

[322] T. Raza and H. Raza, arXiv:cond-mat/0804.2667 (2008).

[323] P. Mavropoulos, N. Papanikolaou, and P. H. Dederichs, Phys. Rev. Lett. **85**, 1088 (2000).

[324] D. Wortmann, H. Ishida, and S. Blügel, Phys. Rev. B **65**, 165103 (2002).

[325] M. Sterrer, M. Heyde, M. Novicki, N. Nilius, T. Risse, H.-P. Rust, G. Pacchioni, and H.-J. Freund, J. Phys. Chem. B (Letter) **110**, 46 (2006).

[326] A. Gibson, R. Haydock, and J. P. LaFemina, Phys. Rev. B **50**, 2582 (1994).

[327] B. M. Klein, W. E. Pickett, L. L. Boyer, and R. Zeller, Phys. Rev. B **35**, 5802 (1987).

[328] G. Pacchioni, J. Chem. Phys. **128**, 182505 (128).

[329] J. P. Velev, K. D. Belashchenko, S. S. Jaswal, and E. Y. Tsymbal, Appl. Phys. Lett. **90**, 072502 (2007).

[330] E. Bauer and R. Belkhou, J. Vac. Sci. Technol. B **25**, 1470 (2007).

[331] Y. Takagaki, L. Däweritz, and K. H. Ploog, Phys. Rev. B **75**, 035213 (2007).

[332] S. Kümmel and J. P. Perdew, Phys. Rev. B **68**, 035103 (2003).

[333] J. B. Krieger, Y. Li, and G. J. Iafrate, Phys. Rev. A **45**, 101 (1992).

[334] O. V. Gritsenko and E. J. Baerends, Phys. Rev. A **54**, 1957 (1996).

[335] C. D. Pemmaraju, S. Sanvito, and K. Burke, Phys. Rev. B **77**, 121204(R) (2008).

[336] N. Agraït, C. Untiedt, G. Rubio-Bollinger, and S. Vieira, Phys. Rev. Lett. **88**, 216803 (2002).

[337] R. H. M. Smit, Y. Noat, C. Untiedt, N. D. Lang, M. C. van Hemert, and J. M. van Ruitenbeek, Nature **419**, 906 (2002).

[338] M. Galperin and M. A. Ratner, J. Chem. Phys. **121**, 11965 (2004).

[339] M. Wimmer and K. Richter, cond-mat/0806.2739 (2008).

[340] M. B. Nardelli, J.-L. Fattebert, and J. Bernholc, Phys. Rev. B **64**, 245423 (2001).

[341] F. R. Krajewski and M. Parrinello, Phys. Rev. B **74**, 125107 (2006).

[342] F. Freimuth, Y. Mokrousov, D. Wortmann, S. Heinze, and S. Blügel, Phys. Rev. B **78**, 035120 (2008).

[343] T. Miyazaki, D. R. Bowler, R. Choudhury, and M. J. Gillan, The Journal of Chemical Physics **121**, 6186 (2004).

[344] M. I. Friswell, AIAA Journal **38**, 2355 (2000).

Appendix A

Heisenberg model for a periodic system

Here we present the approximations involved in mapping DFT total energies to a model Hamiltonian of the form of Eq. (3.3). The most general form of model Hamiltonian able to fit all possible energies for a system of N_s collinear magnetic moments is

$$E_{s_1,s_2,\ldots} = E_0 - \sum_{\nu=1}^{N_s} \frac{1}{\nu!} \sum_{j_1} s_{j_1} \sum_{j_2} s_{j_2} \cdots \sum_{j_\nu} s_{j_\nu} \; J_{j_1,j_2,\ldots,j_\nu} \; , \tag{A.1}$$

where s_j is the moment at site j (normalized to one), and the Js are coupling parameters. The $J_{ij,\ldots}$ parameters are symmetric under permutations of the indices, and each of the $J_{ij,\ldots}$ is zero if any of the indices i and j are equal. The number of independent parameters is therefore 2^{N_s}. Since time reversal symmetry makes the system invariant under a global spin rotation only Js with an even number of indices are not zero, and equation (A.1) reduces to

$$\begin{aligned} E_{s_1,s_2,\ldots} = & E_0 - \frac{1}{2} \sum_{i,j} s_i s_j \; J_{ij} \; - \frac{1}{4!} \sum_{i,j,k,l} s_i s_j s_k s_l \; J_{ijkl} \; - \\ & - \frac{1}{6!} \sum_{i,j,k,l,m,n} s_i s_j s_k s_l s_m s_n \; J_{ijklmn} \; - \ldots \end{aligned} \tag{A.2}$$

E_0 is a constant that can be associated to the energy of a paramagnetic phase. In fact if all the local magnetic moments are randomly aligned the contribution coming from $\sum_{i,j} \mathbf{s}_j \mathbf{s}_j \; J_{ij}$ vanishes. We now neglect 4-moment coupling constants J_{ijkl} and higher, thus reducing the model to an Ising-type. This approximation neglects the dependence of the moment on each Mn atom as well as the small induced magnetic moment over the As atoms on the orientation of the moments of the surrounding Mn atoms. Furthermore it is assumed that the coupling constants are independent from the angle between the magnetic moments, so that the fit can be extrapolated to a

vector magnetic moment $\mathbf{s}_i$ to give a Heisenberg type energy

$$E_{\mathbf{s}_1,\mathbf{s}_2,\ldots} = E_0 - \frac{1}{2}\sum_{i,j} \mathbf{s}_j \mathbf{s}_j \; J_{ij}. \tag{A.3}$$

Appendix B

Special block matrix inversions

We define the matrices A, B, C, D, E, F, G in such a way that the matrices in each of the sets $\{A, B\}$, $\{C, D, E\}$, and $\{F, G\}$ have the same number of rows, and the number of columns is equal for the matrices in the sets $\{A, C\}$, $\{B, D, F\}$ and $\{E, G\}$. We can then build the block matrix M_3:

$$M_3 = \begin{pmatrix} A & B & 0 \\ C & D & E \\ 0 & F & G \end{pmatrix}, \tag{B.1}$$

where the 0s represent zero block matrices of appropriate size. The inverse of such a matrix can be calculated explicitly as

$$M_3^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BL_3CA^{-1} & -A^{-1}BL_3 & A^{-1}BL_3EG^{-1} \\ -L_3CA^{-1} & L_3 & -L_3EG^{-1} \\ G^{-1}FL_3CA^{-1} & -G^{-1}FL_3 & G^{-1} + G^{-1}FL_3EG^{-1} \end{pmatrix}, \tag{B.2}$$

where

$$L_3 = \left(D - CA^{-1}B - EG^{-1}F\right)^{-1}. \tag{B.3}$$

Similarly for a 2×2 block matrix M_2 given by

$$M_2 = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{B.4}$$

the inverse is

$$M_2^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BL_{2,\mathrm{R}}CA^{-1} & -A^{-1}BL_{2,\mathrm{R}} \\ -L_{2,\mathrm{R}}CA^{-1} & L_{2,\mathrm{R}} \end{pmatrix}, \tag{B.5}$$

where

$$L_{2,\mathrm{R}} = \left(D - CA^{-1}B\right)^{-1}. \tag{B.6}$$

The inverse of M_2 can also be written in an equivalent way as

$$M_2^{-1} = \begin{pmatrix} L_{2,\mathrm{L}} & -L_{2,\mathrm{L}} B D^{-1} \\ -D^{-1} C L_{2,\mathrm{L}} & D^{-1} + D^{-1} C L_{2,\mathrm{L}} B D^{-1} \end{pmatrix}, \tag{B.7}$$

where

$$L_{2,\mathrm{L}} = \left(A - B D^{-1} C\right)^{-1}. \tag{B.8}$$

Appendix C

Additional relations for the calculation of the self-energies

C.1 Verification of the recursive relation for the SGF

Here we demonstrate that Σ_{L} calculated using Eq. (5.48) indeed fulfills the recursive relation for g_{L} of Eq. (5.50). Insert Eqs. (5.47) and (5.48) into Eq. (5.50) and take the inverse to obtain

$$K_0 + K_{-1}\bar{T}_{\mathrm{R}} + K_1\bar{T}_{\mathrm{R}}^{-1} = 0. \tag{C.1}$$

Using the definition of the matrix $\bar{T}_{\mathrm{R}}$ [Eq. (5.24)] we write

$$\sum_{n=1}^{N}\left(K_0 + K_{-1}e^{-i\bar{k}_n} + K_1 e^{i\bar{k}_n}\right)\bar{\phi}_{\mathrm{R},n}\tilde{\bar{\phi}}^{\dagger}_{\mathrm{R},n} = 0. \tag{C.2}$$

This equation corresponds to the defining equation for the $\bar{\phi}_{\mathrm{R},n}$ and is therefore fulfilled by definition. The same is therefore true for Eq. (5.50). Eq. (5.51) for g_{R} can be obtained similarly.

C.2 Regularization of K_1 and K_{-1} for $K_1^{\dagger} \neq K_{-1}$

In Sec. 5.3.1 we assume that $K_1 = K_{-1}^{\dagger}$ in order to write the transformed matrices K_1' and K_{-1}' in form of Eq. (5.55). If $K_1^{\dagger} \neq K_{-1}$ the same can be done by performing a generalized SVD of the Hamiltonian and overlap matrices as described in Ref. [68]. Here we present a different approach, based on two standard SVD transformations, one for K_1 and one for $K_{-1}^{\dagger}$

$$\begin{aligned} K_1 &= U_1 S_{\mathrm{a}} V_1^{\dagger}, \\ K_{-1}^{\dagger} &= U_{-1} S_{\mathrm{b}} V_{-1}^{\dagger}. \end{aligned} \tag{C.3}$$

Here U_1, U_{-1}, V_1 and V_{-1} are unitary matrices, and S_a and S_b are diagonal matrices with the singular values on the diagonal. In general there are M_1 singular values of K_1 smaller than $\delta_{\mathrm{SVD}} s_{\mathrm{a,max}}$, and M_{-1} singular values of K_{-1} smaller than $\delta_{\mathrm{SVD}} s_{\mathrm{b,max}}$, with $s_{\mathrm{a,max}}$ and $s_{\mathrm{b,max}}$ being respectively the largest singular values of K_1 and K_{-1}. If $M = \min(M_1, M_{-1})$, we obtain $K_{1,\mathrm{SVD}}$ by setting the smallest M singular values of K_1 to zero. In the same way we obtain $K_{-1,\mathrm{SVD}}$ by setting the smallest M singular values of K_{-1} to zero. A transformation

$$\begin{aligned} K_1' &= U_1^\dagger K_{1,\mathrm{SVD}} U_{-1}, \\ K_{-1}' &= U_1^\dagger K_{-1,\mathrm{SVD}} U_{-1} \end{aligned} \tag{C.4}$$

brings both K_1' and K_{-1}' to the form of Eq. (5.55). All the results of Sec. 5.3.1 are then valid also for $K_1^\dagger \neq K_{-1}$.

If the Hamiltonian and overlap matrices are real and Hermitian, but the energy is complex, then $K_1 = K_{-1}^{\dagger *}$. By using Eq. (C.3), and the fact that S_a and S_b are real, we obtain $S_\mathrm{a} = S_\mathrm{b}$, so that $M = M_1 = M_{-1}$. If the Hamiltonian and overlap matrices are Hermitian but not real, then in general $S_\mathrm{a} \neq S_\mathrm{b}$. However in all the calculations performed the difference between S_a and S_b was very small, so that in practice we always had $M_1 = M_2$.

In Sec. 5.3.2 we limit the singular values of K_1 from below without reducing the size of the system. If $K_1^\dagger \neq K_{-1}$ we simply apply the transformations described in Sec. 5.3.2 to both K_1 and K_{-1} independently.

C.3 Quadratic eigenvalue problem for the right-going states

We find that in the solution of Eq. (5.6) the numerical accuracy for those eigenvalues with $|e^{ik_n}| > 1$ [$\mathrm{Im}(k_n) < 0$] is better than for those with $|e^{ik_n}| < 1$ [$\mathrm{Im}(k_n) > 0$], especially when $|k_n| \gg 1$. For Σ_L we only need the left-going states, for which Eq. (5.6) gives the better accuracy. For Σ_R the right-going states are needed. In this case, in order to increase the accuracy for the right decaying states [$\mathrm{Im}(k_n) > 0$], instead of Eq. (5.6) we solve the equivalent equation

$$\begin{pmatrix} -K_0 & -K_1 \\ I_N & 0_N \end{pmatrix} \Phi_{\mathrm{R},n} = e^{-ik_n} \begin{pmatrix} K_{-1} & 0_N \\ 0_N & I_N \end{pmatrix} \Phi_{\mathrm{R},n}, \tag{C.5}$$

with

$$\Phi_{\mathrm{R},n} = \begin{pmatrix} e^{-i\frac{k_n}{2}} \\ e^{i\frac{k_n}{2}} \end{pmatrix} \frac{\phi_{\mathrm{R},n}}{\sqrt{v_n}}. \tag{C.6}$$

The eigenvalues of the states with $\mathrm{Im}(k_n) > 0$ now have an absolute value larger than one and therefore a higher accuracy.

C.4 Derivative of the self-energies with respect to the energy

Here we present a way to obtain the derivative of the SE with respect to energy, which is needed for example to evaluate Eq. (6.63). We only present the result for the right SE, the derivative of the left SE can be calculated analogously. Using the result for the right SE in Eq. (5.49), the derivative with respect to energy E is

$$\frac{\partial \Sigma_R}{\partial_E} = -S_1 T_R + K_1 \frac{\partial T_R}{\partial_E}, \tag{C.7}$$

where we have also used the definition $K_1 = H_1 + ES_1$. If we now insert the definition of T_R [Eq. (5.23)], and by using the definition of the group velocity v_g [Eq. (5.10)], we obtain

$$\frac{\partial \Sigma_R}{\partial_E} = -S_1 T_R + K_1 \sum_{n=1}^{N} \phi_{\mathrm{R},n} \frac{i e^{ik_n}}{v_n} \tilde{\phi}^\dagger_{\mathrm{R},n} + e^{ik_n} \frac{\partial \phi_{\mathrm{R},n} \tilde{\phi}^\dagger_{\mathrm{R},n}}{\partial E}. \tag{C.8}$$

All quantities in this equation are known, except the derivative of the WFs with respect to energy

$$\frac{\partial \phi_{\mathrm{R},n} \tilde{\phi}^\dagger_{\mathrm{R},n}}{\partial E} = \frac{\partial \phi_{\mathrm{R},n}}{\partial E} \tilde{\phi}^\dagger_{\mathrm{R},n} + \phi_{\mathrm{R},n} \frac{\partial \tilde{\phi}^\dagger_{\mathrm{R},n}}{\partial E}. \tag{C.9}$$

The first step now is to express the derivative of the duals as function of the WFs. If we derive the completeness relation for the WF and duals

$$\sum_{m=1}^{N} \phi_{\mathrm{R},m} \tilde{\phi}^\dagger_{\mathrm{R},m} = \mathbb{1}_N \tag{C.10}$$

with respect to energy, and then multiply the result from the left with $\tilde{\phi}^\dagger_n$, we obtain

$$\frac{\partial \tilde{\phi}^\dagger_{\mathrm{R},n}}{\partial E} = -\tilde{\phi}^\dagger_{\mathrm{R},n} \sum_{m=1}^{N} \frac{\partial \phi_{\mathrm{R},m}}{\partial E} \tilde{\phi}^\dagger_{\mathrm{R},m}. \tag{C.11}$$

This shows that the derivative of one of the duals depends on the derivatives of all the WFs. With this result, Eq. (C.9) becomes

$$\frac{\partial \phi_{\mathrm{R},n} \tilde{\phi}^\dagger_{\mathrm{R},n}}{\partial E} = \frac{\partial \phi_{\mathrm{R},n}}{\partial E} \tilde{\phi}^\dagger_{\mathrm{R},n} - \phi_{\mathrm{R},n} \tilde{\phi}^\dagger_{\mathrm{R},n} \sum_{m=1}^{N} \frac{\partial \phi_{\mathrm{R},m}}{\partial E} \tilde{\phi}^\dagger_{\mathrm{R},m}. \tag{C.12}$$

The last information missing is therefore the value of $\frac{\partial \phi_{\mathrm{R},m}}{\partial E}$. This can be obtained from the defining relation for the $\phi_{\mathrm{R},m}$ [Eq. (5.4)]

$$M_n \phi_n = 0, \tag{C.13}$$

with the newly introduced matrix M_n defined by

$$M_n = K_0 + K_1 e^{ik_n} + K_{-1} e^{-ik_n}. \tag{C.14}$$

Taking the derivative of Eq. (C.13) with respect to E, we obtain

$$M_n \frac{\partial \phi_n}{\partial E} = -\frac{\partial M_n}{\partial E} \phi_n, \tag{C.15}$$

with

$$\frac{\partial M_n}{\partial E} = -\left(S_0 + S_1 e^{ik_n} + S_{-1} e^{-ik_n}\right) + K_1 \frac{ie^{ik_n}}{v_n} + K_{-1} \frac{-ie^{-ik_n}}{v_n}. \tag{C.16}$$

Eq. (C.13) implies that M_n has always an eigenvalue equal to zero, so that M_n is singular. Therefore we can not solve Eq. (C.15) by simple inversion of M_n [344]. One possible solution is to rescale the ϕ_n in such a way, that the element with the largest complex modulus is set to 1 [344]. We denote the index of this element n_{max}. $\tilde{\phi}_n$ has to be adjusted accordingly. The general form of the ψ_n then is

$$\phi_n = \begin{pmatrix} \phi_{n,1} \\ \phi_{n,2} \\ \vdots \\ 1 \\ \vdots \\ \phi_{n,N} \end{pmatrix}. \tag{C.17}$$

Its derivative becomes

$$\frac{\partial \phi_n}{\partial E} = \begin{pmatrix} \frac{\partial \phi_{n,1}}{\partial E} \\ \frac{\partial \phi_{n,2}}{\partial E} \\ \vdots \\ 0 \\ \vdots \\ \frac{\partial \phi_{n,N}}{\partial E} \end{pmatrix}, \tag{C.18}$$

where the chosen normalization results in the n_{max}^{th} element in the derivative to be exactly zero. The non-zero elements can be obtained by solving an equation equivalent to Eq. (C.15) [344], in which M_n is replaced by an invertible matrix $\hat{M}_n$, and $\frac{\partial M_n}{\partial E}$ is replaced by a modified matrix D_n

$$\hat{M}_n \frac{\partial \phi_n}{\partial E} = -D_n \phi_n. \tag{C.19}$$

$\hat{M}_n$ is set equal to M_n, but the n_{max}^{th} row and column are substituted by zeroes, except for the diagonal element, which is substituted by 1. The resulting matrix $\hat{M}_n$ has the form

$$\hat{M}_n = \begin{pmatrix} M_{n,11} & M_{n,12} & \dots & 0 & \dots & M_{n,1N} \\ M_{n,21} & M_{n,22} & \dots & 0 & \dots & M_{n,2N} \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \dots & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ M_{n,N1} & M_{n,N2} & \dots & 0 & \dots & M_{n,NN} \end{pmatrix}. \tag{C.20}$$

D_n is set equal to $\frac{\partial M_n}{\partial E}$, but the n_{max}^{th} row is substituted by a line of zeroes. The derivatives $\frac{\partial \phi_n}{\partial E}$ are then given by

$$\frac{\partial \phi_n}{\partial E} = -\hat{M}_n^{-1} D_n \phi_n. \tag{C.21}$$

By using Eqs. (C.12) and (C.21) in Eq. (C.8), the derivative of the SEs with respect to energy can be evaluated.

Appendix D

Publications stemming from this work

• Lattice Distortion Effects on the Magnetostructural Phase Transition of MnAs, F. Iikawa, M. J. S. P. Brasil, C. Adriano, O. D. D. Couto, C. Giles, P. V. Santos, L. Däweritz, I. Rungger, and S. Sanvito, Phys. Rev. Lett., **95**, 077203 (2005).

• *Ab initio* study on the magnetostructural properties of MnAs, I. Rungger and S. Sanvito, Phys. Rev. B, **74**, 024429 (2006).

• Bias Dependent TMR in Fe/MgO/Fe(100) Tunnel Junctions, Ivan Rungger, Alexandre Reily Rocha, Oleg Mryasov, Olle Heinonen and Stefano Sanvito, MRS Proceedings, 941E, 0941-Q01-03 (2006).

• First Principles Study of the Phase Transitions of MnAs, Ivan Rungger and Stefano Sanvito, MRS Proceedings, 941E, 0941-Q01-03 (2006).

• Electronic transport through Fe/MgO/Fe(100) tunnel junctions, Ivan Rungger, Alexandre Reily Rocha, Oleg Mryasov, Olle Heinonen and Stefano Sanvito, J. Magn. Magn. Mater., **316**, 481 (2007), Proceedings of the Joint European Magnetic Symposia.

• Algorithm for the construction of self-energies for electronic transport calculations based on singularity elimination and singular value decomposition, I. Rungger and S. Sanvito, Phys. Rev. B, **78**, 035407 (2008).

• *I-V* curves of Fe/MgO (001) single- and double-barrier tunnel junctions, J. Peralta-Ramos, A. M. Llois, I. Rungger and S. Sanvito, Phys. Rev. B, **78**, 024430 (2008).

• Resonant electronic states and *I-V* curves of Fe/MgO/Fe(100) tunnel junctions, I. Rungger, O. Mryasov and S. Sanvito, submitted; also arXiv:cond-mat/0808.0902

• Weakly coupled and bound states in electronic transport through Fe/MgO tunnel junctions, I. Rungger and S. Sanvito, in preparation.

• Electronic transport through Fe/MgO(100) tunnel junctions with oxygen deficient MgO barriers, I. Rungger and S. Sanvito, in preparation.

• Bias dependent transmission through quantum well states in double barrier tunnel junctions, I. Rungger, J. Peralta-Ramos, A. M. Llois and S. Sanvito, in preparation.

• *Ab initio* simulations of the transport properties of Mn_{12} based spin-devices. C. D. Pemmaraju, I. Rungger and S. Sanvito, in preparation.

• Use of the embedding method for non-equilibrium transport calculations, C. Toher, M. Khadilkar, I. Rungger, and S. Sanvito, in preparation.