Theory of Non-equilibrium Vertex Correction

Youqi Ke
Centre for the Physics of Materials
Department of Physics
McGill University
Montréal, Québec
Canada

A Thesis submitted to the
Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Contents

List of Figures

Abstract

For realistic nanostructures, there are inevitably some degree of disorder such as impurity atoms, imperfect lattices, surface roughness, etc.. For situations where disorder locate randomly in the nanostructure, any calculated quantum transport results should be averaged over disorder distributions. A brute force approach is to generate many disorder configurations, calculate each of them, and then average the results. For atomistic first principles modeling, such a brute force averaging is computationally prohibitive - if not impossible, to perform. It is therefore very important and useful to develop a theoretical framework where the disorder averaging is done analytically before atomic first principles analysis is carried out.

In this thesis, we have developed such a first principles non-equilibrium quantum transport theory and its associated modeling software for predicting disorder scattering in nano-electronic devices. Our theoretical formalism is based on carrying out density functional theory (DFT) within the Keldysh non-equilibrium Green's function (NEGF) framework, and a non-equilibrium vertex correction (NVC) theory for handling disorder configurational average at the non-equilibrium density matrix level. In our theory, we use the coherent potential approximation to calculate disorder averaging of the device Hamiltonian and one particle Green's functions, and use NVC to calculate correlated multiple impurity scattering at the non-equilibrium density matrix level. After the NEGF-DFT-NVC self-consistent calculation is converged, we calculate the transmission coefficients by a second, unavoidable, vertex correction. The NEGF-DFT-NVC theory allows us to predict non-equilibrium quantum transport properties of nanoelectronic devices with atomistic disorder from first principles without any phenomenological parameters. The theory and implementation details are presented.

We have applied the NEGF-DFT-NVC method to investigate several important problems associated with disorder scattering in nano-electronic device systems. These include interface roughness scattering in Fe/vacuum/Fe magnetic tunnel junctions; the diffusive scattering of carriers due to oxygen vacancies in Fe/MgO/Fe magnetic tunnel junctions; the surface roughness scattering that enhances resistivity of copper interconnect wires; and effects of barrier layer coating for Cu interconnects. Our

investigations reveal very important role played by the atomic level defects and impurities to both equilibrium and nonequilibrium quantum transport properties, and results compare favorably with the corresponding experimental data.

Résumé

Dans le cas de nanostructures concrètes, un certain degré de désordre apparaît inévitable tel que la présence d'impuretés, de structures cristallines imparfaites, de surfaces rugueuses, etc. Dans les situations où le désordre se matérialise aléatoirement dans la nanostructure, tout calcul de transport quantique devrait être réalisé en tant que moyenne sur plusieurs distributions désordonnées. Une approche par force brute consiste à générer plusieurs configurations désordonnées, calculer les propriétés d'intérêt pour chacune d'entre elles, et ensuite effectuer la moyenne des résultats. Dans le cas de la modélisation atomique à partir des principes premiers, une telle moyenne par force brute est prohibitive en terme de temps de calcul - sinon impossible. Il est ainsi très important et utile de développer un cadre théorique où la moyenne de désordre est faite analytiquement avant que l'analyse par les principes premiers ne soit effectuée.

Dans cette thèse, nous avons développé une telle théorie de transport quantique hors équilibre à partir des principes premiers et le logiciel de modélisation associé pour la prédiction de la diffusion par désordre dans des dispositifs nanoélectroniques. Notre formalisme théorique est basé sur l'utilisation de la théorie de la fonctionnelle de densité (DFT) dans le cadre de la fonction de Green hors équilibre de Keldysh (NEGF), et sur l'emploi d'une correction de sommet hors équilibre (NVC) pour le traitement des moyennes configurationnelles de désordre au niveau de la matrice de densité hors équilibre. Dans notre théorie, nous utilisons l'approximation du potentiel cohérent afin de calculer les moyennes de désordre de l'Hamiltonien du dispositif et les fonctions de Green à une particule, et nous utilisons la NVC pour calculer la diffusion par impuretés multiples corrélée au niveau de la matrice de densité horséquilibre. Après que le calcul auto-cohérent NEGF-DFT-NVC ait convergé, nous calculons les coefficients de transmission par le biais d'une seconde correction de sommet inévitable. La théorie NEGF-DFT-NVC nous permet de prédire les propriétés de transport quantique hors équilibre de dispositifs nanoélectroniques avec désordre au niveau atomique à partir des principes premiers sans aucun paramètre phénoménologique. La théorie et les détails d'implémentation sont présentés dans ce travail.

Nous avons appliqué la méthode NEGF-DFT-NVC afin d'examiner plusieurs problèmes

importants associés à la diffusion par désordre dans des systèmes de dispositif nanoélectronique. Cela inclut la diffusion par rugosité de surface dans des jonctions tunnel magnétiques Fe/vide/Fe; la diffusion due à des lacunes d'oxygène dans des jonctions tunnel magnétiques Fe/MgO/Fe; la diffusion par rugosité de surface qui décuple la resistivité de fils de connexion en cuivre; et les effets des revêtements couche barrière pour des connexions en Cu. Notre étude révèle le rôle très important joué par les défauts de niveau atomique et les impuretés vis-à-vis des propriétés de transport quantique à la fois en équilibre et hors équilibre, et les résultats se comparent favorablement aux données expérimentales correspondantes.

Statement of Originality

In this thesis we report our development of a theoretical formalism and its associated software tool for calculating disorder scattering at the nonlinear and non-equilibrium level for quantum transport in nanoelectronic devices. Our theory is based on carrying out density functional theory (DFT) calculations within the Keldysh nonequilibrium Green's function framework (NEGF), and deal with the disorder effects by the nonequilibrium vertex correction (NVC) theory. This work has resulted in a powerful and unique quantum transport technique such that non-equilibrium electron transport properties of realistic devices having atomistic disorder, can be predicted from first principles without using any phenomenological parameter. My original contributions to this work include:

- Derivation and implementation of the NEGF formalism for quantum transport within the tight-binding linear Muffin Tin orbital method (Chapter 4).
- Formulation and implementation of the NVC theory using a many-body perturbation approach within the NEGF-DFT formalism such that the disorder averaging at non-equilibrium can be carried out at the density matrix level (Chapter 5). This is the first time in literature for solving the problem of non-equilibrium disorder scattering, which represents a significant theoretical advance for quantum transport theory. The associated software implementation provides a most efficient, accurate and powerful modeling method for a wide range of applications.
- Design and develop the comprehensive software package with full parallelization that implements the NEGF-DFT-NVC formalism. Many important computational algorithms have been developed and applied for fast and accurate computation, these include: 1) Renormalization-decimation technique: most efficient for calculating surface Green's functions; 2) Real space technique for boundary condition of the device; 3) Techniques for incorporating the coherent potential approximation and NVC self-consistent solution into the entire electronic structure calculation; 4) Using the geometry symmetry of the device to save k-sampling; 5) Full Parallelization over spin, energy and BZ k-points computa-

tion; 6) Recursive Green's function technique for inverse of tri-diagonal matrix; 7) Optimization over calculation procedures, and developing various mixers to realize fast convergence of non-equilibrium computation.

- The investigation of effects of various disordered impurities/defects in a variety of nano-electronic devices: interface roughness scattering in Fe/vacuum/Fe magnetic tunnel junctions; disorder scattering by oxygen vacancies in Fe/MgO/Fe magnetic tunnel junctions; surface roughness scattering that enhances resistivity of copper interconnect wires; and effects of barrier layer coating for Cu interconnects.

  The work in this thesis has resulted in several papers for peer-reviewed journals [1, 2, 3, 4, 5].

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my thesis supervisor, Prof. Hong Guo for his great supervision and encouragement and patience in every step of my PhD studies in McGill University. My experience of working as a student of Prof.Guo is an invaluable treasure which will benefit my whole life. Working with him during the past 5 years has greatly extended my problem solving and creative ability, and strengthened my independent working ability. I am also deeply indebted to Prof. Ke Xia in Beijing Normal University for providing insightful suggestions and help during the course of my research. I also appreciate the opportunity to collaborate with Prof. Dianel Gall in Rensselaer Polytechnic Institute on the project of Copper Interconnect

Dr. Pengxiang Xu deserves acknowledgement for sharing his knowledge about the TB-LMTO method at the very beginning of my PhD study. It is my great pleasure to say 'thank you' to the present and previous members of our research group. Thank especially to Dr. Lei Liu, Dr. Yu Zhu, Zhanyu Ning, Dr. Zahid Ferdows, Jessy Massen, Zi Wang, Mathieu César, Dr. Vladimir Timochevskii for extensive discussions and collaborations and pleasant personal interactions.

Special acknowledgement should be given to 'The Coach' for correcting and editing my thesis so that I could meet the deadline.

At last but not least, I would like to thank my wife Ruili Ma for her great support during my Ph.D study and also my parents and sisters.

Physical Constants and Units

1 Å	=	10^{-10} m
a_0 (Bohr radius)	=	0.5292 Å
m_e (electron mass)	=	9.1096×10^{-31} kg
m_p (proton mass)	=	1.6726×10^{-27} kg
e (electron charge)	=	1.6×10^{-19} C
h (Planck's constant)	=	6.626×10^{-34} J s
k_B (Boltzmann's constant)	=	1.38×10^{-23} K
k_BT (at 1 K)	=	8.616×10^{-5} eV
c (speed of light)	=	2.9979×10^{8} m/s
G_0 (quantum unit of conductance)	=	$7.75 \times 10^{-5}\Omega^{-1} = \frac{1}{12.9\text{k}\Omega}$

Atomic units are used throughout this thesis unless otherwise indicated. In this system of units, $e = m_e = \hbar = 1$.

1 unit of Length	=	a_0	=	0.5292 Å
1 unit of Mass	=	m_e	=	9.1096×10^{-31} kg
1 unit of Charge	=	e	=	1.6×10^{-19} C
1 unit of Angular momentum	=	$\hbar$	=	1.0546×10^{-34} J s
1 unit of Energy	=	1 Hartree	=	27.2 eV
1 unit of Time	=	$\frac{\hbar}{1 \text{ Hartree}}$	=	2.4189×10^{-17} s

List of Abbreviations

TMR	Tunneling Magnetoresistance Ratio
MTJ	Magnetic Tunnel Junction
DFT	Density Functional Theory
EQ	Equilibrium
TB-LMTO	Tight-Binding Linear Muffin Tin Orbital
NVC	Non-Equilibrium Vertex Correction
NEQ	Non-Equilibrium
NEGF	Non-Equilibrium Green's Function
XC	Exchange-Correlation
PC	Parallel [Magnetic] Configuration
APC	Anti-Parallel [Magnetic] Configuration

1

Introduction

Due to the continuing down-scaling of semiconductor electronic devices, modern transistors in today's computer technology have entered the nanometer era. For the past four decades, the electronics industry has followed the Moore's observation that device feature sizes steadily decrease at a rate of roughly a factor of two every eighteen months. According to the *International Technology Roadmap for Semiconductors* (ITRS), 22 nm technology will be achieved by year 2016 when the projected minimum device features will be less than 10 nm and computer chips will have more than six billion transistors[6]. The device miniaturization has brought the technology into a new regime where the quantum phenomena of charge and spin transport become important physics. In addition, the discrete properties of materials are playing increasingly dominant roles in device operation. So far, it has become clear both experimentally and theoretically that quantum transport in nano-electronic devices is closely coupled to the atomic, chemical and materials properties of the device nanostructure. Such a coupling poses new challenges to both experimental and theoretical understanding of nano-electronic device physics.

On the theoretical side, it has been a very difficult problem to simultaneously incorporate non-equilibrium quantum transport effects and atomic scale microscopic details of the material, into the same device physics formalism. We begin by recalling that the well established traditional device theory in micro-electronics has its principle rooted in classical or semi-classical physics. As such it is insufficient or even invalid for quantum effects. The practical modeling methods of micro-electronics are empirical

or semi-empirical where material and electronic parameters are obtained by fitting to experimental data. Such a fitting procedure is becoming increasingly expensive and less reliable as device sizes reach the $\sim$ 20nm regime. The discrete properties of the materials have already produced very large device to device variability that has become a major problem of the device industry. Therefore, there is an urgent need to develop a practically viable quantum transport theoretical formalism that can make predictive modeling of nano-scale devices without using any phenomenological parameter, namely an atomistic first principles theory and method.

For materials theory, the most powerful, practical and widely used atomistic formalism is the density functional theory (DFT) [7]. DFT solves the many-body quantum mechanic model of the material in a mean field manner[8]. The success of DFT is evidenced by the 1998 Nobel Prize awarded to its discoverer, Prof. Walter Kohn. Traditionally, DFT has been applied to two kinds of problems: (i) closed/isolated systems of atoms/molecules, as in quantum chemistry; (ii) periodic systems as in solid state physics. A nano-electronic device is however neither finite nor periodic: it is an open boundary problem where a scattering region is contacted by external metal electrodes and connected to the outside world. Because a current is flowing through the system, the physics is strictly non-equilibrium process.

For non-equilibrium quantum transport theory, the most powerful and systematic technique is the Keldysh non-equilibrium Green's function formalism(NEGF)[9, 10, 11]. NEGF expresses quantum transport properties such as electric/spin current and conductance in terms of various Green's functions which can be calculated for a given Hamiltonian of the device. In a typical NEGF theoretical analysis[10, 11, 12], however, the device Hamiltonian and its potential terms are not calculated but assumed with phenomenological parameters. The aim was to reveal quantum transport features **qualitatively**. One can also adjust the parameters in the Hamiltonian to fit the theoretical results to the experimental data.

During the past one and half decade, tremendous effort has been devoted to develop

a theoretical formalism and associated modeling tool that combine the NEGF quantum transport theory with the DFT materials theory. The aim of such an effort was to achieve the capability for **quantitative** prediction in non-equilibrium transport problems from atomic first principles without any phenomenological parameter. The NEGF-DFT formalism, as first reported in Ref.[13] in 2001, has steadily evolved in to a state-of-art first principles technique that goes beyond (i,ii) above where DFT-like method is applied within the NEGF framework in real space[13]. The basic idea of the real space NEGF-DFT formalism[13] is to calculate device Hamiltonian and electronic structure by a DFT-like self-consistent field theory, populate this electronic structure by NEGF which accounts for the non-equilibrium quantum statistics, and deal with open device boundaries and electrostatic boundaries directly using real space techniques. The fact that NEGF-DFT formalism works for steady state quantum transport has been demonstrated by many direct quantitative comparison to experimental data, as well as by recent literature which puts this formalism onto more rigorous theoretical footing[14]. As a result, many groups in the world have developed similar NEGF-DFT implementations [13, 15, 16, 17, 18, 19, 20, 21, 22] and it has become the *de facto* standard technique for quantitative modeling of nano-electronic devices.

In traditional DFT, the electronic density is determined by the electron wave function,

$$\rho(\mathbf{r}) \;=\; \sum_i f_i |\psi_i(\mathbf{r})|^2 \quad , \tag{1.1}$$

where f_i is the Fermi-Dirac distribution function for orbital i. The appearance of f_i indicates equilibrium physics. In the NEGF-DFT formalism, on the other hand, the electronic density is calculated by the NEGF $G^<$,

$$\rho(\mathbf{r}) \;=\; \frac{1}{2\pi} \int dE G^<(E, \mathbf{r}) \quad , \tag{1.2}$$

where the quantity $G^<$ accounts for non-equilibrium distribution and by integrating the electron energy E, Eq.(1.2) fills the electronic states to build the density ρ at non-equilibrium. After ρ is obtained this way, it is used to determine the Hamiltonian

of the system. In the NEGF-DFT formalism, the electronic Hamiltonian has several terms including the kinetic energy term, the Hartree potential term, the exchange-correlation term, and terms due to external fields. Even though these are the same terms as in the conventional equilibrium DFT theory, the density that determines these terms are from Eq.(1.2) - which is non-equilibrium, and not from Eq.(1.1). Going to equilibrium, it can be proved that Eq.(1.2) reduces to Eq.(1.1) exactly. Therefore, even though 'DFT' is used in naming the NEGF-DFT method, the theories of NEGF-DFT and conventional DFT are qualitatively different. After the NEGF-DFT self-consistently determined the Hamiltonian of the system, transport properties can be calculated. For instance the electron transmission coefficient is obtained by the following formula,

$$T(E) = Tr[\Gamma_L(E)G^r(E)\Gamma_R(E)G^a(E)]. \tag{1.3}$$

where $G^{r,a}$ are the respective retarded and advanced Green's functions, $\Gamma_{R/L}$ describes the coupling between the device and left and right electrodes which can be obtained from the surface Green's functions of the electrode surfaces.

Despite of its successes, so far the NEGF-DFT formalism has been restricted to analyzing quantum transport in perfect nano-structures without random disorder. However, any realistic nano-electronic devices inevitably contain some amount of disorder or imperfections, such as impurities, defects, dopants, dislocations, and so on. Indeed, it is the unintentional impurities sitting at unpredictable random locations that have produced the large device-to-device variability. Effects due to disordered impurities or defects inside realistic devices are very important from both the theoretical and technological points of view, because quantum transport properties can be dramatically influenced by or even built on the impurities. The examples are the disorder effect in magnetic tunneling junctions[23], spin dependent transport in dilute magnetic semiconductors[24], dopant scattering in semiconductor nano-wires and transistors[25], disorder scattering in the memresitive devices[26], and so on. Un-

derstanding the roles of disordered impurities for non-equilibrium quantum transport has become a critical issue for device theory.

The most difficult stumbling block for materials theory of disordered systems is the absence of translational invariance which renders many well established solid state physics methods useless. For example, the Bloch theorem no longer holds for disordered atomic structures and, as a consequence, momentum conservation is lost. This turns the electron scattering from specular to diffusive[1]. In order to determine physical properties, one is faced with a rather disconcerting task of examining individually, one by one, each possible configuration of the disorder. Afterward, random spatial distribution of the disorder require theoretical results to be averaged over a large ensemble of the disorder configurations. For example, for disordered systems, Eqs.(1.2,1.3) are no longer meaningful. Instead, one needs to calculate their configurationally averaged results:

$$\begin{aligned} \overline{\rho} &= \frac{1}{2\pi}\int dE\overline{G^<(E)} \\ \overline{T(E)} &= Tr[\overline{\Gamma_L(E)G^r(E)\Gamma_R(E)G^a(E)}]. \end{aligned} \tag{1.4}$$

where $\overline{(\cdots)}$ denotes configurational average.

In first principles calculation of disordered systems, the above configurational average may be carried out over a super-cell with a translational invariant boundary condition. This super-cell calculation can be performed by choosing an appropriate size of the super-cell, generating many atomic configurations for a given impurity concentration x, computing physical quantities for each configuration, and finally averaging the results. Such a brute force approach is simple but is severely limited due to the prohibitively large computational costs and the fact that the averaged results

[1]Throughout this thesis, specular scattering is sometimes called coherent scattering; while diffusive scattering is sometimes called incoherent scattering. These wording should not be confused with quantum coherence. For both specular and diffusive scattering, our theory assumes quantum coherence, namely no inelastic scattering is considered and the impurities provide elastic scattering to the charge carriers.

depend on the super-cell size. Especially, the super-cell method becomes impractical when treating two-probe nano-electronic device systems such as those shown in Fig.(6.2,7.1,8.1 and 8.2) because the system size becomes impossible to handle even with a modern or even future supercomputer. For instance, there are known situations where millions of disordered configurations must be calculated to obtain accurate averaging to quantum transport properties[27]. It is clear that a practical and efficient technique is desired for addressing the configurational average problem within the NEGF-DFT framework for realizing *ab initio* simulation of the disordered nano-electronic devices.

There are many different ways in which realistic materials deviate from the ideal structure, e.g. substitutional disorder, magnetic disorder, vacancy, interface roughness, etc. The substitutional disorder is known as the simplest type of disorder in which the system is still characterized by a regular lattice and the lattice sites can be occupied by the original host atoms or by impurity atoms. A vacancy can be viewed as replacing a host atom by an empty atom. So far, the greatest progress in theory toward understanding disordered materials is on the substitutional disorder. A widely used technique is the coherent potential approximation (CPA)[28, 29, 30] as implemented in KKR[31] and tight-binding linear Muffin Tin orbital (LMTO) [32, 33, 34, 35] first principles methods. In CPA, an effective medium with translational invariance is constructed self-consistently to describe the substitutionally disordered system. Such a restoration of translational invariance drastically reduces computation complexity and allows us to apply many well established solid state physics techniques. In particular, CPA analytically derives formula for configurational averaged quantities and, afterward, evaluates these averaged quantities numerically only once (instead of many times in super-cell brute force approach). CPA is now a well established formalism that has seen a wide range of applications in materials physics. More recently, CPA theory has even been developed to investigate thermal phase transitions[36].

However, so far the first principles CPA technique can only been applied to **equilibrium** problems[28, 30, 37]. On the other hand, nano-electronic devices operate under **non-equilibrium** conditions due to current flow and, for instance, one often wishes to predict nonlinear current-voltage (I-V) characteristics of a device. Therefore, it is extremely important to develop appropriate theoretical techniques for non-equilibrium disorder averaging within the NEGF-DFT framework, namely, to calculate Eq.(1.4). The purpose of this thesis research is to solve the problem of non-equilibrium disorder average for non-equilibrium and nonlinear quantum transport.

To treat the non-equilibrium quantum transport properties of nano-electronic devices having atomistic substitutional disorder under external bias potential, we have developed a non-equilibrium vertex correction (NVC) theory and implemented it into the NEGF-DFT formalism. In particular, NVC solves the problem for non-equilibrium density averaging, and we apply CPA to self-consistently construct the effective medium that gives the configurationally averaged one-particle Green's function, $\overline{G^{r,a}(E)}$. The configurational average of NEGF which determines the non-equilibrium density matrix is performed by the NVC theory such that:

$$\overline{G^<} = \overline{G^r \Sigma^< G^a} = \overline{G^r}(\Sigma^< + \Omega_{NVC})\overline{G^a} \tag{1.5}$$

where $\Sigma^<$ is the self-energy due to the ordered electrodes, the self-energy Ω_{NVC} is the central quantity of this work which is called the non-equilibrium vertex correction describing the multiple impurity scattering at non-equilibrium. The electric current after configurational average is evaluated using the Landauer formula which involves an unavoidable second vertex correction self-energy,

$$\begin{aligned} \overline{T(E)} &= Tr[\overline{\Gamma_L G^r \Gamma_R G^a}] \\ &= Tr[\Gamma_L \overline{G^r} \Gamma_R \overline{G^a} + \Gamma_L \overline{G^r} \Omega_{VC} \overline{G^a}] \end{aligned} \tag{1.6}$$

where the first term on the right hand side represents specular scattering, the second

term gives the inter-channel diffusive scattering due to the impurities. Our NEGF-DFT-NVC formalism is derived and implemented within the tight binding LMTO first principle framework.

By using the NEGF-DFT-NVC theory, disorder effects to nonlinear and non-equilibrium quantum transport properties can be calculated from atomic first principle in a self-consistent and efficient manner. The capabilities of our NEGF-DFT-NVC method are illustrated by several representative studies of nano-electronic systems where the disorder effects of impurities or defects are found to be very important for electron conduction. These examples include the interface roughness scattering in Fe/vacuum/Fe magnetic tunneling junctions (MTJ)[2]; effects of oxygen vacancies in Fe/MgO/Fe MTJ[3]; and effects of rough surfaces in Cu interconnects[4, 5].

This thesis is organized as follows. In chapter 2, we review the basic theorems of DFT and the Kohn-Sham equation. We introduce the exchange-correlation energy functional, the self-consistent mean field nature of DFT. These serve as a starting point for the first principles materials theory.

In chapter 3, we present the implementation of the tight binding LMTO method for solving the the Kohn-Sham equation. We introduce the Muffin Tin approximation for crystals, the tight binding representation of a quantity α, Muffin tin orbital linearization, the atomic sphere approximation, and finally the Hamiltonian matrix in the nearly orthogonal LMTO basis.

In chapter 4, we derive the NEGF formalism for quantum transport in two-probe electronic device structures and present its implementation in a finite localized basis set. We introduce the useful recursive Green's function method. Finally, we rewrite the NEGF formalism into the tight binding LMTO framework.

In chapter 5, we present details of the NVC theory for non-equilibrium disorder averaging of the density matrix; we also present a numerical implementation of the NEGF-DFT-NVC formalism within the tight binding LMTO framework for quantum

transport [1].

In chapter 6, we apply the our NEGF-DFT-NVC method to investigate the interface roughness scattering effects in the Fe/Vacuum/Fe MTJ. The results of this work can be found in Ref.[2].

In chapter 7, we present a study of the roles of disordered oxygen vacancies inside the Fe/MgO/Fe MTJ. The results of this work has been submitted for publication at the writing of this thesis[3].

In chapter 8, we present a investigation of random surface roughness scattering in the copper interconnect wires. Results of this chapter can be found in Refs.[4, 5].

Finally, chapter 9 is reserved for a short conclusion, outlook and other related applications of our NEGF-DFT-NVC currently underway.

2

Density Functional Theory

As briefly discussed in chapter 1, the density functional theory (DFT)[7] is the most powerful, practical and widely used atomistic formalism for quantitative modeling of materials. DFT solves the many-body quantum mechanic model of the material in a mean field manner[8]. In this chapter, we shall briefly review DFT.

2.1 Basic theorems

In this section, we shall present two basic Hohenberg-Kohn (HK) theorems [7] that have put DFT on firm theoretical footing. The first HK theorem states that the ground-state properties of a N interacting electron system is a functional of the ground electron density. Namely, if the density is known, then all the properties of the ground state of the N-electron system become known. The second HK theorem asserts that the correct ground-state electron density should minimize the total energy functional. The original HK theorems were derived only for the non-degenerate ground state. Then, M. Levy[38] in 1979 provided a more general derivation to encompass the degenerate ground state. Here, we will follow Levy's derivation of these basic theorems of DFT.

We start with the Hamiltonian which describes the N interacting electrons in an

external potential V_{ext}.

$$\hat{H} = \sum_{i=1}^{N} \hat{T}_i + \sum_{i<j} \frac{1}{\mid \mathbf{r}_i - \mathbf{r}_j \mid} + \sum_{i=1}^{N} V_{ext}(\mathbf{r}_i) = \hat{T} + \hat{V}_{ee} + \hat{U}_{ext} \tag{2.1}$$

where $\hat{T}$ and $\hat{V}_{ee}$ are the kinetic energy operator and electron-electron Coulomb interaction operator, respectively. Let's define an universal functional of electron density for the kinetic and electron repulsion energies[38],

$$F[\rho] = min\langle \Psi_\rho \mid \hat{T} + \hat{V}_{ee} \mid \Psi_\rho \rangle \ . \tag{2.2}$$

The minimum is searched over all the antisymmetric wavefunctions Ψ that give the fixed density ρ

$$\rho = \langle \Psi_\rho \mid \Psi_\rho \rangle \tag{2.3}$$

The functional $F[\rho]$ is universal because it is not dependent on either the specific system or the external potential V_{ext}. If we introduce the total energy functional $E[\rho]$,

$$E[\rho] = \int d\vec{r} U_{ext}(\mathbf{r})\rho(\mathbf{r}) + F[\rho] = \langle \Psi_{\rho,min} \mid \hat{T} + \hat{V}_{ee} + U_{ext} \mid \Psi_{\rho,min} \rangle \tag{2.4}$$

where $\Psi_{\rho,min}$ denotes the wave function satisfying Eq.(2.2), then the two basic theorems of DFT become:

$$E[\rho] \geq E_{GS} \tag{2.5}$$

For all the possible choices of ρ, the ground state density ρ_{GS} gives the minimum:

$$E[\rho_{GS}] = E_{GS} \tag{2.6}$$

where E_{GS} is the total energy of the ground state.

The proof of the first theorem is very straightforward. According to the minimum

property of the ground state, we can easily write down

$$\langle \Psi_{\rho,min} \mid \hat{T} + \hat{V}_{ee} + U_{ext} \mid \Psi_{\rho,min} \rangle \geq E_{GS} \ . \tag{2.7}$$

This proves the inequality in the first HK theorem. To prove the second HK theorem, applying the minimum property of the ground-state again, we find

$$E_{GS} = \langle \Psi_{GS} \mid \hat{T} + \hat{V}_{ee} + \hat{U}_{ext} \mid \Psi_{GS} \rangle \leq \langle \Psi_{\rho_{GS},min} \mid \hat{T} + \hat{V}_{ee} + \hat{U}_{ext} \mid \Psi_{\rho_{GS},min} \rangle \tag{2.8}$$

where Ψ_{GS} is the ground-state wave function which yields the correct ground-state electron density $\rho_{GS} = \langle \Psi_{GS} \mid \Psi_{GS} \rangle$. We can subtract the external potential term from both side of above equation and obtain

$$\langle \Psi_{GS} \mid \hat{T} + \hat{V}_{ee} \mid \Psi_{GS} \rangle \leq \langle \Psi_{\rho_{GS},min} \mid \hat{T} + \hat{V}_{ee} \mid \Psi_{\rho_{GS},min} \rangle \tag{2.9}$$

According to the definition of $\Psi_{\rho_{GS},min}$, we can obtain a reverse relation

$$\langle \Psi_{GS} \mid \hat{T} + \hat{V}_{ee} \mid \Psi_{GS} \rangle \geq \langle \Psi_{\rho_{GS},min} \mid \hat{T} + \hat{V}_{ee} \mid \Psi_{\rho_{GS},min} \rangle \tag{2.10}$$

The above two equations are satisfied simultaneously when and only when

$$\langle \Psi_{GS} \mid \hat{T} + \hat{V}_{ee} \mid \Psi_{GS} \rangle = \langle \Psi_{\rho_{GS},min} \mid \hat{T} + \hat{V}_{ee} \mid \Psi_{\rho_{GS},min} \rangle \tag{2.11}$$

Then we have

$$\begin{aligned} E_{GS} &= \langle \Psi_{GS} \mid \hat{T} + \hat{V}_{ee} + U_{ext} \mid \Psi_{GS} \rangle \\ &= \int d\vec{r} U_{ext}(\vec{r}) \rho_{GS}(\vec{r}) + \langle \Psi_{GS} \mid \hat{T} + \hat{V}_{ee} \mid \Psi_{GS} \rangle \\ &= \int d\vec{r} U_{ext}(\vec{r}) \rho_{GS}(\vec{r}) + \langle \Psi_{\rho_{GS},min} \mid \hat{T} + \hat{V}_{ee} \mid \Psi_{\rho_{GS},min} \rangle \\ &= \int d\vec{r} U_{ext}(\vec{r}) \rho_{GS}(\vec{r}) + F[\rho_{GS}] \\ &= E[\rho_{GS}] \end{aligned} \tag{2.12}$$

we can see the possible degeneracy of Ψ_{GS} will not affect the above proof. The HK theorems promise that the electron density is the fundamental variable for describing the many-body problem of interacting electrons. Introducing the kinetic energy functional $T[\rho] = \langle \Psi_{\rho,min} \mid \hat{T} \mid \Psi_{\rho,min} \rangle$ and electron-electron interaction energy functional $V_{ee}[\rho] = \langle \Psi_{\rho,min} \mid \hat{V}_{ee} \mid \Psi_{\rho,min} \rangle$, the total energy functional can be rewritten as

$$E[\rho] = \int d\vec{r} U_{ext}(\mathbf{r})\rho(\mathbf{r}) + T[\rho] + V_{ee}[\rho] \tag{2.13}$$

For $V_{ee}[\rho]$, we may separate it into two terms,

$$V_{ee}[\rho] = \frac{1}{2} \int \frac{\rho(\vec{r})\rho(\mathbf{r}')}{\mid \mathbf{r} - \mathbf{r}' \mid} d\mathbf{r} d\mathbf{r}' + W_{ee}[\rho] \tag{2.14}$$

where the first term is the Hartree energy describing the classical electrostatics, and the functional $W_{ee}[\rho]$ is a non-classic term of the electron-electron interaction which is the major part of the exchange-correlation energy we shall discuss next. If the universal functionals $T[\rho]$ and $W_{ee}[\rho]$ are known, minimization of the total energy functional $E[\rho]$ with respect to the density will yield the ground state energy and the electron density of the system, and all other ground-state properties can then be derived from them.

Unfortunately, the exact form of the functional $T[\rho]$ and $W_{ee}[\rho]$ are not known. Total energy minimization therefore requires approximations for them, this gives rise to the corresponding approximations in E_{GS} and $\rho_{GS}(\vec{r})$ and other derivable ground-state quantities. So far, two different approaches have been constructed for practical applications of DFT. One approach is the orbital free density functional theory [39] which is closely related to the original spirit of the HK theorems. In orbital free formulation, there are many deficiencies that come from the approximate treatment of the kinetic energy functional. The other approach, which is the standard, is due to Kohn and Sham[8] who introduced a different partitioning of the total energy functional so that the ground-state energy can be found by solving a set of single particle equations. In the next section, we will derive the single particle equations

of Kohn and Sham [8] and show how to map the interacting many-body problem of interacting electrons onto the tractable problem of non-interacting electrons moving in an self-consistent field. We should emphasize that DFT based calculations are in principle restricted to systems in their ground states.

2.2 Kohn-Sham equations

Kohn and Sham in 1965 separated[8] the functional $E[\rho]$ into the following form,

$$E[\rho] = T_0[\rho] + \frac{1}{2}\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{\mid \mathbf{r} - \mathbf{r}' \mid} d\mathbf{r}d\mathbf{r}' + \int d\mathbf{r} U_{ext}(\mathbf{r})\rho(\mathbf{r}) + E_{XC}[\rho] \tag{2.15}$$

where $T_0[\rho]$ defines the kinetic energy of a non-interacting electron system that owns the same electron density $\rho(\mathbf{r})$ as the original system of interacting electrons. The second and third terms are the classical Coulomb potential energy and external potential energy, respectively. The exchange-correlation energy E_{XC} contains all other residual interactions,

$$E_{XC}[\rho] = W_{ee}[\rho] + T[\rho] - T_0[\rho] \ . \tag{2.16}$$

Although $T_0[\rho]$ is different from the true many-body kinetic energy $T[\rho]$, it can be treated exactly in the Kohn-Sham (KS) approach and removes the deficiency associated with approximation made to kinetic energy functional - if $T[\rho] - T_0[\rho]$ is known. But since it is not known, some inevitably approximations have to be made (see below).

To find the ground state quantities, we must minimize the total energy $E[\rho]$ with the constraint that the system contains a constant number of electrons. Using the Lagrange multiplier μ, we apply the variational principle to obtain,

$$\frac{\delta}{\delta\rho(\mathbf{r})}\{E[\rho] - \mu\int \rho(\mathbf{r})d\mathbf{r}\} = 0 \ . \tag{2.17}$$

Hence

$$\frac{\delta E[\rho]}{\delta \rho(\mathbf{r})} = \frac{\delta T_0[\rho]}{\delta \rho(\mathbf{r})} + U_{ext}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{\mid \mathbf{r} - \mathbf{r}' \mid} d\mathbf{r} d\mathbf{r}' + \frac{\delta E_{XC}[\rho]}{\delta \rho(\mathbf{r})} = \mu \tag{2.18}$$

If we compare this result to a similar equation for non-interacting electrons moving in an effective potential $V_{eff}(\mathbf{r}')$, namely:

$$\frac{\delta E[\rho]}{\delta \rho} = \frac{\delta T_0[\rho]}{\delta \rho} + V_{eff}(\mathbf{r}) = \mu \ , \tag{2.19}$$

we see that Eqs.(2.18) and (2.19) describe the same mathematical problem if:

$$V_{eff}(\mathbf{r}) = U_{ext}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{\mid \mathbf{r} - \mathbf{r}' \mid} d\mathbf{r} d\mathbf{r}' + \frac{\delta E_{XC}[\rho]}{\delta \rho} \mid_{\rho=\rho_{GS}} \tag{2.20}$$

Hence, the ground-state single-particle density satisfying above equations (Eq.(2.18)-(2.20)) can be found by solving the one-electron Schördinger equation for the non-interacting electrons moving in an effective potential,

$$H_{KS}\psi_i(\mathbf{r}) = [-\nabla^2 + V_{eff}(\mathbf{r})]\psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}) \tag{2.21}$$

such that

$$\rho_{GS}(\mathbf{r}) = \sum_i^N \mid \psi_i(\mathbf{r}) \mid^2 \ . \tag{2.22}$$

Eq.(2.21) is the well known Kohn-Sham (KS) equation where the operator acting on the KS orbital ψ_i is the KS Hamiltonian. Here N is the number of electrons, ϵ_i is the eigenvalue of the KS Hamiltonian. The ground-state total energy of Eq.(2.15) can be rewritten in term of KS orbital energy ϵ_i,

$$E[\rho] = \sum_i^N \epsilon_i - \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{\mid \mathbf{r} - \mathbf{r}' \mid} d\mathbf{r} d\mathbf{r}' - \int \rho(\mathbf{r}) V_{XC}(\mathbf{r}) d\mathbf{r} + E_{XC}[\rho] \tag{2.23}$$

where $V_{XC} = \frac{\delta E_{XC}[\rho]}{\delta \rho}$ is called exchange-correlation potential. The magic of the Kohn-Sham DFT lies in the fact that it transforms the complicated interacting many-electron problem into a tractable problem of non-interacting electrons moving in a

self-consistent effective potential field. The solution of these self-consistent equations (Eqs.(2.20)-(2.23)) yields in principle the exact ground state energy and electron density, provided that the exact exchange-correlation functional is known. Practically, a calculation starts from an initial guess for $\rho(\mathbf{r})$. One then calculates the corresponding V_{eff} and solves the Kohn-Sham equations for the ψ_i. Using the KS orbital one constructs a new density $\rho(\mathbf{r})$ from Eq.(2.22) for the next iteration, and the process is repeated until numerical convergence is achieved.

The Kohn-Sham DFT has been the foundation of modern electronic structure calculations. It should be noted that the Kohn-sham DFT outlined above breaks down for cases of spin polarization, magnetic field, multi-components, relativistic, and time-dependent problems. To extend DFT to these situations, since the original work of Kohn and Sham[8], many further generalizations of DFT have been developed to meet different physical requirements. As an example, for non-relativistic systems having spin polarization[40], E_{XC} may depend on the single-particle densities of electrons with each spin channel:

$$E_{XC} = E_{XC}[\rho_\uparrow, \rho_\downarrow] \ , \tag{2.24}$$

and the total single-particle density is the sum of the spin channels:

$$\rho = \rho_\uparrow + \rho_\downarrow \ . \tag{2.25}$$

In addition, we have a spin dependent exchange-correlation potential:

$$V_{XC,\rho_\sigma} = \frac{\delta E_{XC}[\rho_\uparrow, \rho_\downarrow]}{\delta \rho_\sigma(\mathbf{r})} \tag{2.26}$$

where $\sigma = \uparrow, \downarrow$ is the spin index. The Kohn-Sham equation must then be solved for each spin orientation separately.

From the above equations, the only thing that remains unknown and can not be calculated exactly is the exchange-correlation energy functional E_{XC} which play the central role in the Kohn-Sham DFT calculations. The exchange-correlation energy is a

relatively small part of the total energy in a typical system, but it is the largest part in nature's "glue" that binds atoms together. The exchange-correlation energy contains all the complexities of an interacting many-electron problem, but an exact numerical determination of $E_{XC}[\rho]$ is far too complicated beyond a few simple atoms (by exact diagonalization). In practical applications, approximate expressions for $E[\rho(\mathbf{r})]$ and V_{XC} have been created, each having its own deficiencies. Developing these functionals is a very active field of research but is beyond this thesis. In the next section, we will examine some general aspects of E_{XC} which may place constraints for developing approximate functionals.

2.3 Exchange-Correlation energy E_{XC}

The exchange-correlation effect that contains all the many-body physics of interacting electrons, arises from the fact that normally electrons can not move randomly but try to avoid each other in order to reduce the electron-electron Coulomb interaction. The exchange-correlation energy separated from the total energy above, consists of three different contributions (see Eq.(2.16)). The first term is the potential energy of exchange E_X. The exact exchange energy includes the effect of anti-symmetric wave functions due to the Pauli exclusion principle and it corrects the unphysical self-interactions included in the Hartree potential energy. The second term is the potential energy of correlation E_C that represents the effect of electron-electron repulsion on the interacting many-body wave function that is beyond the Hartree-Fock term. Both potential energies of exchange and correlation must be negative because interacting electron motion reduces the expectation value of electron-electron Coulomb repulsion, and any approximation must give negative values for them. The third term in the exchange-correlation is a smaller positive kinetic energy of correlation due to the extra swerving motion of electrons as they avoid one another. To achieve a deeper understanding about the exchange-correlation energy, we will introduce the concept of exchange-correlation hole that guides the construction of accurate approximations.

In the Khon-Sham DFT, the essential simplification lies in the link drawn between the interacting electrons and the non-interacting electrons. This link can be studied by varying the electron-electron interaction by changing a parameter λ from 0 to 1 in the following Hamiltonian [41],

$$H_\lambda = -\frac{1}{2}\nabla^2 + V_{ext}^\lambda + \lambda V_{ee} \tag{2.27}$$

where V_{ext}^λ is added to keep $\langle \Psi_\lambda \mid \hat{\rho} \mid \Psi_\lambda \rangle = \rho(\mathbf{r})$ and $\langle \Psi_\lambda \mid H_\lambda \mid \Psi_\lambda \rangle = E_{GS}$ during the variation, and Ψ_λ is the corresponding ground-state wave function of H_λ. Then the relation of exchange-correlation energy to the exchange-correlation hole can be found exactly using the integral over the coupling constant λ[42, 43, 44],

$$E_{XC} = \int_0^1 d\lambda \langle \Psi_\lambda \mid \frac{\partial H_\lambda}{\partial \lambda} \mid \Psi_\lambda \rangle - E_{Hartree} = \frac{1}{2}\int d\mathbf{r}\rho(\mathbf{r}) \int d\mathbf{r}' \frac{1}{\mid \mathbf{r} - \mathbf{r}' \mid} \rho_{XC}(\mathbf{r}, \mathbf{r}' - \mathbf{r}) \tag{2.28}$$

where

$$\rho_{XC}(\mathbf{r}, \mathbf{r}' - \mathbf{r}) = \rho(\mathbf{r}') \int_0^1 d\lambda [g(\mathbf{r}, \mathbf{r}', \lambda) - 1] \tag{2.29}$$

is defined as the exchange-correlation hole density. The function $g(\mathbf{r}, \mathbf{r}', \lambda)$ is the normalized pair correlation function of the system of H_λ. The exchange-correlation hole ρ_{XC} arises from the fact that the presence of electron at $\mathbf{r}$ must reduces the possibility of finding another one at $\mathbf{r}'$, thus it is always negative. We can see that the exchange-correlation energy is formulated in term of the potential energy resulting from the electron interacting with its exchange-correlation hole.

Three aspects should be stressed for the exchange-correlation hole. First, since $g(\mathbf{r}, \mathbf{r}' - \mathbf{r}, \lambda) - 1$ tends to be zero as $\mathbf{r}' - \mathbf{r} \to \infty$, we may expect the exchange-correlation energy to be the consequence of short-range effects of the Coulomb interaction. Therefore, it is reasonable to approximate E_{XC} as local or semi-local functional of density in form of,

$$E_{XC} = \int d\mathbf{r}\rho(\mathbf{r})\epsilon_{XC}([\rho], \mathbf{r}) \tag{2.30}$$

$\epsilon_{XC}(\rho, \mathbf{r})$ is the exchange-correlation energy density at $\mathbf{r}$, it is exactly given by ρ_{XC}

$$\epsilon_{XC}(\rho, \mathbf{r}) = \int d\mathbf{r}' \frac{1}{\mid \mathbf{r} - \mathbf{r}' \mid} \rho_{XC}(\mathbf{r}, \mathbf{r}' - \mathbf{r}) \tag{2.31}$$

Second, the definition of the pair correlation function promises that the exchange-correlation hole must obey a sum rule that it integrates to unity. This is the consequence of the fact that if one electron sits at $\mathbf{r}$, then there must be one electron missing in $\mathbf{r}'$ space. The sum rule is expressed as an integral over $\mathbf{r}'$

$$\int d\mathbf{r}' \rho_{XC}(\mathbf{r}, \mathbf{r}' - \mathbf{r}) = -1 \tag{2.32}$$

This sum rule is satisfied for any realistic electron system, it constrains and guides the development of accurate approximations: a construction that violates the sum rule can not be expected to work well. The sum rule has been used to explain why existing approximations for the exchange-correlation (XC) functionals actually work[44].

The last aspect of the XC functional arises from the isotropic nature of Coulomb interaction (translational and rotational invariance). This gives a significant consequence that E_{XC} only depends on the spherical average of n_{XC}. Thus, approximations can give rather accurate E_{XC} as long as the spherical part is close to reality - even though the approximations may be quite bad for the non-spherical part of the exchange-correlation hole.

In the case of spin polarization, the exact form for exchange-correlation energy can be expressed in term of spin density and spin dependent pair correlation function,

$$E_{XC} = \frac{1}{2} \sum_{\sigma,\sigma'} \int d\mathbf{r} \rho_\sigma(\mathbf{r}) \int d\mathbf{r}' \frac{1}{\mid \mathbf{r} - \mathbf{r}' \mid} \rho_{\sigma'}(\mathbf{r}') \int_0^1 d\lambda [g_{\sigma,\sigma'}(\mathbf{r}, \mathbf{r}', \lambda) - 1] \tag{2.33}$$

where σ and σ' denote spin indexes.

Except the three aspects discussed above for the exchange-correlation hole, the exact exchange and correlation energies also satisfy some other known properties

which may also be used to help developing approximate functionals. For example, the coordinate scaling conditions[45], the one electron limit, the Lieb-Oxford bound relation[46], and the derivative discontinuity of E_{XC} in insulators [47, 48, 49], have all been used to constrain the form of E_{XC}. Here we wish to emphasize the derivative discontinuity which is expressed with respect to the total electron number,

$$\Delta_{XC} = \frac{\delta E_{XC}}{\delta n} \mid_{N+\delta} - \frac{\delta E_{XC}}{\delta n} \mid_{N-\delta} = V_{XC}^{+} - V_{XC}^{-} \tag{2.34}$$

The origin of the derivative discontinuity is the fact that the nature of states change discontinuously as a function of $n(\vec{r})$ at the band gap, where $n(\vec{r})$ is the occupation. This derivative discontinuity requires the exchange-correlation potential for all electrons in the system to change by a constant amount when one electron is added. As a consequence, even with an accurate exchange-correlation potential, the Kohn-Sham band gap is still different from the true gap by an amount of Δ_{XC},

$$E_{gap} = \frac{\delta E}{\delta n} \mid_{N+\delta} - \frac{\delta E}{\delta n} \mid_{N-\delta} = \frac{\delta (T_0 + E_{XC})}{\delta n} \mid_{N+\delta} - \frac{\delta (T_0 + E_{XC})}{\delta n} \mid_{N-\delta} = \Delta_{KS} + \Delta_{XC} \tag{2.35}$$

where $\Delta_{KS} = \frac{\delta T_0}{\delta n} \mid_{N+\delta} - \frac{\delta T_0}{\delta n} \mid_{N-\delta}$ (derivative discontinuity in KS kinetic energy) is the KS band gap. Although the above equation is exact in the exact Kohn-Sham theory, because the amplitude of Δ_{XC} is unknown, we still have to ask the question: how much of the error in the prediction of band gap using approximate XC functionals such as the functionals in local density approximation [50, 44] and generalized gradient approximation [51] , is not the intrinsic error of the KS theory. This question might be answered to some extent by the improved XC functionals such as hybrid functionals [52, 53, 54] and self-interaction corrected functionals[55]. The mechanism for making progress in prediction of band gaps with DFT are under intense debate because of its scientific significance. Most recently, it has been explained that the incorrect band-gap prediction with most approximate functionals originates mainly from errors in describing systems with fractional charge [56, 57], and it is possible to have a functional which gives the correct band gap. In the following, we will present a brief

description for the different classes of the approximations in the XC functional.

We can group the popular XC functionals into three types: the local density approximations (LDA), the semi-local approximations (GGA, meta-GGA), and the non-local approximations known as orbital functionals (hybrid functionals[52], optimized effective potential (OEP) schemes [58], and LDA + U methods [59, 60]). In the case of weakly **inhomogeneous** electron gas, $E_{XC}[\rho(\vec{r})]$ can be approximated by the exchange-correlation energy of the **homogeneous** electron gas having a constant density ρ: namely replacing the constant ρ by $\rho(\vec{r})$ to describe the **inhomogeneous** system. This is the well known local density approximation (LDA)[50, 44] which is precise only in the limit of very slow-varying density. LDA has been proven rather successful for a wide range of materials but it suffers a number of deficiencies such as overestimating the binding energy, underestimating the bonding distance and band gaps in semiconductors and insulators. Systematic improvements of LDA can be made by adding more features when building the E_{XC} functional. Better description of reality can be expected by incorporating more constraints to E_{XC}.

The GGA functional[51] adds gradients of local density $\nabla\rho$ to E_{XC}. Meta-GGA includes the local kinetic energy density $\tau(\mathbf{r})$ and Laplacian of the local density $\nabla^2\rho$[61], in addition to $\nabla\rho$ and ρ, to E_{XC}. Using GGA, good results for geometries and total energies have been achieved although the band gap problem still remains. Higher accuracy should be expected from meta-GGA calculations. Recently, a new type of meta-GGA, called the modified Becke-Johnson exchange potential functional (MBJ)[62] has been reported to correcting the band gap in semiconductors and insulators, we have implemented this new exchange potential functional into our transport software package. To achieve more accurate results, some orbital dependent functionals have been developed by building the exact exchange (EXX) energy from the Hartree-Fock theory. This type of orbital functionals are known as the hybrid functionals and OEP-EXX[58] methods. Another type of orbital functionals are the self-interaction corrected functionals[55] which completely remove the self-interaction

error from the approximated density functional. This is critical for the description of the localized electrons and very useful to investigating the atomic limit of impurity atoms in solid state hosts. In particular, a particular approach called LDA (GGA)+U method has been widely used in strongly correlated material calculations with great success[59, 60]. This method incorporates the insight gained from studies of strong correlations using the Hubbard model Hamiltonian. It corrects errors in the local or semi-local functionals thus improves the description of correlation effects in the localized d- and f-bands. We have implemented the LDA+U[59, 60] method into our NEGF-DFT software package for transport calculations for investigating the quantum transport properties in the devices with strong on-site coulomb interaction.

2.4 Using KS-DFT as a mean field theory

As seen from Eq.(2.23), the total energy is not simply a sum of the KS eigenvalues ϵ_i and, as known, the many-body wave function is not the single Slater determinant of the KS single particle orbitals. Therefore we should not assign any physical meaning to the KS eigenvalues and eigenstates. The KS eigenvalues often bear only semi-quantitative representation for the true energy spectrum. However, by comparison to experimental data, in many situations the KS eigenvalues do provide a surprisingly good approximation to the true energy spectrum of extended systems, for instance DFT often results in very good agreement with experimental photoemission and inverse photoemission measurements[63]. Practically, using DFT the band structure calculations in solid state physics are significantly simplified to the calculations of the KS eigenvalues. Nevertheless, we should be aware of that KS-DFT, applied this way, is not rigorously the many-body ground-state theory but is self-consistent mean field theory. In the following, this mean-field nature of KS theory is emphasized again in the application of Keldysh non-equilibrium Green's function based density functional theory which form the basis of this thesis.

Fig.4.2 shows a representative structure of a two-probe nano-electronic device,

where a central device scattering region is contacted by two semi-infinite electrodes. Steady state electric current is driven by a bias voltage applied to the electrodes. In such a system, the applied bias breaks the time reversal symmetry and brings complications to theoretical treatment[9]. The NEGF-DFT method provides a first principle approach to calculate non-equilibrium electronic structure from which quantum transport properties is derived. In the NEGF-DFT approach[13], the KS equation is solved self-consistently using NEGF to populate the electronic states. Importantly, the description of devices under bias requires the physics of non-equilibrium quantum statistics but not a ground-state theory of total energy minimization. In particular, the NEGF-DFT equations cannot be solved by a variational principle but must be solved directly by the differential NEGF-DFT equations. The NEGF-DFT method is restored to exactly the same form of ground-state KS DFT theory when the device is set back to equilibrium at which time reversal symmetry is restored and fluctuation-dissipation theorem is satisfied. Finally, the mean-field nature of the KS equation is utilized in the NEGF-DFT quantum transport approach for the infinite two-probe system.

3

Tight-Binding Linear Muffin Tin Orbital Method

In this chapter, we review the tight binding linear Muffin Tin orbital method (LMTO) for self-consistently solving the Kohn-Sham DFT equations. The LMTO method has a long history[64] and is a very well established method in the DFT community. We choose this method for two main reasons. First, the tight binding version of the LMTO method can solve large number atoms, for instance we have solved transport problems involving several thousand atoms[4, 5] at nonequilibrium. No other DFT method can have such a computational ability for nonequilibrium transport analysis. Note that the phrase "tight binding" only means the method is atomic center based, and the method is fully self-consistent. Second, the mathematical procedure of the LMTO method most naturally allows us to implement CPA and the vertex correction theory (see Chapter 5).

3.1 Solving the Kohn-Sham equations

To solve the Kohn-Sham equation

$$H_{KS}\Psi_j(\mathbf{r}) = [-\Delta + V_{eff}(\mathbf{r})]\Psi_j(\mathbf{r}) = E_j\Psi_j(\mathbf{r}) \tag{3.1}$$

the KS orbital Ψ_j is expanded by a complete basis set $\{\chi_i\}$, namely:

$$\Psi_j(\mathbf{r}) = \sum_i c_{i,j}\chi_i \ . \tag{3.2}$$

The solution of the KS equation Eq.3.1 can be transformed to an eigenvalue problem of the secular equation

$$[H - EO]c = 0 \tag{3.3}$$

where the Hamiltonian matrix is defined as $H_{i,i'} = \langle\chi_i|-\Delta+V_{eff}|\chi_{i'}\rangle$ and the Overlap matrix is defined by $O_{i,i'} = \langle\chi_i|\chi_{i'}\rangle$. For a fast solution of Eq.3.3, the chosen basis set should be mathematically simple for the computation of the Hamiltonian and overlap matrix elements. Moreover, if reasonable accuracy can be guaranteed, it is computationally advantageous to reduce the size of basis set. After solving Eq.3.3, the electron density ρ is constructed from the eigenvectors using Eq.(2.22). Next, the new electron density is used to update the effective potential V_{eff} which forms the new KS equation of the next iteration step. This process is iterated until the KS equation is solved to self-consistency. The various methods for calculating electronic structure of matter may be distinguished by the way in which the basis function χ_i is chosen. One can use fixed basis functions such as the linear combination of atomic orbital (LCAO), Gaussian (LCGO), etc. One can also choose partial-wave basis functions such as the linear augmented plane wave (LAPW), Muffin Tin Orbital (MTO), etc. In this chapter, we will introduce a particular method which is the tight binding linear Muffin Tin orbital (TB-LMTO) method.

TB-LMTO is based on the Muffin Tin (MT) approximation of the crystal potential V_{eff}. In MT approximation, the crystal potential is approximated to be spherically symmetric inside the MT spheres centered at the individual nuclei, and a constant value V_{MTZ} in the interstitial region between the MT spheres[65]. This approximation to the real potential $V_{eff}(\mathbf{r})$ has high physical transparency for the crystal and gives very good accuracy. It should be mentioned that in practical calculations, even if a suitable basis set is used, it is still rather complex to achieve efficient convergence of the self-consistent iteration of the KS equation. Many important mathematical and computational issues such as matrix inversion and diagonalization, mixing schemes in the self-consistent iteration[66], parallel computation techniques, etc are of vital

importance for achieving efficient self-consistent computation.

3.2 Muffin Tin approximation for crystals

According to the MT approximation[65], for a single MT sphere, we can define the potential to be

$$V_{MT}(r_R) = v_R(r_R) - V_{MTZ} \qquad (r_R \leq s_R) \tag{3.4}$$

$$V_{MT}(r_R) = V_{MTZ} \qquad (r_R \geq s_R) \tag{3.5}$$

where $v_R(r_R)$ is the spherically symmetric potential inside the R-th atomic sphere, R denotes the nuclei site, r_R is the distance $\mid \vec{r} - \vec{R} \mid$, and s_R is the radius of the MT sphere. Hence, the KS Hamiltonian of a crystal in MT approximation can be written in the following form,

$$H_{KS} = -\Delta + \sum_R V_{MT}(r_R) \tag{3.6}$$

where the sum extends over the whole crystal. The corresponding KS equation becomes:

$$(H_{KS} - E)\Psi = (-\Delta + \sum_R V_{MT}(r_R) - E)\Psi = 0 \ . \tag{3.7}$$

Now, the question that we face is to choose a suitable basis set for solving such a KS equation with a simple MT potential. Within the MT geometry, the construction of the basis function χ_{RL} involves the choices for basis "head" and "tail" functions within the MT spheres, and a reasonably complete envelop functions in the interstitial region. In principle, from the minimal basis set consideration, one should achieve fast l convergence. The best choice of each function is a reasonable solution of the KS equation in a larger region of the system. Therefore, we seek the solutions of the KS equations inside each single MT sphere and in the interstitial region, based on these solutions we can construct MTO basis functions in the entire space.

The Schrödinger equation inside the MT sphere on site R reads:

$$[-\Delta + v_R(r_R) - E]\varphi(E, \mathbf{r}_R) = 0 \ . \tag{3.8}$$

It is not difficult to find the solution of this differential equation for any energy E. The solution can be given in a general form,

$$\varphi_{RL}(E, \vec{r}_R) = \varphi_{Rl}(E, r_R) Y_L(\hat{\mathbf{r}}_R) \tag{3.9}$$

Y_L is the spherical harmonics. The radial function $\varphi_{Rl}(E, r_R)$ can be obtained by numerical integration of the following radial Schrödinger equation inside the MT sphere,

$$[-\frac{d^2}{dr_R^2} + \frac{l(l+1)}{r_R^2} + v_R(r_R) - E] r_R \varphi_{Rl}(E, r_R) = 0 \tag{3.10}$$

The radial function can be normalized within the MT sphere,

$$\int_{s_R} \varphi_{Rl}^2(E, r_R) r_R^2 dr_R = 1 \tag{3.11}$$

and also satisfies the following orthogonality relation

$$\int_{s_R} \varphi_{Rl}(E, r_R) \dot{\varphi}_{Rl}(E, r_R) r_R^2 dr_R = 0 \tag{3.12}$$

where the functions $\varphi_{Rl}(E, r_R)$ and $\dot{\varphi}_{Rl}(E, r_R) = \partial \varphi_{Rl} / \partial E$ are both truncated inside the R-th sphere of radius s_R. φ_{Rl} and $\dot{\varphi}_{Rl}$ will be used for construction of the basis head and tail functions within the MT spheres in the TB-LMTO method. It can be proven that not only φ_{Rl} but also $\dot{\varphi}_{Rl}$ are orthogonal to the core electron states. This gives an important consequence that the basis set constructed from them are orthogonal to the core electron orbitals.

A good choice of envelop function can be the solution of the Schrödinger equation

in the interstitial region where the potential is flat,

$$[-\Delta - \kappa^2]\chi^e_{RL}(\kappa, \mathbf{r}_R) = 0 \tag{3.13}$$

Here the envelope function may be defined as atomic centered angular momentum functions χ^e_{RL}, $\kappa^2 = E - V_{MTZ}$ is the electron kinetic energy in the interstitial region between the MT spheres. For the above differential equation, there are generally two independent solutions which may be taken as the spherical Bessel and Neumann functions[65] for any energy. These exact solutions are all energy dependent functions which may be used for constructing energy dependent basis functions. However, in a linear method the basis function is required to be independent of energy, hence one takes a constant value of κ. We shall from now on make the simplest choice $\kappa = 0$, then the Schrödinger equation Eq.(3.13) is reduced to the Laplace equation,

$$-\Delta\chi^e_{RL}(\mathbf{r}_R) = 0 \tag{3.14}$$

This is a reasonable choice within the atomic sphere approximation (ASA) which will be introduced to eliminate the interstitial region by filling the space using overlapping spheres (see section 3.7).

3.3 The bare envelop function

We may first only consider the system in the absence of atomic MT spheres so that the Laplace equation is valid in the entire space. The atomic MT spheres will be put back by augmentation in Section 3.5. The Laplace equation of Eq.(3.14) is known to have two independent solutions. According to the asymptotic behavior at $r_R \to \infty$, the regular solution is given by

$$J^0_{RL}(\mathbf{r}_R) = (r_R/w)^l[2(2l+1)]^{-1}Y_L(\hat{\mathbf{r}}_R) \tag{3.15}$$

and the irregular solution is

$$K^0_{RL}(\mathbf{r}_R) = (r_R/w)^{-l-1} Y_L(\hat{\mathbf{r}}_R) \tag{3.16}$$

where the superscript 0 means the bare MTO representation, and w is a scaling factor which makes the two solutions dimensionless. The solution $K^0_{RL}(\mathbf{r}_R)$ is regular except at point R, and has the form of electrostatic field of 2^l poles at R. We may take this multi-pole field solution of $K_{RL}(\mathbf{r}_R)$ as the bare envelope function in the full space. The bare envelop function K^0_{RL} can be expanded into the irregular solutions of $J^0_{R'L}$ about any other site $R' \neq R$ [67], according to:

$$K^0_{RL}(\mathbf{r}_R) = -\sum_{L'} J^0_{R'L'}(\mathbf{r}_{R'}) \mathbf{S}^0_{R'L',RL} \tag{3.17}$$

where the expansion coefficients $\mathbf{S}^0_{R'L',RL}$ are the so-called bare canonical structure constant which is only determined by the lattice geometry of the crystal. The $\mathbf{S}^0$ matrix is independent of the crystal potential. The explicit form of S^0 matrix element is given by [67],

$$\mathbf{S}^0_{RL,R'L'} = \sum_{L''} (-1)^{l'+1} \frac{8\pi(2l''-1)!! C_{LL'L''}}{(2l-1)!!(2l'-1)!!} K^0(R'-R) \tag{3.18}$$

for $R \neq R'$. When $R = R'$, $\mathbf{S}^0_{RL,RL'}$ are equal to zero. The quantities $C_{LL'L''}$ are the Gaunt coefficients. The matrix $\mathbf{S}^0$ is a real and symmetric matrix, and the element $\mathbf{S}^0_{RL,R'L'}$ depends on the distance $\mid R - R' \mid$ according to a inverse power law:

$$\mathbf{S}^0_{RL,R'L'} \propto \left(\frac{1}{\mid R - R' \mid} \right)^{l+l'+1} \tag{3.19}$$

Since the bare envelop function of Eq.(3.16) is a long-range function which decays slowly by inverse power law of r_R, we shall introduce a screening procedure to obtain a localized envelope function which is desired for efficient computation. For this purpose, we introduce a separation of the space using the Wigner-Seitz (WS) cell

labeled by R, and we define the functions $|K^0_{RL}\rangle$ and $|J^0_{RL}\rangle$ to be truncated inside the WS cell around R, and the function $|K^0_{RL}\rangle^\infty$ extends over all space. Then, we can rewrite the bare envelop function in the full space $|K^0_{RL}\rangle^\infty$ in the following form,

$$|K^0_{RL}\rangle^\infty = |K^0_{RL}\rangle - \sum_{R'} \sum_{L'} |J^0_{R'L'}\rangle \mathbf{S}^0_{R'L',RL} \tag{3.20}$$

where the envelop function inside the WS of $R' \neq R$ is replaced by the expansion of Eq.(3.17). The above equation can be rewritten in a more compact format,

$$|K^0\rangle^\infty = |K^0\rangle - |J^0\rangle \mathbf{S}^0 \tag{3.21}$$

The localized envelop function can be obtained by screening the bare function with added multi-poles at the neighboring sites [32, 67] which we shall introduce in the next section.

3.4 Screened envelope function

We can modify the regular solution $|J^0_{RL}\rangle$ discussed in the last section by adding an amount $-\alpha_{Rl}$ of the irregular solution $|K^0_{RL}\rangle$ inside the R-th WS cell by defining:

$$|J^\alpha_{RL}\rangle \equiv |J^0_{RL}\rangle - \alpha_{Rl}|K^0_{RL}\rangle \tag{3.22}$$

where the constant α_{RL} is named screening constant (usually independent of R). The above equation is the α representation of MTO. In analogy with the bare envelope function of Eq.(3.21), we define the screened envelope functions in the full space as:

$$|K^\alpha\rangle^\infty = |K^0\rangle - |J^\alpha\rangle \mathbf{S}^\alpha \tag{3.23}$$

Where matrix $\mathbf{S}^{\alpha}$ is the screened structure constant which will be found in the following. Inserting $|J^{\alpha}\rangle$ into the bare envelope function of Eq.(3.21), we find:

$$|K^{0}\rangle^{\infty} = (|K^{0}\rangle - |J^{\alpha}\rangle\mathbf{S}^{0}[1-\alpha\mathbf{S}^{0}]^{-1})(1-\alpha\mathbf{S}^{0}) \tag{3.24}$$

Inserting $|J^{\alpha}\rangle$ into the screened envelope function, we find:

$$|K^{\alpha}\rangle^{\infty} = (|K^{0}\rangle - |J^{0}\rangle\mathbf{S}^{\alpha}[1+\alpha\mathbf{S}^{\alpha}]^{-1})(1+\alpha\mathbf{S}^{\alpha}) \tag{3.25}$$

Comparing the above two equations, we can find - if and only if

$$\mathbf{S}^{\alpha} = \mathbf{S}^{0}[1-\alpha\mathbf{S}^{0}]^{-1} \tag{3.26}$$

or equivalently

$$\mathbf{S}^{0} = \mathbf{S}^{\alpha}[1+\alpha\mathbf{S}^{\alpha}]^{-1} \tag{3.27}$$

$|K^{\alpha}\rangle^{\infty}$ is the superposition of $|K^{0}\rangle^{\infty}$ that is given by

$$|K^{\alpha}\rangle^{\infty} = |K^{0}\rangle^{\infty}(1+\alpha\mathbf{S}^{\alpha}) = |K^{0}\rangle^{\infty}(1+\alpha\mathbf{S}^{0})^{-1} \ . \tag{3.28}$$

So far, we have defined the envelope functions $|K^{\alpha}\rangle^{\infty}$ for a general α representation of MTO. An exponential decay can be achieved for $|K^{\alpha}\rangle^{\infty}$ with increasing distance. Thus the screened structure constant $\mathbf{S}^{\alpha}_{RL,R'L'}$ must decrease exponentially with the inter-atomic distance $\mid R - R' \mid$ as seen from Eq.(3.23). The fastest decay can be realized by choosing a suitable screening constant for each orbital l. In the most screened α [32] representation, the range of the screened structure constant $\mathbf{S}^{\alpha}$ can even be limited to the first- and second-nearest neighboring sites. Consequently, the screened envelope function provides a great advantage for fast computation. In the subsequent sections, we shall construct the TB-LMTO basis set $\mid \chi^{\alpha}_{RL}\rangle^{\infty}$ by augmenting the wave function $|K^{\alpha}_{RL}\rangle^{\infty}$ inside MT spheres. The LMTO is defined as the linear approximation to the energy-dependent MTO which we shall introduce in the next

section.

3.5 Augmentation to Energy Dependent MTO

To proceed further, we now use the MT geometry of the space instead of the WS cell separation of last section, so that the screened envelop function in Eq.(3.23) can be rewritten in the full space,

$$|K^{\alpha}\rangle^{\infty} = |K^{0}\rangle - |J^{\alpha}\rangle\mathbf{S}^{\alpha} + |K^{\alpha}\rangle^{i} \tag{3.29}$$

where the functions $|K^{0}_{RL}\rangle$ and $|J^{\alpha}_{RL}\rangle$ are truncated inside the MT spheres, and the wave function in the interstitial region is $|K^{\alpha}\rangle^{i}$ which is given by the difference:

$$|K^{\alpha}\rangle^{i} = |K^{\alpha}\rangle^{\infty} - [|K^{0}\rangle - |J^{\alpha}\rangle\mathbf{S}^{\alpha}] \tag{3.30}$$

The envelop function in Eq.(3.29) is continuous and differentiable in the full space. In the following, we will augment the wave function inside MT spheres using the solution of the Schrödinger equation in atomic sphere, and keep the envelop function inside the interstitial region unchanged.

As we have known from Section 3.2, the Schrödinger equation inside the atomic MT sphere on site R can be solved to obtain an energy-dependent regular solution $|\psi_{RL}(E, \mathbf{r}_R)\rangle$ which is truncated inside the MT sphere and normalized to unity within the MT sphere. To construct the energy-dependent MTO, we first augment the wave function $|J^{\alpha}_{RL}\rangle$ in Eq.(3.29) to some regular function $|\tilde{J}^{\alpha}_{RL}\rangle$, namely

$$|\ J^{\alpha}_{RL}\rangle \rightarrow |\ \tilde{J}^{\alpha}_{RL}\rangle \tag{3.31}$$

To keep the wave function continuous and differentiable at the sphere boundary s_R,

we need $|\tilde{J}^{\alpha}_{RL}\rangle$ to satisfy the following:

$$\tilde{J}^{\alpha}_{RL}(s_R) = J^{\alpha}_{RL}(s_R) \tag{3.32}$$

$$[\partial \tilde{J}^{\alpha}_{Rl}(r_R)/\partial r_R]_{s_R} = [\partial J^{\alpha}_{Rl}(r_R)/\partial r_R]_{s_R} \tag{3.33}$$

The wave function $|\tilde{J}^{\alpha}_{RL}\rangle$ will be determined in the next section. The wave function $\mid K^{0}_{RL}\rangle$ inside the MT sphere can be substituted by proper linear combination of the radial function $\varphi_{RL}(E)$ and $\tilde{J}^{\alpha}_{RL}$ in the following form,

$$K^{0}_{RL}(r_R) \rightarrow \varphi_{RL}(E, r_R) N^{\alpha}_{RL}(E) + \tilde{J}^{\alpha}_{RL}(r_R) P^{\alpha}_{RL}(E) \tag{3.34}$$

Here, the functions $N^{\alpha}_{RL}(E)$ and $P^{\alpha}_{RL}(E)$ can be determined using the condition of continuous and differentiable matching at s_R which can be easily realized by using the Wronskian [67],

$$W[f_1(r), f_2(r)] = r^2[f_1(r) f_2^{'}(r) - f_1^{'}(r) f_2(r)]_{r=S_R} \tag{3.35}$$

As a result, the function $P^{\alpha}_{Rl}(E)$ is given by,

$$P^{\alpha}_{RL}(E) = \frac{W[\varphi(E), K^0]_{RL}}{W[\varphi(E), \tilde{J}^{\alpha}]_{RL}} = \frac{W[\varphi(E), K^0]_{Rl}}{W[\varphi(E), J^{\alpha}]_{RL}} \tag{3.36}$$

which is the so-called potential function in the α representation, and the function $N^{\alpha}_{RL}(E)$, which serves as the normalization function, reads as:

$$N^{\alpha}_{RL}(E) = \frac{W[\tilde{J}^{\alpha}, K^0]_{RL}}{W[\tilde{J}^{\alpha}, \varphi(E)]_{RL}} = \frac{W[J^{\alpha}, K^0]_{RL}}{W[J^{\alpha}, \varphi(E)]_{RL}} = [(w/2)\dot{P}^{\alpha}_{Rl}(E)]^{1/2} \tag{3.37}$$

It is easy to see that by augmentation of the functions $\mid J^{\alpha}\rangle$ and $\mid K^{0}\rangle$ in the Eq.(3.29) using Eq.(3.31) and Eq.(3.34), the energy dependent MTO of α representation can be written in the full space in the following form,

$$|\chi^{\alpha}(E)\rangle^{\infty} = |\varphi(E)\rangle N^{\alpha}(E) + |\tilde{J}^{\alpha}\rangle(P^{\alpha}(E) - \mathbf{S}^{\alpha}) + |K^{\alpha}\rangle^{i} \tag{3.38}$$

where $|\rangle^{\infty}$ denotes that the function extends in all space, the function $|\rangle$ is truncated inside MT sphere, and the function $|\rangle^{i}$ is truncated inside the interstitial region.

3.6 Energy linearization

The function $\tilde{J}^{\alpha}$ in Eq.(3.38) may be defined in such a way that the energy dependence of $|\chi^{\alpha}(E)\rangle^{\infty}$ vanishes at the first order of $(E-\epsilon_v)$, where ϵ_v is the energy chosen at the center of interest, namely $|\dot{\chi}^{\alpha}(E)\rangle^{\infty}/_{\epsilon_v} = 0$. Thus we have

$$\begin{aligned}
&|\dot{\chi}^{\alpha}(E)\rangle^{\infty}/_{\epsilon_v} \\
&= [|\varphi(E)\rangle\dot{N}^{\alpha}(E) + |\dot{\varphi}(E)\rangle N^{\alpha}(E) + |\tilde{J}^{\alpha}\rangle\dot{P}^{\alpha}(E)]/_{\epsilon_v} \\
&= [|\dot{\varphi}^{\alpha}(E)\rangle N^{\alpha}(E) + |\tilde{J}^{\alpha}\rangle\dot{P}^{\alpha}(E)]/_{\epsilon_v} \\
&= [|\dot{\varphi}^{\alpha}\rangle N^{\alpha} + |\tilde{J}^{\alpha}\rangle\dot{P}^{\alpha}] = 0
\end{aligned} \tag{3.39}$$

where the omission of the energy variable means $E = \epsilon_v$. We thus can obtain the function,

$$|\tilde{J}^{\alpha}\rangle = -|\dot{\varphi}^{\alpha}\rangle N^{\alpha}/\dot{P}^{\alpha} = -|\dot{\varphi}^{\alpha}\rangle(w/2)/N^{\alpha} \tag{3.40}$$

Here, $|\dot{\varphi}^{\alpha}\rangle$ reads,

$$|\dot{\varphi}^{\alpha}\rangle = |\varphi\rangle o^{\alpha} + |\dot{\varphi}\rangle \tag{3.41}$$

where the quantity o^{α} is given by:

$$o^{\alpha} = \dot{N}^{\alpha}/N^{\alpha} = \langle\varphi|\dot{\varphi}^{\alpha}\rangle \tag{3.42}$$

Up to now, we see that the first energy derivative of $|\varphi\rangle$, namely $|\dot{\varphi}\rangle$, is also a component of the TB-LMTO basis function.

Letting $|\dot{\chi}^{\alpha}(E)\rangle^{\infty}/_{\epsilon_v} = 0$, the energy dependent MTO $|\chi^{\alpha}(E)\rangle^{\infty}$ can be expanded

around the energy center ϵ_v in the form of:

$$|\chi^{\alpha}(E)\rangle^{\infty} = |\chi^{\alpha}\rangle^{\infty} + (E-\epsilon_v)^2|\ddot{\chi}^{\alpha}\rangle^{\infty} + ... \tag{3.43}$$

where $|\chi^{\alpha}\rangle^{\infty} = |\chi^{\alpha}(\epsilon_v)\rangle^{\infty}$. Neglecting the $(E-\epsilon_v)^2$ and higher order terms, we can define $|\chi^{\alpha}\rangle^{\infty}$ as the energy independent MTO in the α representation. Therefore, after normalization, we have the TB-LMTO:

$$\begin{aligned} |\chi^{\alpha}\rangle^{\infty} &= |\chi^{\alpha}(\epsilon_v)\rangle^{\infty}/N^{\alpha} \\ &= |\varphi\rangle + |\dot{\varphi}^{\alpha}\rangle h^{\alpha} + |K^{\alpha}\rangle^{i}/N^{\alpha} \\ &= |\varphi\rangle(1+o^{\alpha}h^{\alpha}) + |\dot{\varphi}\rangle h^{\alpha} + |K^{\alpha}\rangle^{i}/N^{\alpha} \end{aligned} \tag{3.44}$$

where the matrix h^{α} reads as:

$$\begin{aligned} h^{\alpha} &= -\frac{w}{2}\frac{1}{N^{\alpha}}[P^{\alpha}-\mathbf{S}^{\alpha}]\frac{1}{N^{\alpha}} \\ &= -(\dot{P}^{\alpha})^{-1/2}[P^{\alpha}-\mathbf{S}^{\alpha}](\dot{P}^{\alpha})^{-1/2} \\ &= C^{\alpha} - \epsilon_v + (\Delta^{\alpha})^{1/2}\mathbf{S}^{\alpha}(\Delta^{\alpha})^{1/2} \end{aligned} \tag{3.45}$$

after using Eqs. (3.36), (3.37) and (3.40). In Eq.(3.45), only the screened structure constant matrix S^{α} is a non-diagonal matrix. The potential parameter C^{α}_{RL} reads,

$$C^{\alpha}_{RL} = \epsilon_{v,RL} - P^{\alpha}_{RL}/\dot{P}^{\alpha}_{RL} = \epsilon_{v,RL} - W[K,\varphi]_{RL}/W[K,\dot{\varphi}^{\alpha}]_{RL} \tag{3.46}$$

which determines the $Rl-th$ band position and hence is usually called the band center defined in the α representation. The second parameter Δ^{α} is defined by

$$\Delta^{\alpha}_{RL} = 1/\dot{P}^{\alpha}_{RL} = -(w/2)W[K,\dot{\varphi}^{\alpha}]^{-2}_{RL} \tag{3.47}$$

which represents the width and hybridization strength of the $Rl-th$ band in the α representation.

In Eq.(3.44), the envelope function $|K^\alpha\rangle^i$ (see Eq.(3.30)) can be rewritten to:

$$\begin{aligned}|K^\alpha\rangle^i &= |K^\alpha\rangle^\infty - |K^0\rangle + |J^\alpha\rangle\mathbf{S}^\alpha \\ &= |K^0\rangle^\infty(1+\alpha\mathbf{S}^\alpha) - |K^0\rangle + |J^\alpha\rangle\mathbf{S}^\alpha \\ &= (|K^0\rangle^\infty - |K^0\rangle)(1+\alpha\mathbf{S}^\alpha) + |J^0\rangle\mathbf{S}^\alpha \end{aligned} \tag{3.48}$$

according to the Sections 3.3 and 3.4.

Up to now, the only thing we may have questions about is the choice of the energy center ϵ_v for each basis function $|\chi^\alpha\rangle^\infty$. Generally speaking, ϵ_v should be chosen to suit the problem at hand. For example, one can use $\epsilon_{v,Rl} = E_F$ to give a correct Fermi surface and Fermi velocity, and use the center of the occupied part of RL band to achieve accurate charge density in the self-consistent DFT calculations.

As a summary for this section, we have found the general TB-LMTO basis set characterized by the screening constant α. The explicit basis function is expressed in full space by Eq.(3.44) in which the interstitial wave function is given in Eq.(3.48), and all the associated quantities have also been defined in proper detail above. In the following, we shall introduce the atomic sphere approximation to remove the integration over the interstitial region, which greatly simplifies the computation of the Hamiltonian and overlap matrices within the TB-LMTO method.

3.7 Atomic sphere approximation: Hamiltonian and overlap matrices

The atomic sphere approximation (ASA) [64, 65] uses overlapping atomic spheres to fill the entire space so that the interstitial envelop function $|K^\alpha\rangle^i$ in Eq.(3.44) drops out. Thus the basis function of all space within ASA becomes:

$$|\chi^\alpha\rangle^\infty_{ASA} = |\varphi\rangle(1+o^\alpha h^\alpha) + |\dot{\varphi}\rangle h^\alpha \tag{3.49}$$

Using the normalization of partial wave (see Eq.(3.11)) and the orthogonal relation between a partial wave and its first energy derivative (see Eq.(3.12)), the overlap matrix becomes:

$$O^{\alpha}_{ASA} =^{\infty}_{ASA} \langle\chi^{\alpha}|\chi^{\alpha}\rangle^{\infty}_{ASA} = (1+h^{\alpha}o^{\alpha})(1+o^{\alpha}h^{\alpha}) + h^{\alpha}ph^{\alpha} \tag{3.50}$$

where $p = \langle\dot{\varphi}|\dot{\varphi}\rangle$ is a small quantity. The last term is usually neglected in most applications, thus

$$O^{\alpha}_{ASA} \simeq (1+h^{\alpha}o^{\alpha})(1+o^{\alpha}h^{\alpha}) \tag{3.51}$$

Using the relation $[-\Delta + V(r_R) - \epsilon_v]\dot{\varphi}(E)/_{\epsilon_v} = \varphi$, we can also write down the Hamiltonian Matrix without difficulty,

$$\begin{aligned} H^{\alpha}_{ASA} &=^{\infty}_{ASA} \langle\chi^{\alpha}| - \Delta + v(r)|\chi^{\alpha}\rangle^{\infty}_{ASA} \\ &= (1+h^{\alpha}o^{\alpha})h^{\alpha} + (1+h^{\alpha}o^{\alpha})\epsilon_v(1+o^{\alpha}h^{\alpha}) + h^{\alpha}\epsilon_v p h^{\alpha} \end{aligned} \tag{3.52}$$

After neglecting the last small term containing p, we have

$$H^{\alpha}_{ASA} \simeq (1+h^{\alpha}o^{\alpha})h^{\alpha} + (1+h^{\alpha}o^{\alpha})\epsilon_v(1+o^{\alpha}h^{\alpha}) \tag{3.53}$$

We have so far obtained the Hamiltonian and overlap matrices in Eq.(3.53) and Eq.(3.51), respectively, within the TB-LMTO method in ASA. The ASA has been proven to work rather well not only for close packed solids, but also for open systems such as the diamond structure when vacuum spheres are used to fill the empty space [68]. The major error of ASA is due to neglecting the integration over the interstitial region. To correct this error, one can include the contribution of the interstitial region to the overlap and Hamiltonian matrices, using the so-called combined correction. The integral $O^{\alpha}_{i} =^{\infty} \langle\chi^{\alpha,i}|\chi^{\alpha,i}\rangle^{\infty} = \frac{1}{N^{\alpha}}{}^{i}\langle K^{\alpha}|K^{\alpha}\rangle^{i}\frac{1}{N^{\alpha}}$ gives us the correction to the overlap matrix, and hence the correction to the Hamiltonian matrix is $O^{\alpha}_{i}V_{MTZ}$.

The calculation of the matrix O_i is complicated [67]. It should be mentioned that if we include the combined correction term, the subsequent derivation of the NEGF formalism may become impossible within the TB-LMTO method. In practical applications, the ASA error can be minimized by minimizing the interstitial volume or the overlapping volume between spheres. In other words, we need to optimize the choice of the radii of the atomic spheres in the system to overcome the error introduced by ASA.

3.8 Transformation to the nearly orthogonal basis

From the particular form of the overlap matrix in Eq.(3.51), we can introduce the nearly orthogonal basis set by the following transformation in ASA,

$$|\chi^{orth}\rangle^{\infty} = |\chi^{\alpha}\rangle^{\infty}(1+o^{\alpha}h^{\alpha})^{-1} = |\varphi\rangle + |\dot{\varphi}\rangle h^{\alpha}(1+o^{\alpha}h^{\alpha})^{-1} \tag{3.54}$$

Then the overlap matrix becomes unit matrix after neglecting $p = \langle\dot{\varphi}|\dot{\varphi}\rangle$,

$$O^{orth}_{ASA} =^{\infty} \langle\chi^{orth}|\chi^{orth}\rangle^{\infty} \doteq 1 \tag{3.55}$$

The Hamiltonian matrix in nearly orthogonal representation is given by:

$$H^{orth}_{ASA} =^{\infty} \langle\chi^{orth}| - \Delta + v(r)|\chi^{orth}\rangle^{\infty} \doteq \epsilon_v + h^{\alpha}(1+o^{\alpha}h^{\alpha})^{-1} \tag{3.56}$$

Inserting $h^{\alpha} = C^{\alpha} - \epsilon_v + (\Delta^{\alpha})^{1/2}\mathbf{S}^{\alpha}(\Delta^{\alpha})^{1/2}$ and according to the previous definitions of the associated potential parameters C^{α} in Eq.(3.46), Δ^{α} in Eq.(3.47) and o^{α} in Eq.(3.42), the H^{orth}_{ASA} can be simplified to the following form:

$$H^{orth}_{ASA} = C + \sqrt{\Delta}\mathbf{S}^{\alpha}[1-(\gamma-\alpha)\mathbf{S}^{\alpha}]^{-1}\sqrt{\Delta} \tag{3.57}$$

where the potential parameters C, Δ and γ are diagonal matrices and are respectively

defined as,

$$C_{RL} \equiv \epsilon_{v,RL} - W\{K,\varphi\}_{RL}/W\{K,\dot{\varphi}\}_{RL} \tag{3.58}$$

$$\Delta_{RL} \equiv -(w/2)[W\{K,\dot{\varphi}\}_{RL}]^{-2} \tag{3.59}$$

$$\gamma_{RL} \equiv W\{J,\dot{\varphi}\}_{RL}/W\{K,\dot{\varphi}\}_{RL} \tag{3.60}$$

The potential parameters C_{RL}, γ_{RL} and Δ_{RL} are independent of the α constant, they represent the center, width and distortion of the $RL-th$ band respectively. By applying the definition of $\mathbf{S}^{\alpha}$ (see Eq.(3.26)), H^{orth}_{ASA} can be equivalently expressed as the following:

$$H^{orth}_{ASA} = C + \sqrt{\Delta}\mathbf{S}^{0}[1-\gamma\mathbf{S}^{0}]^{-1}\sqrt{\Delta} = C + \sqrt{\Delta}\mathbf{S}^{\gamma}\sqrt{\Delta} \tag{3.61}$$

Hence we see that H^{orth}_{ASA} is independent of the screening constant α.

To conclude this section, we have obtained the nearly orthogonal LMTO basis set in Eq.(3.54), and the corresponding Hamiltonian matrix in Eq.(3.57) or equivalently in Eq.(3.61). Consequently, the KS equation Eq.(3.1) can be solved self-consistently with the TB-LMTO method such that electronic structures of matter is obtained from atomic first principles.

As seen from the derivation procedure in previous sections, the error in TB-LMTO method has three different origins. The first is due to the linearization of the energy dependent MTO inside the atomic spheres. The linearized basis is accurate to first order of $(E-\epsilon_v)$, thus the ASA Hamiltonian matrix of Eq.(3.57) is only accurate to $(E-\epsilon_v)^2$, yielding eigenvalues accurate up to $(E-\epsilon_v)^2$. The second origin of error arises from the choice of the envelop function by using solutions of the Laplace equation Eq.(3.14), this is actually equivalent to the zero-th order approximation of $\kappa^2 = E - V_{VTZ}$ in the envelop function, or approximating the kinetic energy k^2 in the interstitial region to zero. This gives rise to the energy independent structure constant in the Hamiltonian matrix, although ASA removes the interstitial region.

The third source of error lies in the atomic sphere approximation which removes the interstitial integration from the Hamiltonian and overlap matrix, and approximates the real potential and charge density to be spherically symmetrical. This implies that it is much more important to choose the sphere radii in an ASA calculation than in a full potential calculation. Many applications have amply proven that ASA is a good approximation for crystals with symmetries. Moreover, the variational approach allows one to go beyond ASA by incorporating interstitial integral and non-spherical potential. The TB-LMTO with ASA provides very efficient practical computation and, at the same time, very good accuracy. It will be used in this work for quantum transport.

3.9 Summary

To summarize, we have presented the TB-LMTO in ASA for self-consistent solution of the KS equation. In particular, starting from the Muffin Tin approximation to the crystal potential, we presented the mathematical procedure for constructing the general TB-LMTO characterized by α in full space (Sections 3.3-3.6). We introduced ASA to simplify the computation of the Hamiltonian and overlap matrices by eliminating the interstitial region. Finally the nearly orthogonal LMTO in ASA is defined and the corresponding Hamiltonian matrix is obtained. This TB-LMTO-ASA method is particularly useful and reasonably accurate for describing crystals with symmetries. The most important property of this method lies in its simple Hamiltonian matrix elements in which all quantities, such as potential parameters and the structure constant, are defined with very explicit physical meaning (see Eq.(3.61)). The final results obtained in this chapter are the nearly orthogonal LMTO in Eq.(3.54) and the corresponding Hamiltonian matrix in Eq.(3.57) or Eq.(3.61), which will serve as the starting point for our subsequent derivations. In subsequent Chapters, we shall introduce Green's function formalism within the TB-LMTO-ASA method.

4

The non-equilibrium Green's function formalism

In this Chapter, we review the Keldysh Non-equilibrium Green's Function formalism (NEGF) for treating quantum transport in nano-scale devices. The NEGF formalism [9, 10, 11] provides a conceptually simple and computationally powerful approach for taking into account the non-equilibrium quantum statistics for devices under finite bias. It has been widely used in the community of device simulation.

4.1 Basic definitions

For future use, here are the definitions of the retarded, advanced, lesser and greater Green's functions:

$$G^R(\mathbf{r},t,\mathbf{r}',t') = -i\theta(t-t')\langle\{\hat{\psi}(\mathbf{r},t)\hat{\psi}^+(\mathbf{r}',t')\}\rangle \tag{4.1}$$

$$G^A(\mathbf{r},t,\mathbf{r}',t') = i\theta(t'-t)\langle\{\hat{\psi}(\mathbf{r},t)\hat{\psi}^+(\mathbf{r}',t')\}\rangle \tag{4.2}$$

$$G^<(\mathbf{r},t,\mathbf{r}',t') = i\langle\hat{\psi}^+(\mathbf{r}',t')\hat{\psi}(\mathbf{r},t)\rangle \tag{4.3}$$

$$G^>(\mathbf{r},t,\mathbf{r}',t') = -i\langle\hat{\psi}(\mathbf{r},t)\hat{\psi}^+(\mathbf{r}',t')\rangle \tag{4.4}$$

where $\theta(x)$ is the step function, $\theta(x) = 1$ for $x \geq 0$ and $\theta(x) = 0$ otherwise. The operators $\hat{\psi}^\dagger$ and $\hat{\psi}$ are the creation and annihilation field operators for the electrons. The retarded Green's function G^R is nonzero only for $t \geq t'$ and it describes the retarded response at time t to an earlier perturbation of the system at time t'. The

advanced Green's function G^A differs from zero only if $t \leq t'$, and provides the advanced response to a later perturbation. The lesser Green's function $G^<$ provides the distribution function for particles, and the greater Green's function that for the holes. From the above definitions, we see that these four Green's functions are not independent and they satisfy:

$$G^R - G^A = G^> - G^< \tag{4.5}$$

and

$$G^A = [G^R]^\dagger \tag{4.6}$$

It is easy to find the important relation by which the lesser Green's function is directly linked to the electron density,

$$n(\mathbf{r}, t) = -iG^<(\mathbf{r}, t, \mathbf{r}, t) \tag{4.7}$$

The calculation of $G^<$ is the ultimate goal of this chapter.

Generally, the equation of motion for the field operator $\hat{\psi}^\dagger(\mathbf{r}, t)$ is:

$$i\frac{\partial\hat{\psi}(\mathbf{r}, t)}{\partial t} = \hat{\psi}(\mathbf{r}, t)H - H\hat{\psi}(\mathbf{r}, t) \tag{4.8}$$

where H is the Hamiltonian which we shall assume to not depend on t explicitly, and $\hat{\psi}(\mathbf{r}, t)$ can be written as $e^{iHt/\hbar}\hat{\psi}(\mathbf{r})e^{-iHt/\hbar}$ in the Heisenberg picture.

Because our goal is to find self-consistent solutions to the KS equation, we will only concentrate on deriving relations in the mean field approximation where the Hamiltonian can be written in terms of one-body operators:

$$H = -\nabla^2 + V(\mathbf{r}) \tag{4.9}$$

Equivalently, H can be expressed in term of field operators:

$$H = \int d\mathbf{r}' \hat{\psi}^+(\mathbf{r}')[-\nabla^2 + V(\mathbf{r}')]\hat{\psi}(\mathbf{r}') \tag{4.10}$$

Thus we can write,

$$\psi(\mathbf{r})H = \int d\mathbf{r}' \hat{\psi}(\mathbf{r})\hat{\psi}^+(\mathbf{r}')[-\nabla^2 + V(\mathbf{r}')]\hat{\psi}(\mathbf{r}') \tag{4.11}$$

and

$$\begin{aligned} H\psi(\mathbf{r}) &= \int d\mathbf{r}' \hat{\psi}^+(\mathbf{r}')[-\nabla^2 + V(\mathbf{r}')]\hat{\psi}(\mathbf{r}')\hat{\psi}(\mathbf{r}) \\ &= -\int d\mathbf{r}' \hat{\psi}^+(\mathbf{r}')[-\nabla^2 + V(\mathbf{r}')]\hat{\psi}(\mathbf{r})\hat{\psi}(\mathbf{r}') \\ &= -\int d\mathbf{r}' \hat{\psi}^+(\mathbf{r}')\hat{\psi}(\mathbf{r})[-\nabla^2 + V(\mathbf{r}')]\hat{\psi}(\mathbf{r}') \end{aligned} \tag{4.12}$$

by using the relation $\hat{\psi}(\mathbf{r})\hat{\psi}(\mathbf{r}') + \hat{\psi}(\mathbf{r}')\hat{\psi}(\mathbf{r}) = 0$. From the above, we find,

$$\hat{\psi}(\mathbf{r})H - H\hat{\psi}(\mathbf{r}) = \int d\mathbf{r}'[\hat{\psi}(\mathbf{r})\hat{\psi}^+(\mathbf{r}') + \hat{\psi}^+(\mathbf{r}')\hat{\psi}(\mathbf{r})][-\nabla^2 + V(\mathbf{r}')]\hat{\psi}(\mathbf{r}') \tag{4.13}$$

$$= \int d\mathbf{r}'\delta(\mathbf{r} - \mathbf{r}')[-\nabla^2 + V(\mathbf{r}')]\hat{\psi}(\mathbf{r}') \tag{4.14}$$

$$= [-\nabla^2 + V(\mathbf{r})]\hat{\psi}(\mathbf{r}) \tag{4.15}$$

Finally, the Heisenberg equation of motion Eq.(4.8) can be rewritten for the mean field system as follows

$$i\frac{\partial\hat{\psi}(\mathbf{r},t)}{\partial t} = e^{iHt/\hbar}[\hat{\psi}(\mathbf{r})H - H\hat{\psi}(\mathbf{r})]e^{-iHt/\hbar} = e^{iHt/\hbar}[H\hat{\psi}(\mathbf{r})]e^{-iHt/\hbar} = H\hat{\psi}(\mathbf{r},t) \tag{4.16}$$

We will start from this equation of motion to derive the lesser Green's function for the two-probe devices discussed in Section 4.2. In particular, the retarded Green's function can be proven to satisfy the following equation for one-body Hamiltonian:

$$[i\frac{\partial}{\partial t} - H]G^R(\mathbf{r},t,\mathbf{r}',t') = \delta(\mathbf{r} - \mathbf{r}')\delta(t - t') \tag{4.17}$$

When there is no explicit time dependence in the Hamiltonian, the Green's function depends only on the time difference, namely $G(\mathbf{r}, t, \mathbf{r}', t') = G(\mathbf{r}, \mathbf{r}', t-t')$. In this case, one usually prefers to work in energy space by Fourier transforming the variable $(t-t')$ to energy E:

$$G(\mathbf{r}, \mathbf{r}', E) = \int_{-\infty}^{+\infty} G(\mathbf{r}, \mathbf{r}', t-t') e^{iE(t-t')} dt \tag{4.18}$$

It is straightforward to prove that energy dependent retarded Green's function satisfies the following equation:

$$[E^+ - H(\mathbf{r})]\mathbf{G^R}(\mathbf{r}, \mathbf{r}', \mathbf{E}) = \delta(\mathbf{r} - \mathbf{r}') \tag{4.19}$$

where $E^+ \equiv E + i\eta$ and η is a positive infinitesimal number.

If the system has translational invariance, Green's function only depends on the difference $\mathbf{r} - \mathbf{r}'$, $G(\mathbf{r}, \mathbf{r}', E) = G(\mathbf{r} - \mathbf{r}', E)$. For this situation, one usually prefers to work in k-space by Fourier transforming the variable $\mathbf{r} - \mathbf{r}'$ to momentum $\mathbf{k}$,

$$G(\mathbf{k}, E) = \int d\mathbf{r} G(\mathbf{r} - \mathbf{r}', E) e^{i\mathbf{k}(\mathbf{r}-\mathbf{r}')} \tag{4.20}$$

Here, $G(\mathbf{k}, E)$ satisfies:

$$[E^+ - H(\mathbf{k})]G^R(\mathbf{k}, E) = \mathbf{I} \tag{4.21}$$

where $\mathbf{I}$ is a diagonal matrix.

A very useful relation, i.e. the fluctuation-dissipation theorem, is worth mentioning here. If we define a spectral function $A(\mathbf{k}, E)$ as:

$$A(\mathbf{k}, E) = i[G^R(\mathbf{k}, E) - G^A(\mathbf{k}, E)] = i[G^>(\mathbf{k}, E) - G^<(\mathbf{k}, E)] \ , \tag{4.22}$$

then, at equilibrium, the lesser Green's function satisfies the following equation[11]

$$G^<(\mathbf{k}, E) = if(E)A(\mathbf{k}, E) \tag{4.23}$$

where $f(E)$ is the Fermi-Dirac distribution function. This relation is the fluctuation-dissipation theorem which is the most important property of the equilibrium theory. The fluctuation-dissipation theorem provides great conveniences for the calculation of equilibrium systems because, from the above, we obtain

$$G^<(\mathbf{k}, E) = f(E)[G^A(\mathbf{k}, E) - G^R(\mathbf{k}, E)] = f(E)[G^{R,+}(\mathbf{k}, E) - G^R(\mathbf{k}, E)] \quad (4.24)$$

Since the retarded Green's function is analytical on the upper half energy plane, the above relation greatly simplifies the theoretical calculation of $G^<$ for equilibrium situations (see later). In particular, when we compute the equilibrium density matrix from Eq.(1.2), the energy integration can be completed by a contour going through the upper half plane - avoiding the many poles of the Green's functions on the real energy axis.

4.2 Non-equilibrium Green's Function

In this Section, we formulate the NEGF formalism for two-probe device structures as shown in Fig.4.1, in which the central device scattering region is sandwiched by two semi-infinite electrodes. These electrodes extend to electron reservoirs at $z = \pm\infty$. The two reservoirs are at equilibrium having their respective electrochemical potentials μ_l and μ_r. The electron current flowing through the central scattering region is driven by the difference of the chemical potentials, namely by the bias voltage $eV_b = \mu_l - \mu_r$. For calculating the material properties, we only consider the mean field Hamiltonian which is consistent with the Kohn-Sham DFT, and neglect any inelastic scattering process.

To derive NEGF equations, we start from Eq.(4.16), namely

$$i\frac{\partial\hat{\psi}(\mathbf{r}, t)}{\partial t} = H\hat{\psi}(\mathbf{r}, t) \quad (4.25)$$

Dividing the two-probe system into three parts: left and right electrodes and central device scattering region, the Hamiltonian of the device becomes

$$H = H_l + \tau_l + \tau_l^+ + H_c + \tau_r + \tau_r^+ + H_r \tag{4.26}$$

Here, as shown in Fig.4.1, terms H_c, H_l and H_r describe the central device region, the

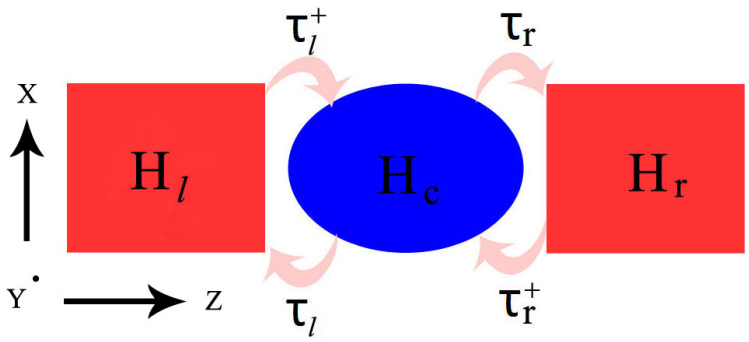

Figure 4.1: Schetch of a two probe device. Here, H_c, H_l and H_r describe the central device region, the left and right electrodes respectively; τ_l, τ_l^+, τ_r and τ_r^+ correspond to the interactions between the electrodes and the central device region. The electrodes extend to $z = \pm\infty$.

left and right electrodes respectively; τ_l, τ_l^+, τ_r and τ_r^+ correspond to the interactions between the electrodes and the central region. The field operator is separated as,

$$\hat{\psi}(\mathbf{r},t) = \begin{pmatrix} \hat{\psi}_l(\mathbf{r},t) \\ \hat{\psi}_c(\mathbf{r},t) \\ \hat{\psi}_r(\mathbf{r},t) \end{pmatrix} \tag{4.27}$$

Therefore we can directly rewrite the equation of motion Eq.(4.25) in a matrix format,

$$i\frac{\partial}{\partial t}\begin{pmatrix} \hat{\psi}_l \\ \hat{\psi}_c \\ \hat{\psi}_r \end{pmatrix} = \begin{pmatrix} H_l & \tau_l^+ & 0 \\ \tau_l & H_c & \tau_r \\ 0 & \tau_r^+ & H_r \end{pmatrix}\begin{pmatrix} \hat{\psi}_l \\ \hat{\psi}_c \\ \hat{\psi}_r \end{pmatrix} \tag{4.28}$$

From the above equation, it is straightforward to write down the following coupled equations,

$$i\frac{\partial \hat{\psi}_l}{\partial t} = H_l\hat{\psi}_l + \tau_l^+\hat{\psi}_c \tag{4.29}$$

$$i\frac{\partial\hat{\psi}_c}{\partial t} = \tau_l\hat{\psi}_l + H_c\hat{\psi}_c + \tau_r\hat{\psi}_r \tag{4.30}$$

$$i\frac{\partial\hat{\psi}_r}{\partial t} = H_r\hat{\psi}_r + \tau_r^+\hat{\psi}_c \tag{4.31}$$

Because the electrodes are assumed at equilibrium where the fluctuation-dissipation theorem Eq.(4.24) holds, our task is to find the lesser Green's function for describing the central scattering region of the device which is out of equilibrium. Therefore, we shall solve for $\hat{\psi}_c$ by eliminating electrode field operators $\hat{\psi}_l$ and $\hat{\psi}_r$. This can be done by the above three differential equations Eq.(4.29)-Eq.(4.30).

For Eq.(4.29), it is easy to write down the solution for the left electrode,

$$\hat{\psi}_l(\mathbf{r},t) = \hat{\Psi}_l(\mathbf{r},t) + \int d\mathbf{r}' \int_{-\infty}^{+\infty} dt' g_{ll}^R(\mathbf{r},t,\mathbf{r}',t')\tau_l^+\hat{\psi}_c(\mathbf{r}',t') \tag{4.32}$$

where Ψ_l is given by

$$i\frac{\partial\hat{\Psi}_l}{\partial t} = H_l\hat{\Psi}_l \tag{4.33}$$

which represents the field operator in the isolated left electrode before attaching to the central scattering region of the device. The Green's function g_{ll}^R satisfies the following differential equation,

$$[i\frac{\partial}{\partial t} - H_l(\mathbf{r})]g_{ll}^R(\mathbf{r},t,\mathbf{r}',t') = \delta(\mathbf{r}-\mathbf{r}')\delta(t-t') \ . \tag{4.34}$$

Similarly, for the right electrode, we can write down the solution according to Eq.(4.31),

$$\hat{\psi}_r(\mathbf{r},t) = \hat{\Psi}_r(\mathbf{r},t) + \int d\mathbf{r}' \int_{-\infty}^{+\infty} dt' g_{rr}^R(\mathbf{r},t,\mathbf{r}',t')\tau_r^+\hat{\psi}_c(\mathbf{r}',t') \tag{4.35}$$

where $\hat{\Psi}_r$ and g_{rr}^R are defined in the same way as that of left electrode in Eq.(4.33) and Eq.(4.34) respectively .

To solve the Green's function of the central region, inserting Eq.(4.32) and Eq.(4.35)

into Eq.(4.30), we obtain

$$i\frac{\partial\hat{\psi}_c(\mathbf{r},t)}{\partial t} - H_c\hat{\psi}_c(\mathbf{r},t) - \int d\mathbf{r}' \int_{-\infty}^{+\infty} dt'\Sigma(\mathbf{r},t,\mathbf{r}',t')\hat{\psi}_c(\mathbf{r}',t') = \hat{S} \tag{4.36}$$

where

$$\Sigma(\mathbf{r},t,\mathbf{r}',t') = \tau_l g_{ll}^R(\mathbf{r},t,\mathbf{r}',t')\tau_l^+ + \tau_r g_{rr}^R(\mathbf{r},t,\mathbf{r}',t')\tau_r^+ \tag{4.37}$$

$$\hat{S}(\mathbf{r},t) = \tau_l\hat{\Psi}_l(\mathbf{r},t) + \tau_r\hat{\Psi}_r(\mathbf{r},t) \tag{4.38}$$

Now, let's define the Green's function $G_{cc}^R(\mathbf{r},t,\mathbf{r}',t')$ by the following differential equation,

$$\begin{aligned} &i\frac{\partial G_{cc}^R(\mathbf{r},t,\mathbf{r}',t')}{\partial t} - H_c G_{cc}^R(\mathbf{r},t,\mathbf{r}',t') \\ &- \int d\mathbf{r}'' \int_{-\infty}^{+\infty} dt''\Sigma^R(\mathbf{r},t,\mathbf{r}'',t'')G_{cc}^R(\mathbf{r}'',t'',\mathbf{r}',t') = \delta(\mathbf{r}-\mathbf{r}')\delta(t-t') \end{aligned} \tag{4.39}$$

and define a new field operator $\hat{\Psi}_c(\mathbf{r},t)$ satisfying,

$$i\frac{\partial\hat{\Psi}_c(\mathbf{r},t)}{\partial t} - H_c\hat{\Psi}_c(\mathbf{r},t) - \int d\mathbf{r}' \int_{-\infty}^{+\infty} dt'\Sigma^R(\mathbf{r},t,\mathbf{r}',t')\hat{\Psi}_c(\mathbf{r}',t') = 0 \tag{4.40}$$

Here $\hat{\Psi}_c(\mathbf{r},t)$ is decoupled from the electrodes compared to Eq.(4.36): it represents electrons in the central region of the device. The solution for the central region can be expressed in term of $\hat{\Psi}_c$ and G_{cc},

$$\hat{\psi}_c(\mathbf{r},t) = \hat{\Psi}_c(\mathbf{r},t) + \int d\mathbf{r}' \int_{-\infty}^{+\infty} dt' G_{cc}(\mathbf{r},t,\mathbf{r}',t')\hat{S}(\mathbf{r}',t') \tag{4.41}$$

It is easy to check that the above solution satisfies Eq.(4.36). The second term in Eq.(4.41) includes the contribution of electrons coupled to the electrodes. According to the definition of Eq.(4.3), the lesser Green's function is thus obtained as,

$$\begin{aligned} &G_{cc}^<(\mathbf{r},t,\mathbf{r}',t') = i < \hat{\psi}_c(\mathbf{r}',t')^+\hat{\psi}_c(\mathbf{r},t) >= i < \hat{\Psi}_c^+(\mathbf{r}',t')\hat{\Psi}_c(\mathbf{r},t) > \\ &+ \int d\mathbf{r}''dt'' \int dt'''d\mathbf{r}'''G_{cc}^{R,*}(\mathbf{r}',t',\mathbf{r}'',t'')G_{cc}^R(\mathbf{r},t,\mathbf{r}''',t''')i < \hat{S}^+(\mathbf{r}'',t'')\hat{S}(\mathbf{r}''',t''') > \end{aligned} \tag{4.42}$$

$$=< \hat{\Psi}_c^+ \hat{\Psi}_c > + \int d\mathbf{r}''dt'' \int dt'''d\mathbf{r}''' G^R_{cc}(\mathbf{r}, t, \mathbf{r}''', t''')\Sigma^<(\mathbf{r}''', t''', \mathbf{r}'', t'')G^A_{cc}(\mathbf{r}'', t'', \mathbf{r}', t')$$

Because $< \hat{\Psi}_c(\mathbf{r}, t)^+ \hat{S}(\mathbf{r}', t') >=< \hat{S}^+(\mathbf{r}', t')\hat{\Psi}_c(\mathbf{r}, t) >= 0$, the isolated systems do not respond to each other. Here $\Sigma^<(\mathbf{r}''', t''', \mathbf{r}'', t'') = i < \hat{S}^+(\mathbf{r}'', t'')\hat{S}(\mathbf{r}''', t''') >$ is the lesser self-energy which will be described in detail in next section, and we have used $G^A_{cc}(\mathbf{r}'', t'', \mathbf{r}', t') = G^{R,*}_{cc}(\mathbf{r}', t', \mathbf{r}'', t'')$.

Now, we may want to know the contribution of the bound states, namely the first term in Eq.(4.42). To derive an expression for $< \hat{\Psi}_c^+ \hat{\Psi}_c >$, we start from the definition of a new Green's function $G_{cc0}(\mathbf{r}, t, \mathbf{r}', t')$ that describes the completely isolated central region (i.e. no electrodes), comparing with Eq.(4.39),

$$i\frac{\partial G^R_{cc0}(\mathbf{r}, t, \mathbf{r}', t')}{\partial t} - H_c G^R_{cc0}(\mathbf{r}, t, \mathbf{r}', t') = \delta(\mathbf{r} - \mathbf{r}')\delta(t - t') \tag{4.43}$$

and the corresponding field operator $\Psi_{c0}(\mathbf{r}, t)$ satisfies

$$i\frac{\partial \hat{\Psi}_{c0}(\mathbf{r}, t)}{\partial t} - H_c \hat{\Psi}_{c0}(\mathbf{r}, t) = 0 \tag{4.44}$$

It can be easily found that the following $\hat{\Psi}_c$ satisfies Eq.(4.40),

$$\hat{\Psi}_c(\mathbf{r}, t) = \hat{\Psi}_0(\mathbf{r}, t) + \int d\mathbf{r}'dt' \int d\mathbf{r}''dt'' G^R_{cc0}(\mathbf{r}, t, \mathbf{r}'', t'')\Sigma^R(\mathbf{r}'', t'', \mathbf{r}', t')\hat{\Psi}_c(\mathbf{r}', t') \tag{4.45}$$

For convenience, in the following we shall rewrite all the quantities and operators in the form of matrices [in the space of $(\mathbf{r}, t)$]. For example, Eq.(4.45) can be rewritten as;

$$\hat{\Psi}_c = \hat{\Psi}_0 + \hat{\Psi}_c G^R_{cc0}\Sigma^R \tag{4.46}$$

G_{cc} and G_{cc0} can be connected by the Dyson equation,

$$G^R_{cc} = G^R_{cc0} + G^R_{cc0}\Sigma^R G^R_{cc} \tag{4.47}$$

With above relations, we can find

$$i < \hat{\Psi}_c^+ \hat{\Psi}_c >= (1 + G_{cc}^R \Sigma^R) G_{cc0}^< (\Sigma^A G_{cc}^A + 1) \tag{4.48}$$

where $G_{cc0}^< = i < \hat{\Psi}_{c0}^+ \hat{\Psi}_{c0} >$ is the lesser Green's function of the isolated central region of the device.

Finally, we can write down the complete form of the lesser Green's function for the central region of interest[11],

$$G_{cc}^< = (1 + G_{cc}^R \Sigma^R) G_{cc0}^< (\Sigma^A G_{cc}^A + 1) + G_{cc}^R \Sigma^< G_{cc}^A \tag{4.49}$$

where all the quantities have already been defined in detail above. This is the well known Keldysh equation. It can be proven that the first term of Eq.(4.49) is nonzero only at true bound states. But for open systems, the bound states in the central region acquire some width due to coupling to the electrodes. It does not matter how small the width is - as long as it is nonzero, the first term vanishes identically[69]. Hence for our problem $G_{cc}^<$ becomes:

$$G_{cc}^< = G_{cc}^R \Sigma^< G_{cc}^A \tag{4.50}$$

This is equivalent to neglecting Ψ_c in Eq.(4.41), thus

$$\hat{\psi}_c(\mathbf{r}, t) = \int d\mathbf{r}' \int_{-\infty}^{+\infty} dt' G_{cc}(\mathbf{r}, t, \mathbf{r}', t') S(\mathbf{r}', t') \tag{4.51}$$

So far all the quantities are defined in the space of $(\mathbf{r}, t)$. Because in steady state, all the quantities only depend on $t - t'$, it will be very useful to Fourier transform into energy space to obtain:

$$G_{cc}^<(\mathbf{r}, \mathbf{r}', E) = \int d\mathbf{r}'' d\mathbf{r}''' G_{cc}^R(\mathbf{r}, \mathbf{r}'', E) \Sigma^<(\mathbf{r}'', \mathbf{r}''', E) G_{cc}^A(\mathbf{r}''', \mathbf{r}', E) \tag{4.52}$$

and $G^R_{cc}(\mathbf{r}, \mathbf{r}'', E)$ satisfies the equation

$$(E^+ - H_c)G^R_{cc}(\mathbf{r}, \mathbf{r}', E) - \int d\mathbf{r}''\Sigma^R(\mathbf{r}, \mathbf{r}'', E)G^R_{cc}(\mathbf{r}'', \mathbf{r}', E) = \delta(\mathbf{r} - \mathbf{r}') \tag{4.53}$$

Eq.(4.52) and Eq.(4.53) form the basic equations of our NEGF formalism where the self energies Σ and $\Sigma^<$ describe the coupling between electrodes and the central region.

4.3 The self-energies $\Sigma^<$ and Σ

In Section 4.2, it has been shown that,

$$\Sigma^<(\mathbf{r}, t, \mathbf{r}', t') = i < \hat{S}^+(\mathbf{r}', t')\hat{S}(\mathbf{r}', t) > \tag{4.54}$$

and

$$\hat{S}(\mathbf{r}, t) = \tau_l\hat{\Psi}_l(\mathbf{r}, t) + \tau_r\hat{\Psi}_r(\mathbf{r}, t) \tag{4.55}$$

Consequently,

$$\begin{aligned}\Sigma^<(\mathbf{r}, t, \mathbf{r}', t') &= \tau_l^*\tau_l i < \hat{\Psi}_l^+(\mathbf{r}', t')\hat{\Psi}_l(\mathbf{r}, t) > + \tau_r^*\tau_r i < \hat{\Psi}_r^+(\mathbf{r}', t')\hat{\Psi}_r(\mathbf{r}, t) > \\ &= \tau_l g^<_{ll}(\mathbf{r}, t, \mathbf{r}', t')\tau_l^+ + \tau_r g^<_{rr}(\mathbf{r}, t, \mathbf{r}', t')\tau_r^+ \end{aligned} \tag{4.56}$$

where $g^<_{ll} = i < \hat{\Psi}_l^+\hat{\Psi}_l >$ and $g^<_{rr} = i < \hat{\Psi}_r^+\hat{\Psi}_r >$ are the lesser Green's functions for the isolated left and right electrodes, respectively. For steady state situations, Fourier transformation can be carried out to obtain an energy dependent self-energy,

$$\Sigma^<(\mathbf{r}, \mathbf{r}', E) = \tau_l g_l^<(\mathbf{r}, \mathbf{r}', E)\tau_l^+ + \tau_r g_r^<(\mathbf{r}, \mathbf{r}', E)\tau_r^+ \tag{4.57}$$

In our consideration, electron reservoir is in equilibrium state. The electrode which connects to the reservoir is made of good metal and is equilibrated with the reservoir. Hence the fluctuation-dissipation theorem, Eq.(4.24), is applicable to the electrode,

namely $g^{<}_{ll/rr} = f_{l/r}(E)[g^{A}_{ll/rr} - g^{R}_{ll/rr}]$ with $f_{l/r}(E)$ corresponding to the Fermi-Dirac distribution function in the left and right electrodes respectively. The self-energy in Eq.(4.57) can be rewritten as

$$\begin{aligned}
&\Sigma^{<}(\mathbf{r},\mathbf{r}',E) = f_l(E)[\Sigma^{A}_{l}(\mathbf{r},\mathbf{r}',E) - \Sigma^{R}_{l}(\mathbf{r},\mathbf{r}',E)] \\
&+f_r(E)[\Sigma^{A}_{r}(\mathbf{r},\mathbf{r}',E) - \Sigma^{R}_{r}(\mathbf{r},\mathbf{r}',E)] \\
&= f_l(E)\Gamma_l(\mathbf{r},\mathbf{r}',E) + f_r(E)\Gamma_r(\mathbf{r},\mathbf{r}',E)
\end{aligned} \tag{4.58}$$

where $\Gamma_{l/r} = \Sigma^{A}_{l/r} - \Sigma^{R}_{l/r}$ and $\Sigma^{R/A}_{l/r} = \tau_{l/r} g^{R/A}_{ll/rr} \tau^{+}_{l/r}$. $\Sigma^{<}$ basically describes the coupling of the electrodes to the central device region, and it broadens the electronic states there.

4.4 Electric current

The electric current flowing from the left electrode to the central region is given by the time derivative of the charge,

$$\begin{aligned}
&J_l(r,t) = -e\frac{d}{dt} < \hat{\psi}^{+}_{l}(\mathbf{r},t)\hat{\psi}_l(\mathbf{r},t) > \\
&= \frac{-e}{i\hbar}\{< \hat{\psi}^{+}_{l}(\mathbf{r},t)[i\hbar\frac{d}{dt}\hat{\psi}_l(\mathbf{r},t)] > - < [-i\hbar\frac{d}{dt}\hat{\psi}^{+}_{l}(\mathbf{r},t)]\hat{\psi}_l(\mathbf{r},t) >\}
\end{aligned} \tag{4.59}$$

where $< \hat{\psi}^{+}_{l}(\mathbf{r},t)\hat{\psi}_l(\mathbf{r},t) >$ represents the electron density matrix of the left electrode. From the formulation in the last section, it is known that (see Eq.(4.29))

$$i\hbar\frac{\partial\hat{\psi}_l(\mathbf{r},t)}{\partial t} = H_l\hat{\psi}_l(\mathbf{r},t) + \tau^{+}_{l}\hat{\psi}_c(\mathbf{r},t)\ . \tag{4.60}$$

Therefore,

$$-i\hbar\frac{\partial\hat{\psi}^{+}_{l}(\mathbf{r},t)}{\partial t} = H^{+}_{l}\hat{\psi}^{+}_{l}(\mathbf{r},t) + \tau_l\hat{\psi}^{+}_{c}(\mathbf{r},t)\ . \tag{4.61}$$

Substituting the two equations above, Eq.(4.59) can be rewritten in the following form,

$$\begin{aligned} J_l(\mathbf{r},t) &= -e\frac{d}{dt} < \hat{\psi}_l^+(\mathbf{r},t)\hat{\psi}_l(\mathbf{r},t) > \\ &= \frac{-e}{i\hbar}\{\tau_l^+ < \hat{\psi}_l^+(\mathbf{r},t)\hat{\psi}_c(\mathbf{r},t) > -\tau_l < \hat{\psi}_c^+(\mathbf{r},t)\hat{\psi}_l(\mathbf{r},t) >\} \end{aligned} \tag{4.62}$$

The electric current is formulated as the difference of the flowing-in and flowing-out electrons of the left electrodes. From the Eqs.(4.32,4.51), we know,

$$\hat{\psi}_l^+(\mathbf{r},t) = \hat{\Psi}_l^+(\mathbf{r},t) + \int d\mathbf{r}' \int dt' g_{ll}^{R,*}(\mathbf{r},t,\mathbf{r}',t')\tau_l\hat{\psi}_c^+(\mathbf{r}',t') \tag{4.63}$$

$$\hat{\psi}_c^+(\mathbf{r},t) = \int d\mathbf{r}' \int dt' G_{cc}^{R,*}(\mathbf{r},t,\mathbf{r}',t')\hat{S}^+(\mathbf{r}',t') \tag{4.64}$$

Hence, we can write

$$\begin{aligned} &\tau_l^+(\mathbf{r}) < \hat{\psi}_l^+(\mathbf{r},t)\hat{\psi}_c(\mathbf{r},t) >= \tau_l^+ \int d\mathbf{r}' \int dt' G_{cc}(\mathbf{r},t,\mathbf{r}',t') < \hat{\Psi}_l^+(\mathbf{r},t)\hat{S}(\mathbf{r}',t') > \\ &+\tau_l^+ \int d\mathbf{r}' \int dt' g_{ll}^{R,*}(\mathbf{r},t,\mathbf{r}',t')\tau_l < \hat{\psi}_c^+(\mathbf{r}',t')\hat{\psi}_c(\mathbf{r},t) > \\ &= -i \int d\mathbf{r}' \int dt' [G_{cc}^R(\mathbf{r},t,\mathbf{r}',t')\Sigma_l^<(\mathbf{r}',t',\mathbf{r},t) + G_{cc}^<(\mathbf{r},t,\mathbf{r}',t')\Sigma_l^A(\mathbf{r}',t',\mathbf{r},t)] \end{aligned} \tag{4.65}$$

Where $\Sigma_l^<$ is the less self-energy contributed from the left electrode. Similarly, it is found that

$$\begin{aligned} &\tau_l < \hat{\psi}_c^+(\mathbf{r},t)\hat{\psi}_l(\mathbf{r},t) > \\ &= -i \int d\mathbf{r}' \int dt' [\Sigma_l^<(\mathbf{r},t,\mathbf{r}',t')G_{cc}^A(\mathbf{r}',t',\mathbf{r},t) + \Sigma_l^R(\mathbf{r},t,\mathbf{r}',t')G_{cc}^<(\mathbf{r}',t',\mathbf{r},t)] \end{aligned} \tag{4.66}$$

The electric current of Eq.(4.62) is thus obtained as

$$\begin{aligned} J_l(\mathbf{r},t) &= -e\frac{d}{dt} < \hat{\psi}_l^+(\mathbf{r},t)\hat{\psi}_l(\mathbf{r},t) > \\ &= \frac{e}{\hbar}[G_{cc}^R\Sigma_l^< + G_{cc}^<\Sigma_l^A - \Sigma_l^<G_{cc}^A - \Sigma_l^RG_{cc}^<]_{(\mathbf{r},t,\mathbf{r},t)} \end{aligned} \tag{4.67}$$

Here, to keep the formula simple, we have written all quantities as matrices in space of $(\mathbf{r}, t)$. The electric current $J(\mathbf{r}, t)$ can be viewed as the diagonal matrix element. In steady state, it is easier to work with energy dependent quantities. By Fourier transformation, we find

$$\begin{aligned} J_l(\mathbf{r}) &= \int dE J_l(\mathbf{r}, E) \\ &= \frac{e}{\hbar}\int dE[G^R_{cc}(E)\Sigma^<_l(E) + G^<_{cc}(E)\Sigma^A_l(E) - \Sigma^<_l(E)G^A_{cc}(E) - \Sigma^R_l(E)G^<_{cc}(E)]_{(\mathbf{r},\mathbf{r})} \end{aligned} \tag{4.68}$$

The total current is obtained by integrating $J_l(\vec{r})$,

$$\begin{aligned} I_{total,l} &= \int d\vec{r} J_l(\vec{r}) \\ &= \frac{e}{\hbar}\int dE Tr[G^R_{cc}(E)\Sigma^<_l(E) + G^<_{cc}(E)\Sigma^A_l(E) - \Sigma^<_l(E)G^A_{cc}(E) - \Sigma^R_l(E)G^<_{cc}(E)] \\ &= \frac{e}{\hbar}\int dE Tr\{\Sigma^<_l(E)(G^R_{cc}(E) - G^A_{cc}(E)) + (\Sigma^A_l(E) - \Sigma^R_l(E))G^<_{cc}(E)\} \\ &= \frac{e}{\hbar}\int dE Tr\{\Gamma_l[f_l(E)(G^R_{cc}(E) - G^A_{cc}(E)) + G^<_{cc}(E)]\} \\ &= \frac{e}{\hbar}\int dE Tr\{(f_l(E) - f_r(E))\Gamma_l G^R_{cc}\Gamma_r G^A_{cc}\} \end{aligned} \tag{4.69}$$

Here we have used the relations: $\Gamma_l = \Sigma^A_l - \Sigma^R_l$, $G^{A,-1}_{cc} - G^{R,-1}_{cc} = \Sigma^R - \Sigma^A = -(\Gamma_l + \Gamma_r)$, and $\Sigma^< = f_l(E)\Gamma_l + f_r(E)\Gamma_r$ and also $\Sigma^<_l = f_l(E)\Gamma_l$. Finally, in exactly the same way we can derive the electric current from the right electrode, namely

$$I_{total,r} = \frac{e}{\hbar}\int dE Tr\{(f_l(E) - f_r(E))\Gamma_l G^R\Gamma_r G^A\} \tag{4.70}$$

For non-equilibrium steady state, we know

$$I_{total,l} + I_{total.r} = 0 \tag{4.71}$$

and the total current flow through the two probe device is

$$I = I_{total,l} = \frac{e}{\hbar}\int dE Tr\{(f_l(E) - f_r(E))\Gamma_l G^R\Gamma_r G^A\} = \frac{e}{\hbar}\int dE T(E) \tag{4.72}$$

This is the well known Landauer formula where we can define a transmission coefficient $T(E)$, as

$$T(E) = Tr\{\Gamma_l G^R \Gamma_r G^A\} \tag{4.73}$$

Note that $T(E)$ is also a function of external bias voltage because the Green's functions depend on the bias.

4.5 Implementation with local orbital basis set

In previous sections, we have provided a detailed derivation of the general NEGF formalism for non-equilibrium quantum transport calculations of two-probe devices. The numerical implementation of the formulation is based on local orbital basis set which provides good accuracy and also manageable sized matrices. In the following, we discuss the implementation of NEGF within a general basis set. The choice of different basis functions only changes implementation details.

Let $\{\chi_{RL}\}$ to denote a complete set of local orbital basis functions which are centered at the atomic sites labelled by R and having quantum number L. All the quantities can be expanded using this basis set. For instance, eigen-functions ϕ_i can be expanded as:

$$\phi_i(\mathbf{r}) = \sum_{RL} c_{i,RL}\chi_{RL}(\mathbf{r}) \quad . \tag{4.74}$$

Similarly, the Hamiltonian matrix can be constructed as $\mathbf{H}_{RL,R'L'} = \langle\chi_{RL}|H|\chi_{R'L'}\rangle$; the overlap matrix $\mathbf{O}_{RL,R'L'} = \langle\chi_{RL}|\chi_{R'L'}\rangle$. From now on, I will use the black bold characters to represent matrices. Using the basis sets, the retarded Green's function satisfies a matrix equation, or equivalently

$$\mathbf{G}^R(E) = \Big(E^+\mathbf{O} \; - \; \mathbf{H}\Big)^{-1} \tag{4.75}$$

where $E^+ = E + i\eta$ and η is a positive infinitesimal. Hence, the retarded Green's function is obtained by inverting the matrix $E^+\mathbf{O} - \mathbf{H}$.

Equivalently, the matrix Eq.(4.75) can also be written as:

$$(E^{+}\mathbf{O}-\mathbf{H})\mathbf{G}^{R}(E)=\mathbf{I}\quad . \tag{4.76}$$

where $\mathbf{I}$ is the identity matrix.

For open systems such as the two-probe device shown in Fig.4.1, Eq.(4.76) is actually an infinite matrix equation since the electrodes are semi-infinitely long. To deal with this problem, we divide the system into three parts: left electrode, central region and right electrode. With this division, Eq.(4.76) can be rewritten in block matrix form as follows,

$$\begin{pmatrix} E^{+}\mathbf{O}_{ll}-\mathbf{H}_{ll} & E^{+}\mathbf{O}_{lc}-\mathbf{H}_{lc} & 0 \\ E^{+}\mathbf{O}_{cl}-\mathbf{H}_{cl} & E^{+}\mathbf{O}_{cc}-\mathbf{H}_{cc} & E^{+}\mathbf{O}_{cr}-\mathbf{H}_{cr} \\ 0 & E^{+}\mathbf{O}_{rc}-\mathbf{H}_{rc} & E^{+}\mathbf{O}_{rr}-\mathbf{H}_{rr} \end{pmatrix} \begin{pmatrix} \mathbf{G}^{R}_{ll} & \mathbf{G}^{R}_{lc} & \mathbf{G}^{R}_{lr} \\ \mathbf{G}^{R}_{cl} & \mathbf{G}^{R}_{cc} & \mathbf{G}^{R}_{cr} \\ \mathbf{G}^{R}_{rl} & \mathbf{G}^{R}_{rc} & \mathbf{G}^{R}_{rr} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & 0 & 0 \\ 0 & \mathbf{I} & 0 \\ 0 & 0 & \mathbf{I} \end{pmatrix} \tag{4.77}$$

Now, our goal is to obtain the expression for $\mathbf{G}_{cc}$ of the central region. From the last matrix equation, it is straightforward to write down the following set of equations for various blocks:

$$(E^{+}\mathbf{O}_{ll}-\mathbf{H}_{ll})\mathbf{G}^{R}_{lc}+(E^{+}\mathbf{O}_{lc}-\mathbf{H}_{lc})\mathbf{G}^{R}_{cc}=0 \tag{4.78}$$

$$(E^{+}\mathbf{O}_{cl}-\mathbf{H}_{cl})\mathbf{G}^{R}_{lc}+(E^{+}\mathbf{O}_{cc}-\mathbf{H}_{cc})\mathbf{G}^{R}_{cc}+(E^{+}\mathbf{O}_{cr}-\mathbf{H}_{cr})\mathbf{G}^{R}_{cr}=\mathbf{I} \tag{4.79}$$

$$(E^{+}\mathbf{O}_{rc}-\mathbf{H}_{rc})\mathbf{G}^{R}_{cc}+(E^{+}\mathbf{O}_{rr}-\mathbf{H}_{rr})\mathbf{G}^{R}_{rc}=0 \tag{4.80}$$

Eliminating the Green's function blocks of $\mathbf{G}^{R}_{lc}$ and $\mathbf{G}^{R}_{rc}$, we find the solution for $\mathbf{G}^{R}_{cc}$:

$$\mathbf{G}^{R}_{cc}(E)=[E^{+}\mathbf{O}_{cc}-\mathbf{H}_{cc}-\mathbf{\Sigma}^{R}_{l}(E)-\mathbf{\Sigma}^{R}_{r}(E)]^{-1} \tag{4.81}$$

where

$$\mathbf{\Sigma}^{R}_{l}(E)=(E^{+}\mathbf{O}_{cl}-\mathbf{H}_{cl})\mathbf{G}^{R,0}_{ll}(E^{+}\mathbf{O}_{lc}-\mathbf{H}_{lc}) \tag{4.82}$$

$$\mathbf{G}^{R,0}_{ll}(E) = [E^{+}\mathbf{O}_{ll} - \mathbf{H}_{ll}]^{-1} \tag{4.83}$$

$$\mathbf{\Sigma}^{R}_{r}(E) = (E^{+}\mathbf{O}_{cr} - \mathbf{H}_{cl})\mathbf{G}^{R,0}_{rr}(E^{+}\mathbf{O}_{rc} - \mathbf{H}_{rc}) \tag{4.84}$$

$$\mathbf{G}^{R,0}_{rr}(E) = (E^{+}\mathbf{O}_{rr} - \mathbf{H}_{rr})^{-1} \tag{4.85}$$

With Eq.(4.81), the retarded Green's function of the central region $\mathbf{G}^{R}_{cc}(E)$ is now expressed in terms of the Hamiltonian and overlap matrices of the central region, plus the self-energies $\mathbf{\Sigma}^{R}_{l}$ and $\mathbf{\Sigma}^{R}_{r}$ due to electrodes. Since all these quantities are matrices of finite size, Eq.(4.81) can be calculated by matrix inversion. The Eq.(4.78) and Eq.(4.80) will be further used in Section 4.7 below.

From the matrix form Eq.(4.81), the Green's function in real space $G^{R}_{cc}(\mathbf{r}, \mathbf{r}', E)$ in Eq.(4.53), can be obtained by an outer product of the basis set functions. Hence,

$$G^{R}_{cc}(\mathbf{r}, \mathbf{r}', E) = \sum_{RL,R'L'} \mathbf{G}^{R}_{cc,RL,R'L'}(E)\chi_{RL}(\mathbf{r})\chi^{*}_{R'L'}(\mathbf{r}') \tag{4.86}$$

where atomic sites R and R' are defined within the central region of the device.

Next, we define the NEGF matrix using our basis set $\{\chi_{RL}\}$ for the central device region. By inserting Eq.(4.86) into Eq.(4.52), we can rewrite the NEGF Eq.(4.52) in terms of basis functions and Green's function matrices,

$$\begin{aligned}
&G^{<}_{cc}(\mathbf{r}, \mathbf{r}', E) = \int d\mathbf{r}_1 d\mathbf{r}_2 G^{R}_{cc}(\mathbf{r}, \mathbf{r}_1, E)\Sigma^{<}(\mathbf{r}_1, \mathbf{r}_2)G^{A}_{cc}(\mathbf{r}_2, \mathbf{r}', E) \\
&= \int d\mathbf{r}_1 d\mathbf{r}_2 [\sum_{im} \mathbf{G}^{R}_{cc,im}(E)\chi_i(\mathbf{r})\chi^{*}_{m}(r_1)]\Sigma^{<}(\mathbf{r}_1, \mathbf{r}_2)[\sum_{nj} \mathbf{G}^{A}_{cc,nj}(E)\chi_n(\mathbf{r}_2)\chi^{*}_{j}(\mathbf{r}')] \\
&= \sum_{ij} \chi_i(\mathbf{r})\chi^{*}_{j}(\mathbf{r}') \sum_{mn} \mathbf{G}^{R}_{cc,im}(E)[\int d\mathbf{r}_1 d\mathbf{r}_2 \chi^{*}_{m}(\mathbf{r}_1)\Sigma^{<}(\mathbf{r}_1, \mathbf{r}_2)\chi_n(\mathbf{r}_2)]\mathbf{G}^{A}_{cc,nj}(E) \\
&= \sum_{ij} \chi_i(\mathbf{r})\chi^{*}_{j}(\mathbf{r}') \sum_{mn} \mathbf{G}^{R}_{cc,im}(E)\mathbf{\Sigma}^{<}_{mn}(E)\mathbf{G}^{A}_{cc,nj}(E) \\
&= \sum_{ij} \chi_i(\mathbf{r})\chi^{*}_{j}(\mathbf{r}')[\mathbf{G}^{R}_{cc}(E)\Sigma^{<}(E)\mathbf{G}^{A}_{cc}(E)]_{ij}
\end{aligned} \tag{4.87}$$

where all the indexes i, j, m, n run over the atomic positions in the central region.

Defining the NEGF matrix $\mathbf{G}^<(E)$ as:

$$\mathbf{G}^<_{cc}(E) = \mathbf{G}^R_{cc}(E)\mathbf{\Sigma}^<(E)\mathbf{G}^A_{cc}(E) \tag{4.88}$$

where the matrices $\mathbf{G}^R_{cc}(E)$ and $\mathbf{G}^A_{cc}(E)$ are connected by the matrix $\mathbf{\Sigma}^<(E)$, and $\mathbf{G}^A_{cc}(E) = \mathbf{G}^{R,\dagger}_{cc}(E)$. Then,

$$G^<_{cc}(\mathbf{r},\mathbf{r}',E) = \sum_{RL,R'L'} \mathbf{G}^<_{cc,RL,R'L'}(E)\chi_{RL}(\mathbf{r})\chi^*_{R'L'}(\mathbf{r}') \tag{4.89}$$

From the definition of the lesser self-energy in Section 4.3 (see Eq.(4.57)), the matrix $\Sigma^<(E)$ becomes:

$$\begin{aligned}\mathbf{\Sigma}^<(E) &= f_l(E)\mathbf{\Gamma}_l(E) + f_r(E)\mathbf{\Gamma}_r(E) \\ &= f_l(E)(\mathbf{\Sigma}^A_l(E) - \mathbf{\Sigma}^R_l(E)) + f_r(E)(\mathbf{\Sigma}^A_r(E) - \mathbf{\Sigma}^R_r(E))\end{aligned} \tag{4.90}$$

where the line width function matrix $\Gamma_{l/r}(E)$ of the electrodes are giving by the following expression:

$$\mathbf{\Gamma}_{l/r}(E) = \mathbf{\Sigma}^A_{l/r}(E) - \mathbf{\Sigma}^R_{l/r}(E) \tag{4.91}$$

in which the self-energy $\mathbf{\Sigma}^R_{l/r}(E)$ is defined in Eqs.(4.82,4.84). At equilibrium, one can confirm that the NEGF matrix in Eq.(4.88) satisfies the fluctuation-dissipation theorem using Eqs.(4.81,4.82,4.84,4.90), namely,

$$\mathbf{G}^<_{cc}(E) = \mathbf{G}^A_{cc}(E) - \mathbf{G}^R_{cc}(E) \tag{4.92}$$

Therefore,

$$G^<_{cc}(\mathbf{r},\mathbf{r}',E) = \sum_{RL,R'L'} \chi_{RL}(\mathbf{r}')\chi^*_{R'L'}(\mathbf{r})[\mathbf{G}^A_{cc}(E) - \mathbf{G}^R_{cc}(E)]_{RL,R'L'} \tag{4.93}$$

which provides great simplicity for the equilibrium theory since the electron density is directly given by $\rho(\mathbf{r}) \sim \int G^<_{cc}(\mathbf{r},\mathbf{r},E)dE$. The complexity of the theory for non-

equilibrium can be seen by comparing Eq.(4.92) (for equilibrium) and Eq.(4.88) and will be further discussed in Chapter 5.

Similarly to NEGF, we can express the electric current formula, Eq.(4.72) where all quantities are defined in real space, into a matrix form. By inserting Eq.(4.89) and Eq.(4.86) into Eq.(4.72), we obtain,

$$\begin{aligned}
I_{total} &= \frac{e}{\hbar}\int dE \int d\mathbf{r}d\mathbf{r}'\Gamma_l(\mathbf{r},\mathbf{r}',E)\{f_l(E)[G^R_{cc}(\mathbf{r}',\vec{r},E) - G^A_{cc}(\mathbf{r}',\mathbf{r},E)] + G^<_{cc}(\mathbf{r}',\mathbf{r},E)\} \\
&= \frac{e}{\hbar}\int dE \int d\mathbf{r}d\mathbf{r}'\Gamma_l(\mathbf{r},\mathbf{r}',E)\sum_{ij}\{f_l(E)[\mathbf{G}^R_{cc,ij}(E) - G^A_{cc,ij}(E)] + \mathbf{G}^<_{cc,ij}(E)\}\chi_i(\vec{r})\chi^*_j(\vec{r'}) \\
&= \frac{e}{\hbar}\int dE \sum_{ij}[\int d\mathbf{r}d\mathbf{r}'\Gamma_l(\mathbf{r},\mathbf{r}',E)\chi_i(\mathbf{r})\chi^*_j(\mathbf{r}')]\{f_l[\mathbf{G}^R_{cc,ij}(E) - \mathbf{G}^A_{cc,ij}(E)] + \mathbf{G}^<_{cc,ij}(E)\} \\
&= \frac{e}{\hbar}\int dE \sum_{ij}\mathbf{\Gamma}_{l,ji}(E)\{f_l(E)[\mathbf{G}^R_{cc,ij}(E) - \mathbf{G}^A_{cc,ij}(E)] + \mathbf{G}^<_{cc,ij}(E)\} \\
&= \frac{e}{\hbar}\int dE \sum_{j}\{\mathbf{\Gamma}_l(E)\{f_l(E)[\mathbf{G}^R_{cc}(E) - \mathbf{G}^A_{cc}(E)] + \mathbf{G}^<_{cc}(E)\}\}_{jj} \\
&= \frac{e}{\hbar}\int dE Tr\{\mathbf{\Gamma}_l(E)\{f_l(E)[\mathbf{G}^R(E) - \mathbf{G}^A(E)] + \mathbf{G}^<(E)\}\}
\end{aligned} \tag{4.94}$$

Using Eqs.(4.90,4.88) in the above equation, we obtain,

$$I_{total} = \frac{e}{\hbar}\int dE Tr\{(f_r(E) - f_l(E))\mathbf{\Gamma}_l(E)\mathbf{G}^R_{cc}(E)\mathbf{\Gamma}_r(E)\mathbf{G}^A_{cc}(E)\} \tag{4.95}$$

which is the Landauer formula. We can define a transmission coefficient $T(E)$ as follows in term of matrices,

$$T(E) = Tr[\mathbf{\Gamma}_l(E)\mathbf{G}^R_{cc}(E)\mathbf{\Gamma}_r(E)\mathbf{G}^A_{cc}(E)] \tag{4.96}$$

The equilibrium conductance G is obtained from $T(E)$ at the Fermi energy of the system,

$$G = \frac{e^2}{\hbar}T(E_F) \tag{4.97}$$

So far, we have rewritten equations of NEGF into a matrix form using a general local orbital basis set $\{\chi_{RL}\}$. Since χ_{RL} is an atomic orbital centered at the atomic site

R, it decays away from R. Therefore the overlap of two basis functions χ_{RL} and $\chi_{R'L'}$ can be neglected when $R - R'$ is large. This approximation makes the Hamiltonian and overlap sparse matrices. Furthermore, we shall carry out the following procedure. The two-probe device system is divided into principle layers (PL) from one end to the other end as shown in Fig.4.2, the PLs are labeled by an index p: the central device region contains $p = 1, 2, ..., n-1, N$ PLs, the left and right electrodes contain PLs labeled by $p = -\infty, ..., -1, 0$ and $p = N+1, N+2, ..., +\infty$, respectively. The width of PL is chosen such that atoms in a PL only have direct orbital overlap to atoms in the same PL, as well as those in the nearest neighbor PLs. Such a division into PLs makes the matrix $(E\mathbf{O} - \mathbf{H})$ to have a tri-diagonal form. Denoting $\mathbf{M} \equiv (E\mathbf{O} - \mathbf{H})$, we have:

$$\begin{pmatrix}
..... & \mathbf{M}_{-10} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{M}_{0-1} & \mathbf{M}_{00} & \mathbf{M}_{01} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \mathbf{M}_{10} & \mathbf{M}_{11} & \mathbf{M}_{12} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \mathbf{M}_{21} & \mathbf{M}_{22} & \mathbf{M}_{23} & 0 & 0 & 0 & 0 \\
 & & & & ..... & ..... & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mathbf{M}_{N-1,N-2} & \mathbf{M}_{N-1,N-1} & \mathbf{M}_{N-1,N} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \mathbf{M}_{N,N-1} & \mathbf{M}_{N,N} & \mathbf{M}_{N,N+1} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \mathbf{M}_{N+1,N} & \mathbf{M}_{N+1,N+1} & \mathbf{M}_{N+1,N+2} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .....
\end{pmatrix} \tag{4.98}$$

This immediately results in the following important simplification for the self-energy matrices $\mathbf{\Sigma}_l$ and $\mathbf{\Sigma}_r$,

$$\mathbf{\Sigma}_l^R = (E^+\mathbf{O}_{cl} - \mathbf{H}_{cl})\mathbf{G}_{ll}^{R,0}(E^+\mathbf{O}_{lc} - \mathbf{H}_{lc}) = \begin{pmatrix} \mathbf{\Sigma}_{11}^{R,l} & 0 & .... & 0 \\ & & ..... & \\ 0 & 0 & .... & 0 \end{pmatrix}_{cc} \tag{4.99}$$

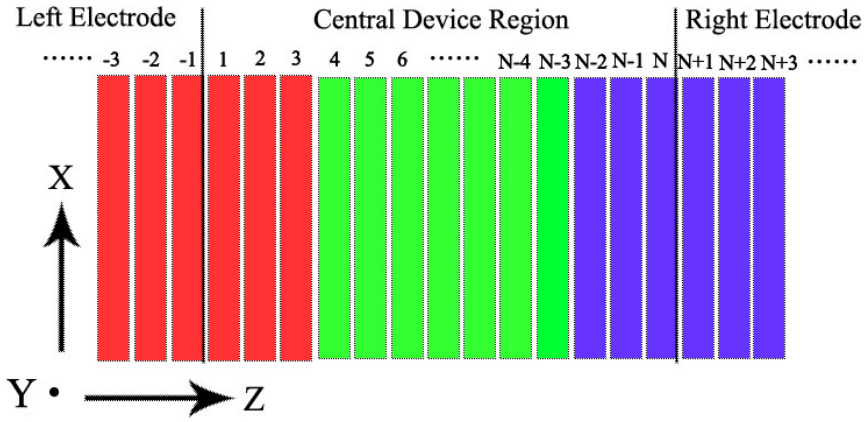

Figure 4.2: Schetch of two probe device in principle layers. A portion of the electrode is included in the central device region for screening, the central device region contains N principle layers.

$$\mathbf{\Sigma}_r^R = (E^+\mathbf{O}_{cr} - \mathbf{H}_{cr})\mathbf{G}_{rr}^{R,0}(E^+\mathbf{O}_{rc} - \mathbf{H}_{rc}) = \begin{pmatrix} 0 & 0 & .... & 0 \\ & & ..... & \\ 0 & 0 & .... & \mathbf{\Sigma}_{NN}^{R,r} \end{pmatrix}_{cc} \tag{4.100}$$

where the only non-zero blocks are $\mathbf{\Sigma}_{11}^{R,l}$ for $\mathbf{\Sigma}_l^R$, and $\mathbf{\Sigma}_{NN}^{R,r}$ for $\mathbf{\Sigma}_r^R$. These nonzero blocks are:

$$\mathbf{\Sigma}_{11}^{R,l} = (E^+\mathbf{O}_{10} - \mathbf{H}_{10})\mathbf{G}_{ll,00}^{R,0}(E^+\mathbf{O}_{01} - \mathbf{H}_{01}) \tag{4.101}$$

$$\mathbf{\Sigma}_{NN}^{R,r} = (E^+\mathbf{O}_{N,N+1} - \mathbf{H}_{N,N+1})\mathbf{G}_{rr,N+1,N+1}^{R,0}(E^+\mathbf{O}_{N+1,N} - \mathbf{H}_{N+1,N}) \tag{4.102}$$

Here, quantities $\mathbf{G}_{00}^{R,0}$ and $\mathbf{G}_{N+1,N+1}^{R,0}$ are the surface Green's functions of the left and right electrodes which we will describe in detail in the next section.

We can finally rewrite the NEGF matrix elements of Eq.(4.88) by using Eqs.(4.99, 4.100),

$$\mathbf{G}_{cc,pp'}^{<}(E) = f_l(E)\mathbf{G}_{cc,p1}^{R}\mathbf{\Gamma}_{11}^{l}(E)\mathbf{G}_{cc,1p'}^{A}(E) + f_r(E)\mathbf{G}_{cc,pN}^{R}(E)\mathbf{\Gamma}_{NN}^{r}(E)\mathbf{G}_{cc,Np'}^{A}(E) \tag{4.103}$$

where $\mathbf{\Gamma}_{11}^{l} = \mathbf{\Sigma}_{11}^{A,l} - \mathbf{\Sigma}_{11}^{R,l}$ and $\mathbf{\Gamma}_{NN}^{r} = \mathbf{\Sigma}_{NN}^{A,r} - \mathbf{\Sigma}_{NN}^{R,r}$. Noting that $\mathbf{G}_{1p'}^{A}(E) = \mathbf{G}_{p'1}^{R,\dagger}$, and $\mathbf{G}_{Np'}^{A}(E) = \mathbf{G}_{p'N}^{R,\dagger}(E)$, the computation of the full matrix $\mathbf{G}_{cc}^{<}$ only needs the elements $\mathbf{G}_{cc,p1}^{R}$ of the first column and $\mathbf{G}_{cc,pN}^{R}$ of the last column. Similarly, the transmission

coefficient can be rewritten as,

$$T(E) = Tr[\mathbf{\Gamma}_{11}(E)\mathbf{G}^{R}_{cc,1N}(E)\mathbf{\Gamma}_{NN}(E)\mathbf{G}^{A}_{cc,N1}(E)] \tag{4.104}$$

where $\mathbf{G}^{A}_{cc,N1}(E) = \mathbf{G}^{R,\dagger}_{cc,1N}(E)$. Only the block $\mathbf{G}^{R}_{cc,1N}(E)$ is needed for the calculation of $T(E)$.

In this section, we discussed the NEGF theory for calculating electron density matrix and electric current for two-probe device systems. It should be noted that we shall implement this formalism in the framework of KS DFT which is a mean field materials theory. For strongly correlated systems, more complicated expressions are needed to calculate quantum transport. In the next section, we will introduce the recursive Green's function technique for the inversion of the tri-diagonal matrix like Eq.(4.98), which is extensively used in NEGF simulation of nano-electronic devices.

4.6 The recursive Green's function technique

From the discussion in the last section, we see the retarded Green's function of the central device region in Eq.(4.81) is explicitly written as inverse of a tri-diagonal matrix, according to Eqs.(4.98,4.99,4.100),

$$\mathbf{G}^{R}_{cc} = \begin{pmatrix} \mathbf{M}_{11} - \mathbf{\Sigma}^{R,l}_{11} & \mathbf{M}_{12} & 0 & 0 & 0 \\ \mathbf{M}_{21} & \mathbf{M}_{22} & \mathbf{M}_{23} & 0 & 0 \\ & & ..... & 0 & 0 \\ 0 & 0 & \mathbf{M}_{N-1,N-2} & \mathbf{M}_{N-1,N-1} & \mathbf{M}_{N-1,N} \\ 0 & 0 & 0 & \mathbf{M}_{N,N-1} & \mathbf{M}_{NN} - \mathbf{\Sigma}^{R,r}_{NN} \end{pmatrix}^{-1} \tag{4.105}$$

The recursive Green's function technique provides efficient inversion of this tri-diagonal matrix. Here, we directly write down the equations for computing each block of the Green's function[70]. First, the diagonal blocks of the Green's function

are given by,

$$\mathbf{G}^R_{ii} = [\mathbf{M}_{ii} - \mathbf{\Sigma}^l_i - \mathbf{\Sigma}^r_i]^{-1}, \quad (1 \leq i \leq N) \tag{4.106}$$

where $\mathbf{\Sigma}^l_i$ and $\mathbf{\Sigma}^r_i$ can also be called self-energies which are recursively calculated by starting from the left side and right side of the central region: $\mathbf{\Sigma}^l_i$ and $\mathbf{\Sigma}^r_i$ are calculated independently. For example, by calculating from the left side, we obtain

$$\mathbf{\Sigma}^{R,l}_1 = \mathbf{\Sigma}^{R,l}_{11} \tag{4.107}$$

$$\mathbf{\Sigma}^{R,l}_{i+1} = \mathbf{M}_{i+1,i}[\mathbf{M}_{ii} - \mathbf{\Sigma}^l_i]^{-1}\mathbf{M}_{i,i+1} \ . \tag{4.108}$$

Calculating from the right side, we obtain

$$\mathbf{\Sigma}^{R,r}_N = \mathbf{\Sigma}^{R,r}_{NN} \tag{4.109}$$

$$\mathbf{\Sigma}^{R,r}_{i-1} = \mathbf{M}_{i-1,i}[\mathbf{M}_{ii} - \mathbf{\Sigma}^r_i]^{-1}\mathbf{M}_{i,i-1} \tag{4.110}$$

Once the diagonal block is found, the off-diagonal block of the Green's function can be derived in the following manner,

$$\mathbf{G}^R_{ij} = -[\mathbf{M}_{ii} - \mathbf{\Sigma}^r_i]^{-1}\mathbf{M}_{i,i-1}\mathbf{G}_{i-1,j}, \quad (i > j) \tag{4.111}$$

$$\mathbf{G}^R_{ij} = -[\mathbf{M}_{ii} - \mathbf{\Sigma}^l_i]^{-1}\mathbf{M}_{i,i+1}\mathbf{G}_{i+1,j}, \quad (i < j) \tag{4.112}$$

These results are general for inverting tri-diagonal block matrices. Such recursive Green's function technique provides great advantages in computing the NEGF matrix of Eq.4.93 for which only the first and last columns of the retarded Green's function, namely $\mathbf{G}^R_{p1}$ and $\mathbf{G}^R_{pN}$ for $1 \leq p \leq N$, are needed.

In the following, we present recursive equations for the surface Green's function that is needed for calculating self-energies of the electrodes (see Eqs.(4.101,4.102)).

For the left electrode, we can write down the following equation according to Eq.(4.82),

$$\begin{pmatrix} \mathbf{M}_{DD} & \mathbf{M}_{D0} \\ \mathbf{M}_{0D} & \mathbf{M}_{00} \end{pmatrix} \begin{pmatrix} \mathbf{G}^{R,0}_{ll,DD}(E) & \mathbf{G}^{R,0}_{ll,D0}(E) \\ \mathbf{G}^{R,0}_{ll,0D}(E) & \mathbf{G}^{R,0}_{ll,00}(E) \end{pmatrix} = \mathbf{I} \tag{4.113}$$

Here, the subscript D includes all the principle layers from $-\infty$ to -1. One can solve the block $\mathbf{G}^{R,0}_{ll,00}$,

$$\mathbf{G}^{R,0}_{ll,00}(E) = [\mathbf{M}_{00} - \mathbf{M}_{0D}\mathbf{M}^{-1}_{DD}\mathbf{M}_{D0}]^{-1} = [\mathbf{M}_{00} - \mathbf{M}_{0-1}[\mathbf{M}^{-1}_{DD}]_{-1-1}\mathbf{M}_{-10}]^{-1} \tag{4.114}$$

In practical applications, we always assume that the semi-infinite electrode is a perfect crystal, namely all principle layers in the electrode are the same. As a result, in the left electrode where $p \leq 0$, $\mathbf{M}_{pp} = \mathbf{M}_{00}$, $\mathbf{M}_{p,p-1} = \mathbf{M}_{0-1}$, $\mathbf{M}_{p-1,p} = \mathbf{M}_{-10}$, and $[\mathbf{M}^{-1}_{UU}]_{-1-1} = \mathbf{G}^{R,0}_{ll,-1-1} = \mathbf{G}^{R,0}_{ll,00} = \mathbf{G}^{R,0}_{left}$. Therefore, the above equation for the surface Green's function $\mathbf{G}^{0}_{00}$ is changed into a self-consistent equation,

$$\mathbf{G}^{R,0}_{left}(E) = [\mathbf{M}_{00} - \mathbf{M}_{0-1}\mathbf{G}^{R,0}_{left}(E)\mathbf{M}_{-10}]^{-1} \tag{4.115}$$

It is similar to find the surface Green's function for the right electrode, letting $\mathbf{G}^{R,0}_{N+1,N+1} = \mathbf{G}^{0}_{right}$ we have:

$$\mathbf{G}^{R,0}_{right}(E) = [\mathbf{M}_{N+1,N+1} - \mathbf{M}_{N+1,N+2}\mathbf{G}^{R,0}_{right}(E)\mathbf{M}_{N+2,N+1}]^{-1} \tag{4.116}$$

These two non-linear self-consistent equations provide a way to compute the surface Green's functions $\mathbf{G}^{R,0}_{left}(E)$ and $\mathbf{G}^{R,0}_{right}(E)$. In the literature, several different approaches were invented to compute them iteratively[71]. The most efficient method appears to be the renormalization-decimation technique [71] which we have implemented in our quantum transport software package. In this technique, 2^n number of principle layers can be included after n iterations.

4.7 Green's function method in TB-LMTO-ASA

In previous sections, we have introduced a general NEGF formulation and a general localized orbital basis set for practical calculations. In the rest of this thesis, we shall focus on a particular implementation within the tight binding LMTO-ASA framework as described in Chapter 3. As shown there, the Hamiltonian matrix that corresponds to the nearly orthogonal LMTO basis takes the following form (see Eq.(3.57)),

$$\mathbf{H}^{orth} = C + \sqrt{\Delta}\left\{\mathbf{S}^{\alpha}\left[1 + (\alpha - \gamma)\,\mathbf{S}^{\alpha}\right]^{-1}\right\}\sqrt{\Delta} \tag{4.117}$$

Here the diagonal matrices C , Δ and γ are the potential parameters that represents the center, the width and the distortion of the RL-th "band" (see Eqs. (3.58, 3.59,3.60)); α is the screening constant for the TB-LMTO representation (see Eq.(3.22)); and the matrix $\mathbf{S}^{\alpha}$ is the screened structure constant in α representation (see Eq.(3.26)). It should be mentioned again that the Hamiltonian matrix constructed by the orthogonal LMTO is invariant with respect to the screening constant α.

Because the orthogonal LMTO basis are rather long-ranged functions, it is computationally demanding if we apply the NEGF formalism described above directly. For practical purposes, we therefore rewrite the Green's function matrix by the following procedures,

$$\begin{aligned}
&\mathbf{G}(z) = [z - \mathbf{H}^{orth}]^{-1} \\
&= [z - C - \sqrt{\Delta}\mathbf{S}^{\alpha}[1 + (\alpha - \gamma)\mathbf{S}^{\alpha}]^{-1}\sqrt{\Delta}]^{-1} \\
&= \sqrt{\Delta}^{-1}\left\{\frac{z - C}{\Delta} - \mathbf{S}^{\alpha}[1 + (\alpha - \gamma)\mathbf{S}^{\alpha}]^{-1}\right\}^{-1}\sqrt{\Delta}^{-1} \\
&= \sqrt{\Delta}^{-1}[1 + (\alpha - \gamma)\mathbf{S}^{\alpha}]\left\{\frac{z - C}{\Delta}[1 + (\alpha - \gamma)\mathbf{S}^{\alpha}] - \mathbf{S}^{\alpha}\right\}^{-1}\sqrt{\Delta}^{-1} \\
&= \sqrt{\Delta}^{-1}[1 + (\alpha - \gamma)\mathbf{S}^{\alpha}]\left\{\frac{z - C}{\Delta} - \frac{\Delta + (z - C)(\gamma - \alpha)}{\Delta}\mathbf{S}^{\alpha}\right\}^{-1}\sqrt{\Delta}^{-1} \\
&= \sqrt{\Delta}^{-1}[1 + (\alpha - \gamma)\mathbf{S}^{\alpha}]\left\{\frac{z - C}{\Delta + (\gamma - \alpha)(z - C)} - \mathbf{S}^{\alpha}\right\}^{-1}\frac{\sqrt{\Delta}}{\Delta + (z - C)(\gamma - \alpha)}
\end{aligned}$$

$$= \frac{\gamma - \alpha}{\Delta + (z - C)(\gamma - \alpha)} + \frac{\sqrt{\Delta}}{\Delta + (z - C)(\gamma - \alpha)} \left\{ \frac{z - C}{\Delta + (\gamma - \alpha)(z - C)} - \mathbf{S}^{\alpha} \right\}^{-1}$$

$$\frac{\sqrt{\Delta}}{\Delta + (z - C)(\gamma - \alpha)} = \lambda^{\alpha}(z) + u^{\alpha}(z)[P^{\alpha}(z) - \mathbf{S}^{\alpha}]^{-1} u^{\alpha}(z) \tag{4.118}$$

This is the final form of the Green's function matrix in LMTO-ASA. Here, the matrices $\lambda^{\alpha}(z)$, $u^{\alpha}(z)$ and $P^{\alpha}(z)$ are all diagonal matrices with elements λ^{α}_{RL},u^{α}_{RL} and P^{α}_{RL} respectively,

$$P^{\alpha}_{RL}(z) = \frac{z - C_{RL}}{\Delta_{RL} + (\gamma_{RL} - \alpha_{RL})(z - C_{RL})}$$

$$\lambda^{\alpha}_{RL}(z) = \frac{\gamma_{RL} - \alpha_{RL}}{\Delta_{RL} + (\gamma_{RL} - \alpha_{RL})(z - C_{RL})}$$

$$u^{\alpha}_{RL}(z) = \frac{\sqrt{\Delta_{RL}}}{\Delta_{RL} + (\gamma_{RL} - \alpha_{RL})(z - C_{RL})} \ .$$

We introduce an auxiliary Green's function in the α representation of TB-LMTO for our calculations:

$$\mathbf{g}^{\alpha}(z) = [P^{\alpha}(z) - \mathbf{S}^{\alpha}]^{-1} \ . \tag{4.119}$$

From now on, to distinguish from the auxiliary quantities $\mathbf{g}^{\alpha}$ and $\mathbf{g}^{\alpha,<}$ which we shall introduce below, we call $\mathbf{G}^{\delta}$ where $\delta = r, a, <$ the physical Green's functions. From Eq.(4.118), we can observe that the physical Green's function is simply connected with the auxiliary Green's function by the following relation,

$$\mathbf{G}_{RL,R'L'}(z) = \lambda^{\alpha}_{RL}\delta_{RL,R'L'} + u^{\alpha}_{RL}\mathbf{g}^{\alpha}_{RL,R'L'}u^{\alpha}_{R'L'} \tag{4.120}$$

Although all quantities are α dependent, the Green's function is invariant with respect to the screening constant because of the α independent Hamiltonian matrix. The screening constant α can thus be considered as a free parameter which can be exploited to simplify numerical computation. Inside a particular form of the auxiliary Green's function $\mathbf{g}^{\alpha}$ of Eq.(4.119), the lattice geometry is only described by the screened structure constant $\mathbf{S}^{\alpha}$, and the chemical occupants on lattice sites are represented by the quantity P^{α}, thus the physical representation of a real atomistic

system is greatly simplified in this method. The structure constant matrix $\mathbf{S}^\alpha$ in the most screened representation is a short-ranged matrix[32], it provides a substantial computation advantage by calculating the auxiliary $\mathbf{g}^\alpha$ before obtaining the physical Green's functions.

When we consider a two-probe system, the Green's function for the central scattering region can be written in term of $\mathbf{g}^\alpha$, according to Eq.(4.120),

$$\mathbf{G}_{cc}(z) = \lambda_c^\alpha(z) + u_c^\alpha(z)\mathbf{g}_{cc}^\alpha(z)u_c^\alpha(z) \tag{4.121}$$

Similar to Eq.(4.77), we can write down the following equation for $\mathbf{g}_{cc}^\alpha$ from Eq.(4.119),

$$\begin{pmatrix} P_l^\alpha - \mathbf{S}_{ll}^\alpha & -\mathbf{S}_{lc}^\alpha & 0 \\ -\mathbf{S}_{cl}^\alpha & P_c^\alpha - \mathbf{S}_{cc}^\alpha & -\mathbf{S}_{cr}^\alpha \\ 0 & -\mathbf{S}_{rc}^\alpha & P_r^\alpha - \mathbf{S}_{rr}^\alpha \end{pmatrix} \begin{pmatrix} \mathbf{g}_{ll}^\alpha & \mathbf{g}_{lc}^\alpha & \mathbf{g}_{lr}^\alpha \\ \mathbf{g}_{cl}^\alpha & \mathbf{g}_{cc}^\alpha & \mathbf{g}_{cr}^\alpha \\ \mathbf{g}_{rl}^\alpha & \mathbf{g}_{rc}^\alpha & \mathbf{g}_{rr}^\alpha \end{pmatrix} = \begin{pmatrix} \mathbf{I} & 0 & 0 \\ 0 & \mathbf{I} & 0 \\ 0 & 0 & \mathbf{I} \end{pmatrix} \tag{4.122}$$

Thus

$$(P_l^\alpha - \mathbf{S}_{ll}^\alpha)\mathbf{g}_{lc}^\alpha - \mathbf{S}_{lc}^\alpha\mathbf{g}_{cc}^\alpha = 0 \tag{4.123}$$

$$-\mathbf{S}_{cl}^\alpha\mathbf{g}_{lc}^\alpha + (P_c^\alpha - \mathbf{S}_{cc}^\alpha)\mathbf{g}_{cc}^\alpha - \mathbf{S}_{cr}^\alpha\mathbf{g}_{cr}^\alpha = \mathbf{I} \tag{4.124}$$

$$-\mathbf{S}_{rc}^\alpha\mathbf{g}_{cc}^\alpha + (P_r^\alpha - \mathbf{S}_{rr}^\alpha)\mathbf{g}_{cr}^\alpha = 0 \tag{4.125}$$

We can then find:

$$\mathbf{g}_{cc}^\alpha = [P_{cc}^\alpha - \mathbf{S}_{cc}^\alpha - \mathbf{\Sigma}_l^\alpha - \mathbf{\Sigma}_r^\alpha]^{-1} \tag{4.126}$$

where

$$\mathbf{\Sigma}_l^\alpha(z) = \mathbf{S}_{cl}^\alpha\mathbf{g}_{ll}^{a,0}(z)\mathbf{S}_{lc}^\alpha \tag{4.127}$$

$$\mathbf{g}_{ll}^{a,0}(z) = [P_l^\alpha(z) - \mathbf{S}_{ll}^\alpha]^{-1} \tag{4.128}$$

$$\mathbf{\Sigma}_r^\alpha(z) = \mathbf{S}_{cr}^\alpha\mathbf{g}_{rr}^{a,0}(z)\mathbf{S}_{rc}^\alpha \tag{4.129}$$

$$\mathbf{g}_{rr}^{a,0}(z) = [P_r^{\alpha}(z) - \mathbf{S}_{rr}^{\alpha}]^{-1} \tag{4.130}$$

Here $\mathbf{\Sigma}_{l/r}^{\alpha}$ and $\mathbf{g}_{ll/rr}^{\alpha,0}$ are the respective self-energies and surface Green's function that correspond to the auxiliary Green's function.

In the following, we rewrite the physical NEGF in Eq.(4.88) in terms of the auxiliary Green's functions. If we only consider orthogonal basis sets, according to Eq.(4.82) to Eq.(4.85), where the overlap matrix $O = I$, Eq.(4.88) can be rewritten into the following form,

$$\begin{aligned}\mathbf{G}_{cc}^{<}(E) &= f_l(E)\mathbf{G}_{cc}^{R}\mathbf{H}_{cl}(\mathbf{G}_{ll}^{A,0} - \mathbf{G}_{ll}^{R,0})\mathbf{H}_{lc}\mathbf{G}_{cc}^{A}(E) \\ &+ f_r(E)\mathbf{G}_{cc}^{R}\mathbf{H}_{cr}(\mathbf{G}_{rr}^{A,0} - \mathbf{G}_{rr}^{R,0})\mathbf{H}_{rc}\mathbf{G}_{cc}^{A}(E)\end{aligned} \tag{4.131}$$

According to Eq.(4.78) and Eq.(4.80), we can obtain a similar equation for $\mathbf{G}^A$ by replacing E^+ with E^-,

$$[E^- - \mathbf{H}_{ll}]\mathbf{G}_{lc}^{A} - \mathbf{H}_{lc}\mathbf{G}_{cc}^{A} = 0 \tag{4.132}$$

and

$$[E^- - \mathbf{H}_{rr}]\mathbf{G}_{rc}^{A} - \mathbf{H}_{rc}\mathbf{G}_{cc}^{A} = 0 \tag{4.133}$$

where $\mathbf{O}_{lc/rc} = 0$ and $\mathbf{O}_{ll/rr} = 1$ for orthogonal basis set. Substituting these two equations into Eq.(4.131), and using the relation $\mathbf{G}^R = \mathbf{G}^{A,\dagger}$, we rewrite the physical NEGF as,

$$\begin{aligned}\mathbf{G}_{cc}^{<}(E) &= f_l(E)\mathbf{G}_{cl}^{R}(E^+ - \mathbf{H}_{ll})(\mathbf{G}_{ll}^{A,0} - \mathbf{G}_{ll}^{R,0})(E^- - \mathbf{H}_{ll})\mathbf{G}_{lc}^{A}(E) \\ &+ f_r(E)\mathbf{G}_{cr}^{R}(E^+ - \mathbf{H}_{rr})(\mathbf{G}_{rr}^{A,0} - \mathbf{G}_{rr}^{R,0})(E^- - \mathbf{H}_{rr})\mathbf{G}_{rc}^{A}(E) \\ &= f_l(E)\mathbf{G}_{cl}^{R}(E^+ - E^-)\mathbf{G}_{lc}^{A}(E) + f_r(E)\mathbf{G}_{cr}^{R}(E^+ - E^-)\mathbf{G}_{rc}^{A}(E)\end{aligned} \tag{4.134}$$

Now, apply the block Green's function using Eq.(4.120), we obtain

$$\mathbf{G}_{cc}^{<}(E) = f_l(E)u_c^{\alpha}(E^+)\mathbf{g}_{cl}^{\alpha,R}u_l^{\alpha}(E^+)(E^+ - E^-)u_l^{\alpha}(E^-)\mathbf{g}_{lc}^{\alpha,A}u_c^{\alpha}(E^-)$$

$$+f_r(E)u_c^\alpha(E^+)\mathbf{g}_{cr}^{\alpha,R}u_r^\alpha(E^+)(E^+-E^-)u_r^\alpha(E^-)\mathbf{g}_{rc}^{\alpha,A}u_c^\alpha(E^-) \tag{4.135}$$

Noting:

$$\begin{aligned}
&u_l^\alpha(E^+)(E^+-E^-)u_l^\alpha(E^-) = P_l^\alpha(E^+) - P_l^\alpha(E^-)\\
&= [P_l^\alpha(E^+) - \mathbf{S}_{ll}^\alpha][\mathbf{g}_{ll}^{\alpha,A,0} - \mathbf{g}_{ll}^{\alpha,R,0}][P_l^\alpha(E^-) - \mathbf{S}_{ll}^\alpha]\\
&u_r^\alpha(E^+)(E^+-E^-)u_r^\alpha(E^-) = P_r^\alpha(E^+) - P_r^\alpha(E^-)\\
&= [P_r^\alpha(E^+) - \mathbf{S}_{rr}^\alpha][\mathbf{g}_{rr}^{\alpha,A,0} - \mathbf{g}_{rr}^{\alpha,R,0}][P_r^\alpha(E^-) - \mathbf{S}_{rr}^\alpha]
\end{aligned} \tag{4.136}$$

and from Eqs.(4.123, 4.125), we obtain the following relations,

$$[P_l^\alpha(E^-) - \mathbf{S}_{ll}^\alpha]\mathbf{g}_{lc}^{\alpha,A} = \mathbf{S}_{lc}^\alpha\mathbf{g}_{cc}^{\alpha,A} \tag{4.137}$$

$$[P_r^\alpha(E^-) - \mathbf{S}_{rr}^\alpha]\mathbf{g}_{rc}^{\alpha,A} = \mathbf{S}_{rc}^\alpha\mathbf{g}_{cc}^{\alpha,A} \tag{4.138}$$

and equivalently the relations

$$\mathbf{g}_{cl}^{\alpha,R}[P_l^\alpha(E^+) - \mathbf{S}_{ll}^\alpha] = \mathbf{g}_{cc}^{\alpha,R}\mathbf{S}_{cl}^\alpha \tag{4.139}$$

$$\mathbf{g}_{cr}^{\alpha,R}[P_r^\alpha(E^+) - \mathbf{S}_{rr}^\alpha] = \mathbf{g}_{cc}^{\alpha,R}\mathbf{S}_{cr}^\alpha \tag{4.140}$$

The NEGF formula in Eq.(4.131) is written into the following final form:

$$\begin{aligned}
&\mathbf{G}_{cc}^<(E) = f_l(E)u_c^\alpha(E^+)\mathbf{g}_{cc}^{\alpha,R}[\mathbf{\Sigma}_l^{\alpha,A} - \mathbf{\Sigma}_l^{\alpha,R}]\mathbf{g}_{cc}^{\alpha,A}u_c^\alpha(E^-)\\
&+f_r(E)u_c^\alpha(E^+)\mathbf{g}_{cc}^{\alpha,R}[\mathbf{\Sigma}_l^{\alpha,A} - \mathbf{\Sigma}_l^{\alpha,R}]\mathbf{g}_{cc}^{\alpha,A}u_c^\alpha(E^-)\\
&= u_c^\alpha(E^+)\mathbf{g}_{cc}^{\alpha,<}u_c^\alpha(E^-)
\end{aligned} \tag{4.141}$$

This is the final expression of the physical NEGF within the tight binding LMTO-ASA method. The physical NEGF is also independent of the α representation. In

Eq.(4.141), we defined the a auxiliary NEGF $\mathbf{g}^{\alpha,<}$ in the α representation,

$$\mathbf{g}_{cc}^{\alpha,<} = \mathbf{g}_{cc}^{\alpha,R}\{f_l(E)\mathbf{\Gamma}_l^{\alpha} + f_r(E)\mathbf{\Gamma}_r^{\alpha}\}\mathbf{g}_{cc}^{\alpha,A} = \mathbf{g}_{cc}^{\alpha,R}\mathbf{\Sigma}^{\alpha,<}\mathbf{g}_{cc}^{\alpha,A} \tag{4.142}$$

where $\mathbf{\Gamma}_{l/r}^{\alpha} = \mathbf{\Sigma}_{l/r}^{\alpha,A} - \mathbf{\Sigma}_{l/r}^{\alpha,R}$ and

$$\mathbf{\Sigma}^{\alpha,<} = f_l(E)\mathbf{\Gamma}_l^{\alpha} + f_r(E)\mathbf{\Gamma}_r^{\alpha} \tag{4.143}$$

By computation of $\mathbf{g}^{\alpha,<}$, one can directly determine the physical NEGF $\mathbf{G}^{<}$.

Finally, we need to express the transmission coefficient in terms of the auxiliary Green's function. Similar to the above derivation of NEGF, we first rewrite T(E) of Eq.(4.96) as:

$$T(E) = Tr[\Gamma_l\mathbf{G}_{cr}^{R}(E^+ - E^-)\mathbf{G}_{rc}^{A}] = Tr[\mathbf{G}_{rc}^{A}\Gamma_l\mathbf{G}_{rc}^{R}(E^+ - E^-)] \tag{4.144}$$

Because from Eq.(4.77), we can obtain that, for the orthogonal representation,

$$(E - \mathbf{H}_{ll})\mathbf{G}_{lr} = \mathbf{H}_{lc}\mathbf{G}_{cr} \tag{4.145}$$

Then it is easy to find

$$T(E) = Tr[\mathbf{G}_{rl}^{A}(E^+ - E^-)\mathbf{G}_{lr}^{R}(E^+ - E^-)] \tag{4.146}$$

Using the Green's function of Eq.(4.120), the above $T(E)$ can be rewritten as

$$\begin{aligned} T(E) &= Tr[\mathbf{g}_{rl}^{\alpha}u_l^{\alpha}(E^+ - E^-)u_l^{\alpha}\mathbf{g}_{lr}^{\alpha}u_r^{\alpha}(E^+ - E^-)u_r^{\alpha}] \\ &= Tr\{\mathbf{g}_{rl}^{\alpha}[P_l^{\alpha}(E^+) - P_l^{\alpha}(E^-)]\mathbf{g}_{lr}^{\alpha}[P_r^{\alpha}(E^+) - P_r^{\alpha}(E^-)]\} \end{aligned} \tag{4.147}$$

Algebraically, this expression looks exactly the same as Eq.(4.96) with $P^{\alpha} - \mathbf{S}^{\alpha}$ taking the place of $z - \mathbf{H}$. The final expression for transmission coefficient in tight binding

LMTO-ASA method becomes:

$$T(E) = Tr(\mathbf{\Gamma}_l^{\alpha} \mathbf{g}_{cc}^{\alpha,R} \mathbf{\Gamma}_r^{\alpha} \mathbf{g}_{cc}^{\alpha,A}) \tag{4.148}$$

So far, we have rewritten the Green's function formalism for quantum transport in terms of the auxiliary Green's functions within the framework of tight binding LMTO-ASA method, see Eq.(4.148) for electric current, Eqs.(4.141 ,4.142) for the physical NEGF, and Eqs.(4.120,4.119) for the physical Green's function.

4.8 Summary

As a summary for this Chapter, we reviewed the general NEGF formalism and its practical implementation for treating non-equilibrium quantum transport in nano-electronic devices. Starting from the KS Hamiltonian, in Sections 4.2,4.3,4.4, we derived the Keldysh NEGF equation in Eq.(4.52) and the Landauer formula in Eq.(4.72). In Section 4.5, we presented a practical implementation of the NEGF quantum transport formalism within a general local orbital basis set by rewriting the NEGF formalism in terms of the Green's function matrices. The recursive Green's function technique is introduced for fast computation. In Section 4.7, a particular implementation of NEGF formalism within the TB-LMTO-ASA approach is introduced by reformulating the physical Green's function and NEGF in terms of their auxiliary counterparts, and rewriting the Landauer formula using the auxiliary Green's function. The NEGF-DFT approach provides the most powerful technique for simulating quantum transport properties of nano-electronic devices from atomic first principles. Based on NEGF-DFT, in the next chapter, we will develop a non-equilibrium vertex correction theory for handling the non-equilibrium disorder averaging problem. The results obtained in this Chapter, Eqs.(4.119, 4.120, 4.141, 4.142, 4.148) will serve as the starting point of Chapter 5.

5

Theory of non-equilibrium vertex correction

As we discussed in Chapter 1, it is inevitable that all nano-electronic devices produced in laboratories contain some amount of disorder or imperfections such as impurities, defects, dopants, dislocations, and so on. These unintentional impurities sit at unpredictable random locations in the host lattice and significantly affect quantum transport and electron scattering. There are experimental evidence that even a single impurity can some times affect charge transport substantially[72]. Understanding the roles of disordered impurities for nonequilibrium quantum transport has become a critical issue for device theory. The main intellectual contribution of this thesis is to develop a non-equilibrium vertex correction (NVC) theory to calculate disorder averaging in the NEGF-DFT formalism. In this Chapter, we present the details of the NVC theory that allows us to calculate disorder averaging at the density matrix level, Eq.(1.5), and the transmission coefficient level, Eq.(1.6).

So far in the theoretical literature, understanding effects due to substitutional disorder have achieved the greatest progress and we shall focus on this type of disorder as well. Substitutional disorder is produced by replacing a randomly selected host atom in a perfect crystal with an impurity atom. The atomic potential is changed at the replacement site inducing a charge rearrangement that affect the potential landscape at neighboring sites. Such a potential change provides electron scattering which alters quantum transport properties of the system. As a simplest assumption, atomic substitution occurs on sites of an unperturbed rigid lattice, namely the host lattice is assumed to not undergo significant structural change after atomic substitution. This

is an approximation because the impurity atom may not have the same size as the host atom and local strain may induce some structural relaxation near the impurity. The lattice relaxation can be important in disordered systems and it is itself a large topic of active research. Since our interest is on electron scattering and quantum transport, we shall not consider structural issues due to impurity and assume a rigid lattice model.

With these in mind, the probability of finding an atom of a given species on a lattice site is independent of the nature of the atoms occupying the neighboring sites. For example, in a binary alloy A_xB_{1-x}, the probability that a given site is occupied by an A atom is x, which is the atomic concentration of the A-species relative to the total number of atoms. In reality, various effects such as atomic size, electrostatic effects, chemical nature of the atoms etc., may cause atoms to aggregate or to preferentially choose neighboring atoms. Even though such clustering is usually short range for small concentration x, it produces some degree of non-randomness. Again, we shall not consider such structural issues.

We emphasize that our NVC theory is atomic center based (see below), hence it can be applied to systems where impurities preferentially locate at certain sites. For instance, perhaps there are more impurities near the interface of a heterostructure than deep inside the lattice, namely the impurities are distributed randomly but not uniformly. For such non-uniformly distributed impurity structures, one may carry out DFT total energy calculations to determine the likely low energy atomic configurations and then carry out quantum transport calculations. We will use the coherent potential approximation (CPA) within the multiple scattering theory to calculate the configurationally averaged retarded or advanced Green's functions, and develop the NVC theory for the configurational average of NEGF.

5.1 Basic definitions

We consider a binary disorder system of two different atomic elements $Q = A, B$, which are distributed randomly over a regular ideal lattice site R with probabilities c_R^A and $c_R^B = 1 - c_R^A$, and the site label R runs over the entire lattice. The probability c_R^Q can depend on R thus impurity can be non-uniformly distributed. In the Muffin Tin approximation, the corresponding Hamiltonian can be written in the form of (see Eq.(3.6)),

$$H = -\Delta + \sum_R V_{MT}(r_R) \tag{5.1}$$

$$V_{MT}(r_R) = \eta_R^A V_{MT}^A(r_R) + \eta_R^B V_{MT}^B(r_R) \tag{5.2}$$

Here η_R^Q denotes the occupation of site R by an atom of type Q, for example $\eta_R^Q = 1$ if the R-site is occupied by an atom of type Q, otherwise $\eta_R^Q = 0$. The potential $V_{MT}^Q(r_R)$ represents the component dependent (Q dependent) potential within the Q atomic sphere on site R. For random disorder, the probability for $\eta_R^A = 1$ is c_R^A; for $\eta_R^A = 0$ is $1 - c_R^A$. In the above Hamiltonian, the potential $V_{MT}(r_R)$ can randomly take on two different potentials $V_{MT}^A(r_R)$ and $V_{MT}^B(r_R)$ with the respective probabilities $c_R^A$and c_R^B. As a direct consequence, within the tight binding LMTO-ASA approach, the potential parameters C_R, Δ_R and γ_{RL} at the site R can have two different values randomly, C_R^A and C_R^B, Δ_R^A andΔ_{RL}^B, γ_R^A and γ_R^B with the respective probabilities c_R^A and c_R^B. Please distinguish the capital C_R from the lower case c_R.

With these statistical quantities, we can rewrite the Hamiltonian and the physical Green's function matrices of Eqs.(3.57, 4.120), $\mathbf{H}_{R,R'}$ and $\mathbf{G}_{R,R'}$, in terms of the occupation indices,

$$\mathbf{H}_{R,R'} = \sum_Q C_R^Q \eta_R^Q \delta_{R,R'} + \sum_{Q,Q'} \sqrt{\Delta_R^Q} \eta_R^Q \left\{ \mathbf{S}^\alpha \left[1 + (\alpha - \gamma)\mathbf{S}^\alpha \right]^{-1} \right\}_{RR'} \eta_{R'}^{Q'} \sqrt{\Delta_{R'}^{Q'}} \tag{5.3}$$

where $\gamma_R = \sum_Q \eta_R^Q \gamma_R^Q$, and

$$\mathbf{G}_{R,R'}(z) = \sum_Q \lambda_R^{\alpha,Q}(z)\eta_R^Q \delta_{R,R'} + \sum_{Q,Q'} u_R^{\alpha,Q}(z)\eta_R^Q \mathbf{g}_{RR'}^{\alpha}\eta_{R'}^{Q'} u_{R'}^{\alpha,Q'}(z) \tag{5.4}$$

Here g^α is the auxiliary Green's function in α representation (see Eq.(4.119)),

$$\mathbf{g}^\alpha(z) = [P^\alpha - \mathbf{S}^\alpha]^{-1} \tag{5.5}$$

where $P_R^\alpha = \sum_Q \eta_R^Q P_R^{\alpha,Q}$. In the above Eqs.(5.3,5.4,5.5), α is the screening constant which is independent of the atomic species. The diagonal matrices λ_{RL}^α, μ_{RL}^α and P_{RL}^α in addition to those potential parameters, are all statistical quantities which depend on the random occupation of the given site R by atoms A or B. The only non-random quantity is the tight-binding structure constant matrix $\mathbf{S}^\alpha$ which is corresponding to the lattice geometry. The Green's function $\mathbf{G}(z)$ and the corresponding auxiliary Green's function $\mathbf{g}^\alpha(z)$ are thus statistical quantities.

For a two-probe binary system, the physical NEGF matrix elements of the central region within the tight binding LMTO-ASA approach can be rewritten in the following form (using Eq.(4.141)),

$$\mathbf{G}_{cc,RR'}^{<}(z) = \sum_{Q,Q'} \mu_R^{\alpha,Q}(z)\eta_R^Q \mathbf{g}_{cc,RR'}^{<,\alpha}(z)\eta_{R'}^{Q'}\mu_{R'}^{\alpha,Q'}(z) \tag{5.6}$$

and the auxiliary NEGF $\mathbf{g}^{<,\alpha}$ in the α representation is given by the expression (see Eq.(4.142)),

$$\mathbf{g}_{cc}^{\alpha,<} = \mathbf{g}_{cc}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\mathbf{g}_{cc}^{\alpha,\mathcal{A}} \tag{5.7}$$

where the index cc denotes the central region, $\mathbf{\Sigma}^{\alpha,<}$ is the auxiliary self-energy. These quantities were discussed in Section 4.7. We shall assume that the electrodes of the device contain no disorder, hence $\mathbf{\Sigma}^{\alpha,<}$ is not a statistical quantity. It is evident that the physical NEGF and auxiliary NEGF are both statistical matrices.

The statistical quantities must be averaged over configurations of disorder in order to obtain physically meaningful results. Namely, we need to calculate the following:

$$\overline{\mathbf{G}_{R,R'}(z)} = \sum_{Q} c_R^Q \lambda_R^{\alpha,Q} \delta_{R,R'} + \sum_{Q,Q'} \mu_R^{\alpha,Q} \overline{\eta_R^Q \mathbf{g}_{RR'}^{\alpha} \eta_{R'}^{Q'}} \mu_{R'}^{\alpha,Q'} \tag{5.8}$$

and

$$\overline{\mathbf{G}_{cc,RR'}^{<}(z)} = \sum_{Q,Q'} \mu_R^{\alpha,Q}(z) \overline{\eta_R^Q \mathbf{g}_{cc,RR'}^{<,\alpha}(z) \eta_{R'}^{Q'}} \mu_{R'}^{\alpha,Q'}(z) \tag{5.9}$$

where $\overline{(\cdots)}$ denotes configurational average, and the relation $\overline{\eta_R^Q} = c_R^Q$ has been used.

Using the relation $\eta_R^Q \eta_R^{Q'} = \eta_R^Q \delta_{Q,Q'}$, we can rewrite the configurationally averaged site-diagonal elements as:

$$\begin{aligned}
\overline{\mathbf{G}}_{RR}(z) &= \sum_{Q} c_R^Q \lambda_R^{\alpha,Q} + \sum_{Q,Q'} \mu_R^{\alpha,Q} \overline{\eta_R^Q \mathbf{g}_{RR}^{\alpha} \eta_R^{Q'}} \mu_R^{\alpha,Q'} \\
&= \sum_{Q} c_R^Q \lambda_R^{\alpha,Q} + \mu_R^{\alpha,Q} \overline{\eta_R^Q \mathbf{g}_{RR}^{\alpha}} \mu_R^{\alpha,Q} \\
&= \sum_{Q} c_R^Q \overline{\mathbf{G}}_{RR}^{Q}(z)
\end{aligned} \tag{5.10}$$

$$\overline{\mathbf{G}}_{RR}^{Q}(z) = \lambda_R^{\alpha,Q} + \mu_R^{\alpha,Q} \overline{\mathbf{g}}_{RR}^{\alpha,Q} \mu_R^{\alpha,Q} \tag{5.11}$$

where $\mathbf{G}_{R,R}^{Q}$ is the the **conditionally** averaged site-diagonal Green's function which is connected to its auxiliary counterpart defined as,

$$\overline{\mathbf{g}}_{RR}^{\alpha,Q} = \overline{\eta_R^Q \mathbf{g}_{RR}^{\alpha}} / c_R^Q \tag{5.12}$$

For the site-diagonal elements of NEGF, we have a similar decomposition,

$$\overline{\mathbf{G}}_{cc,RR}^{<}(z) = \sum_{Q} \mu_R^{\alpha,Q}(z) \overline{\eta_R^Q \mathbf{g}_{cc,RR}^{<,\alpha}(z)} \mu_R^{\alpha,Q}(z) = \sum_{Q} c_R^Q \overline{\mathbf{G}}_{cc,RR}^{<,Q} \tag{5.13}$$

$$\overline{\mathbf{G}}_{cc,RR}^{<,Q} = \mu_R^{\alpha,Q}(z) \overline{\mathbf{g}}_{cc,RR}^{<,\alpha,Q}(z) \mu_R^{\alpha,Q}(z) \tag{5.14}$$

Here we have defined the conditionally averaged site-diagonal physical NEGF, $\overline{\mathbf{G}}^{<,Q}_{cc,RR}$ which has a simple connection to its auxiliary counterpart,

$$\overline{\mathbf{g}}^{<,\alpha,Q}_{cc,RR}(z) = \overline{\eta^Q_R \mathbf{g}^{<,\alpha}_{cc,RR}(z)}/c^Q_R \tag{5.15}$$

From above relations, it is seen that the averaged site-diagonal part of physical Green's function and NEGF, $\overline{\mathbf{G}}_{RR}$ and $\overline{\mathbf{G}}^{<}_{cc,RR}$, are both partitioned into contributions corresponding to the occupation of the site R by atoms $Q = A, B$. The conditional average over site-diagonal quantities, such as $\overline{\mathbf{G}}^Q_{RR}$, $\overline{\mathbf{g}}^Q_{RR}$, $\overline{\mathbf{G}}^{<,Q}_{cc,RR}$ and $\overline{\mathbf{g}}^{\alpha,<,Q}_{cc,RR}$, has the following meaning: the occupation on site R is fixed to atom Q, and the averaging is restricted to all configurations of the remaining sites in the system. This is usually called "restricted ensemble average". The averaged local density matrices of the atoms of type Q, which are needed in charge self-consistency in the atomic sphere approximation, are directly provided by the conditionally averaged quantities $\overline{\mathbf{G}}^Q_{RR}$ and $\overline{\mathbf{G}}^{<,Q}_{cc,RR}$ at equilibrium and non-equilibrium, respectively (see Section 5.6). This is why we only concentrated on the site-diagonal elements so far.

For the averaged auxiliary quantities, one can check the following equations to be true,

$$\sum_{Q=A,B} C^Q_R \overline{\mathbf{g}}^{\alpha,Q}_{RR} = \overline{\mathbf{g}}^{\alpha}_{RR} = \{\overline{[P^\alpha - \mathbf{S}^\alpha]^{-1}}\}_{RR} \tag{5.16}$$

$$\sum_{Q=A,B} C^Q_R \overline{\mathbf{g}}^{\alpha,<,Q}_{cc,RR} = \overline{\mathbf{g}}^{\alpha,<}_{cc,RR} = \{\overline{\mathbf{g}^{\alpha,\mathcal{R}}_{cc} \mathbf{\Sigma}^{\alpha,<} \mathbf{g}^{\alpha,\mathcal{A}}_{cc}}\}_{RR} \tag{5.17}$$

These are guaranteed by the relation $\sum_{Q=A,B} \eta^Q_R = 1$.

The configurational average produces averaged quantities which are non-random and most importantly, the averaged quantities such as $\overline{\mathbf{G}}$, $\overline{\mathbf{g}}^\alpha$, $\overline{\mathbf{G}}^<$ and $\overline{\mathbf{g}}^{\alpha,<}$, have the full crystal symmetry of the underlying ideal lattice. Therefore the configurational average has restored translational invariance which is critical for applying many techniques and theories of solid state physics (such as the Bloch theorem). This averaging procedure renders an intractable problem of disorder to a calculable theory.

In the following, we will derive the averaged auxiliary Green's function $\overline{\mathbf{g}}^{\alpha}$ under the coherent potential approximation (CPA) within multiple scattering theory in Section 5.2. After obtaining $\overline{\mathbf{g}}^{\alpha}$, the averaged auxiliary NEGF $\overline{\mathbf{g}}^{\alpha,<}$ will be formulated by deriving the non-equilibrium vertex correction in Section 5.3. The derivation of the conditionally averaged site-diagonal element of the auxiliary green's function and NEGF, $\overline{\mathbf{g}}^{\alpha,Q}_{RR}$ and $\overline{\mathbf{g}}^{\alpha,<,Q}_{RR}$ respectively, will be presented in section 5.4.

5.2 Coherent potential approximation

The coherent potential approximation (CPA) was independently introduced by Soven[28] and Taylor[29], and mathematically worked out by introducing the single-site approximation by Velicky et al[30]. The CPA was first applied within the KKR method to do charge-self-consistent calculation by Stocks and Winters[31]. Kudrnovsky et al.[34, 35] implemented CPA within the tight binding LMTO method. There are extensive literature on the CPA method and its implementations[73, 74]. CPA provides a self-consistent method to construct a translationally invariant effective Hamiltonian characterizing the configurational averaged material. In this section,we will apply the CPA method to calculate the configurationally averaged auxiliary Green's function in the tight binding LMTO-ASA method,

$$\overline{\mathbf{g}}^{\alpha}(z) = \overline{[P^{\alpha} - \mathbf{S}^{\alpha}]^{-1}} \tag{5.18}$$

In the above equation, the structure constant $\mathbf{S}^{\alpha}$ possesses all the symmetry of the underlying lattice, and the only statistical quantity is the diagonal potential function P^{α}_{R} which describes atoms occupying the site R statistically. According to the basic idea of CPA[28, 29], let us introduce a coherent potential function $\mathcal{P}^{\alpha}$ by defining it in the following way:

$$\overline{\mathbf{g}}^{\alpha} \equiv [\mathcal{P}^{\alpha} - \mathbf{S}^{\alpha}]^{-1} \tag{5.19}$$

The unknown coherent potential function $\mathcal{P}^\alpha$ describes an effective medium after configurational average which recovers the translational invariance of the system.

The coherent potential $\mathcal{P}^\alpha$ is in general a site non-diagonal and non-statistical quantity. Its determination can be significantly simplified by using the site-diagonal approximation[34, 35],

$$\mathcal{P}^\alpha{}_{RR'}(z) = \mathcal{P}^\alpha{}_R(z)\delta_{R,R'} \tag{5.20}$$

where all the site non-diagonal matrix elements are set to zero. The coherent potential function describes the properties of effective non-random atoms which characterize the system after configurational average. The main task of CPA is to find the unknown site-diagonal coherent potential function $\mathcal{P}^\alpha_R$. To do so, we proceed by the following identity:

$$P^\alpha - \mathbf{S}^\alpha \equiv (\mathcal{P}^\alpha - \mathbf{S}^\alpha) + P^\alpha - \mathcal{P}^\alpha \ . \tag{5.21}$$

Using $\mathbf{g}^\alpha = [P^\alpha - \mathbf{S}^\alpha]^{-1}$, and $\overline{\mathbf{g}}^\alpha = [\mathcal{P}^\alpha - \mathbf{S}^\alpha]^{-1}$, we obtain

$$\mathbf{g}^{\alpha,-1} = \overline{\mathbf{g}}^{\alpha,-1} + P^\alpha - \mathcal{P}^\alpha \ . \tag{5.22}$$

Therefore,

$$\mathbf{g}^\alpha = \overline{\mathbf{g}}^\alpha + \overline{\mathbf{g}}^\alpha(\mathcal{P}^\alpha - P^\alpha)\mathbf{g}^\alpha = \overline{\mathbf{g}}^\alpha + \overline{\mathbf{g}}^\alpha \mathbf{T} \overline{\mathbf{g}}^\alpha \tag{5.23}$$

where T is the T-matrix which contains all the disorder scattering processes in the system:

$$\mathbf{T} = (\mathcal{P}^\alpha - P^\alpha) + (\mathcal{P}^\alpha - P^\alpha)\overline{\mathbf{g}}^\alpha \mathbf{T} \tag{5.24}$$

It is clear that the T-matrix has a functional dependence on $\mathcal{P}^\alpha$. Performing configurational average on the above equation, we immediately obtain:

$$\overline{\mathbf{T}} = 0 \tag{5.25}$$

which provides a self-consistent condition for calculating the coherent potential function $\mathcal{P}^\alpha$. The self-consistent solution leads to physically meaningful results for all the

physical properties of the disordered system, and preserves all the analytical and symmetry properties of the averaged material. However, it is rather difficult to solve this self-consistent equation exactly and some approximations are needed. We will now introduce the widely used single-site approximation to determine the site-diagonal matrix $\mathcal{P}^\alpha$ self-consistently.

Introducing a single-site scattering matrix $\mathbf{t}_R$ associated with the site R,

$$\mathbf{t}_R = (\mathcal{P}_R^\alpha - P_R^\alpha) + (\mathcal{P}_R^\alpha - P_R^\alpha)\overline{\mathbf{g}}^\alpha \mathbf{t}_R \tag{5.26}$$

namely $\mathbf{t}_R = [1 - (\mathcal{P}_R^\alpha - P_R^\alpha)\overline{\mathbf{g}}]^{-1}(\mathcal{P}_R^\alpha - P_R^\alpha) = (\mathcal{P}_R^\alpha - P_R^\alpha)[1 - \overline{\mathbf{g}}(\mathcal{P}_R^\alpha - P_R^\alpha)]^{-1}$, we can express the total $\mathbf{T}$-matrix in terms of the single-site property $\mathbf{t}_R$. Let $\triangle_R \equiv \mathcal{P}_R^\alpha - P_R^\alpha$, we obtain

$$\mathbf{t}_R = [1 - \triangle_R \overline{\mathbf{g}}^\alpha]^{-1} \triangle_R = \triangle_R [1 - \overline{\mathbf{g}}^\alpha \triangle_R]^{-1} \tag{5.27}$$

Hence

$$\mathbf{T} = \sum_R (\mathcal{P}_R^\alpha - P_R^\alpha)(1 + \overline{\mathbf{g}}^\alpha \mathbf{T}) = \sum_R \triangle_R (1 + \overline{\mathbf{g}}^\alpha \mathbf{T}) = \sum_R \mathcal{Q}_R \tag{5.28}$$

where the $\mathcal{Q}_R \equiv \triangle_R(1 + \overline{\mathbf{g}}^\alpha \mathbf{T}) = \triangle_R(1 + \overline{\mathbf{g}}^\alpha \mathcal{Q}_R + \overline{\mathbf{g}}^\alpha \sum\limits_{R' \neq R} \mathcal{Q}_{R'})$, or:

$$\mathcal{Q}_R = \mathbf{t}_R(1 + \overline{\mathbf{g}}^\alpha \sum_{R' \neq R} \mathcal{Q}_{R'}) = (1 + \sum_{R' \neq R} \mathcal{Q}_{R'} \overline{\mathbf{g}}^\alpha)\mathbf{t}_R \tag{5.29}$$

Here, $\mathcal{Q}_R$ describes the total contribution of the atom on site R to the $\mathbf{T}$-matrix and it is different from the single-site quantity $\mathbf{t}_R$. The above equation expresses the strength of an individual scatterer as a product of an isolated scattering event and a factor describing the transformation of an unperturbed wave incident on site R into an effective wave due to the multiple scattering events. Using Eq.(5.29) yields the multiple scattering equation:

$$\begin{aligned} \mathbf{T} = \sum_R \mathcal{Q}_R = \sum_R \mathbf{t}_R + \sum_R \mathbf{t}_R \overline{\mathbf{g}}^\alpha \sum_{R' \neq R} \mathbf{t}_{R'} + \\ \sum_R \mathbf{t}_R \overline{\mathbf{g}}^\alpha \sum_{R' \neq R} \mathbf{t}_{R'} \overline{\mathbf{g}}^\alpha \sum_{R'' \neq R'} \mathbf{t}_{R''} + \cdots \end{aligned} \tag{5.30}$$

where $\mathbf{T}$ is written as a sum of single-site scattering contributions. In this multiple scattering equation, the electron can not scatter twice in sequence on the same atomic site. The Eq.(5.26) to Eq.(5.30) are exact and represent a closed form of the multiple scattering theory. They lead to the exact configurational average,

$$\overline{\mathbf{T}} = \sum_R \overline{\mathcal{Q}}_R \tag{5.31}$$

$$\overline{\mathcal{Q}}_R = \overline{\mathbf{t}_R(1 + \overline{\mathbf{g}}^{\alpha} \sum_{R' \neq R} \mathcal{Q}_{R'})} \tag{5.32}$$

To proceed further, we introduce a single site approximation (SSA)[30] to decouple the individual single-site scattering events in Eq.(5.30). Application of SSA means the configurational average over the T-matrix is done in the following approximate manner:

$$\begin{aligned}
\overline{\mathcal{Q}}_R &= \overline{\mathbf{t}_R(1 + \overline{\mathbf{g}}^{\alpha} \sum_{R' \neq R} \mathcal{Q}_{R'})} \\
&= \overline{\mathbf{t}}_R(1 + \overline{\mathbf{g}}^{\alpha} \sum_{R' \neq R} \overline{\mathcal{Q}}_{R'}) + \overline{\mathbf{t}_R(1 + \overline{\mathbf{g}}^{\alpha} \sum_{R' \neq R} \mathcal{Q}_{R'} - \overline{\mathcal{Q}}_{R'})} \\
&\approx \overline{\mathbf{t}}_R(1 + \overline{\mathbf{g}}^{\alpha} \sum_{R' \neq R} \overline{\mathcal{Q}}_{R'})
\end{aligned} \tag{5.33}$$

Here, the first term describes the average scattered wave by the individual atom on site R; the second term contains fluctuations away from the average wave. Neglecting the second term forms the single-site approximation. The physical meaning of SSA is that during the multiple scattering process, the electron wave scatters off one impurity at a time. In other words, the scattering events from the sites at different times are independent of each other. Since the probability is small for scattering off multiple impurities at the same time, SSA is a good approximation. The SSA becomes exact in dilute limit of disorder.

Applying SSA, we can rewrite the averaged T-matrix as:

$$\overline{\mathbf{T}} \approx \sum_R \overline{\mathbf{t}}_R + \sum_R \overline{\mathbf{t}}_R \overline{\mathbf{g}}^{\alpha} \sum_{R' \neq R} \overline{\mathbf{t}}_{R'} + \sum_R \overline{\mathbf{t}}_R \overline{\mathbf{g}}^{\alpha} \sum_{R' \neq R} \overline{\mathbf{t}}_{R'} \overline{\mathbf{g}}^{\alpha} \sum_{R'' \neq R'} \overline{\mathbf{t}}_{R''} + \cdots \quad (5.34)$$

Finally, the CPA self-consistent condition Eq.(5.25) is immediately simplified to $\overline{\mathbf{t}}_R = 0$ for each site in the system. This simplified self-consistent condition of CPA reads

$$\overline{\mathbf{t}}_R = C_R^A \mathbf{t}_R^A + C_R^B \mathbf{t}_R^B = 0 \quad (5.35)$$

where $\mathbf{t}_R^Q = (\mathcal{P}_R^{\alpha} - P_R^{\alpha,Q})[1 - \overline{\mathbf{g}}(\mathcal{P}_R^{\alpha} - P_R^{\alpha,Q})]^{-1}$.

Therefore, the coherent potential function is found to satisfy the following self-consistent equation:

$$\mathcal{P}_R^{\alpha} = \overline{P}_R^{\alpha} + (\mathcal{P}_R^{\alpha} - P_R^{\alpha,A}) \overline{\mathbf{g}}^{\alpha} (\mathcal{P}_R^{\alpha} - P_R^{\alpha,B}) \quad (5.36)$$

where $\overline{P}_R^{\alpha} = (C_R^A P_R^{\alpha,A} + C_R^B P_R^{\alpha,B})$. In Eq.(5.36), only the site-diagonal matrix element of $\overline{\mathbf{g}}_{RR}^{\alpha}$ is required. Thus, Eq.(5.36) and Eq.(5.19) consist of a closed system of self-consistent equations. Due to the translational invariance after configurational average, the elements of the quantity $\overline{\mathbf{g}}_{RR}^{\alpha}$ within an unit cell can be calculated using computational techniques developed for crystals such as the lattice Fourier transformation.

To solve Eq.(5.36) self-consistently, one starts by setting $\mathcal{P}_R^{\alpha} = \overline{P}_R^{\alpha}$; then calculates the site-diagonal elements of $\overline{\mathbf{g}}_{RR}^{\alpha}$ which gives a new coherent potential function $\mathcal{P}_R^{\alpha}$ by Eq.(5.36) for the next iteration step, and this process is repeated until self-consistency is achieved. In practical first principle application of CPA, a quantity called coherent interactor (see Chapter 4.2 in Ref.[75]) is introduced for simplifying the solution of the CPA equations. From Eq.(5.36), we observe that SSA allows us to achieve the average over the whole system self-consistently by carrying out calculation on each individual site one by one.

In summary, the single site approximation within CPA [30] provides a very powerful way to calculate the average auxiliary Green's function $\overline{\mathbf{g}}^{\alpha}$. In particular the coherent potential function $\mathcal{P}_R^{\alpha}$ can be calculated self-consistently from Eq.(5.36). In the next Section, we apply SSA again to decouple the individual scattering events t_R in order to calculate $\overline{\mathbf{g}}^{a,<}$. A new quantity will emerge which is the nonequilibrium vertex correction (NVC).

5.3 Non-equilibrium vertex correction

Consider a two-probe device such as that shown in Fig.4.1, the configurational average over the auxiliary non-equilibrium Green's function give us,

$$\overline{\mathbf{g}}_{cc}^{\alpha,<}(z) = \overline{\mathbf{g}_{cc}^{\alpha,\mathcal{R}}\boldsymbol{\Sigma}^{\alpha,<}\mathbf{g}_{cc}^{\alpha,\mathcal{A}}} \tag{5.37}$$

where the averaging is carried out over the product of three quantities: $\mathbf{g}_{cc}^{\alpha,\mathcal{R}}$, $\boldsymbol{\Sigma}^{\alpha,<}$, and $\mathbf{g}_{cc}^{\alpha,\mathcal{A}}$ which may all be statistical quantities for disordered systems. The mathematical method for averaging the product of three quantities is much more complicated than that for averaging the product of two random quantities. For our two-probe system, we shall assume that impurities only exists inside the central scattering region of the device and not inside the left/right electrodes. This way, the self-energy $\boldsymbol{\Sigma}^{<,\alpha}$ (see Eq.(4.143)) is not a statistical quantity. This assumption is not limiting because the scattering region contains a portion of electrodes (see Fig.4.2), namely our formalism allows disorder to appear in the that portion of the electrodes as well. Within this assumption, the configurational average of the above equation becomes the average over two statistical quantities $\mathbf{g}_{cc}^{\alpha,\mathcal{R}}$ and $\mathbf{g}_{cc}^{\alpha,\mathcal{A}}$, connected by the non-statistical self-energy matrix $\boldsymbol{\Sigma}^{\alpha,<}$.

For a two-probe system, the auxiliary Green's function of the central region can be expressed as follows for a particular disorder configuration of the central region

(see Eq.(4.126)),

$$\mathbf{g}^{\alpha}_{cc}(z) = [P^{\alpha}_{cc}(z) - \mathbf{S}^{\alpha}_{cc} - \mathbf{\Sigma}^{\alpha}_{L}(z) - \mathbf{\Sigma}^{\alpha}_{R}(z)]^{-1} \tag{5.38}$$

and using CPA of Section 5.2, we can write down its configurational average,

$$\overline{\mathbf{g}}^{\alpha}_{cc}(z) = [\mathcal{P}^{\alpha}_{cc}(z) - \mathbf{S}^{\alpha}_{cc} - \mathbf{\Sigma}^{\alpha}_{L}(z) - \mathbf{\Sigma}^{\alpha}_{R}(z)]^{-1} \tag{5.39}$$

Even though the lesser self-energy $\mathbf{\Sigma}^{\alpha,<}$ is independent of impurity configurations of the central region, $\overline{\mathbf{g}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\mathbf{g}^{\alpha,\mathcal{A}}} \neq \overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}$, because impurity scattering correlates $\mathbf{g}^{\alpha,\mathcal{R}}$ and $\mathbf{g}^{\alpha,\mathcal{A}}$. To find the accurate expression for $\overline{\mathbf{g}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\mathbf{g}^{\alpha,\mathcal{A}}}$, we can start by making the substitutions,

$$\mathbf{g}^{\alpha,\mathcal{R}} = \overline{\mathbf{g}}^{\alpha,\mathcal{R}} + \overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{T}^{\mathcal{R}}\overline{\mathbf{g}}^{\alpha,\mathcal{R}} \tag{5.40}$$

$$\mathbf{g}^{\alpha,\mathcal{A}} = \overline{\mathbf{g}}^{\alpha,\mathcal{A}} + \overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathbf{T}^{\mathcal{A}}\overline{\mathbf{g}}^{\alpha,\mathcal{A}} \tag{5.41}$$

where $\mathbf{T}^{\mathcal{A}}$ and $\mathbf{T}^{\mathcal{R}}$ are the total T-matrices corresponding to the retarded and advanced auxiliary Green's functions, and they satisfy the relation $\mathbf{T}^{\mathcal{A}} = [\mathbf{T}^{\mathcal{R}}]^{\dagger}$. The average auxiliary NEGF can be immediately written,

$$\begin{aligned}
\overline{\mathbf{g}}^{\alpha,<}_{cc}(z) &= \overline{\mathbf{g}^{\alpha,\mathcal{R}}_{cc}\mathbf{\Sigma}^{\alpha,<}\mathbf{g}^{\alpha,\mathcal{A}}_{cc}} \\
&= \overline{[\overline{\mathbf{g}}^{\alpha,\mathcal{R}} + \overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{T}^{\mathcal{R}}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}]\mathbf{\Sigma}^{\alpha,<}[\overline{\mathbf{g}}^{\alpha,\mathcal{A}} + \overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathbf{T}^{\mathcal{A}}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}]} \\
&= \overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}} + \overline{\mathbf{g}}^{\alpha,\mathcal{R}}\overline{[\mathbf{T}^{\mathcal{R}}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathbf{T}^{\mathcal{A}}]}\overline{\mathbf{g}}^{\alpha,\mathcal{A}} \\
&= \overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}} + \overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Omega}^{\alpha}_{NVC}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}
\end{aligned} \tag{5.42}$$

$$\mathbf{\Omega}^{\alpha}_{NVC} = \overline{[\mathbf{T}^{\mathcal{R}}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathbf{T}^{\mathcal{A}}]} \ . \tag{5.43}$$

Here the self-consistent condition of CPA, $\overline{\mathbf{T}} = 0$ (Eq.(5.25)), is applied. The configurationally averaged quantities $\overline{\mathbf{g}}^{\alpha,\mathcal{R}}_{cc}$ and $\overline{\mathbf{g}}^{\alpha,\mathcal{A}}_{cc}$ are obtained by CPA using the single site approximation. The quantity $\mathbf{\Omega}^{\alpha}_{NVC}$ is the central quantity which is the non-

equilibrium vertex correction (NVC). The average in $\mathbf{\Omega}^{\alpha}_{NVC}$ is an average over the pair of $\mathbf{T}^{\mathcal{R}}$ and $\mathbf{T}^{\mathcal{A}}$ connected by the already averaged quantity $\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}$ which do not depend on specific impurity configurations any more. The NVC is a self-energy that represents multiple impurity scattering in the device under non-equilibrium conditions. The equation Eq.(5.42) is exact for the averaged auxiliary NEGF. Its first term is the coherent part whose calculation was discussed in detail in Chapter 4. The NVC term can play very important and even dominanting roles for quantum transport. Once $\mathbf{\Omega}^{\alpha}_{NVC}$ is calculated, we obtain the averaged non-equilibrium density matrix that is used to construct the Hamiltonian of the system.

We now derive the NVC self-energy $\mathbf{\Omega}^{\alpha}_{NVC}$. Inserting Eq.(5.28), namely $\mathbf{T} = \sum_{R} \mathcal{Q}_R$, into Eq.(5.43), we obtain

$$\mathbf{\Omega}_{NVC} = \sum_{R,R'} \overline{\mathcal{Q}^{\mathcal{R}}_{R}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathcal{Q}^{\mathcal{A}}_{R'}} \ , \tag{5.44}$$

$$\overline{\mathcal{Q}^{\mathcal{R}}_{R}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\Sigma^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathcal{Q}^{\mathcal{A}}_{R'}} = \overline{\mathbf{t}^{\mathcal{R}}_{R}(1+\sum_{M\neq R}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathcal{Q}^{\mathcal{R}}_{M})\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathcal{Q}^{\mathcal{A}}_{R'}} \tag{5.45}$$

where $\mathcal{Q}_R = \mathbf{t}_R(1+\mathbf{g}^{\alpha}\sum_{M\neq R}\mathcal{Q}_M) = (1+\sum_{N\neq R}\mathcal{Q}_N\mathbf{g}^{\alpha})\mathbf{t}_R$. We apply SSA to decouple the single-site scattering matrix $\mathbf{t}_R$ and the quantity $\mathcal{Q}_{R'}$ when $R \neq R'$. Therefore, for $R \neq R'$, we have the average

$$\overline{\mathcal{Q}^{\mathcal{R}}_{R}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathcal{Q}^{\mathcal{A}}_{R'}} = \overline{\mathbf{t}}^{\mathcal{R}}_{R}\overline{(1+\sum_{M\neq R}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathcal{Q}^{\mathcal{R}}_{M})\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathcal{Q}^{\mathcal{A}}_{R'}} = 0 \tag{5.46}$$

Hence, in SSA, $\mathbf{\Omega}^{\alpha}_{NVC}$ is actually a site-diagonal matrix:

$$\mathbf{\Omega}_{NVC} = \sum_{R}\mathbf{\Omega}_{NVC,R} \tag{5.47}$$

Namely, NVC is a sum of single-site contributions:

$$\mathbf{\Omega}_{NVC,R} = \overline{\mathcal{Q}^{\mathcal{R}}_{R}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathcal{Q}^{\mathcal{A}}_{R}}$$

$$= \overline{\mathbf{t}_R^{\mathcal{R}}(1+\sum_{M\neq R}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathcal{Q}_M^{\mathcal{R}})\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}(1+\sum_{N\neq R}\mathcal{Q}_M^{\mathcal{A}}\overline{\mathbf{g}}^{\alpha,\mathcal{A}})\mathbf{t}_R^{\mathcal{A}}} \tag{5.48}$$

Again, apply SSA, $\mathbf{\Omega}_{NVC,R}$ can be approximated by the following procedure,

$$\begin{aligned}
&\mathbf{\Omega}_{NVC,R}\\
&= \overline{\mathbf{t}_R^{\mathcal{R}}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathbf{t}_R^{\mathcal{A}}} + \overline{\mathbf{t}_R^{\mathcal{R}}\sum_{M,N\neq R}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathcal{Q}_M^{\mathcal{R}}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathcal{Q}_N^{\mathcal{A}}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathbf{t}_R^{\mathcal{A}}}\\
&= \overline{\mathbf{t}_R^{\mathcal{R}}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathbf{t}_R^{\mathcal{A}}} + \overline{\mathbf{t}_R^{\mathcal{R}}\sum_{M,N\neq R}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\overline{\mathcal{Q}_M^{\mathcal{R}}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathcal{Q}_N^{\mathcal{A}}}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathbf{t}_R^{\mathcal{A}}}\\
&= \overline{\mathbf{t}_R^{\mathcal{R}}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathbf{t}_R^{\mathcal{A}}} + \overline{\mathbf{t}_R^{\mathcal{R}}\sum_{M\neq R}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\overline{\mathcal{Q}_M^{\mathcal{R}}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathcal{Q}_M^{\mathcal{A}}}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathbf{t}_R^{\mathcal{A}}}\\
&= \overline{\mathbf{t}_R^{\mathcal{R}}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathbf{t}_R^{\mathcal{A}}} + \overline{\mathbf{t}_R^{\mathcal{R}}\sum_{M\neq R}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Omega}_{NVC,M}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathbf{t}_R^{\mathcal{A}}}
\end{aligned} \tag{5.49}$$

Finally, we obtained a closed set of linear equations for the unknown NVC. In Eq.(5.49), the average is over the pair of scattering events on the same site. In other words the scattering from different sites is regarded as statistically uncorrelated and the motion of two particles in the medium is correlated only if they both scatter from the same site. This two-particle decoupling is consistent with the assumption that disorder is completely random.

By solving Eq.(5.49), we obtain the averaged auxiliary NEGF. The configurational average over the auxiliary NEGF is given by

$$\overline{\mathbf{g}}^{\alpha,<} = \overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}} + \sum_R \overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Omega}_{NVC,R}^{\alpha}\overline{\mathbf{g}}^{\alpha,\mathcal{A}} \tag{5.50}$$

where the vertex correction to the average auxiliary NEGF is a sum of single-site contribution. Using Eq.(5.49), we can express $\overline{\mathbf{g}}^{\alpha,<}$ by expanding it as an infinite series:

$$\begin{aligned}
\overline{\mathbf{g}}^{\alpha,<} = \overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}} + \sum_R \overline{\mathbf{g}}^{\alpha,\mathcal{R}}\overline{\mathbf{t}_R^{\mathcal{R}}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathbf{t}_R^{\mathcal{A}}}\overline{\mathbf{g}}^{\alpha,\mathcal{A}} +\\
\sum_{R\neq R'}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\overline{\mathbf{t}_R^{\mathcal{R}}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\overline{\mathbf{t}_{R'}^{\mathcal{R}}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathbf{t}_{R'}^{\mathcal{A}}}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\mathbf{t}_R^{\mathcal{A}}}\overline{\mathbf{g}}^{\alpha,\mathcal{A}} + \cdots
\end{aligned} \tag{5.51}$$

This expression may be written diagrammatically as in Fig.5.1 in which the solid lines are the averaged propagators $\overline{g}^{\alpha,\mathcal{A}/\mathcal{R}}$ and the dotted lines are associated with two particle scattering from a single site denoted by a cross. The rules for obtaining this diagram are: 1) no crossed line is allowed - this is equivalent to neglect spatial correlations; 2) two adjacent unconnected dotted lines represent scattering from two different sites. This reflects the fact that a particle is not allowed to scatter twice in succession at the same site; 3) all unconnected lines are averaged over separately. These three rules represent the implementation of the single-site decoupling to the average auxiliary NEGF. The first line in Fig.5.1 gives the sequence of scattering events that contribute to $\overline{\mathbf{g}}^{\alpha,<}$. The single-site decoupling technique can be observed in this figure: the diagram on the left hand side represents $\mathbf{g}^{\alpha,<}$; the first diagram on right side is the first term in Eq.(5.51); the second diagram on the right hand side represents two-particle scattering at a single site, and the third diagram represents two-particle scattering at two different sites. These diagrams are the second and third terms of Eq.(5.51). The infinite series of diagrams are summed diagrammatically in the second line of Fig.5.1 in term of the vertex correction defined in Eq.(5.49).

Since any $\overline{(\cdots)}$ is the average over single-site quantity, the final result of Eq.(5.49) can be rewritten explicitly as a sum of contributions from each atomic species on the site:

$$\begin{aligned}\mathbf{\Omega}_{NVC,R} = & \sum_{Q=A,B} C_R^Q \mathbf{t}_R^{Q,\mathcal{R}} [\overline{\mathbf{g}}^{\alpha,\mathcal{R}} \mathbf{\Sigma}^{\alpha,<} \overline{\mathbf{g}}^{\alpha,\mathcal{A}}]_{RR} \mathbf{t}_R^{Q,\mathcal{A}} \\ & + \sum_{Q=A,B} C_R^Q \mathbf{t}_R^{Q,\mathcal{R}} [\overline{\mathbf{g}}^{\alpha,\mathcal{R}} \sum_{R' \neq R} \mathbf{\Omega}_{NVC,R'} \overline{\mathbf{g}}^{\alpha,\mathcal{A}}]_{RR} \mathbf{t}_R^{Q,\mathcal{A}} \ . \end{aligned} \tag{5.52}$$

This site decomposition is similar to that done in equilibrium calculations[30, 37].

To proceed further, we note that after impurity average, the two-probe device system in Fig.4.2 restores translational invariance along the transverse (x,y) directions. Therefore $\mathbf{\Omega}_{NVC,R}$ can be calculated within an unit cell of the central scattering region plus a k-sampling in the 2-dimensional (2D) transverse Brillouin zone (BZ). To do so,

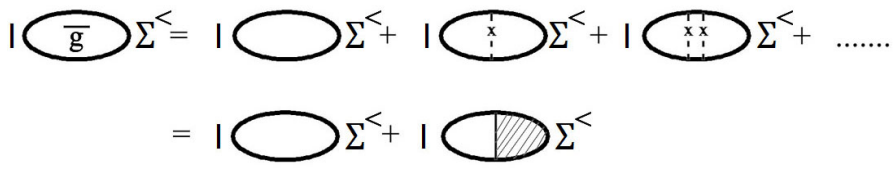

Figure 5.1: Diagrammatical representation of the average auxiliary NEGF of Eq.(5.51)

we add and subtract a $R = R'$ term in the second term of Eq.(5.52),

$$\begin{aligned}
\mathbf{\Omega}_{NVC,R} &= \sum_{Q=A,B} C_R^Q \mathbf{t}_R^{Q,\mathcal{R}} [\overline{\mathbf{g}}^{\alpha,\mathcal{R}} \mathbf{\Sigma}^{\alpha,<} \overline{\mathbf{g}}^{\alpha,\mathcal{A}}]_{RR} \mathbf{t}_R^{Q,\mathcal{A}} \\
&- \sum_{Q=A,B} C_R^Q \mathbf{t}_R^{Q,\mathcal{R}} \overline{\mathbf{g}}_{RR}^{\alpha,\mathcal{R}} \mathbf{\Omega}_{NVC,R} \overline{\mathbf{g}}_{RR}^{\alpha,\mathcal{A}} \mathbf{t}_R^{Q,\mathcal{A}} \\
&+ \sum_{Q=A,B} C_R^Q \mathbf{t}_R^{Q,\mathcal{R}} [\overline{\mathbf{g}}^{\alpha,\mathcal{R}} \sum_{R'} \mathbf{\Omega}_{NVC,R'} \overline{\mathbf{g}}^{\alpha,\mathcal{A}}]_{RR} \mathbf{t}_R^{Q,\mathcal{A}}
\end{aligned} \tag{5.53}$$

After a 2D Fourier transform, the above equation changes to:

$$\begin{aligned}
\mathbf{\Omega}_{NVC,R} &= \sum_{Q=A,B} C_R^Q \mathbf{t}_R^{Q,\mathcal{R}} \frac{1}{N_{k_\parallel}} \sum_{k_\parallel} [\overline{\mathbf{g}}^{\alpha,\mathcal{R}}(k_\parallel, E) \mathbf{\Sigma}^{\alpha,<} \overline{\mathbf{g}}^{\alpha,\mathcal{A}}(k_\parallel, E)]_{RR} \mathbf{t}_R^{Q,\mathcal{A}} \\
&- \sum_{Q=A,B} C_R^Q \mathbf{t}_R^{Q,\mathcal{R}} \overline{\mathbf{g}}_{RR}^{\alpha,\mathcal{R}} \mathbf{\Omega}_{NVC,R} \overline{\mathbf{g}}_{RR}^{\alpha,\mathcal{A}} \mathbf{t}_R^{Q,\mathcal{A}} \\
&+ \sum_{Q=A,B} C_R^Q \mathbf{t}_R^{Q,\mathcal{R}} [\sum_{R'} \frac{1}{N_{k_\parallel}} \sum_{k_\parallel} \overline{\mathbf{g}}_{RR'}^{\alpha,\mathcal{R}}(k_\parallel, E) \mathbf{\Omega}_{NVC,R'} \times \overline{\mathbf{g}}_{R'R}^{\alpha,\mathcal{A}}(k_\parallel, E)] \mathbf{t}_R^{Q,\mathcal{A}}
\end{aligned} \tag{5.54}$$

where the site-diagonal property of $\mathbf{\Omega}_{NVC}$ has been applied in the Fourier transformation, all the site labels R and R' are now defined within the unit cell of the central region of the device. In the above equation, all the quantities have been defined previously: $\mathbf{g}^\alpha(k_\parallel, E)$ is defined in Eq.(5.27), site-diagonal element $\mathbf{g}_{RR}^\alpha$ and $\mathbf{t}_R$ are given by Eq.(5.27) and Eq.(5.26), respectively. Finally, the retarded and advanced quantities are not independent, they obey $\overline{\mathbf{g}}^{\alpha,\mathcal{A}}(k_\parallel, E) = \overline{\mathbf{g}}^{\alpha,\mathcal{R},\dagger}(k_\parallel, E)$, $\mathbf{t}^{\mathcal{A}} = \mathbf{t}^{\mathcal{R},\dagger}$.

After finding all the relevant quantities, Eq.(5.54) can be solved explicitly by linear algebra or implicitly by iteration. Obviously, for a clean sites R with only one atomic

specie, $\mathbf{\Omega}_{NVC,R} = 0$ because $\mathbf{t}_R = 0$ on clean sites. Therefore, Eq.(5.54) can be simplified to only include disordered sites. Hence, for the third term of Eq.(5.54), only the elements $\mathbf{g}^{\alpha}_{R,R'}(k_{\parallel}, E)$ associated two disordered sites R and R' are required in the evaluation of $\mathbf{\Omega}_{NVC}$. This is very useful for simplifying numerical computation. For devices possessing some geometrical symmetry, one can reduce the BZ integration in Eq.(5.54).

In summary, based on SSA within CPA, the averaged auxiliary Green's function has been formulated and NVC self-energy $\mathbf{\Omega}_{NVC}$ is derived. These are necessary in order to obtain configurationally averaged auxiliary NEGF $\overline{\mathbf{g}}^{\alpha,<}$. In the NEGF-DFT self-consistent calculation, the component-projected local electron density matrix $\overline{\mathbf{n}}^{Q}_{R,LL'}$ (See section 5.6) is obtained by the conditionally averaged physical quantities $\overline{\mathbf{G}}^{Q}_{RR}$ for equilibrium and $\overline{\mathbf{G}}^{<,Q}_{RR}$ for non-equilibrium which are simply connected with $\mathbf{g}^{\alpha,Q}$ and $\mathbf{g}^{\alpha,<,Q}$ through Eqs.(5.11,5.14). Thus the next key step is to formulate the conditionally averaged auxiliary Green's functions $\overline{\mathbf{g}}^{\alpha,Q}$ and$\overline{\mathbf{g}}^{\alpha,<,Q}$.

5.4 Conditionally averaged site-diagonal quantities: $\overline{\mathbf{g}}^{\alpha,Q}_{RR}$ and $\overline{\mathbf{g}}^{\alpha,<,Q}_{RR}$

To calculate the conditionally averaged site-diagonal auxiliary Green's function $\overline{\mathbf{g}}^{\alpha,Q}_{RR}$ in Eq.(5.12) and $\overline{\mathbf{g}}^{\alpha,<,Q}_{RR}$ in Eq.(5.15), we first employ the projection operator-like property [34] of η^{Q}_{R} to derive $\overline{\eta^{Q}_{R}\mathbf{g}^{\alpha}_{RR}}$ of Eq.(5.12) and $\overline{\eta^{Q}_{R}\mathbf{g}^{\alpha,<}_{RR}}$ of Eq.(5.15).

According to the definition of occupations η^{A}_{R} and η^{B}_{R} (see Eq.(5.2) Section 5.1), we have:

$$\eta^{A}_{R} + \eta^{B}_{R} = 1 \tag{5.55}$$

$$\eta^{A}_{R} P^{\alpha,A}_{R} + \eta^{B}_{R} P^{\alpha,B}_{R} = P^{\alpha}_{R} \tag{5.56}$$

which are always satisfied in a particular disorder configuration of impurities. Here,

the potential function P_R^α is a statistical quantity. We solve to obtain:

$$\eta_R^A = \Delta(P_R^\alpha)^{-1}\left(P_R^{\alpha,B} - P_R^\alpha\right) \tag{5.57}$$

$$\eta_R^B = -\Delta(P_R^\alpha)^{-1}\left(P_R^{\alpha,A} - P_R^\alpha\right) \tag{5.58}$$

where $\Delta(P_R^\alpha) \equiv (P_R^{\alpha,B} - P_R^{\alpha,A})$, the occupation labels η_R^Q are handled as site-diagonal matrices which may be a zero matrix or an identity matrix. Using these expressions, we obtain:

$$\begin{aligned}\overline{\eta^A \mathbf{g}^\alpha} &= \overline{\Delta(P^\alpha)^{-1}(P^{\alpha,B} - P^\alpha)\mathbf{g}^\alpha} \\ &= \Delta(P^\alpha)^{-1}[(P^{\alpha,B}\overline{\mathbf{g}}^\alpha) - \overline{(P^\alpha \mathbf{g}^\alpha)}]\end{aligned} \tag{5.59}$$

This expression is simplified by using Eq.(5.22),

$$\overline{\eta^A \mathbf{g}^\alpha} = \Delta(P^\alpha)^{-1}(P^{\alpha,B} - \mathcal{P}^\alpha)\overline{\mathbf{g}}^\alpha \tag{5.60}$$

where $\mathcal{P}^\alpha$ is the coherent potential introduced in Section 5.2 to describe the averaged effective medium. With this result, we can write down the conditionally averaged auxiliary Green's function as follows:

$$\begin{aligned}\overline{\mathbf{g}}_{RR}^{\alpha,A} &= \overline{\eta_R^A \mathbf{g}_{RR}^\alpha}/C_R^A \\ &= \Delta(P_R^\alpha)^{-1}(P_R^{\alpha,B} - \mathcal{P}_R^\alpha)\overline{\mathbf{g}}_{RR}^\alpha/C_R^A\end{aligned} \tag{5.61}$$

Using $C_R^A = \overline{\eta_R^A}$ and Eq.(5.57), apply the CPA self-consistent equation (5.36), we finally obtain the conditionally averaged auxiliary Green's function for atomic specie A as follows:

$$\overline{\mathbf{g}}_{RR}^{\alpha,A} = [1 - \overline{\mathbf{g}}_{RR}^\alpha(\mathcal{P}_R^\alpha - P_R^{\alpha,A})]^{-1}\overline{\mathbf{g}}_{RR}^\alpha \ . \tag{5.62}$$

Similarly, for atomic specie B we have:

$$\overline{\mathbf{g}}^{\alpha,B}_{RR} = [1 - \overline{\mathbf{g}}^{\alpha}_{RR}(\mathcal{P}^{\alpha}_{R} - P^{\alpha,B}_{R})]^{-1}\overline{\mathbf{g}}^{\alpha}_{RR} \, . \tag{5.63}$$

From these results, one can confirm that the relation $\overline{\mathbf{g}}^{\alpha}_{RR} = C^{A}_{R}\overline{\mathbf{g}}^{\alpha,A}_{RR} + C^{B}_{R}\overline{\mathbf{g}}^{\alpha,B}_{RR}$ is, indeed, satisfied.

Next, we calculate the conditionally averaged site-diagonal NEGF $\overline{\mathbf{g}}^{\alpha,<,Q}_{RR}$ by Eq.(5.15). We start from:

$$\begin{aligned}
\overline{\eta^{A}\mathbf{g}^{\alpha,<}} &= \Delta(P^{\alpha})^{-1}\overline{(P^{\alpha,B} - P^{\alpha,+})\mathbf{g}^{\alpha,<}} \\
&= \Delta(P^{\alpha})^{-1}[P^{\alpha,B}\overline{\mathbf{g}}^{\alpha,<} - \overline{P^{\alpha,+}\mathbf{g}^{\alpha,<}}]
\end{aligned}$$

here $\mathbf{g}^{\alpha,<} = \mathbf{g}^{\alpha,\mathcal{R}}\mathbf{\Sigma}^{\alpha,<}\mathbf{g}^{\alpha,\mathcal{A}}$, $\overline{\mathbf{g}}^{\alpha,<} = \overline{\mathbf{g}}^{\alpha,\mathcal{R}}(\mathbf{\Sigma}^{\alpha,<} + \mathbf{\Omega}_{NVC})\overline{\mathbf{g}}^{\alpha,\mathcal{A}}$ and $P^{\alpha,+} = \mathcal{P}^{\alpha,+} - \overline{\mathbf{g}}^{\alpha,\mathcal{R},-1} + \mathbf{g}^{\alpha,\mathcal{R},-1}$ where $\mathcal{P}^{\alpha,+}$ is the coherent potential function corresponding to the $\overline{\mathbf{g}}^{\alpha,\mathcal{R}}$. Using these relations we can derive the following expression for $\overline{\eta^{A}\mathbf{g}^{\alpha,<}}$:

$$\begin{aligned}
\overline{\eta^{A}\mathbf{g}^{\alpha,<}} &= \Delta(P^{\alpha})^{-1}[(P^{\alpha,B} - \mathcal{P}^{\alpha,+})\overline{\mathbf{g}}^{\alpha,\mathcal{R}} \\
&\times(\mathbf{\Sigma}^{\alpha,<} + \mathbf{\Omega}_{NVC})\overline{\mathbf{g}}^{\alpha,\mathcal{A}} + \mathbf{\Omega}_{NVC}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}] \\
&= \Delta(P^{\alpha})^{-1}[(P^{\alpha,B} - \mathcal{P}^{\alpha,+})\overline{\mathbf{g}}^{\alpha,<} + \mathbf{\Omega}_{NVC}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}]
\end{aligned} \tag{5.64}$$

With this result, Eq.(5.15) gives the conditionally averaged auxiliary NEGF for atomic specie A:

$$\begin{aligned}
\overline{\mathbf{g}}^{\alpha,<,A}_{RR} &= \Delta(P^{\alpha}_{R})^{-1}[(P^{\alpha,B}_{R} - \mathcal{P}^{\alpha,+}_{R})\overline{\mathbf{g}}^{\alpha,<}_{RR} \\
&+\mathbf{\Omega}_{NVC,R}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}_{RR}]/C^{A}_{R}
\end{aligned} \tag{5.65}$$

A similar derivation gives, for atomic specie B:

$$\begin{aligned}
\overline{\mathbf{g}}^{\alpha,<,B}_{RR} &= -\Delta(P^{\alpha}_{R})^{-1}[(P^{\alpha,A}_{R} - \mathcal{P}^{\alpha,+}_{R})\overline{\mathbf{g}}^{\alpha,<}_{RR} \\
&+\mathbf{\Omega}_{NVC,R}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}_{RR}]/C^{B}_{R}
\end{aligned} \tag{5.66}$$

Finally, it can be checked that $\overline{\mathbf{g}}^{\alpha,<}_{RR} = C^A_R \overline{\mathbf{g}}^{\alpha,<,A}_{RR} + C^B_R \overline{\mathbf{g}}^{\alpha,<,B}_{RR}$.

Because configurational average restores translational invariance, the auxiliary quantities $\overline{\mathbf{g}}^{\alpha}_{RR}$ and $\overline{\mathbf{g}}^{\alpha,<}_{RR}$ in the above equations can be calculated by a lattice Fourier transformation. Using these results we can calculate the component-projected average local electron density matrix $\overline{\mathbf{n}}^{A/B}_R$. At equilibrium, the averaged local density matrix $\overline{\mathbf{n}}^{A/B}_R$ is given by the physical Green's function $\mathbf{G}^{A/B}_{RR}$ which are related to $\mathbf{g}^{\alpha,A/B}_{RR}$ through Eq.(5.11). At non-equilibrium, the averaged local density matrix is determined by the physical NEGF $\overline{\mathbf{G}}^{<,A/B}_{RR}$ which are related to $\overline{\mathbf{g}}^{\alpha,<,A/B}_{RR}$ through Eq.(5.14). The electronic density is obtained from the density matrix and the DFT Hamiltonian is then determined by Eq.(2.21) of Chapter 2. This process is repeated until numerical convergence of the entire NEGF-DFT-NVC iteration is reached. This way, the DFT self-consistent calculation of disordered systems becomes feasible using the NVC theory under non-equilibrium quantum transport conditions (see section 5.7 for implementation details).

5.5 Testing NVC

A stringent test to the NEGF-DFT-NVC formalism and its numerical implementation is to check if the fluctuation-dissipation theorem is satisfied at *equilibrium*:

$$\mathbf{G}^{<,Q,\sigma}_{RR} = \mathbf{G}^{\mathcal{A},Q,\sigma}_{RR} - \mathbf{G}^{\mathcal{R},Q,\sigma}_{RR} \tag{5.67}$$

which is equivalent to the following relation within LMTO-ASA approach,

$$\mathbf{g}^{\alpha,<,Q,\sigma}_{RR} = \mathbf{g}^{\alpha,\mathcal{A},Q,\sigma}_{RR} - \mathbf{g}^{\alpha,\mathcal{R},Q,\sigma}_{RR} \tag{5.68}$$

where we explicitly indicated the spin index σ. In the above relations, evaluating the left hand side needs Ω_{NVC} while evaluating the right hand side does not.

As an example, we consider a spin polarized magnetic structure. The central

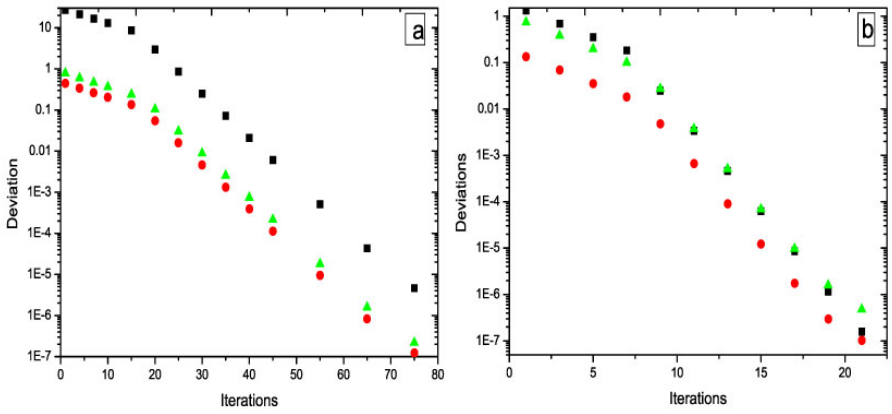

Figure 5.2: Results for spin-up (a) and spin-down (b) electron in the test structure at *equilibrium* plotted against the iterative steps in solving Eq.(5.54). Black Square: the deviation $\Delta\mathbf{\Omega}^{\sigma}_{NVC}$; The deviation $\Delta\mathbf{g}^{Q,\sigma}_{RR}$ (red circles for the Cu atom, green up-triangles for the Co atom). All quantities are iterated to zero as dictated by the fluctuation-dissipation theorem.

scattering region consists of 5 monolayers (ML) of Cu, connecting to 5 ML $Co_{0.8}Cr_{0.2}$ alloy, further connecting to 20 ML of Cu, then connecting to 2 ML of Co and finally connecting to 5 ML Cu. The system is connected to two semi-infinite Cu electrodes from left and right. In this rather complicated system, the disorder is due to the $Co_{0.8}Cr_{0.2}$ alloy where 20% Cr substitute the Co atoms. In the numerical calculation, all atomic spheres are put on the ideal Cu FCC lattice in (111) direction. We then carry out the NEGF-DFT-NVC two probe self-consistent calculation to obtain the electron potential inside the device scattering region at *equilibrium* (no bias voltage), then we focus on the Fermi energy $E = E_f$ of the Cu electrode using a 200×200 k-sampling grid for the BZ integration.

In the test, $\mathbf{g}^{\mathcal{A},Q,\sigma}_{RR} - \mathbf{g}^{\mathcal{R},Q,\sigma}_{RR}$ serves as the benchmark accurate result which is obtained by CPA from the converged electronic potentials. The NVC self-energy $\mathbf{\Omega}^{\sigma}_{NVC}$ is evaluated iteratively to obtain $\mathbf{g}^{<,Q,\sigma}_{RR}$. At equilibrium, the difference $\Delta\mathbf{g}^{Q,\sigma}_{RR} \equiv \sum_{LL'} |$ $\mathbf{g}^{<,Q,\sigma}_{RR} - (\mathbf{g}^{\mathcal{A},Q,\sigma}_{RR} - \mathbf{g}^{\mathcal{R},Q,\sigma}_{RR}) \,|_{L,L'}$ must approach zero due to the fluctuation-disspation relation Eq.(5.68). Furthermore, the deviation of NVC self-energy $\mathbf{\Omega}^{\sigma}_{NVC}$ between two iteration steps, namely $\Delta\mathbf{\Omega}^{\sigma}_{NVC} \equiv \sum_{R,LL'} | \, \mathbf{\Omega}^{\sigma,n}_{NVC,R} - \mathbf{\Omega}^{\sigma,n-1}_{NVC,R} \,|_{L,L'}$ where n is the number of iteration, must also approach zero when n is large. We monitor these quantities on each atom of the system. We present them for the Cu atom on the clean layer-4 of

the above test structure and the Co atom on the disordered layer-6 (counting from the left). The results are shown for spin-up channel in Fig.5.2(a) and spin-down channel in Fig.5.2(b). We can clearly see that both $\Delta\mathbf{g}^{Q,\sigma}_{RR}$ and $\Delta\mathbf{\Omega}^{\sigma}_{NVC}$ approach zero as the iteration proceeds, indicating that the fluctuation-dissipation theorem is numerically well satisfied after certain iteration steps. For this spin polarized test system, different spin converges in different ways.

We now investigate the diagonal elements of $\mathbf{g}^{<,Q,\sigma}_{RR}$ with and without the NVC self-energy $\mathbf{\Omega}^{\sigma}_{NVC}$ for the above Cu and Co atoms, to compare with the benchmark results given by $(\mathbf{g}^{\mathcal{A},Q,\sigma}_{RR} - \mathbf{g}^{\mathcal{R},Q,\sigma}_{RR})$. The comparison is given in Table-5.1. Without NVC, quantity $\mathbf{g}^{<,Q,\sigma}_{RR}$ shows a large difference from that with NVC. With NVC, $\mathbf{g}^{<,Q,\sigma}_{RR}$ is essentially the same as the benchmark result given by the fluctuation-dissipation theorem $\mathbf{g}^{\mathcal{A},Q,\sigma}_{RR} - \mathbf{g}^{\mathcal{R},Q,\sigma}_{RR}$ for all the elements. These tests provide confidence of the NEGF-DFT-NVC formalism and its numerical implementation.

Table 5.1: Some diagonal matrix elements of $\mathbf{g}^{<,Q,\sigma}_{RR}$ for spin-up of the Cu atom on layer-4, and the Co atom on layer-6 of the test structure. Method 1: calculated $\mathbf{g}^{<,Q,\sigma}_{RR}$ without NVC self-energy; Method 2: $\mathbf{g}^{<,Q,\sigma}_{RR}$ with NVC; and Method 3: using fluctuation-dissipation theorem $\mathbf{g}^{<,Q,\sigma}_{RR} = \mathbf{g}^{\mathcal{A},Q,\sigma}_{RR} - \mathbf{g}^{\mathcal{R},Q,\sigma}_{RR}$. Clearly, results of method-2 and method-3 are essentially identical, giving a strong test of the NEGF-DFT-NVC theory and implementation; while results without NVC have large errors.

Matrix Elements	Method 1	Method 2	Method 3
11(ss)/Cu	0.1795688	0.3424076	0.3424076
44(pp)/Cu	0.0663471	0.1251638	0.1251638
99(dd)/Cu	0.0026016	0.0048582	0.0048583
11(ss)/Co	0.0650346	0.2369839	0.2369839
44(pp)/Co	0.0207483	0.1144417	0.1144417
99(dd)/Co	0.0029164	0.0590015	0.0590015

5.6 Charge density

The full spin-dependent and component-dependent local charge density within the Q atomic sphere on site R, $\rho^{Q,\sigma}_{R}$, is given by

$$\rho^{Q,\sigma}_{R}(\mathbf{r}_R) = \rho^{Q,\sigma,val}_{R}(\mathbf{r}_R) + \rho^{Q,\sigma,core}_{R}(\mathbf{r}_R) \tag{5.69}$$

$$\rho_R^{Q,\sigma,val}(\mathbf{r}_R) = \sum_{R,LL'} \int_{E_B}^{E_F} \varphi_{R,L}^{Q,\sigma}(\mathbf{r}_R, E)\mathbf{n}_{R,LL'}^{Q,\sigma}(E)\varphi_{RL'}^{Q,\sigma}(\mathbf{r}_R, E) \tag{5.70}$$

From now on, the spin index is restored. The charge density is the sum of contributions from valence electrons and core electrons, denoted by $\rho_R^{Q,\sigma,val}(\mathbf{r}_R)$ and $\rho_R^{Q,\sigma,core}(\mathbf{r}_R)$, respectively. In Eq.(5.70), the energy integration in the first term is carried out between the bottom of the valence band E_B and the highest occupied band which is usually represented by the Fermi Level, $\varphi_{RL}^{Q,\sigma}(\mathbf{r}_R, E) = \varphi_{RL}^{Q,\sigma}(r_R, E)Y_L(\hat{\mathbf{r}}_R)$ is the partial wave function which is the regular solution of the Schördinger equation at energy E normalized to unity within the Q atomic sphere on site R; $\mathbf{n}_{R,LL'}^{Q,\sigma}(E)$ is the component projected local density of state (DOS) matrix which can be determined by the conditionally averaged site-diagonal physical Green's function at equilibrium,

$$\mathbf{n}_{R,LL}^{Q,\sigma}(E) = \frac{1}{2\pi}(\mathbf{G}_{RL,RL}^{Q,\mathcal{A},\sigma} - \mathbf{G}_{RL,RL}^{Q,\mathcal{R},\sigma}) = -\frac{1}{\pi} Im\mathbf{G}_{RL,RL}^{Q,\mathcal{R},\sigma}(E) \tag{5.71}$$

and by conditionally averaged site-diagonal physical NEGF at non-equilibrium,

$$\mathbf{n}_{R,LL}^{Q,\sigma}(E) = \frac{1}{2\pi}\mathbf{G}_{RL,RL}^{Q,<,\sigma} \tag{5.72}$$

The conditionally averaged site-diagonal physical Green's function and NEGF are given by Eq.(5.11) and Eq.(5.14).

For the two-probe device system under bias, the valence density can be rewritten

$$\begin{aligned}\rho_R^{Q,\sigma,val}(\mathbf{r}_R) = &\sum_{R,LL'} \int_{E_B}^{\mu_r} \varphi_{RL}^{Q,\sigma}(\mathbf{r}_R, E)[-\frac{1}{\pi} Im\overline{\mathbf{G}}_{RL,RL'}^{Q,\mathcal{R},\sigma}(E)]\varphi_{RL'}^{Q,\sigma}(\mathbf{r}_R, E) \\ &+ \sum_{R,LL'} \int_{\mu_r}^{\mu_l} \varphi_{RL}^{Q,\sigma}(\mathbf{r}_R, E)[\frac{1}{2\pi}\overline{\mathbf{G}}_{RL,RL'}^{Q,<,\sigma}]\varphi_{RL'}^{Q,\sigma}(\mathbf{r}_R, E)\end{aligned} \tag{5.73}$$

where $\mu_{r/l}$ are the chemical potentials of the left and right electrodes such that $eV_b = \mu_l - \mu_r$. Numerically, we separate the energy integration into two different parts. The first part satisfies the fluctuation-dissipation theorem, and the integration is done in the energy range of E_B to μ_r. The second part is for the non-equilibrium density

where the integration range is between μ_r and μ_l. The core electron density is actually spherically symmetric $\rho_R^{Q,\sigma,core}(\mathbf{r}_R) = \rho_R^{Q,\sigma,core}(r_R)$.

As an important approximation in the tight binding LMTO method, we perform a spherical average over the valence charge density to obtain spherical total density,

$$\rho_R^{Q,\sigma,val}(r_R) = \frac{1}{4\pi}\sum_{R,L}\{\int_{E_B}^{\mu_r} dE[\varphi_{RL}^{Q,\sigma}(r_R,E)]^2[-\frac{1}{\pi}Im\overline{\mathbf{G}}_{RL,RL}^{Q,\mathcal{R},\sigma}(E)] + \int_{\mu_r}^{\mu_l} dE[\varphi_{RL}^{Q,\sigma}(r_R,E)]^2[\frac{1}{2\pi}\overline{\mathbf{G}}_{RL,RL}^{Q,<,\sigma}]\} \tag{5.74}$$

which can be calculated using the radial function $\varphi_{R,L,\sigma}^{Q}(r_R,E)$ and diagonal elements of the conditionally averaged quantities, all the off-diagonal contributions are integrated to be zero. In practical calculations, the energy dependent radial function is replaced by a Taylor expansion at an energy $E_{\nu,RL\sigma}^{Q}$ which may be chosen as the center of the occupied part of the valence band,

$$\varphi_{RL}^{Q,\sigma}(r_R,E) = \varphi_{R,L}^{Q,\sigma}(r_R) + \dot{\varphi}_{RL}^{Q,\sigma}(r_R)(E - E_{\nu,RL}^{Q,\sigma}) \tag{5.75}$$

As a direct consequence, the valence charge density can be rewritten as:

$$\rho_R^{Q,\sigma,val}(r_R) = \frac{1}{4\pi}\sum_{R,L} m_{RL}^{Q,\sigma,0}[\varphi_{RL}^{Q,\sigma}(r_R)]^2 + 2m_{RL}^{Q,\sigma,1}\varphi_{RL}^{Q,\sigma}(r_R)\dot{\varphi}_{RL}^{Q,\sigma}(r_R) + m_{RL}^{Q,\sigma,2}[\dot{\varphi}_{RL}^{Q,\sigma}(r_R)]^2 \tag{5.76}$$

where the quantities $m_{RL}^{Q,\sigma,k}$ (k=0,1,2) denote the energy moments of the $QRL\sigma$-projected valence densities of state,

$$m_{RL}^{Q,\sigma,k} = \frac{1}{4\pi}\sum_{R,L}\{-\frac{1}{\pi}\int_{E_B}^{\mu_r} dE Im\overline{\mathbf{G}}_{RL,RL}^{Q,\mathcal{R},\sigma}(E)(E - E_{\nu,RL}^{Q,\sigma})^k + \frac{1}{2\pi}\int_{\mu_r}^{\mu_l} dE\overline{\mathbf{G}}_{RL,RL}^{Q,<,\sigma}(E - E_{\nu,RL}^{Q,\sigma})^k\} \tag{5.77}$$

The two energy integrations can be calculated along two separate energy contours. From the bottom of valence band E_B to μ_r, a complex energy contour can be used

because of the analytical property of the retarded Green's function in the upper complex energy plane. The choice of this complex energy contour can be many, such as a semi-circle, a rectangle etc.. The use of complex energy contour presents two most important advantages: a few energy points are enough to converge the integral and a small number of k-points is enough to converge the BZ integration on each of the complex energy point.

The integration from μ_r to μ_l, unfortunately, has to be done along the real energy axis because the integrand NEGF is not analytic on complex energy plane except on the real energy axis. As such, a much denser energy mesh as well as a much larger number of k-points are required to obtain accurate results and numerical convergence. The details depend on the research problem and will be presented in Chapters 6-8 for several applications.

5.7 Electronic potential

Within the tight binding LMTO-ASA method, the component dependent electron potential in Eq.(5.2) can be determined by[33]

$$V_R^{Q,\sigma}(r_R) = -\frac{2Z_R^Q}{r_R} + \int_{(R)} \frac{2\rho_R^Q(r_R)}{\mid \mathbf{r}_R - \mathbf{r'}_R \mid} d\mathbf{r}_R + V_{xc,R}^{Q,\sigma}(r_R) + V_{M,R} \tag{5.78}$$

where all terms are spherically symmetric. The first term is the nuclear potential in which Z_R^Q denotes the nuclear charge in the Q atom at site R; the second term is the Hartree potential where $\rho_R^Q = \rho_{R\uparrow}^Q + \rho_{R\downarrow}^Q$ is the total charge density; the third term is the exchange-correlation potential. These three terms represent the intra-atomic interaction. The inter-atomic interaction is included in the last term which is a constant for a particular site R. The component independent term $V_{M,R}$ is usually called the Madelung potential which depends on the averaged charge density and

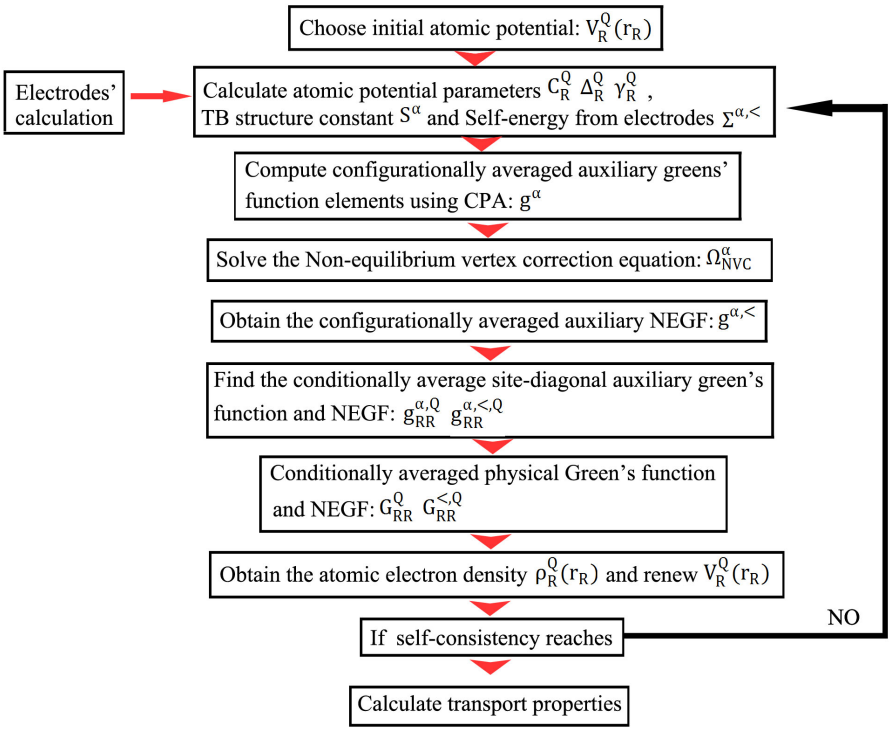

Figure 5.3: Flowchart of the NEGF-DFT-NVC self-consistent calculation.

multipole moments corresponding to all the other lattice sites,

$$V_{M,R} = \sum_{R'L'} M_{Rs,RL'} \overline{q}_{R'L'} \tag{5.79}$$

where $M_{Rs,RL'}$ is the Madelung constant and

$$\overline{q}_{RL} = \sum_{Q} C_R^Q q_{RL}^Q \tag{5.80}$$

$$q_{RL}^Q = Q_{RL}^Q - Z_R^Q \delta_{L,s} \tag{5.81}$$

$$Q_{RL}^Q = \sqrt{\frac{4\pi}{2l+1}} \int_{(R)} r^l Y_L(\hat{\mathbf{r}}_R) \rho_R^Q(\mathbf{r}_R) d\mathbf{r}_R \ . \tag{5.82}$$

The charge density $\rho_R^Q(\mathbf{r}_R)$ is the density before the spherical average and the real Harmonics can be used for Y_L. For $L = 0$, Q_{R0}^Q is the total electron charge of the Q atom on site R which is simply given by the moment $Q_{R0}^Q = \sum_{L\sigma} m_{RL\sigma}^{Q,0}$. For $L > 0$, q_{RL}^Q are the multipole moments. In our calculations, contributions of $L = 0, 1$ are

included in Eq.(5.79) and higher multipole moments are neglected because of the relatively much smaller effect for most cases.

Finally, Fig.5.3 pltos the implementation flowchart that illustrates our NEGF-DFT-NVC self-consistent calculation procedure.

5.8 Transmission coefficient

Using the NEGF-DFT-NVC formalism presented so far, the electronic structure of disordered systems at non-equilibrium can be calculated self-consistently. After the NEGF-DFT-NVC self-consistent calculation is converged, we calculate current-voltage (I-V) characteristics by the Landauer formula, Eq.(4.95). Within LMTO-ASA, in Section 4.7 we have proven that the Landauer formula takes the following form in terms of the auxiliary Green's functions, after disorder average:

$$\overline{I} = \frac{e}{h}\int_{\mu_r}^{\mu_l} Tr[\overline{\mathbf{\Gamma}_l^{\alpha}\mathbf{g}^{\alpha,\mathcal{R}}\mathbf{\Gamma}_r^{\alpha}\mathbf{g}^{\alpha,\mathcal{A}}}]dE \tag{5.83}$$

The integrand of Eq.(5.83) is the transmission coefficient $\overline{T}(E)$ where impurity average, once again, correlates the retarded and advanced auxiliary Green's functions $\mathbf{g}^{\mathcal{R}}$ and $\mathbf{g}^{\mathcal{A}}$. The calculation of $\overline{T}(E)$ involves a vertex correction even at equilibrium. In particular, $\overline{T}(E)$ is written into a specular contribution and a vertex correction contribution. The specular contribution is sometimes called "coherent" contribution, but not to be confused with quantum coherence. The vertex contribution describes inter-channel scattering events[37]. Therefore,

$$\overline{T}(E) = Tr\left[\mathbf{\Gamma}_l\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Gamma}_r^{\alpha}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\right] + Tr\left[\mathbf{\Gamma}_l^{\alpha}\overline{\mathbf{g}}^{\alpha,\mathcal{R}}\mathbf{\Omega}'_{VC}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}\right] \tag{5.84}$$

where the average auxiliary Green's function $\overline{\mathbf{g}}^{\alpha}$ is determined in Section 5.2 (see Eq.(5.19)); $\mathbf{\Omega}'_{VC}$ is obtained from the expression of $\mathbf{\Omega}_{NVC}$ by replacing $\mathbf{\Sigma}^{<,\alpha}$ with $\mathbf{\Gamma}_r^{\alpha}$ in Eq.(5.54). The equilibrium conductance for a spin channel is given by $G = (e^2/h)\overline{T}$.

Similar to the diagrammatical expansion of the average auxiliary NEGF in Fig.5.1, the transmission coefficient can also be expanded diagrammatically as in Fig.5.4, In

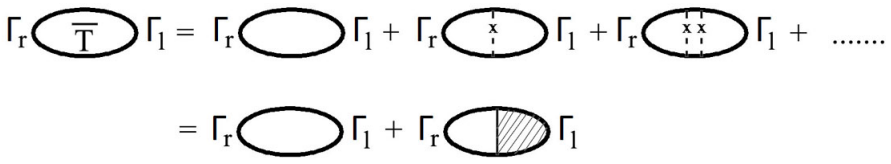

Figure 5.4: Diagrammatical representation of the transmission coefficient of Eq.(5.84)

the first line, the first term on the right hand corresponds to the specular part of Eq.(5.84), and all other terms describe the infinite multiple scattering expansion. The infinite series of the diagrams for describing multiple impurity scattering is summed up diagrammatically in the second line in term of the vertex correction.

Due to the translational invariance after configurational average, we can perform Fourier transformation to obtain transmission of the unit cell, $\overline{T}(E) = \frac{1}{N_{k_\parallel}} \sum_{k_\parallel} \overline{T}(E, k_\parallel)$, where $\overline{T}(E, k_\parallel)$ is the $k_\parallel$ resolved transmission coefficient which is given by,

$$\begin{aligned}\overline{T}(E, k_\parallel) = Tr\left[\mathbf{\Gamma}_l^\alpha(E, k_\parallel)\overline{\mathbf{g}}^{\alpha,\mathcal{R}}(E, k_\parallel)\mathbf{\Gamma}_r^\alpha(E, k_\parallel)\overline{\mathbf{g}}^{\alpha,\mathcal{A}}(E, k_\parallel)\right] \\ +Tr\left[\mathbf{\Gamma}_l^\alpha(E, k_\parallel)\overline{\mathbf{g}}^{\alpha,\mathcal{R}}(E, k_\parallel)\mathbf{\Omega}'_{VC}\overline{\mathbf{g}}^{\alpha,\mathcal{A}}(E, k_\parallel)\right] \qquad (5.85)\end{aligned}$$

In the above expression, $\mathbf{\Omega}'$ is independent of $k_\parallel$ since it is a site-diagonal matrix. It can be seen that the electron energy is conserved during the transport. The first term is contributed by the specular transport, in which the electron transports through the central device region by conserving the momentum represented by $k_\parallel$. In the second term which is the vertex correction, electrons may transport through without the restriction of momentum conservation, an electron enters the device from one electrode having momentum $k_\parallel$, and goes out of other electrode with a different momentum $k'_\parallel$. In other words, disorder scattering makes electrons traversing the device diffusively without conserving $k_\parallel$. Fig.5.5 shows a cartoon for illustrating the specular

transport and diffusive transport by disorder scattering. A sophisticated theoretical approach for analyzing realistic devices must include the effects of multiple scattering by randomly distributed impurities, as we have done here.

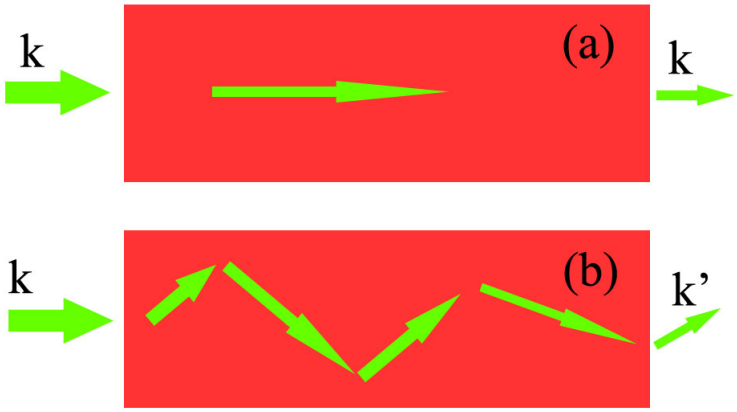

Figure 5.5: Schematic plot for illustrating specular transport (a) and diffusive transport due to random disorder scattering (b).

5.9 Summary

We have developed a non-equilibrium vertex correction theory and its associated software package for analyzing quantum transport properties of disordered nanoelectronic devices at non-equilibrium. The impurity averaging of the non-equilibrium density matrix is facilitated by NVC that is related to quantum statistical information of the device scattering region. Our NEGF-DFT-NVC theory has several desired features, including atomistic first principle, non-equilibrium, efficient configurational average and self-consistency. This allows us to analyze non-equilibrium quantum transport of realistic device structures including realistic atomic substitutional impurities. The next several Chapters will present some typical applications of our theory.

6

Effects of interface roughness in magnetic tunnel junctions

From this Chapter on, we shall apply the NEGF-DFT-NVC theory as implemented within the tight binding LMTO-ASA self-consistent DFT scheme, presented in Chapter 4-5, for several practical applications of quantum transport modeling. As a first application, in this Chapter we investigate disorder effects due to interface roughness in magnetic tunnel junctions (MTJ).

The simplest MTJ consists of two ferromagnetic metal electrodes sandwiching a very thin non-magnetic insulting barrier layer. The barrier is typically several nanometers thick, as shown schematically in Fig.6.1. It was experimentally found (for reviews, see Ref.[76, 77]) that the resistance of the MTJ is large when the magnetic moments of the two ferromagnetic layers are in parallel configuration (PC), and it is small when they are in anti-parallel configuration (APC). A MTJ is therefore a digital device registering two values of resistance depending on the magnetic properties. Information can therefore be stored in the MTJ and magnetic random access memory (MRAM) device can therefore be produced[78]. The digital nature of MTJ can also be applied to produce magnetic sensors, spin pumps and spin transistors[78]. In fact, MTJ is just one example of a larger problem of spintronics which exploit the spin degrees of freedom of the electron, in addition to charge, for electronic device applications[79]. At present, small scale commercial application of MRAM has already happened.

A most important device merit of MTJ is the tunnel magnetoresistance ratio

(TMR) which is defined as

$$TMR \equiv \frac{I_{PC} - I_{APC}}{I_{APC}} \tag{6.1}$$

where I_{PC} and I_{APC} are the total current in the parallel PC and APC. The tunneling current in a MTJ is spin polarized, contributed by both spin-up (majority) and -down (minority) electrons. Hence $I_{PC,APC} = I^{\uparrow}_{PC,APC} + I^{\downarrow}_{PC,APC}$ for total PC and APC currents, respectively.

The magnetoresistance effect was first experimentally demonstrated by Jullier in 1975 in a Fe/Ge/Co junction [80], producing a TMR of 14% at low temperatures. It took two decades for this effect to be measured at room temperature. In 1995, Moordera[81] and Miyazaki[82] independently observed significant TMR at the order of 20% in MTJs made of amorphous aluminium oxide AlO_x barriers. Since then, tremendous effort has been spent on improving TMR value. TMR value of 70% has been achieved in $FeCoB/AlO_x/FeCoB$ amorphous structure[83]. The situation changed quite dramatically with the demonstration of large TMR value in MgO based MTJ at room temperature. In 2004, giant room temperature TMR up to 200% have been reported experimentally in epitaxial $Fe/MgO/Fe$ structure[84] and textured FeCo/MgO/FeCo MTJ[85]. The highest TMR in MgO based MTJ reported to date is about 600% at room temperature and slightly above 1000% at low temperatures [86].

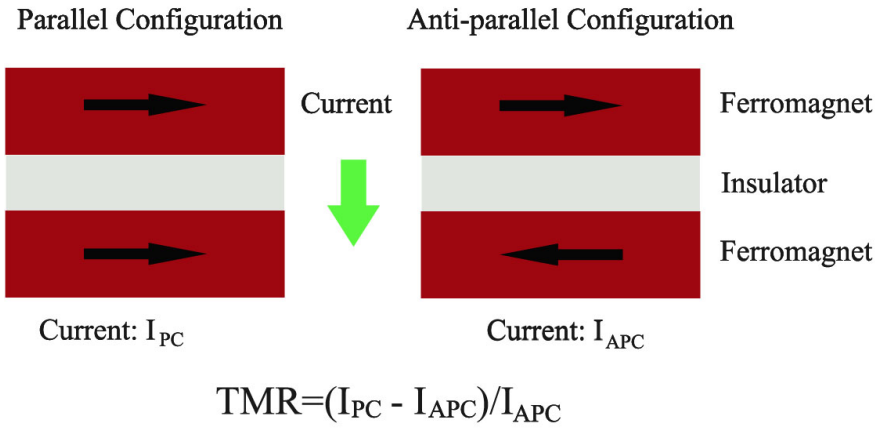

Figure 6.1: Schetch of magnetic configuration in magnetic tunnel junction

The value of TMR is determined by spin polarization of the electrons tunneling through the barrier from one ferromagnetic electrode to the other. It was thought that the spin polarization of the tunneling electron is an intrinsic property of the ferromagnetic electrode, and thus given by the spin polarization of the electronic density of state (DOS) at Fermi level of the magnetic metal[80]. This provided a reasonable explanation for TMR values in conventional MTJ based on amorphous material AlO_x.

However, as revealed by experiments[84, 87], the spin polarization is determined by the whole properties of the MTJ. It is not an intrinsic property of the ferromagnets metal alone, but it depends on the structural and electronic properties of the tunnel barrier material and the ferromagnetic/barrier interfaces[87]. The degree of spin polarization, its sign and the bias voltage dependence of the properties of a MTJ may all be changed by use of different barrier materials and magnetic electrodes, and even by using the same material but with different fabrication conditions. Although extensive experimental and theoretical research have been carried out for understanding spin dependent tunneling, the mechanisms governing the spin polarization in MTJ is still far from being completely understood.

In particular, the effects of interface roughness at the metal/barrier interface is expected to be important. In practical MTJ, some interface roughness is inevitable since the growth of one material over another material is usually accompanied by some degree of strain. This is especially true for metal/insulator interfaces. It is therefore an interesting and important problem to understand how interface roughness influence the TMR ratio.

In this Chapter, we shall investigate a tunnel junction made of Fe electrodes sandwiching a vacuum (VA) barrier in the form of Fe/VA/Fe heterostructure. Such a system has attracted many experimental studies already[88, 89, 90, 91]. For our problem of interface roughness effects, using vacuum as barrier excludes other disorder effects which might exist in real solid state insulators, hence any disorder scattering

of the spin polarized electrons is unambiguously due to the interface roughness. The NEGF-DFT-NVC formalism developed in Chapter 5 allows us to carry out disorder averaging of transport results which is critical for this problem. A short account of this work has been published before[2].

6.1 Calculation overview

The atomic structure of the Fe/VA/Fe MTJ is shown schematically in Fig.6.2, where two semi-infinite ferromagnetic Fe (100) electrodes are separated by a vacuum spacer. In our calculations, the device scattering region consists of the vacuum spacer plus some layers of Fe (100) on either side, as shown in Fig.6.2. The atomic sphere approximation (ASA, discussed in Section 3.7 of Chapter 3) is used with sites of a bcc lattice occupied by either Fe atoms or vacuum spheres. The atomic spheres of Fe and vacuum are kept the same. The thickness of vacuum spacer is thus measured by the number of mono-layers of vacuum spheres. The experimental lattice constant of bulk Fe, $a_{Fe} = 2.866$ angstrom, is used throughout our calculation.

The disordered roughness on the Fe/VA interfaces is represented by randomly replacing Fe atoms by vacuum spheres. Thus the roughness can be modeled by a binary alloy model Fe_xVA_{1-x} where x measures the degree of disorder. We only consider the roughness to be present in the surface Fe layer. For simplicity, any possible lattice relaxation at the surface is neglected.

Let us fix the left/right Fe/VA interfaces to have $x\%$ and $(1-x)\%$ Fe atoms respectively, and the rest are vacuum spheres. Namely we have Fe_xVa_{1-x} on left, and $Fe_{1-x}Va_x$ on right, such that the width of the vacuum spacer is a constant on average with respect to the variable x. The scattering region consists of ten perfect atomic layers of Fe on the left and right ending with the rough interface sandwiching four pure vacuum layers. By setting $x = 0$, the junction is restored to the ideal structure having perfect interface and 5 vacuum layers. The ideal structure as well as the

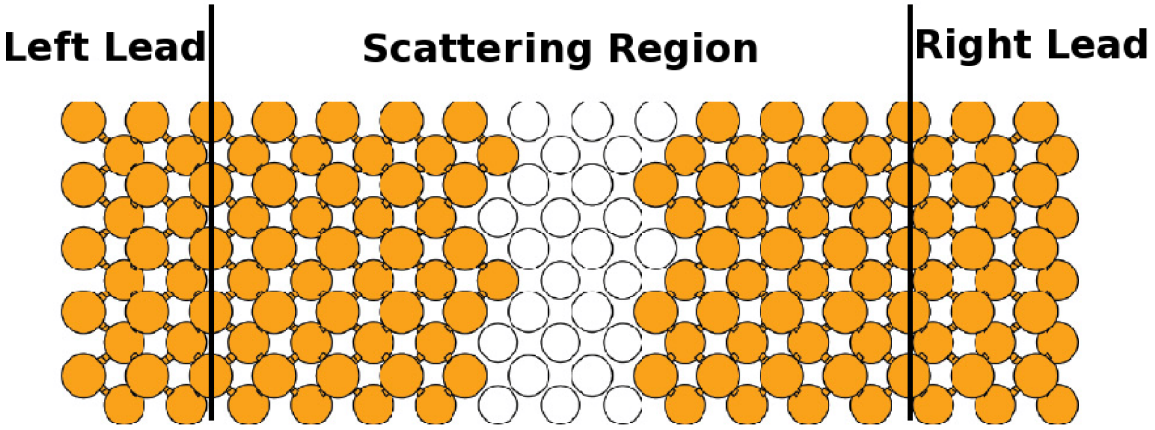

Figure 6.2: Schematic of atomic structure of the Fe/VA/Fe magnetic tunnel junction. The two Fe/VA interfaces have roughness disorder. Fe: yellow spheres; vacuum: white spheres.

configurationally averaged structures have translational symmetry in the transverse x, y directions (see discussions in Section 5.2), and a k-sampling technique is used to deal with the x,y periodicity.

We use s,p,d basis set to expand all physical quantiteis and the exchange-correlation potential is treated at the local spin density approximation (LSDA) level[50]. In the self-consistent electronic structure calculation, to obtain excellent convergence with respect to the 2 dimensional BZ integration, we use 100×100 $k_{\|}$ mesh for converging the equilibrium density matrix; and a 200×200 mesh for converging the non-equilibrium density matrix. For the I-V curve calculation, using Eq.(5.85), 300×300 BZ k-mesh is used for converging the transmission coefficient.

6.2 Perfect junctions

We first consider an ideal perfect Fe/VA/Fe junction in which electron tunnels by conserving the transverse momentum $k_{\|}$, namely the scattering is specular.

We calculated spin dependent conductance of the ideal junction as a function of vacuum spacer thickness at equilibrium. Fig.6.3 plots the spin-up conducance $G^{\uparrow}_{PC}$ in PC, spin-down conductance $G^{\downarrow}_{PC}$ in PC, and conductance in APC, G^{σ}_{APC}. In APC,

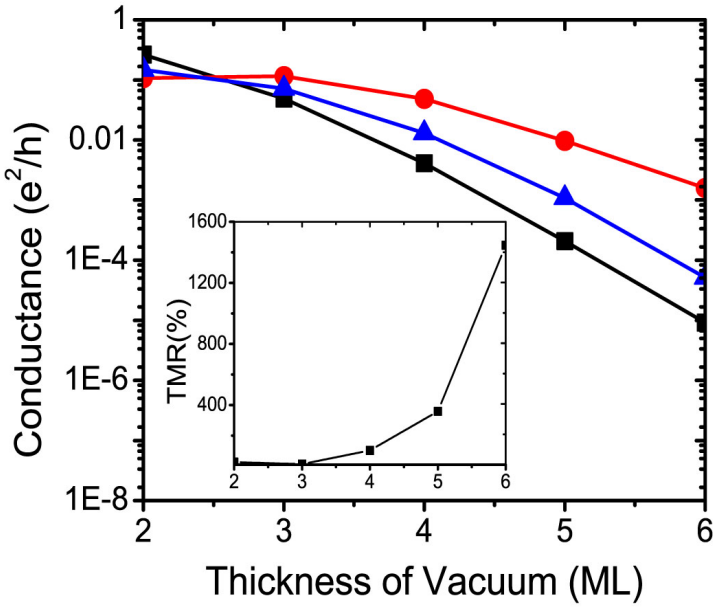

Figure 6.3: Conductance versus thickness of the vacuum spacer for spin-up and spin-down channels in PC and APC of ideal MTJs. Red circles: $G^{\uparrow}_{PC}$; blue up-triangles: $G^{\downarrow}_{PC}$; black squares: G^{σ}_{APC}. Note $G^{\uparrow}_{APC} = G^{\downarrow}_{APC}$ for the symmetric junctions we calculated. Inset: TMR versus thickness of vacuum spacer for the perfect junctions.

conductances of the two spin channels are calculated to be exactly the same, which is actually expected for our symmetical ideal junction. In all cases, an exponential dependence on the vacuum spacer thickness can be reached after about 3 ML. It is seen that spin-down channel dominates the conductance in PC, giving rise to the large and negative spin polarization of the tunneling electron. We found that $G^{\uparrow}_{PC}$ and G^{σ}_{APC} decay faster than $G^{\downarrow}_{PC}$. As a direct consequence, the TMR value calculated from Eq.(6.1) and shown in the inset of Fig.6.3, is enhanced significantly by orders of magnitude with increasing spacer thickness, reaches about 1500% at when the vacuum is 6 ML thick.

Fig.6.4 plots the $k_{\|}$ resolved transmission coefficient (transmission hot spots) for spin channels in PC and APC of the junctions with 3, 5, 6 ML vacuum spacers at the Fermi energy of the Fe electrodes. All the spin channels show a 4-fold symmetry which is due to the geometry of the junction. For the spin-up channel in PC and for all three junctions [Fig.6.4(b,c,i)], the total conductance is mainly contributed by

transmission hot spots surrounding the center of the 2D BZ (the Γ-point). In Fig.6.4, the transmission possesses a maximum at the center and decays rapidly away from the center. As the thickness of the vacuum spacer is increased, the transmission is reduced exponentially and becomes more and more concentrated around the center. We note the nearly circular symmetry of the transmission around the Γ point in Fig.6.4(b,e,i), which reflects the symmetry of the Δ_1 band of spin up electrons in the Fe electrode. The dominant role of the spin-down channel in PC is evident as illustrated in Fig.6.4(c,f,j). The spin-down channel transmission in PC, for all the thickness, is dominated by some hot spots having high transmission values as large as unity. We can see there are two different sets of hot spots, one set is closely surrounding the Γ-point, and the other is far from Γ. When thickness of the vacuum spacer is increased, the size of these hot spots is rapidly shrunk especially for those outer hot spots. At 6 ML vacuum, the hot spots surrounding the Γ point become dominant.

The hot spots of resonant peaks in $k_{\|}$ resolved transmission was previously investigated[92, 93]. The resonance peaks in the spin-down channel were attributed to the resonant states on the surface of the Fe electrodes[92, 93] and confirmed experimentally in Fe (001) surfaces[88]. In PC, both Fe surfaces in the ideal junction have surface states for spin-down electrons and these states form bonding and anti-bonding pairs across the junction[92], giving rise to a resonant tunneling behavior. The increase of vacuum thickness weakens the coupling between the surface resonances and shrink the hot-spots significantly. Finally, the transmission hot spots in APC are quite similar to those of the spin-down channel in PC, but with much smaller amplitudes.

So far, we have seen that the large TMR value in Fe/VA/Fe is a direct result of the important resonant tunneling transmission arising from the resonant surface states of spin-down electrons in PC. Clearly, the resonance will be reduced by applying a bias voltage which mis-aligns resonant peaks on one Fe surface relative to those on the other Fe surface. Fig.6.5(a) plots the spin polarized electron current a function

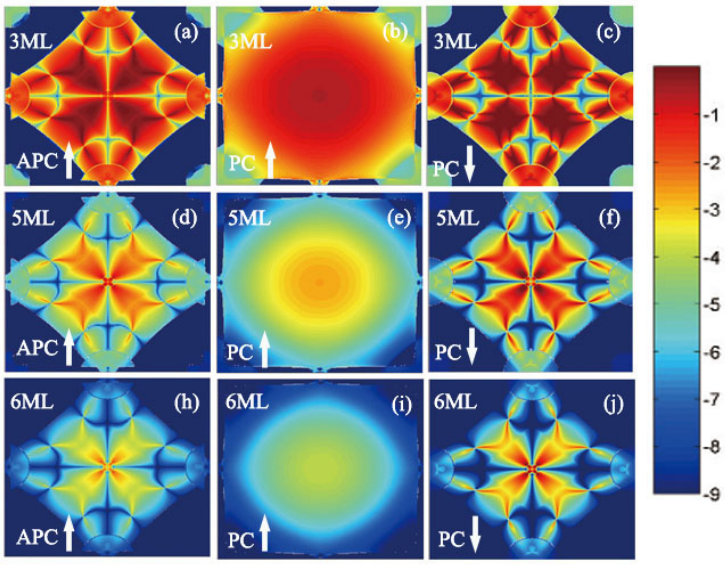

Figure 6.4: $k_{\|}$ resolved transmission coefficient $T = T(E_F, k_x, k_y)$ in 2D BZ for the spin-up and spin-down channels in PC and APC of perfect junctions having 3,5,6 ML vacuum spacer, shown in logarithmic scale . (a)(d)(h): APC spin-up channel. The APC spin-down channel is the same as the spin-up channel; (b)(e)(i): PC spin-up channel; (c)(f)(j):PC spin-down channel.

of bias for 5 ML vacuum spacer in PC and APC. For all spin channels, the calculated current is symmetric about zero bias due to the symmetric nature of the junctions. Applying a bias breaks the time reversal symmetry, thus transmission coefficients in spin-up and spin-down channels in APC are no longer the same. In both PC and APC, the current is dominated by the spin-down channel which has a nonlinear bias dependence. Currents of spin-up channels in PC and APC show almost a linear behavior.

We can calculate the TMR value from the spin polarized I-V curves in Fig6.5(a) for PC and APC. At zero bias when all currents vanish, we use the results of conductance at equilibrium to compute the TMR value. Fig6.5(b) shows the TMR versus bias from -0.7 V to 0.7 V. We found $TMR = 366\%$ at equilibrium, it is significantly quenched to about 100% by applying a small bias of 0.1 V and reaches 50% at 0.2 V after which TMR decreases at a slower rate. This TMR vs bias phenomena is referred to as the "zero bias anomaly" which is often found experimentally in many different tunneling junctions[94, 95, 96, 97]. This zero bias anomaly in our perfect junction can be easily

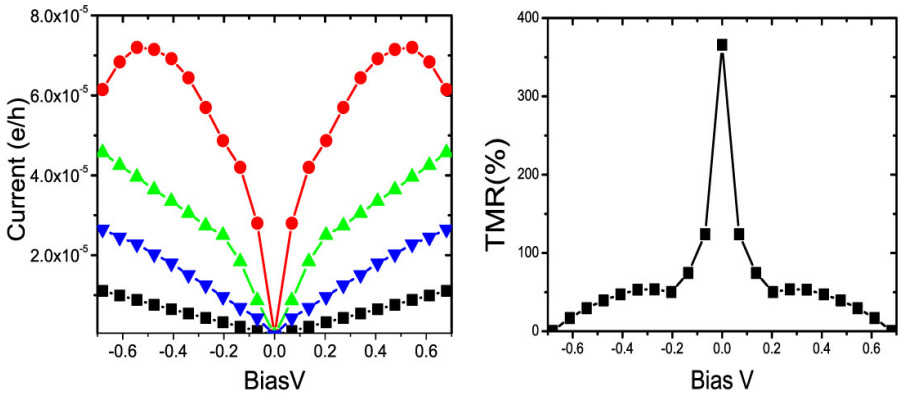

Figure 6.5: (a) Spin polarized current versus bias voltage for spin-up and spin-down channels in PC and APC of a junction having 5 ML vacuum. Black squares: $I^{\uparrow}_{PC}$; red circles: $I^{\downarrow}_{PC}$; up-triangles: $I^{\downarrow}_{APC}$; down-triangles: $I^{\uparrow}_{APC}$. (b) TMR value as a function of bias voltage.

understood: the bias shifts the resonance peaks on one Fe surface relative to the other Fe surface, thus drastically quenching the resonant tunneling of the spin-down channel in PC, thereby reducing the TMR value.

6.3 Rough interface junctions

Since perfect junction interface is very difficult to realize experimentally, it is important to investigate spin polarized transport in junctions with rough surface layers. We model the interface roughness by replacing some Fe atoms randomly with vacuum spheres on the interface Fe layer.

A rough interface may influence transport in three different ways. First, the roughness destroys the point group symmetry of the junction and therefore breaks the surface resonant states. As a result it is expected to decrease the large resonant transmission found in ideal junctions for spin-down channel of PC. Second, roughness may effectively reduce the thickness of the vacuum spacer thereby enhance the transmission which depends exponentially on the barrier thickness. This is likely to influence transmission in spin-up channel of PC. Third, roughness induces inter-channel scattering so that electrons may tunnel through the junction by changing its

momentum to couple to the simplest modes in the vacuum spacer.

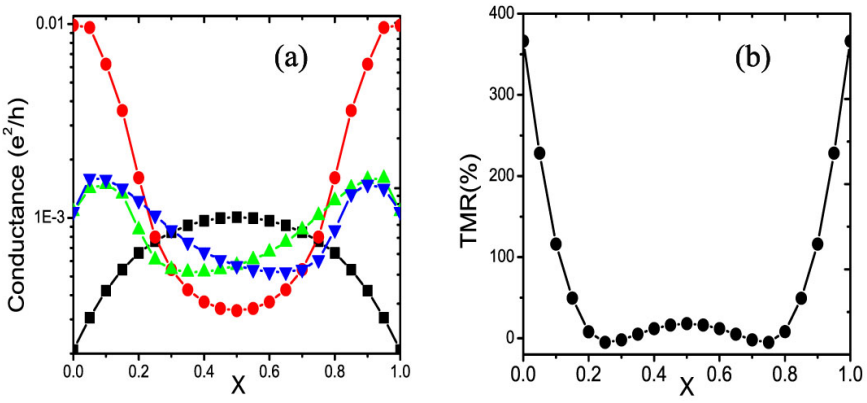

Figure 6.6: (a) total conductance $G^{\uparrow,\downarrow}$ versus disorder x at equilibrium. Red circles: $G^{\downarrow}_{PC}$; black squares: $G^{\uparrow}_{PC}$ in PC. Blue down-triangles: $G^{\downarrow}_{APC}$ in APC; green up-triangles: $G^{\uparrow}_{APC}$ in APC. (b) TMR value versus x.

Fig.6.6(a) shows the disorder dependence of configurationally averaged total conductance for both spin-up and -down channels in PC and APC of junctions having 5 ML vacuum spacer at equilibrium ($V_b = 0$). Since the left interface is chosen to be Fe_xVA_{1-x} while the right $Fe_{1-x}VA_x$, $G^{\uparrow,\downarrow}_{PC}$ are both symmetric about $x = 0.5$ in PC (black squares and red circles). For APC they are not symmetric but satisfy $G^{\uparrow}_{APC}(x) = G^{\downarrow}_{APC}(1-x)$, as expected. It is apparent that $G^{\downarrow}_{PC}$ is greatly reduced by the interface roughness, namely reduced by about two orders of magnitude when $x = 50\%$. The roughness has less effect for $G^{\uparrow}_{PC}$ and for APC. As x is increased from zero, $G^{\uparrow}_{PC}$ increases steadily and reaches its maximum at $x = 50\%$. Conductance of both spin channels in APC show a fast increase at very small x and then decreases.

The TMR value can be calculated using the equilibrium total conductance in Fig.6.6(a). Fig.6.6(b) plots TMR versus x showing a dramatic effect of disorder. In particular, as x is increased from zero, TMR drops to very small values, even to slightly negative values at about $x = 25\%$. These TMR features are consistent with previous super-cell calculations[98] for a thicker vacuum barrier.

To see how the interface roughness affect spin polarized tunneling, Fig.6.7 plots the specular part and vertex correction part of the conductance - the first term and

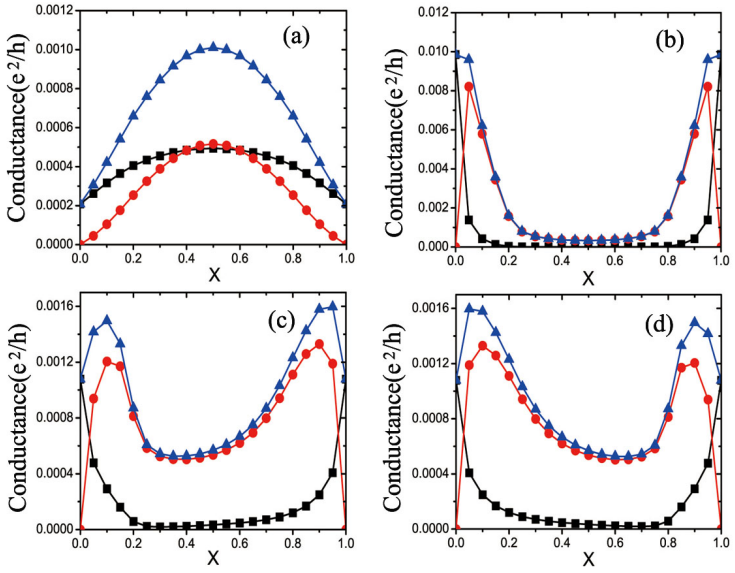

Figure 6.7: Conductance versus disorder x at equilibrium. Black squares: specular part, the first term of Eq.(5.84). Red circles: vertex correction part, the second term of Eq.(5.84). Blue up-triangles: total conductance. (a): $G_{\uparrow}^{PC}$. (b): $G_{\downarrow}^{PC}$. (c): $G_{\uparrow}^{APC}$. (d): $G_{\downarrow}^{APC}$.

second term of Eq.(5.84), as a function of x for spin-up and -down channels in PC and APC. Most evidently, the vertex correction part plays an important role in all spin channels. As shown in Fig.6.7(b,c,d), for the spin-down channel in PC and all spin channels in APC, the total conductance is dominated by contributions from the vertex correction; while the specular parts in these channels are dramatically quenched by even a very small amount of interface roughness. In other words, a very small amount of disorder on the Fe interface may turn the specular tunneling into diffusive tunneling. At very weak disorder, the surface resonant state of spin-down electrons quite possibly survives but it can be broadened by the disorder. These resonance states have high DOS around the Fermi energy, they may contribute to tunneling through inter-channel scattering. This is likely responsible for the sudden jump of the vertex correction from zero to a large value by a small amount of disorder x, as shown in Fig.6.7(b)(c)(d).

We now investigate non-equilibrium properties when $V_b \neq 0$ so that current flows through the disordered junctions. First, to show the importance of NVC, we have calculated I-V curves at $x = 0.05$ by including vertex correction only at the level of

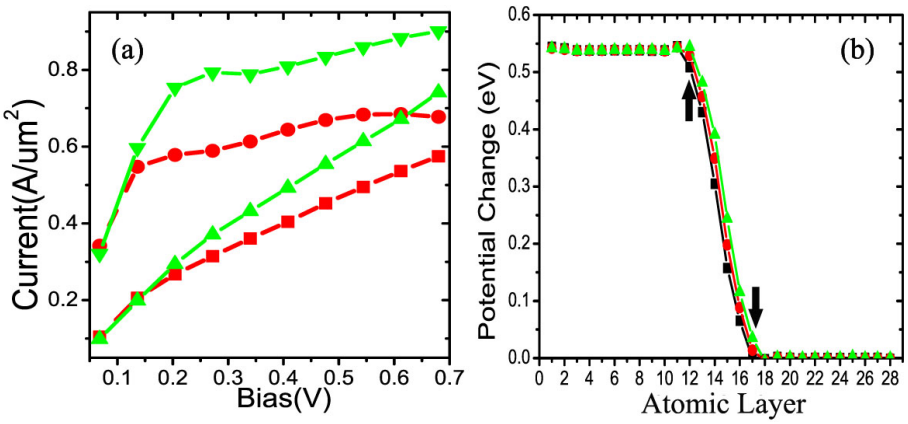

Figure 6.8: (a) Comparison of I-V curves with disorder $x = 0.05$. Solid lines (green): current for PC (up-triangles) and APC (down-triangles) without using NVC in the density matrix self-consistent iteration. Dashed lines (red): current for PC (circles) and APC (squares) using the full NVC formalism.(b) Bias induced electrostatic potential change versus atomic layer of the disordered junctions at $V_b = 0.544V$. Up-triangle: x=0.8; Red Circle: x=0.5; Black Square: x=0.2. the marked layers are the two interfacial disordered atomic layers.

transmission coefficient, *i.e.* without NVC in the NEGF-DFT self-consistent iterations of the density matrix: the solid lines (green) in Fig.6.8 plots this result. In comparison, the dashed lines (red) plot the full results where NVC is included. The substantial difference indicates that NVC is extremely important for obtaining correct results at non-equilibrium. Fig.6.8(b) plots the bias induced electrostatic potential change in each atomic layer of the disordered junction when applying $V_b = 0.544V$, using the NVC in the self-consistent NEGF-DFT iterations. The potential change arises from the bias induced charge redistribution in which NVC plays extremely important roles. As expected, deep inside the left and right electrodes, the potential changes by essentially two constants whose difference is exactly the value of the applied bias voltage. The potential drops in the middle region of the junction in nearly linear manner. The disorder effect on the potential profile, especially in the middle region, is apparent. It is worth to emphasize that the use of NVC in NEGF-DFT self-consistent calculation is critical to obtain the expected potential change as shown in Fig.6.8(b).

Fig.6.9 plots the total current and TMR versus bias for $x = 0.05, 0.3, 0.5$, obtained using the full NVC formalism of Chapter 5. The non-symmetric bias dependence in

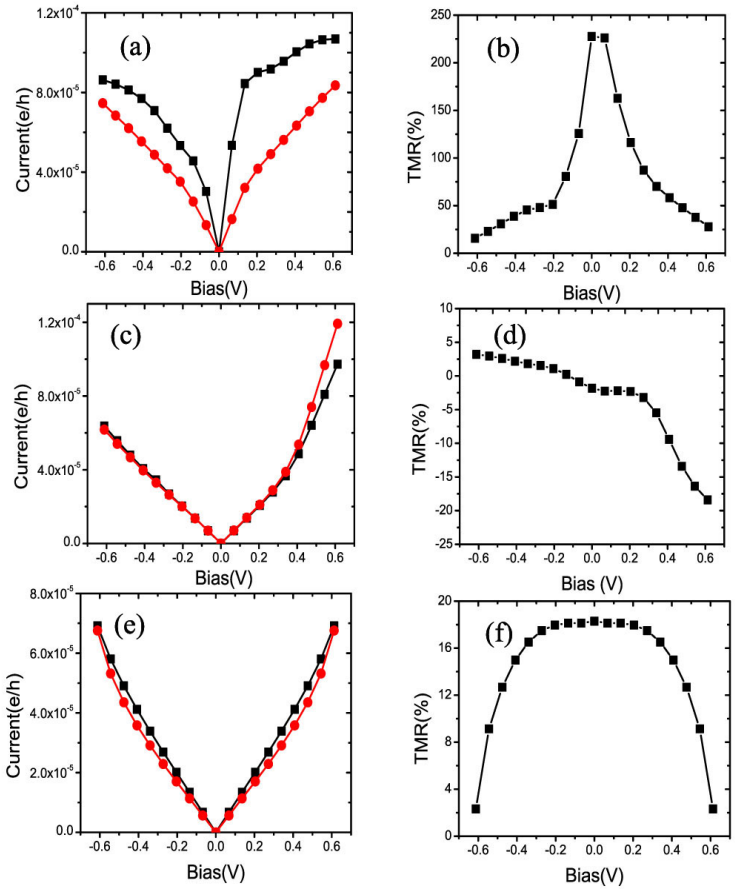

Figure 6.9: Total current for PC and APC and TMR versus bias voltage for different value of x. (a)(b) for x=0.05; (c)(d) for x=0.3, and (e)(f) for x=0.5. in (a)(c)(e), black squares: total current for PC I^{PC}; red circles: total current for APC I^{APC}.

(a,b,c,d) is the result of the left-right asymmetry of MTJs. The effect of disorder on the I-V characteristics is apparent in Fig.6.9(a,c,e). For small value of $x = 5\%$, the zero bias anomaly appears in Fig.6.9(b), and TMR reduces rapidly with increasing V_b. For larger x, the situation is completely changed and the zero bias anomaly is absent in Fig.6.9(d,f). In particular, for $x = 0.5$, TMR maintains almost a constant at small bias less than $0.3V$, then it reduces slowly with the increase of bias. For $x = 0.3$, when applying a positive bias, the TMR is negative and the absolute value of TMR is increased with increasing bias in the positive direction. On the other hand, the calculated TMR goes to positive value very slowly at negative bias. Our theoretical observations at larger x is consistent with the experimental measurement of spin dependent vacuum tunneling in Co(0001)[90] which showed the absence of

zero bias anomaly.

In the vacuum spacer based MTJ, mechanisms that may influence tunneling includes scattering by interface disorder, by disorder in the electrodes, and by exciting magnons that propagate away from the Fe interfaces. Zero bias anomaly is usually attributed to magnon scattering[96] even though there is so far no experimental confirmation. According to our calculation of ideal and rough junctions, zero bias anomaly can be attributed to weak disorder on the interface.

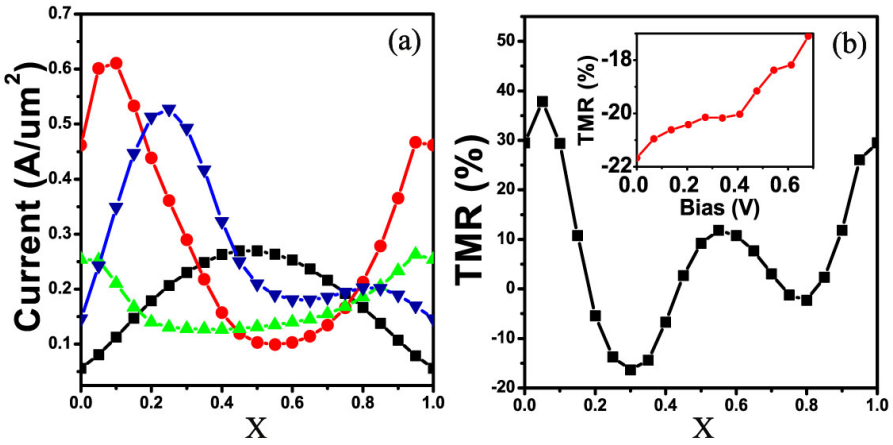

Figure 6.10: (a) Spin currents versus disorder x at bias $V_b = 0.544$V, for PC and APC. Red circles and black squares: spin currents for spin-up and -down in PC; green up-triangles and blue down-triangles: spin currents for spin-up and -down in APC. (b) TMR versus x at the same V_b. Inset of (b): TMR versus V_b for a device where left and right interfaces have different values of x, on the left interface $x = 0.3$, on the right $x = 0.05$.

Fig.6.10 plots spin currents and TMR versus disorder x at $V_b = 0.544$V. This is to be compared with Fig.6.6 where $V_b = 0$. A finite bias breaks left-right symmetry of the atomic structure and therefore, the spin currents do not have a symmetric behavior about $x = 0.5$ anymore. Both spin currents (Fig.6.10(a)) and TMR (Fig.6.10(b)) varies with disorder x in substantial ways. In particular, TMR rapidly dips to negative values when x is increased to about 20%. So far we have focused on devices where the left has a Fe_xVac_{1-x} interface while the right has $Fe_{1-x}Vac_x$. We have also applied the NEGF-DFT-NVC formalism to devices where the left and right interfaces are disordered totally differently. The inset of Fig.6.10(b) plots TMR for such a system where left interface has $x = 0.3$ while the right interface has $x = 0.05$. For this

system TMR is negative and its absolute value decreases as V_b is increased which is qualitatively similar to what discussed above.

6.4 Summary

In summary, we have applied the NEGF-DFT-NVC first principle quantum transport method to investigate effects of interface roughness on spin polarized electron tunneling in Fe/vacuum/Fe MTJ. It is found that roughness very efficiently turns the specular scattering at the interface to diffusive scattering. In particular, the TMR value is dramatically influenced by the interface disorder. Moreover, the interface disorder can alter the bias dependence of TMR in significant ways: the zero bias anomaly can be observed in junctions having weak or no disorder, but it disappears when the interface disorder is increased. The NVC is crucial for the self-consistent calculation of non-equilibrium electronic and transport through the disordered device.

7

Role of oxygen vacancy in Fe/MgO/Fe MTJ

As discussed in the last Chapter, one of the most important device merits of magnetic tunnel junctions (MTJ) is the tunnel magneto-resistance ratio (TMR). The higher the TMR, the more sensitive the device. For typical MTJs such as the amorphous AlO_x based systems[81, 82], the TMR ratio is $\sim$ 20%. About a decade ago, using density functional theory, Butler et al.[99] and Mathon and Umerski[100] predicted that MgO can be an excellent barrier material for MTJ. In particular, they found that Fe/MgO/Fe junctions can have huge TMR values, up to 10,000%. The physical reason is the very efficient coherent spin filtering effect in such a device (see below). MTJ is the fundamental device element for practical systems such as the magnetic random access memory and other spintronic devices[79, 78, 101]. More recently, experiments[85, 84] have achieved TMR ratio exceeding 200% at room temperature in epitaxial or textured Fe/MgO/Fe MTJs. The significant advance in this field was due to a deeper microscopic understanding of the physical mechanism that governs the spin dependent tunneling in the ideal structures of Fe/MgO/Fe MTJ. Since the experimental work of Parkin et al.[85] and Yuasa et al.[84], the experimental record of TMR value has been continuously increasing in MgO based MTJ. To date, the highest room temperature TMR value is $\sim$ 600% reported in textured CoFeB/MgO/CoFeB junctions grown by sputtering[86], and it is slightly above 1000% at low temperatures.

Nevertheless, it has been more than ten years since the original theoretical predictions[99, 100] of huge TMR $\sim 10,000\%$, and more than six years since experimentalists[85, 84] achieved TMR $\sim$ 200%: there is still a very large gap between predicted and mea-

sured TMR values. A very important problem is to identify and therefore rectify detrimental effects that prevent experimental TMR ratio from reaching much higher values.

For ideal clean Fe/MgO/Fe junctions, as explained clearly before[99], by symmetry the minority-spin d-states having transverse momentum $k_{\|} \neq (0,0)$ in Fe, cannot couple to the slowly decaying Δ_1 band of MgO at $k_{\|} = (0,0)$. These Fe states are therefore filtered out by MgO. Furthermore, the majority-spin channel in the left Fe cannot tunnel through in APC, because the right Fe is in the antiparallel state. The overall result is a very small APC current and a large spin polarized PC current, giving rise to the huge TMR in the ideal limit.

It is therefore generally believed that atomic defects in the experimental MTJ is the likely cause for not reaching the ideal theoretical TMR limit, because defect scattering can destroy the tunneling symmetry. One such effect[102, 103, 104, 105] is the possibility of extra oxygen atoms at the Fe/MgO interface forming a FeO layer. On the other hand, experiments on devices with clean, none oxidized Fe/MgO interfaces[106, 107, 108, 109] still report a TMR ratio far from the theoretical ideal limit. Calculations showed that small Fe/MgO interface structural randomness also drops TMR[103], but not enough to reach the current experimental range. More recently, experiments were carried out to investigate another kind of defects, oxygen vacancies (OV) inside the MgO barrier[110], and Ref.[111] provided direct experimental evidence of localized defect states inside the MgO energy gap which was attributed to the OV. In the experiment of Miao et al.[110], by introducing oxygen vacancies into the MgO barrier, the resistances of the Fe/MgO/Fe MTJ was dramatically enhance by about 50 times and the TMR value is significantly reduced. The oxygen vacancy within Fe/MgO/Fe MTJ is likely due to the compressive strain of Fe/MgO interface mismatch during the crystal growth[111]. Theoretical investigation of OV effects on TMR is rather limited. Ref.[112] reported a supercell density functional theory (DFT) calculation in which one oxygen atom in the MgO barrier was removed,

this defect reduced the TMR ratio from the ideal limit by a factor of roughly two.

In this Chapter, we provide a systematic analysis of the disorder effect due to oxygen vacancies (OV) in MgO barrier from atomistic first principle. Our results provide very good insight to spin dependent tunneling within the diffusive scattering regime. By placing the OVs at interfacial layers or interior layers in the MgO, a general trend is discovered: (i) merely a few percent of OV at the interfacial layer drop the TMR value from the theoretical limit 10000% to the experimental range; OV at the next layer to has similarly dramatic effect; (ii) effect of the interior OVs on the TMR is far less important, although they significantly increase the junction resistance; (iii) interfacial OV efficiently reduces the specular scattering in favor of diffusive scattering, causing the spin channels in Fe to scattering into the $\Delta 1$ band of MgO, thereby dramatically reducing the TMR; (iv), by filling the OVs with nitrogen atoms, the ideal TMR limit is partially recovered.

7.1 Calculation overview

The Fe/MgO/Fe MTJ is schematically shown in Fig.7.1, where two semi-infinite Fe electrodes are separated by 13 atomic layers of MgO barrier in the (100) direction. We adopt the same atomic structure as the Ref.[99] for the Fe/MgO/Fe junction, in which the interfacial Fe atoms sit on the top of the O atoms and the Fe-O distance is 2.16 angstrom. The experimental lattice constant $a = 2.866$ angstrom is used for the Fe lattice, and the MgO lattice constant is taken to be a factor of $\sqrt{2}$ larger than that of the Fe lattice, so that these two lattice matches perfectly at the interface. The nearly 3% lattice mismatching is neglected in our calculations.

In order to apply the NEGF-DFT-NVC formalism as implemented within the LMTO-ASA self-consistent scheme (see Chapters 3,5), we have carefully chosen the atomic spheres inside the device. For Fe, the atomic sphere has radius 1.411 angstrom which space fills the bcc lattice. Inside MgO, we take radius 1.354 angstrom for oxygen

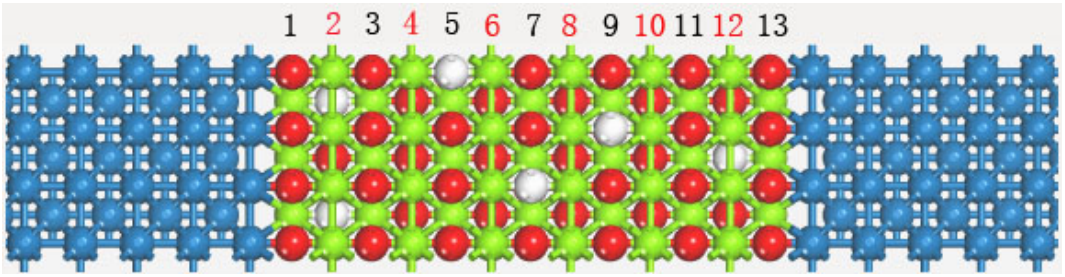

Figure 7.1: Atomic structure of the Fe/MgO/Fe MTJ with 13 ML MgO. Blue sphere: Fe; red: O; Green: Mg; White: Oxygen Vacancy. the junction is periodically extended in the transverse x,y directions. the numbers label the MgO layers from left to right.

atoms and 0.960 angstrom for Mg atoms. Vacuum sphere of radius 0.673 angstrom is added at the center of the cube with 4 O-atoms and 4-Mg atoms to fill the space. The calculated MgO band gap is 5.8eV within LSDA. At the Fe/MgO interface, two vacuum spheres are inserted exactly above the vacuum spheres inside the MgO with the same radius, and a vacuum sphere of radius 0.814 angstrom is added exactly above the Mg atom to fill the total volume of the transport junction. Positions of the interfacial vacuum spheres are arranged to minimize overlap. The quality of these choices of sphere size will be tested by reproducing results of previous first principle calculations for perfect ideal junctions. The oxygen vacancy is represented by replacing oxygen atoms with vacuum spheres (Va) of the same size, where we neglect the small structural distortion[113]. For the substitutional disorder of oxygen vacancies, we use the binary alloy model $O_{1-x}Va_x$ with the x as an input parameter.

We use the LSDA exchange-correlation potential of Ref.[50], and use s,p,d basis sets to expand physical quantities. For all the self-consistent calculations of the MTJ, a 100×100 k-mesh is used to sample the transverse two-dimensional BZ for converging the equilibrium density matrix (see Eq.(5.74)in Chapter 5). A 200×200 k-mesh is used to converge the non-equilibrium density matrix (see Eq.(5.74) in Chapter 5). To converge the transmission coefficients of all spin channels, a 400×400 k-mesh is used for most MTJs. The only exception is the case where the oxygen vacancies are on the MgO layer next to the interface: we use a much denser k-mesh of 1600×1600.

7.2 Coherent tunneling in ideal junction

We begin by investigating the equilibrium coherent spin dependent tunneling in perfect Fe/MgO/Fe junctions. Fig.7.2 plots the conductances as a function of MgO thickness L for up-spin channel $G^{\uparrow}_{PC}$ and down-spin channel $G^{\downarrow}_{PC}$ in parallel configuration (PC); and G^{σ}_{APC} in APC. For the symmetric junction of Fig.7.1, $G^{\uparrow}_{APC} = G^{\downarrow}_{APC}$.

The excellent spin filtering effect[99, 100] in the ideal Fe/MgO/Fe junction is reproduced. In particular, as barrier thickness L increases, the conductances of all spin channels exponentially decay, but $G^{\downarrow}_{PC}$ decays the fastest while $G^{\uparrow}_{PC}$ the slowest. As a result, the calculated TMR value shown in the inset of Fig.7.2 increases rapidly as a function of the MgO thickness, and reaches about 10000% when the thickness $L = 13$ monolayers (ML).

Since MgO has a large band gap, electron wave functions tunnelling into MgO exponentially decay but the decay rate takes a minimum value at $k_{\parallel} = (0, 0)$ with the Δ_1 symmetry. Thus, in perfect junctions where electron tunnels and maintains transverse momentum $k_{\parallel}$ conservation, those with $k_{\parallel} = (0, 0)$ and Δ_1 symmetry will contribute dominantly to the conductance $G^{\uparrow}_{PC}$; while electrons having $k_{\parallel} \neq (0, 0)$ contribute to the conductances $G^{\downarrow}_{PC}$ and G^{σ}_{APC} because the spin down state in Fe electrodes is empty at $k_{\parallel} = 0$ on the Fermi level. Therefore, we can observe that $G^{\uparrow}_{PC}$ decays slower than $G^{\downarrow}_{PC}$ and G^{σ}_{APC}.

These behaviors can be confirmed by examining the $k_{\parallel}$ resolved transmission coefficient (transmission hot spots) $T^{\sigma}(E_F, k_x, k_y)$ distributed in the 2D BZ, shown in Fig.7.3, on the Fermi level. Fig.7.3 shows the results for spin channels in PC and APC of the MTJ with 3 ML and 7 ML thick MgO barriers. For the spin-up channel in PC, $T^{\uparrow}_{PC}$ - as seen in Fig.7.3(a)(d), is dominated by the transmission around the center of BZ ($k_{\parallel} = (0, 0)$) with a circular symmetry, while the spin-down channel in PC is contributed by the hot spots at the BZ boundary. These hot spots in $T^{\downarrow}_{PC}$ is due to interface resonance states existing away from the $k_{\parallel} = (0, 0)$. This is also why we can

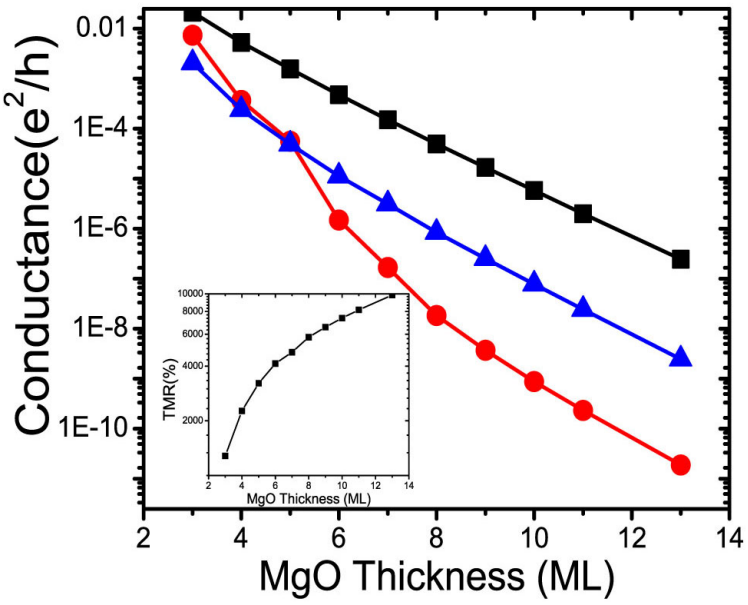

Figure 7.2: Conductance versus thickness of MgO barirer for spin up and spin down channels in PC and APC for perfect MTJ. Black square: $G^{\uparrow}_{PC}$; red circle: $G^{\downarrow}_{PC}$; blue up-triangle: G^{σ}_{APC}. In APC, $G^{\uparrow}_{APC} = G^{\downarrow}_{APC}$ for the symmetric perfect junctions. Inset: TMR versus thickness of MgO barrier.

see that for $L < 6$ ML, $G^{\downarrow}_{PC}$ is larger than G^{σ}_{APC}. As L increases, $T^{\uparrow}_{PC}$ increasingly becomes more concentrated around the BZ center and its amplitude decreases exponentially. The hot spot $T^{\uparrow}_{PC}$ is shrunk dramatically to quench $G^{\downarrow}_{PC}$ in Fig.7.2, and the APC channel becomes dominated by the transmissions surrounding the $k_{\|} = (0, 0)$ point. As shown in Fig.7.2, for all the thickness we calculated ($L = 3$ to 13 ML) the conductance in PC is dominated by the spin-up channel. All these observations agree very well with previous first principle calculations[99, 103, 100], they provide confirmation on the high quality of our chosen atomic spheres in the electrodes, in the MgO barrier and on the Fe/MgO interfaces of the MTJ.

7.3 Disorder effects of oxygen vacancy

From now on, we investigate effects of disordered oxygen vacancies in the MgO barrier of the MTJ. We will focus on the MTJ with 13 ML MgO which was the experimental system in Ref.[110]. The effects of OV are revealed by putting different concentrations

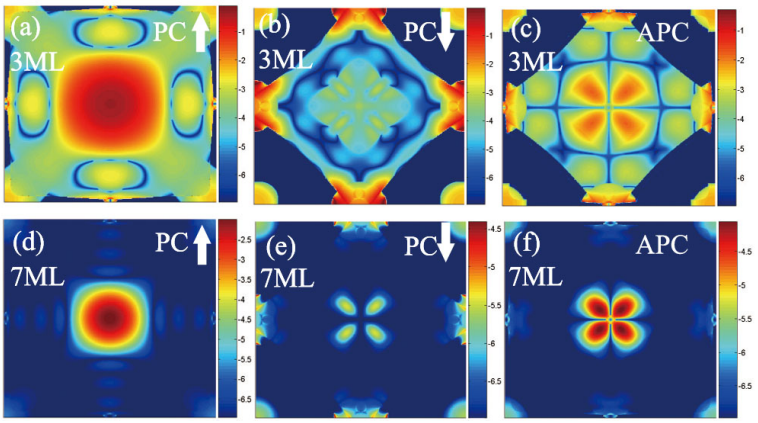

Figure 7.3: $k_{\|}$ resolved transmission coefficient $T = T(E_F, k_x, k_y)$ in 2D BZ for the spin-up and -down channels in PC and APC of perfect junctions with 3 and 7 MLs thick MgO barrier, shown on logarithmic scale . (a)(d): PC spin-up channel, (b)(e): PC spin-down channel; (c)(f): APC spin-up and -down channels. Due to the symmetry in perfect junctions, the spin-down channel transmission is the same as that of spin-up channel in APC.

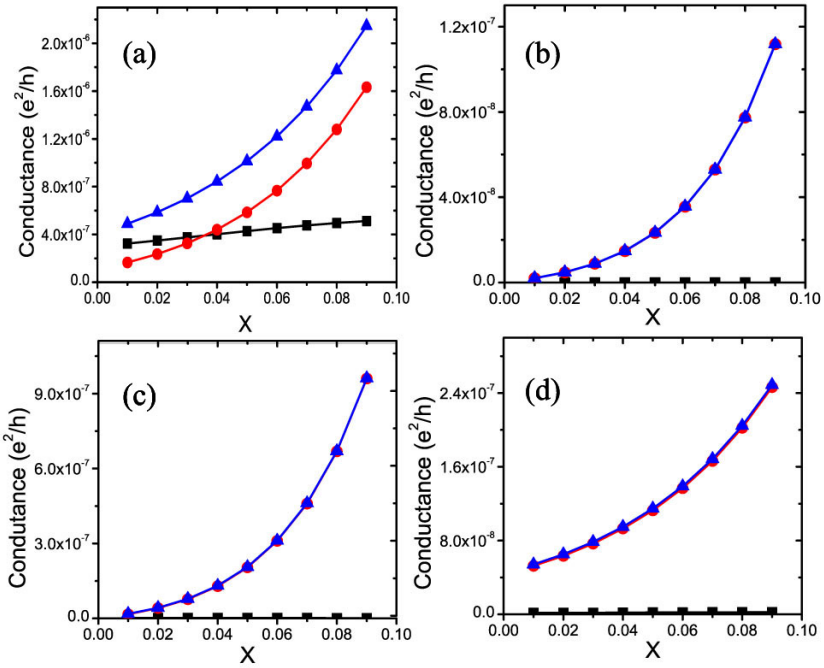

Figure 7.4: Conductance versus interfacial OV concentration x at equilibrium for spin-up (a)(c) and -down(b)(d) channels in PC (a)(b) and APC (c)(d). Layer-1 of the MgO is fixed with 3% OV; Layer-13 with x%. Blue up-triangle: total conductance; Red Circle: Vertex Correction; Black Square: Coherent part.

of OVs in different MgO layers.

Interfacial oxygen vacancy. First, we investigate the influence of interfacial OVs located on layer-1 and -13 of the MgO. We do not consider trapping Fe atoms into OV sites on the surface of MgO because of the blind adsorption effect discussed in Ref. [114]. Fixing 3% OV on the layer 1 of the MgO and putting x% on layer-13, leaving the rest MgO layers perfect, the results are shown in Fig.7.4 which plots the conductance versus x for PC and APC ranging from 1% to 9%. The most striking result is that the vertex correction part of the transmission, for example, the second term of Eq.(5.84), plays a dominant role in all spin channels for both PC and APC. As shown in Fig.7.4(b)(c)(d), the spin-down channel in PC and both spin channels in APC are almost entirely contributed by diffusive scattering. The OVs assist minority-spin d-states in Fe to traverse the MTJ by introducing inter-channel scattering which couples these states to the slowly decaying Δ_1 band of the MgO. As a result the coherent spin filtering effect[99] discussed above is drastically reduced. The conductance in APC increases significantly resulting to a drastic reduction of the TMR ratio.

Shown in Fig.7.5, for several OV distributions, TMR reduces dramatically from the ideal limit of $\sim 10,000$% to the experimental range of ~ 250%, when x is merely 4%. The blue stars in Fig.7.5 are the TMR values versus x for junctions having 7 ML MgO barrier and the same disorder distribution as the black circles. It is clearly seen that the results of 7 ML thick MgO barrier have the almost same magnitude as that of 13 ML MgO. This indicates that the TMR value has very weak dependence on the thickness of MgO for when junctions have interfacial disorder. This result is consistent with the experimental measurements[84]. To emphasize the dominating roll of interfacial OV, we have calculated a junction where layers-1 and -13 have 3% OV while layer-7 has a varying x%, the results are shown in Fig.7.7. It is clear that all spin channels are decreased at the same rate. Consequently, As the inset of Fig.7.7 shows, the TMR essentially stays at ~ 350% independent of the layer-7 x.

Next-neighbor oxygen vacancy. When the OVs are located at layer-2 and -12,

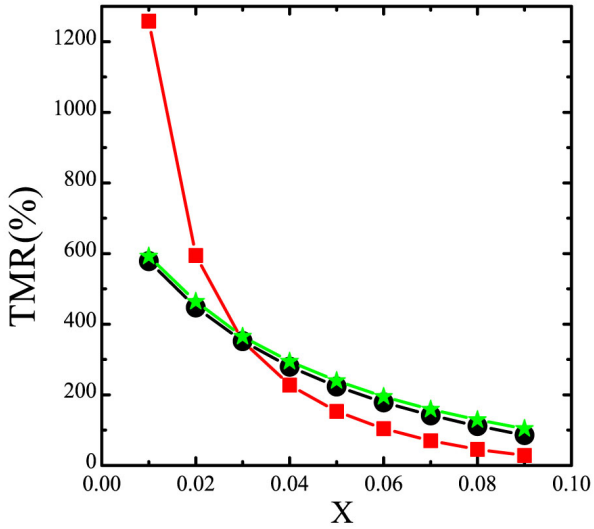

Figure 7.5: TMR versus disorder x for three types of MTJ. Red Square: for symmetrical junctions with x% OV on both MgO layers of 1 and 13; Black Circles: for asymmetric junctions with 3% OV fixed on layer layer-1 and x% on layer 13; Blue star: same disorder distribution as the black circles but for a junction with 7 ML thick MgO barrier.

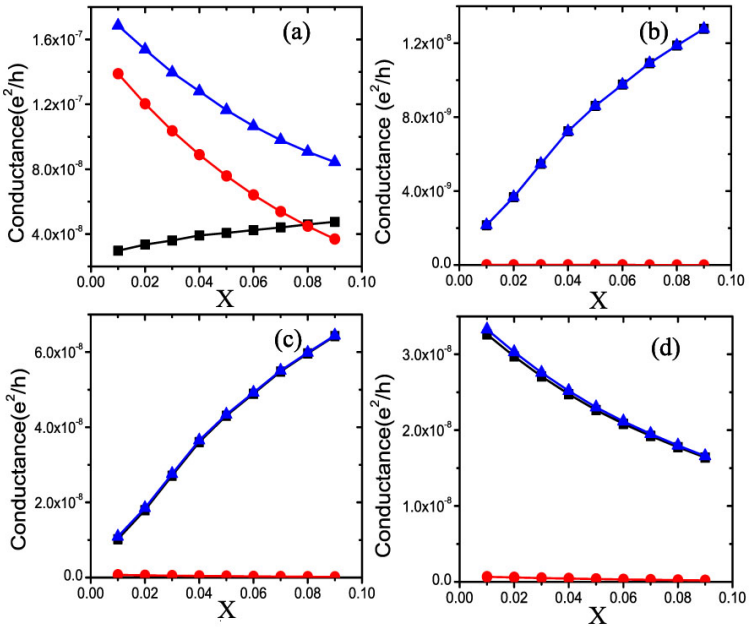

Figure 7.6: Conductance versus interfacial OV concentration x at equilibrium for spin-up (a)(c) and -down(b)(d) channels in PC (a)(b) and APC (c)(d). Layer-2 of the MgO is fixed with 3% OV; Layer-12 with x%. Blue up-triangle: total conductance; Red Circle: Vertex Correction; Black Square: Coherent part.

and all other MgO layers are clean, the disorder effects are found to be qualitatively the same as that of the interfacial OV. Again, the TMR is diminished very quickly from the theoretical ideal limit to $\sim 250\%$ when OV concentration x is less than a few percent (similar to Fig.7.5). With 3% OV on layer-2 and -12, adding further OVs in the middle layers of MgO does not significantly reduce TMR any further indicating, again, the importance of OVs near the Fe/MgO interface. We have further calculated junctions with both interfacial OV and next-neighbor OV, the TMR drops more quickly. The only main difference between the interfacial OV and the next-neighbor OV, is the behavior of the conductance of spin-up channel in PC and spin-down channel in APC. For the interfacial OV, Fig.7.4(a)(d) show an increasing conductance versus x; for the next-neighbor OV, it is a decreasing conductance. For the conductances of the other two spin channels behave the same as that of Fig.7.4(b)(c). The existence of an interfacial OV layer in effect reduces the width of the perfect MgO tunnel barrier, thus enhancing the tunneling probability. On the other hand, OVs (especially the interior OVs) provide scattering centers that reduce the tunneling probability.

Interior OV and interface roughness. By Leaving the interfacial and next-neighbor MgO layer clean, a few percent interior OVs still reduces the TMR but much less drastically. For example, when 5% OVs are put on the layer-3, we found that the total conductance in APC is almost the same as the perfect junction, while the spin up channel in PC is decreased by a factor of 2. Hence TMR is reduced from 10000% to 4067%. The interior OV has much less effect on TMR because the spin down electron in Fe has more difficulty to reach the interior OV to cause inter-channel scattering that couple it to the evanescent state at $k_{\|} = (0, 0)$ in the MgO barrier. In all the interior OV (on layers 3-11) configurations, the spin-up channel in PC is found to decrease for increasing OV concentrations.

For comparison, we have also investigated the surface roughness scattering by replacing randomly the Fe atoms at the Fe/MgO interface with $x\%$ of the vacuum

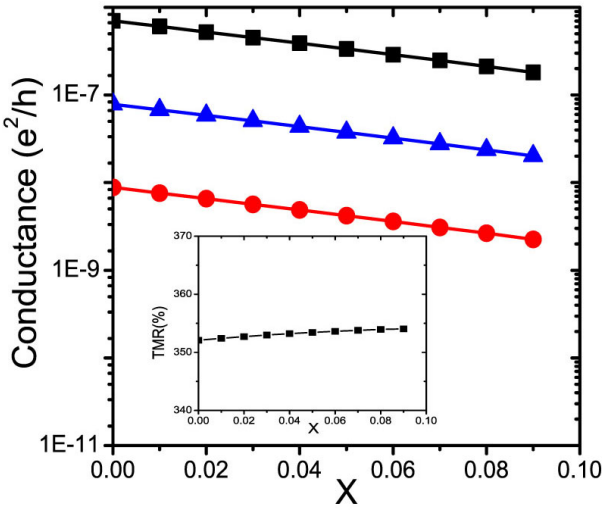

Figure 7.7: Conductance versus OV concentration x for spin-up and -down channels in PC and single spin channel in APC. 3% OV is fixed on interfacial MgO layer of 1 and 13, x% OV is put on the layer-7. Black Square: $G^{\uparrow}_{PC}$; Red Circle: $G^{\downarrow}_{PC}$; Blue up-triangle: G^{σ}_{APC}.

spheres with the same atomic size, so that the roughness is simulated using the binary alloy model$Fe_{1-x}O_x$. We consider a symmetrical case in which the roughness presents on both the left and right Fe/MgO interfaces, and find TMR of the MTJ is reduced steadily as x increases, reaching to 2300% at $x = 9\%$. This shows that the interface roughness is much less effective than the interfacial OV and next-neighbor OV in reducing the TMR magnitude.

Junction resistance. Experimentally, introducing OV to MgO layers can cause junction resistance to increase by 50 times[110]. We have calculated resistance $1/G_{P\uparrow}$ as a function of disordered layers of MgO, results shown in the inset of Fig.7.8. 10 junctions with $x = 5\%$ (circles) and $x = 3\%$ (squares) OVs existing on layers 3, or on 3-4, or 3-4-5, ..., or 3-4-...-10-11, are calculated. The resistances exhibit an exponentially fast increase: 5% OV causes 220-fold increase while 3% OV causes a 50-fold increase. This is consistent with experiment observations[110].

$k_{\|}$ resolved transmission coefficient. So far, a general trend emerges: small amount of OVs on or near the Fe/MgO interface very efficiently turn the specular

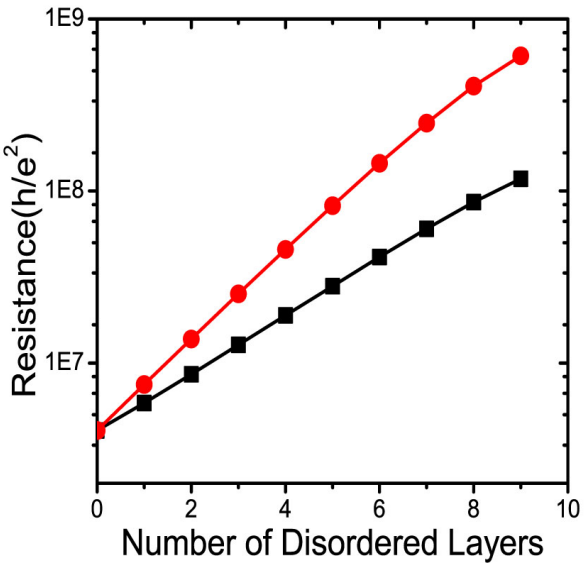

Figure 7.8: $\frac{1}{G_{P\uparrow}}$ versus number of disordered layers (3,3-4,3-4-5,.....,3-4-...-10-11) in logarithmic scale. Red Circle: for 5% OV; black square: for 3% OV.

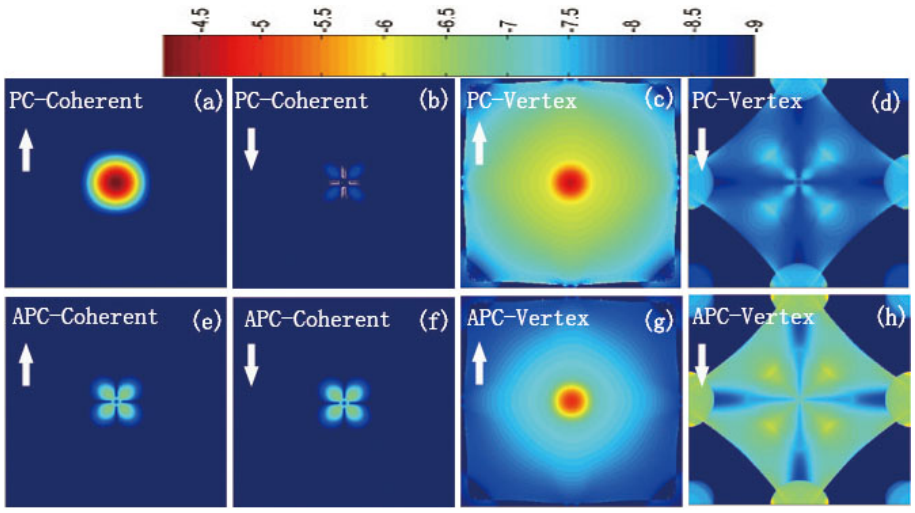

Figure 7.9: $k_{||}$ resolved transmission distribution in 2D BZ for coherent and vertex correction parts of spin ↑ and spin ↓ in PC and APC of the junction with 3% on both layers 1 and 13, in logarithmic scale. (a)-(d): PC; (e)-(f):APC. (a)(c)(e)(g): Spin ↑;(b)(d)(f)(h):Spin ↓. (a)(b)(e)(f): Coherent part;(c)(d)(g)(h):vertex correction part.

scattering into diffusive scattering, causing the spin-down channel in Fe to couple to the minimum decaying Δ_1 band of the MgO barrier, leading to a dramatic reduction of TMR value. This trend is vividly depicted in Fig7.9 which plots the $k_{\|}$ resolved coherent and vertex correction parts of the transmission coefficient, i.e., the 1st and 2nd terms of Eq.(5.84), for a symmetric junction with 3% OV at the layer-1 and -13. In the coherent part of all spin channels, electron tunnels through the MgO barrier by conserving $k_{\|}$. The coherent part is largely contributed by the transmission around $k_{\|} = (0, 0)$ for all channels calculated. The spin-up channel in PC (Fig7.9 (a)) is circularly distributed surrounding the BZ center, while PC spin-down and APC spin channels have a 4-fold symmetry ((Fig7.9 (b)(e)(f)).The coherent parts of spin up (e) and spin down (f) channels in APC are completely the same for the geometrical symmetry. The vertex correction part of all spin channels shows diffusive feature due to the inter-channel scattering at the interface. In the vertex correct part, a similar shape but different amplitude can be found between (c) and (g), and between (d) and (f). Most importantly, the APC spin-up vertex correction part have a circular symmetry around $k_{\|} = (0, 0)$, which matches the symmetry of the state at $k_{\|} = (0, 0)$ inside MgO barrier. Consequently, the APC spin up diffusive channel can easily pass through MgO barrier via this state, which increases APC transmission and drastically reduces the TMR magnitude. It is worth to mention that for APC, Fig7.9 (g) and (h) show completely different transmission distribution and amplitude in BZ, but the total transmission coefficient after integrating over the entire BZ gives exactly the same value: this is actually guaranteed by the left/right symmetry of the junction after configurational average.

TMR Dependence on the Bias. So far, we have investigated important OV effect on the equilibrium transport properties of the junction. We may want to ask what's the OV effect on the TMR dependence on bias voltage. Applying the NEGF-DFT NVC method, we calculated the non-equilibrium electronic structure for a 7-layer MgO junctions with and without OVs self-consistently, and then the TMR

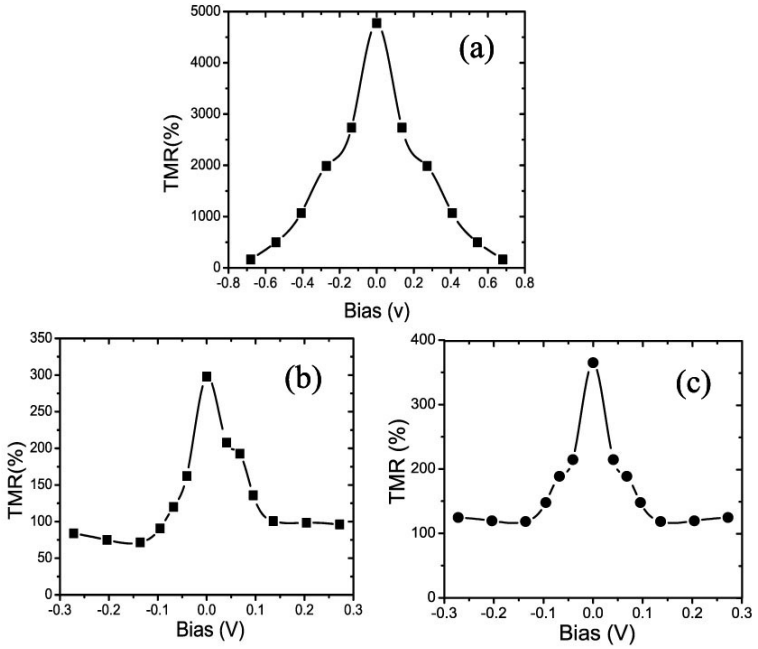

Figure 7.10: TMR versus bias voltage for a MTJ with 7 MLs thick MgO barrier. (a) Perfect Junction without OVs; (b) a unsymmetrical MRJ with 2% and 3% OVs on the interfacial layer-1 and -7 respectively; (c) symmetrical MTJ with 3% OVs on both layer-1 and -7.

versus bias is obtained by calculating the spin dependent currents, results are shown in Fig.7.10(a,b,c) for both clean, asymmetrical and symmetrical disordered interfacial layers. The asymmetrical MTJ shows a asymmetrical dependence on bias voltage. For all three cases the external bias reduces TMR ratio significantly at small bias voltage and the existence of OV does not qualitatively alter the general voltage dependence, even though the vertical scales of Fig.7.10(a,b,c) are very different. Similar voltage dependence of TMR has been reported experimentally[85, 84] and theoretically[103].

Nitrogen doping. Recently it has been reported that nitrogen can be doped into MgO[115, 116] as substitutional atoms to oxygen. Since OV causes a dramatic reduction of TMR, filling the OV with nitrogen may partially restore it toward the clean theoretical limit. To exploit this possibility, we have calculated a 13-layer Fe/MgO/Fe junction where the interfacial OVs are filled with nitrogen atoms. Fig.7.11 plots the TMR versus the nitrogen concentration: TMR remains above 8000% when 4% of

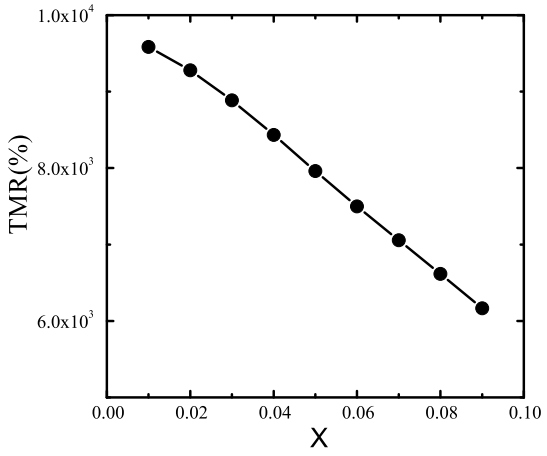

Figure 7.11: TMR versus disorder x for a junction where interfacial oxygen vacancy is filled by nitrogen atoms, namely x% nitrogen atoms replace the oxygen atoms on the 1st and 13th MgO layers.

interfacial oxygen atoms are randomly replaced by nitrogen. This means the diffusive scattering of spin-down electrons injected from Fe is much less effective by nitrogen impurity than by OV. Therefore, if viable experimental methods can be found to fill the almost unavoidable OVs near the Fe/MgO interface, it is possible to reach extremely high TMR ratio.

7.4 Summary

In summary, by using the NEGF-DFT-NVC first principle quantum transport method, we have investigated the significant effect of disordered oxygen vacancies in the MTJ of Fe/MgO/Fe. It was found that interchannel scattering by disordered oxygen vacancies on interface result in substantial reduction of the tunnel magnetoresistance ratio (TMR). Diffusive scattering by the oxygen vacancies inside the MgO barrier was found to exponentially enhance the the junction resistance. The TMR value of the junction with disordered interfacial oxygen vacancies can be rapidly reduced by applying the bias. The inclusion of disordered oxygen vacancies significantly improves the agreement between first principle calculation and experiments.

8

Surface roughness scattering in Copper interconnects

One of the key issues for integrated circuit technology is the increase of Cu interconnect resistivity with decreasing wire cross section[6], a phenomenon typically referred to as "size effect"[117]. The "size effect" becomes appreciable when the interconnect line-width approaches $\sim$ 100 nm, namely when it reaches 2-3 times the mean free path which is 39 nm for Cu at room temperature. The size effect becomes severe below 50 nm, giving rise to $\sim$ 100% increase in the resistivity[118, 119, 120]. This size effect in copper wires has been a great challenge for the continued down scaling of electronic devices because increased resistivity dramatically enhances heat dissipation and interconnect delay in the integrated circuits.

Experimentally, among the several electron scattering mechanisms that contribute to the resistivity of Cu interconnects[117, 118, 119, 121, 122, 123] including the scattering by phonons, random impurities, grain boundaries and surface roughness, the surface roughness scattering has been identified as a major source to the size effect when the line width is less than 50 nm. A 50% increase in Cu resistivity due to surface roughness scattering has been reported in a recent experimental study[119].

Theoretical investigations of bulk film resistivity has a very long history starting with the well known semiclassical model of surface roughness scattering by Fuchs[124] in the 30's and by Sondheimer[125] in the 50's. The Fuchs-Sondheimer (FS) model is still widely used by circuit engineers today. In the FS model, a phenomenological specularity parameter p is used to characterize electron scattering at the surface:

$p = 1$ means perfectly specular scattering while $p = 0$ means completely diffusive scattering. Different scattering mechanisms in metal films have also been described by empirical models[126, 127, 128, 129] and by advanced analytic models that take into account quantum effects prevalent in very thin films[130, 131, 132]. The long history and extensive investigation have provided our current understanding of thin film resistivity.

While the empirical and analytic models have provided useful knowledge on the influence of different scattering mechanisms in Cu interconnects and are appealing for their simplicity, there is a clear need that calls for more accurate quantitative methods to directly calculate the resistivity for realistic atomic configurations without employing any phenomenological parameter and without fitting to experimental data. Atomistic first principles approaches can be very useful in this regard. Recently, several *ab initio* studies [133, 134] of the resistance of Cu films and nanowires have been reported where a super-cell approach was employed on periodic atomic structures. In an earlier study[133], we have shown a 30% to 40% reduction in the conductance of thin Cu films due to surface roughness and the reduction was attributed to the destruction of isotropic Fermi surface sheets by atomic mounds on the surface. However, one limitation of the super-cell approach is that it can be applied only to calculate the resistance, but not the resistivity. Besides, within this approach the roughness on the surface cannot be completely random due to the periodic atomic arrangements.

In this chapter, using the NEGF-DFT-NVC quantum transport method, we investigate surface roughness scattering induced resistivity in Cu film from atomic first principles.

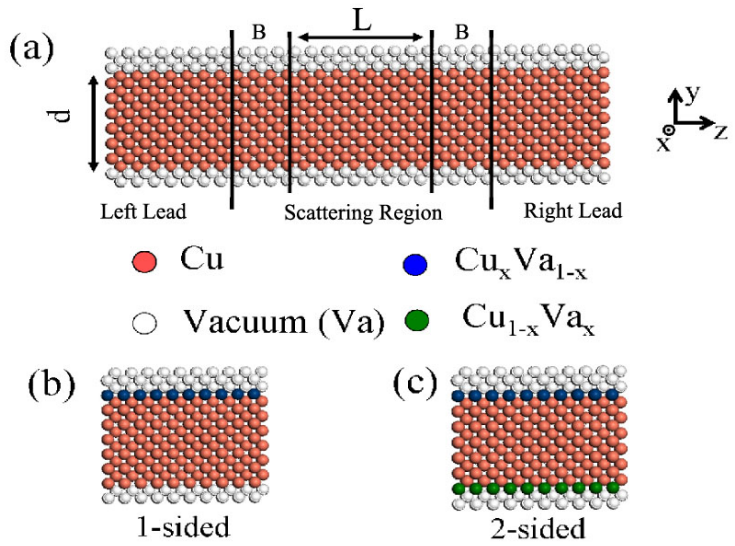

Figure 8.1: (a) Atomic structure of the Cu thin film. The two leads and the buffers (denoted by B) are perfect Cu films without any disorder. The buffer regions connect the leads to the scattering region. The atomic models used for surface roughness are shown in (b) for 1-sided roughness and (c) for 2-sided roughness.

8.1 Calculation overview

Fig.8.1(a) shows the atomic structure of the copper film. We treat the copper film as a two-probe device having two semi-infinite perfect films as electrodes separated by a scattering region of specific length (L) and thickness (d). The device is periodic in the width direction. The atomic structure is formed as such that the (010) direction is along the thickness (denoted by d) of the film while the (101) and $(10\bar{1})$ directions are along the length (denoted by L) and the width of the film respectively. The thickness and length of the copper film will be expressed as the number of monolayers (MLs) of Cu planes in each direction. We use the periodic boundary condition in the thickness direction by including enough vacuum space in the calculation box such that images of the films do not interact. The experimental lattice constant of fcc bulk Cu $a = 3.61$ angstrom is used in all our calculations.

For the surface roughness, we employ two models: 1-sided roughness (Fig.8.1(b)) and 2-sided roughness (Fig.8.1(c)). In both cases, the surface roughness is modeled by randomly replacing a fraction $(1 - x)$ of Cu atoms on the outermost layer by

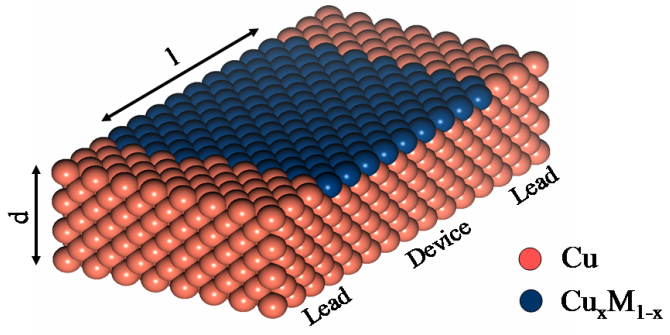

Figure 8.2: (a) Atomic structure of the Cu thin film treated as a two probe device of length L and thickness d. The barrier metal coating is shown for the 1-layer coating model. For the 4-layer coating model, three additional pure metal layers are added on top of the 1-layer model.

vacuum spheres of the same size. In order to solely focus on the surface roughness scattering, we consider single crystal Cu film without any impurity atom but with surface roughness. This way, the only contribution to the resistivity in our calculation is the surface roughness scattering. Later on (see below), to investigate the effects of coating material as shown in Fig.8.2, we fill the randomly distributed vacuum spheres on the outermost layers with some other metal atoms.

We will investigate transport properties in these copper films at equilibrium, because for good metal the nonequilibrium (finite bias) effects are weak at typical voltages applied to interconnects. Since we consider Cu films to have infinite extent in one transverse, say x-direction, and a finite width d in the other transverse y-direction, the 2D BZ is sampled by $(k_x, k_y) = (60, 1)$ k-mesh for each energy point. The energy integration for the equilibrium electron density matrix, Eq.(5.74), is performed with 28 energy points along a complex energy contour in the upper half plane[2]. We have checked that these computational parameters produced converged numerical values. In the transport calculations, both the specular and the vertex parts in Eq.(5.84) are evaluated with (100,1) k-mesh in the k-sampling. One of the main advantages of our NEGF-DFT-NVC formalism is that it can handle quite large systems: we have performed calculations of Cu films with up to around 1800 Cu atoms without any difficulty. The main results of this Chapter can also be found in Refs.[4, 5].

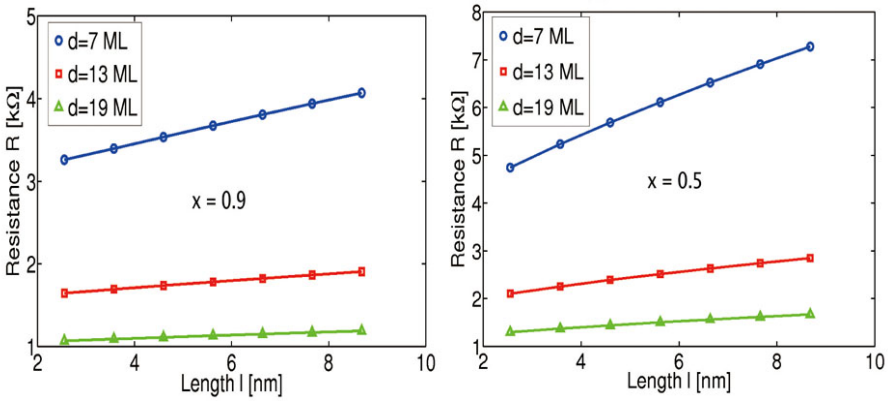

Figure 8.3: Copper film resistance R as a function of length L of the film for two cases. (a) $x = 0.9$, *i.e.* 10% disorder and 1-sided roughness. (b) $x = 0.5$, *50%* disorder and 2-sided roughness. the Resistance is calculated by $1/G$ (G is the equilibrium conductance).

8.2 Surface roughness scattering

We begin by calculating the resistance $R = 1/G$ of the Cu film for different thicknesses and lengths. The results are presented in Fig.8.3 for 1-sided roughness having 10% disorder ($x = 0.9$) and for 2-sided roughness with 50% disorder ($x = 0.5$). The former is very conductive with little surface roughness whereas the latter is low conducting with high roughness concentration. In both cases, resistance increases rather linearly with the length for all thickness of the film, showing an expected Ohmic behavior. A resistance of several thousand Ohms (Ω) suggests significant contributions from the surface roughness.

From the slope of resistance versus length curves in Fig. 8.3, we obtain the resistivity ρ, the results are presented in Fig. 8.4 as a function of disorder x for different thickness of the film. The resistivity shown here are solely due to surface roughness scattering. The surface is perfect if $x = 1$ (no disorder, *i.e.* roughness is zero); it is also perfect if $x = 0$, *i.e.* when the topmost layer atoms are all replaced by vacuum sites so that the next perfect layer of Cu becomes the top surface. Since the scattering is completely specular at a perfect surface, the resistivity caused by roughness scattering is zero at both $x = 0$ and $x = 1$, as shown in Fig.8.3. We observe that the resistivity is not quite symmetric around $x = 0.5$ for the 1-sided roughness model: this

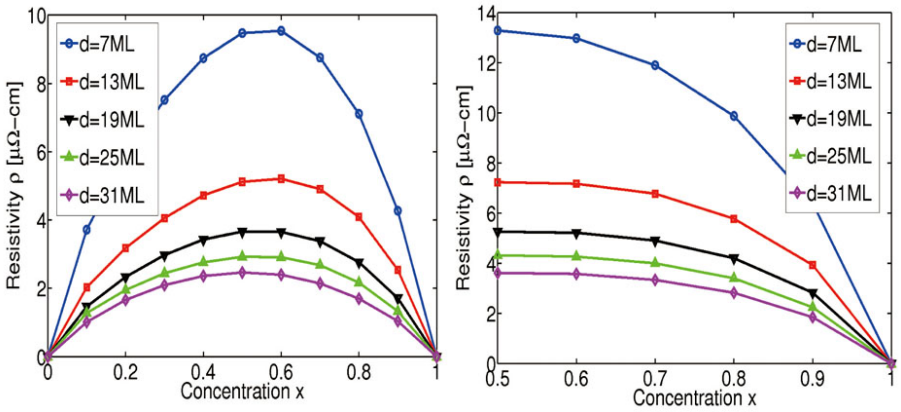

Figure 8.4: Surface roughness induced resistivity ρ versus x for different thickness d of the Cu films. (a): 1-sided roughness,(b): 2-side roughness. the resistivity is obtained by fitting the linear slope of resistance versus length

is expected since the atomic potentials of a Cu atom and a vacuum site are different. Interestingly, resistivity is maximum when the disorder is 40% (*i.e.* $x = 0.6$). On the other hand, for 2-sided roughness the resistivity is completely symmetric around $x = 0.5$ (the ρ values for $x < 0.5$ are not shown in Fig. 8.4b). We also notice that the resistivity is much higher at around $x = 0.5$ for $d = 7$ ML compared to the resistivity for the other thickness values, showing a substantial size effect when the film thickness is smaller.

Fig. 8.5 plots the increase in resistivity with decreasing film thickness, this is the size effect mentioned earlier. Our calculated values for resistivity ($\sim$ 2-14 $\mu\Omega$-cm) are quite substantial compared to the room temperature *bulk* Cu resistivity value of $\rho_b = 1.67$ $\mu\Omega$-cm obtained experimentally. This shows that surface roughness scattering can have a significant effect on the resistivity of very thin Cu films. When the surface roughness is small (*i.e.* $x \sim 1$), the resistivity becomes less dependent on the thickness of the film which is expected for specular scattering. On the other hand, for high level of surface roughness ($x \sim 0.5$) the resistivity shows strong dependence on thickness. For the 2-sided roughness the resistivity becomes almost twice as large as for the 1-sided roughness. It is not exactly twice as large because the roughness concentrations are not symmetric on both sides for the 2-sided roughness model (see Fig. 8.1c), and the films are so thin that the top and the bottom surfaces can have

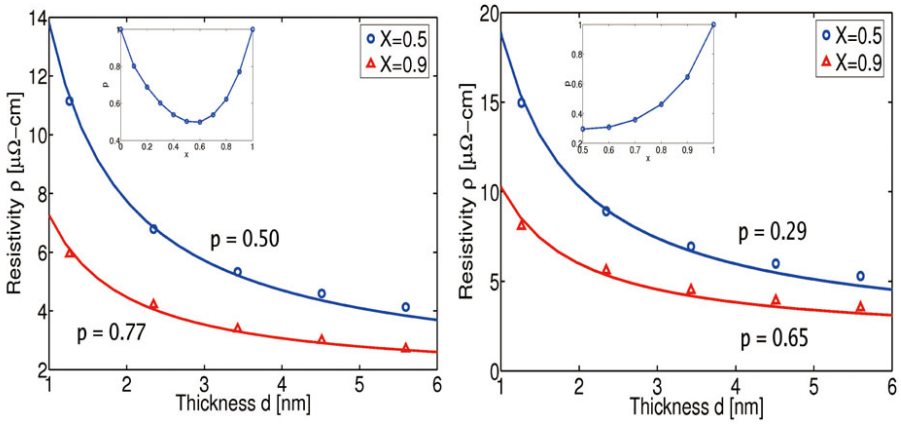

Figure 8.5: Resistivity ρ as a function of thickness (d) of the Cu film for different roughness concentrations (1-x). The ρ values (which equal our calculated resistivity due to surface roughness scattering plus the bulk resistivity value of 1.67 $\mu\Omega$-cm) are shown by the circles and the triangles. Solid lines are the corresponding fit of our data with the Fuchs-Sondheimer analytic model where p is the specularity parameter. The p values obtained from the fitting are shown as a function of x in the inset.

some correlations. We wish to point out that quantum oscillation in resistivity versus thickness d is difficult to observe in Cu films even though the thickness we have used are within the quantum regime. This is because a necessary condition to observe quantum oscillation is that the layer spacing should be much smaller than $\frac{\lambda_F}{2}$ where λ_F is the Fermi wavelength[130]. For Cu film, the layer spacing is 1.81 Å while $\frac{\lambda_F}{2}$ is 2.33 Å as obtained from the calculated Fermi energy of -6.9 eV for Cu. These two length scales are too close for appreciable quantum oscillations to be observable.

It is somewhat difficult to compare our calculated results quantitatively with experimental data available in the literature for two reasons. (i) Almost all the available experimental data were obtained for Cu film of thickness greater than 10 nm which is too large a system for atomistic *ab initio* calculations; (ii) due to the presence of all scattering mechanisms (phonon, impurity, surface roughness and grain boundary) in experimental systems, it is quite difficult to identify the contribution from each scattering process independently. Nevertheless, in a recent experiment[121], great care was taken to minimize the bulk impurity and grain boundary scattering effects by growing thin Cu film of pure and single crystalline nature, and the reported experimental value of resistivity is 8.35 $\mu\Omega$-cm at $d = 6.6$ nm. In comparison, our

(a) 1-sided Roughness

x	p	±Δp	RMSE
0.1	0.80	0.016	0.14
0.5	0.50	0.024	0.21
0.9	0.77	0.015	0.13
0.99	0.97	0.002	0.01

(b) 2-sided Roughness

x	p	±Δp	RMSE
0.5	0.29	0.049	0.43
0.7	0.36	0.056	0.49
0.9	0.65	0.043	0.37
0.99	0.94	0.004	0.03

Table 8.1: Values for the specular parameter p obtained from the Fuchs-Sondheimer fitting where ±Δp is the error bound for the p value and RMSE is the Root Mean Squared Error for the fit. Low values for both ±Δp and RMSE indicate the goodness of our fit.

calculated value at $d = 5.6$ nm is 5.30 $\mu\Omega$-cm (which includes a bulk resistivity value of 1.67 $\mu\Omega$-cm to take into account of the phonon scattering). The consistency is quite reasonable and the remaining discrepancy can be attributed to the nature of the roughness model used in our calculations.

As mentioned before, surface scattering is conventionally described by a semi-classical model developed by Fuchs[124] and Sondheimer[125] and later modified by Rossnagel and Kuan[117] to take into account surface roughness. This model proposes a relationship between the resistivity ρ and the thickness d of a thin film:

$$\frac{\rho}{\rho_b} = 1 + 0.375\frac{\lambda}{d}(1-p)S \tag{8.1}$$

where ρ_b is the bulk resistivity, λ is the room temperature electron mean free path, and p is called "specularity parameter" whose value ranges from 0 to 1 for completely diffusive to completely specular scattering, respectively. The surface roughness factor S is an empirical constant which quantifies the contribution due to *macroscopic* surface roughness. It can have a value greater than or equal to 1.0 with $S = 1$ representing a perfect surface. It is obvious that the two phenomenological parameters p and S are related since diffusive scattering is a direct consequence of the surface roughness. Therefore, it is almost impossible to independently determine p and S.

In order to estimate the value for p which is rather useful for experimental characterization of resistivity of rough films, we fit our calculated results with the Fuchs-Sondheimer (FS) equation (Eq. 8.1) by tuning the specularity parametr p, and we

employ the following values for the other parameters in Eq. 8.1: $\rho_b = 1.67$ $\mu\Omega$-cm, $\lambda = 39$ nm, $\rho = \rho_b + \rho_s$ (ρ_s is our calculated resistivity), and $S = 1$ (in this study, we choose to focus solely on the parameter p). We observe a good fit (see the solid lines in Fig. 8.5) of our data to the FS model by adjusting p. The resulting p values are shown in Table 8.1 (also, see the insets of Fig. 8.5). The fit is very satisfactory as suggested by the low values of $\pm\Delta$p and RMSE in Table 8.1. The minimum values of p for 1-sided and 2-sided roughness are 0.5 and 0.29 near $x = 0.5$, respectively. We note here that these p-values are obtained by directly fitting the data using Eq.(8.1), which assumes identical top and bottom surfaces. This assumption is likely reasonable for our 2-sided roughness layers which have comparable morphologies on top and bottom surfaces. Thus, we expect that the p-values in Table I(b) correspond to the specularity of the simulated surfaces. In contrast, the p values shown in table I(a) represent an average of the perfectly flat bottom surface and the rough top surface of the 1-sided layers. Assuming specular scattering ($p = 1$) for the bottom perfect surface, the p-value for the rough top-surface at x = 0.5 drops exactly to p = 0, corresponding to completely diffuse scattering. This is in agreement with various experimental studies which reported completely diffuse surface scattering[117, 121, 122]. Here, we would like to mention clearly that the good fit of our calculated *ab initio* data with the semi-classical FS model does not endorse or validate the FS formula for any quantum effects. In order to validate the FS model properly, one needs to perform a fitting with resistivity data for a much wider range of thickness values which includes both the quantum and semi-classical regimes. In addition we note that the F-S relation in Eq.(8.1) is an approximation which becomes inaccurate for $d/\lambda < 0.1$. Therefore, one could argue that Eq.(8.1) is not applicable to our computational approach since in the absence of true bulk scattering $\lambda = \infty$. However, a more sensible λ-value for our simulated thin layers may be the system length L. In that case, Eq.(8.1) holds true, since $\lambda \sim L < 10nm$ and thus $d/\lambda > 0.1$ for all simulated thicknesses $d > 1nm$. Despite these uncertainties in how to interpret the F-S model within our results, we believe that there is value in the presented fitting, as it provides an estimate of the

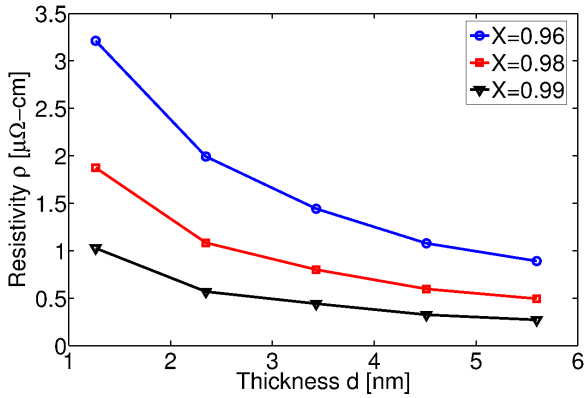

Figure 8.6: Resistivity (ρ) as a function of thickness (d) when the surface is almost perfect, i.e. the vacuum concentration $(1-x)$ is very low. The data presented here are for the 2-sided roughness model. As expected, resistivity shows less dependence on the thickness at low roughness concentrations, even though there is still substantial contribution from the surface roughness scattering.

specularity parameter p.

Finally, we plot resistivity as a function of thickness in Fig. 8.6 for *diluted* concentrations of roughness, *i.e.* for nearly perfect Cu surfaces. We expect that for perfect surface the resistivity will be independent of thickness and it will not deviate much from the bulk value. However, we observe in Fig. 8.6 that the resistivity is still increasing with decreasing thickness even when x is as high as 0.99 (*i.e.* 1% disorder). In other words, our calculations show that a thin Cu film which misses 1 atom for 100 (*i.e.* $x = 0.99$) is still rough enough to produce significant scattering at the surface. We also notice that at these diluted concentrations the resistivity is almost directly proportional to the disorder concentration.

8.3 Searching for coating material

Since Cu wires will continue to dominate the interconnect technology in any foreseeable future, it is of critical importance to find ways to minimize roughness scattering. Experimental growth of Cu films having a perfectly flat surface has not been possible so far, as even annealed single crystal Cu(001) layers still show a peak-to-valley roughness of more than 1 nm[135]. By coating barrier atoms on the Cu film,

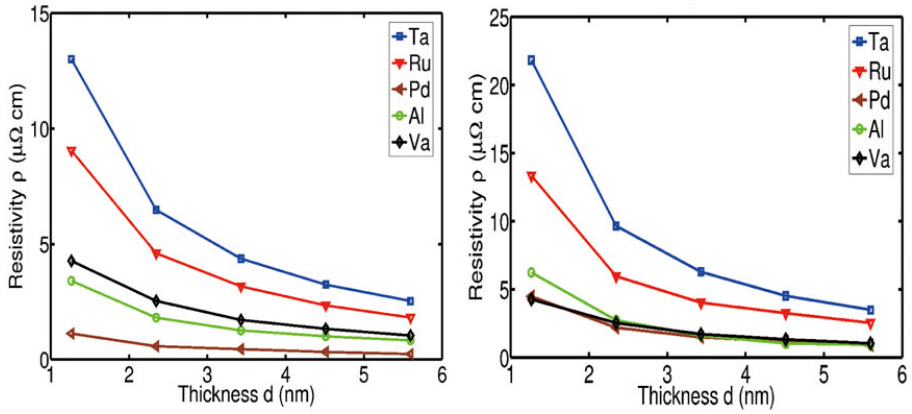

Figure 8.7: Cu thin film resistivity ρ as a function of film thickness d for different metal barriers with concentration $x = 0.9$ for the 1-layer (a) and the 4-layer (b) coating model. The resistivity for bare rough Cu films is denoted as 'Va'. The maximum thickness value of $d = 5.59$ nm corresponds to 31 MLs of Cu film.

the geometrical roughness may be filled by barrier atoms resulting in a smoother *geometry* thus possibly less diffuse scattering[117]. Barrier metals being examined experimentally[117, 136, 137, 138, 139] include Ti, Ta, Ru, Al and Pd: results indicate that some barrier metals actually increase resistivity compared with bare Cu films while others do reduce it, and there has been little theoretical understanding of the trends. In this Section we investigate effects of coating material from atomic first principles. The results can be found in Ref.[5].

Fig. 8.7 plots the calculated Cu resistivity ρ as a function of film thickness d for four different barrier metals Ta, Ru, Pd, and Al, as well as rough Cu films without a barrier coating denoted by 'Va'. The results for Ti coating are not presented to make the plot less crowded, since they are very similar to that of Ta with little difference in the resistivity values. It is clear that barrier metals make a significant difference in the resistivity. For Ta and Ru coating, the size effect is very pronounced and the resistivity is higher than bare Cu films. On the other hand, the resistivity with Al and Pd barriers are lower than the bare Cu films (see Fig. 8.7). These results qualitatively agree with experimental observations where the Cu sheet resistance was reported to increase by up to 15% using Ta barrier coating[136], and was always reduced with Al barrier coating[117]. A comparison between the two coating models (1-layer or

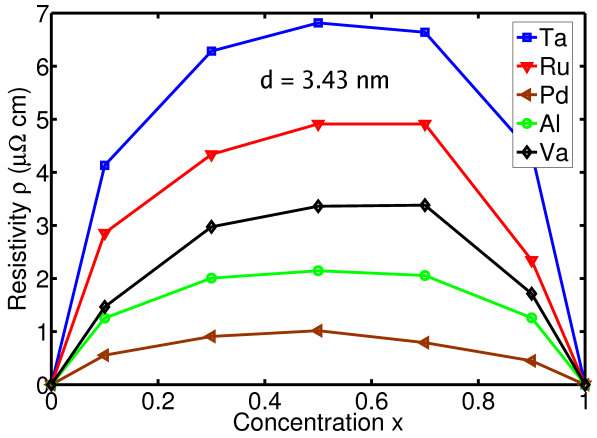

Figure 8.8: Resistivity (ρ) of Cu thin film at thickness $d = 3.43$ nm and 1-layer coating model versus disorder parameter x.

4-layer barrier) does not show a significant qualitative difference, suggesting that the most important contribution to resistivity comes from the Cu-barrier metal interface.

For all four barrier metals, the 4-layer coating model has higher resistivity which can be attributed to the increased mismatch of the pure Cu leads with the thicker barrier layers in the scattering region, causing electron scattering into the barrier layer. Since there is no qualitative difference, in the following we will focus on the 1-layer coating model.

Fig. 8.8 is a plot of the resistivity versus the disorder concentration parameter $(1-x)$ for a film of thickness $d = 3.43$ nm. The resistivity is not completely symmetric around $x = 0.5$ which is expected because the atomic potential of the Cu atom and the barrier atom are not the same. The maximum resistivity value is observed in the range between $x = 0.5$ to 0.7 for all barrier materials. The resistivity is zero at both limits $x = 0$ and $x = 1$ where the Cu surface is perfect such that scattering is completely specular. Very importantly, for any x value the resistivity with Al and Pd barrier coating is always lower than the bare Cu film; but it becomes higher with Ta and Ru barriers. These results suggest that Al and Pd barrier layers should be effective in suppressing the size effect in Cu films.

The effect of the barrier coating can be due to several factors including localized

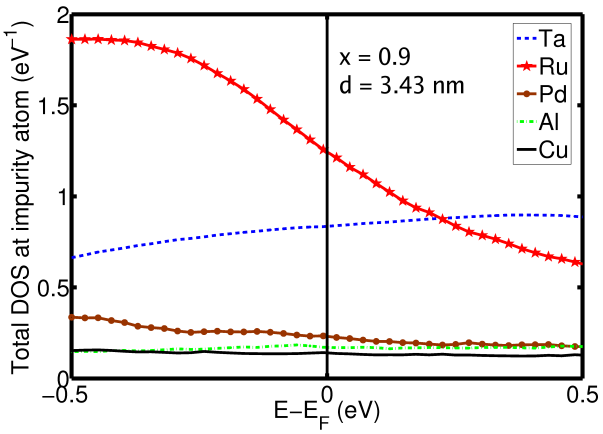

Figure 8.9: Density of states (DOS) as a function of energy E at the metal impurity atoms on the Cu surface. The solid line denoted by 'Cu' represents the DOS for Cu atom on a perfect Cu surface without any impurity.

d states, a change in the Fermi surface, and/or a difference in the lattice structure and crystal potentials. To better understand this effect, we calculated the density of states (DOS) at the barrier metal atoms on the Cu-barrier interface layer. The results are presented in Fig. 8.9. We found that the DOS for Al/Pd match very well with that for Cu atoms on a perfect Cu surface, while the DOS for Ta/Ru does not match. Therefore, the effect of Al/Pd barriers is to effectively smooth out the electronic structure of a rough Cu film toward that of a perfect film, thereby reducing the overall resistivity. For Ta/Ru barriers, even though the DOS around E_F is much higher than that for Cu, it is mostly composed of the rather localized d orbital. We may thus argue that the d states of the Ta/Ru atoms do not contribute as well to the overall conductance, and the mismatch of the DOS with that of Cu does not smooth out the electronic structure of the rough Cu film. To further confirm the behavior of DOS of Fig. 8.9, we have carried out additional calculations[140] using a projector augmented wave (PAW) DFT method as implemented in the electronic package VASP[141, 142] on *periodic* structures of the barrier coated films, and the results (not presented) show very good qualitative agreement with our two-probe results in Fig. 8.9.

8.4 Summary

In summary, we have investigated surface roughness induced resistivity of thin Cu films by using the NEGF-DFT-NVC *ab initio* formalism. Our calculations for bare rough films demonstrate that electron scattering due to surface roughness significantly enhances resistivity. Even in the low roughness regime, the resistivity of very thin Cu films remains substantial as compared with the bulk value. Our results can be fit to the Fuchs-Sondheimer semi-classical formula which allows us to estimate the degree of specular scattering, and for the first time in literature we have obtained the $p = p(x)$ characteristics.

It is found coating copper films by other barrier atoms may reduce or enhance resistivity. Compared with the bare rough surface, the resistivity value of Ti, Ta and Ru coated surface are higher while for Al and Pd the resistivity is lower. These different effects can be attributed to difference in the electronic states of the coating metals on the surface. Our results strongly suggest that it is possible to electronically smooth out the effects of roughness of thin Cu films such that the resulting Cu-barrier metal interface becomes more specular to electron scattering.

9

Conclusion

We have developed a non-equilibrium vertex correction (NVC) theory within the NEGF-DFT framework for analyzing quantum transport properties of disordered nanoelectronic devices under finite external bias potential that drives current flow. In the NEGF-DFT-NVC quantum transport method, the impurity averaging at the non-equilibrium density matrix level is carried out by self-consistently including the NVC self-energy that describes the change of the non-equilibrium distribution due to multiple impurity scattering. After the NEGF-DFT-NVC self-consistent calculation is converged for the Hamiltonian, a second, unavoidable vertex correction is needed to calculate the transmission coefficient that includes inter-channel disorder scattering. The entire theoretical formalism is implemented successfully in the linear Muffin Tin orbital DFT framework. At this point, it is worth to recall the well known field theoretical calculations of the Kubo formula for impurity scattering in metal[143]. There, the vertex correction is calculated at the equilibrium conductivity level[143]. In our work, the NVC theory is developed at the non-equilibrium density matrix level which is necessary for nonequilibrium quantum transport problems.

The NEGF-DFT-NVC theory has several desired features, including atomistic first principle, non-equilibrium, accurate, self-consistent and efficient computation. These features allow us to analyze nonlinear and non-equilibrium quantum transport properties of realistic device structures from atomic first principle without any phenomenological parameters.

We have applied the NEGF-DFT-NVC method to investigate effects of disordered impurities/defects to the quantum transport properties of several typical nanoelectronic device systems. First, it is found that for Fe/Vacuum/Fe magnetic tunnel junction (MTJ), the interface disorder can alter the bias dependence of TMR in significant ways: the zero bias anomaly can be observed in junctions having weak or no disorder, but it disappears when the interface disorder is increased. Second, we investigated the important role of disordered oxygen vacancies in Fe/MgO/Fe MTJ. The inclusion of oxygen vacancies significantly improves the agreement between first principle calculation and experiments: the TMR value is substantially quenched from the ideal theoretical limit of $10,000\%$ to the experimental range of around 250% by just a few percent of vacancies. Diffusive scattering by the oxygen vacancies inside the MgO barrier was found to exponentially enhance the the junction resistance, which was observed in recent experiments[110]. The TMR value of junctions having disordered interfacial oxygen vacancies can be rapidly reduced by bias. Third, we have calculated surface roughness scattering induced resistivity in copper interconnects. It is demonstrated that disordered surface roughness scattering significantly enhanced the resistivity of very thin copper lines. On the other hand, the resistivity can be lowered by coating the rough film using Al and Pd, while metals Ta, Ti and Ru increase the resistivity.

The research on effects of disordered impurities/defects is of vital importance for technological applications because all realistic systems have a degree of disorder. We believe our NEGF-DFT-NVC formalism and its associated software tool provide a powerful first principle approach for many future directions. Several immediate and very important problems can be tackled including quantum transport in dilute magnetic semiconductors[144]; non-uniform doping effects in nanoscale Si field effect transistors[145]; spin injection from ferromagnetic metal into GaAs where diffusion of metal atoms into GaAs create spin flipping scattering centers[146]; disorder effects in graphene and carbon nanostructures[147]; impurity scattering in the newly discovered topological insulators[148], etc.. At the writing of this thesis, we are carrying out all

of these calculations and results will be published in the near future.

On a more theoretical level, there is a need to improve certain issues in our existing NEGF-DFT-NVC method. These include relaxing the linear energy expansion in the LMTO and realize the MTO computation; relaxing the atomic sphere approximation and the single site approximation; as well as other issues in DFT itself such as the exchange-correlation functionals. With these improvements, more accurate results are expected and more complicated systems can be investigated. We hope to make these advances in the future.

Bibliography

[1] Y. Ke, K. Xia, and H. Guo. Theory of nonequilibrium vertex correction. *To be submitted to Phys. Rev. B*, 2010.

[2] Y. Ke, K. Xia, and H. Guo. Disorder scattering in magnetic tunnel junctions: Theory of nonequilibrium vertex correction. *Phys. Rev. Lett.*, 100:166805, 2008.

[3] Y. Ke, K. Xia, and H. Guo. Oxygen vacancy induced diffusive scattering in fe/mgo/fe magnetic tunnel junctions. *Being reviewed at Phys. Rev. Lett.*, 2010.

[4] Y. Ke and *et. al.* Resistivity of thin cu films with surface roughness. *Phys. Rev. B*, 79:155406, 2009.

[5] F. Zahid, Y. Ke, D. Gall, and H. Guo. Resistivity of thin cu films coated with ta, ti, ru, al, and pd barrier layers from first principles. *Phys. Rev. B*, 81:045406, 2010.

[6] see the website: http://www.itrs.net/.

[7] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys.Rev.*, 136.

[8] W. Kohn and L.J. Sham. Self-consistent equations including exchange and correlation effects. *Phys.Rev.*, 140, 1965.

[9] L.V. Keldysh. Diagram technique for non-equilibrium processes. *Soviet Phys.JETP*, 20:1018, 1965.

[10] S. Datta. Electronic transport in mesoscopic systems. *(Cambridge University Press, 1997).*

[11] H. Haug and A.-P. Jauho. Quantum kinetics in transport and optics of semiconductors. *(Springer; 2nd, 2007).*

[12] A.-P. Jauho, N.S. Wingreen, and Y. Meir. Time-dependent transport in interacting and noninteracting resonant-tunneling systems. *Phys. Rev. B*, 50:5528, 1994.

[13] J. Taylor, H. Guo, and J. Wang. Ab initio modeling of open systems: Charge transfer, electron conduction, and molecular switching of a c60 device. *Phys. Rev. B*, 63:121104.

[14] X. Zheng and *et. al.* Time-dependent density-functional theory for open systems. *Phys.Rev.B*, 75:195127, 2007.

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

[16] S.V. Faleev, F. Leonard, D.A. Stewart, and V.M. Schilfgaarde. Ab initio tight-binding lmto method for nonequilibrium electron transport in nanosystems. *Phys. Rev. B*, 71:195422, 2005.

[17] D. Waldron, P. Haney, B. Larade, A. MacDonald, and H. Guo. Nonlinear spin current and magnetoresistance of molecular tunnel junctions. *Phys. Rev. Lett.*, 96:166804, 2006.

[18] Y. Xue, S. Datta, and M.A. Ratner. Charge transfer and 'band lineup' in molecular electronic devices: A chemical and numerical interpretation. *J. Chem. Phys.*, 115:4292, 2002.

[19] M. Brandbyge and *et. al.* Density-functional method for nonequilibrium electron transport. *Phys. Rev. B*, 65:165401, 2002.

[20] S.-H. Ke, H.U. Baranger, and W.T. Yang. Electron transport through molecules:self-consistent and non-self-consistent approaches. *Phys. Rev. B*, 70:085410, 2004.

[21] A.R. Rocha and *et. al.* Towards molecular spintronics. *Nature Materials*, 4:335, 2005.

[22] W.C. Lu, M. Vicent, and J. Bernholc. Nonequilibrium quantum transport properties of organic molecules on silicon. *Phys. Rev. Lett.*, 95:206805, 2005.

[23] E.Y. Tsymbal, A. Sokolov, I.F. Sabirianov, and B. Doudin. Resonant inversion of tunneling magnetoresistance. *Phys. Rev. Lett.*, 90:186602, 2003.

[24] H. Ohno. Making nonmagnetic semiconductors ferromagnetic. *Science*, 281:951, 1998.

[25] T. Markussen, R. Rurali, A.P. Jauho, and M. Brandbyge. Scaling theory put into practice: First-principles modeling of transport in doped silicon nanowires. *Phys. Rev. Lett.*, 99:076803, 2007.

[26] D.B. Strukov, G.S. Snider, D.R. Stewart, and R.S. Williams. The missing memristor found. *Nature*, 453:80, 2008.

[27] W. Ren, Z. Qiao, J. Wang, Q. Sun, and H. Guo. Universal spin-hall conductance fluctuations in two dimensions. *Phys. Rev. Lett.*, 97:066603, 2006.

[28] P. Soven. Coherent-potential model of substitutional disordered alloys. *Phys. Rev.*, 156:809, 1967.

[29] D.W. Taylor. Vibrational properties of imperfect crystals with large defect concentrations. *Phys. Rev.*, 156:1017, 1967.

[30] B. Velicky, S. Kirkpatrick, and H. Ehrenreich. Single-site approximations in the electronic theory of simple binary alloys. *Phys. Rev.*, 175:747, 1968.

[31] G.M. Stocks and H. Winter. Self-consistent-field-korringa-kohn-rostoker-coherent-potential approximation for random alloys. *Z. Phys. B: Condens. Matter*, 46:95, 1982.

[32] O.K. Anderson and O. Jepsen. Explicit, first-principles tight-binding theory. *Phys. Rev. Lett.*, 53:2571, 1984.

[33] I. Turek and *et. al. Electronic Structure of the Disordered Alloys, Surfaces and Interfaces. (Kluwer, Boston, 1997).*

[34] J. Kudrnovsky and V. Drchal. Electronic structure of random alloys by the linear band-structure methods. *Phys. Rev. B*, 41:7515, 1990.

[35] J. Kudrnovsky, V. Drchal, and J. Masek. Canonical description of electron states in random alloys. *Phys. Rev. B*, 35:2487, 1987.

[36] R.V. Chepulskii and W.H. Butler. Temperature and particle-size dependence of the equilibrium order parameter of fept alloys. *Phys. Rev. B*, 72:134205, 2005.

[37] K. Carva, I. Turek, J. Kudrnovsky, and O. Bengone. Disordered magnetic multilayers: Electron transport within the coherent potential approximation. *Phys. Rev. B*, 73:144421, 2006.

[38] M. Levy. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem. *Proc.Natl.Acad.Sct.USA*, 76:6062, 1979.

[39] for a review, Y.A. Wang, and E.A. Carter. Chapter 5 of.

[40] U.V. Barth and L. Hedin. A local exchange-correlation potential for the spin polarized case: I. *J. Phys. C: Solid State Phys.*, 5:1629, 1972.

[41] J. Harris and R.O. Jones. The surface energy of a bounded electron gas. *J.Phys.F*, 4:1170, 1974.

[42] J. Harris. Adiabatic-connection approach to kohn-sham theory. *Phys.Rev.A*, 29:1648, 1984.

[43] D.C. Langreth and J.P. Perdew. The exchange-correlation energy of a metallic surface. *Solid State Communication*, 17:1425, 1975.

[44] O. Gunnarsson and B.I. Lundqvist. Exchange and correlation in atoms, molecules and solids by the spin-density-functional formalism. *Phys.Rev.B*, 13:4274, 1976.

[45] M. Levy and J.P. Perdew. Hellmann-feynman, virial, and scaling requisites for the exact universal density functionals. shape of the correlation potential and diamagnetic susceptibility for atoms. *Phys.Rev.A*, 32:2010, 1985.

[46] E.H. Lieb and S. Oxford. Improved lower bound on the indirect coulomb energy. *Int. J. Quantum Chem.*, 19:427, 1981.

[47] J.P. Perdew, R.G. Parr, M. Levy, and J.L. Balduz. Density-functional theory for fractional particle number: Derivative discontinuities of the energy. *Phys. Rev. Lett.*, 49:1691, 1982.

[48] L.J. Sham and M. Schluter. Density-functional theory of the energy gap. *Phys. Rev. Lett.*, 51:1888, 1983.

[49] J.P. Perdew and M. Levy. Physical content of the exact kohn-sham orbital energies: Band gaps and derivative discontinuities. *Phys. Rev. Lett.*, 51:1884, 1983.

[50] U.Von BARTH and L. HEDIN. A local exchange-correlation potential for the spin polarized case: I. *J. Phys. C: Solid State Phys*, 5:1629, 1972.

[51] J.P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865, 1996.

[52] A.D. Becke. A new mixing of hartreeĺcfock and local density functional theories. *J. Chem. Phys.*, 98:1372, 1993.

[53] B.G. Janesko, T.M. Henderson, and G.E. Scuseria. Screened hybrid density functionals for solid-state chemistry and physics. *Phys. Chem. Chem. Phys.*, 11:443, 2009.

[54] J. Muscat, A. Wander, and N.M. Harrison. On the prediction of band gaps from hybrid functional theory. *Chem.Phys.Lett.*, 342:397, 2001.

[55] J.P. Perdew and A. Zunger. Self-interaction correction to density functional approximations for many-electron systems. *Phys.Rev.B*, 23:5048, 1981.

[56] A.J. Cohen, P. Mori-Sanchez, and W. Yang. Fractional charge perspective on the band gap in density-functional theory. *Phys.Rev.B*, 77:115123, 2008.

[57] A.J. Cohen, P. Mori-Sanchez, and W. Yang. Insights into current limitations of density functional theory. *Science*, 321:792, 2008.

[58] J.D. Talman and W.F. Shadwick. Optimized effective atomic central potential. *Phys. Rev. A*, 14:36, 1976.

[59] V.I. Anisimov, J. Zaanen, and O.K. Andersen. Band theory and mott insulators: Hubbard u instead of stoner i. *Phys. Rev. B*, 44:943, 1991.

[60] A.I. Liechtenstein, V.I. Anisimov, and J. Zaanen. Density-functional theory and strong interactions: Orbital ordering in mott-hubbard insulators. *Phys. Rev. B*, 52, 1995.

[61] J.P. Perdew and *et. al.* Accurate density functional with correct formal properties: A step beyond the generalized gradient approximation. *Phys. Rev. Lett.*, 82:2544, 1999.

[62] F. Tran and P. Blaha. Accurate band gaps of semiconductors and insulators with a semilocal exchange-correlation potential. *Phys. Rev. Lett.*, 102:226401, 2009.

[63] M. L*ü*ders and *et. al.* Ab initio angle-resolved photoemission in multiple-scattering formulation. *J. Phys. Cond. Mat.*, 13:8587, 2001.

[64] O.K. Andersen. Linear method in band theory. *Phys. Rev. B*, 12:3060, 1975.

[65] H.L. Skriver. The linear muffin tin orbital method. *(Springer, Berlin, 1984).*

[66] V. Eyert. A comparative study on methods for convergence acceleration of iterative vector sequences. *J. Comput. Phys.*, 124:271, 1996.

[67] O.K. Andersen, O. Jepsen, and D. Glötzel. in 'highlights of condensed matter theory'. *edited by F. Bassani, F. Fumi, and M.P. Tosi (North Holland, New York, 1985).*

[68] D. Glotzel, B. Segall, and O.K. Andersen. Self-consistent electronic structure of si, ge and diamond by the lmto-asa method. *Solid State Communications*, 36:403, 1980.

[69] J. Maciejko, J. Wang, and H. Hong Guo. Time-dependent quantum transport far from equilibrium: An exact nonlinear response theory. *Phys.Rev.B*, 74:085324, 2006.

[70] M.E. Godfrin. A method to compute the inverse of an n-block tridiagonal quasi-hermitian matrix. *J. Phys.: Condens. Matter*, 3:7843, 1991.

[71] M.P.L. Sancho, J.M.L. Sancho, and J. Rubio. Highly convergent schemes for the calculation of bulk and surface green functions. *I. Phys. F: Met. Phys.*, 15:851, 1985.

[72] H. H. Mehrez and *et. al.* I-v characteristics and differential conductance fluctuations of au nanowires. *Phys.Rev.B*, 65:195419, 2002.

[73] A. Gonis. Green's functions for ordered and disordered systems. *(North-Hilland, Amsterdam, 1992).*

[74] P. Weinberger. Electron scattering theory for ordered and disordered matter. *(Clarendon Press, Oxford, 1990).*

[75] I. Turek and *et. al.* Electronic structure of disordered alloys, surfaces and interfaces. *(Kluwer, Dordrecht, 1997).*

[76] G. Reiss and *et. al.* Magnetic tunnel junctions. *STMP*, 227:291, 2007.

[77] E.Y. Tsymbal, O.N. Mryasov, and P.R. LeClair. Spin-dependent tunnelling in magnetic tunnel junctions. *J. Phys.: Condens. Matter*, 15, 2003.

[78] S.S.P. Parkin and *et al.* Magnetically engineered spintronic sensors and memory. *PROCEEDINGS OF THE IEEE*, 91:661, 2003.

[79] S.A. Wolf and *et. al.* Spintronics: A spin-based electronics vision for the future. *Science*, 294:1488, 2001.

[80] M. Jullier. Tunneling between ferromagnetic films. *Phys. Lett.*, 54.

[81] J.S. Moodera and *et. al.* Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions. *Phys. Rev. Lett.*, 74:3273, 1995.

[82] T. Miyazaki and N. Tezuka. Giant magnetic tunneling effect in fe/al2o3/fe junction. *J.Magn.Magn.Mater.*, 139, 1995.

[83] A. Kaufler and *et. al.* Tunnel-magnetoresistance system with an amorphous detection layer. *J. Appl. Phys.*, 91:1701, 2002.

[84] S. Yuasa and *et al.* Giant room-temperature magnetoresistance in single-crystal fe/mgo/fe magnetic tunnel junctions. *Nat. Mater.*, 3:868, 2004.

[85] S.S.P. Parkin and *et al.* Giant tunnelling magnetoresistance at room temperature with mgo (100) tunnel barriers. *Nat. Mater.*, 3:862, 2004.

[86] S. Ikeda and *et. al.* Tunnel magnetoresistance of 604diffusion in cofeb/mgo/cofeb pseudo-spin-valves annealed at high temperature. *J. Appl. Phys.*, 93:082508, 2008.

[87] J.M.De Teresa and *et. al.* Role of metal-oxide interface in determining the spin polarization of magnetic tunnel junctions. *Science*, 286:507, 1999.

[88] J.A. Stroscio and *et. al.* Tunneling spectroscopy of bcc (001) surface states. *Phys. Rev. Lett.*, 75:2960, 1995.

[89] S.N. Okuno, T. Kishi, and K. Tanaka. Spin-polarized tunneling spectroscopy of co(0001) surface states. *Phys. Rev. Lett.*, 88:066803, 2002.

[90] H.F. Ding and *et. al.* Absence of zero-bias anomaly in spin-polarized vacuum tunneling in co(0001). *Phys. Rev. Lett.*, 90:116603, 2003.

[91] M.M.J. Bischoff and *et. al.* Local electronic structure of fe(001) surfaces studied by scanning tunneling spectroscopy. *Phys. Rev. B*, 68:045422, 2003.

[92] O. Wunnicke and *et. al.* Effects of resonant interface states on tunneling magnetoresistance. *Phys. Rev. B*, 65:064425, 2002.

[93] C. Uiberacker and P.M. Levy. Role of symmetry on interface states in magnetic tunnel junctions. *Phys. Rev. B*, 64:193404, 2001.

[94] S. Zhang, P.M. Levy, A.C. Marley, and S.S.P. Parkin. Quenching of magnetoresistance by hot electrons in magnetic tunnel junctions. *Phys. Rev. Lett.*, 79:3744, 1997.

[95] J. Zhang and R.M. White. Voltage dependence of magnetoresistance in spin dependent tunneling junctions. *J. Appl. Phys.*, 83:6512, 1998.

[96] J.S. Moodera, J. Nowak, and J.M. van de Veerdonk. Interface magnetism and spin wave scattering in ferromagnet-insulator-ferromagnet tunnel junctions. *Phys. Rev. Lett.*, 80:2941, 1998.

[97] R. Jansen and J.S. Moodera. Magnetoresistance in doped magnetic tunnel junctions: Effect of spin scattering and impurity-assisted transport. *Phys. Rev. B*, 61:9047, 2000.

[98] P.X. Xu and *et. al.* Influence of roughness and disorder on tunneling magnetoresistance. *Phys. Rev. B*, 73:180402, 2006.

[99] W.H. Butler and *et.al.* Spin-dependent tunneling conductance of fe|mgo|fe sandwiches. *Phys. Rev. B*, 63:054416, 2001.

[100] J. Mathon and A. Umerski. Theory of tunneling magnetoresistance of an epitaxial fe/mgo/fe(001) junction. *Phys. Rev. B*, 63:220403.

[101] A.A. Tulapurkar and *et al.* Spin-torque diode effect in magnetic tunnel junctions. *nature*, 483:339, 2005.

[102] X.-G. Zhang, W.H. Butler, and A. Bandyopadhyay. Effects of the iron-oxide layer in fe-feo-mgo-fe tunneling junctions. *Phys. Rev. B*, 68:092402, 2003.

[103] D. Waldron and *et. al.* First principles modeling of tunnel magnetoresistance of fe/mgo/fe trilayers. *Phys. Rev. Lett.*, 97:226802, 2006.

[104] H.L. Meyerheim and *et. al.* Geometrical and compositional structure at metal-oxide interfaces: Mgo on fe(001). *Phys. Rev. Lett.*, 87:076102, 2001.

[105] S.G. Wang and *et. al.* Evidence for feo formation at the fe/mgo interface in epitaxial tmr structure by x-ray photoelectron spectroscopy. *J. Magn. Magn. Mater.*, 310:1935, 2007.

[106] M. Sicot and *et. al.* On the quality of molecular-beam epitaxy grown fe/mgo and co/mgo(001) interfaces. *J. Appl. Phys.*, 99:08.

[107] K. Miyokawa and *et. al.* X-ray absorption and x-ray magnetic circular dichroism studies of a monatomic fe(001) layer facing a single-crystalline mgo(001) tunnel barrier. *Jpn. J. Appl. Phys., Part 1*, 44, 2005.

[108] L. Plucinski and *et. al.* Mgo/fe(100) interface: A study of the electronic structure. *Phys. Rev. B*, 75:214411, 2007.

[109] P. Luches and *et. al.* Absence of oxide formation at the fe/mgo(001) interface. *Surf. Sci.*, 583:191, 2005.

[110] G.X. Miao, Y.J. Park, and J.S. Moodera. Disturbance of tunneling coherence by oxygen vacancy in epitaxial fe/mgo/fe magnetic tunnel junctions. *Phys. Rev. Lett.*, 100:246803, 2008.

[111] P. Mather, J. Read, and R. Buhrman. Disorder, defects, and band gaps in ultrathin (001) mgo tunnel barrier layers. *Phys. Rev. B*, 73:205412, 2006.

[112] J.P. Velev and *et. al.* Effect of oxygen vacancies on spin-dependent tunneling in fe/mgo/fe magnetic tunnel junctions. *Appl. Phys. Lett.*, 90:072502, 2007.

[113] J. Carrasco, N. Lopez, and F. Illas. First principles analysis of the stability and diffusion of oxygen vacancies in metal oxides. *Phys. Rev. Lett.*, 93:225502, 2004.

[114] B.D. Yu. Ab initio study of the adsorption of fe atoms on a defective mgo(001) surface: Blind adsorption. *Phys. Rev. B*, 71:193403, 2005.

[115] C.H. Yang, M. Samant, and S.S. Parkin. American physical society march meeting, 2009, abstract v22, p00004.

[116] M. Pesci and *et. al.* Nature of defect states in nitrogen-doped mgo. *J. Chem. Phys. C*, 114:1350, 2010.

[117] S.M. Rossnagel and T.S. Kuan. Alteration of cu conductivity in the size effect regime. *J. Vac. Sci. Technol. B*, 22:240, 2004.

[118] T.S. Kuan and *et. al.* in materials, technology, and reliability for advanced interconnects and low-k dielectrics, mrs symposia proceedings no. 612. *(Materials Research Society, Pittsburgh, 2000), p. D7.1.1.*

[119] J.J. Plombon and *et. al.* Influence of phonon, geometry, impurity, and grain size on copper line resistivity. *Appl. Phys. Lett.*, 89:113124, 2006.

[120] S.M. Aouadi, P.K. Shreeman, and M. Williams. Real-time spectroscopic ellipsometry study of ultrathin diffusion barriers for integrated circuits. *J. Appl. Phys.*, 96:3949, 2004.

[121] J.M. Purswani and D. Gall. Electron scattering at single crystal cu surfaces. *Thin Solid Films*, 516:465, 2007.

[122] H.-D. Liu and *et. al.* Thickness dependent electrical resistivity of ultrathin (<40 nm) cu films. *Thin Solid Films*, 384:151, 2001.

[123] E.V. Barnat and *et. al.* Real time resistivity measurements during sputter deposition of ultrathin copper films. *J. Appl. Phys.*, 91:1667, 2002.

[124] K. Fuchs. The conductivity of thin metallic films according to the electron theory of metals. *Proc. Cambridge Philos. Soc.*, 34:100, 1938.

[125] E.H. Sondheimer. The mean free path of electrons in metals. *Adv. Phys.*, 1:1, 1952.

[126] M.S.P. Lucas. Electrical conductivity of thin metallic films with unlike surfaces. *J. Appl. Phys.*, 36:1632, 1965.

[127] A.F. Mayadas and M. Shatzkes. Electrical-resistivity model for polycrystalline films: the case of arbitrary reflection at external surfaces. *Phys. Rev. B*, 1:1382, 1970.

[128] Y. Namba. Resistivity and temperature coefficient of thin metal films with rough surface. *Jap. J. Appl. Phys.*, 9:1326, 1970.

[129] H. Hoffmann and J. Vancea. Critical assessment of thickness-dependent conductivity of thin metal films. *Thin Solid Films*, 85:147, 1981.

[130] X.-G. Zhang and W.H. Butler. Conductivity of metallic films and multilayers. *Phys. Rev. B*, 51:10085, 1995.

[131] N. Trivedi and N.W. Ashcroft. Quantum size effects in transport properties of metallic films. *Phys. Rev. B*, 38:12298, 1988.

[132] Z. Tesanovic, M.V. Jaric, and S. Maekawa. Quantum transport and surface scattering. *Phys. Rev. Lett.*, 57:2760, 1986.

[133] V. Timoshevskii *et. al.* . The influence of surface roughness on electrical conductance of thin cu films: An ab initio study. *J. Appl. Phys.*, 103:113705, 2008.

[134] Y. Zhou and *et. al.* The influence of surface roughness on electrical conductance of thin cu films: An ab initio study. *J. Phys. Conden. Matt.*, 20:095209, 2008.

[135] J.M. Purswani and D. Gall. Surface morphological evolution during annealing of epitaxial cu(001) layers. *J. Appl. Phys.*, 104:044305, 2008.

[136] J.S. Chawla and D. Gall. Specular electron scattering at single-crystal cu(001) surfaces. *Appl. Phys. Lett.*, 94:252101, 2009.

[137] K.L. Ou and *et. al.* Comparative study of polycrystalline ti, amorphous ti, and multiamorphous ti as a barrier film for cu interconnect. *J. Vac. Sci. Technol. B*, 23:229, 2005.

[138] S. Tsukimoto and *et. al.* Fabrication of cu(ti) alloy interconnects with self-formation of thin barrier metal layers using a high-pressure annealing process. *J. Electron. Mater.*, 36:1658, 2007.

[139] D.Y. Shih and *et. al.* Thin film interdiffusions in cu/pd, cu/pt, cu/ni, cu/nib, cu/co, cu/cr, cu/ti, and cu/tin bilayer films: Correlations of sheet resistance with rutherford backscattering spectrometries. *J. Appl. Phys.*, 70:3052, 1991.

[140] For the dos calculations with vasp, we used a supercell consists of a 7 layer (4×4) cu (110) periodic atomic structure with one impurity metal atom replacing a cu atom at the surface layer. the vacuum region in the z-direction of the supercell is set to be around 15 å. the four top layers are allowed to relax while the three bottom layers are fixed at the bulk position with a bulk cu lattice constant 3.61 å, consistent with our lmto calculations. an energy cutoff of 275 ev is used in the plane wave basis. the monkhorst-pack scheme is adopted for integration of brillouin zone with a k-mesh of $8\times8\times1$. note that, the vasp calculations are only possible for peridodic structures whereas our lmto calculations are performed treating the cu film as a two probe device of specific length.

[141] G. Kresse and J. Hafner. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B*, 47:558, 1993.

[142] G. Kresse and J. Furthmuller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169, 1996.

[143] G.D. Mahan. Many particle physics. *(Plenum, New York, 1990)*.

[144] T. Jungwirth and *et. al.* Theory of ferromagnetic (iii,mn)v semiconductors. *Rev. Mod. Phys.*, 78:809, 2006.

[145] P. Zhang and *et. al.* Electronic transport in nanometre-scale silicon-on-insulator membranes. *Nature*, 439:703, 2006.

[146] M. Oestreich. Materials science: Injecting spin into electronics. *Nature*, 402:735, 1999.

[147] R.van Noorden. Moving towards a graphene world. *Nature*, 442:228, 2006.

[148] G. Brumfiel. Topological insulators: Star material. *Nature*, 466:310, 2010.