Atomistic modeling of

random impurity effects in nano devices

Qing Shi 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 Science

Contents

Ab	stract		vi
Ré	sumé		viii
Sta	Statement of Originality		
Ac	knowle	dgments	xii
1 Int	roduct	ion	1
2 De	ensity F	unctional Theory	7
2.1	Basic	Theorems and Assumptions of DFT	8
	2.1.1	Born-Oppenheimer Approximation	8
	2.1.2	Hohenberg-Kohn Theorem	12
	2.1.3	Kohn-Sham Equation	15
	2.1.4	Exchange Correlation Potential	16
	2.1.5	Self-consistent Procedure of DFT	22
2.2	Linea	Combination of Atomic Orbital Implementation	23
	2.2.1	Atomic Orbitals	23
	2.2.2	Pseudopotential approximation	25
2.3	Linea	Muffin Tin Orbital Implementation	26
	2.3.1	Atomic Sphere Approximation (ASA)	26
	2.3.2	Constructing Muffin Tin orbitals	28
	2.3.3	Energy-independent Muffin Tin orbitals	33
	2.3.4	Overlap and Hamiltonian Matrices	34

	2.3.5 Potential Parameters	36
	2.3.6 Representation Transformation	38
2.4	Summary	39
3 Ap	oplication: Band Engineering of the GaSbN Alloy	42
3.1	Computational Method	44
3.2	Computational Setup	47
3.3	Results and Discussion	48
3.4	Summary	53
4 Gr	een's Function Theory for Quantum Transport	54
4.1	Landauer Picture for Quantum Transport	54
4.2	NEGF	56
4.3	NEGF-DFT Implementation	60
	4.3.1 Self-Consistent Procedure	60
	4.3.2 Energy Integration	60
4.4	NEGF-DFT in TB-LMTO Method	63
4.5	Summary	65
5 Di	sorder Scattering	67
5.1	Coherent Potential Approximation	69
5.2	Nonequilibrium Coherent Potential Approximation and Physical Quan- tities	78
	5.2.1 Nonequilibrium Coherent Potential Approximation	78
	5.2.2 Self-consistent solution of NECPA equations	81
	5.2.3 Charge Density and Transmission Coefficient	82
5.3	NECPA with Transverse Periodicity	83
5.4	NECPA with LMTO-ASA	85

	5.5	Summary	86
6	Ap	plication: Diffusive Transport in Graphene Tunnel-FET	87
	6.1	Device Structure	89
	6.2	Disordered Electronic Materials	90
	6.3	Transport with Disorder Scattering	93
	6.4	Summary	96
7	NE	CPA with LCAO Basis	97
	7.1	Tight Binding Verification	99
	7.2	DFT Implementation	104
		7.2.1 Determination of impurity Hamiltonian	106
		7.2.2 Code implementation	108
		7.2.3 Verification by doped graphene system	115
	7.3	Application: Dopant Limited Mobility of MoS_2	122
	7.4	Summary	125
8	Co	nclusions	126
A	ST	RUCTURE CONSTANTS	130
B	8 W	RONSKIAN	131
C	C LN	ATO GREEN'S FUNCTIONS	132
	C.1	Retarded/Advanced Green's function	132
	C.2	Two-probe retarded/advanced Green's function $\ldots \ldots \ldots \ldots$	134
Ľ) AN	NEQUIVALENT FORM OF THE CPA EQUATION SET	137
	D.1	Rewriting the CPA Equations with the Conditional Greens function	137
	D.2	The Introduction of Interactor Ω_i	139

E	BA	C Model and Molecular Dynamics	140
	E.1	Extraction of the coupling constant	140
	E.2	Molecular Dynamics	141
\mathbf{F}	NE	CPA EQUATION SET IN LMTO-ASA	144
	Ref	erences	145

LIST OF FIGURES

4	Diagram of a simulated device in NEGF-DFT method. This is also called a "two- probe" model. It consists of two semi-infinite leads (left lead and right lead outside the simulation box) and a central region (in rectangular box). Parts of the leads, called buffer layers are included in the simulation box to ensure a smooth match of the potential at the edges of the simulation box. The grey part in the center is the device part (also called "scattering region"). The simulation is periodically repeated in the cross-section of the device	1.1
20	Comparison of the band gaps calculated by RESCU with HSE06 functional and experimental data. Group V, III-V and II-VI semiconductors are plotted in the figure. Dotted line refers to straight line $y = x$. From left to right, the materials are InAs, InSb, Ge, GaSb, Si, GaAs, CdSe, InP, CdTe, AlSb, ZnTe, CdS, ZnSe, AlP, AlAs, GaP, ZnS, C and BN.	2.1
22	Flowchart of self-consistent solution procedure in DFT calculation	2.2
27	Diagram for Muffin Tin potentials. The whole space is divided into two regions: the slightly overlapped atomic spheres (showed in dashed spheres) and the in-between interstitial region. In atomic spheres the potentials are chosen to be spherical symmetric while in interstitial region potential is chosen to be a constant V_0 .	2.3
43	Schematic of the electrochemical water splitting process. A semiconductor is placed in neutral water. The red lines are the CBM and VBM of the semiconductor mea- sured with respect to the vacuum energy level. The blue dashed lines are the redox potential of neutral water. When light $(h\nu)$ is shined onto the semiconductor, elec- tron hole pairs are generated and captured by H^+ and OH^- in water respectively.	3.1
45	Atomic supercells of the simulated GaNSb lattice under different Sb concentration. The lattices is in parallelepiped shape with the three vector angles of $\pi/2$, $\pi/2$ and $\pi/3$. (a) Sb concentration of 2.7%. (b) Sb concentration of 0.35%.	3.2
45	Surface calculation of a GaN surface. The bottom figure is the lattice structure of a GaN slab. The top figure plots the electrostatic potential along the slab. In the center of the slab, a unit cell of GaN is picked up to calculate the average potential (red dashed line). ΔV measures the difference between the calculated average potential and the vacuum energy.	3.3

3.4	Band gap $E_g(x)$ of dilute $GaSb_xN_{1-x}$ alloy vs. Sb concentration x by KS-DFT with the HSE06 hybrid functional. Inset: E_g vs. lattice mismatch for $GaSb_xN_{1-x}$ (green) and $In_xGa_{1-x}N$ (red). The mismatch is measured w.r.t. wurtzite GaN crystal	48
3.5	DOS of $GaSb_xN_{1-x}$ at the valence band side of the Fermi level. (a) Total DOS normalized to the same number of Sb atoms at different x . The band edge of GaN, $E_{v,GaN}$, is used as reference to align the DOS at different Sb concentration. (b) Total DOS per atom. For clarity, (a, b) do not include the calculated data at $x = 0.7\%, 2.7\%$ and 5.6% as they are similar to the presented curves of $x = 0.35\%$ and 1.4%.	49
3.6	Projected DOS of GaSb _x N _{1-x} for $x = 2.7\%$. DOS is projected onto s, p orbitals of N atoms and s, p, d orbitals of Ga atoms. The Fermi level is shifted to 0 eV.	50
3.7	Band alignments relative to the vacuum level as a function of antimony concentra- tion x . Band positions are aligned to vacuum energy by surface calculations. The blue (red) dots/dashed lines are for relaxed (unrelaxed) surface cases. Black lines refer to the neutral water redox potentials of H^+/H_2 and O_2/H_2O	52
4.1	Diagram for Landauer picture. A two-probe structure is presented in the figure. The grey ellipse in the center denotes the central region of the device where scattering occurs. The device is connected to the outside two reservoirs through two leads (the two horizontal light blue bars). The two reservoirs stays in equilibrium respectively with independent chemical potentials $\mu_{L/R}$ and distribution functions $f_{L/R}$. We use a battery symbol to show that the electrochemical potential of the two reservoirs can be different.	55
4.2	Diagram for self-consistent procedure of NEGF-DFT technique. All physical quan- tities including self-energies and Green's functions are matrices under some type of basis, e.g. LCAO basis or LMTO basis.	61
4.3	Diagram of integration path when calculating electron density. The red curve refers to the path we use to integrate lesser Green's function	62
5.1	Schematic plot of a two-probe system with some impurity sites in the central scat- tering region. The left/right leads extend to $z = \pm \infty$. The device extends to $x = \pm \infty$ in the transverse direction. The black dots are sites of electrodes, white circles are pure sites in the scattering region, white sites with a cross in the center are impurity sites	68
5.2	Schematic plot of a isolated systems. (a) isolated system with some random impurities. The white circles are pure sites in the material and the white sites with a cross in the center are impurity sites. (b) isolated system with CPA. The circles in light blue are the stema with extended and the stema is the stema of the stema and the stema of	70
	ngnt blue are the atoms with concrent potentials	10

5.3

6.1

6.2

6.3

6.4

6.5

6.6

Schematic plot of the complex-time contour that goes above the real-time axis from $\tau = -\infty$ to $+\infty$, and returns below the real-time axis to $\tau = -\infty$. The contour-ordered Green's function is defined on the complex-time contour	80
Lattice structure of graphene. (a) plots the top view of the graphene primitive cell. (b) plots the structure of graphene 4×4 super cell. The green (grey) sites refer to the graphene A (B) sublattice.	89
Schematic of the co-doped graphene TFET. The top figure shows the side view of the TFET atomic structure. The channel is made of co-doped graphene and is attached to metal gate through a layer of Al_2O_3 dielectric material. The source/drain are connected to the outside via copper contacts. The red box highlights the part to simulate. The bottom left figure shows the top view of the graphene/copper interface. The bottom right figure shows the atomic structure of the co-doped graphene. Grey spheres refer to carbon atoms and green (silver) spheres refer to boron (nitrogen) dopants.	90
Schematic figures for co-doped graphene. The grey spheres refer to carbon host atoms and the green (silver) spheres correspond to doped boron (nitrogen) atoms. (a) A particular configuration of B-N substitution co-doping in graphene where the co-doping concentration is $x\%$. (b) A top view of the effective medium of co-doped graphene. The effective medium by CPA where a site has $x\%$ probability to be impurity and $(1 - x)\%$ probability to be the host atom. The dopant atoms are plotted with a much smaller radius to make the host atoms (carbon) visible from top view.	91
Electronic band structure of graphene. (a) Band structure of pristine graphene. (b) CPA band of disordered graphene with B-N co-doping at 10%. The CPA bands has a finite width reflecting the disorder broadening. A band gap opened around Dirac cone is observed. Fermi levels are fixed at "0" eV for both figures	92
Density of states (DOS) of the co-doped graphene. (a) DOS versus energy for several co-doping concentrations. A near linear dependence of the average band gaps on the concentration in the interested range is showed in the inset. (b) The averaged DOS versus energy of the B-N co-doped graphene with extra individual boron (p-type) or nitrogen (n-type) impurity atoms. Fermi levels move accordingly to the doping type, as expected. The change in band gap is very small with the extra individual dopants as investigated.	93
Top views of the simulated structure of the double-gate TFET (the active region). The leads provide self-energy to the central region in the NEGF-DFT self-consistent analysis. (a) The structure where doping is done by VCA. (b) The structure where CPA accounts for multiple disorder scattering.	93

94	⁷ Band diagrams of the TFET along the channel. (a) Energy profiles and Fermi level in the (B-N) co-doped graphene TFET obtained from the first principle simulation method with NECPA. Scattering region I includes (B-N) co-doping and extra p- type doping, and region II includes the co-doping. (b) LDOS of the co-doped graphene TFET in Fig. 6.6(b). The interband tunneling is clearly seen from LDOS in the band gap. A direct tunneling current is also observed	6.7
95	Transport properties of the simulated co-doped graphene TFET. (a) Transmissions of the modeled graphene TFET with VCA doping and CPA doping. The significant decrease in the NECPA results is caused by the diffusive inter-band tunneling due to the strong disorder scattering. (b) Drain current of the 10 nm long graphene TFET with (blue circles) and without (black squares) disorder scattering. Disorder scattering dramatically reduces the inter-band tunneling current	6.8
98	Illustrating ASA in some LMTO-ASA calculations. Vacuum spheres are labeled by "Vac". Different vacuum spheres are labeled with subscript numbers. (a) ASA for graphene lattice. It has been used in Chapter 6; (b) ASA for monolayer black phosphorus lattice. The spheres are slightly overlapped in space. The radii of all spheres has been reduced in the figure to make the figure less crowded	7.1
100	2 Schematic plot of 1D two-probe system with some impurity sites in the central scattering region investigated by tight binding. The left/right leads extend to $z = \pm \infty$. The black dots are sites of leads and grey circles are disorder sites in the scattering region.	7.2
102	3 Transmission coefficient $vs.$ energy. The energy is in arbitrary unit. The left column figures are full plots and the right column are the zoomed in ones in proper energy range. The black dots and curve refer to the result of NECPA-LCA. The other shapes and colours are brute force calculation for different disorder strengths t_{11} ranging from 0.5 to 1.5. From up to down, the concentration varies from 0.1 to 0.001, as shown in (a), (c), (e) and (g).	7.3
103	Transmission difference (ΔT) vs. impurity coupling (t_{11}) and impurity concentra- tion (x) . (a) Dependence of ΔT on t_{11} . The blue, red, yellow and purple curves correspond to $x = 0.1, 0.05, 0.01, 0.001$ respectively. (b) Dependence of ΔT on x . The blue, red, yellow, purple curves correspond to $t_{11} = 0.5, 0.9, 0.11, 1.5$ respec- tively.	7.4
105	5 Plots for S blocks (diagonal). (a) S for carbon atom. (b) S for boron atom. The side colour bar indicates the element values. (c) is a side view of carbon S matrix, seen along the diagonal direction. (d) is similar to (c), but for boron. The horizontal axis of the figures is the index of the atomic orbitals, namely s , s , y , z , x , y , z , x , xy , yz , $(3z^2 - 1)/2/\sqrt{3}$, zx and $(x^2 - y^2)/2$ orbitals, where repeated orbitals represent the double- ζ polarizations.	7.5

7.6	Determination of energy shift Δ . The brown atoms are the host atoms and the green one is the impurity. (a) atomic structure of the two-probe system. (b) atomic structure of the bulk structure. The structure is chosen to be very large so that atoms far away from the impurity is not affected. (c) a schematic plot of how to determine the energy shift. h_1 is the host atom Hamiltonian from the bulk calculation and h_2 is the host atom Hamiltonian from the two-probe calculation. s is the overlap matrix of the host atom.	107
7.7	A flowchart of solving NECPA within LCAO. The three coloured blocks refers to the three main steps: preparing data (in red), solving the CPA equations self-consistently (in blue) and solving corresponding less quantities (quantities with superscript $<$) self-consistently (in green). Quantities in light grey are initial guess that enters the self-consistent loops.	108
7.8	Execution time vs. number of PL's. All matrix operations are performed for 50 times. The vertical axis refers to the time for 50 times of operations. Each block is 40×40 . (a) matrix inverse. The PL method is compared with the LU decomposition method implemented in MATLAB. (b) $inv(H) \cdot D \cdot inv(H^{\dagger})$. The PL method is compared with the direct method.	111
7.9	Rate of convergence using the Anderson mixing (blue curves) and no mixing (red curves). (a) testing with a simple numerical example $f_i = -d_{ii}x_i - cx_i^3$ for $i = 1, \dots, 5$. <i>M</i> is set to be 3 for the Anderson mixing. (b) testing with a 5 nm graphene lattice. The lattice is doped with 1% of boron. <i>M</i> is set to be 5 for the Anderson mixing. $\dots \dots \dots$	113
7.10	Lattice structure of the simulated graphene two-probe system. The two leads ex- tend to $z = \pm \infty$ in the transport direction. The system extends to $x = \pm \infty$ in the transverse direction. A vacuum of 10 Å is added to y direction. The two-probe channel is along zigzag direction of graphene and has a length of L. Boron dopants distribute uniformly inside the central region (showed by orange atoms). Both the boron concentration and channel length L are to vary according to our requirement.	115
7.11	Plots for Δ matrices of the two impurities. (a) Δ for boron atom that occupies type A graphene site. (b) Δ for boron atom that occupies type B graphene site. The 2 figures at the bottom show their corresponding structures. Both matrices shows very good diagonality. The off-diagonal elements of Δ is smaller than 0.002. The numbers \ldots	115
7 1 9	Plots to varify the Weld identity relation. The vertical axis is the error of the Weld	

7.12 Plots to verify the Wald identity relation. The vertical axis is the error of the Wald identity relation. It is defined as the max element of the matrix $|G^r - G^a - G^> + G^<|$. The horizontal axis refers to the positive infinitesimal number added to the energy. 117

7.13	Transmission vs. energy. The boron concentration is 5%. The black curve is the transmission function of the clean system (directly calculated by NanoDcal). The red curve with circular dots is the transmission function calculated by our NECPA-LCA code. The blue dots are results calculated by the NanoDsim package. The three are shifted so that their Dirac cones are at $E = 0$.	118
7.14	Transmission at E_f vs. Boron concentration. The boron concentration varies from 0.05 to 0.001. The blue curve with circular dots is for the clean graphene. The yellow curve with crosses is for the results calculated by the NanoDsim package. The red curve with square dots is for our code. Only the transmission at Fermi energy E_f is plotted, where E_f is determined by DFT calculations.	119
7.15	Resistance vs. structure length. We calculate the resistance of graphene devices at various lengths $(L = 3 \text{ nm}, 4 \text{ nm}, 5 \text{ nm}, 6 \text{ nm} \text{ and } 7 \text{ nm})$. The central region is doped with boron at a concentration of 0.1%. The extracted mobility is shown in the figure too.	121
7.16	Lattice structures of the simulated MoS_2 . (a) and (b) are the top view and side view of the MoS_2 lattice. The black box is the primitive cell of monolayer MoS_2 . (c) and (d) are top views of the simulated two-probe MoS_2 structures. In (c)/(d), the S/Mo atoms in the central region are doped with Cl/Re atoms.	122
7.17	Band diagrams of the MoS_2 MOSFET at low bias. The source/drain (leads) are high <i>n</i> -doped (1%) and the channel (central region) is lightly <i>n</i> -doped. (a) low V_G . The barrier at the central region blocks the electron transport path so that the current can not flow through, referring to an "off" state. (b) high V_G . By tuning V_G , the barrier gets lower so that the electrons can pass through, referring to an "on" state.	123
7.18	Resistance vs. structure length. We calculate the resistance of MoS_2 devices at various lengths ($L = 2.8 \text{ nm}$, 3.7 nm , 4.7 nm , 5.6 nm and 6.6 nm). The central region is doped with 1) Cl at S sites with a concentration of 0.1% (red dots) and 2) Re at Mo sites with a concentration of 0.2% (purple dots). The two cases gives the same number of impurities in the lattice. The extracted mobilities are shown in the text boxes beside the curves.	125

LIST OF TABLES

2.1	Comparison between LCAO and LMTO used by this thesis. VS refers to vacuum sphere.	39
E.1	Tersoff potentials for GaNSb alloys (Part 1). The table is organized according to the favor of LAMMPS	141
E.2	Tersoff potentials for GaNSb alloys (Part 2)	142

Abstract

In the past five decades, the size of electronic devices in a computer chip has shrunk exponentially according to the Moore's law. As electronic devices become smaller and smaller, the continual down-scaling of the technology is increasingly challenged by many emerging problems. One of the most important issues is the device-todevice variation caused by random discrete dopants (RDD) (also known as random impurities). The RDD effect can render various variations to device properties such as the material electronic band structures and transport. In this thesis, we investigate several random impurity induced problems in semiconductor devices by using and extending first principles parameter-free method.

For bulk system with RDD, the Kohn-Sham density functional theory (KS-DFT) method can be applied to investigate its electronic properties. We investigated the band gap and band alignment of $GaSb_x N_{1-x}$ (0 < x $\leq 1\%$) alloys, which can be seen as a GaN lattice doped with random Sb impurities. Systems with over 1000 atoms were calculated at the HSE06 hybrid functional level. By calculating the band gap and natural band alignment at different x, we found that with proper amount of Sb doping $GaSb_xN_{1-x}$ should be ideal for photochemical water splitting in solar fuel applications. For open system (e.g., two-probe system) with RDD, we combine the non-equilibrium Green's function density functional theory (NEGF-DFT) method and the non-equilibrium coherent potential approximation (NECPA) theory to calculate disorder scattering and configuration averaging of physical results. The NECPA is a powerful method for solving random disordered systems by analytically averaging the physical properties before a numerical simulation is done. Choosing the linear Muffin-tin orbitals (LMTO) with atomic sphere approximation (ASA) to implement KS-DFT, we simulate carrier transport in a boron-nitrogen (B-N) co-doped graphene Tunnel field effect transistor (TFET). We found that the B-N co-doping opened a substantial gap that linearly scaled with the co-doping concentration which is appropriate for making TFET. We also observed that impurity scattering in the graphene TFET reduces the band-to-band tunneling current by a substantial factor. When the impurity concentration in the open system is low (<1%), we show that the NECPA calculation is simplified and an approximate method for implementing the NECPA

with LCAO basis is presented. Low concentration approximation (LCA) was made that the off-diagonal disorder can be neglected when evaluating the transmission. The validity of this assumption was verified by a tight binding model and the DFT implementation was verified by a boron doped graphene system. Both showed that NECPA-LCA gives acceptable results on transport when the impurity concentration is low. The method was employed to predict the dopant (Cl/Re) limited mobility of monolayer MoS₂.

Résumé

Au cours des cinq dernières décennies, la taille des dispositifs électroniques dans une puce a diminué de manière exponentielle conformément à la loi de Moore. Alors que les appareils électroniques deviennent de plus en plus petits, la réduction progressive de la technologie est de plus en plus mise à mal par de nombreux problèmes émergents. L'un des problèmes les plus importants est la variation d'un dispositif à l'autre provoquée par des dopants discrets aléatoires (RDD) (également appelés impuretés aléatoires). L'effet RDD peut entraîner diverses variations des propriétés du dispositif, telles que les structures de bandes électroniques matérielles et le transport. Dans cette thèse, nous étudions plusieurs problémes aléatoires induits par les impuretés dans les dispositifs à semi-conducteurs en utilisant et en étendant la méthode sans paramètre de premiers principes.

Pour les systèmes en vrac avec RDD, la méthode de la théorie de la densité fonctionnelle Kohn-Sham (KS-DFT) peut être appliquée pour étudier ses propriétés électroniques. Nous avons étudié la bande interdite et l'alignement des alliages $GaSb_xN_{1-x}$ (0 < x $\leq 1\%$), qui peuvent être vus comme un réseau GaN dopé avec des impuretés Sb aléatoires. Les systèmes de plus de 1000 atomes ont été calculés au niveau fonctionnel hybride HSE06. En calculant l'intervalle de bande et l'alignement naturel de bande à différents x, nous avons constaté qu'avec une quantité appropriée de dopage Sb, $GaSb_xN_{1-x}$ devrait être idéal pour le fractionnement photochimique de l'eau dans les applications de carburant solaire. Pour un système ouvert (par exemple, un système à deux sondes) avec RDD, nous combinons la méthode de la théorie de la densité fonctionnelle de la fonction de Green (NEGF-DFT) non-équilibrée à la théorie de l'approximation du potentiel cohérent non-à l'équilibre (NECPA) pour calculer la dispersion des désordres et la moyenne de configuration des résultats physiques. La NECPA est une méthode puissante pour résoudre des systèmes désordonnés aléatoires en faisant la movenne analytique des propriétés physiques avant la simulation numérique. En choisissant les orbitales linéaires Muffin-étain (LMTO) avec approximation de sphère atomique (ASA) pour mettre enœuvre la KS-DFT, nous simulons le transporteur dans un transistor à effet de champ (TFET) graphène tunnel dopé graphène bore-azote (B-N). Nous avons constaté que le co-dopage B-N ouvrait un espace important qui diminuait de manière linéaire avec la concentration de co-dopage appropriée pour la fabrication de TFET. Nous avons également observé que la diffusion des impuretés dans le TFET de graphène réduisait considérablement le courant de tunnel entre bandes. Lorsque la concentration en impuretés dans le système ouvert est faible (<1 %), nous montrons que le calcul NECPA est simplifié et qu'une méthode approximative de mise enœuvre de la NECPA avec LCAO est présentée. Une approximation basse concentration (LCA) a été faite pour que le désordre hors diagonal puisse être négligé lors de l'évaluation de la transmission. La validité de cette hypothèse a été vérifiée par un modèle de liaison étroit et la mise enœuvre de la DFT par un système de graphène dopé au bore. Les deux ont montré que la NECPA-LCA donne des résultats acceptables sur le transport lorsque la concentration en impuretés est faible. La méthode a été utilisée pour prédire la mobilité limitée du dopant (Cl/Re) de la monocouche MoS₂. In this thesis, I apply, extend and further develop state-of-the-art first-principles techniques to investigate electronic devices and materials with random impurities. My original contributions to this work include:

- Applied the KS-DFT theory at the level of hybrid functional to study the the band gap and band alignment of $GaSb_xN_{1-x}$ alloys.
- Applying the NECPA-LMTO-ASA method to investigate the B-N co-doped graphene TFET. The results were presented at the IEEE International Electron Devices Meeting (IEDM), the world's top tier forum for breakthroughs in semiconductor and electronic device technology.
- Extending the NECPA formalism to reduce its complexity at low impurity concentration (NECPA-LCA). We verified the extension by a tight-binding model and its DFT implementation was verified by a boron doped graphene.
- Implemented the NECPA-LCA method into software and applied it to investigate the mobility of Cl/Re doped monolayer MoS₂.

The following lists research articles published during course of this thesis research:

1. Qing Shi, Ying-Chih Chen, Faqrul A. Chowdhury, Zetian Mi, Vincent Michaud-Rioux, and Hong Guo, "Band engineering of GaSbN alloy for solar fuel applications", Physical Review Materials 1.3 (2017): 034602.

2. Qing Shi, L Zhang, Y Zhu, L Liu, M Chan and H Guo, "Atomic disorder scattering in emerging transistors by parameter-free first principle modeling", 2014 IEEE International Electron Devices Meeting (IEDM 2014), p.733-736.

3. Qing Shi, Yu Zhu, Lei Liu and Hong Guo, "Analyzing short channel quantum transport variability from atomistic first principles", Phys. Rev. Applied 3, 064008 (2015).

4. Qing Shi, Fei Liu and Hong Guo, "Dielectric Material for Monolayer Black Phosphorus Transistors: A First-Principles Investigation", 2015 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD 2015), pp. 56 (2015).

5. Fei Liu, Qing Shi, Jian Wang and Hong Guo, "Device Performance Simulations of Multilayer Black Phosphorus Tunneling Transistors", Appl. Phys. Lett., 107, 203501 (2015).

6. Bohloul S, Shi Q, Wolkow RA, Guo H., "Quantum transport in gated danglingbond atomic wires", Nano Lett., 2017, 17 (1), pp 322-327.

7. Lining Zhang, Qing Shi, Zubair Ahmed, Hong Guo, and Mansun Chan, "Atomistic Simulations Verified Electrostatic Model of Monolayer MoS₂ MOSFETs", Electron Device Letters, under review (2019).

8. F. A. Chowdhury, S. M. Sadaf, Q. Shi, Y.-C. Chen, H. Guo, and Z. Mi, "Optically active dilute-antimonide III-nitride nanostructures for optoelectronic devices", Appl. Phys. Lett. 111, 061101 (2017).

9. Chen, Ying-Chih, Jing-Zhe Chen, Vincent Michaud-Rioux, Qing Shi, and Hong Guo. "Efficient evaluation of nonlocal operators in density functional theory", Physical Review B 97, no. 7 (2018): 075139.

Acknowledgments

First and foremost, I would like to thank my supervisor, Professor Hong Guo, for giving me the opportunity to work in this group and giving me such an exciting research project. In addition to providing guidance to my project, he has been a great source of inspiration for me by his extensive knowledge on both physics and engineering. Doing research under his supervision has been a truly rewarding experience.

I would also like to thank Prof. Aashish Clerk, Prof. Bill Coish and Prof. Tami Pereg-Barnea of the McGill physics department, from whom I have learned important and useful techniques on condensed matter physics.

Next, I would like to thank Dr. Y.C. Chen, Dr. Vincent Michaud-Rioux, Mr. Chenyi Zhou and Mr. Raphaël Prentki from our group, Dr. Eric Zhu and Dr. Lei Liu from Nanoacademic and Dr. Paul Zhang from Hong Kong University of Science and Technology for their continuous help on my research project. Discussing with them not only improves my techniques but also broadens my horizons on physics and electrical engineering.

Last but not least, I would like to thank my parents, for their great love and support on me. Without their support, I cannot survive my Ph.D research in McGill.

Physical Constants and Units

=	$10^{-10} {\rm m}$
=	$0.5292~{\rm \AA}$
=	$9.1096 \times 10^{-31} \text{ kg}$
=	$1.6726 \times 10^{-27} \text{ kg}$
=	$1.6 \times 10^{-19} {\rm C}$
=	$6.626\times 10^{-34}~{\rm J\cdot s}$
=	$1.38\times10^{-23}~{\rm K}$
=	$0.026~{\rm eV}$
=	$2.9979\times 10^8~{\rm m/s}$
=	$7.75 \times 10^{-5} \Omega^{-1} = \frac{1}{12.9 \mathrm{k} \Omega}$

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	=	$0.5292 \ { m \AA}$
1 unit of Mass	=	m_e	=	$9.1096 \times 10^{-31} \mathrm{kg}$
1 unit of Charge	=	e	=	$1.6 \times 10^{-19} {\rm C}$
1 unit of Angular momentum	=	\hbar	=	1.0546 $\times 10^{-34}~\mathrm{J\cdot s}$
1 unit of Energy	=	1 Hartree	=	$27.2~{\rm eV}$
1 unit of Time	=	$\frac{\hbar}{1 \text{ Hartree}}$	=	$2.4189 \times 10^{-17} s$

List of Abbreviations

- ASA Atomic Sphere Approximation
- BAC Band Anti-Crossing
- BEB Blackman-Esterling-Berk
- CPA Coherent Potential Approximation
- DFT Density Functional Theory
- GGA Generalized Gradient Approximation
- HK Hohenberg-Kohn
- HSE Heyd-Scuseria-Ernzerhof
- KS Kohn-Sham
- LCA Low Concentration Approximation
- LCAO Linear Combination of Atomic Orbitals
- LDA Local Density Approximation
- LMTO Linear Muffin-Tin Orbitals
- NBA Natural Band Alignment
- NECPA Non-Equilibrium Coherent Potential Approximation
- NEGF Non-Equilibrium Green's Function
- NVC Non-Equilibrium Vertex Correction
- RDD Random Discrete Dopants
- SC Self-Consistent
- SQS Special Quasirandom Structures
- SSA Single Site Approximation
- TB Tight-Binding
- VCA Virtual Crystal Approximation
- XC Exchange-Correlation

1

Introduction

In the past five decades, the size of electronic devices in a computer chip has shrunk exponentially according to the Moore's law—which states that the number of transistors on integrated circuit chips doubles approximately every 18 months [1]. As electronic devices become smaller and smaller, the continual down-scaling of the technology is increasingly challenged by many emerging problems. These include the channel length modulation effect, drain induced barrier lowering (DIBL), punch through effect, field dependent mobility, velocity saturation, oxide breakdown, device reliability, etc. These effects have detrimental influence on device performance [2, 3, 4, 5, 6, 7]. One of the most important issues is the device-to-device variation caused by random discrete dopants (RDD) (also known as random impurities). The RDD effect can render a series of fluctuations to device properties including threshold voltage, which is the voltage needed to turn on a transistor, and sub-threshold slope which is the gate voltage increment when threshold current increases by one order of magnitude [8, 9, 10]. In general, RDD fluctuation is induced by some randomness in the dopant concentration and/or variation of the dopant locations in the device material. It has been reported that RDD has become a key challenge in modern integrated circuits industry [11, 12]. Investigating RDD effects is highly important not only to understand its influence on device characteristics, but also to controll it. Random impurities have also been used in band engineering, for instance the $GaSb_xN_{1-x}$ alloy is a promising material for electrochemical water splitting application [13], where randomly doped Sb replaces some N atoms. It outperforms the ZnO based alloys [14] since nitride is chemically stable in water [15]; it is superior to $In_xGa_{1-x}N$ alloy because far fewer Sb impurities are needed than In to produce the correct band gap.

There are substantial literature of theoretical work on random impurity effects in electronic devices. These works can be classified into two kinds: those based on parameterized model calculations and those based on first principles calculations. Take RDD effects in electronic devices as an example, for model calculations, a classical drift-diffusion approach is applied to determine the transport current, in which dopant atoms are simulated by some scattering potentials which are considered a perturbation to the potential of the intrinsic channel material [16, 17, 18]. In first principles methods, devices are treated as a large group of atoms and everything is calculated from quantum mechanics [19, 20, 21]. In parameterized model analysis, the computation complexity is low hence a large number of samples can be simulated to obtain an ensemble average of results. But due to lack of quantum effects, there are discrepancies between the simulation results and experimental data particularly for devices with $\sim 10 \,\mathrm{nm}$ size or smaller [17]. In first principles methods, the simulated results are more reliable but due to the high computational cost, the maximum number of simulated atoms is very limited. Thus, for instance, to simulate a dopant concentration of 0.1%, one must calculate 1000 atoms in order to accommodate just one dopant atom. Since this single dopant can locate at 1000 possible or even more locations, the calculation becomes prohibitive for first principles. This is why first principles calculations typically use very large dopant concentrations—much larger than that in real devices. Even with such unrealistic doping concentrations, one can only calculate at most a few disorder configurations to obtain the average. As a result, RDD induced variation has not been investigated to satisfaction from quantum mechanical first principles.

In this thesis, we apply and extend a newly developed first principles method that overcomes the above technical difficulties to investigate the influence of random impurities. Our technique is based on carrying out the Kohn Sham density functional

theory (KS-DFT) within the Keldysh non-equilibrium Green's function formalism (NEGF), where the disorder scattering of impurity atoms is simplified by the coherent potential approximation (CPA) theory at the single Green's function level, and by the non-equilibrium CPA (NECPA) theory at the two-Green's function correlator level. Our method allows us to predict the performance of devices with random disorder impurities from atomic first principles without any phenomenological parameter accurately and efficiently. NEGF-DFT is currently the most powerful method to calculate non-equilibrium quantum transport properties of nanostructures. A simpler version of this approach is to replace KS-DFT by a tight-binding (TB) Hamiltonian [22] that removes the difficult self-consistent calculation of the equilibrium part of the Hamiltonian and focuses only on solving the potential induced by non-equilibrium effects (e.g. external bias). The NEGF-TB technique was applied to describe many quantum transport problems and achieved success [23, 24, 25, 26]. Nevertheless, the TB parameters often cannot accurately account for materials properties, especially when dealing with interfaces between different materials. KS-DFT is a technique that is widely used in materials physics to predict mechanical and electronic properties and is, in principle, parameter-free. Traditionally, KS-DFT was applied to systems with periodic boundary conditions such as crystals or to finite systems such as isolated molecules, but not to open systems such as the transport junctions. Due to current flow in devices, transport problems are intrinsically non-equilibrium which cannot be solely described by KS-DFT. In 2001, Taylor, Wang and Guo [27] established a first principles technique that combined NEGF and KS-DFT to study non-equilibrium quantum transport problems of nano-scale devices. In the NEGF-DFT formalism, the density matrix that enters the KS-DFT is constructed at non-equilibrium by NEGF. as such the meaning of KS-DFT in the NEGF-DFT formalism is no-longer the usual ground state KS-DFT. Fig. 1.1 shows a simple sketch of a typical "two-probe" device structure analyzed by NEGF-DFT. It consists of a central region (in rectangular box) and two semi-infinite leads (the blue parts outside the box). Parts of the leads, called buffer layers, are included in the simulation box to ensure a smooth match of



simulation box

Figure 1.1: Diagram of a simulated device in NEGF-DFT method. This is also called a "two-probe" model. It consists of two semi-infinite leads (left lead and right lead outside the simulation box) and a central region (in rectangular box). Parts of the leads, called buffer layers are included in the simulation box to ensure a smooth match of the potential at the edges of the simulation box. The grey part in the center is the device part (also called "scattering region"). The simulation is periodically repeated in the cross-section of the device.

the potential at the edges of the simulation box. The grey part in the center is the device part (also called "scattering region"). This technique has been used to study a wide range of transport problems in molecular transport junctions [28, 29, 30], carbon nanotubes [31], tunneling junctions [32, 33], spin transport [34], heterojunctions [35], metal wires [36], etc. In this thesis, NEGF-DFT will be applied to investigate transport of boron-nitrogen (B-N) doped graphene Tunnel-FET (TFET) and mobility of chlorine (Cl) or rhenium (Re) doped MoS₂.

To handle disorder scattering due to RDD, we shall apply the coherent potential approximation (CPA) and non-equilibrium vertex correction (NVC) formalism developed by Ke, Xia and Guo [37] in 2008 or its mathematically equivalent - the non-equilibrium coherent potential approximation (NECPA) formalism developed by Zhu, Liu and Guo [38] in 2013. This technique allows one to deal with non-equilibrium transport problems of disordered systems. CPA is a successful and widely used technique for studying random substitutional disorders in solids [39, 40]. It is usually implemented with Korringa-Kohn-Rostoker (KKR) [41], a method for calculating electronic structures, and linear Muffin Tin orbital (LMTO) [42, 43, 44] method with atomic sphere approximation (ASA) of KS-DFT. In CPA, an effective medium is constructed self-consistently to determine the configuration averaged physical properties of random systems. CPA drastically reduces the computational complexity for studying disordered system. To simulate a disordered system having N possible

disorder configurations, in principle one has to carry out N calculations to perform the disorder configuration averaging. In CPA, one derives an averaged formula for physical quantities, averaged density matrix, averaged Hamiltonian, averaged conductance, etc., and evaluates them only once. Traditionally, CPA is only applicable to equilibrium problems [45, 46]. The extension of CPA to solve non-equilibrium problems such as quantum transport, is achieved by a further formulation of the NVC [37] or NECPA theory [38] which allows one to calculate the configuration average of the non-equilibrium density matrix (e.g. NEGF), transmission coefficients and other quantities (details in later parts of the thesis). To date, the CPA-NVC and NECPA formalism have produced many exciting results [36, 47, 48, 49, 50, 51]. Finally, a drawback of these CPA/NECPA implementations is that they have requirements on the lattice structures. For example, the LMTO-ASA is solely suitable for close-packed lattices. For non close-packed structures, vacuum spheres in space are used to make the structure close-packed (see Fig. 7.1 in Chapter 7 for ASA of graphene and monolayer black phosphorus). Such complicated vacuum space filling must be worked out for each problem. It is highly demanded the CPA/NECPA be implemented with the linear combination of atomic orbitals (LCAO) basis which requires no vacuum space filling. The difficulty comes from the fact that the Hamiltonian has off-site disorders in LCAO unlike in KKR and LMTO-ASA, where disorders only appear on-site. There have been several attempts to overcome this difficulty. One of the most important attempts is the Blackman-Esterling-Berk (BEB) method [52]. In BEB, the off-diagonal disorders are projected onto a higher dimensional Hamiltonian space so that they occupy only the diagonal part of the new Hamiltonian. However, it appears very difficult to reduce the BEB theory to calculate transport. In this work, we propose the NECPA-LCA (low concentration approximation) that has been implemented with LCAO basis. It is verified that the off-site disorder under LCAO can be neglected

when the impurity concentration is low, which is the case for most semiconductor problems. CPA/NECPA applies well to cases where lattice local distortion caused by impurities can be neglected. This is satisfied by most cases where impurity and host atoms have similar radii, such as boron in graphene. When impurity atom is much larger/smaller than host atom (e.g., Sb in GaN), lattice distortion around impurities plays an important role in the modification of the band structure. In these case, KS-DFT should be applied with the special quasirandom structures (SQS) method [53] which we shall use. This method gives the best periodic supercell that approximates the true disordered state for a given number of atoms per supercell.

In this thesis, we start by investigating impurity effect of a semiconductor problem, the dilute doping by Sb atoms to GaN, using highly accurate KS-DFT at the hybrid exchange-correlation potential level where over 1000 atoms are calculated to directly predict band bowing due to Sb doping. The KS-DFT theory will be introduced in Chapter 2 and its application to $GaSb_xN_{1-x}$ will be detailed in Chapter 3. Then, for transport at non-equilibrium, such direct calculation cannot be carried out and approximate schemes are introduced, including the CPA theory, the NVC theory, and the NECPA theory. These theories were most easily implementable - as done in the past, in site-oriented DFT methods such as the LMTO approach [54], which we shall apply for a number of investigations. Related theories and implementations will be presented in Chapter 4 and 5 followed by Chapter 6 where we apply them to investigate the doped graphene TFET. For semiconductor devices where the impurity concentration is typically low, we shall show in Chapter 7 that these equilibrium and non-equilibrium CPA methods can be extended to DFT methods based on LCAO, and we achieved such a technical advance (the NECPA-LCA method) which proved to be useful in practical semiconductor simulations. Finally, in Chapter 8 a brief summary of this thesis will be presented and future works suggested.

 $\mathbf{2}$

Density Functional Theory

The topic of this thesis concerns the parameter-free transport simulation of materials that has random disorders inside. The formalism that allows us to perform such calculations is built on the foundation of two mature theories, namely density functional theory (DFT) and non-equilibrium Green's functions (NEGF) theory. Within this transport formalism, DFT is responsible for constructing the material Hamiltonian and density of states, whereas NEGF describes how the electronic states are populated via non-equilibrium quantum statistics. Based on the two, the non-equilibrium coherent potential approximation deals with the random disorder scattering by finding the Green's function of an effective medium of the disordered material. The subject of this chapter is to present DFT.

DFT is a successful quantum theory for predicting mechanical and electronic properties of materials widely used in condensed matter physics, material science, computational physics and quantum chemistry [55, 56, 57, 58]. We employ DFT to construct the Hamiltonian of the material at equilibrium. This chapter is devoted to present a brief but comprehensive introduction of DFT and some of its implementation such as the Linear Combination of Atomic Orbitals (LCAO) and Linear Muffin Tin Orbital (LMTO) implementations. The contents of this chapter are outlined as follows. Section 2.1 is devoted to presenting the basic theories of DFT. Section 2.2 presents the LCAO basis and its implementation. Section 2.3 will be devoted to an introduction of LMTO basis and implementation. Finally, section 2.4 is a brief summary of this chapter. From now on, atomic units are applied to all derivations and equations throughout the thesis unless otherwise specified.

2.1 Basic Theorems and Assumptions of DFT

In general, quantum chemistry of materials requires solving a many body Schrödinger equation which consists of many nucleis and even more electrons. This is a complicated problem and its numerical solution is prohibitive for any realistic materials according to today's computation power. By employing the Born-Oppenheimer approximation, the problem is reduced to a "much simpler" electron Schrödinger equation due to the fact that the motion of nuclei is way slower than that of electrons [59]. In this section we present some basic but important theories related to solving the electron Schrödinger equation. Section 2.1.1 will discuss the Born-Oppenheimer approximation which separate the dynamics of the ions and electrons. Section 2.1.2 will be devoted to the Hohenberg-Kohn (HK) theorem [60]—the most basic theorem of DFT. It simplifies the N-interacting electrons many body problem into a variation problem expressed by the ground state single particle density $\rho(\mathbf{r})$, where **r** denotes the real space position. Section 2.1.3 is dedicated to the Kohn-Sham (KS) equations [61] which provide a feasible method to solve N-interacting electrons problem numerically by a "mean field"-like approach. Section 2.1.4 is an introduction to the exchange-correlation (XC) functionals used in the projects presented in this thesis. Finally, Section 2.1.5 presents a brief description and framework on how DFT works in numerical implementations.

2.1.1 Born-Oppenheimer Approximation

Born-Oppenheimer approximation is the most basic approximation to solve multiparticle problems. It was first introduced to solve time-dependent Schrödinger equation but it also gives hint on how to solve time-independent one. We start from the Schrödinger equation of the many particle system:

$$\left[\sum_{i} -\frac{1}{2}\nabla_{i}^{2} + \sum_{I} -\frac{1}{2M_{I}}\nabla_{I}^{2} + \sum_{i>j} \frac{1}{|\mathbf{r}_{i} - \mathbf{r}_{j}|} - \sum_{i,I} \frac{Z_{I}}{|\mathbf{r}_{i} - \mathbf{R}_{I}|} + \sum_{I>J} \frac{Z_{I}Z_{J}}{|\mathbf{R}_{I} - \mathbf{R}_{J}|} - E\right] \Psi_{all}(\{\mathbf{r}_{i}\}, \{\mathbf{R}_{I}\}) = 0,$$
(2.1)

where five summation terms on the left side are the electronic kinetic energy, the nuclear kinetic energy, the electronic coulomb interaction energy, the electron-nucleus coulomb interaction energy and the nuclear coulomb interaction energy, respectively. We have used the lowercase letters for electron indices and position vectors, and uppercase letters for nucleus. M_I denotes the mass of the nucleus and $Z_{I/J}$ denotes the nuclear charge. $\mathbf{r}_{i/j}$ and $\mathbf{R}_{I/J}$ are the electron and nucleus positions in space. Ψ_{all} refers to the system wave function (including both electrons and nuclei) and E the system eigen energy. The lowest eigen energy E is called the ground state energy. As will be shown in the following subsection, the system ground state plays a significant role in DFT.

In any solid state material, both the ions and electrons will move rapidly from a classical point of view: ions vibrate around their equilibrium positions and electrons move around their corresponding ions. Since ions have much larger masses than electrons (more than 1000 times), they tend to move much slower than electrons, i.e., in a dynamical sense, the electrons can be regarded as particles that follow the nuclear motion adiabatically, meaning that they are "dragged" along with the nuclei without requiring a finite relaxation time. This picture arises the Born-Oppenheimer ansatz, it is assumed that ion motion is much slower than electron motion that they can be considered as fixed in space when one solves the electron motion. This way the electrons can be treated as to move in an external potential generated by the "frozen" ions and the system ground state wave function Ψ_{all} is able to be factorized

into an electronic part and a nuclear part:

$$\Psi_{all}(\{\mathbf{r}_i\}, \{\mathbf{R}_I\}) = \Psi_e(\{\mathbf{r}_i\}, \{\mathbf{R}_I\})\Psi_c(\{\mathbf{R}_I\}), \qquad (2.2)$$

where we have used the subscript e(c) to denote the electronic (nuclear) part. The electronic part Ψ_e depends on both the electronic spatial coordinates and the nuclear positions and the nuclear part Ψ_c depends only on the nuclear positions. For the sake of simplicity, we rewrite the original Schrödinger equation (2.1) into a more compact form by defining some symbols for the five terms in Eq. (2.1):

$$[T_e + T_n + U_{ee}(\{\mathbf{r}_i\}) + U_{en}(\{\mathbf{r}_i\}, \{\mathbf{R}_I\}) + U_{nn}(\{\mathbf{R}_I\}) - E]$$

$$\Psi_e(\{\mathbf{r}_i\}, \{\mathbf{R}_I\})\Psi_c(\{\mathbf{R}_I\}) = 0.$$
(2.3)

This equation is still too complicated to solve since the coordinates of electrons and nuclei couple throughout the equation. One useful trick to simplify the problem is to separate $\{\mathbf{r}_I\}$ and $\{\mathbf{R}_I\}$ as much as possible. Eq. (2.3) can be grouped into the following form

$$\frac{[T_e + U_{ee}(\{\mathbf{r}_i\}) + U_{en}(\{\mathbf{r}_i\}, \{\mathbf{R}_I\})]\Psi_e(\{\mathbf{r}_i\}, \{\mathbf{R}_I\})}{\Psi_e(\{\mathbf{r}_i\}, \{\mathbf{R}_I\})} = E - \frac{[T_n + U_{nn}(\{\mathbf{R}_I\})]\Psi_c(\{\mathbf{R}_I\})}{\Psi_c(\{\mathbf{R}_I\})},$$
(2.4)

where the approximation that the nuclear wave function Ψ_c is more localized than the electronic wave function Ψ_e

$$\left|\nabla_{I}\Psi_{c}\right| \gg \left|\nabla_{I}\Psi_{e}\right| \tag{2.5}$$

is applied. This relation originates from the mass difference between electrons and nuclei. As mentioned in previous paragraph, nuclei are several thousand times heavier than electrons, therefore the nuclear components of the wave function are more localized in space than the electronic component of the wave function. Considering this in the classical limit, the nuclei are fully localized about single points representing classical point particles. Thus, it follows that the nuclear wave function rises more steeply than the electronic wave function, which is exactly what is implied in Eq. (2.5). The left hand side of Eq. (2.4), the electronic ground state eigenvalue, must be of { \mathbf{R}_I } since the right hand side is a function of { \mathbf{R}_I } only. Let it be E_e { \mathbf{R}_I }, the system Schrödinger equation can be separated into two equations:

$$[T_e + U_{ee}(\{\mathbf{r}_i\}) + U_{en}(\{\mathbf{r}_i\}, \{\mathbf{R}_I\})] \Psi_e(\{\mathbf{r}_i\}, \{\mathbf{R}_I\}) = E_e\{\mathbf{R}_I\} \Psi_e(\{\mathbf{r}_i\}, \{\mathbf{R}_I\}), \quad (2.6)$$

$$[T_n + U_{nn}(\{\mathbf{R}_I\}) + E_e\{\mathbf{R}_I\}] \Psi_c(\{\mathbf{R}_I\}) = E\Psi_c(\{\mathbf{R}_I\}), \qquad (2.7)$$

Eq. (2.6) gives a set of electronic states $\{\Psi_e^n(\{\mathbf{r}_i\}, \{\mathbf{R}_I\})\}\$ and corresponding eigenvalues $\{E_e^n\{\mathbf{R}_I\}\}\$. n indices the solution of Eq. (2.6). $\{E_e^n\{\mathbf{R}_I\} + U_{ee}\{\mathbf{R}_I\}\}\$ depend parametrically on the nuclear positions $\{\mathbf{R}_I\}\$ and are called Born-Oppenheimer energy surfaces. The Born-Oppenheimer approximation is invalid when the Born-Oppenheimer energy surfaces are coupled, e.g., the system reaches a nuclear configuration where the energy difference between two energy surfaces $|E_e^i\{\mathbf{R}_I\} - E_e^j\{\mathbf{R}_I\}|\$ is close to or smaller than the thermal energy k_BT , where k_B is the Boltzmann constant and T is the system temperature.

To sum up, Born-Oppenheimer approximation is reasonable for most situations where electrons follow the nuclei motion adiabatically. This is ensured in most situations by the fact that nucleus has much larger mass than the electron. For the sake completeness, we point out that in certain chemical processes involving light elements, Born-Oppenheimer approximation can break down and the dynamics of both ions and electrons must be treated on equal footing [59]. We do not consider such situations throughout this thesis. In the following subsections, we focus on the electronic Schrödinger equation (Eq. 2.6). The subscripts e for Ψ_e and E_e are omitted for the sake of simplicity. Indices \mathbf{r}_i and \mathbf{R}_I are neglected when not necessary.

2.1.2 Hohenberg-Kohn Theorem

In previous subsection, we introduced the Born-Oppenheimer approximation which reduces the nuclear degrees of freedom as parameters in the electronic Schrödinger equation. While it brings a significant reduction in complexity, the many-electron problem still remains unapproachable, especially for large system calculation, e.g., real materials computation. In 1964, Hohenberg and Kohn proposed the Hohenberg-Kohn (HK) theorem which further reduces the complexity of the problem [60]. HK theorem reduces the fully interacting many-electron problem to determining the ground state single particle density $\rho(\mathbf{r})$,

$$\rho(\mathbf{r}) = N \prod_{i=2\dots N} \int d\mathbf{r}_i \Psi^*(\{\mathbf{r}_i\}) \Psi(\{\mathbf{r}_i\}).$$
(2.8)

where N is the number of electrons and $\Psi(\mathbf{r})$ is the electronic wave function. HK theorem states that the non-degenerate electronic ground state energy (E) of an Nelectron system is a unique functional of the system ground state electron density $\rho(\mathbf{r})$,

$$E = E\left[\rho\left(\mathbf{r}\right)\right].\tag{2.9}$$

As proven by Hohenberg and Kohn, all properties of the system, including excited state properties, are in principle exact functionals of the ground state electronic density. The reason for this is that there is a one-to-one mapping between the ground state density and the external potential. A simple proof to this theorem is as follows. For a system with N interacting electrons, the Hamiltonian is given as

$$H = \sum_{i=1}^{N} \frac{-\nabla_i^2}{2} + \sum_{i>j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i=1}^{N} V_{ext}(\mathbf{r}_i), \qquad (2.10)$$

where ∇_i is the gradient operator to the i^{th} electron wave function and $V_{ext}(\mathbf{r})$ refers to the external potential generated by the nuclear ions. It is obvious that the system is uniquely determined by the number of electrons N and the external potential $V_{ext}(\mathbf{r})$. Moreover, electron number N can also be determined by the electron density $\rho(\mathbf{r})$ via $N = \int d\mathbf{r}\rho(\mathbf{r})$. Therefore the proof to this theorem is to prove that the ground state electron density $\rho(\mathbf{r})$ can uniquely and universally determine the external potential $V_{ext}(\mathbf{r})$. If this claim is proven, physical properties including the total energy E and wave function $\Psi(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)$ can be determined uniquely and universally by the electron density $\rho(\mathbf{r})$.

Suppose we can have two different external potentials denoted as $V_{ext}(\mathbf{r})$ and $V'_{ext}(\mathbf{r})$ corresponding to the same electron density $\rho(\mathbf{r})$. These two different potentials have their corresponding energies E and E', Hamiltonians H and H' as well as wave functions Ψ and Ψ' , respectively. Note E is for non-degenerate ground state, we have

$$E = \langle \Psi | H | \Psi \rangle$$

$$< \langle \Psi' | H | \Psi' \rangle$$

$$= \langle \Psi' | H' | \Psi' \rangle + \langle \Psi' | H - H' | \Psi' \rangle$$

$$= E' + \frac{1}{N} \sum_{i}^{N} \int d\mathbf{r}_{i} \rho (\mathbf{r}_{i}) (V_{ext} (\mathbf{r}_{i}) - V'_{ext} (\mathbf{r}_{i}))$$

$$= E' + \int d\mathbf{r} \rho (\mathbf{r}) (V_{ext} (\mathbf{r}) - V'_{ext} (\mathbf{r})), \qquad (2.11)$$

where the last equal sign has used the fact that all the $\rho(\mathbf{r}_i)$'s corresponds to the same electronic density so that the N terms in the summation equal each other. We also have a similar inequality for E',

$$E' = \langle \Psi' | H' | \Psi' \rangle$$

$$< \langle \Psi | H' | \Psi \rangle$$

$$= \langle \Psi | H | \Psi \rangle + \langle \Psi | H' - H | \Psi \rangle$$

$$= E + \int d\mathbf{r} \rho (\mathbf{r}) (V'_{ext} (\mathbf{r}) - V_{ext} (\mathbf{r})). \qquad (2.12)$$

From these two inequalities we arrive at E + E' < E' + E, a contradiction. Therefore,

for non-degenerate ground state, there can be only one external potential $V_{ext}(\mathbf{r})$ corresponding to the ground state electron density $\rho(\mathbf{r})$. Thus the HK theorem is proven for a non-degenerate ground state case. In fact the HK theorem is also correct for the case of degenerate ground state as shown in Ref. [62].

From the above discussion, we see that the ground state electronic density uniquely and universally determines all properties of an *N*-electron system. This is also true for the electronic total energy $E[\rho(\mathbf{r})]$. Therefore we can obtain all the properties of a system by minimizing a unique, universal energy functional $E[\rho(\mathbf{r})]$ which is the central spirit of DFT. Usually the total energy functional has the form of

$$E\left[\rho\left(\mathbf{r}\right)\right] = T\left[\rho\left(\mathbf{r}\right)\right] + \frac{1}{2}\int d\mathbf{r}d\mathbf{r}'\frac{\rho\left(\mathbf{r}\right)\rho\left(\mathbf{r}'\right)}{|\mathbf{r} - \mathbf{r}'|} + \int d\mathbf{r}\rho\left(\mathbf{r}\right)V_{ext}\left(\mathbf{r}\right) + \tilde{E}_{XC}\left[\rho\left(\mathbf{r}\right)\right].$$
 (2.13)

The first term refers to the kinetic energy, the second term is the classical electronelectron Coulomb interaction energy (or the Hartree energy) and the third term is the electric energy due to the electron-nucleus interaction. The last term is a non-classical term including the exchange energy (X) and the correlation energy (C). The exchange energy refers to the quantum effect between identical particles and is known as the Pauli repulsion effect for electrons. The correlation effect is a result of the collective behavior of electrons to screen and decrease the Coulombic interaction. More details about the $\tilde{E}_{XC}[\rho(\mathbf{r})]$ term are described in Section 2.1.4. Up to this point, it seems that our problem can be solved by simply minimizing the total energy using Eq. (2.13). Unfortunately, the exact expression for the kinetic energy functional $T[\rho(\mathbf{r})]$ and $\tilde{E}_{XC}[\rho(\mathbf{r})]$ are generally unknown for the many body Coulomb problem and the numerical minimization process of Eq. (2.13) is hard to achieve. Therefore some further approximations are needed.
2.1.3 Kohn-Sham Equation

Eq. (2.13) indicates that the original many-electron problem is equivalent to minimizing the energy functional $E[\rho(\mathbf{r})]$. Kohn and Sham proposed a simple way to achieve this in 1965 [61] by mapping the problem of interacting electrons onto a fictitious system of non-interacting "electrons". By introducing a series of auxiliary wave functions $\{\varphi_i\}$, the electron density can be formulated as

$$\rho\left(\mathbf{r}\right) = \sum_{i}^{N} |\varphi_{i}\left(\mathbf{r}\right)|^{2}, \qquad (2.14)$$

where N denotes the number of electrons. We should mention that these auxiliary wave functions satisfy $\langle \varphi_i | \varphi_j \rangle = \delta_{ij}$ if they are orthonormal. Now it is clear that the total energy in Eq. (2.13) is a functional of auxiliary wave function $\{\varphi_i\}$. To minimize the total energy functional $E[\rho(\mathbf{r})]$ under the constraint $\langle \varphi_i | \varphi_j \rangle = \delta_{ij}$, we apply the variational principle. This is done by the Lagrange multiplier method. By introducing the Lagrange multiplier ε_i , we arrive at the condition for minimizing Eq. (2.13)

$$\frac{\delta E\left[\rho\left(\mathbf{r}\right)\right]}{\delta\varphi_{i}} - \varepsilon_{i}\varphi_{i} = 0.$$
(2.15)

After inserting Eqs. (2.13) and (2.14), we obtain:

$$\left(\frac{-\nabla^2}{2} + \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + V_{ext}(\mathbf{r}) + V_{XC}(\mathbf{r})\right) \varphi_i(\mathbf{r}) = \varepsilon_i \varphi_i(\mathbf{r}), \qquad (2.16)$$

which is the well-known Kohn-Sham equation. Here we have absorbed the difference between the exact kinetic energy $T[\rho(\mathbf{r})]$ and the sum of single-particle kinetic energy into the \tilde{E}_{XC} and denote it as E_{XC} . $V_{XC}(\mathbf{r})$ is defined as

$$V_{XC}\left(\mathbf{r}\right) = \frac{\delta E_{XC}\left[\rho\left(\mathbf{r}\right)\right]}{\delta\rho\left(\mathbf{r}\right)}.$$
(2.17)

Introducing

$$V_{eff}\left(\mathbf{r}\right) = \int d\mathbf{r}' \frac{\rho\left(\mathbf{r}'\right)}{\left|\mathbf{r} - \mathbf{r}'\right|} + V_{ext}\left(\mathbf{r}\right) + V_{XC}\left(\mathbf{r}\right), \qquad (2.18)$$

Eq. (2.16) can be written in a simpler form

$$\left(\frac{-\nabla^2}{2} + V_{eff}\left(\mathbf{r}\right)\right)\varphi_i\left(\mathbf{r}\right) = \varepsilon_i\varphi_i\left(\mathbf{r}\right).$$
(2.19)

This is a mean field equation. All many-body effects are replaced by a simple mean field. Thus we see that the original many-body problem has been reduced to a single particle non-interacting Schrödinger-like equation. If the exact expression of $V_{eff}(\mathbf{r})$ is worked out, by solving the Eq. (2.19) the energy levels $\{\varepsilon_i\}$ and wave functions $\{\varphi_i\}$ can be obtained. The KS eigen functions $\{\varphi_i\}$ are introduced as auxiliary functions that help us to solve the original many-electron Schrödinger problem. Their physical meaning is not clear but in general literature of DFT and what follows it is standard practice to assume they represent real quasiparticles eigen functions [63]. We shall proceed to interpret the eigenvalues and eigen functions as those of real quasiparticles in the rest of the thesis.

2.1.4 Exchange Correlation Potential

As discussed in the previous section, the only remaining issue for solving the KS Eq. (2.19) is the determination of the exchange-correlation potential of Eq. (2.17). Here we present a brief introduction to the exchange-correlation functionals used in this thesis.

The local density approximation (LDA) [64, 65] and the generalized gradient approximation (GGA) [66] are the most widely used XC functionals to date. In LDA the XC energy is a sole function of the local electron density $\rho(\mathbf{r})$

$$E_{XC}^{LDA} = E_{XC}^{LDA} \left[\rho \left(\mathbf{r} \right) \right]. \tag{2.20}$$

The GGA goes one step further than LDA by considering that the XC energy is a functional of both the local electron density $\rho(\mathbf{r})$ and its gradient in real space $\nabla \rho(\mathbf{r})$

$$E_{XC}^{GGA} = E_{XC}^{GGA} \left[\rho \left(\mathbf{r} \right), \nabla \rho \left(\mathbf{r} \right) \right].$$
(2.21)

Because more information for electron density is provided in GGA, it may give more accurate results than LDA but is slightly more difficult to calculate.

In LDA, the XC energy is expressed as

$$E_{XC}^{LDA}\left[\rho\left(\mathbf{r}\right)\right] = \int d\mathbf{r}\rho\left(\mathbf{r}\right)\varepsilon_{XC}^{LDA}\left[\rho\left(\mathbf{r}\right)\right],\tag{2.22}$$

where $\varepsilon_{XC}^{LDA}[\rho(\mathbf{r})]$ is the density of XC energy in real space. In commonly applied form [67, 68], ε_{XC}^{LDA} is frequently split into an exchange part and a correlation part $\varepsilon_{XC}^{LDA} = \varepsilon_X^{LDA} + \varepsilon_C^{LDA}$. The expression of ε_X^{LDA} is derived from homogeneous electron gas and takes the form

$$\varepsilon_X^{LDA}[\rho(\mathbf{r})] = -\frac{3}{4} \left(\frac{3}{2\pi}\right)^{\frac{2}{3}} \frac{1}{r_s}, \quad r_s = \left(\frac{3}{4\pi\rho(\mathbf{r})}\right)^{\frac{1}{3}}$$
 (2.23)

as proposed by Dirac [69]. The correlation part ε_C^{LDA} is more complicated. There is no exact formula for correlation up to date. Its expression is worked out by fitting parameterized formula to higher level computation results such as quantum Monte Carlo simulation results (the Ceperley-Alder data) [70]. Such parametrizations are widely used in quantum-chemical codes. For example, in 1981, Perdew and Zunger suggested the following formula for ε_C^{LDA} [67]

$$\varepsilon_{C}^{LDA}(r_{s}) = \begin{cases} \frac{\gamma}{1 + \beta_{1} r_{s}^{1/2} + \beta_{2} r_{s}}, & \text{if } r_{s} \ge 1, \\ A \ln r_{s} + B + C r_{s} \ln r_{s} + D r_{s}, & \text{if } r_{s} < 1, \end{cases}$$
(2.24)

where γ , β_1 , β_2 , A, B, C and D are parameters that has to be fitted. This formula is usually named as "PZ81" in many DFT or NEGF-DFT codes. The PZ81 parametrization has several shortcomings, such as an artificial discontinuity of second and higher derivatives at $r_s = 1$. Other forms of LDA are proposed to overcome this shortcoming, e.g., the "VWN" by Vosko, Wilk and Nusair [71], and "PW92" by Perdew and Wang [68]. We have used the PZ81 for the LDA XC function throughout the thsis since it works well with our problems and it is computationally cheap.

In GGA, the spirit is to include local gradient $\nabla \rho(\mathbf{r})$ information in the XC formula so that it describes better the inhomogeneity of the system. Early attempts suggest that a naive expansion in orders of $|\nabla \rho(\mathbf{r})|$ was less accurate than LDA [66]. This can be understood in terms of exact sum rules and constraints on E_{XC} derived from the exchange-hole [66]. It turns out that naive gradient expansion violated some of these constraints while the LDA does not. The reason is that LDA corresponds to a physical system (homogeneous gas) while the gradient expansions correspond to unphysical system. The GGA was introduced by explicitly constructing a functional that met many of the known constraints imposed on E_{XC} [66]. PBE is the most successful version of GGA and is proposed by Perdew, Burke and Ernzerhof in 1996 [72]. We briefly introduce the PBE functional as below.

The XC energy is partitioned into two parts like in LDA

$$E_{XC}^{PBE} = E_X^{PBE} + E_C^{PBE}.$$
(2.25)

PBE exchange energy has the form

$$E_X^{PBE}(\rho, \nabla \rho) = \int d\mathbf{r} \rho \varepsilon_X^{LDA} F_X^{PBE}(s), \qquad (2.26)$$

where

$$F_X^{PBE} = 1 + \frac{\mu s^2}{1 + \frac{\mu s^2}{\kappa}},\tag{2.27}$$

is a dimensionless function. s is the dimensionless density gradient and takes the form

$$s = \frac{|\nabla \rho|}{2(3\pi^2)^{1/3}\rho^{4/3}}.$$
(2.28)

 κ and μ are parameters to be determined from known constraints [66]. The second term of F_X^{PBE} is the density gradient contribution to the exchange energy.

The formula for the correlation part is more complicated and we list them as follows:

$$E_{C}^{PBE}(\rho, \nabla \rho) = \int d\mathbf{r} \left[\varepsilon_{C}^{LDA} + H_{C}^{PBE}(r_{s}, \eta, t) \right],$$

$$H_{C}^{PBE}(r_{s}, \eta, t) = \gamma' \phi^{3} \ln \left[1 + \frac{\beta}{\gamma'} t^{2} \left(\frac{1 + At^{2}}{1 + At^{2} + A^{2}t^{4}} \right) \right],$$

$$A(r_{s}, \eta) = \frac{\beta}{\gamma'} \frac{1}{e^{-\varepsilon_{C}^{LDA}/\gamma' \phi^{3}} - 1},$$

$$\phi(\eta) = \frac{1}{2} \left[(1 + \eta)^{2/3} + (1 - \eta)^{2/3} \right],$$

(2.29)

 H_C^{PBE} denotes the density gradient contribution to the correlation energy. γ' and β are parameters determined from physical constraints [66]. $\eta = (\rho_{\uparrow} - \rho_{\downarrow})/\rho$ is the relative spin polarization. $t = \nabla \rho/2\phi\rho k_s$ is the reduced dimensionless density gradient. $k_s = \sqrt{4k_F/\pi}$ is the Thomas-Fermi screening wave number. $k_F = (3\pi^2\rho)^{1/3}$ is the Fermi wave vector. Compared to other GGA forms, PBE provides an accurate description of the linear response of the uniform electron gas, correct behavior under uniform scaling, and a smoother potential. We note that in the PBE correlation formula the LDA correlation energy is required. It can be taken from PZ81, PW92 or VWN. PBE is used for GGA throughout the thesis unless otherwise specified.

Though LDA and GGA earned a great success in the past decades for predicting atomic structures of condensed phase materials (bond lengths, total energies, etc.), they fail to quantitatively predict band gaps for semiconductors and insulators as well as physical quantities associated with the gaps (e.g. exciton binding energy). Several attempts have been devoted to overcome this issue by making some empirical



Figure 2.1: Comparison of the band gaps calculated by RESCU with HSE06 functional and experimental data. Group V, III-V and II-VI semiconductors are plotted in the figure. Dotted line refers to straight line y = x. From left to right, the materials are InAs, InSb, Ge, GaSb, Si, GaAs, CdSe, InP, CdTe, AlSb, ZnTe, CdS, ZnSe, AlP, AlAs, GaP, ZnS, C and BN.

assumption (e.g., the modified Becke-Johnson semi-local exchange potential [73]) or taking into account part of the XC energy at some higher level (e.g., the hybrid functional [74, 75]). The later is considered to be more reasonably at physics level and one of its most popular member, HSE06 developed by Heyd-Scuseria-Ernzerhof, has been implemented into the DFT code RESCU developed at McGill University [76].

HSE06 is a type of hybrid functional which expresses the complicated XC functional in terms of the pieces from PBE and exact exchange

$$E_{XC} = aE_X^{ex} + (1-a)E_X^{PBE} + E_C^{PBE}, (2.30)$$

where E_X^{ex} is the exact exchange functional. It differs from the exchange functionals in Eqs. (2.23) and (2.26). It is calculated exactly. E_X^{PBE} and E_C^{PBE} are exchange and correlation functionals from PBE. *a* is the mixing parameter which tells the percentage of each functional and is determined by perturbation theory [77]. E_X^{ex} is calculated from $\langle \Psi | \hat{V}_{cou} | \Psi \rangle$, where $V_{cou} = \sum_{i>j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$ is the coulomb operator. In section 2.1.3 we have introduced a series of auxiliary functions $\{\varphi_i\}$ and the system ground state wave function Ψ should be expanded in terms of these single particle functions. Finally the calculation of E_X^{ex} involves calculating many electron repulsion integrals

$$\langle ij|mn\rangle = \int \int \frac{\varphi_i(\mathbf{r})\varphi_j(\mathbf{r})\varphi_m(\mathbf{r}')\varphi_n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r} d^3\mathbf{r}'.$$
 (2.31)

However, the coulomb operator is a non-local operator and it decays very slowly in distance, which makes the computation of the exact exchange very expensive. In HSE06, the coulomb operator is split into short-range (SR) and long-range (LR) components

$$\frac{1}{r} = \frac{\operatorname{erfc}(\omega r)}{r} + \frac{\operatorname{erf}(\omega r)}{r}, \qquad (2.32)$$

where $\operatorname{erf}()$ and $\operatorname{erfc}()$ are the error function and complementary error function. $\operatorname{erfc}(\omega r)/r$ decays very fast with r and ω is the parameter which controls the decay rate of the SR term. For $\omega = 0$, the LR term becomes zero and the SR term is equivalent to the full Coulomb operator. The opposite is true for $\omega \to \infty$. In HSE06, the LR contribution is neglected since it is found that the LR term only contributes less than a few percent to the total exchange energy. Neglecting LR terms in exact exchange gives

$$E_{XC}^{HSE06} = a E_X^{ex,SR}(\omega) - a E_X^{PBE,SR}(\omega) + E_{XC}^{PBE}, \qquad (2.33)$$

where $E_X^{PBE,SR}$ is the SR component of the PBE exchange energy. It is the modification of the original PBE XC functional and details can be found in Ref. [74]. In HSE06, ω is usually set to be 0.15 and *a* is chosen to be 0.25. HSE06 is particularly successful especially in predicting the band gap of semiconductors. In Fig. 2.1 we have plotted the band gaps of some popular group V, III-V and II-VI semiconductors calculated by RESCU [78] with HSE06 functional [74] and compared them with experimental data. The experimental data are extracted from Ref. [79]. We note



Figure 2.2: Flowchart of self-consistent solution procedure in DFT calculation.

that some calculated band gaps are slightly smaller than experimental ones for wide band gap materials due to the fact that we are using the same mixing parameter *a* for all materials. Good agreement has been observed between the two and meanwhile RESCU is superior in computational speed than other DFT codes such as VASP due to the incorporation of the computationally efficient technique described in Ref. [76]. In our later application chapter, RESCU with HSE06 functional will be used to study the band gap engineering of electrochemical materials.

2.1.5 Self-consistent Procedure of DFT

We summarize the DFT self-consistent procedure in Fig. 2.2. The KS Eq. (2.16) tells us that the KS Hamiltonian is solely determined by the electron density. Therefore, given a proper initial guess to the electron density, a series of wave functions are determined by solving the KS equation. In this thesis, the initial guess is chosen to be the neutral electron density $\rho_{NA}(\mathbf{r})$ which is a sum over electron density of isolated atoms that form the material. Then a new electron density is obtained from these KS wave functions via Eq. (2.14). This process is repeated until self-consistency is achieved, namely until results of the present iteration step differ from the previous step by less than a pre-specified numerical tolerance, typically at the 10^{-5} Hartree level for the monitored physical quantities (e.g. elements of density matrix, Hamiltonian matrix, total energy, etc.). which the Hamiltonian (wave function) is expressed in a matrix (vector) form. In this thesis, we employ the atomic centered basis such as the LCAO and LMTO. The reason that we choose atomic centered basis is that the Green's functions are needed in the later chapters (Chapter 4 and 5) when dealing with transport. With atomic centered basis, the Hamiltonian is a sparse matrix so that the Green's function - inverse of Hamiltonian matrix - can be calculated efficiently. The numerical program used in this thesis is implemented by the technical computing platform Matlab.

2.2 Linear Combination of Atomic Orbital Implementation

This section is dedicated to introduce some details of the atomic orbital basis and its implementation with DFT. The choice of the basis determines the matrix size and the sparsity of the Hamiltonian so that a compact and short-ranged basis set can greatly reduce the computational cost. The atomic orbital basis set is one of such basis that are popularly employed in DFT and has been implemented in RESCU [78]. The single particle KS wave function is expanded on the isolated atomic orbital with different angular momentum numbers, magnetic momentum numbers and spins. In the rest of the thesis, we omit the spin index to make the formula look less crowded.

2.2.1 Atomic Orbitals

The linear combination of atomic orbitals (LCAO) method employs a basis that is similar to the wave functions of the single isolated atom, which are the product of a radial part $u_{\mathbf{R}l}(r)$ and an angular part (i.e., spherical harmonics) $Y_L(\hat{\mathbf{r}})$:

$$\phi_{\mathbf{R}L}(\mathbf{r}) = u_{\mathbf{R}l}(r)Y_L(\hat{\mathbf{r}}),\tag{2.34}$$

where **R** designates a particular atom at position **R** and L = (l, m) corresponds to the combination of angular and magnetic quantum numbers. The $\{\phi_{\mathbf{R}L}\}$ orbitals are built with a cutoff radius $r_{\mathbf{R}l}^c$, outside of which the orbital vanishes to zero. The procedure to construct the atomic orbitals will not be discussed here, but the detail can be found in Ref. [80].

The type and number of orbitals one uses depends on the desired accuracy and the chemical nature of the atom. Generally, a basis set includes one or more s, p, d and f orbitals. Since the LCAO orbitals are decent approximations to the real local charge distribution, a small number of orbitals are generally sufficient to achieve reasonable accuracy, making the LCAO basis rather compact. Another advantage is that the orbitals are localized near the atomic centers, which results in sparse block-diagonal matrices. Such matrices can be inverted by efficient numerical procedures.

In DFT codes for calculating crystals, the periodic boundary condition is employed to the lattice unit cell. Therefore it is convenient to consider the Bloch boundary conditions via the periodic basis

$$\phi_{\mathbf{R}l}^{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{T}} \exp(i\mathbf{k} \cdot \mathbf{T}) \phi_{\mathbf{R}l}(\mathbf{r} - \mathbf{T}), \qquad (2.35)$$

where **k** is the Bloch wave vector and **T** is the lattice vector in reciprocal space. $N = \sum_{\mathbf{T}} 1$ is a normalization factor. In a periodic system, the KS states (in Eq. (2.16)) are denoted by the wave vector **k** and can be expanded on basis $\{\phi_{\mathbf{R}l}^{\mathbf{k}}(\mathbf{r})\}$'s as

$$\varphi_i^{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}l} c_{\mathbf{R}l}^{\mathbf{k}i} \phi_{\mathbf{R}l}^{\mathbf{k}}(\mathbf{r}).$$
(2.36)

The sum is taken over all orbitals in the unit cell considered. Eqs. (2.36) and (2.16) together give the matrix form of the KS equation

$$\mathbf{H}^{\mathbf{k}}\mathbf{C}^{\mathbf{k}} = \mathbf{S}^{\mathbf{k}}\mathbf{C}^{\mathbf{k}}\mathbf{E}^{\mathbf{k}} \tag{2.37}$$

where the matrices elements are defined by

$$H_{\mathbf{R}l,\mathbf{R}'l'}^{\mathbf{k}} = \int d\mathbf{r}\phi_{\mathbf{R}l}^{\mathbf{k}*}(\mathbf{r}) \left[-\frac{\nabla^2}{2} + \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + V_{ext}(\mathbf{r}) + V_{XC}(\mathbf{r}) \right] \phi_{\mathbf{R}'l'}^{\mathbf{k}}(\mathbf{r}) \quad (2.38)$$

$$S_{\mathbf{R}l,\mathbf{R}'l'} = \int d\mathbf{r} \phi_{\mathbf{R}l}^{\mathbf{k}*}(\mathbf{r}) \phi_{\mathbf{R}'l'}^{\mathbf{k}}(\mathbf{r})$$
(2.39)

$$C_{\mathbf{R}l,i}^{\mathbf{k}} = c_{\mathbf{R}l}^{\mathbf{k}i} \tag{2.40}$$

$$E_{ij}^{\mathbf{k}} = \delta_{ij}\varepsilon_i^{\mathbf{k}}.$$
(2.41)

Details of calculating the above elements can be found in Ref. [81]. Eq. (2.37) is also called the secular equation. All DFT solvers are based on solving this equation and the difference is the basis on which those matrices are expanded. Once the KS equation (2.37) is obtained, the system is solved by solving the eigen pair problems of matrices $\mathbf{H}^{\mathbf{k}}$ and $\mathbf{S}^{\mathbf{k}}$.

2.2.2 Pseudopotential approximation

In most practical situations, core electrons of the individual atoms rarely participate in chemical bonding and can very often be treated as "frozen". In this case, one can replace the core electrons by an artificial potential that mimics their effect on the valence electrons. This is known as the pseudopotential (pp) approximation. The pp will include, not only the potential associated to the core electrons, but also that of the nucleus. This technique has several advantages: i) avoiding the potential singularities associated with the point-like nuclear charges, ii) excluding the core electrons from the Hamiltonian and overlap matrices thus reducing the complexity of the problem, and iii) replacing the "true" orbitals, which are rapidly oscillating functions near the core, by smooth effective pseudo-orbitals (i.e., the LCAO basis) which reduces the difficulty in numerical solution.

In pp approximation, a cutoff radius r_{core} is applied so that the artificial pp matches the potential obtained from the all-electron calculation at $r > r_{core}$. This guarantees that the pseudo-orbitals (the eigenstates of the pp) are identical to the all-electron eigenstates for $r > r_{core}$. Note that the pseudo-orbitals are nodeless inside r_{core} , which is easier to represent on a real-space grid. In this thesis we employ the Optimized Norm-Conserving Vanderbilt (ONCV) pp [82, 83] when needed. The ONCV is a competitive choice for both accuracy and computational efficiency compared to ultrasoft [84] and projector-augmented-wave (PAW) potentials [85].

2.3 Linear Muffin Tin Orbital Implementation

This section is dedicated to introduce the implementation of KS-DFT in LMTO-ASA framework. In section 2.3.1, we will introduce the atomic sphere approximation (ASA) based on which the Muffin Tin orbitals are constructed. Like LCAO, LMTO is another type of atomic basis but it can make DFT much more efficient due to further approximation. With LMTO the Hamiltonian matrix is sparser than LCAO, so that the computation is faster. In section 2.3.2 we will present the details of constructions of MTO's. Section 2.3.3 is devoted to the linearization of energy. It allows us to reduce the energy dependent orbitals to energy independent ones to reduce the computational complexity. In section 2.3.4, we will present how to construct overlap and Hamiltonian matrices under the framework of LMTO. Then in section 2.3.5 we will introduce the potential parameters through which the Hamiltonian can have a much simpler form.

2.3.1 Atomic Sphere Approximation (ASA)

To introduce the ASA, we shall start from the Muffin Tin potential. The Muffin Tin potential is expressed by a group of spherically symmetric potentials centered at each atomic site. All these spherically symmetric potentials are limited within some fixed radii $(s_{\mathbf{R}})$ without overlapping (see solid spheres in grey in Fig. 2.3). And the remaining interstitial region is chosen to have a constant potential V_0 [86].

In this thesis, the atomic sphere approximation (ASA) [87, 88, 89, 54] for the



Figure 2.3: Diagram for Muffin Tin potentials. The whole space is divided into two regions: the slightly overlapped atomic spheres (showed in dashed spheres) and the in-between interstitial region. In atomic spheres the potentials are chosen to be spherical symmetric while in interstitial region potential is chosen to be a constant V_0 .

potential, is employed. It is developed from the traditional Muffin Tin potential. The basic concept of ASA is to neglect the computationally troublesome interstitial region. This is achieved by increasing the size of the atomic spheres used by Muffin Tin potential or adding new "vacuum" spheres to the interstitial region so that the total volume of all the spheres equals that of the solid cell. The introduction of the vacuum spheres increases the difficulty of the use of the method since the positions as well as their radius must be tuned carefully so that the method gives correct eigen energies. Details will be discussed in Chapter 6. The solid cell is divided into a group of Wigner-Seitz cells so that when calculating physical quantities such as the total energy, we only need to sum up the energies in each Wigner-Seitz cell. Thus the ASA consists of two spirits: i) the use of spherically symmetric potentials inside slightly overlapping atomic spheres (see dashed spheres in Fig. 2.3) centered at individual nuclei, and ii) a complete neglect of the kinetic energy of the interstitial region. It is worth noting that the overlap of the spheres is generally chosen to be around 15%-20% [90]. In ASA, the potential is of the form of [89]

$$V\left(\mathbf{r}\right) = \sum_{\mathbf{R}} V_{\mathbf{R}}\left(\mathbf{r}\right),\tag{2.42}$$

where $V_{\mathbf{R}}(\mathbf{r})$ is spherically symmetric potential inside the **R**-th atomic sphere with a radius of $s_{\mathbf{R}}$. Then the single electron Schrödinger equation can be written as

$$\left[-\frac{\nabla^2}{2} + \sum_{\mathbf{R}} V_{\mathbf{R}}(\mathbf{r}) - E\right] \psi(\mathbf{r}) = 0, \qquad (2.43)$$

where $\psi(\mathbf{r})$ is the single electron wave function. In MTO method, this wave function is expanded as [89]

$$\psi\left(\mathbf{r}\right) = \sum_{i} c_{i} \chi_{i}\left(\mathbf{r}\right),\tag{2.44}$$

where $\{\chi_i(\mathbf{r})\}\$ are the MT orbitals.

The Muffin Tin orbitals are a group of orbitals determined within each atomic sphere. In these atomic spheres, interactions from other atomic spheres are neglected therefore these Muffin Tin orbitals can be solved accurately from a mathematical point of view.

2.3.2 Constructing Muffin Tin orbitals

As written in Eq. (2.44), the single electron wave function is constructed with a group of Muffin Tin orbitals. This section presents in detail how to construct these orbitals.

The Muffin Tin orbitals are determined from the single-site (or single-sphere) problem, where only one atomic sphere is considered in the whole space. To solve the corresponding Schrödinger equation, boundary conditions are needed. These orbitals $\{\chi_{\mathbf{R}}(\mathbf{r})\}$ must satisfy four boundary conditions: i) they must be finite as $r \to 0$, ii) they must be zero as $r \to \infty$, iii) they must be continuous at the spherical boundary $r = s_{\mathbf{R}}$, and iv) the slope of them must be continuous at the spherical boundary $r = s_{\mathbf{R}}.$

We start with the region outside the atomic sphere, i.e. the interstitial region. In ASA, the kinetic energy in the interstitial region is neglected, $E - V_0 = 0$ where V_0 is the constant potential of the interstitial region. The single electron Schrödinger equation in the interstitial region of Fig. 2.3 becomes

$$-\nabla^2\psi\left(\mathbf{r}\right) = 0,\tag{2.45}$$

which is a Laplace equation. Due to its invariance with respect to rotation, $\psi(\mathbf{r})$ should have a form of

$$\psi\left(\mathbf{r}\right) = \sum_{L} a_{l}\left(r\right) Y_{L}\left(\hat{\mathbf{r}}\right)$$
(2.46)

where $\hat{\mathbf{r}} = \mathbf{r}/r$ and the label L = (l, m) labels the angular momentum and its zcomponent quantum numbers like in LCAO. $a_l(r)$ is the radial part of the $\psi(\mathbf{r})$. $Y_L(\hat{\mathbf{r}})$ is the spherical harmonics. After inserting Eq. (2.46) into the Laplace Eq. (2.45), we arrive at the equation for radial wave function,

$$\left[-\frac{\partial^2}{\partial r^2} - \frac{2}{r}\frac{\partial}{\partial r} + \frac{l\left(l+1\right)}{r^2}\right]a_l\left(r\right) = 0.$$
(2.47)

This equation leads to two solutions of the original Laplace equation: the regular solution

$$J_L(\mathbf{r}) = J_l(r) Y_L(\hat{\mathbf{r}}), \quad J_l(r) = \frac{1}{2(2l+1)} \left(\frac{r}{s_{\mathbf{R}}}\right)^l$$
(2.48)

and the irregular solution

$$K_L(\mathbf{r}) = K_l(r) Y_L(\hat{\mathbf{r}}), \quad K_l(r) = \left(\frac{s_{\mathbf{R}}}{r}\right)^{l+1}.$$
 (2.49)

Due to the second boundary condition mentioned at the beginning of this section that $\psi(\mathbf{r})$ should be zero as $r \to \infty$, we only keep the irregular solution $K_L(\mathbf{r})$. Also the irregular solution $K_L(\mathbf{r})$ centered at \mathbf{R} can be expanded in terms of regular solutions

centered at \mathbf{R}' ($\mathbf{R}' \neq \mathbf{R}$) according to

$$K_L(\mathbf{r}_{\mathbf{R}}) = -\sum_{L'} S_{\mathbf{R}L,\mathbf{R}'L'} J_{L'}(\mathbf{r}_{\mathbf{R}'}).$$
(2.50)

Here the subscript $\mathbf{r}_{\mathbf{R}}$ indicates that the position vector is centered at the atomic sphere located at \mathbf{R} and so as to $\mathbf{r}_{\mathbf{R}'}$. $S_{\mathbf{R}L,\mathbf{R}'L'}$ in Eq. (2.50) are called the canonical structure constants. For more details about the structure constants, see Appendix A. With the help of Eq. (2.50), the irregular solutions centered at \mathbf{R} can be expressed by regular solutions centered at \mathbf{R}' . Note that this expansion is only valid when $r_{\mathbf{R}'} < |\mathbf{R} - \mathbf{R}'|$.

Then we consider the solution inside the atomic sphere (see Fig. 2.3 dashed spheres) where the Schrödinger equation is

$$\left[-\frac{\nabla^2}{2} + V_{\mathbf{R}}\left(\mathbf{r}\right) - E\right]\psi_{\mathbf{R}L}\left(\mathbf{r}\right) = 0.$$
(2.51)

Again, the solutions should be spherically symmetric inside the atomic sphere. Therefore it is convenient to break $\psi_{\mathbf{R}L}(\mathbf{r})$ into the radial component and angular component:

$$\psi_{\mathbf{R}L}\left(\mathbf{r}, E\right) = u_{\mathbf{R}l}\left(r, E\right) Y_L\left(\hat{\mathbf{r}}\right) \tag{2.52}$$

Here the energy E is written out explicitly. By inserting this into the Schrödinger Eq. (2.51), we obtain the Schrödinger equation for the radial component

$$\left[-\frac{\partial^2}{\partial r^2} - \frac{2}{r}\frac{\partial}{\partial r} + \frac{l\left(l+1\right)}{r^2} + V_{\mathbf{R}}\left(r\right) - E\right]u_{\mathbf{R}l}\left(r, E\right) = 0$$
(2.53)

For this equation $u_{\mathbf{R}l}(r, E)$ cannot be solved analytically. But we can obtain the solution by analyzing the asymptotic behavior for $r \to 0$. When $r \to 0$, $[V_{\mathbf{R}}(r) - E]$ can be dropped in favour of the other terms in Eq. (2.53) and we obtain two solutions, the regular one $u_{\mathbf{R}l}(r, E) \propto r^l$ and the irregular one $u_{\mathbf{R}l}(r, E) \propto r^{-l-1}$. Becasue inside the atomic sphere the MTO should remain finite when $r \to 0$, only the regular solution is kept.

Having determined the wave functions in and outside the atomic sphere, let's consider the boundary conditions at $r = s_{\mathbf{R}}$. As discussed before, four conditions must be satisfied for MTO's. We have chosen the regular solution for MTO inside the atomic spheres and the irregular solution for outside, therefore the first two conditions are satisfied now. However, our solutions do not yet satisfy the remaining two conditions. In order to do so, one more function is included in the MTO inside the atomic spheres. We will see that the additional function is taken to be the energy derivative of the regular solution inside the atomic sphere $\dot{u}_{\mathbf{R}l}(r, E)$. Throughout this thesis, we shall use the notation that a dot over an energy-dependent function refers to its energy derivative.

To fulfill the other two boundary conditions at $r = s_{\mathbf{R}}$, we use the Wronskian to match a function to a linear combination of two other functions [54]. For details about Wronskian, please refer to Appendix B. With the help of this matching condition in Eq. (B.4), the regular and irregular solutions can be expanded in terms of $u_{\mathbf{R}l}(r, E)$ and $\dot{u}_{\mathbf{R}l}(r, E)$ [54]:

$$J_l(r) \to -\{J, \dot{u}\}_{\mathbf{R}l} u_{\mathbf{R}l}(r, E) + \{J, u\}_{\mathbf{R}l} \dot{u}_{\mathbf{R}l}(r, E)$$

$$(2.54)$$

$$K_l(r) \to -\{K, \dot{u}\}_{\mathbf{R}l} u_{\mathbf{R}l}(r, E) + \{K, u\}_{\mathbf{R}l} \dot{u}_{\mathbf{R}l}(r, E)$$

$$(2.55)$$

where the relation

$$\{u_{\mathbf{R}l}, \dot{u}_{\mathbf{R}l}\} = -1 \tag{2.56}$$

is used. $\{\cdots\}$ is the Wronskian (see Appendix B for its definition). The notation $\{K, u\}_{\mathbf{R}l}$ refers to the Wronskian value at the boundary $r = s_{\mathbf{R}}$,

$$\{K, u\}_{\mathbf{R}l} = \{K_l(r), u_{\mathbf{R}l}(r, E)\}_{\mathbf{R}l}|_{r=s_{\mathbf{R}}},$$
(2.57)

and so as to $\{K, \dot{u}\}_{\mathbf{R}l}$, $\{J, u\}_{\mathbf{R}l}$ and $\{J, \dot{u}\}_{\mathbf{R}l}$. The Wronskians appearing in Eqs.

(2.54) and (2.55) are coefficients used to guarantee a smooth matching of the wave functions at the boundary $r = s_{\mathbf{R}}$.

Now the MTOs can be constructed. The starting point is the decaying solution $K_L(\mathbf{r})$ of the Laplace equation centered at \mathbf{R} and it can be solved exactly. By applying Eq. (2.50), $K_L(\mathbf{r}_{\mathbf{R}})$ can be rewritten as the form of

$$K_{L}(\mathbf{r}_{\mathbf{R}}) = \begin{cases} K_{L}(\mathbf{r}_{\mathbf{R}}), & r_{\mathbf{R}} \leq s_{\mathbf{R}} \\ -\sum_{L'} S_{\mathbf{R}L,\mathbf{R}'L'} J_{L'}(\mathbf{r}_{\mathbf{R}'}), & r_{\mathbf{R}'} \leq s_{\mathbf{R}'} (\mathbf{R}' \neq \mathbf{R}) \\ K_{L}(\mathbf{r}_{\mathbf{R}}), & \mathbf{r} \in I, \end{cases}$$
(2.58)

where the space is divided into three regions: i) the region inside the atomic sphere **R** in which the wave function should be matched with the solution of Eq. (2.53); ii) the region inside other spheres, where $K_L(\mathbf{r}_R)$ can be expanded by $J_{L'}(\mathbf{r}_{R'})$ centered at $\mathbf{R'}$ and iii) the interstitial region where $K_L(\mathbf{r}_R)$ is the exact solution. The only issue we need to consider now is region i). Region i) refers to the region inside the atomic sphere where no analytical solution can be obtained. But since we have mentioned before that the solutions can be expressed by $\psi_{\mathbf{R}L}$ and $\dot{\psi}_{\mathbf{R}L}$, the inner sphere solutions can be constructed by matching conditions. This can be easily achieved by applying the relation shown in Eq. (2.55). Therefore, the Muffin Tin orbitals in all space have the form of [54]

$$\chi_{\mathbf{R}L}(r, E) = \begin{cases} -\{K, \dot{u}\}_{\mathbf{R}l} \psi_{\mathbf{R}L}(\mathbf{r}_{\mathbf{R}}, E) \\ +\{K, u\}_{\mathbf{R}l} \dot{\psi}_{\mathbf{R}L}(\mathbf{r}_{\mathbf{R}}, E), & r_{\mathbf{R}} \leq s_{\mathbf{R}} \\ \sum_{L'} S_{\mathbf{R}L, \mathbf{R}'L'} \left[\{J, \dot{u}\}_{\mathbf{R}'l'} \psi_{\mathbf{R}'L'}(\mathbf{r}_{\mathbf{R}'}, E) \\ -\{J, u\}_{\mathbf{R}'l'} \dot{\psi}_{\mathbf{R}'L'}(\mathbf{r}_{\mathbf{R}'}, E)\right], & r_{\mathbf{R}'} \leq s_{\mathbf{R}'} (\mathbf{R}' \neq \mathbf{R}) \\ K_L(\mathbf{r}_{\mathbf{R}}), & \mathbf{r} \in I. \end{cases}$$
(2.59)

One last thing to mention is that the interstitial region is neglected in ASA. Here in

Eq. (2.59) we keep it just for the sake of completeness.

2.3.3 Energy-independent Muffin Tin orbitals

The MTO described in Eq. (2.59) are energy dependent. This is an undesirable drawback for numerical eigen-energy calculation. We wish to find a way to drop the energy dependence. This is done by the energy linearization.

Energy linearization is performed by taking a Taylor expansion of radial solution inside an atomic sphere centered at \mathbf{R} [54]

$$u_{\mathbf{R}l}(r, E) = \phi_{\mathbf{R}l}(r) + \dot{\phi}_{\mathbf{R}l}(r) \left(E - E^{0}_{\mathbf{R}l}\right) + \dots, \qquad (2.60)$$

where the notations of

$$\phi_{\mathbf{R}l}\left(r\right) = u_{\mathbf{R}l}\left(r, E_{\mathbf{R}l}^{0}\right), \quad \dot{\phi}_{\mathbf{R}l}\left(r\right) = \dot{u}_{\mathbf{R}l}\left(r, E_{\mathbf{R}l}^{0}\right) \tag{2.61}$$

are introduced. $E_{\mathbf{R}l}^{0}$ is the energy around which the radial wave function $u_{\mathbf{R}l}(r, E)$ is expanded. Usually $E_{\mathbf{R}l}^{0}$ is taken to be the center point of the occupied part of the $\mathbf{R}l$ -th projected valence density of states [89, 54]. It is interesting that $\phi_{\mathbf{R}l}(r)$ and $\dot{\phi}_{\mathbf{R}l}(r)$ clearly satisfy all the relations valid for $u_{\mathbf{R}l}(r, E)$ and $\dot{u}_{\mathbf{R}l}(r, E)$. Therefore, the linearized (energy-independent) MTO's, called LMTO's, can be obtained by simply replacing $u_{\mathbf{R}l}(r, E)$ and $\dot{u}_{\mathbf{R}l}(r, E)$ with functions $\phi_{\mathbf{R}l}(r)$ and $\dot{\phi}_{\mathbf{R}l}(r)$ [54]

$$\chi_{\mathbf{R}L}(r, E) = \begin{cases} -\{K, \dot{\phi}\}_{\mathbf{R}l} \phi_{\mathbf{R}L}(\mathbf{r}_{\mathbf{R}}) \\ +\{K, \phi\}_{\mathbf{R}l} \dot{\phi}_{\mathbf{R}L}(\mathbf{r}_{\mathbf{R}}), & r_{\mathbf{R}} \leq s_{\mathbf{R}} \\ \sum_{L'} S_{\mathbf{R}L, \mathbf{R}'L'} \left[\{J, \dot{\phi}\}_{\mathbf{R}'l'} \phi_{\mathbf{R}'L'}(\mathbf{r}_{\mathbf{R}'}) \\ -\{J, \phi\}_{\mathbf{R}'l'} \dot{\phi}_{\mathbf{R}'L'}(\mathbf{r}_{\mathbf{R}'})\right], & r_{\mathbf{R}'} \leq s_{\mathbf{R}'} \left(\mathbf{R}' \neq \mathbf{R}\right), \end{cases}$$
(2.62)

where the interstitial part is neglected in the expression above.

2.3.4 Overlap and Hamiltonian Matrices

In section 2.3.3, we presented the expression for linear Muffin Tin orbitals. In this section, we will present how the overlap and Hamiltonian matrices are constructed under these orbitals. The overlap and Hamiltonian matrices are defined by

$$O_{ij} = \int \chi_i \left(\mathbf{r} \right) \chi_j \left(\mathbf{r} \right) d^3 \mathbf{r}$$
(2.63)

and

$$H_{ij} = \int \chi_i \left(\mathbf{r} \right) \left[-\frac{\nabla^2}{2} + V \left(\mathbf{r} \right) \right] \chi_j \left(\mathbf{r} \right) d^3 \mathbf{r}.$$
 (2.64)

 $\{\chi_i(\mathbf{r})\}\$ are LMTO orbitals. Notice that ASA is employed in this thesis hence the integrals appearing in Eqs. (2.63) and (2.64) are restricted to integrals within atomic spheres. We start with the overlap matrix. By inserting Eq. (2.62) into the definition of overlap matrix (2.63), we find that three type of integrals need to be done [54],

$$\int_{0}^{s_{\mathbf{R}}} \phi_{\mathbf{R}L}\left(\mathbf{r}\right) \phi_{\mathbf{R}L'}\left(\mathbf{r}\right) d^{3}\mathbf{r} = \delta_{LL'}, \qquad (2.65)$$

$$\int_{0}^{s_{\mathbf{R}}} \phi_{\mathbf{R}L}\left(\mathbf{r}\right) \dot{\phi}_{\mathbf{R}L'}\left(\mathbf{r}\right) d^{3}\mathbf{r} = 0 \qquad (2.66)$$

and

$$\int_{0}^{s_{\mathbf{R}}} \dot{\phi}_{\mathbf{R}L} \left(\mathbf{r} \right) \dot{\phi}_{\mathbf{R}L'} \left(\mathbf{r} \right) d^{3} \mathbf{r} = p_{\mathbf{R}l} \delta_{LL'}.$$
(2.67)

 $p_{\mathbf{R}l}$ is defined as

$$p_{\mathbf{R}l} = \int_0^{s_{\mathbf{R}}} \dot{\phi}_{\mathbf{R}L}^2(r) r^2 d^3 r.$$
 (2.68)

Therefore, the elements of overlap matrix in the LMTO framework has the form of [54]

$$O_{\mathbf{R}'L',\mathbf{R}''L''} = \{K, \dot{\phi}\}^{2}_{\mathbf{R}'L'} \delta_{\mathbf{R}'L',\mathbf{R}''L''} - \{K, \dot{\phi}\}_{\mathbf{R}'l'} \{J, \dot{\phi}\}_{\mathbf{R}'l'} S_{\mathbf{R}'L',\mathbf{R}''L''} - S_{\mathbf{R}'L',\mathbf{R}''L''} \{J, \dot{\phi}\}_{\mathbf{R}''l''} \{K, \dot{\phi}\}_{\mathbf{R}''l''} + \sum_{\mathbf{R}L} S_{\mathbf{R}'L',\mathbf{R}L} \{J, \dot{\phi}\}^{2}_{\mathbf{R}l} S_{\mathbf{R}L,\mathbf{R}''L''} + O^{(p)}_{\mathbf{R}'L',\mathbf{R}''L''}.$$
(2.69)

The last term includes all contributions from the quantity $p_{\mathbf{R}l}$. This relation can be rewritten into a matrix form

$$O = \left(\{K, \dot{\phi}\} - S\{J, \dot{\phi}\}\right) \left(\{K, \dot{\phi}\} - \{J, \dot{\phi}\}S\right) + O^{(p)}.$$
 (2.70)

The construction of the Hamiltonian matrix is done similarly. We insert Eq. (2.62) into Eq. (2.64) and after some algebra we obtain the Hamiltonian matrix:

$$H = \left(\{K, \dot{\phi}\} - S\{J, \dot{\phi}\}\right) E^{0} \left(\{K, \dot{\phi}\} - \{J, \dot{\phi}\}S\right) - \left(\{K, \dot{\phi}\} - S\{J, \dot{\phi}\}\right) \left(\{K, \phi\} - \{J, \phi\}S\right) + H^{(p)}$$
(2.71)

where $H^{(p)}$ contains all contributions from the quantities $p_{\mathbf{R}l}$.

If we introduce an auxiliary matrix

$$M = \{K, \dot{\phi}\} - \{J, \dot{\phi}\}S, \tag{2.72}$$

the overlap and Hamiltonian matrices can be rewritten as

$$O = M^T M \tag{2.73}$$

$$H = M^{T} E^{0} M - M^{T} \left(\{K, \phi\} - \{J, \phi\} S \right).$$
(2.74)

where the contribution from $p_{\mathbf{R}l}$ is neglected since it is small.

Finally, we list the matrices in orthogonal basis,

$$O^{orth} = (M^T)^{-1} O M^{-1} = I$$
 (2.75)

$$H^{orth} = (M^T)^{-1} H M^{-1} = E^0 - (\{K, \phi\} - \{J, \phi\} S) \left(\{K, \dot{\phi}\} - \{J, \dot{\phi}\} S\right)^{-1}.$$
 (2.76)

An equivalent expression for H^{orth} follows Eq. (2.76) is

$$H^{orth} = E^0 - \frac{\{K, \phi\}}{\{K, \dot{\phi}\}} + \frac{\omega}{2} \frac{1}{\{K, \dot{\phi}\}} S\left(1 - \frac{\{J, \dot{\phi}\}}{\{K, \dot{\phi}\}}\right)^{-1} \frac{1}{\{K, \dot{\phi}\}}, \qquad (2.77)$$

which clearly shows that the orthogonal Hamiltonian matrix is real and symmetric [54].

2.3.5 Potential Parameters

Potential parameter is widely used in MTO or LMTO theory [54, 89]. After introducing potential parameters, the H^{orth} of Eq. (2.77) can have a much simpler expression. We start by introducing the definition of potential parameters and its derivatives [54]:

$$P_{\mathbf{R}l}\left(E\right) = \frac{\{K, \psi\left(E\right)\}_{\mathbf{R}l}}{\{J, \psi\left(E\right)\}_{\mathbf{R}l}},\tag{2.78}$$

$$\dot{P}_{\mathbf{R}l}(E) = \frac{\omega}{2} \frac{1}{\{J, \psi(E)\}_{\mathbf{R}l}^2}$$
(2.79)

and

$$\ddot{P}_{\mathbf{R}l}(E) = -\omega \frac{\{J, \psi(E)\}_{\mathbf{R}l}}{\{J, \psi(E)\}_{\mathbf{R}l}^3}.$$
(2.80)

The expressions for functions $\dot{P}_{\mathbf{R}l}(E)$ and $\ddot{P}_{\mathbf{R}l}(E)$ are obtained by directly taking derivatives on Eq. (2.78) and using the relation $\{J_l(r), K_l(r)\} = -\omega/2$. As discussed in the previous section, $\psi(E)$ is expanded around some energy point, so by inserting Eq. (2.60) into Eqs. (2.78) (2.79) and (2.80), we have

$$P_{\mathbf{R}l}\left(E\right) = \frac{E - C_{\mathbf{R}l}}{\Delta_{\mathbf{R}l} + \gamma_{\mathbf{R}l}\left(E - C_{\mathbf{R}l}\right)},\tag{2.81}$$

$$\dot{P}_{\mathbf{R}l}\left(E\right) = \frac{\Delta_{\mathbf{R}l}}{\left[\Delta_{\mathbf{R}l} + \gamma_{\mathbf{R}l}\left(E - C_{\mathbf{R}l}\right)\right]^2} \tag{2.82}$$

and

$$\ddot{P}_{\mathbf{R}l}\left(E\right) = -\frac{2\Delta_{\mathbf{R}l}\gamma_{\mathbf{R}l}}{\left[\Delta_{\mathbf{R}l} + \gamma_{\mathbf{R}l}\left(E - C_{\mathbf{R}l}\right)\right]^3},\tag{2.83}$$

where parameters $C_{\mathbf{R}l}$, $\Delta_{\mathbf{R}l}$ and $\gamma_{\mathbf{R}l}$ have the forms of [54]

$$C_{\mathbf{R}l} = E_{\mathbf{R}l}^{0} - \frac{\{K, \phi\}_{\mathbf{R}l}}{\{K, \dot{\phi}\}_{\mathbf{R}l}},$$
(2.84)

$$\Delta_{\mathbf{R}l} = \frac{\omega}{2} \frac{1}{\{K, \dot{\phi}\}_{\mathbf{R}l}^2} \tag{2.85}$$

and

$$\gamma_{\mathbf{R}l} = \frac{\{J, \dot{\phi}\}_{\mathbf{R}l}}{\{K, \dot{\phi}\}_{\mathbf{R}l}}.$$
(2.86)

We can also obtain the expressions for the Wronskians appearing in above equations by inversely written them in the form of

$$\{J,\phi\}_{\mathbf{R}l} = \sqrt{\frac{\omega}{2\Delta_{\mathbf{R}l}}} \left[\Delta_{\mathbf{R}l} + \gamma_{\mathbf{R}l} \left(E^{0}_{\mathbf{R}l} - C_{\mathbf{R}l}\right)\right], \qquad (2.87)$$

$$\left\{J,\dot{\phi}\right\}_{\mathbf{R}l} = \sqrt{\frac{\omega}{2\Delta_{\mathbf{R}l}}}\gamma_{\mathbf{R}l},\tag{2.88}$$

$$\{K,\phi\}_{\mathbf{R}l} = \sqrt{\frac{\omega}{2\Delta_{\mathbf{R}l}}} \left(E^{0}_{\mathbf{R}l} - C_{\mathbf{R}l} \right)$$
(2.89)

and

$$\{K, \dot{\phi}\}_{\mathbf{R}l} = \sqrt{\frac{\omega}{2\Delta_{\mathbf{R}l}}}.$$
(2.90)

Inserting Eqs. (2.87) to (2.90) into the expression of H^{orth} in Eq. (2.76), we arrive at a much simpler expression [54]

$$H^{orth} = C + \sqrt{\Delta}S(1 - \gamma S)^{-1}\sqrt{\Delta}.$$
(2.91)

In this expression, C, Δ and γ are diagonal matrices constructed by the potential parameters $C_{\mathbf{R}l}$, $\Delta_{\mathbf{R}l}$ and $\gamma_{\mathbf{R}l}$.

2.3.6 Representation Transformation

The LMTO-ASA structure matrix S can be made more sparse if we introduce the representation transformation. The representation transformation allows us to transform the LMTO basis into a short-ranged localized basis without loss of any information [42, 91] and, as a result, we obtain very sparse structure constant matrix ($S_{\mathbf{R}L,\mathbf{R}'L'}$) which greatly enhances the efficiency of numerical computation. As presented in Eqs. (2.49) and (2.58), the LMTO basis decays according to a power law r^{-l-1} for large r. But after employing the representation transformation, the transformed basis, which is called the tight-binding (TB) LMTO basis, decays exponentially with respect to r. Therefore, using TB-LMTO, interactions beyond the second nearest neighbor can be safely neglected which greatly boosts the calculation speed.

For potential functions and structure constants, the transformation is defined as [54]

$$P^{\beta} = P^{\delta} + P^{\delta} (\beta - \delta) P^{\beta},$$

$$S^{\beta} = S^{\delta} + S^{\delta} (\beta - \delta) S^{\beta},$$
(2.92)

where physical quantities are transformed from representation δ to β . β (δ) is a diagonal matrix with elements of $\beta_{\mathbf{R}l}$ ($\delta_{\mathbf{R}l}$). By some algebra, Eq. (2.92) can be rewritten as

$$P^{\beta} = \left[1 - P^{\delta} \left(\beta - \delta\right)\right]^{-1} P^{\delta},$$

$$S^{\beta} = \left[1 - S^{\delta} \left(\beta - \delta\right)\right]^{-1} S^{\delta}.$$
(2.93)

By using a special set of **R**-independent but *l*-dependent screening constants $\{\alpha_l\}$, the LMTO basis is transformed into the TB-LMTO basis [42]. For general cases only *s*, *p* and *d* orbitals are taken into consideration, so only three screening constants are included These values can be found in reference [42]. Finally let's end this section by giving out the Hamiltonian matrix after the transformation:

$$H^{orth,\beta} = C + \sqrt{\Delta} S^{\beta} \sqrt{\Delta},$$

$$S^{\beta} = S^{\delta} (1 - (\beta - \delta) S^{\delta})^{-1}.$$
(2.95)

In the self-consistent procedure, we start from some initial guess for ϕ 's (e.g., solutions from isolated spheres). Then the Hamiltonian is calculated from Eqs. (2.84), (2.85), (2.86), (2.91), (2.93), (2.94) and (2.95). After solving the corresponding secular equation, a new charge density can be calculated by Eq. (2.14). The potential at each sphere is then updated by Eq. (2.42). ϕ 's are updated according to the potential. The new ϕ 's are then used to calculate a new Hamiltonian. Such process is repeated until required accuracy is achieved.

2.4 Summary

In Summary, this chapter introduced the density functional theory and its implementations with LCAO and LMTO. The Born-Oppenheimer approximation separated the dynamics of electrons and ions. By neglecting the latter, only electron degrees of freedom need to be focused. HK theorem helped to reduce the *N*-interacting electrons problem into a variation problem with respect to the single electron density and reduced to solving the Kohn-Sham equation (2.16). The electron density $\rho(\mathbf{r})$ is calculated by Eq. (2.14). Hence, the complicated many body problem is solved

	LCAO	LMTO
Formalism	Simple	Complicated
Pseudo potential needed?	Yes	No
Computational efficiency	Fairly fast	Very fast
VS needed ?	no	yes
Easy to implement CPA?	No	Yes

by Eqs. (2.19) and (2.14) self-consistently after choosing appropriate XC functionals such as LDA, GGA or hybrid functionals. Then we derived formula for both LCAO and LMTO implementation of DFT. In the LCAO method, valence electron orbitals are constructed from corresponding isolated atomic orbitals. The core electrons are treated by pseudo potential method, which reduces the computational cost. In LMTO method, orbitals are constructed from a group of slightly overlapped muffin tin spheres. By neglecting the kinetic energy in the interstitial region (ASA), the wave functions outside the sphere were solved analytically. Meanwhile, the wave functions inside the sphere were obtained by solving the spherical symmetric Schrödinger equation in each atomic sphere numerically. By applying continuous boundary conditions at the sphere surface, these wave functions determined the Muffin tin orbitals in the whole space. Then, energy linearization was applied to the Muffin tin orbitals, rendering a group of energy independent Muffin tin orbitals, the LMTO's. With the orbitals being set, the KS Hamiltonian and overlap matrices for both LCAO and LMTO are derived. The KS Hamiltonian must be solved self-consistently using the procedure described in section 2.1.5. Both LCAO and LMTO are atomic centered basis with which Hamiltonian is in a sparse matrix form so that the KS-DFT procedure can be carried out more efficiently compared to other nonlocal basis such as plane wave [92]. Table 2.1 lists a comparison between the LCAO and LMTO (with ASA) used in this thesis. LCAO is superior to LMTO in the following aspects. LCAO is atomic orbital based so that its implementation with KS-DFT is very straightforward and simple. Vacuum spheres are not necessary in most cases with LCAO because the potential far away from nuclei are calculated exactly. LMTO requires the use of vacuum spheres in space so that it is difficult to apply to materials with low lattice symmetry such as crystals with non close-packed structure, interfaces and surfaces. LMTO outclasses LCAO in the following aspects. With LMTO, the Hamiltonian is sparser and the computation is faster than with LCAO, which allows to calculate larger systems. LCAO can handle transport problems with several hundreds atoms while LMTO can solve systems with more than several thousands atoms. In LMTO, all electrons are included in the computation so that no pseudo potential is needed. With LMTO, the CPA theory is very natural to implement (see Chapter 5) while with LCAO, further approximation is needed and the formalism is rather complicated [52].

Application: Band Engineering of the GaSbN Alloy

III-nitride material is one of the most important semiconductors after silicon [93, 94, 95]. Due to its exceptional optical properties and extraordinary ability of resisting photo-oxidation and corrosion in harsh photocatalytic conditions [96], recent works have demonstrated the exciting results that III-nitride *nanostructures* possess ideal attributes for solar fuel generation through natural water splitting [97, 98, 99, 100] and CO_2 reduction [101]. Fig. 3.1 plots the electrochemical process of water splitting. The reaction is first initiated by photon absorption ($h\nu$ in the Fig. 3.1), which generates numerous electron-hole pairs in the semiconductor's conduction band minimum (CBM) and valence band maximum (VBM) (red lines in Fig. 3.1). If the water redox potentials (blue dashed lines in Fig. 3.1) stays in between the CBM and VBM, the photo-generated electrons reduce water to form hydrogen, and the holes oxidize water molecules to give oxygen. InGaN is one of the most advanced candidates for electrochemical water splitting. By tuning the concentration x of In, the band gap E_g of $In_xGa_{1-x}N$ can be continuously lowered from about 3.4 eV at x = 0 to about 0.7 eV at x = 1, covering broadband of the solar spectra thereby producing high solar-to-energy conversion rates [15]. To achieve the best catalytic effects for water splitting, the band gap of the catalyst needs to properly straddle the redox potential of water molecules which is about 1.23 eV [102]. With appropriate electrochemical overpotentials at both the valence and conduction band edges, the optimal band gap is around $2.2 \, \text{eV}$ [103]. This can be achieved for InGaN by alloying high In concentration at about 40% into GaN. With such concentrations, the resulting InGaN typically contains significant



Figure 3.1: Schematic of the electrochemical water splitting process. A semiconductor is placed in neutral water. The red lines are the CBM and VBM of the semiconductor measured with respect to the vacuum energy level. The blue dashed lines are the redox potential of neutral water. When light $(h\nu)$ is shined onto the semiconductor, electron hole pairs are generated and captured by H^+ and OH^- in water respectively.

density of defects due to the large lattice mismatch of around 11% between InN and GaN [104, 105, 106], as well as large strain-induced polarization field [107, 108] due to the piezoelectric property of the materials. These detrimental effects are some of the serious challenges for realizing large scale practical applications of InGaN as an advanced photo catalyst for solar fuel. It is therefore important to search for other elements that can serve the same purpose as indium.

A few recent experimental investigations found that doping a small amount of antimony (Sb), at $x = 1 \sim 8\%$, the band gap of GaN was substantially reduced from 3.4 eV to about 1.9 eV at x = 1% and to about 1.5 eV at x = 8% [13, 109, 110], namely equivalent to the effect of incorporating more than 40% of indium into GaN. Obviously, the relevant regime for solar fuel is in 0 < x < 1%. Such a dilute antimonide nitride provides extraordinary opportunities for band engineering, strain engineering and polarization engineering, which may well overcome the materials challenges associated with InGaN in solar fuel applications. To date, the bowing shape of band reduction of $GaSb_xN_{1-x}$ at x < 1% - relevant for solar fuel application, has not been investigated either by experimental or theoretical investigations; the underlying cause for the sharp band gap reduction in the dilute Sb limit remained to be verified [111]; and the band bending properties of GaSbN surface for the purpose of solar fuel application have not been investigated. It is the purpose of this chapter to understand these important issues from atomic first principles presented in Chapter 2.

3.1 Computational Method

On the theoretical side, the key issue of the problem is to calculate the band gap and band edge positions of the $GaSb_xN_{1-x}$ materials. Calculations of band gap by KS-DFT at relatively large Sb concentration [110], x > 2%, found that E_g is almost saturated there. The part we want to study is the bowing regime 0 < x < 1%, in which the gap of GaN is reduced to the proper value to straddle the redox potential of water. Such low Sb concentration is very difficult for direct KS-DFT analysis. Simulating small x requires relatively large super-cells: for x = 0.1%, at least 2,000 GaN atoms must be included in the super-cell to accommodate just a single Sb atom. To obtain correct E_g for semiconductors, higher level theories such as the hybrid exchangecorrelation (XC) functional [74], are required which is computationally extremely expensive for over 2000 atoms.

To overcome these difficulties, we employ a powerful KS-DFT method [78] developed by our group with the HSE06 hybrid XC functional [74] that accurately and efficiently predicts electronic structure of semiconductors [76] which allows us to investigate $GaSb_xN_{1-x}$ in the dilute Sb concentration limit. Double-zeta polarized LCAO basis up to *d* orbitals with the ONCV pseudo potentials is employed and they produce very accurate description of the band structures and electrostatic potentials, which are necessary for our calculations. Another difficulty is the randomness of Sb atoms in $GaSb_xN_{1-x}$. The electronic band structure of the alloy depends on the arrangement of Sb atoms in the lattice. Therefore the alloy band gap has to be calculated by



Figure 3.2: Atomic supercells of the simulated GaNSb lattice under different Sb concentration. The lattices is in parallelepiped shape with the three vector angles of $\pi/2$, $\pi/2$ and $\pi/3$. (a) Sb concentration of 2.7%. (b) Sb concentration of 0.35%.



Figure 3.3: Surface calculation of a GaN surface. The bottom figure is the lattice structure of a GaN slab. The top figure plots the electrostatic potential along the slab. In the center of the slab, a unit cell of GaN is picked up to calculate the average potential (red dashed line). ΔV measures the difference between the calculated average potential and the vacuum energy.

taking the average of all possible configurations, which is prohibitive for hybrid functional calculations. To overcome the difficulty, we employ the special quasirandom structures (SQS) method [53] implemented in the Alloy Theoretic Automated Toolkit (ATAT) [112]. The method gives the best periodic supercell that approximates the true disordered state for a given number of atoms per supercell. The chosen supercells (see Fig. 3.2 for x = 2.7% and x = 0.35%) here give a band gap accuracy within 0.05 eV as tested by SQS method.

To find the band edge positions, it is useful to determine the natural band alignment (NBA) of GaSb_xN_{1-x}. In particular, the NBA reflects lineup between unstrained systems which can be determined from surface calculations [15], and for solar fuel production the NBA between the semiconductor and vacuum provides useful information. For the surface calculation, we construct a repeated mirror symmetric slab super-cell in the *m*-plane orientation (1010) of the wurtzite GaSb_xN_{1-x} structure (see Fig. 3.3). In this work we consider *m*-plane as NBA appears to be not sensitive to the crystal orientation[113]. The number of layers contained in the slab is increased until the electrostatic potential difference, ΔV , between the crystal and vacuum converges. The surface calculation allows us to align the average electrostatic potential in the crystal to the vacuum potential which acts as a common reference. Here, ΔV is calculated at the PBE level of the XC functional and for technical details we refer interested readers to Ref. [113]. Once ΔV is obtained, the absolute positions of the band edges ($E_{c/v,abs}$) with respect to the vacuum level is obtained by

$$E_{c,abs} = E_c - \Delta V - E_0, \qquad (3.1)$$

$$E_{v,abs} = E_v - \Delta V - E_0, \tag{3.2}$$

where $E_{c/v}$ are the CB minimum and VB maximum of the bulk material. E_0 is the average electrostatic potential of the bulk. $E_{c/v}$ and E_0 are obtained by HSE06 bulk calculations. In the NBA calculation, structural relaxation of the slab is difficult for KS-DFT at low Sb concentration: e.g. for x = 0.35% the slab contains about 1500 atoms. We therefore employ classical molecular dynamics to obtain the relaxed structure, by the LAMMPS method [114] with the Tersoff potential [115] found in Ref. [116]. We calibrated the potential parameters so that they produced the correct surface structures of the alloy (see Appendix E for the parameters).

3.2 Computational Setup

In KS-DFT bulk calculation, $7 \times 7 \times 5$ k-point mesh is used for sampling the Brillouin zone (BZ) in the self-consistent calculation of intrinsic GaN/GaSb which has a wurtzite primitive cell. The calculated super-cell is scaled up according to the Sb concentration: at x = 5.6, 2.7, 1.35, 0.7, 0.35, 0.175%, it contains 72, 72, 144, 288, 576 and 1152 atoms respectively, 5.6% with two Sb atoms and all the others with one Sb atom. For low Sb concentration (x < 0.7%) thus large super-cell (e.g., see Fig. 3.2)(b), Γ point is adequate for BZ sampling. The band gap $E_g(x)$ is correctly determined by employing the Heyd-Scuseria-Ernzerhof (HSE06) hybrid-XC functional [74], which predicts $E_g = 3.46$ eV for GaN and 0.55 eV for GaSb with spin-orbit coupling (SOC), in very good agreement with previous theoretical and experimental values [117, 118]. There is no reported experimental band gap for wurtzite GaSb to the best of our knowledge so we only compare with previous reported theoretical values.

For solar fuel application, the relevant compound $GaSb_xN_{1-x}$ is at the dilute limit, $0 < x \leq 1\%$, which has only small lattice mismatch to GaN so that high quality materials can be readily grown experimentally [119]. As mentioned above, experimentally Sb is surprisingly efficient in reducing the GaN band gap [110, 120]. At x > 8%, $GaSb_xN_{1-x}$ becomes indirect band gap material. At small x, electronic property of $GaSb_xN_{1-x}$ such as $E_g(x)$ is not sensitive to the atomic arrangements so long as the Sb atoms are distributed uniformly in the material. The super-cell is therefore prepared as follows. Take x=5.6% as an example, we prepare a super-cell with 72 atoms and two of the N atoms are replaced by Sb. The smallest x we investigated is 0.175% for which there are 1152 atoms inside the supercell with one Sb atom. In the dilute limit,



Figure 3.4: Band gap $E_g(x)$ of dilute $GaSb_xN_{1-x}$ alloy vs. Sb concentration x by KS-DFT with the HSE06 hybrid functional. Inset: E_g vs. lattice mismatch for $GaSb_xN_{1-x}$ (green) and $In_xGa_{1-x}N$ (red). The mismatch is measured w.r.t. wurtzite GaN crystal.

the distance between Sb atoms is ~ 10 Å so that little direct interaction exists between them. The atoms in the super-cell are relaxed with the PBESol XC functional [121] using RESCU [78]. In the following, SOC is not considered to reduce computational cost as it only make a difference on the band gap and electrostatic potential for less than 20 meV.

3.3 Results and Discussion

Fig. 3.4 plots the calculated $E_g(x)$ of $\operatorname{GaSb}_x N_{1-x}$ versus x. As x increases, $E_g(x)$ drops sharply in the range of 0 < x < 1% and approaches a saturated value of about 1.8 eV for x > 1%, consistent with experiment observations [110, 120]. To the best of our knowledge, such a strong band bowing is unseen in other III-V materials - for instance, for $\operatorname{In}_x \operatorname{Ga}_{1-x} N$ the E_g drops slowly with x (see the inset of Fig. 3.4). This distinct band bowing in Fig. 3.4 cannot be fit by conventional bowing formula involving a single bowing parameter b, $E_g(\operatorname{GaSb}_x N_{1-x}) = (1-x)E_g^{\operatorname{GaN}} + xE_g^{\operatorname{GaSb}} - bx(1-x)$. The



Figure 3.5: DOS of $GaSb_xN_{1-x}$ at the valence band side of the Fermi level. (a) Total DOS normalized to the same number of Sb atoms at different x. The band edge of GaN, $E_{v,GaN}$, is used as reference to align the DOS at different Sb concentration. (b) Total DOS per atom. For clarity, (a, b) do not include the calculated data at x = 0.7%, 2.7% and 5.6% as they are similar to the presented curves of x = 0.35% and 1.4%.

strong bowing in $GaSb_xN_{1-x}$ suggests strong interactions between the atoms and/or orbitals (see below). The predicted bowing in Fig. 3.4 is quantitatively comparable to the measured data [110]. The inset of Fig. 3.4 plots the calculated E_g of GaSbN and InGaN versus the relative lattice mismatch to GaN: GaSbN has a much smaller lattice mismatch than InGaN to GaN which is important from material fabrication point of view.

What is the microscopic origin of the strong band bowing in $GaSb_xN_{1-x}$? To interpret their experimental data, Refs. [110, 120] suggested the bowing to come from the band anti-crossing (BAC). Here we investigate this picture from an *ab initio* point of view by calculating the density of states (DOS) of the materials. We find that at small x, Sb atoms only contribute to the valence band (VB) of $GaSb_xN_{1-x}$ and introduce impurity states inside the GaN band gap, it has very little influence on the conduction band (CB). Therefore in Fig. 3.5 we plot the calculated DOS at VB side of the Fermi level, for x=1.35%, 0.35% and 0.175%, with the curves aligned with respect to $E_{v,GaN}$ which is the VB edge of GaN. We also found that VB DOS of $GaSb_xN_{1-x}$ below the zero point of the horizontal axis in Fig. 3.5, is very similar to that of the VB DOS of pure GaN, which means the main effect of Sb is to introduce



Figure 3.6: Projected DOS of $GaSb_xN_{1-x}$ for x = 2.7%. DOS is projected onto s, p orbitals of N atoms and s, p, d orbitals of Ga atoms. The Fermi level is shifted to 0 eV.

impurity states inside the original GaN gap, namely, the sharp reduction of $E_g(x)$ and strong bowing of GaSb_xN_{1-x} is due to these impurity states. As the Sb concentration x decreases, width of the impurity states becomes narrower and the main peak moves closer to $E_{v,\text{GaN}}$ (see Fig. 3.5(a)). This agrees reasonably well with the BAC theory [111]

$$E_{\pm}(k) = \frac{1}{2} \left\{ E^{\nu}(k) + E^{Sb} \pm \sqrt{[E^{Sb} - E^{\nu}(k)]^2 + 4V^2 x} \right\}$$

where E^v is the original GaN VB edge, E^{Sb} is the Sb impurity energy, and V the coupling parameter between impurity states and the GaN valence states. At the HSE06 level of the XC functional, $E^{Sb} \approx 0.5$ eV above the GaN VB. Then we obtain V = 6.7 eV (see Appendix E for details) which is significantly larger than that of other III-V materials [111], e.g., 2.7 eV for GaAs_xN_{1-x} and 3.5 eV for InP_xN_{1-x} etc..

It is also interesting to reveal how atoms interact in $GaSb_xN_{1-x}$ by projecting DOS (PDOS) on to each atom. Fig. 3.6 shows the calculated pdos for x = 2.7%. At the VB edge, the *p*-orbital of N and *p*, *d*-orbital of Ga are found to dominate. For impurity states, we found the *p*-orbital of Sb and N, as well as the *p*, *d*-orbital of Ga,
dominate. This indicates a strong interaction between the impurity states introduced by Sb and the GaN VB, consistent with the fitted large V parameter in the BAC model discussed in the previous paragraph. As expected, it is the Ga atoms nearest neighboring to Sb that contribute to the impurity states. It is however very surprising that the *p*-orbital of N has the largest contribution to the impurity states even though in wurzite GaSbN, the closest N atoms to Sb is only the *next* nearest neighbor. Indeed, by calculating the real space projected charge density of the impurity states, we find that Sb orbitals are mixed with those of its nearest neighbor Ga atoms and nextnearest neighbor N atoms. This can be understood by the fact that Sb has a large radius so that its orbitals can overlap and interact strongly with the next-nearest N atoms.

Having understood that the sharp reduction of $E_g(x)$ in the dilute Sb limit is due to efficient establishment of the impurity states in the GaN gap, it remains to be understood why the E_g reduction becomes very slow when x > 1%, i.e. the "turning" behavior in the bowing curve Fig. 3.4. At large x, different Sb atoms interact with the same N atom, our investigation suggests this to weaken the interaction between Sb and N, therefore slows down the E_g decrease. Inspecting the effective p-orbital radius of Sb and N, we find the "critical" distance between two Sb atoms to be approximately 15 Å - below which both Sb interact with the same N atom. This is very close to the average distance between Sb atoms in the case of x = 0.7%, where the "turning" point is located (see Fig. 3.4). Another important point is the effective DOS of the impurity states: it should be comparable to that of the GaN VB to act as a "second VB". Fig. 3.5(b) shows the total DOS per atom. We find that even at 0.175\%, the DOS is on the same order as that of the GaN VB, which indeed indicates the possibility for the impurities states acting as a valence band.

Fig. 3.7 plots the calculated VB and CB alignments relative to the vacuum level. We found that surface structural relaxation tends to lower the band position by approximately $0.25 \,\text{eV}$. This is caused by the surface dipole of the $(01\overline{10})$ plane of



Figure 3.7: Band alignments relative to the vacuum level as a function of antimony concentration x. Band positions are aligned to vacuum energy by surface calculations. The blue (red) dots/dashed lines are for relaxed (unrelaxed) surface cases. Black lines refer to the neutral water redox potentials of H^+/H_2 and O_2/H_2O .

GaSbN. On GaN (0110) surface, Ga atoms tend to go inside the lattice so that the polar Ga-N bonds have a perpendicular component to the surface which lowers down electron affinity of GaN. In GaSbN, Sb atoms close to the surface has a similar effect according to our calculation, so that a similar reduction in band alignments is obtained. We observe that at low Sb concentration, the VB edge varies strongly at x < 1% but approaches a saturated value beyond 1%. The CB edge varies much less. This is consistent with the finding presented above, namely impurity Sb states interact with GaN VB but not CB. Although there appears to be no simple band bowing model to describe the VB alignment as that in Ref. [15], we can find such a model to describe the CB very well. As we are only interested in the low x range $0 < x \le 5.6\%$, we use the normalized concentration $x_n \equiv x/0.056$, so that,

$$E_{c,\text{GaSbN}} = (1 - x_n)E_{c,\text{GaN}} + x_n E_{c,5.6\%} - bx_n(1 - x_n)$$

where $E_{c,5.6\%}$ is the absolute CB edge at x = 5.6% and b the bowing parameter.

Fitting the calculated data gives the bowing parameter b=0.62 eV.

To achieve unassisted solar water splitting, it is essential that band edges of the semiconductor photocatalyst straddle the water redox potential which, for neutral water, is shown as the horizontal solid black lines in Fig. 3.7 (obtained from Ref. [15]). As we can see, in order to realize neutral water splitting, Sb content must be controlled to 0.3% or slightly lower if overpotentials are taken into considered. At this doping level, the lattice mismatch is less than 0.1% which is ideal for achieving good crystal quality experimentally. Higher Sb content (> 1%) may be suitable for acid water splitting.

3.4 Summary

In summary, we have investigated the band gap and band alignment of $GaSb_xN_{1-x}$ alloys with KS-DFT. The band gaps are predicted for the dilute Sb limit at the HSE06 hybrid functional level. We find that the band gap decreases sharply in the Sb concentration range of $0 < x \leq 1\%$. The fundamental mechanism of such a sharp band bowing is found to be due to impurity states inside the band gap of GaN and a strong quantum interaction between the orbitals of Sb impurity and their surrounding neighbors and, in particular, the interaction between Sb and its next nearest neighbor N atom plays a significant role. Natural band alignments are calculated by combining the HSE06 bulk calculation and PBE surface calculation. Using the calculated natural band alignments, we predict that $x \approx 0.3\%$ and slightly lower of Sb doping should be ideal for photochemical water splitting in solar fuel applications. Green's Function Theory for Quantum Transport

This chapter is a review of the non-equilibrium Green's function (NEGF) theory used in this thesis. In order to perform parameter-free first principles simulation for non-equilibrium quantum transport, we shall combine the density functional theory (DFT) with the non-equilibrium Green's function theory: DFT is used to construct the Hamiltonian of the device; NEGF is employed to obtain the non-equilibrium quantum statistics of the device. In Chapter 2, we have discussed DFT and its implementation. This chapter is dedicated to the presentation of NEGF theory, based on which disorder scattering will be performed in the chapter followed.

This chapter is organized as follows. In Section 4.1, the Laudauer formalism will be reviewed. In section 4.2, we will present how to obtain important physical quantities, including the non-equilibrium electron density matrix, current and transmission coefficients in the framework of NEGF theory. Then the computational implementation of NEGF-DFT will be outlined in section 4.3. Section 4.4 is devoted to present the NEGF formalism in the LMTO framework. Section 4.5 is a brief summary for this chapter.

4.1 Landauer Picture for Quantum Transport

In this thesis, We shall consider a transport system schematically shown in Fig. 4.1 which is an open system. The grey ellipse in the center represents the center region of



Figure 4.1: Diagram for Landauer picture. A two-probe structure is presented in the figure. The grey ellipse in the center denotes the central region of the device where scattering occurs. The device is connected to the outside two reservoirs through two leads (the two horizontal light blue bars). The two reservoirs stays in equilibrium respectively with independent chemical potentials $\mu_{L/R}$ and distribution functions $f_{L/R}$. We use a battery symbol to show that the electrochemical potential of the two reservoirs can be different.

our device. The device is connected to the outside world (two reservoirs) through two leads (two light blue horizontal bars in the Fig. 4.1). The reservoirs stay on their own equilibrium states with their own chemical potential $\mu_{L/R}$ and a bias voltage $V_{L/R}$ is applied across. In the Landauer picture, electrons coming out of a reservoir follow the equilibrium distribution of that reservoir. After scattering in the device region, electrons arriving at the reservoirs will empty out without any reflection (reflection is allowed before electrons entering the reservoirs). It is assumed that the distribution in the left reservoir f_L is independent with that of the right reservoir f_R because of the fact that they are far apart. As whole the entire system is in non-equilibrium since these Fermi functions are not equal when the bias voltage V is applied to drive a current flow. Finally the current flowing through the device is calculated by

$$I = \frac{e}{h} \int T(\varepsilon) \left[f_L(\varepsilon) - f_R(\varepsilon) \right] d\varepsilon, \qquad (4.1)$$

where $T(\varepsilon)$ is the transmission function and related to the probability that electron injected from the left lead will transmit to the right lead. It is worth nothing that the bias information $(V_{L/R})$ is included in the two Fermi distribution functions in Eq. (4.1) and the transmission function $T(\varepsilon)$ depends on the bias also. It is calculated from the Green's functions:

$$T(\varepsilon) = \operatorname{Tr} \left\{ \Gamma_L(\varepsilon) \, G^r(\varepsilon) \, \Gamma_R(\varepsilon) \, G^a(\varepsilon) \right\}.$$
(4.2)

 $G^{r/a}$ and $\Gamma_{L/R}$ are the Green's functions and linewidth functions respectively. More details about them will be talked about in Section 4.2. Derivations of Eqs. 4.1 and 4.2 are performed by the NEGF theory and can be found in Refs. [122, 123].

Finally, we would like to point out that Eq. (4.2) applies as long as the Hamiltonian has a quadratic form [123]. No higher order terms like electron-electron scattering or phonon scattering should exist in the Hamiltonian if we use Eq. (4.2) to calculate the transmission.

4.2 NEGF

When modeling a nano device as a two-probe structure, a goal is to compute electronic current using Eq. (4.1). Let's recall the details for a two-probe model: the center region is connected to the outside world via two semi-infinite long leads extending to the reservoirs (e.g. Fig. 4.1). These leads respect equilibrium Fermi-Dirac statistics, because they are in contact with thermal reservoirs. Moreover, they have their own chemical potential $\mu_{L/R}$, which satisfy $\mu_L - \mu_R = qV$. q is the charge of the carrier and V is the applied voltage (see Fig. 4.1).

In NEGF theory, the most important physical quantity is the non-equilibrium electron density matrix. In the following we will present in detail how this physical quantity is calculated. We emphasize that the formalism in this section are derived in the framework of LCAO basis. The subscripts for basis are omitted for the sake of cleanness of the expressions.

We start with the calculation of electron density ρ . In equilibrium, the density is calculated via Eq. (2.14). In non-equilibrium, it is calculated by the Keldysh NEGF [123] via the Keldysh equation

$$G^{<}(\varepsilon) = G^{r}(\varepsilon) \Sigma^{<}(\varepsilon) G^{a}(\varepsilon), \qquad (4.3)$$

such that

$$\rho = -\frac{i}{2\pi} \int_{-\infty}^{\infty} G^{<}(\varepsilon) \, d\varepsilon. \tag{4.4}$$

Here G^r and G^a are the retarded and advanced Green's functions. $\Sigma^<$ is the lesser self-energy which will be explained in detail later. Eqs. (4.3) and (4.4) are applicable to both equilibrium and non-equilibrium systems. The Keldysh equation connects the distribution Green's function $G^<$ with the retarded and advanced Green's functions and the latter can be calculated by the following formula [123]

$$G^{r,a}(\varepsilon) = \left[\varepsilon O - H_0 - \Sigma^{r,a}(\varepsilon)\right]^{-1}.$$
(4.5)

We can also express them by using the Dyson equation [123]

$$G^{r,a}\left(\varepsilon\right) = G_0^{r,a}\left(\varepsilon\right) + G_0^{r,a}\left(\varepsilon\right)\Sigma^{r,a}\left(\varepsilon\right)G^{r,a}\left(\varepsilon\right).$$
(4.6)

By combining (4.5) and (4.6), the retarded/advanced Green's function can be rewritten as

$$G^{r,a}(\varepsilon) = \left[\left(G_0^{r,a}(\varepsilon) \right)^{-1} - \Sigma^{r,a}(\varepsilon) \right]^{-1}$$

$$G_0^{r,a}(\varepsilon) = \left[\left(\varepsilon \pm i0^+ \right) O - H_0 \right]^{-1}.$$
(4.7)

Here matrix O is the overlap matrix, H_0 is the Hamiltonian of the isolated central region of the device. Quantity 0^+ is a positive infinitesimal.

The lesser self-energy $\Sigma^{<}$ describes the impact from the two leads of the device (see Fig. 4.1). Since the two leads are independent with each other, $\Sigma^{<}$ is contributed by two components

$$\Sigma^{<}(\varepsilon) = \Sigma_{L}^{<}(\varepsilon) + \Sigma_{R}^{<}(\varepsilon), \qquad (4.8)$$

where $\Sigma_{L,R}^{\leq}$ are the self-energy from the left/right lead. As stated before, the two leads stay on their respective equilibrium states, Σ^{\leq} can be further written in the form [122, 123]:

$$\Sigma^{<}(\varepsilon) = if_{L}(\varepsilon)\Gamma_{L}(\varepsilon) + if_{R}(\varepsilon)\Gamma_{R}(\varepsilon)$$

$$f_{L,R}(\varepsilon) = \frac{1}{e^{(\varepsilon - \mu_{L,R})/k_{B}T} + 1}$$

$$\Gamma_{L,R}(\varepsilon) = i\left(\Sigma_{L,R}^{r}(\varepsilon) - \Sigma_{L,R}^{a}(\varepsilon)\right)$$
(4.9)

where $\Gamma_{L,R}$ is the linewidth function. The linewidth function describes the coupling between the leads and the center region of the device (see Fig. 4.1). The retarded/advanced self-energy appearing in Eq. (4.5) is obtained from the contribution of the two leads:

$$\Sigma^{r,a}\left(\varepsilon\right) = \Sigma_{L}^{r,a}\left(\varepsilon\right) + \Sigma_{R}^{r,a}\left(\varepsilon\right).$$
(4.10)

To calculate $\Sigma_L^{r,a}$ and $\Sigma_R^{r,a}$, we write the Hamiltonian and overlap matrices of the total system in the form:

$$H = \begin{pmatrix} H_{LL} & H_{LC} & 0 \\ H_{CL} & H_{CC} & H_{CR} \\ 0 & H_{RC} & H_{RR} \end{pmatrix}$$

$$O = \begin{pmatrix} O_{LL} & O_{LC} & O_{LR} \\ O_{CL} & O_{CC} & O_{CR} \\ O_{RL} & O_{RC} & O_{RR} \end{pmatrix},$$
(4.11)

where the central region Hamiltonian H_{CC} is a $N \times N$ matrix, the lead Hamiltonian $H_{LL/RR}$ is a $\infty \times \infty$ matrix and the lead-center coupling Hamiltonian $H_{L/R,C}$ $(H_{C,L/R})$

is a $\infty \times N$ or $N \times \infty$ matrix. The same dimensional properties apply to the overlap matrices. Combining Eq. (4.11) with Eqs. (4.7)-(4.9) and set the isolated central Hamiltonian H_0 to H_{CC} , the self-energies can be calculated by the relation

$$\Sigma_{\tau=L,R}^{r,a} = \left(\varepsilon O_{\tau C} - H_{\tau C}\right)^{\dagger} \left[\left(\varepsilon \pm i0^{+}\right) O_{\tau\tau} - H_{\tau\tau} \right]^{-1} \left(\varepsilon O_{\tau C} - H_{\tau C}\right)$$
(4.12)

where the matrix $[(\varepsilon \pm i0^+) O_{\tau\tau} - H_{\tau\tau}]^{-1}$ is the Green's function from the semiinfinite lead, namely the surface Green's function. Note that $H_{\tau\tau}$ and $H_{\tau C}$ have infinite dimensions so the inverse and multiplication operation in Eq. (4.12) are ill defined and further treatment are needed to compute the self-energies.

To overcome such difficulties in computing self-energies, we note that in numerical simulations, the electrodes of the device are modeled as periodic crystal structures, usually metals or degenerately doped semiconductors. Hence the Hamiltonian H_{LL} (we take the left lead as an example) for the electrodes can be written in the form:

$$H_{LL} = \begin{pmatrix} \ddots & \ddots & 0 \\ \ddots & H_{00} & H_{01} \\ 0 & H_{10} & H_{00} \end{pmatrix}.$$
 (4.13)

This translational symmetry enables one to calculate the surface Green's function iteratively without directly inverting the matrix with infinite dimensions [124]. The essential idea is to compute the Green's function of one unit cell of the electrode: this is possible since a unit cell has only finite number of atoms so that the Hamiltonian is a finite matrix that can be inverted. Afterward, one adds a second unit cell using Dyson's equation and solves it to get the Green's function of two unit cells. This is also possible since one is still dealing with finite number of atoms. This process is repeated until one has added very large number of unit cells and the resulting Green's function converges to that of the infinite chain of the unit cells.

4.3 NEGF-DFT Implementation

In the previous section, we introduced how to calculate non-equilibrium quantities in the NEGF framework. Here, some issues related to its implementation will be discussed. Again, the formalism in this section are derived under LCAO basis. The subscripts for basis are omitted for the sake of simplicity.

4.3.1 Self-Consistent Procedure

Fig. 4.2 is a flowchart showing the process of self-consistent iteration in the NEGF-DFT technique. All physical quantities including Green's functions and self-energies are matrices when a basis set is employed. This basis can be LCAO (see Section 2.2), LMTO (see Section 2.3) or others. In the algorithm shown in Fig. 4.2, we start by an initial guess for electron density usually using the neutral atom density (section 2.1.5). Then the Hamiltonian of the device is constructed according to Eqs. (2.18) and (2.19). Using this Hamiltonian, the corresponding Green's function is determined after the self-energies from the leads are obtained separately. By integrating the $G^{<}$ over energy according to Eq. (4.4), a new electron density is determined. If this electron density is within some pre-specified tolerance from the density of the previous self-consistent step, the self-consistency is deemed achieved and the iteration loop ends. If not, the new electron density is used for the next iteration step. This process is repeated until self-consistency is achieved.

4.3.2 Energy Integration

Eq. (4.4) states that the electron density is obtained by integrating the lesser Green's function over the whole energy range from $-\infty$ to $+\infty$. Unfortunately, this integration is hard to do in numerical simulations. This is because the Green's function integrand has many singularities near the integration axis (the real energy axis).



Figure 4.2: Diagram for self-consistent procedure of NEGF-DFT technique. All physical quantities including self-energies and Green's functions are matrices under some type of basis, e.g. LCAO basis or LMTO basis.

To solve this issue, integration range is reduced due to the Fermi-Dirac distribution function in the integrand of Eq. (4.4) via Eq. (4.9). Note that Fermi distribution decreases exponentially to zero as the energy goes above the Fermi energy, hence the upper limit of the integration range is practically bounded, and set to $\varepsilon_h =$ $\max(\mu_L, \mu_R) + nk_BT$ where T is the temperature, k_B the Boltzmann constant and n is usually taken to be 15~30 [90]. When choosing the lower limit, we have to make sure it is low enough to account for the electronic density of states all the way from the band bottom. In numerical simulations, this lower limit (ε_l) is often chosen to be many atomic units below the min (μ_L, μ_R).

Having reduced the integration range of Eq. (4.4), to deal with the singularities of the Green's function, we break the integration into two separate parts

$$\rho(\mathbf{r}) = -\frac{i}{2\pi} \int_{-\infty}^{\infty} G^{<}(\mathbf{r}, \mathbf{r}, \varepsilon) \, d\varepsilon = -\frac{i}{2\pi} \int_{\varepsilon_{l}}^{\varepsilon_{h}} G^{<}(\mathbf{r}, \mathbf{r}, \varepsilon) \, d\varepsilon$$
$$= -\frac{i}{2\pi} \int_{\varepsilon_{l}}^{\varepsilon_{m}} G^{<}(\mathbf{r}, \mathbf{r}, \varepsilon) \, d\varepsilon - \frac{i}{2\pi} \int_{\varepsilon_{m}}^{\varepsilon_{h}} G^{<}(\mathbf{r}, \mathbf{r}, \varepsilon) \, d\varepsilon.$$
(4.14)

where the lower limit ε_l is a very negative number. One interesting property of the integrand is that when below the energy $\varepsilon_m = \min(\mu_L, \mu_R) - nk_BT$, the two Fermi distribution functions satisfy $f_L = f_R = 1$. Therefore by Eqs. (4.3), (4.5) and (4.9) $G^{<}$ satisfies



Figure 4.3: Diagram of integration path when calculating electron density. The red curve refers to the path we use to integrate lesser Green's function.

and Eq. (4.14) becomes

$$\rho(\mathbf{r}) = -\frac{1}{\pi} \int_{\varepsilon_l}^{\varepsilon_m} \operatorname{Im}\left(G^r\left(\mathbf{r}, \mathbf{r}, \varepsilon\right)\right) d\varepsilon - \frac{i}{2\pi} \int_{\varepsilon_m}^{\varepsilon_h} G^{<}\left(\mathbf{r}, \mathbf{r}, \varepsilon\right) d\varepsilon.$$
(4.16)

Now the first part of the integration can be easily integrated because the singularities of the function Im $(G^r(\mathbf{r}, \mathbf{r}, \varepsilon))$ only appear in the negative half of the complex energy plane. Therefore by constructing a new integration contour (see the red contour C in Fig. 4.3) in the positive half of the complex energy plane (Fig. 4.3), the first term of Eq. (4.16) can be calculated without difficulty by the theorem of residual. Then Eq. (4.15) becomes

$$\rho(\mathbf{r}) = -\frac{1}{\pi} \int_{C} \operatorname{Im} \left[G^{r}\left(\mathbf{r}, \mathbf{r}, z\right) \right] dz - \frac{i}{2\pi} \int_{\varepsilon_{m}}^{\varepsilon_{h}} G^{<}\left(\mathbf{r}, \mathbf{r}, \varepsilon\right) d\varepsilon, \qquad (4.17)$$

where z is the complex energy. Now the troublesome part is the second integral in Eq. (4.17). Because $G^{<}$ has singularities on both upper and lower half of the complex energy planes, this integral can only be done along the real energy axis with large number of energy points. To smooth out the sharp features around the singularities, typically a very small imaginary number is added to the energy.

4.4 NEGF-DFT in TB-LMTO Method

In Section 4.3, we discussed NEGF-DFT within the LCAO implementation. The NEGF-DFT method can also be implemented in LMTO-ASA, which is presented in this section under the TB-LMTO framework. Since the Hamiltonian (H^{orth} in Eq. (2.95)) is expressed under a set of orthogonal basis, the overlap matrix (O) in this section is set to be 'I' by default.

We start with the retarded and advanced Green's functions G^r and G^a . Green's function are calculated from the Hamiltonian,

$$G^{r}(\varepsilon) = \left[\varepsilon - H^{orth} + i0^{+}\right]^{-1},$$

$$G^{a}(\varepsilon) = \left[\varepsilon - H^{orth} - i0^{+}\right]^{-1},$$

$$H^{orth} = C + \sqrt{\Delta}S^{\alpha}(1 - (\gamma - \alpha)S^{\alpha})^{-1}\sqrt{\Delta},$$
(4.18)

where we have assumed that the Hamiltonian is transformed to the α representation. The superscript α on the Hamiltonian is omitted for simplicity of notation. The Green's functions are from Eq. (4.5). The self-energy terms disappear above since the Hamiltonian is for the whole system. An infinitesimal 0⁺ is added to distinguish the retarded and advanced Green's function. Combining these two equations together we have

$$G^{r,a}(\varepsilon) = \left[\varepsilon - C - \sqrt{\Delta}S^{\alpha}(1 - (\gamma - \alpha)S^{\alpha})^{-1}\sqrt{\Delta}\right]^{-1}.$$
(4.19)

After some algebra (see Appendix C.1) the Green's functions can be written as

$$G^{r,a} = \frac{\gamma - \alpha}{\Delta + (\gamma - \alpha) \left(\varepsilon^{\pm} - C\right)} + \frac{\sqrt{\Delta}}{\Delta + (\gamma - \alpha) \left(\varepsilon^{\pm} - C\right)} \left[\frac{\varepsilon^{\pm} - C}{\Delta + (\gamma - \alpha) \left(\varepsilon^{\pm} - C\right)} - S^{\alpha}\right]^{-1} \times \frac{\sqrt{\Delta}}{\Delta + (\gamma - \alpha) \left(\varepsilon^{\pm} - C\right)} = \lambda^{\alpha} \left(\varepsilon^{\pm}\right) + \mu^{\alpha} \left(\varepsilon^{\pm}\right) \left[P^{\alpha} \left(\varepsilon^{\pm}\right) - S^{\alpha}\right]^{-1} \mu^{\alpha} \left(\varepsilon^{\pm}\right), \qquad (4.20)$$

where $\varepsilon^{\pm} = \varepsilon \pm i0^+$. P^{α} , λ^{α} and μ^{α} are diagonal matrices with elements of

$$P_{\mathbf{R}l}^{\alpha}\left(\varepsilon\right) = \frac{\varepsilon - C_{\mathbf{R}l}}{\Delta_{\mathbf{R}l} + (\gamma_{\mathbf{R}l} - \alpha_{l})\left(\varepsilon - C_{\mathbf{R}l}\right)},$$

$$\lambda_{\mathbf{R}l}^{\alpha}\left(\varepsilon\right) = \frac{\gamma_{\mathbf{R}l} - \alpha_{l}}{\Delta_{\mathbf{R}l} + (\gamma_{\mathbf{R}l} - \alpha_{l})\left(\varepsilon - C_{\mathbf{R}l}\right)},$$

$$\mu_{\mathbf{R}l}^{\alpha}\left(\varepsilon\right) = \frac{\sqrt{\Delta_{\mathbf{R}l}}}{\Delta_{\mathbf{R}l} + (\gamma_{\mathbf{R}l} - \alpha_{l})\left(\varepsilon - C_{\mathbf{R}l}\right)}.$$
(4.21)

To proceed further, we introduce an auxiliary Green's function:

$$g^{\alpha}(\varepsilon) = \left[P^{\alpha}(\varepsilon) - S^{\alpha}\right]^{-1}.$$
(4.22)

Then the Green's function becomes

$$G^{r,a} = \lambda^{\alpha} \left(\varepsilon^{\pm} \right) + \mu^{\alpha} \left(\varepsilon^{\pm} \right) g^{\alpha} \left(\varepsilon^{\pm} \right) \mu^{\alpha} \left(\varepsilon^{\pm} \right).$$
(4.23)

In addition, g^{α} is transformed according to the relation

$$g^{\beta} = (\beta - \delta) \frac{P^{\delta}}{P^{\beta}} + \frac{P^{\delta}}{P^{\beta}} g^{\delta} \frac{P^{\delta}}{P^{\beta}}$$
(4.24)

in the representation transformation.

Like in LCAO framework, when calculate the transport of a two-probe system, we have to focus on the Green's function in the central region. The influence from the leads has to be considered. Reference [37, 91, 125] shows that equation Eq. (4.23) remains valid but the effect of L/R electrodes modifies the expression for the auxiliary Green's function. The auxiliary Green's function becomes

$$\tilde{g}_{CC}^{\alpha,r/a}\left(\varepsilon\right) = \left[P_{C}^{\alpha}\left(\varepsilon\right) - S_{CC}^{\alpha} - \Pi_{L}^{\alpha,r/a}\left(\varepsilon\right) - \Pi_{R}^{\alpha,r/a}\left(\varepsilon\right)\right]^{-1}.$$
(4.25)

where the modified self-energies are

$$\Pi_L^{\alpha,r/a}(\varepsilon) = S_{CL}^{\alpha} [P_L^{\alpha}(\varepsilon) - S_{LL}^{\alpha}]^{-1} S_{LC}^{\alpha},$$

$$\Pi_R^{\alpha,r/a}(\varepsilon) = S_{CR}^{\alpha} [P_R^{\alpha}(\varepsilon) - S_{RR}^{\alpha}]^{-1} S_{RC}^{\alpha}.$$
(4.26)

For details of these relations, please refer to Appendix C.2. We emphasize that a " \sim " symbol was added on top of the two-probe g to distinguish with the bulk Green's function g.

The next important quantity is the lesser Green's function. This can be easily obtained by inserting quantities in LMTO form into Eq. (4.9),

$$G_{CC}^{<}(\varepsilon) = \mu_{C}^{\alpha,r}(\varepsilon) \,\tilde{g}_{CC}^{\alpha,<}(\varepsilon) \,\mu_{C}^{\alpha,a}(\varepsilon) ,$$

$$\tilde{g}_{CC}^{\alpha,<}(\varepsilon) = \tilde{g}_{CC}^{\alpha,r}(\varepsilon) \left[\Pi_{L}^{\alpha,<}(\varepsilon) + \Pi_{R}^{\alpha,<}(\varepsilon) \right] \,\tilde{g}_{CC}^{\alpha,a}(\varepsilon) ,$$

$$\Pi_{\tau=L,R}^{\alpha,<}(\varepsilon) = i f_{\tau}(\varepsilon) \cdot \Lambda_{\tau}^{\alpha}(\varepsilon) ,$$

$$\Lambda_{\tau}^{\alpha}(\varepsilon) = i \left[\Pi_{\tau}^{\alpha,r}(\varepsilon) - \Pi_{\tau}^{\alpha,a}(\varepsilon) \right] .$$
(4.27)

Finally, the transmission function can be written in the form [126]:

$$T\left(\varepsilon\right) = \operatorname{Tr}\left(\Lambda_{L}^{\alpha}\tilde{g}_{CC}^{\alpha,r}\Lambda_{R}^{\alpha}\tilde{g}_{CC}^{\alpha,a}\right).$$

$$(4.28)$$

4.5 Summary

In this chapter we introduced NEGF-DFT theory for solving quantum transport problems. The Landauer picture, discussed in Section 4.1 can be used to describe the quantum transport of two-probe devices out of equilibrium. We presented the basic formalism for NEGF-DFT from which quantities such as electron density ρ and transmission coefficient T are calculated. Later, the NEGF-DFT formula in LMTO framework was derived. The NEGF-DFT method correctly accounts for the nonequilibrium quantum statistics that are important for transport calculations. The NEGF-DFT technique is computationally efficient and does not rely on empirical parameters. Furthermore, the NEGF-DFT framework is formulated in the language of many body theory and hence is well suited to include new physical effects in its theory such as impurity scattering which will be discussed in detail in the next chapter.

Disorder Scattering

As discussed in Chapter 1, it is inevitable that any device or material fabricated in lab or industry contains some amount of disorder or imperfections such as impurities, defects, dopants and so on. These unintentional disorders sit at unpredictable random locations in the host lattice and significantly affect quantum transport through the device. Such disorders also causes variations in device properties from device to device [127]. Therefore, theories and techniques that can investigate random disordered system from first principles are highly desired. Previously, such disorder problem is solved by the "brute force" method. Consider the most common substitutional disorder problem (see Fig. 5.1), impurities randomly substitute the host atoms in the center region of the two-probe system. In principle, to calculate the average transport property, one has to perform calculation (usually by NEGF method) for each individual disorder configuration. Suppose there are N sites in the center and 1 type of impurity atom, there would be 2^N configurations to calculate. When N is large, which is usually true in real case, the computation is forbidden. Another problem is, if the impurity concentration is low (say < 0.1%), a large super lattice (more than 1,000 atoms) has to be considered to accommodate a single impurity. Such calculation is extremely hard for NEGF-DFT. Therefore, a method that allows us to obtain the average property by only calculating the system once is highly in demand. It is timely that such theories have been developed in 2008 at McGill University [37] and later in 2013 its more general form the NECPA been developed by the same group [38]. The main purpose of this chapter is to present the details of the NECPA



Figure 5.1: Schematic plot of a two-probe system with some impurity sites in the central scattering region. The left/right leads extend to $z = \pm \infty$. The device extends to $x = \pm \infty$ in the transverse direction. The black dots are sites of electrodes, white circles are pure sites in the scattering region, white sites with a cross in the center are impurity sites.

theory in the NEGF-DFT formalism, which we will need in Chapter 7.

The NECPA is formulated by contour-ordered nonequilibrium Green's function where the disorder average is carried out within the coherent potential approximation on the complex-time contour. NECPA has been implemented with LMTO-ASA. One is now able to calculate the average quantum transport property of a random disordered two-probe system from first principles. It was successfully applied to solve the diffusive transport problem in co-doped graphene Tunnel FET (TFET) [51], which will be presented in detail in Chapter 6. One more thing irrelevant to this thesis but worth mentioning is that under the same framework the same group reported a first principles formulation to calculate the fluctuation of quantum transport [128] around the mean and it was successfully applied to solve the variability problem in silicon nanoFET [50].

This chapter will be organized as follows. We first introduce the CPA theory in section 5.1. To make it simple and clean, the formula are derived and presented in the tight binding framework. In section 5.2 we show the non-equilibirum version of the CPA theory by using the analytic continuation method. Section 5.3 will touch

the details of NECPA formalism when the system has periodic boundary conditions. In section 5.4 we present the NECPA formalism with LMTO-ASA implementation. Finally section 5.5 is a short summary of the chapter.

5.1 Coherent Potential Approximation

We start from a two-probe system schematically shown in Fig. 5.1. There are disorder sites randomly located in the scattering region indicated by the crossed circles. Consider a simple but not trivial case that each atom has only one orbital and orbitals are orthonormal with each other, theoretically, one may mimic disorder effects by assigning the on-site energy to a random discrete variable. It is assumed that on a disorder site *i*, the on-site energy ε_i takes the value value ε_{iq} with probability x_{iq} , where q = 1, 2, ... labels multiple impurity species and, clearly, $\sum_q x_{iq} = 1$ is satisfied for any site *i*.

The Hamiltonian of the two-probe system in the second quantization representation can be written as

$$H = H_C + \sum_{\beta = L,R} H_\beta + \sum_{\beta = L,R} H_{C\beta}, \qquad (5.1)$$

$$H_C = \sum_i \varepsilon_i c_i^{\dagger} c_i + \sum_{i < j} t_{ij} c_i^{\dagger} c_j + t_{ij}^* c_j^{\dagger} c_i, \qquad (5.2)$$

$$H_{\beta} = \sum_{k} \epsilon_{\beta k} a^{\dagger}_{\beta k} a_{\beta k}, \qquad (5.3)$$

$$H_{C\beta} = \sum_{ik} t_{ik} c_i^{\dagger} a_{\beta k} + t_{ik}^* a_{\beta k}^{\dagger} c_i, \qquad (5.4)$$

where H_C is the Hamiltonian of the central scattering region (see Fig. 5.1, the region with white circles or circles with a cross in the center), H_{β} ($\beta = L, R$) is the Hamiltonian of the left or right lead, and $H_{C\beta}$ is the coupling between the central scattering region and the β lead. Note that the above Hamiltonian is in a quadratic form, thus analytical solution of quantum transport can be obtained if the on-site energy ε_i is



Figure 5.2: Schematic plot of a isolated systems. (a) isolated system with some random impurities. The white circles are pure sites in the material and the white sites with a cross in the center are impurity sites. (b) isolated system with CPA. The circles in light blue are the atoms with coherent potentials.

a definite variable. The complexity of our problem comes from the fact that ε_i is a random variable and hence any physical quantities must be averaged over disorder configurations.

To address the CPA theory, we first consider the system to be an isolated one (with no leads connecting to the outside reservoirs, see Fig. 5.2(a)). The system Hamiltonian is reduced to Eq. (5.2). The retarded Green's function of the isolated system is calculated by

$$G^{r} = \left(EI - H_{C} + i0^{+}\right)^{-1}, \qquad (5.5)$$

which is similar to Eq. (4.5). I is the identity matrix and comes from the orbital orthonormality. To solve the disorder problem, we need to calculate the configurational average of G^r ,

$$\overline{G^r} = \overline{\left(EI - H_C + i0^+\right)^{-1}},\tag{5.6}$$

where we have use the symbol $\overline{(\cdots)}$ to denote the configurational average. As a basic idea of the CPA theory, the difficult configuration average on $(EI - H_C + i0^+)^{-1}$ can

be taken off when a coherent potential function $\tilde{\varepsilon}^r$ is introduced [129]

$$\overline{G^r} = \left(EI - H_C - \hat{\varepsilon}^r\right)^{-1},\tag{5.7}$$

where we have absorbed the $+i0^+$ term into the coherent potential function. The coherent potential function $\tilde{\varepsilon}^r$ describes an effective medium after configuration average and it retains the translational invariance (thus Fourier transform can be done, even though the disordered crystal has lost its translational invariance). Changing from Eq. (5.6) to (5.7) represents a huge simplification of theory if $\tilde{\varepsilon}^r$ can be calculated. In general the function $\tilde{\varepsilon}^r$ is a site non-diagonal quantity, but its determination can be simplified by employing the following useful approximation [43, 44, 129]

$$\tilde{\varepsilon}_{ii'}^r = \tilde{\varepsilon}_i^r \delta_{i,i'},\tag{5.8}$$

which assumes that the coherent potential $\tilde{\varepsilon}^r$ is a site-diagonal quantity. We further absorb the diagonal part of H_C into the coherent potential and we obtain

$$\overline{G^r} = \left(EI - H_C^0 - \tilde{\varepsilon}^r\right)^{-1},\tag{5.9}$$

where H_C^0 is the off-diagonal part of H_C . Changing from Eq. (5.7) to Eq. (5.9) does nothing but some mathematical rearrangement. Now the configurational averaged Green's function is expressed by an on-site potential, which means an "average" on-site potential is introduced to describe the disordered system (see Fig. 5.2(b)). Instead of ε_{iq} , "average" potential $\tilde{\varepsilon}^r$ describes the on-site potential of the averaged system. The main task of CPA is to determine the unknown quantity $\tilde{\varepsilon}^r$.

We proceed from the Dyson equation [122] that connects Eq. (5.5) and (5.9),

$$G^{r} = \overline{G^{r}} + \overline{G^{r}} \left(\varepsilon^{r} - \tilde{\varepsilon}^{r} \right) G^{r}, \qquad (5.10)$$

where we have done similar modification to Eq. (5.5) with to Eq. (5.9), i.e., extracting

the diagonal part of H_C out and denote it as ε^r .

To proceed, we introduce the T-matrix of the disordered system. By definition the matrix T satisfies the relation [54]

$$\left(\varepsilon^r - \tilde{\varepsilon}^r\right)G^r = T\overline{G^r},\tag{5.11}$$

from which Eq. (5.10) can be further written as

$$G^r = \overline{G^r} + \overline{G^r} T \overline{G^r}, \tag{5.12}$$

which indicates that T matrix contains all disorder scattering effects of the random system. By combining Eqs. (5.11) and (5.12), we find that T also satisfies:

$$T = (\varepsilon^r - \tilde{\varepsilon}^r) + (\varepsilon^r - \tilde{\varepsilon}^r) \overline{G^r} T$$

= $(\varepsilon^r - \tilde{\varepsilon}^r) \left(I + \overline{G^r} T \right).$ (5.13)

To make it clear what physics T contains, we denote $\Delta \varepsilon = \varepsilon^r - \tilde{\varepsilon}^r$. Recursively replacing T matrix in Eq. (5.13) by itself, we arrive at

$$T = \Delta \varepsilon + \Delta \varepsilon \overline{G^r} \Delta \varepsilon + \Delta \varepsilon \overline{G^r} \Delta \varepsilon \overline{G^r} \Delta \varepsilon + \cdots, \qquad (5.14)$$

which clearly indicates that the T matrix contains all possible scattering process likely to happen in the transport of disordered system. Performing the configuration average on Eq. (5.12), we arrive at

$$\overline{T} = 0, \tag{5.15}$$

which provides a clear self-consistent condition for determining the potential function $\tilde{\varepsilon}^r$. However, this process of determining $\tilde{\varepsilon}^r$ is still difficult since T is a matrix that all scattering processes couple with each other. Therefore further approximations are

demanded. Here we will introduce a widely accepted and used method called the single-site approximation (SSA) [54]. In SSA, we consider only the scattering events that involve one impurity at a time. Then similar to Eq. (5.13) the scattering matrix associated with that impurity (at site i) can be written as,

$$t_i^r = \left(\varepsilon_{iq}^r - \tilde{\varepsilon}_i^r\right) + \left(\varepsilon_{iq}^r - \tilde{\varepsilon}_i^r\right) \overline{G_i^r} t_i^r$$
$$= \left(\varepsilon_{iq}^r - \tilde{\varepsilon}_i^r\right) \left(I + \overline{G_i^r} t_i^r\right), \qquad (5.16)$$

from which we have

$$t_{i}^{r} = \left[I - \left(\varepsilon_{iq}^{r} - \tilde{\varepsilon}_{i}^{r}\right)\overline{G_{i}^{r}}\right]^{-1}\left(\varepsilon_{iq}^{r} - \tilde{\varepsilon}_{i}^{r}\right)$$

$$= \left[I - \Delta\varepsilon_{iq}\overline{G_{i}^{r}}\right]^{-1}\Delta\varepsilon_{iq}$$

$$= \left(\varepsilon_{iq}^{r} - \tilde{\varepsilon}_{i}^{r}\right)\left[I - \overline{G_{i}^{r}}\left(\varepsilon_{iq}^{r} - \tilde{\varepsilon}_{i}^{r}\right)\right]^{-1}$$

$$= \Delta\varepsilon_{iq}\left[I - \overline{G_{i}^{r}}\Delta\varepsilon_{iq}\right]^{-1}$$

$$= \left[\left(\varepsilon_{iq}^{r} - \tilde{\varepsilon}_{i}^{r}\right)^{-1} - \overline{G_{i}^{r}}\right]^{-1}.$$
(5.17)

Here we have defined a quantity $\Delta \varepsilon_{iq} = \varepsilon_{iq}^r - \tilde{\varepsilon}_i^r$. $\overline{G_i^r}$ refers to the propagator at site i for the system where all sites are described by the effective medium $\tilde{\varepsilon}^r$. $\tilde{\varepsilon}_i$ is the coherent potential at site i. t_i^r is the single site scattering matrix at site i. We put a superscript r on t_i since it is calculated all by retarded quantities. The next thing we are interested in is the relation between the single site scattering matrix t_i^r and the full scattering matrix T. We try to express the T matrix in terms of single-site scattering matrix t_i^r defined above. It is straight forward to rewrite Eq. (5.13) in form of

$$T = (\varepsilon^{r} - \tilde{\varepsilon}^{r}) + (\varepsilon^{r} - \tilde{\varepsilon}^{r}) \overline{G^{r}}T$$

$$= \sum_{i} (\varepsilon^{r}_{i} - \tilde{\varepsilon}^{r}_{i}) (I + \overline{G^{r}}T)$$

$$= \sum_{i} \Delta \varepsilon_{i} (I + \overline{G^{r}}T) = \sum_{i} Q_{i}, \qquad (5.18)$$

where Q_i is defined as

$$Q_{i} \equiv \Delta \varepsilon_{i} \left(I + \overline{G^{r}}T \right)$$
$$= \Delta \varepsilon_{i} \left(I + \overline{G^{r}}Q_{i} + \overline{G^{r}}\sum_{i' \neq i}Q_{i'} \right).$$
(5.19)

Attention that we have omitted the subscript q for both ε_{iq}^r and $\Delta \varepsilon_{iq}$ since they are not important here. The second equality of Eq. (5.19) comes from replacing T with sum of Q_i . Eqs. (5.17), (5.18) and (5.19) also give

$$Q_i = t_i^r \left(I + \overline{G^r} \sum_{i' \neq i} Q_{i'} \right) = \left(I + \sum_{i' \neq i} Q_{i'} \overline{G^r} \right) t_i^r.$$
(5.20)

 Q_i describes the contribution of random atom at site *i* to the scattering. *T* matrix is different with the single-site scattering matrix t_i^r . This difference is due to multiple impurities scattering which can be shown clearly after inserting expression for $Q_{i'}$ into Eq. (5.20),

$$Q_{i} = t_{i}^{r} \left(I + \overline{G^{r}} \sum_{i' \neq i} Q_{i'} \right) = t_{i}^{r} \left(I + \overline{G^{r}} \sum_{i' \neq i'} t_{i'}^{r} \left(I + \overline{G^{r}} \sum_{i'' \neq i'} Q_{i''} \right) \right)$$
$$= t_{i}^{r} + t_{i} \overline{G^{r}} \sum_{i' \neq i} t_{i'}^{r} + t_{i}^{r} \overline{G^{r}} \sum_{i' \neq i} t_{i'}^{r} \overline{G^{r}} \sum_{i'' \neq i'} Q_{i''}.$$
(5.21)

If we continue to insert expression for $Q_{i''}$ and so forth, we can have

$$Q_{i} = t_{i}^{r} + t_{i}^{r}\overline{G^{r}}\sum_{i'\neq i} t_{i'}^{r} + t_{i}\overline{G^{r}}\sum_{i'\neq i} t_{i'}^{r}\overline{G^{r}}\sum_{i''\neq i'} \left(t_{i''}^{r} + t_{i''}^{r}\overline{G^{r}}\sum_{i''\neq i''}Q_{i'''}\right)$$
$$= t_{i}^{r} + t_{i}\overline{G^{r}}\sum_{i'\neq i} t_{i'}^{r} + t_{i}^{r}\overline{G^{r}}\sum_{i'\neq i} t_{i'}^{r}\overline{G^{r}}\sum_{i''\neq i'} t_{i''}^{r} + \dots \qquad (5.22)$$

Then the multiple scattering equation gives:

$$T = \sum_{i} \left(t_{i}^{r} + t_{i}^{r} \overline{G^{r}} \sum_{i' \neq i} t_{i'}^{r} + t_{i}^{r} \overline{G^{r}} \sum_{i' \neq i} t_{i'}^{r} \overline{G^{r}} \sum_{i'' \neq i'} t_{i''}^{r} + \cdots \right)$$
$$= \sum_{i} t_{i}^{r} + \sum_{i} t_{i}^{r} \overline{G^{r}} \sum_{i' \neq i} t_{i'}^{r} + \sum_{i} t_{i}^{r} \overline{G^{r}} \sum_{i' \neq i} t_{i'}^{r} \overline{G^{r}} \sum_{i'' \neq i'} t_{i''}^{r} + \cdots,$$
(5.23)

where the T is expressed in terms of single-site scattering contributions. Having known the contribution of each site to T, we consider the configurational average of the scattering (average of Eq. (5.20)). To go forward, a single-site approximation [46] is employed,

$$\overline{Q_i} = t_i^r \left(I + \overline{G^r} \sum_{i' \neq i} Q_{i'} \right)$$
$$= \overline{t_i^r} \left(I + \overline{G^r} \sum_{i' \neq i} \overline{Q_{i'}} \right) + \overline{t_i^r \overline{G^r} \sum_{i' \neq i} \left(Q_{i'} - \overline{Q_{i'}} \right)}.$$
(5.24)

In SSA the second term above is neglected hence we arrive at

$$\overline{Q_i} \approx \overline{t_i^r} \left(I + \overline{G^r} \sum_{i' \neq i} \overline{Q_{i'}} \right).$$
(5.25)

The physical meaning of SSA is that when electrons flow through a disordered system, they can only scatter off one impurity at a time. In other words, the scattering events happening at different sites are independent of each other. SSA is a good approximation since the probability of scattering off multiple impurities simultaneously is small. The probability of multiple impurities scattering can be measured by comparing the impurity scattering mean free path (L_{im}) and the average distance between impurities (L_{av}) . SSA beaks when impurity concentration is large so that L_{im} and L_{av} are comparable. Notice that when SSA is employed, a similar expansion to Eq. (5.23) can be derived:

$$\overline{T} \approx \sum_{i} \overline{t_{i}^{r}} + \sum_{i} \overline{t_{i}^{r}} \overline{G^{r}} \sum_{i' \neq i} \overline{t_{i'}^{r}} + \sum_{i} \overline{t_{i}^{r}} \overline{G^{r}} \sum_{i' \neq i} \overline{t_{i'}^{r}} \overline{G^{r}} \sum_{i'' \neq i'} \overline{t_{i''}^{r'}} + \dots \quad (5.26)$$

Recalling the condition Eq. (5.15), Eq. (5.26) vanishing to 0 gives the new selfconsistent condition

$$\overline{t_i^r} = \sum_q x_{iq} t_{iq}^r = 0, \qquad (5.27)$$

where the single site scattering matrix $t_{iq}^r = (\varepsilon_{iq}^r - \tilde{\varepsilon}_i^r) \left[I - \overline{G_i^r} (\varepsilon_{iq}^r - \tilde{\varepsilon}_i^r)\right]^{-1}$ is from Eq. (5.17). Note that we have specified a subscript q to t_i^r and it does nothing but make the information more complete. Eqs. (5.9), (5.17) and (5.27) form a close set of self-consistent equations from which $\tilde{\varepsilon}^r$ can be solved. We rewrite them below

$$\overline{t_i^r} = \sum_q x_{iq} t_{iq}^r = 0,$$

$$t_{iq}^r = \left[\left(\varepsilon_{iq}^r - \tilde{\varepsilon}_i^r \right)^{-1} - \overline{G_i^r} \right]^{-1},$$

$$\overline{G_i^r} = \left[\overline{G^r} \right]_{ii},$$

$$\overline{G^r} = \left(EI - H_C^0 - \tilde{\varepsilon}^r \right)^{-1}.$$
(5.28)

 $[\cdots]_{ii}$ means taking the *i*th diagonal element of the matrix. Eqs. (5.28) are used by most of the CPA work. At the very beginning of the calculation, $\tilde{\varepsilon}_i^r$ is set to be some initial guess (usually the average of ε_{iq}^r). Then $\overline{G_i^r}$ can be obtained from the last two equations of Eqs. (5.28). With $\overline{G_i^r}$ at hand, the first two equations of Eqs. (5.28) together produce a new $\tilde{\varepsilon}_i^r$, which will be used for the next iteration step. This process will be repeated until self-consistency is achieved. Finally the configuration average of Green's function $\overline{G^r}$ is obtained from which all physical quantities can be calculated. However, Eqs. (5.28) are sometimes not convenient to use especially when there are more than two types of atoms possible to occupy the lattice sites where solving $\tilde{\varepsilon}_i^r$ from the first two equations becomes extremely complicated. A simple way to overcome the problem is to rewrite them by using the conditional Greens function defined as

$$\overline{G_{iq}^r} = \left[\left(EI - H_C^0 - \tilde{\varepsilon}_{iq}^r \right)^{-1} \right]_{ii}, \qquad (5.29)$$

where $\tilde{\varepsilon}_{iq}^r$ means to replace the *i*th diagonal element by ε_{iq}^r . The physical meaning of the conditional Greens function \overline{G}_{iq}^r is that the propagator from site *i* to *i* under the condition that the site is occupied by species *q* while all other sites remain disordered. By using the conditional Green's function, the CPA equation set [38] can be rewritten as below,

$$\overline{G_i^r} = \sum_q x_{iq} \overline{G_{iq}^r},$$

$$\overline{G_i^r} = \left[EI - H_C^0 - \tilde{\varepsilon}^r\right]^{-1},$$

$$\overline{G_i^r} = \left[\overline{G^r}\right]_{ii},$$

$$\overline{G_i^r} = \left[EI - \tilde{\varepsilon}_i^r - \Omega_i^r\right]^{-1},$$

$$\overline{G_{iq}^r} = \left[EI - \varepsilon_{iq}^r - \Omega_i^r\right]^{-1}.$$
(5.30)

 Ω_i^r is a site diagonal quantity introduced here to make the equation sets simpler. It describes the effective coupling of a given site *i* to all other sites in a system. Eqs. (5.30) are equivalent to Eqs. (5.28). A simple proof can be found in Appendix D. Eqs. (5.30) is a closed equation set that can be solved self-consistently. This will be discussed in the following section along with its non-equilibrium form. One last thing to mention is that Eqs. (5.30) are obtained for isolated system. For open system (e.g., two-probe system), we can simply add a self energy term Σ to the second line of the above equation set to make them applicable to the open system.

We end this section by summarizing the properties of the CPA, its advantages and limitations: i) CPA creates an effective lattice that describes the average electronic properties of the random system. The lattice site is occupied by an "effective" atom, which is some combination of the host atom and impurity atom. The "effective" atom is described by the site-diagonal coherent potential, which is a complex number or matrix; ii) for system of form $A_x B_{1-x}$, CPA gives the average of 2^N possible configurations if there are N atomic sites in the lattice. This average mathematically coincides with the one that average over all possible configurations having xN A atoms in the lattice; iii) the CPA becomes exact in three important limiting cases namely in the low concentration limit, in the weak scattering limit, and in the split-band limit (the separation of constituents bands is large as compared to their bandwidths) [45]. CPA also provides a good interpolation procedure between these limits, and it is more accurate than any other single-site theory.

5.2 Nonequilibrium Coherent Potential Approximation and Physical Quantities

Section 5.1 derives the CPA equation set for the isolated systems. In current section we go one step further to apply it to non-equilibrium state of a two-probe system (open system). All the details follow Refs. [38, 123]. Firstly we will present in detail the NECPA theory from which the configurational average of Green's functions can be calculated. These quantities are needed for calculating physical quantities such as the nonequilibrium charge density and the transmission coefficient.

5.2.1 Nonequilibrium Coherent Potential Approximation

In this subsection, we start from the Langreth theorem and then apply it to the equilibrium CPA equation set (Eqs. 5.30) to get its nonequilibrium counterpart.

In the equilibrium Green's function theory of solving many body problem, we let the time indices of the Green's function evolve to asymptotic past $(t = -\infty)$ and future $(t = +\infty)$ under the adiabatic process so that the initial and final states of the Green's function are known states. Since the system is in equilibrium and the process is adiabatic, both states are eigen states of the solvable non-interacting Hamiltonian and they differ by a phase factor at most. Therefore the Feynman diagram technique can be applied to solve the many body problem [123].

However, when the system is driven out of equilibrium, the asymptotic past and future states are not necessarily related in most of the cases. Consider, for example, an important problem in surface physics, where atoms or molecules impinging on a surface to exchange charge with the surface, and hence the initial state at $t = -\infty$ is very different from the final state at $t = +\infty$. Such difference from equilibrium case poses difficulties on the application of Feynman diagram technique. Recall that in nonequilibrium problems we are not really interested in the asymptotic future so that the key problem is to get rid of referring to the asymptotic future state. A simple but solid way to do this is the Keldysh contour-ordered Green's function theory (see Fig. 5.29). People find that it is always simpler to consider our problem on the complex time contour instead of the real time axis since contour-ordered Green's function themselves has very similar mathematical structure to the equilibrium ones. In this way, the convenient and powerful Feynman diagram technique can be applied with no difficulties. After solving the complex time contour-ordered Green's function, the final step is to find their corresponding real time ones since all physical quantities and observable are related to the real time Green's functions. This process is done by the Langreth theorem which simply maps the complex contour-ordered Green's functions to the real time Green's functions and is usually called "analytic continuation" [123]. Consider the "matrix products" of two contour-ordered quantities (such as Green's functions or self-energies) C = AB, its real time correspondings can be calculated by [123]

$$(AB)^{<,>} = A^r B^{<,>} + A^{<,>} B^a, (5.31)$$

$$(AB)^{r,a} = A^r B^a, (5.32)$$

where the "matrix multiplication" consists of summations over internal degrees of freedom (space and spin) and convolution integrals on the real axis from $-\infty$ to $+\infty$.



Figure 5.3: Schematic plot of the complex-time contour that goes above the real-time axis from $\tau = -\infty$ to $+\infty$, and returns below the real-time axis to $\tau = -\infty$. The contour-ordered Green's function is defined on the complex-time contour.

Now we derive the non-equilibrium version of the CPA equations (5.30) using the Langreth theorem (5.31) and (5.32). One difficulty of applying the Langreth theorem to the CPA equations (5.30) is that Eqs. (5.30) involves the inverse of some quantities, which does not fit into the form of the Langreth theorem. To overcome this difficulty, we apply the generalized Langreth theorem, which can be found in Ref. [38]. Two sets of equations can be obtained for G^r and $G^<$,

$$\overline{G_i^r} = \sum_q x_{iq} \overline{G_{iq}^r},$$

$$\overline{G_i^r} = \left[EI - H_C^0 - \tilde{\varepsilon}^r - \Sigma^r\right]^{-1},$$

$$\overline{G_i^r} = \left[\overline{G^r}\right]_{ii},$$

$$\overline{G_i^r} = \left[EI - \tilde{\varepsilon}_i^r - \Omega_i^r\right]^{-1},$$

$$\overline{G_{iq}^r} = \left[EI - \varepsilon_{iq}^r - \Omega_i^r\right]^{-1}.$$
(5.33)

and

$$\overline{G_i^{<}} = \sum_{q} x_{iq} \overline{G_{iq}^{<}},$$

$$\overline{G^{<}} = \overline{G^r} \left[\tilde{\varepsilon}^{<} + \Sigma^{<} \right] \overline{G^a},$$

$$\overline{G_i^{<}} = \left[\overline{G^{<}} \right]_{ii},$$

$$\overline{G_i^{<}} = \overline{G_i^r} \left(\tilde{\varepsilon_i^{<}} + \Omega_i^{<} \right) \overline{G_i^a},$$

$$\overline{G_{iq}^{<}} = \overline{G_{iq}^r} \Omega_i^{<} \overline{G_{iq}^a}.$$
(5.34)

The above equation sets (5.33) and (5.34) are the NECPA equations [38].

5.2.2 Self-consistent solution of NECPA equations

In previous subsection we have derived the non-equilibrium version of CPA equations. In this subsection we show the steps of solving these equations. The equation set is solved iteratively and the process mainly consists of two parts: 1) solving $G^{r,a}$ from Eqs. (5.33); 2) solving $G^{<}$ from Eqs. (5.34). Detailed steps are listed as follows [38]. We first solve $G^{r,a}$:

(1) Make an initial guess of Ω^r .

(2) Determine $\tilde{\varepsilon}$ from the first, fourth and fifth lines of Eq. (5.33) and the result is:

$$\tilde{\varepsilon}_{i}^{r} = E - \Omega_{i} - \left[\sum_{q} x_{iq} \left(E - \varepsilon_{iq} - \Omega_{i}^{r}\right)^{-1}\right]^{-1}$$

(3) Determine $\overline{G^r}$ from the second line of Eq. (5.33):

$$\overline{G^r} = \left[E - H_C^0 - \tilde{\varepsilon}^r - \Sigma^r \right]^{-1}.$$

(4) Update Ω^r by solving it from the fourth line of Eq. (5.33) and the result is:

$$\Omega_i^r = E - \tilde{\varepsilon}_i^r - \left(\overline{G^r}\right)_{ii}^{-1}.$$

(5) Go back to step (2) to repeat the process until Ω^r is fully converged. Usually a criteria is set to measure the accuracy of the convergence.

Once Ω^r is solved, $G^<$ can be solved by the following steps:

- (1) Make an initial guess of $\Omega^{<}$.
- (2) Determine $\tilde{\varepsilon}^{<}$ from the first, fourth and fifth lines of Eqs. (5.34) and the result

is:

$$\tilde{\varepsilon}_{i}^{<} = \left(\overline{G_{i}^{r}}\right)^{-1} \left[\sum_{q} x_{iq} \overline{G_{iq}^{r}} \Omega_{i}^{<} \overline{G_{iq}^{a}}\right] \left(\overline{G_{i}^{a}}\right)^{-1} - \Omega_{i}^{<}.$$

(3) Determine $\overline{G^{<}}$ from the second line of Eq. (5.34):

$$\overline{G^{<}} = \overline{G^r} \left[\tilde{\varepsilon}^{<} + \Sigma^{<} \right] \overline{G^a}.$$

(4) Update $\Omega^{<}$ by solving it from the fourth line of Eq. (5.34) and the result is:

$$\Omega_i^{<} = \left(\overline{G_i^r}\right)^{-1} \left[\overline{G^{<}}\right]_{ii} \left(\overline{G_i^a}\right)^{-1} - \tilde{\varepsilon}_i^{<}.$$

(5) Go back to step (2) to repeat the process until $\Omega^{<}$ is converged. Like before, a criteria is set to measure the accuracy of the convergence.

Having shown the procedure of solving the NECPA equation set, there are two more important issues remaining to address. 1) Computational complexity of the above process, i.e., does the computational time scales "reasonably" with the number of sites/atoms or central region length? Step (3) of above process is proportional to N^3 , where N is the number of disorder sites. In Chapter 7, we will show that by using some special matrix algorithm, the computational time can be drastically reduced, making it viable for real material application. 2) How to accelerate the converging process. In Chapter 7, we show that the Anderson method [54] largely accelerate the converging speed.

5.2.3 Charge Density and Transmission Coefficient

Recall that in Chapter 4, the key step of the NEGF-DFT is to calculate the system charge density ρ , which is calculated from $G^{<}$. Similarly, for the disorder system, the average charge density is calculated by

$$\overline{\rho} = -\frac{i}{2\pi} \int_{-\infty}^{\infty} \overline{G^{<}}(\varepsilon) \, d\varepsilon, \qquad (5.35)$$

where the $\overline{G^{<}}$ is obtained from solving the NECPA equation set (5.33) and (5.34).

As discussed in Section 4.1 the transmission coefficient is calculated from Eq. (4.2). Our goal is to obtain its disorder configuration average

$$\overline{T(E)} = \overline{\operatorname{Tr}\left(\Gamma_L G^r \Gamma_R G^a\right)}.$$
(5.36)

This quantity can be calculated through a somewhat tricky approach. We go from Eq. (5.36)

$$\overline{T(E)} = \overline{\operatorname{Tr}\left(\Gamma_L G^r \Gamma_R G^a\right)} = \operatorname{Tr}\left(\overline{\Gamma_L G^r \Gamma_R G^a}\right)$$
$$= \operatorname{Tr}\left(\Gamma_L \overline{G^r \Gamma_R G^a}\right).$$
(5.37)

Since the linewidth function Γ_L has been solved from the electrodes (see Eq. (4.9)), we need to calculate the quantity $\overline{G^r \Gamma_R G^a}$ in which two random quantities G^r and G^a are connected by a non-random quantity Γ_R . This is much alike what we have done for solving $\overline{G^{<}}$ in section 5.2.1. From a mathematical point of view, $\overline{G^{<}}$ and $\overline{G^r \Gamma_R G^a}$ have identical structure. To solve the quantity $\overline{G^r \Gamma_R G^a}$, we only need to replace the quantity $\Sigma^{<}$ by Γ_R in the NECPA equations of $\overline{G^{<}}$ [38]. Once the corresponding NECPA equation is solved, we just need to insert $\overline{G^r \Gamma_R G^a}$ back to Eq. (5.37) to get the final transmission coefficient.

5.3 NECPA with Transverse Periodicity

In some applications, one may need to study two-probe systems whose dimension is infinite in the transverse direction (see Fig. 5.1). The way to deal with such infinity is to apply periodic boundary conditions originating from the lattice periodicity. For two-probe systems without disorder, the transverse periodicity allows one to apply the Bloch theorem. This way, the calculation of transverse periodic two-probe system is reduced to the calculation of a small unit cell with appropriate k sampling. For two-probe systems with random disorder, the translational symmetry is broken in the transverse dimensions and Bloch theorem does not hold. Nevertheless, NECPA is an effective medium theory whose application restores the translational symmetry of G^r and $G^<$. One can still work with the small unit cell with k sampling to calculate these quantities.

For two-probe systems with transverse periodicity, NECPA equations (5.33) and (5.34) need to be modified slightly to include k sampling:

$$\overline{G_i^r} = \sum_q x_{iq} \overline{G_{iq}^r},$$

$$\overline{G^r}(k) = \left[EI - H_C^0(k) - \tilde{\varepsilon}^r - \Sigma^r(k)\right]^{-1},$$

$$\overline{G^r} = \int_{-\infty}^{+\infty} \frac{dk}{2\pi} \overline{G^r}(k),$$

$$\overline{G_i^r} = \left[\overline{G^r}\right]_{ii},$$

$$\overline{G_i^r} = \left[E - \tilde{\varepsilon_i^r} - \Omega_i^r\right]^{-1},$$

$$\overline{G_{iq}^r} = \left[E - \epsilon_{iq} - \Omega_i^r\right]^{-1}.$$
(5.38)

and

$$\overline{G_i^{<}} = \sum_q x_{iq} \overline{G_{iq}^{<}},$$

$$\overline{G^{<}}(k) = \overline{G^r}(k) \left[\tilde{\varepsilon}^{<} + \Sigma^{<}(k) \right] \overline{G^a}(k),$$

$$\overline{G^{<}} = \int_{-\infty}^{+\infty} \frac{dk}{2\pi} \overline{G^{<}}(k),$$

$$\overline{G_i^{<}} = \left[\overline{G^{<}} \right]_{ii},$$

$$\overline{G_i^{<}} = \overline{G_i^r} \left(\tilde{\varepsilon}_i^{<} + \Omega_i^{<} \right) \overline{G_i^a},$$

$$\overline{G_{iq}^{<}} = \overline{G_{iq}^r} \Omega_i^{<} \overline{G_{iq}^a}.$$
(5.39)

In the above equations, k is the dimensionless wave vector. For systems with periodicity in one transverse dimension (i.e. 2D system), k is defined as $\mathbf{k} \cdot \mathbf{a}$ in which \mathbf{k} is the wave vector and \mathbf{a} is the unit cell vector of the periodic transverse dimension. For systems with periodicity in two transverse dimensions (3D system), k is defined as $(\mathbf{k_1}, \mathbf{k_2}) = (\mathbf{k} \cdot \mathbf{a_1}, \mathbf{k} \cdot \mathbf{a_2})$ in which $\mathbf{a_1}$ and $\mathbf{a_2}$ are the two unit-cell vectors of the periodic transverse directions and $\int_{-\infty}^{+\infty} \frac{dk}{2\pi}$ should be understood as $\int_{-\infty}^{+\infty} \frac{dk_1}{2\pi} \int_{-\infty}^{+\infty} \frac{dk_2}{2\pi}$.

Finally, we end the tight binding CPA and NECPA section by mentioning that the above equations can be easily extended to the multiple orbital case. On-site quantities like $\tilde{\varepsilon}_i^r$, ε_{iq}^r and so on become $N_o \times N_o$ matrix blocks, where N_o is the number of orbitals of each site.

5.4 NECPA with LMTO-ASA

In section 5.1 and 5.2, we have presented the details of the CPA and NECPA formalism in the tight binding framework. In practical implementation, the formalism is usually implemented with LMTO-ASA [54] or KKR method [130], under which the system Hamiltonian has no off-site disorders. In this section, we present NECPA formalism in the framework of LMTO-ASA (see section 2.3 for its detail). Since the basic formulas follow the ones presented in section 5.1 and 5.2, we only address some key difference compared to the tight binding ones.

The biggest difference of LMTO-ASA implementation is the Green's function. As presented in section 4.4, in LMTO-ASA both Green's functions (G^r and $G^<$) are expressed by the auxiliary Green's functions (see Eq. (4.23) and Eq. (4.27)). The CPA is applied to the auxiliary Green's function [54] (see Eq. (4.22)) by

$$\overline{g^{\alpha}} = \left(\mathcal{P}^{\alpha} - S^{\alpha}\right)^{-1},\tag{5.40}$$

where $\mathcal{P}^{\alpha}_{\mathbf{R},\mathbf{R}'} = \mathcal{P}^{\alpha}_{\mathbf{R}} \delta_{\mathbf{R},\mathbf{R}'}$ [54, 129], is the introduced on-site coherent potential in

LMTO-ASA, like what is done in Eq. (5.7). Other quantities in G^r and $G^<$ are generally on-site, whose configurational average can be performed easily. Once the configurational average Green's functions are solved, everything else can be solved following similar formulas as in the tight binding. Interested readers can refer to my master's thesis [131] for more details on the formulas. The NECPA equation set in LMTO-ASA is not listed in Ref. [131], so we put them in Appendix F of this thesis.

5.5 Summary

In summary, this chapter has introduced theories for solving quantum transport in systems having random disorder impurities. We first presented the tight binding CPA formalism that can be used to predict configuration averaged quantities at the single particle Green's function and Hamiltonian level based on the Green's function theory derived in Chapter 4. Then by using the analytic continuation theory, we extend the CPA equations to its nonequilibrium form, the NECPA formalism for calculating the configuration average of the distribution Green's function $G^{<}$ and the transmission coefficient which involved the average over two Green's function correlators. Lastly, we briefly discussed their implementation in LMTO-ASA. Experience has shown that LMTO-ASA form of NECPA is difficult to use when the lattice is not close packed. In Chapter 7, we will discuss its implementation in LCAO by applying a futher approximation.
6

Application: Diffusive Transport in Graphene Tunnel-FET

As the device size scales down to sub-100 nm regime, modeling of emerging and nanoscale electronic devices requires quantum models beyond the capability of traditional technology computer aided design (TCAD) methods [132]. An important consideration in device modeling is how to deal with the atomic scale disorder. Examples include the random grain boundary scattering in copper and graphene [133] leading to significantly higher resistance of the material; the discrete atomic dopants scattering in transition metal dichalgogenides that changes the contact resistance with metal electrodes [134, 135]; the device-to-device variability due to fluctuations of random dopants [9]. Experimental results have demonstrated that atomic disorder can seriously affect and may even dominate device performance [133, 134, 135, 136] if not controlled. Traditionally there are several methods in device physics to simulate atomic disorder effects including the density gradient correction in the drift-diffusion model [137], the random-alloy model in the tight-binding approach for computing band structures of alloys [138], and so on. Microscopically, disordered impurities interacts with the host material to alter its electronic property so it is desirable to simulate disorder effects from first principles that determine the atomic potential selfconsistently and parameter-free. For disorder-free device structures, first principles transport methodology has been realized by carrying out the NEGF-DFT method [27]. With disorder, however, repetitive computation of many disorder configurations by NEGF-DFT in order to perform disorder averaging, is a bottleneck which has so far prevented realistic devices (e.g. with inevitable disorders) to be simulated by first principles. The NECPA-LMTO-ASA method presented in Chapter 5 overcomes this bottleneck.

Graphene (Fig. 6.1(a)) has been studied intensively due to its impressive electronic property such as extremely high carrier mobility (over 10,000 $cm^2 \cdot V^{-1} \cdot s^{-1}$ for suspended samples) [139]. However graphene has no band gap. For electronic devices application such as the FET, it is essential to open up a band gap in graphene to realize the on/off switch function. One way to open a band gap is by the quantum confinement effect, e.g., tailoring graphene into one-dimensional nanoribbons [140, 141, 142, 143]. In experiments, to achieve even a small band gap of $\sim 0.2 \, \text{eV}$ (on/off ratio >10²) requires the nanoribbons to be as narrow as ~10 nm, which is challenging to achieve [141, 142, 143] and difficult for large-scale production. Moreover, the carrier mobility of graphene nanoribbons (about hundreds of $cm^2 \cdot V^{-1} \cdot s^{-1}$) [142, 143] is several orders of magnitudes lower than that of a graphene sheet due to the (intrinsic) band folding and phonon scattering [144] as well as (extrinsic) difficulty in controlling the edge roughness [142, 143]. Here we consider another way to open up a band gap in graphene by breaking the inversion symmetry of the A, B sublattices of graphene (see Fig. 6.1(b)), e.g., by applying different potentials onto the two sublattices. This can be done by either placing the graphene sheet onto materials lacking inversion symmetry such as aluminum oxide (Al_2O_3) and boron nitride [145, 146, 147]or co-doping the graphene with both boron and nitrogen [148, 149, 150]. However, the band gap introduced by the former method is an extrinsic one and highly relies on the experimental process, which poses inevitable barriers on its application in FETs. The latter method, although introduces a intrinsic band gap in graphene, encounters the variability problem induced by the random dopants. The random positions of dopants in the graphene lattice cause unpredictable randomness on the band gap as well as transport. The feasibility of employing this method to build transistor remains unknown. Another difficulty was that up to the publication of our paper (Ref. [51]), there was no work studying transistor properties using a parameterfree, *ab initio* method with disorder scattering to the best of our knowledge. In this



Figure 6.1: Lattice structure of graphene. (a) plots the top view of the graphene primitive cell. (b) plots the structure of graphene 4×4 super cell. The green (grey) sites refer to the graphene A (B) sublattice.

chapter, we investigate the influence of impurities scattering on the performance of tunnel-FET (TFET) [151] made from boron-nitrogen (B-N) co-dopded graphene [51] by using the advanced parameter-free NECPA-LMTO-ASA method implemented in the NanoDsim package [152].

6.1 Device Structure

Fig. 6.2 shows a schematic of B-N co-doped graphene TFET. It consists of a channel material (the graphene part as shown by the grey spheres), two metal contacts (copper in the figure) and a gate dielectric material (Al_2O_3 in the figure). The channel in between two contacts is called "the active region" in device modeling. In real devices, dopants are introduced to the channel so that the device can be switched between "on" and "off" states [153]. Such random dopants strongly affect the electronic characteristics in both the active region and also the contacts (as shown in the two sub-figures in Fig. 6.2). For quantitative modeling, multiple disorder scatterings of carriers must be considered in calculating the electronic structure and quantum transport. For device applications, the two most significant problems are the contact resistivity between the metal and semiconductor, and the electronic transport property in the channel when both bias and gate voltages are applied. The two problems can be investigated



Figure 6.2: Schematic of the co-doped graphene TFET. The top figure shows the side view of the TFET atomic structure. The channel is made of co-doped graphene and is attached to metal gate through a layer of Al_2O_3 dielectric material. The source/drain are connected to the outside via copper contacts. The red box highlights the part to simulate. The bottom left figure shows the top view of the graphene/copper interface. The bottom right figure shows the atomic structure of the co-doped graphene. Grey spheres refer to carbon atoms and green (silver) spheres refer to boron (nitrogen) dopants.

separately - the former by KS-DFT and the latter by NEGF-DFT. Here we focus on the latter problem, i.e., the electronic transport in the channel. The structure to simulate is highlighted by the red box in Fig. 6.2 and it has an n-i-p junction structure, where the source (the graphene under the left metal in Fig. 6.2) is n-type doped, the channel is intrinsic and the drain (the graphene under the right metal in Fig. 6.2) is p-type doped.

6.2 Disordered Electronic Materials

We first investigate the electronic properties of the disordered graphene. To introduce a gap suitable for TFET application, the inversion symmetry of graphene lattice has to be broken. Electronic and atomic structures of the device material are calculated by KS-DFT total energy relaxation. Again, when there are impurity atoms or other types of disorder, the CPA method is applied and one obtains the averaged properties (e.g. density of states, retarded Green's functions, density matrix elements, etc.) of the disordered material. To see the essence of CPA we consider the B-N co-doped



Figure 6.3: Schematic figures for co-doped graphene. The grey spheres refer to carbon host atoms and the green (silver) spheres correspond to doped boron (nitrogen) atoms. (a) A particular configuration of B-N substitution co-doping in graphene where the co-doping concentration is x%. (b) A top view of the effective medium of co-doped graphene. The effective medium by CPA where a site has x% probability to be impurity and (1-x)% probability to be the host atom. The dopant atoms are plotted with a much smaller radius to make the host atoms (carbon) visible from top view.

graphene where boron (B) and nitrogen (N) are impurity atoms. Co-doping means the concentration of B and N is the same. Fig. 6.3(a) plots one of possible atomic configurations for a given B-N concentration x. The electronic structure varies from one configuration to another, but their average is well represented by the effective medium model of CPA in Fig. 6.3(b) where boron/nitrogen occupies the A/B site of graphene with a probability of x%.

Fig. 6.4(a) shows the the calculated band structure of pristine graphene by NanoDsim transport package [152]. The band structure is plotted along K-G-M-K path in the reciprocal space. A maximum angular momentum of 2 is used for all carbon, boron and nitrogen atoms. The calculated band structure agrees very well with the results from plane wave methods (e.g., VASP program), which indicates a proper setting of all LMTO-ASA parameters. Fig. 6.4(b) plots the disorder averaged CPA band structure [152] of a 10% B-N co-doped graphene. With B-N co-doping which breaks the inversion symmetry of graphene lattice, a band gap of 0.33 eV is opened at graphene Dirac cone and the band edge electron mass is 0.1 m₀ (m₀ is the electron bare mass). Note that the CPA bands have a finite width (Fig. 6.4(b)) which is the disorder broadening. These material properties can also be obtained by brute force KS-DFT calculations of many disorder configurations, as we verified by the VASP



Figure 6.4: Electronic band structure of graphene. (a) Band structure of pristine graphene. (b) CPA band of disordered graphene with B-N co-doping at 10%. The CPA bands has a finite width reflecting the disorder broadening. A band gap opened around Dirac cone is observed. Fermi levels are fixed at "0" eV for both figures.

program. In VASP verification, we have randomly generated 20 samples of co-doped graphene, each of which contains 60 atoms. The average band gap is 0.31 eV, which comparable with the CPA result 0.33 eV.

We then calculate the density of states (DOS) of the disordered system. DOS of co-doped graphene is shown in Fig. 6.5(a). The band gap of the co-doped graphene increases with the doping concentration. An interesting result is that the band gap linearly increases with co-doping concentration (as shown in the inset of Fig. 6.5(a)), which provides a feasible way to control the graphene band gap. To make the co-doped graphene p-type or n-type (as required by device engineering), additional boron or nitrogen atoms are introduced to the lattice on both the A/B sites, the corresponding DOS at 1% additional doping is shown in Fig. 6.5(b) where the Fermi level shifts up or down in energy according to the doping type, as expected.



Figure 6.5: Density of states (DOS) of the co-doped graphene. (a) DOS versus energy for several co-doping concentrations. A near linear dependence of the average band gaps on the concentration in the interested range is showed in the inset. (b) The averaged DOS versus energy of the B-N co-doped graphene with extra individual boron (p-type) or nitrogen (n-type) impurity atoms. Fermi levels move accordingly to the doping type, as expected. The change in band gap is very small with the extra individual dopants as investigated.

6.3 Transport with Disorder Scattering

Having understood the electronic properties of the disorderd material, we investigate the transport property of TFET built on the co-doped material. The inputs of our quantum transport modeling are solely atomic positions of the device structure and the applied voltages (both source/drain bias and gate bias). We consider double-gate TFET [151] structures whose top view is shown in Fig. 6.6. On the left, the disorder



Figure 6.6: Top views of the simulated structure of the double-gate TFET (the active region). The leads provide self-energy to the central region in the NEGF-DFT self-consistent analysis. (a) The structure where doping is done by VCA. (b) The structure where CPA accounts for multiple disorder scattering.



Figure 6.7: Band diagrams of the TFET along the channel. (a) Energy profiles and Fermi level in the (B-N) co-doped graphene TFET obtained from the first principle simulation method with NECPA. Scattering region I includes (B-N) co-doping and extra p-type doping, and region II includes the co-doping. (b) LDOS of the co-doped graphene TFET in Fig. 6.6(b). The interband tunneling is clearly seen from LDOS in the band gap. A direct tunneling current is also observed.

is treated by the widely used virtual crystal approximation (VCA) which essentially shifts the electronic potential due to charge contributions of impurity atoms [154]. On the right, the disorder is treated by NECPA presented previously that accounts for multiple impurity scattering in addition to the potential change. The two structures shown in Fig. 6.6 serve as a comparison between the VCA and NECPA. A B-N co-doped graphene TFET shown in Fig. 6.6(b) includes the whole channel and part of source/drain treated by NECPA, and we consider practical dimensions where the channel length $L_g=10$ nm, gate dielectric thickness EOT=1.7 nm [153, 155]. In the simulation, the width of the FET is considered infinite (since graphene is a film) which we treat by using periodic boundary conditions in the NECPA-LMTO-ASA analysis. Devices containing 2160 atoms are calculated with the VCA method or 4080 atoms with the NECPA method. The gate modulation on transport is included by solving the real space three-dimensional Poisson equation with the Dirichlet boundary condition at the gate electrodes and floating boundary condition (i.e. Neumann boundary condition) at other surfaces [156].

Before calculating transport, we determine the electrostatic profile of the TFET, which helps us to establish a better understand the TFET physics. The calculated



Figure 6.8: Transport properties of the simulated co-doped graphene TFET. (a) Transmissions of the modeled graphene TFET with VCA doping and CPA doping. The significant decrease in the NECPA results is caused by the diffusive inter-band tunneling due to the strong disorder scattering. (b) Drain current of the 10 nm long graphene TFET with (blue circles) and without (black squares) disorder scattering. Disorder scattering dramatically reduces the inter-band tunneling current.

average band profile under $V_{ds}=0$ (V_{ds} : the bias between the source and the drain) is plotted in Fig. 6.7(a). The band profile is calculated by plotting local density of states (LDOS) projected onto the real space grids. The potential changes within the VCA and NECPA themes are matched excellently as confirmed by the smooth energy profile. Region I indicates the source depletion while region II is the non-inverted channel of the co-doped graphene TFET which are both treated with NECPA. The interband tunneling process across the source/channel junction in the simulated TFET is diffusive and the atomic disorder scattering comes from the co-doping as well as the extra source doping. The transmission spectrum given by NECPA is shown in Fig. 6.8(a) which clearly confirms the inter-band tunneling physics, for example the transmission peak is achieved at the energy level corresponding almost to the smallest tunneling distance.

The device structure constructed as Fig. 6.6 (a) with VCA doping is also simulated. Results show that disorder scattering plays a very significant role as indicated by the NECPA curve (black) which represents a much smaller transmission as compared to that of VCA curve where disorder scattering is absent. Microscopically, the decrease in transmissions comes from the diffusive scattering in the tunneling process by the atomic disorders. It is the first evidence that the inter-band tunneling is affected by the atomic disorders which should be taken into consideration in TFET modeling. The transmission peak only moves slightly in the tunneling energy window which is not affected by the disorder scattering. The local density of states (LDOS) is plotted in Fig. 7.17(b) in which the inter-band tunneling is clearly seen in the band gap. In the 10 nm TFET we also observe a direct tunneling current flowing across the whole channel which is also affected by the disorder scattering.

Finally, non-equilibrium quantum transport of TFETs in Fig. 6.6(a) or (b) has also been obtained from first principles. Transfer characteristics of the TFETs are calculated and disorder scattering reduces inter-band tunneling current as seen in Fig. 6.8(b). Even with scattering the driving current of the co-doped graphene TFET is in the range of mA/ μ m due to the small band gap and effective mass. We also observe that such disorder scattering has larger influence on the off-state current than on-state current as shown in Fig. 6.6. It indicates an increase of 20% in the subthreshold slope [153] due to disorder scattering.

6.4 Summary

We have applied the NECPA-LMTO-ASA method to simulate a graphene TFET by co-doping B-N impurity atoms. The co-doping is found to open a substantial gap that linearly scales with the co-doping concentration which is appropriate for making TFET and/or other transistor applications. We also observe that the diffusive disorder scattering in the graphene TFET reduce the band-to-band tunneling current by a substantial factor. We note that even if the potential change due to doping is accounted for (e.g. by VCA), it is not adequate to obtain correct transport result if impurity scattering is not explicitly included. Finally, because NECPA-LMTO-ASA is a parameter-free first principles modeling method, with only a linear scalability of its computational cost (along the transport direction), the methodology provides a promising TCAD solution to determine disorder induced physical effects.

7

NECPA with LCAO Basis

In Chapter 6, we have shown a practical application of the NECPA-LMTO-ASA method [51]. Other applications can be found in Refs. [37, 36, 50, 157, 158, 159]. The NECPA-LMTO-ASA is a very successful implementation version of the CPA theory to solve material and quantum transport problems with atomistic disorders. However, the NECPA-LMTO-ASA has several drawbacks. First, it has requirements on the lattice structures. In principle, LMTO-ASA is solely suitable for close-packed lattices, which indicates that if the lattice structure is NOT tightly packed in space, LMTO-ASA can NOT produce the correct eigen-states of the systems. Examples are, for example, 2D materials, surfaces and interfaces. In practical applications, this problem is achieved by using vacuum spheres in space to make the structure closepacked. Fig. 7.1 illustrates two examples of ASA for non-close-packed lattices. Fig. 7.1(a) shows the ASA for a graphene lattice. The lattice consists 4 carbon atoms (in black) and 20 vacuum spheres (in green and blue). The 24 spheres together form a hexagonal close-packed (HCP) structure. Similarly, Fig. 7.1(b) (taken from Ref. [157]) shows the ASA for monolayer black phosphorus. Such complicated vacuum space filling has to be worked out for each problem. Second, all these vacuum spheres have flexible radii and tunable spacial positions, which have to be determined by fitting the correct band structures of the investigated system. Last, the incredibly large number of vacuum spheres drastically reduce the efficiency of computation. For example, in graphene lattice, 83% of the atomic spheres are vacuums.

In the literature, there have been several attempts to implement CPA besides



Figure 7.1: Illustrating ASA in some LMTO-ASA calculations. Vacuum spheres are labeled by "Vac". Different vacuum spheres are labeled with subscript numbers. (a) ASA for graphene lattice. It has been used in Chapter 6; (b) ASA for monolayer black phosphorus lattice. The spheres are slightly overlapped in space. The radii of all spheres has been reduced in the figure to make the figure less crowded.

LMTO-ASA. The Korringa-Kohn-Rostoker (KKR) is one of the earliest methods to implement CPA [130]. In KKR, the muffin-tin approximation (see section 2.3.2) is used for the orbitals. Without ASA, the spheres filling is easier than LMTO-ASA but the formalism is very complicated so that NEGF is difficult to implement. The Blackman-Esterling-Berk (BEB) theory is another attempt to implement CPA [52]. It uses LCAO as the basis. The off-diagonal disorders are projected onto a higher dimensional Hamiltonian space so that they occupy only the diagonal part of the new Hamiltonian. However, it appears very difficult to reduce the BEB theory to implement NECPA. In this chapter, we develop a straight forward LCAO-based CPA that avoids the complexity of both the KKR and the BEB theory. The difficulty of applying CPA directly in LCAO is that the quantity (ES - H) has off-diagonal disorders, where S is the basis overlap matrix of the LCAO and H is the system Hamiltonian. The off-diagonal disorders come from two aspects: 1) S, when the atomic orbitals are nonorthogonal. 2) H, which in most cases has disorder dependent off-diagonal parts. These off-diagonal disorders place obstacles in the path of applying CPA in LCAO first principles calculation since the CPA method is only compatible with onsite disorders (see Chapter 5). The basic ideal of our method is that we neglect the off-diagonal disorder and consider the diagonal disorder only. This becomes possible from the fact that in most semiconductor problems, the concentration of impurities are typically very low. Take silicon as an example, the highest doping concentration is typically below 10^{20} cm⁻³ (i.e. <1% in percentage). We show that at such low number of impurities in the lattice, off-diagonal disorder is not important and can be neglected.

This chapter will be organized as follows. In section 7.1, we verify the validity of the above assumption that the off-diagonal disorder can be neglected when evaluating the transmission at low concentration by a tight binding model. Section 7.2 is devoted to the DFT implementation of our method. We show in detail of several important issues of our implementation, such as the determination of impurity Hamiltonian, the computational procedure, matrix algorithm and convergence acceleration. We use boron doped graphene as an example to verify the validity of our method. Section 7.3 is an application of our method to study the dopant limited mobility of monolayer MoS_2 . Finally section 7.4 is a short summary of the chapter.

7.1 Tight Binding Verification

We have assumed that the off-diagonal disorder can be neglected when the number of impurities in the lattice is low. We name it as "low concentration approximation" (LCA) in this thesis. In this section we use simple tight binding models to test the LCA and demonstrate its numerical accuracy by comparing the transmission coefficient calculated by NECPA with LCA (NECPA-LCA) and by "brute force" method. Brute force method means we calculated the result for each individual disorder configuration and then take their statistical average. We do this for a 1D tight binding chain.



Figure 7.2: Schematic plot of 1D two-probe system with some impurity sites in the central scattering region investigated by tight binding. The left/right leads extend to $z = \pm \infty$. The black dots are sites of leads and grey circles are disorder sites in the scattering region.

Fig. 7.2 illustrates the 1D chain. In the model, the leads extends to $z \pm \infty$ and contains two scatterers in the central region. The black dots represent clean sites having on-site energy ε_0 ; the gray dots represent the disorder sites having on-site energy ε_i which is a discrete random variable taking values ε_{iq} $(q = 0, 1, 2, \dots)$ with probability x_{iq} . For simplicity but without loss of generality, only nearest neighbors have interactions with a coupling strength $t_{q_1q_2}$, which is a specie dependent quantity. For NECPA-LCA calculation, disorder averaged Green's functions are calculated by solving Eqs. (5.33) and (5.34) with the off-diagonal elements t setting as the coupling between host atoms. For brute force enumeration, disorder-averaged Green's functions are calculated directly from its definition, namely,

$$\overline{G^{r}} = \sum_{q_{1}} \sum_{q_{2}} x_{1q_{1}} x_{2q_{2}} G^{r}_{q_{1}q_{2}},$$

$$G^{r}_{q_{1}q_{2}} = \left[E - \begin{pmatrix} \varepsilon_{1q_{1}} & t_{q_{1}q_{2}} \\ t_{q_{1}q_{2}} & \varepsilon_{2q_{2}} \end{pmatrix} - \begin{pmatrix} \Sigma_{0}^{r} & 0 \\ 0 & \Sigma_{0}^{r} \end{pmatrix} \right]^{-1},$$

$$\overline{G^{<}} = \sum_{q_{1}} \sum_{q_{2}} x_{1q_{1}} x_{2q_{2}} G^{<}_{q_{1}q_{2}},$$

$$G^{<}_{q_{1}q_{2}} = G^{r}_{q_{1}q_{2}} \left(\begin{pmatrix} if_{L}\Gamma_{0} & 0 \\ 0 & if_{R}\Gamma_{0} \end{pmatrix} \right) G^{a}_{q_{1}q_{2}},$$

in which $f_{L,R}$ are Fermi functions of the left and right electrodes, $\Gamma_0 = -2 \text{Im} \Sigma_0^r$ is the linewidth function of the leads. Σ_0^r is the retarded self-energy of the lead which can be evaluated analytically for the semi-infinite 1D chain [38]

$$\Sigma_0^r = \xi \left(\frac{E + i0^+ - \epsilon_0}{t} \right) t,$$

where

$$\xi(x) = \frac{z - i\sqrt{4 - z^2}}{2},$$

in which the branch of the square root is chosen as $\operatorname{Re}\sqrt{z} > 0$.

For numerical calculation, we consider 2 disorder types on each disorder site. We use '0" to denote the host atoms (the one has larger concentration) and "1" to denote the "impurity". The host on-site energy is set to 0 ($\varepsilon_0 = \varepsilon_{10} = \varepsilon_{20} = 0$) and the impurity on-site energy is set to 1 ($\varepsilon_1 = \varepsilon_{11} = \varepsilon_{21} = 1$). The coupling (off-diagonal element) between host sites is set to 1 ($t_{00} = 1$). To make the problem simple, we assume the coupling between the host site and impurity site equals the coupling between impurity sites ($t_{01} = t_{11}$). In the simulation, we choose $t_{11} = 0.5, 0.9, 1, 1.1, 1.5$ and the impurity concentration x = 0.1, 0.05, 0.01, 0.001.

The simulated results are showed in Fig. 7.3. The left column of figures are the full plots of the transmission as a function of energy. The right column ones are their corresponding zooming-in's. We first focus on Fig. 7.3(a) and (b). The red, blue, purple and green points and curves plot the brute force calculation results when the off-diagonal disorder $t_{11} = 0.5, 0.9, 1.1, 1.5$ respectively. The black points and curve are the result calculated by NECPA-LCA, where the off-diagonal disorder is neglected (i.e., $t_{11} = t_{00} = 1$) as required by the LCA. We see that the NECPA-LCA result deviates from the brute force results by some amount. The smaller $|t_{11} - t_{00}|$ is, the more accurate NECPA-LCA result is. We mention that the brute force result for $t_{11} = 1$ coincides with the NECPA-LCA curve since NECPA-LCA reduces to the normal NECPA in this case. As the impurity concentration decreases (Fig. 7.3 from top to bottom), the deviation becomes smaller for all t_{11} values.

Fig. 7.4 shows the transmission difference (ΔT) as a function of both the impurity



Figure 7.3: Transmission coefficient vs. energy. The energy is in arbitrary unit. The left column figures are full plots and the right column are the zoomed in ones in proper energy range. The black dots and curve refer to the result of NECPA-LCA. The other shapes and colours are brute force calculation for different disorder strengths t_{11} ranging from 0.5 to 1.5. From up to down, the concentration varies from 0.1 to 0.001, as shown in (a), (c), (e) and (g).



Figure 7.4: Transmission difference (ΔT) vs. impurity coupling (t_{11}) and impurity concentration (x). (a) Dependence of ΔT on t_{11} . The blue, red, yellow and purple curves correspond to x = 0.1, 0.05, 0.01, 0.001 respectively. (b) Dependence of ΔT on x. The blue, red, yellow, purple curves correspond to $t_{11} = 0.5, 0.9, 0.11, 1.5$ respectively.

coupling t_{11} and impurity concentration x. The transmission difference is defined as the average difference between the NECPA-LCA result and the brute force result on some specified energy range:

$$\Delta T = \frac{1}{|E_1 - E_2|} \int_{E_1}^{E_2} |T_{\text{NECPA-LCA}}(E) - T_{\text{BF}}(E)| dE, \qquad (7.1)$$

where $T_{\text{NECPA-LCA}}$ and T_{BF} are transmission coefficients calculated by NECPA-LCA method and brute force method accordingly. $[E_1, E_2]$ is the energy range where the transmission is non-zero. ΔT defines how far away the NECPA-LCA result is to the brute force (exact) result. Fig. 7.4 clearly summarizes the results shown in Fig. 7.3. An interesting phenomenon observed in Fig. 7.4(b) is a linear dependence of ΔT on x. We conclude that 1) for any given impurity concentration, the closer the impurity coupling is to the host coupling, the more accurate the NECPA-LCA is; 2) for any given impurity coupling, the lower the impurity concentration is, the more accurate the NECPA-LCA is. Therefore, in practical applications, one has to check if the following is satisfied before applying NECPA-LCA: 1) the concentration is low; 2) impurity coupling is close to the host coupling. For 1), according to the tight binding test, the concentration should be lower than 1%, which is typically satisfied for semiconductors. For 2), in practical material calculations, t_{00} and t_{11} are matrices instead of single numbers. A simple but useful way to measure if they are close to each other is to compare the average value of all matrix elements.

7.2 DFT Implementation

Having understood the accuracy of the NECPA-LCA with tight binding models, we now implement it within DFT using the LCAO basis.

There are several issues in the LCAO implementation. First, all the matrix elements (quantities with subscript i or iq in Eqs. (5.33) and (5.34)) become matrix blocks instead of single numbers. The block size equals the number of orbitals considered in LCAO basis for each element. Take the carbon atom as an example, in our implementation, s, p and d orbitals are considered with double- ζ polarization [80]. There are in total 13 orbitals for carbon atom. The matrix block for carbon is then 13×13 . Second, the LCAO bases are generally nonorthogonal. We need to replace the identity matrix I with the overlap matrix S in the NECPA equation set (Eqs. 5.33) and (5.34)). However, S is disorder dependent on both diagonal and off-diagonal blocks. For off-diagonal blocks, we neglect its disorder in the spirit of LCA. In principle, the diagonal block disorders can be dealt with in a similar way to Hamiltonian. Here we make one more approximation that the diagonal disorders in Scan be neglected as long as the LCAO bases are normalized. The diagonal elements of a S block equals 1 since the orbitals are normalized. The off-diagnoal elements, which describe the orbital overlaps within an atom, should not differ much between atoms if they are not too far away in the periodic table. Fig. 7.5 plots the S matrices of carbon and boron atoms. Fig. 7.5(a) and (b) show their nonzero elements distributions. The side colour bar shows their values. Fig. 7.5(c) and (d) show the side view of matrices along the diagonal direction, which clearly show their element values. We find that their values differs by less than 0.005, which is negligible. We see that the S matrices of carbon and boron atoms are very close in both nonzero



Figure 7.5: Plots for S blocks (diagonal). (a) S for carbon atom. (b) S for boron atom. The side colour bar indicates the element values. (c) is a side view of carbon S matrix, seen along the diagonal direction. (d) is similar to (c), but for boron. The horizontal axis of the figures is the index of the atomic orbitals, namely s, s, y, z, x, y, z, x, xy, yz, $(3z^2 - 1)/2/\sqrt{3}$, zx and $(x^2 - y^2)/2$ orbitals, where repeated orbitals represent the double- ζ polarizations.

elements distributions and values of elements. Therefore, the NECPA equation set with S matrix can be written by replacing the identity matrix I in Eqs. (5.33) with the overlap matrix S,

$$\overline{G_i^r} = \sum_q x_{iq} \overline{G_{iq}^r},$$

$$\overline{G^r}(k) = \left[ES - H_C^0(k) - \tilde{\varepsilon}^r - \Sigma^r(k)\right]^{-1},$$

$$\overline{G^r} = \int_{-\infty}^{+\infty} \frac{dk}{2\pi} \overline{G^r}(k),$$

$$\overline{G_i^r} = \left[\overline{G^r}\right]_{ii},$$

$$\overline{G_i^r} = \left[ES_i - \tilde{\varepsilon_i^r} - \Omega_i^r\right]^{-1},$$

$$\overline{G_{iq}^r} = \left[E_{iq} - \epsilon_{iq} - \Omega_i^r\right]^{-1}.$$
(7.2)

and

$$\overline{G_i^{<}} = \sum_q x_{iq} \overline{G_{iq}^{<}},$$

$$\overline{G^{<}}(k) = \overline{G^r}(k) \left[\tilde{\varepsilon}^{<} + \Sigma^{<}(k) \right] \overline{G^a}(k),$$

$$\overline{G^{<}} = \int_{-\infty}^{+\infty} \frac{dk}{2\pi} \overline{G^{<}}(k),$$

$$\overline{G_i^{<}} = \left[\overline{G^{<}} \right]_{ii},$$

$$\overline{G_i^{<}} = \overline{G_i^r} \left(\tilde{\varepsilon}_i^{<} + \Omega_i^{<} \right) \overline{G_i^a},$$

$$\overline{G_{iq}^{<}} = \overline{G_i^r} \Omega_i^{<} \overline{G_{iq}^a}.$$
(7.3)

Eqs. (7.2) and (7.3) are the key equations we are working with in the following part of this chapter.

7.2.1 Determination of impurity Hamiltonian

The quantities used in the NECPA equations are obtained from LCAO-based DFT and NEGF-DFT calculations performed by the NanoDcal package [152]. The quantities that needs to be calculated by NEGF-DFT are the S, H_C^0 and Σ^r matrices for the two-probe system. Quantities related to the impurities, S_{iq} and ε_{iq}^r are not within the two-probe calculation and are obtained from DFT calculation. We can not directly use the ε_{iq}^r from DFT calculation of an isolated impurity atom to solve NECPA equations due to two reasons: 1) the impurity atom is influenced by its surrounding host atoms. When placing in a host lattice, the impurity Hamiltonian is different from that when it is isolated; 2) impurity Hamiltonian from DFT calculation does not match the Hamiltonian from a NEGF-DFT calculation in energy. When performing DFT and NEGF-DFT calculation, the boundary conditions are different. DFT uses periodic boundary condition while the NEGF-DFT uses open boundary condition, resulting in a mismatch in their energies.



Figure 7.6: Determination of energy shift Δ . The brown atoms are the host atoms and the green one is the impurity. (a) atomic structure of the two-probe system. (b) atomic structure of the bulk structure. The structure is chosen to be very large so that atoms far away from the impurity is not affected. (c) a schematic plot of how to determine the energy shift. h_1 is the host atom Hamiltonian from the bulk calculation and h_2 is the host atom Hamiltonian from the two-probe calculation. s is the overlap matrix of the host atom.

showed in Fig. 7.6. First, we perform a NEGF-DFT calculation to get the Hamiltonian of the two-probe clean system (no impurities) (as shown in Fig. 7.6(a)). The host atom Hamiltonian h_2 is obtained from the two-probe Hamiltonian. Second, find the raw Hamiltonian of the impurity. This is done by performing a DFT bulk calculation for a large clean system accommodating one impurity at the center (Fig. 7.6(b)). The size of the system has to be large enough so that the Hamiltonian of the host atom far away from the impurity coincides with that in the clean system. The impurity sits inside the material so that the influence on its Hamiltonian from the host atoms is included. The raw impurity Hamiltonian h_q and host Hamiltonian h_1 can be extracted from the calculated results. h_1 has to be chosen far away from the impurity atom. Third, the energy shift between the bulk and two-probe system is calculated by

$$\Delta = (h_2 - h_1)/s, \tag{7.4}$$

where s is the overlap matrix of the host atom. "/" denotes the matrix division. In principle, Δ is a matrix from Eq. (7.4). Δ will have the form $c \cdot I$ if the bulk system is large enough in the second step, where c is a constant and I the identity matrix.



Figure 7.7: A flowchart of solving NECPA within LCAO. The three coloured blocks refers to the three main steps: preparing data (in red), solving the CPA equations self-consistently (in blue) and solving corresponding less quantities (quantities with superscript <) self-consistently (in green). Quantities in light grey are initial guess that enters the self-consistent loops.

In practical application, the "diagonality" of Δ can be used as an rule to check if the bulk system is large enough in the second step. Last, we determine the impurity Hamiltonian from the results obtained (Fig. 7.6(c)). The new impurity Hamiltonian h_{qnew} is calculated by

$$h_{qnew} = h_q + \Delta s_q, \tag{7.5}$$

where s_q is the impurity overlap matrix.

7.2.2 Code implementation

Having found all the quantities necessary for solving the NECPA equations (Eqs. (7.2) and (7.3)), we present the complete steps of solving NECPA with LCAO. In principle, we need to solve the charge density of the system when impurities are introduced, i.e., solve the NEGF equations with the NECPA equations self-consistently. However, the

charge density of impurity system differs very little when the impurity concentration is low. The NECPA equations can be solved independently. Fig. 7.7 shows the flowchart of solving the NECPA equations with LCAO under LCA. It takes 4 steps to solve the problem at a given energy E:

(1) For the two-probe system without impurity, calculate its Hamiltonian H_C^0 , the overlap matrix S, the host/impurity atom Hamiltonian ε_{iq} and the lead self-energies $\Sigma^{<}$ from a DFT supercell calculation and a NEGF-DFT two-probe calculation (red block in Fig. 7.7). Details can be found in subsection 7.2.1.

(2) With the quantities solved from (1), solve Eqs. (7.2) by following the selfconsistent steps discussed in subsection 5.2.2. (blue block in Fig. 7.7)

(3) With the solved $\tilde{\varepsilon}^r$ and $\overline{G^r}$, solve Eqs. (7.3) by following the self-consistent steps discussed in subsection 5.2.2 to get $\tilde{\varepsilon}^<$ and $\overline{G^<}$. (green block in Fig. 7.7)

(4) Calculate the disorder averaged transmission T(E) by Eq. (5.37).

These calculations have to be performed for all energy points to produce a full T(E) profile.

In the following, we discuss some problems in the implementation. When solving the NECPA equations, the most time-consuming parts are 2 steps (see subsection 5.2.2):

- (1) $[\overline{G^r}]_{ii} = [(E H_C^0 \tilde{\varepsilon}^r \Sigma^r)^{-1}]_{ii};$
- (2) $[\overline{G^{<}}]_{ii} = [\overline{G^{r}}(\tilde{\varepsilon}^{<} + \Sigma^{<})\overline{G^{a}}]_{ii},$

where (1) involves the inverse of a large matrix, to obtain the diagonal blocks of the inverse; (2) involves the multiplication of three large matrices, two of which comes from the inverse of large matrices. Again, only the diagonal blocks of the results are necessary. The computational cost of full matrix inversion and multiplication is

of $O(N^3)$ which becomes the bottleneck when the number of atomic sites is large. The computational cost can be reduced drastically by applying the principle-layer algorithm [160].

In a two-probe system, the device is partitioned into principal layers (PL) along the transport direction such that each PL only interacts with its two nearest-neighbor PLs. This is always viable due to the fact that the LCAO basis has a finite spacial range. Once a two-probe system is partitioned into PLs, the Hamiltonian H_C^0 , the overlap matrix S and the coherent potentials $\tilde{\varepsilon}^{r/<}$ are all in block tri-diagonal shape. By taking the advantage of the sparsity of the matrices, the cost of calculating the diagonal blocks is of O(n) [160], where n is the number of principle layers. The computational cost is drastically reduced if the number of atomic sites per principle layer N/n is small.

Following Ref. [160], the calculation of diagonal blocks of $\overline{G^r}$ ($[\overline{G^r}]_{ii} = A_i, i = 1, 2, \dots, N$) can be efficiently completed by the following recursive processes:

- (1) calculate the inverse of the 1st block by $C_1 = H_{11}^{-1}$;
- (2) calculate $C_{i+1} = (H_{i+1,i+1} H_{i+1,i}C_iH_{i,i+1})^{-1}$ for $i = 1, 2, \cdots, n-1$;
- (3) let $A_n = C_n$;
- (4) calculate $A_i = C_i + C_i H_{i,i+1} A_{i+1} H_{i+1,i} C_i$ for $i = n 1, n 2, \dots, 1$,

where we have denoted $H \equiv ES - H_C^0 - \tilde{\varepsilon}^r - \Sigma^r$. The high efficiency of this algorithm is because the large matrix is broken into small matrices whose inverse and multiplication are cheap. Similarly, the calculation of diagonal blocks of $\overline{G^{<}}$ ($[\overline{G^{<}}]_{ii} = -iB_i$) can be computed by the following steps:

- (1) let $Y_1 = D_1$;
- (2) calculate $Y_i = D_i + H_{i,i-1}C_{i-1}Y_{i-1}C_{i-1}^{\dagger}H_{i,i-1}^{\dagger}$ for $i = 2, 3, \cdots, n$;



Figure 7.8: Execution time vs. number of PL's. All matrix operations are performed for 50 times. The vertical axis refers to the time for 50 times of operations. Each block is 40×40 . (a) matrix inverse. The PL method is compared with the LU decomposition method implemented in MATLAB. (b) $inv(H) \cdot D \cdot inv(H^{\dagger})$. The PL method is compared with the direct method.

(3) calculate $B_n = C_n Y_n C_n^{\dagger}$;

(4) calculate $B_i = C_i H_{i,i+1} B_{i+1} H_{i,i+1}^{\dagger} C_i^{\dagger} - C_i Y_i C_i^{\dagger} + A_i Y_i C_i^{\dagger} + (A_i Y_i C_i^{\dagger})^{\dagger}$ for $i = n-1, n-2, \cdots, 2, 1,$

where we have denoted $D \equiv i(\Sigma^{<} + \tilde{\varepsilon}^{<})$.

To test the efficiency of the above algorithm, we compare them with the default functions implemented in MATLAB. We set the matrix block size to be 40×40 , a size common in practical applications. Since the PL method is highly efficient, for each test, we do 50 times of operations to get the computational time. Fig. 7.8(a) shows the comparison of matrix inverse calculation when the number of blocks n varies from 50 to 200. The PL method is compared with the default inverse function (inv) in MATLAB, which uses the LU decomposition method. We see that the PL method is more than 10^3 times faster than the default MATLAB function. Moreover, we find that the PL method is nearly of O(n), which is ideal for practical calculations. Fig. 7.8(b) shows the comparison of computing $inv(H) \cdot D \cdot inv(H^{\dagger})$. Similar to (a), the direct method uses the LU decomposition for matrix inverse. We see that the PL method is nearly linear with n while the default function is approximately of $O(n^{2.8})$. Like in (a), the linear scaling with n is ideal for practical calculations. We conclude that by applying the PL method, the computational cost of solving the NECPA equations are drastically reduced.

The last issue to consider in code implementation is how to accelerate the converging procedure of solving the NECPA equations. A common approach is to apply mixing methods. In our code, we have implemented the Anderson mixing method [161]. We briefly introduce the Anderson mixing here. In general all multidimensional iterative procedures may be defined by a nonlinear operator which, when applied to an input vector $x^{(l)}$ of length N, produces an output vector $y^{(l)}$ of the same length. Here the superscript l denotes the iteration number, $l \geq 1$. Self-consistency is achieved when both vectors coincide or, equivalently, when the residual vector defined by

$$\left|F^{(l)}\right\rangle \equiv \left|y^{(l)}\right\rangle - \left|x^{(l)}\right\rangle \tag{7.6}$$

vanishes. Without any mixing method, the input vector of the $(l+1)^{\text{th}}$ step is exactly the output vector of the l^{th} step

$$\left|x^{(l+1)}\right\rangle = \left|x^{(l)}\right\rangle + \left|F^{(l)}\right\rangle = \left|y^{(l)}\right\rangle.$$
(7.7)

No mixing in most cases leads to oscillations and even to a divergent iteration process.

In Anderson mixing method, the mixing step takes the vectors of several previous iterations into account and combines them in a "optimal" way. Anderson method considers the linear combination of previous M steps

$$|x^{(l+1)}\rangle = |x^{(l)}\rangle + \sum_{j=1}^{M} \theta_{j}^{(l)} \left(|x^{(l-j)}\rangle - |x^{(l)}\rangle \right),$$
 (7.8)

where the new input is given in the space spanned by the input vectors. Note that $0 \le M \le l-1$ and the coefficients $\theta_j^{(l)}$ are to be specified. We prefer working with



Figure 7.9: Rate of convergence using the Anderson mixing (blue curves) and no mixing (red curves). (a) testing with a simple numerical example $f_i = -d_{ii}x_i - cx_i^3$ for $i = 1, \dots, 5$. *M* is set to be 3 for the Anderson mixing. (b) testing with a 5 nm graphene lattice. The lattice is doped with 1% of boron. *M* is set to be 5 for the Anderson mixing.

the residual vectors instead of the output vectors

$$|F^{(l+1)}\rangle = |F^{(l)}\rangle + \sum_{j=1}^{M} \theta_{j}^{(l)} \left(|F^{(l-j)}\rangle - |F^{(l)}\rangle\right).$$
 (7.9)

The coefficients $\theta_j^{(l)}$ are determined by the requirement that they yield that particular linear combination which minimizes the norm of the residual vector $F^{(l+1)}$. With the condition $\partial \langle F^{(l+1)} | F^{(l+1)} \rangle / \partial \theta_j^{(l)} = 0$, we are led to the linear equation set

$$\sum_{j=1}^{M} \left\langle F^{(l)} - F^{(l-i)} \right| F^{(l)} - F^{(l-j)} \right\rangle \theta_{j}^{(l)} = \left\langle F^{(l)} - F^{(l-i)} \right| F^{(l)} \right\rangle,$$
(7.10)

where $i = 1, 2, \dots, M$. The linear equations are solved at every step in order to yield the coefficients $\theta_j^{(l)}$ and hence the linear combination with the shortest residual vector. Having found this optimal linear combination the input is updated by Eq. (7.8).

We use 2 examples to test the Anderson mixing. We first test the mixing with a simple numerical example $f_i = -d_{ii}x_i - cx_i^3$ where $i = 1, \dots, 5, d_{ii} = (3.0, 2.0, 1.5, 1.0, 0.5)$ and c = 0.01. We start from the initial guess $x_i = 1$. The error of the l^{th} step is measured by $|f(x^{(l)}) - x^{(l)}|$. M = 3 is used for the Anderson mixing method. Fig.

7.9(a) plots the rate of convergence (in log scale) as a function of iteration number. We find that the error diverges with no mixing (the red curve) while converges very fast with the Anderson mixing (the blue curve). We also test the mixing method with a practical example, i.e., a $5 \,\mathrm{nm}$ long graphene two-probe device. The device has 80 lattice sites and the central region is doped with boron atoms with a concentration of 1%. The mixing is applied to the interactor Ω^r . The error of the l^{th} step is measured by the maximum element of matrix $|\Omega^{r(l)} - \Omega^{r(l-1)}|$. M = 5 is used for the Anderson mixing method. Fig. 7.9(b) plots the rate of convergence as a function of iteration number with both method. We find that with no mixing, the NECPA equation only reaches 10^{-2} and oscillates around it for more than 100 steps of iteration (the red curve). In contrast, with the Anderson mixing, it reaches the required accuracy after only 25 steps of iteration (the blue curve). It is also interesting that in the Anderson method the error converges slowly for the first ~ 10 steps where the error is large. After these steps, it converges faster. This is due to the fact that the Anderson method is more effective when the variables are close to their exact solutions. In practical application, it is always preferred to combine the Anderson method together with some less effective but more stable method (e.g., the linear mixing method [162]). For example, we first run tens of steps with the linear mixing. When the error falls below some point (in our graphene example, say 0.1), we switch to the Anderson mixing method to accelerate the convergence rate.

We conclude that the Anderson mixing allows to converge the NECPA equations efficiently according to our test. We mention that the Broyden method [162] is also implemented in our code. However, we do not find it more effective than the Anderson method. In Ref. [162] the author argues that the Broyden method is equivalent to the Anderson method. In the rest of this chapter, we use the Anderson method for all the calculations.



Figure 7.10: Lattice structure of the simulated graphene two-probe system. The two leads extend to $z = \pm \infty$ in the transport direction. The system extends to $x = \pm \infty$ in the transverse direction. A vacuum of 10 Å is added to y direction. The two-probe channel is along zigzag direction of graphene and has a length of L. Boron dopants distribute uniformly inside the central region (showed by orange atoms). Both the boron concentration and channel length L are to vary according to our requirement.



Figure 7.11: Plots for Δ matrices of the two impurities. (a) Δ for boron atom that occupies type A graphene site. (b) Δ for boron atom that occupies type B graphene site. The 2 figures at the bottom show their corresponding structures. Both matrices shows very good diagonality. The off-diagonal elements of Δ is smaller than 0.002. The numbers

7.2.3 Verification by doped graphene system

Having presented our code implementation, we use a boron doped graphene system as an example for verification. Fig. 7.10 shows the atomic structure of the graphene twoprobe system. The transport is along the zigzag direction of graphene. The central region length L = 5 nm with 80 atoms in the unit cell. The system is infinitely wide in the transverse direction (x direction). The central region is doped with boron impurities uniformly (see orange sites in Fig. 7.10). The calculation is performed by using the NanoDcal package [152] with the ONCV pseudo potential [82, 83] and double- ζ polarization atomic basis. First we determine the impurity Hamiltonian by DFT calculation. A unit cell of 40 atoms (39 carbon atoms and 1 boron atom) is used for the bulk calculation. Note that it is necessary to consider two situations (see the bottom two figures in Fig. 7.11) since graphene has two type of lattice sites with different chemical environments. Such a difference will result in different impurity Hamiltonians for even the same atomic specie. Both types are considered in our calculation and the Δ matrices for both are showed in Fig. 7.11. It shows that both of them present very good diagonality. The off-diagonal elements are smaller than 0.002, which is negligible compared to the diagonal elements (around 0.1). This proves that the 40-atom unit cell is large enough for determining the impurity Hamiltonian as discussed in subsection 7.2.1. After the impurity Hamiltonian is obtained, we perform NEGF-DFT calculation for the two-probe clean graphene system (see Fig. 7.10) where a uniform k-sampling of 31×1 is used. Afterwards, our code is used to perform the NECPA-LCA calculation to compute the average transmission coefficient of the disordered systems. The result is to be compared with the result calculated by the LMTO-ASA-NECPA code NanoDsim package [152].

Before showing the result, we check the Wald identity [123] of our code. In the Green's function theory, the four Green's functions satisfy the relation $G^r - G^a = G^> - G^<$ rigorously. In Ref. [38], it is shown that this relation also holds for disordered system, i.e.

$$\overline{G^r} - \overline{G^a} = \overline{G^>} - \overline{G^<}.$$
(7.11)

In numerical calculation, a positive infinitesimal number η is always added to the imaginary part of the energy in the Green's function G^r to avoid the singularity of



Figure 7.12: Plots to verify the Wald identity relation. The vertical axis is the error of the Wald identity relation. It is defined as the max element of the matrix $|G^r - G^a - G^> + G^<|$. The horizontal axis refers to the positive infinitesimal number added to the energy.

matrices. This way, the Wald identity becomes

$$\overline{G^{>}} - \overline{G^{<}} - \left(\overline{G^{r}} - \overline{G^{a}}\right) = \overline{G^{r}} \cdot 2i\eta \cdot \overline{G^{a}}, \qquad (7.12)$$

where $\overline{G^{>}}$ is calculated by the Keldysh equation $\overline{G^{>}} = \overline{G^r \Sigma^{>} G^a}$ [123]. All the definitions for above Green's functions and self-energies can be found in Chapter 4. Eq. 7.12 indicates that when η decreases, the accuracy of the Wald identity increases. We thus vary the value of η from 10^{-4} to 10^{-10} and calculate the four Green's functions correspondingly at some given energy point. Then we calculate the error by

$$err = max \left| \overline{G^{>}} - \overline{G^{<}} - \left(\overline{G^{r}} - \overline{G^{a}} \right) \right|, \qquad (7.13)$$

where the function $max(\dots)$ means finding the maximum matrix element of (\dots) and results shown in Fig. 7.12 versus η . We find that when η decreases to zero, *err* decreases to zero accordingly. The result provides a very strict verification of the Green's functions calculations in our code implementation. In the following calcula-



Figure 7.13: Transmission vs. energy. The boron concentration is 5%. The black curve is the transmission function of the clean system (directly calculated by NanoDcal). The red curve with circular dots is the transmission function calculated by our NECPA-LCA code. The blue dots are results calculated by the NanoDsim package. The three are shifted so that their Dirac cones are at E = 0.

tion, η is set to 10^{-8} .

We first investigate the effect of impurity scattering on the system transmission coefficient. Impurity scattering tends to have a negative impact on the transport. To make the effect prominent, a high impurity concentration is considered here. Fig. 7.13 shows the calculated transmission as a function of electron energy when the boron impurity concentration is 5%. We calculated three transmission functions for comparison. The black curve is for clean graphene, the blue curve is for 5% doping calculated by the LMTO implementation of NECPA (NanoDsim [152]) and the red curve is calculated by our code. The NanoDsim results serve as the benchmark. All transmission functions are calculated with a $171 \times 1 k$ -sampling. We observe a decrease in the transmission function in the whole energy range as expected. Compared to the NanoDsim results, the NECPA-LCA accounts for part of the the impurity scattering. This is expected since the impurity concentration 5% is too high for NECPA-LCA (see section 7.1). We then investigate the results for low impurity concentration. Fig. 7.14



Figure 7.14: Transmission at E_f vs. Boron concentration. The boron concentration varies from 0.05 to 0.001. The blue curve with circular dots is for the clean graphene. The yellow curve with crosses is for the results calculated by the NanoDsim package. The red curve with square dots is for our code. Only the transmission at Fermi energy E_f is plotted, where E_f is determined by DFT calculations.

shows the transmission at Fermi energy $(T(E_f))$ versus the impurity concentration. We observe that when the concentration is large (>1%), the difference between results from NECPA-LCA (this work) and NanoDsim are quite large. When the impurity concentration is below 1%, the difference becomes acceptable. Therefore, the NECPA-LCA should be applied to systems whose impurity concentration is smaller than 1%.

In the following, we apply NECPA-LCA to calculate the graphene mobility at low impurity concentration. According to the Drude model [163], mobility in a material is given by

$$\mu = \frac{1}{\rho e n_{e/h}},\tag{7.14}$$

where ρ is the resistivity, e is the charge of a single carrier, and $n_{e/h}$ is the density of electron (e) or hole (h) carriers. At low doping concentration and room temperature, almost all dopants are ionized so that $n_{e/h}$ equals the doping concentration. The resistivity ρ of an electronic structure is connected to the resistance R through the Ohm's law,

$$\rho = R \frac{A}{L},\tag{7.15}$$

where L and A representing the length and cross-section area of the structure, respectively. In 2-dimension case, A becomes length along transverse direction. The resistance is defined as the inverse of the conductance G

$$R = \frac{1}{G},\tag{7.16}$$

where G in a non-polarized spin system is given by

$$G = \frac{2e^2}{h}T(E_f). \tag{7.17}$$

The formula is derived from Eq. (4.1) at small bias. Note that unlike the other equations in this thesis, this formula has been translated into SI units. The number calculated by Eq. (7.17) is in unit Ω^{-1} . Hence, we can use the NECPA-LCA method to calculate $T(E_f)$, using Eq. (7.14)-(7.17) the mobility is eventually determined by the slope of a linear fit of low bias resistance versus structure length

$$R = \frac{1}{\mu e n_{e/h} A} L. \tag{7.18}$$

We vary the length of the central region L from 3 nm to 7 nm by 1 nm step. For each case, the NECPA-LCA calculation is carried out for the transmission coefficient at the Fermi energy, which is $E_f = -0.1867 \text{ eV}$ when the boron concentration is at 0.1%. By applying Eq. (7.16) and (7.17), we obtain the resistance as a function of structure length L as shown in Fig. 7.15. We observe a linear dependence of the resistance R on the length L, which is the Ohm's law in Eq. (7.18). n_h (boron doping gives a p-type graphene) can be calculated from the density of carbon atoms in graphene, which is $3.82 \times 10^{19} m^{-2}$. $n_h = 3.82 \times 10^{16} m^{-2}$ when boron concentration is 0.1%. The transverse length $A = 4.26 \times 10^{-10} m$ and the hole charge $e = 1.6 \times 10^{-19} C$. We



Figure 7.15: Resistance vs. structure length. We calculate the resistance of graphene devices at various lengths (L = 3 nm, 4 nm, 5 nm, 6 nm and 7 nm). The central region is doped with boron at a concentration of 0.1%. The extracted mobility is shown in the figure too.

use linear regression method to obtain the slope of the fitting line. From the slope, we extract the mobility $\mu = 4.4 \times 10^4 \, cm^2 \cdot V^{-1} \cdot s^{-1}$ by Eq. (7.18). The number is comparable to the one $\mu = 2 \times 10^4 \, cm^2 \cdot V^{-1} \cdot s^{-1}$ calculated by the LMTO-ASA-CPA method reported in Ref. [164]. Suggesting that the NECPA-LCA method is able to provide reasonable estimation for the semiconductor mobility at low impurity concentration.

It is worth mentioning that the computational cost of our NECPA-LCA code. For the two-probe system with 80 carbon atoms, the self-consistent iterations for solving G^r takes ~ 20 seconds per iteration step. Usually 20 to 40 iterations are needed for converging G^r . For the $G^<$ iterations, it takes ~ 40 seconds per step. 20 to 40 steps are needed to converge $G^<$. Therefore the NECPA-LCA code is a relatively efficient method for estimating disorder effects in quantum transport simulation.



Figure 7.16: Lattice structures of the simulated MoS_2 . (a) and (b) are the top view and side view of the MoS_2 lattice. The black box is the primitive cell of monolayer MoS_2 . (c) and (d) are top views of the simulated two-probe MoS_2 structures. In (c)/(d), the S/Mo atoms in the central region are doped with Cl/Re atoms.

7.3 Application: Dopant Limited Mobility of MoS_2

Dopants play important roles in determining the physical and chemical properties of solids. For semiconductors, doping is necessary for adjusting the carrier densities and tailor the electronic property. The effects of doping are also significant in 2D materials. Recently, much attention has been attracted to the family of layered inorganic transition-metal dichalcogenides (TMDCs). Among them, single-layer MoS_2 possesses a direct band gap [165, 166] and exhibits carrier mobility comparable to graphene nano-ribbons with high current on-off ratios [167]. The electrical property of MoS_2 may be further modulated by substitutional doping, such as Re (*n*-type) and Nb (*p*-type) on Mo atoms, or Cl (*n*-type) on S atoms [168, 169, 170]. Researches also report that *n*-type doping can effectively reduce the contact resistance between


Figure 7.17: Band diagrams of the MoS_2 MOSFET at low bias. The source/drain (leads) are high n-doped (1%) and the channel (central region) is lightly n-doped. (a) low V_G . The barrier at the central region blocks the electron transport path so that the current can not flow through, referring to an "off" state. (b) high V_G . By tuning V_G , the barrier gets lower so that the electrons can pass through, referring to an "on" state.

layered TMDCs and metals [170, 171], which is very useful for electronic device applications. So far there have been very few experimental and theoretical works that determine the impurity limited mobility in *n*-type doped MoS₂. One widely used method to calculate the impurity limited mobility numerically is the LMTO-ASA-NECPA method, as presented in Refs. [164, 157]. However, since the monolayer MoS₂ has a hexagonal lattice with layered structure stacked by S-Mo-S (see Fig. 7.16 (a) and (b)), considerable effort is necessary to determine the correct ASA. Here, we use the NECPA-LCA code to calculate the impurity limited mobility of *n*-doped MoS₂. Two kinds of *n*-type doping are considered, i.e., S sites doped by Cl dopants and Mo sites doped by Re dopants.

The computation is similar to the one for boron doped graphene presented in subsection 7.2.3. We take Cl doping as an example since the computation process for Mo doping is the same. We first perform a DFT calculation for a large Cl doped MoS_2 super cell and then an NEGF-DFT calculation for a two-probe clean MoS_2 structure. The transport is along the zigzag direction of MoS_2 . From the two, the impurity Hamiltonian, the clean system Hamiltonian and overlap matrices are extracted. Afterwards, an NECPA-LCA calculation is then carried out to obtain the configurational averaged transmission function. The only difference to graphene case of subsection 7.2.3 is the energy where the transmission is calculated. For the simulated graphene device, the transmission is calculated at the system Fermi energy E_f , which can be done for any impurity concentration since graphene has zero band gap. For MoS₂, the transmission at E_f vanishes when impurity concentration is low because the E_f lies inside the band gap of MoS₂. Here we consider the device to be a field effect transistor (FET), where the source and drain (leads) are extremely highly *n*-doped (e.g. 1%). The channel (central region) is lightly *n*-doped (e.g., 0.1%). Fig. 7.17 is a schematic of the band diagram when gate voltage V_G is tuned. V_G shifts the band edges (E_C and E_V in the figure) of the central region down when the device is turned "on". The energy where electrons flow through enters into MoS₂ conduction band under appropriate V_G . Transmission needs to be calculated at some energy above the E_C . We choose the energy to be 0.1 eV above E_C , which is a reasonable for a tuned on FET. The mobility is then extracted from a linear fitting of resistance Rversus length L of the scattering region.

Fig. 7.18 plots the relation between R and L. Linear regression gives the mobility of both cases. The mobility is $310\,cm^2\cdot V^{-1}\cdot s^{-1}$ for 0.1% Cl doping and 390 $\,cm^2\cdot$ $V^{-1} \cdot s^{-1}$ for 0.2% Re doping. The calculated mobility value for of Cl doping is higher than the experimentally reported $60 \sim 80 \, cm^2 \cdot V^{-1} \cdot s^{-1}$ in Ref. [170], where a 4 times higher doping level than ours was applied. Our predicted mobility should decrease but still larger than the experimental value when calculated at the same doping level. Other reported values [171, 167] range from 40 to more than $200\,cm^2\cdot V^{-1}\cdot s^{-1}$ with uncontrollable impurities in the material. Our results should be taken as an upper bound of impurity limited mobility. For Re doped MoS_2 , to the best of our knowledge there has been no published values. According to our results, Re appears to be a better n-type dopant than Cl in terms of electron mobility. This can be understood by looking at the density of states (DOS) close to E_C . Re doping tends to provide a higher DOS than Cl doping at energies close to E_C [168] and therefore provide more transport channels. It is interesting to observe that the two fitted lines intersect very close to L = 0 (not shown in the figure). This is because L = 0 refers to the ballistic transport limit, where the two cases coincide.



Figure 7.18: Resistance vs. structure length. We calculate the resistance of MoS_2 devices at various lengths (L = 2.8 nm, 3.7 nm, 4.7 nm, 5.6 nm and 6.6 nm). The central region is doped with 1) Cl at S sites with a concentration of 0.1% (red dots) and 2) Re at Mo sites with a concentration of 0.2% (purple dots). The two cases gives the same number of impurities in the lattice. The extracted mobilities are shown in the text boxes beside the curves.

7.4 Summary

In this chapter we proposed the NECPA-LCA method to solve transport problems in disordered materials with low impurity concentration. In this theory, assumption is made that at low impurity concentration, the off-diagonal disorder can be neglected when evaluating the transmission. The the validity of this assumption was verified by a tight binding model. The DFT implementation is verified by a boron doped graphene system and compared with the results obtained using the LMTO-ASA-CPA method. Both showed that NECPA-LCA gives acceptable results on the transport when the impurity concentration is low (<1%). Finally, the method is employed to predict the dopant limited mobility of monolayer MoS₂, where we found that Re is a better *n*-type dopants than Cl for monolayer MoS₂.

Conclusions

In this thesis we focused on investigating the random impurity effects in semiconductors. We started by investigating the band gap and band alignment of $GaSb_xN_{1-x}$ alloys with KS-DFT. The band gaps were predicted for the dilute Sb limit (<1%) at the HSE06 hybrid functional level where over 1000 atoms were calculated to directly determine the band bowing. Then, for transport with disorders at non-equilibrium, approximated schemes such as the CPA theory and the NECPA theory were introduced along with their most easily implementable site-oriented DFT methods - the LMTO-ASA. We applied these methods to simulate a graphene TFET where codoping B-N impurity atoms were introduced. For semiconductor devices where the impurity concentration is typically low (<1%), we showed that these equilibrium and non-equilibrium CPA methods can be extended to DFT methods based on LCAO, NECPA-LCA which has been shown to be useful in practical semiconductor simulations. We calculated the mobilities of monolayer Cl and Re doped MoS₂. All of our work was from atomistic first principles and parameter-free.

The KS-DFT calculation on $GaSb_xN_{1-x}$ alloy was based on the HSE06 hybrid functional implemented with LCAO basis. With LCAO, the long range Coulomb operator matrix can be represented by a sparse form, which made possible the expensive hybrid functional calculations for very large crystal systems. We found that in $GaSb_xN_{1-x}$ alloy the band gap decreased sharply in the Sb concentration range of $0 < x \leq 1\%$. The fundamental mechanism of such a sharp band bowing was found to be due to impurity states inside the band gap of GaN and a strong quantum interaction between the orbitals of Sb impurity and their surrounding neighbors and, in particular, the interaction between Sb and its next nearest neighbor N atom plays a significant role. Using the calculated natural band alignments, we predicted that $x \approx 0.3\%$ and slightly lower of Sb doping should be ideal for photochemical water splitting in solar fuel applications.

The transport calculations were based on a first principles quantum transport method that combines NEGF-DFT and NECPA theory for analyzing disorder scattering and configuration averaging of the physical results. It is important to point out that combining NEGF with DFT provides us a parameter free method to determine the Hamiltonian matrix of the device under non-equilibrium quantum transport conditions for open device structures. The NECPA, on the other hand, was a very powerful method for solving random disordered systems by analytically averaging the physical properties before a numerical simulation was done. This way one avoids the lengthy and usually prohibitive numerical computation in quantum transport simulations. Choosing TB-LMTO to implement KS-DFT, the NECPA theory was naturally combined with the NEGF-DFT formalism. In graphene TFET, the B-N co-doping was found to open a substantial gap that linearly scaled with the co-doping concentration which was appropriate for making TFET. We also observed that the impurity scattering in the graphene TFET reduces the band-to-band tunneling current by a substantial factor. We noted that even if the potential change due to doping was accounted for (e.g. by VCA), it was not adequate to obtain correct transport result if impurity scattering was not explicitly included. The NECPA-LMTO-ASA method provided a promising parameter-free first principles TCAD solution to determine disorder induced physical effects. Finally, at low impurity concentration (<1%), assumption was made that the off-diagonal disorder can be neglected when evaluating the transmission. The validity of this assumption was verified by a tight binding model. The DFT implementation was verified by a boron doped graphene system and compared with the results obtained using the LMTO-ASA-CPA method. Both showed that NECPA-LCA gives acceptable results on the transport when the impurity concentration is low. The method was employed to predict the dopant limited mobility of monolayer MoS_2 , where we found that Re was a better *n*-type dopants than Cl for monolayer MoS_2 .

Our work indicates that even at low impurity concentration, off-site disorders must be included if one wants to most accurately predict the transport of a disordered system. One possible way is to implement them with bases without ASA, such as the exact muffin-tin orbitals (EMTO) [172] which has been proven feasible with the CPA theory [173]. EMTO has major improvements over the LMTO-ASA that the large overlapping spheres produce more accurate description of the material potential and that the interstitial region is taken into account in contrary to LMTO-ASA. Such advantages makes EMTO suitable to investigate systems with low symmetries. Recently, EMTO has been implemented with CPA-NVC to calculate quantum transport in disordered systems [174]. The viability of implementing NECPA, which allows to calculate not only average but also fluctuation of transport in disordered systems [128], would be an interesting research direction. Another promising direction is to employ the BEB approximation [52] and implement with LCAO basis. Like EMTO, the BEB has been proven viable with the CPA theory [52] but not with quantum transport formula yet. How to reduce the BEB theory to implement NECPA would be a second promising extension of the thesis. EMTO is expected to treat disorder more accurately than BEB with LCAO due to the fact that MTO is more naturally compatible to implement CPA. BEB with LCAO will be more complicated in formalism but less complicated to use when investigating systems with little symmetries.

On a more theoretical level, it is important to go beyond the CPA and NECPA which are at single site approximation (SSA) level. In mesoscopic transport problems such as the weak localization in disorder electronic system, nonlocal interferenceinduced effects play a significant role in quantum transport which is missing in CPA/NECPA with SSA. A promising way to account for the nonlocal effects is to employ the dual-fermion method with the non-equilibrium Keldysh formalism [175].

However, to the best of our knowledge, it is only investigated with tight-binding model but with no *ab initio* method. Considering that the computation is expensive even with tight-binding, how to efficiently implement the dual-fermion method with *ab initio* method is another possible extension of the work.

A

STRUCTURE CONSTANTS

This appendix is to provide further details on the structure constant appearing in Eq. (2.50) in Chapter 2. As mentioned in Eq. (2.50), the irregular solution of Laplace Eq. (2.45) centered at **R** can be expanded by the regular solution centered at **R'** through the help of the structure constants. The structure constants are defined as [89]:

$$S_{\mathbf{R}L,\mathbf{R}'L'} = \sum_{L} (-1)^{l''+1} \frac{8\pi \left(2l-1\right)!! C_{LL'L''}}{(2l'-1)!!} K_L \left(\mathbf{R}''-\mathbf{R}'\right), \tag{A.1}$$

where the sum is restricted by the condition of l = l' + l'' and the double factorial is defined recursively: (2l + 1)!! = (2l + 1)(2l - 1)!! and (-1)!! = 1. The quantities $C_{LL'L''}$ are called Gaunt coefficients and are defined by [54, 89]

$$C_{LL'L''} = \int Y_L\left(\hat{\mathbf{r}}\right) Y_{L'}\left(\hat{\mathbf{r}}\right) Y_{L''}\left(\hat{\mathbf{r}}\right) d^2\hat{\mathbf{r}}.$$
(A.2)

The structure constants are symmetric with respect to the indices $\mathbf{R}L$ and $\mathbf{R}'L',$ i.e.,

$$S_{\mathbf{R}L,\mathbf{R}'L'} = S_{\mathbf{R}'L',\mathbf{R}L},\tag{A.3}$$

and depend on the distance according to the inverse power law [89]

$$S_{\mathbf{R}L,\mathbf{R}'L'} \propto \left(\frac{r_{\mathbf{R}}}{|\mathbf{R}-\mathbf{R}'|}\right)^{l+l'+1}$$
 (A.4)

Β

WRONSKIAN

The Wronskian is used to match a function to a linear combination of two other functions at $r = s_{\mathbf{R}}$. As described in reference [89], the Wronskian is defined as

$$\{f_1(r), f_2(r)\} \equiv r^2 [f_1(r) f'_2(r) - f'_1(r) f_2(r)].$$
(B.1)

This definition implies the following relations (argument r omitted):

$$\{f_1, f_2\} = -\{f_2, f_1\}$$
(B.2)

and

$$\{f_1, f_2\} \{f_3, f_4\} = \{f_1, f_3\} \{f_2, f_4\} - \{f_1, f_4\} \{f_2, f_3\}.$$
 (B.3)

If function g(r) is to be matched smoothly (both the function value and its first derivative) at $r = s_{\mathbf{R}}$ to a linear combination of two functions $f_1(r)$ and $f_2(r)$, the matching condition has the form of [89]

$$g(r) \to \frac{\{g(r), f_2(r)\} f_1(r) - \{g(r), f_1(r)\} f_2(r)}{\{f_1(r), f_2(r)\}}$$
(B.4)

where the Wronskians are evaluated at the matching radius $r = s_{\mathbf{R}}$. We mention a useful property of Wronskian [176]: if two functions $f_1(r)$ and $f_2(r)$ satisfy the same radial differential equation, then their Wronskian $\{f_1(r), f_2(r)\}$ is independent of r.

C

LMTO GREEN'S FUNCTIONS

C.1 Retarded/Advanced Green's function

This appendix is to provide further details on the derivation for Eq. (4.20) in Section 4.4. Let's start with the expression for retarded/advanced Green's function from Eq. (4.18)

$$G^{r,a}\left(\varepsilon\right) = \left[\varepsilon^{\pm} - C - \sqrt{\Delta}S^{\alpha}(1 - (\gamma - \alpha)S^{\alpha})^{-1}\sqrt{\Delta}\right]^{-1}, \quad (C.1)$$

where the subscript " \pm " has been added to the variable ε to distinguish retarded and advance Green's function. "+" applies to retarded Green's function and "-" to the advanced Green's function. Notice that the matrix Δ can be written in the form of

$$\Delta = \sqrt{\Delta} \left\{ \left[I - (\gamma - \alpha) S^{\alpha} \right] \left[I - (\gamma - \alpha) S^{\alpha} \right]^{-1} \right\} \sqrt{\Delta}$$
$$= \sqrt{\Delta} \left\{ \left[I - (\gamma - \alpha) S^{\alpha} \right]^{-1} - (\gamma - \alpha) S^{\alpha} \left[I - (\gamma - \alpha) S^{\alpha} \right]^{-1} \right\} \sqrt{\Delta}.$$
(C.2)

Then, the Green's function can be derived in the following way

$$G^{r,a}(\varepsilon) = \left[\varepsilon^{\pm} - C - \sqrt{\Delta}S^{\alpha}(I - (\gamma - \alpha)S^{\alpha})^{-1}\sqrt{\Delta}\right]^{-1}$$

= $\left[\Delta + (\gamma - \alpha)(\varepsilon^{\pm} - C)\right]^{-1}\left[\Delta + (\gamma - \alpha)(\varepsilon^{\pm} - C)\right]$
 $\times \left[\varepsilon^{\pm} - C - \sqrt{\Delta}S^{\alpha}(I - (\gamma - \alpha)S^{\alpha})^{-1}\sqrt{\Delta}\right]^{-1}$
= $\left[\Delta + (\gamma - \alpha)(\varepsilon^{\pm} - C)\right]^{-1}\left\{\sqrt{\Delta}[I - (\gamma - \alpha)S^{\alpha}]^{-1}\sqrt{\Delta} - \sqrt{\Delta}(\gamma - \alpha)S^{\alpha}[I - (\gamma - \alpha)S^{\alpha}]^{-1}\sqrt{\Delta} + (\gamma - \alpha)(\varepsilon^{\pm} - C)\right\}$

$$\times \left[\varepsilon^{\pm} - C - \sqrt{\Delta} S^{\alpha} (I - (\gamma - \alpha) S^{\alpha})^{-1} \sqrt{\Delta} \right]^{-1}$$

$$= \left[\Delta + (\gamma - \alpha) (\varepsilon^{\pm} - C) \right]^{-1} \left\{ \sqrt{\Delta} [I - (\gamma - \alpha) S^{\alpha}]^{-1} \sqrt{\Delta} + (\gamma - \alpha) [\varepsilon^{\pm} - C - \sqrt{\Delta} S^{\alpha} [I - (\gamma - \alpha) S^{\alpha}]^{-1} \sqrt{\Delta}] \right\}$$

$$\times \left[\varepsilon^{\pm} - C - \sqrt{\Delta} S^{\alpha} (I - (\gamma - \alpha) S^{\alpha})^{-1} \sqrt{\Delta} \right]^{-1}, \quad (C.3)$$

where "1" is replaced by "I". After some further algebra, Eq. (C.3) reduces to

$$\begin{aligned} G^{r,a}\left(\varepsilon\right) &= \left[\Delta + (\gamma - \alpha)\left(\varepsilon^{\pm} - C\right)\right]^{-1} \left\{\sqrt{\Delta}\left[I - (\gamma - \alpha)S^{\alpha}\right]^{-1}\sqrt{\Delta} \right. \\ &+ (\gamma - \alpha)\left[\varepsilon^{\pm} - C - \sqrt{\Delta}S^{\alpha}\left[I - (\gamma - \alpha)S^{\alpha}\right]^{-1}\sqrt{\Delta}\right] \right\} \\ &\times \left[\varepsilon^{\pm} - C - \sqrt{\Delta}S^{\alpha}\left(I - (\gamma - \alpha)S^{\alpha}\right)^{-1}\sqrt{\Delta}\right]^{-1} \\ \left[\Delta + (\gamma - \alpha)\left(\varepsilon^{\pm} - C\right)\right]^{-1} \left\{(\gamma - \alpha) \right. \\ &+ \sqrt{\Delta}\left[I - (\gamma - \alpha)S^{\alpha}\right]^{-1}\sqrt{\Delta}\left[\varepsilon^{\pm} - C - \sqrt{\Delta}S^{\alpha}\left(I - (\gamma - \alpha)S^{\alpha}\right)^{-1}\sqrt{\Delta}\right]^{-1} \right\} \\ &= \left[\Delta + (\gamma - \alpha)\left(\varepsilon^{\pm} - C\right)\right]^{-1} \left\{(\gamma - \alpha) \right. \\ &+ \sqrt{\Delta}\left[\left(\varepsilon^{\pm} - C\right)\left[I - (\gamma - \alpha)S^{\alpha}\right] - \Delta S^{\alpha}\right]^{-1}\sqrt{\Delta} \right\} \\ &= \left[\Delta + (\gamma - \alpha)\left(\varepsilon^{\pm} - C\right)\right]^{-1} \left\{(\gamma - \alpha) \right. \\ &+ \sqrt{\Delta}\left[\left(\varepsilon^{\pm} - C\right) - \left[\Delta + \left(\varepsilon^{\pm} - C\right)\left(\gamma - \alpha\right)\right]S^{\alpha}\right]^{-1}\sqrt{\Delta} \right\} \\ &= \frac{\gamma - \alpha}{\Delta + (\gamma - \alpha)\left(\varepsilon^{\pm} - C\right)} + \frac{\sqrt{\Delta}}{\Delta + (\gamma - \alpha)\left(\varepsilon^{\pm} - C\right)} \left[\frac{\varepsilon^{\pm} - C}{\Delta + (\gamma - \alpha)\left(\varepsilon^{\pm} - C\right)} - S^{\alpha}\right]^{-1} \\ &\times \frac{\sqrt{\Delta}}{\Delta + (\gamma - \alpha)\left(\varepsilon^{\pm} - C\right)}. \end{aligned}$$

Here for the third equal sign, we have used the matrix relation $AB^{-1}AC^{-1} = A(CA^{-1}B)^{-1}$. If we recall the notation from Eq. (4.21), we obtain:

$$G^{r,a}(\varepsilon) = \frac{\gamma - \alpha}{\Delta + (\gamma - \alpha)(\varepsilon^{\pm} - C)} + \frac{\sqrt{\Delta}}{\Delta + (\gamma - \alpha)(\varepsilon^{\pm} - C)} \left[\frac{\varepsilon^{\pm} - C}{\Delta + (\gamma - \alpha)(\varepsilon^{\pm} - C)} - S^{\alpha} \right]^{-1} \times \frac{\sqrt{\Delta}}{\Delta + (\gamma - \alpha)(\varepsilon^{\pm} - C)} = \lambda^{\alpha} + \mu^{\alpha} [P^{\alpha} - S^{\alpha}]^{-1} \mu^{\alpha}, \qquad (C.5)$$

which is exactly the Eq. (4.20) in Section 4.4.

C.2 Two-probe retarded/advanced Green's function

This part is to give further details for Eq. (4.25) in Section 4.4. Let's show how twoprobe Green's function is modified under representation transform. The two-probe Green's function is given by Eqs. (4.5) and (4.12) [123]

$$G_{CC} = \left[\varepsilon - H_{CC} - H_{CL}(\varepsilon - H_{LL})^{-1}H_{LC} - H_{CR}(\varepsilon - H_{RR})^{-1}H_{RC}\right]^{-1}.$$
 (C.6)

To rewrite it in the LMTO formalism, let's check each term in Eq. (C.6). The first two terms have the form of

$$\varepsilon - H_{CC} = \varepsilon - C_C - \sqrt{\Delta_C} S^{\gamma}_{CC} \sqrt{\Delta_C}$$
 (C.7)

in LMTO by Eq. (2.91) [54]. For the third and fourth term, use the notation $\tau = L, R$ we have

$$H_{C\tau}(\varepsilon - H_{\tau\tau})^{-1}H_{\tau C}$$

$$= \sqrt{\Delta_C}S_{C\tau}^{\gamma}\sqrt{\Delta_\tau} \left(\varepsilon - C_{\tau} - \sqrt{\Delta_\tau}S_{\tau\tau}^{\gamma}\sqrt{\Delta_\tau}\right)^{-1}\sqrt{\Delta_\tau}S_{\tau C}^{\gamma}\sqrt{\Delta_C}$$

$$= \sqrt{\Delta_C}S_{C\tau}^{\gamma} \left(\frac{\varepsilon - C_{\tau}}{\Delta_\tau} - S_{\tau\tau}^{\gamma}\right)^{-1}S_{\tau C}^{\gamma}\sqrt{\Delta_C}$$

$$= \sqrt{\Delta_C}S_{C\tau}^{\gamma} (P_{\tau}^{\gamma} - S_{\tau\tau}^{\gamma})^{-1}S_{\tau C}^{\gamma}\sqrt{\Delta_C}, \qquad (C.8)$$

where we have applied the fact that $C_{\tau C} = C_{C\tau} = 0$ and $S^{\gamma} = [(S^{\alpha})^{-1} - (\gamma - \alpha)]^{-1}$ [54]. Therefore the Green's function can be rewritten as

$$G_{CC} = \left[\varepsilon - C_C - \sqrt{\Delta_C} S_{CC}^{\gamma} \sqrt{\Delta_C} - \sqrt{\Delta_C} S_{CL}^{\gamma} (P_L^{\gamma} - S_{LL}^{\gamma})^{-1} S_{LC}^{\gamma} \sqrt{\Delta_C} - \sqrt{\Delta_C} S_{CR}^{\gamma} (P_R^{\gamma} - S_{RR}^{\gamma})^{-1} S_{RC}^{\gamma} \sqrt{\Delta_C} \right]^{-1}$$
$$= \left\{\varepsilon - C_C - \sqrt{\Delta_C} \left[S_{CC}^{\gamma} + S_{CL}^{\gamma} (P_L^{\gamma} - S_{LL}^{\gamma})^{-1} S_{LC}^{\gamma} + S_{CR}^{\gamma} (P_R^{\gamma} - S_{RR}^{\gamma})^{-1} S_{RC}^{\gamma} \right] \sqrt{\Delta_C} \right\}^{-1}.$$
(C.9)

Comparing this equation with Eq. (C.1), we find that if we define

$$\tilde{S}_{CC}^{\gamma} = S_{CC}^{\gamma} + S_{CL}^{\gamma} (P_L^{\gamma} - S_{LL}^{\gamma})^{-1} S_{LC}^{\gamma} + S_{CR}^{\gamma} (P_R^{\gamma} - S_{RR}^{\gamma})^{-1} S_{RC}^{\gamma},$$
(C.10)

everything stays unchanged except for a " \sim " hat on the S matrix. Therefore, similar to Eq. (C.5), we obtain:

$$G_{CC}^{r,a}\left(\varepsilon\right) = \lambda_{C}^{\alpha}\left(\varepsilon^{\pm}\right) + \mu_{C}^{\alpha}\left(\varepsilon^{\pm}\right)\tilde{g}_{CC}^{\alpha}\left(\varepsilon^{\pm}\right)\mu_{C}^{\alpha}\left(\varepsilon^{\pm}\right),\qquad(C.11)$$

where we have defined the modified auxiliary Green's function

$$\tilde{g}_{CC}^{\alpha}(\varepsilon) = \left[P_{CC}^{\alpha}(\varepsilon) - \tilde{S}_{CC}^{\alpha}\right]^{-1}$$

$$= \left[P_{CC}^{\alpha}(\varepsilon) - S_{CC}^{\alpha} - S_{CL}^{\alpha}(P_{L}^{\alpha} - S_{LL}^{\alpha})^{-1}S_{LC}^{\alpha} - S_{CR}^{\alpha}(P_{R}^{\alpha} - S_{RR}^{\alpha})^{-1}S_{RC}^{\alpha}\right]^{-1}.$$
(C.12)

By further defining a modified self-energy

$$\Pi^{\alpha}_{\tau} = S^{\alpha}_{C\tau} (P^{\alpha}_{\tau} - S^{\alpha}_{\tau\tau})^{-1} S^{\alpha}_{\tau C}, \qquad (C.13)$$

Eq. (C.12) becomes:

$$\tilde{g}_{CC}^{\alpha}\left(\varepsilon\right) = \left[P_{CC}^{\alpha}\left(\varepsilon\right) - S_{CC}^{\alpha} - \Pi_{L}^{\alpha} - \Pi_{R}^{\alpha}\right]^{-1}.$$
(C.14)

This result is the Eq. (4.25) in Section 4.4.

D

AN EQUIVALENT FORM OF THE CPA EQUATION SET

This appendix is to present further details for Eqs. (5.30). We first rewrite Eqs. (5.28) with conditional Green's function and then introduce the interactor to make the equation sets computationally simpler. This appendix follows the details presented in Ref. [38].

D.1 Rewriting the CPA Equations with the Conditional Greens function

In this section we rewrite Eqs. (5.28) with the conditional Greens function.

We first prove a lemma for the block matrix inverse. Suppose that A and A' are the inverse of the two 2×2 block matrices:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1},$$
$$A' = \begin{pmatrix} a'_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1},$$

from which we can obtain the expression of the upper left corner of matrices A and

A'

$$A_{11} = \left(a_{11} - a_{12}a_{22}^{-1}a_{21}\right)^{-1},$$

$$A'_{11} = \left(a'_{11} - a_{12}a_{22}^{-1}a_{21}\right)^{-1}.$$

It follows that

$$(A_{11})^{-1} - (A'_{11})^{-1} = a_{11} - a'_{11},$$
 (D.1)

which is the conclusion of the lemma.

Second, we apply the lemma to $\overline{G_i^r}$ and $\overline{G_{iq}^r}$ and obtain a useful relation between them. $\overline{G_i^r}$ is defined as Eq. (5.29) and $\overline{G_{iq}^r}$ is defined by the last two lines of Eqs. (5.28). Let $A = \overline{G_i^r}$ and $A' = \overline{G_{iq}^r}$, and reorder $\overline{G_i^r}$ and $\overline{G_{iq}^r}$ such that the block of site *i* is in the location of a_{11} and a'_{11} respectively. We get

$$a_{11} - a'_{11} = (\tilde{\varepsilon}^r_{iq})_{ii} - (\tilde{\varepsilon}^r)_{ii} = \varepsilon^r_{iq} - \tilde{\varepsilon}^r_i.$$

By using the lemma, it is derived

$$\left(\overline{G_i^r}\right)^{-1} - \left(\overline{G_{iq}^r}\right)^{-1} = \varepsilon_{iq}^r - \tilde{\varepsilon}_i^r.$$
(D.2)

Third, we derive the first line of Eqs. (5.30). We substitute Eq. (D.2) into the second line of Eq. (5.28) and obtain

$$t_{iq}^{r} = \left\{ \left[\left(\overline{G_{i}^{r}}\right)^{-1} - \left(\overline{G_{iq}^{r}}\right)^{-1} \right]^{-1} - \overline{G_{i}^{r}} \right\}^{-1}.$$
 (D.3)

Substitute Eq. (D.3) into the first line of Eq. (5.28) and we obtain

$$\sum_{q} x_{iq} \left\{ \left[\left(\overline{G_i^r}\right)^{-1} - \left(\overline{G_{iq}^r}\right)^{-1} \right]^{-1} - \overline{G_i^r} \right\}^{-1} = 0.$$
 (D.4)

By using $\sum_{q} x_{iq} = 1$, Eq. (D.4) can be simplified as

$$\overline{G_i^r} = \sum_q x_{iq} \overline{G_{iq}^r},\tag{D.5}$$

which is the first line of Eq. (5.30). Up to here, the Eq. (5.28) has been written in terms of the conditional Green's function.

D.2 The Introduction of Interactor Ω_i

In this section we introduce the interactor Ω_i^r to make Eq. (5.28) computationally easier to implement [54]. Eq. (D.2) and Eq. (5.28) shows that not all elements of the Greens function (and conditional Greens function) are necessary during the selfconsistent process. Therefore, we can always find a proper site diagonal quantity Ω_i^r such that

$$\overline{G_i^r} = \left[EI - \tilde{\varepsilon}_i^r - \Omega_i^r\right]^{-1},$$

$$\overline{G_{iq}^r} = \left[EI - \varepsilon_{iq}^r - \Omega_i^r\right]^{-1},$$

(D.6)

which are nothing but just the last two equations of Eq. (5.30).

BAC MODEL AND MOLECULAR DYNAMICS

This appendix is to present further details for Chapter 3. We first introduce the band anti-crossing model and then presents how to extract band bowing parameter from the band bowing figure. Finally we listed the parameters for molecular dynamics used in the project.

E.1 Extraction of the coupling constant

The variation of the band edge of GaNSb can be modeled by the band anti-crossing model as [110]

$$E_{\pm}(k) = \frac{1}{2} \left\{ E^{\nu}(k) + E^{Sb} \pm \sqrt{\left[E^{Sb} - E^{\nu}(k)\right]^2 + 4V^2 x} \right\}$$

where E^{Sb} is the energy of impurity state, $E^{v}(k)$ is the original valence band, V the coupling constant between the valence states and impurity states, x the impurity concentration and finally E_{\pm} is the corresponding impurity states energy and the valence band energy after the quantum interaction. Pay attention that this formula only applies to low x case so that it can be used to describe the band gap of the GaNSb alloy at the low x limit. Take the differential over x we get the slope $V^2/(E^{Sb} - E^v)$ when $x \to 0$. As stated in the text, compared to the valence band edge the change of the conduction band edge E_c is negligible so that it is reasonable to use the same formula to model the variation of the band gap. We use the two points $x \to 0$ and x = 0.175% to estimate the slope at the low x limit and the results from HSE06 that E^{Sb} is about 0.5 eV above E^{v} and finally we get V = 6.7 eV. It is worth mentioning that the band gap of $x \to 0$ is 2.96 eV (3.46 eV, the band gap of GaN, minus 0.6 eV).

E.2 Molecular Dynamics

We use the Tersoff empirical potential [115] for the molecular dynamics calculations. The parameters we used are listed in Tab. E.1 and E.2. We found that these parameters can well produce the lattice constant as well as the surface structure of GaNSb alloys. We have compared ΔV 's from molecular dynamics with those from pure DFT results (i.e., both the relaxation and ΔV are computed by DFT) for x = 0, 2.7%, 5.6%and they differ by <40 meV.

Table E.1: Tersoff potentials for GaNSb alloys (Part 1). The table is organized according to the favor of LAMMPS.

	m	γ	λ_3	С	d	$\cos heta_0$	n	β	λ_2
Ga Ga Ga	1.0	0.04787	1.846	1.918	0.75	-0.2813	1.0	1.0	1.4497
ΝΝΝ	1.0	0.7661	0.000	0.1785	0.2017	-0.04524	1.0	1.0	2.3843
$\mathrm{Sb} \ \mathrm{Sb} \ \mathrm{Sb}$	1.0	0.00	0.00	0.00	1.00	0.00	1.0	1.0	0.0
$Ga \ N \ N$	1.0	0.001632	0.000	65.207	2.821	-0.518	1.0	1.0	2.6391
N Ga Ga	1.0	0.001632	0.000	65.207	2.821	-0.518	1.0	1.0	2.6392
Ga Sb Sb	3.0	0.3630	0.9687	1.2088	0.8398	-0.4277	4.6022	1.0	1.7452
Sb Ga Ga	3.0	0.3630	0.9687	1.2088	0.8398	-0.4277	4.6022	1.0	1.7452
Sb N N	1.0	0.00	0.00	0.00	1.00	0.00	1.0	1.0	0.0
N Sb Sb	1.0	0.00	0.00	0.00	1.00	0.00	1.0	1.0	0.0
Ga Ga N	1.0	0.001632	0.000	65.207	2.821	-0.518	1.0	0.0	0.00

N Ga N	1.0	0.7661	0.000	0.1785	0.2017	-0.04524	1.0	0.0	0.00
N N Ga	1.0	0.001632	0.000	65.207	2.8210	-0.518	1.0	0.0	0.00
Ga N Ga	1.0	0.007874	1.846	1.9180	0.7500	-0.3013	1.0	0.0	0.00
$\mathrm{Sb}~\mathrm{Sb}~\mathrm{N}$	1.0	0.00	0.00	0.00	1.00	0.00	1.0	1.0	0.0
N Sb N	1.0	0.00	0.00	0.00	1.00	0.00	1.0	1.0	0.0
N N Sb	1.0	0.00	0.00	0.00	1.00	0.00	1.0	1.0	0.0
$\rm Sb~N~Sb$	1.0	0.00	0.00	0.00	1.00	0.00	1.0	1.0	0.0
Ga Ga Sb	1.0	0.00	0.00	0.00	1.00	0.00	1.0	1.0	0.0
Sb Ga Sb	1.0	0.00	0.00	0.00	1.00	0.00	1.0	1.0	0.0
Ga Sb Ga	1.0	0.00	0.00	0.00	1.00	0.00	1.0	1.0	0.0
Sb Sb Ga	1.0	0.00	0.00	0.00	1.00	0.00	1.0	1.0	0.0
Ga N Sb	1.0	0.001632	0.000	65.207	2.821	-0.518	1.0	1.0	2.6391
Ga Sb N	3.0	0.3630	0.9687	1.2088	0.8398	-0.4277	4.6022	1.0	1.7452
N Ga Sb	1.0	0.0016	0.000	65.2070	2.8210	-0.518	1.0	1.0	2.6391
N Sb Ga	1.0	0.00	0.00	0.00	1.00	0.00	1.0	1.0	0.0
Sb N Ga	1.0	0.00	0.00	0.00	1.00	0.00	1.0	1.0	0.0
Sb Ga N	3.0	0.3630	0.9687	1.2088	0.8397	-0.4277	4.6022	1.0	1.7452

Table E.2: Tersoff potentials for GaNSb alloys (Part 2).

	В	R	D	λ_1	A	
Ga Ga Ga	410.132	2.87	0.15	1.6092	535.199	
ΝΝΝ	423.769	2.20	0.20	3.5578	1044.77	
Sb Sb Sb	0.00	2.87	0.15	0.00	0.00	
Ga N N	3864.27	2.90	0.20	2.9352	6136.44	
N Ga Ga	3864.27	2.90	0.20	2.9352	6136.44	
Ga Sb Sb	544.9039	3.5	0.1	2.5024	2521.7569	

Sb Ga Ga	544.9039	3.5	0.1	2.5024	2521.7569	
Sb N N	0.00	2.87	0.15	0.00	0.00	
N Sb Sb	0.00	2.87	0.15	0.00	0.00	
Ga Ga N	0.00	2.90	0.20	0.00000	0.00000	
N Ga N	0.00	2.20	0.20	0.00000	0.00000	
N N Ga	0.00	2.90	0.20	0.00000	0.00000	
Ga N Ga	0.00	2.87	0.15	0.00000	0.00000	
Sb Sb N	0.00	2.87	0.15	0.00	0.00	
N Sb N	0.00	2.87	0.15	0.00	0.00	
N N Sb	0.00	2.87	0.15	0.00	0.00	
Sb N Sb	0.00	2.87	0.15	0.00	0.00	
Ga Ga Sb	0.00	2.87	0.15	0.00	0.00	
Sb Ga Sb	0.00	2.87	0.15	0.00	0.00	
Ga Sb Ga	0.00	2.87	0.15	0.00	0.00	
Sb Sb Ga	0.00	2.87	0.15	0.00	0.00	
Ga N Sb	3864.27	2.90	0.20	2.9352	6136.44	
Ga Sb N	544.9039	3.5	0.1	2.5024	2521.7569	
N Ga Sb	3864.27	2.90	0.20	2.9352	6136.44	
N Sb Ga	0.00	2.87	0.15	0.00	0.00	
Sb N Ga	0.00	2.87	0.15	0.00	0.00	
Sb Ga N	544.9039	3.5	0.1	2.5024	2521.7569	

\mathbf{F}

NECPA EQUATION SET IN LMTO-ASA

The NECPA equation set in LMTO-ASA is listed as follows:

$$\overline{g_{\mathbf{R},\mathbf{R}}^{\alpha,r}} = \sum_{Q} c_{\mathbf{R}}^{Q} \overline{g_{\mathbf{R},\mathbf{R}}^{\alpha,Q,r}},$$

$$\overline{g^{\alpha,r}} = [\mathcal{P}^{\alpha,r} - S^{\alpha} - \Sigma^{r}]^{-1},$$

$$\overline{g_{\mathbf{R},\mathbf{R}}^{\alpha,r}} = [\overline{g^{\alpha,r}}]_{\mathbf{R},\mathbf{R}},$$

$$\overline{g_{\mathbf{R},\mathbf{R}}^{\alpha,r}} = [\mathcal{P}_{\mathbf{R}}^{\alpha,r} - \Omega_{\mathbf{R}}^{r}]^{-1},$$

$$\overline{g_{\mathbf{R},\mathbf{R}}^{\alpha,Q,r}} = \left[P_{\mathbf{R}}^{\alpha,Q} - \Omega_{\mathbf{R}}^{r}\right]^{-1}.$$
(F.1)

and

$$\overline{g_{\mathbf{R},\mathbf{R}}^{\alpha,<}} = \sum_{Q} c_{\mathbf{R}}^{Q} \overline{g_{\mathbf{R},\mathbf{R}}^{\alpha,Q,<}},$$

$$\overline{g^{\alpha,<}} = \overline{g^{\alpha,r}} \left[-\mathcal{P}^{\alpha,<} + \Sigma^{<} \right] \overline{g^{\alpha,a}},$$

$$\overline{g_{\mathbf{R},\mathbf{R}}^{\alpha,<}} = \left[\overline{g^{\alpha,<}} \right]_{\mathbf{R},\mathbf{R}},$$

$$\overline{g_{\mathbf{R},\mathbf{R}}^{\alpha,<}} = \overline{g_{\mathbf{R},\mathbf{R}}^{\alpha,r}} \left(-\mathcal{P}_{\mathbf{R}}^{\alpha,<} + \Omega_{\mathbf{R}}^{<} \right) \overline{g_{\mathbf{R},\mathbf{R}}^{\alpha,a}},$$

$$\overline{g_{\mathbf{R},\mathbf{R}}^{\alpha,Q,<}} = \overline{g_{\mathbf{R},\mathbf{R}}^{\alpha,Q,r}} \Omega_{\mathbf{R}}^{<} \overline{g_{\mathbf{R},\mathbf{R}}^{\alpha,Q,a}}.$$
(F.2)

 $\overline{g_{\mathbf{R},\mathbf{R}}^{\alpha,r/<}}$ is retarded/less auxiliary Green's function. The ones with superscript Q are their corresponding conditional average Green's functions. $c_{\mathbf{R}}^{Q}$ is the concentration of specie Q at site \mathbf{R} .

BIBLIOGRAPHY

- Gordon E. Moore. Cramming more components onto integrated circuits. *Electronics*, pages 114–117, APRIL 1965.
- [2] Y.P. Tsividis. Operation and modeling of the MOS transistor. McGraw-Hill, New York, 1987.
- [3] Ronald R. Troutman. Vlsi limitations from drain-induced barrier lowering. IEEE JOURNAL OF SOLID-STATE CIRCUITS, 14(2):383–391, APRIL 1979.
- [4] Tomasz Skotnicki, Gerard Merckel, and Thierry Pederon. The voltage-doping transformation: A new approach to the modeling of mosfet s hort-c hannel effects. *IEEE ELECTRON DEVICE LETTERS*, 9(3), MARCH 1988.
- [5] R. H. Crawford. MOSFET in CircuitDesign. McGraw-Hill, New York, 1967.
- [6] G. Baum and H. Beneking. Drift velocity saturation in mos transistors. IEEE TRANSACTIONS ON ELECTRON DEVICES, 17(6), JUNE 1979.
- [7] J. H. Stathis and D. J. DiMaria. Reliability projection for ultra-thin oxides at low voltage. In *IEDM Tech. Dig.*, pages 167–170, 1998.
- [8] Yiming Li, Kuo-Fu Lee, Chun-Yen Yiu, Yung-Yueh Chiu, and Ru-Wei Chang. Dual-material gate approach to suppression of random-dopant-induced characteristic fluctuation in 16 nm metal-oxide-semiconductor field-effect-transistor devices. Japanese Journal of Applied Physics, 50, 2011.
- [9] A. Asenov, Andrew R. Brown, John H. Davies, Savas Kaya, and Gabriela Slavcheva. Simulation of intrinsic parameter fluctuations in decananometer

and nanometer-scale mosfets. *IEEE TRANSACTIONS ON ELECTRON DE-VICES*, 50(9), SEPTEMBER 2003.

- [10] Hsin-Wen Su, Yiming Li, Yu-Yu Chen, Chieh-Yang Chen, and Han-Tung Chang. Drain-induced-barrier lowering and subthreshold swing fluctuations in 16-nm-gate bulk finfet devices induced by random discrete dopants. In *Device Research Conference (DRC)*, pages 131–134, 2012.
- [11] K. Takeuchi, T. Fukai, T. Tsunomura, A. T. Putra, A. Nishida, S. Kamohara, and T. Hiramoto. Understanding random threshold voltage fluctuation by comparing multiple fabs and technologies. In *Electron Devices Meeting*, 2007. IEDM 2007. IEEE International, pages 467–470, 2007.
- [12] Fu-Liang Yang, Jiunn-Ren Hwang, Hung-Ming Chen, Jeng-Jung Shen, Shao-Ming Yu, Yiming Li, and D. D. Tang. Discrete dopant fluctuated 20 nm/15 nm gate planar cmos. In VLSI Technology, 2007 IEEE Symposium on, pages 208– 209, JUNE 2007.
- [13] R Michael Sheetz, Ernst Richter, Antonis N Andriotis, Sergey Lisenkov, Chandrashekhar Pendyala, Mahendra K Sunkara, and Madhu Menon. Visible-light absorption and large band-gap bowing of gan 1- x sb x from first principles. *Physical Review B*, 84(7):075304, 2011.
- [14] Xunyu Yang, Abraham Wolcott, Gongming Wang, Alissa Sobo, Robert Carl Fitzmorris, Fang Qian, Jin Z Zhang, and Yat Li. Nitrogen-doped zno nanowire arrays for photoelectrochemical water splitting. *Nano letters*, 9(6):2331–2336, 2009.
- [15] Poul Georg Moses and Chris G Van de Walle. Band bowing and band alignment in ingan alloys. Applied Physics Letters, 96(2):021908, 2010.
- [16] A. Asenov. Simulation of statistical variability in nano mosfets. In VLSI Technology, 2007 IEEE Symposium on, pages 86–87, JUNE 2007.

- [17] A. Asenov. Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 μm mosfet's: A 3-d "atomistic" simulation study. *IEEE TRANSAC-TIONS ON ELECTRON DEVICES*, 45(12):2505–2513, DECEMBER 1998.
- [18] Changhwan Shin, Xin Sun, and Tsu-Jae King Liu. Study of randomdopantfluctuation (rdf) effects for the trigate bulk mosfet. *IEEE TRANSAC-TIONS ON ELECTRON DEVICES*, 56(7):1538–1542, JULY 2009.
- [19] Troels Markussen, Riccardo Rurali, Antti-Pekka Jauho, and Mads Brandbyge. Scaling theory put into practice: First-principles modeling of transport in doped silicon nanowires. *Phys. Rev. Lett.*, 99(7):076803, AUGUST 2007.
- [20] E. Cruz-Silva, Z. M. Barnett, B. G. Sumpter, and V. Meunier. Structural, magnetic, and transport properties of substitutionally doped graphene nanoribbons from first principles. *Phys. Rev. B*, 83(15):155445, APRIL 2011.
- [21] Subhasish Mandal and Ranjit Pati. Structural, magnetic, and transport properties of substitutionally doped graphene nanoribbons from first principles. *Phys. Rev. B*, 83(19):195420, MAY 2011.
- [22] Zhibin Ren. NANOSCALE MOSFETS: PHYSICS, SIMULATION AND DE-SIGN. PhD thesis, Purdue University, Western Lafayette, USA, 2001.
- [23] Farzin Assad Assad, Zhibin Ren, Dragica Vasileska, Supriyo Datta, and Mark Lundstrom. On the performance limits for si mosfet's: A theoretical study. *IEEE TRANSACTIONS ON ELECTRON DEVICES*, 47(1):232–240, JAN-UARY 2000.
- [24] Zhibin Ren Ren, Ramesh Venugopal, Supriyo Datta, and Mark Lundstrom. The ballistic nanotransistor: A simulation study. In *Electron Devices Meeting*, 2000. *IEDM '00. Technical Digest. International*, pages 715–718, 2000.

- [25] Farzin Assad, Zhibin Ren, Supriyo Datta, and Mark Lundstrom. Performance limits of silicon mosfet's. In *Electron Devices Meeting*, 1999. IEDM '99. Technical Digest. International, pages 547–550, 1999.
- [26] Zhibin Ren, Ramesh Venugopal, Supriyo Datta, and Mark Lundstrom. Examination of design and manufacturing issues in a 10 nm double gate mosfet using nonequilibrium green's function simulation. In *Electron Devices Meeting*, 2001. *IEDM '01. Technical Digest. International*, pages 107–110, 2001.
- [27] Jeremy Taylor, Hong Guo, and Jian Wang. Ab initio modeling of quantum transport properties of molecular electronic devices. *Phys. Rev. B*, 63(24):245407, JUNE 2001.
- [28] Jeremy Taylor, Hong Guo, and Jian Wang. Ab initio modeling of open systems: Charge transfer, electron conduction, and molecular switching of a c60 device. *Phys. Rev. B*, 63(12):121104, MARCH 2001.
- [29] Manuel Smeu, Robert A. Wolkow, and Hong Guo. Conduction pathway of 1/4-stacked ethylbenzene molecular wires on si(100). Journal of the American Chemical Society, 131(31):11019–11026, 2009.
- [30] Chao-Cheng Kaun and Hong Guo. Ab initio modeling of open systems: Charge transfer, electron conduction, and molecular switching of a c60 device. Nano Letters, 3(11):1521–1525, 2003.
- [31] Yi Liu and Hong Guo. Current distribution in b- and n-doped carbon nanotubes. *Phys. Rev. B*, 69(11):115401, MARCH 2004.
- [32] Derek Waldron, Paul Haney, Brian Larade, Allan MacDonald, and Hong Guo. Nonlinear spin current and magnetoresistance of molecular tunnel junctions. *Phys. Rev. Lett.*, 96(16):166804, APRIL 2006.
- [33] Derek Waldron, Lei Liu, and Hong Guo. Ab initio simulation of magnetic tunnel junctionss. Nanotechnology, 18(42):424026, 2007.

- [34] Zhanyu Ning, Yu Zhu, Jian Wang, and Hong Guo. Nonlinear spin current and magnetoresistance of molecular tunnel junctions. *Phys. Rev. Lett.*, 100(5):056803, FEBRUARY 2008.
- [35] Jesse Maassen, Wei Ji, and Hong Guo. Ab initio modeling of open systems: Charge transfer, electron conduction, and molecular switching of a c60 device. Nano Letters, 11(1):151–155, 2010.
- [36] Youqi Ke, Ferdows Zahid, V. Timoshevskii, Ke Xia, D. Gall, and Hong Guo. Resistivity of thin cu films with surface roughness. *Phys. Rev. B*, 79(15):155406, APRIL 2009.
- [37] Youqi Ke, Ke Xia, and Hong Guo. Disorder scattering in magnetic tunnel junctions: Theory of nonequilibrium vertex correction. *Phys. Rev. Lett.*, 100(16):166805, 2008.
- [38] Yu Zhu, Lei Liu, and Hong Guo. Quantum transport theory with nonequilibrium coherent potentials. *Physical Review B*, 88(20):205415, 2013.
- [39] U. V. Barth and L. Hedin. A local exchange-correlation potential for the spin polarized case: I. J. Phys. C: Solid State Phys., 5(13):1629, 1972.
- [40] J. Harris and R. O. Jones. The surface energy of a bounded electron gas. J. Phys. F: Met. Phys., 4(8):1170, 1974.
- [41] J. Harris and R. O. Jones. Self-consistent-field-korringa-kohn-rostokercoherentpotential approximation for random alloys. Z. Phys. B: Condens. Matter, 46(2):95–98, 1982.
- [42] O.K. Anderson and O. Jepsen. Explicit, first-principles tight-binding theory. *Phys. Rev. Lett.*, 53(27):2571, 1984.
- [43] J. Kudrnovsky and V. Drchal. Electronic structure of random alloys by the linear band-structure methods. *Phys. Rev. B*, 41(11):7515, 1990.

- [44] J. Kudrnovsky, V. Drchal, and J. Masek. Canonical description of electron states in random alloys. *Phys. Rev. B*, 35(5):2487, 1987.
- [45] K. Carva, I. Turek, J. Kudrnovsky, and O. Bengone. Disordered magnetic multilayers: Electron transport within the coherent potential approximation. *Phys. Rev. B*, 73(14):144421, 2006.
- [46] B. Velicky, S. Kirkpatrick, and H. Ehrenreich. Single-site approximations in the electronic theory of simple binary alloys. *Phys. Rev.*, 175(3):747, 1968.
- [47] Youqi ke, Ke Xia, and Hong Guo. Single-site approximations in the electronic theory of simple binary alloys. *Phys. Rev. Lett.*, 105(23):236801, 2010.
- [48] Ferdows Zahid, Youqi Ke, Daniel Gall, and Hong Guo. Resistivity of thin cu films coated with ta, ti, ru, al, and pd barrier layers from first principles. *Phys. Rev. B*, 81(4):045406, 2010.
- [49] Jesse Maassen and Hong Guo. Suppressing leakage by localized doping in si nanotransistor channels. *Phys. Rev. Lett.*, 109(26):266803, 2012.
- [50] Qing Shi, Hong Guo, Yu Zhu, and Lei Liu. Analyzing variability in shortchannel quantum transport from atomistic first principles. *Physical Review Applied*, 3(6):064008, 2015.
- [51] Qing Shi, Lining Zhang, Yu Zhu, Lei Liu, Mansun Chan, and Hong Guo. Atomic disorder scattering in emerging transistors by parameter-free first principle modeling. In *Electron Devices Meeting (IEDM), 2014 IEEE International*, pages 30–6. IEEE, 2014.
- [52] Klaus Koepernik, B Velickỳ, Roland Hayn, and Helmut Eschrig. Self-consistent lcao-cpa method for disordered alloys. *Physical Review B*, 55(9):5717, 1997.
- [53] Alex Zunger, S-H Wei, LG Ferreira, and James E Bernard. Special quasirandom structures. *Physical Review Letters*, 65(3):353, 1990.

- [54] I. Turek, V. Drchal, J. Kudrnovshy, M. Sob, and P. Weinberger. *Electronic structure of disordered alloys, surfaces and interfaces*. Kluwer Academic Publishers, Boston, 1997.
- [55] Federico Rastrelli and Alessandro Bagno. Predicting the nmr spectra of paramagnetic molecules by dft: Application to organic free radicals and transitionmetal complexes. *Chemistry - A European Journal*, 15(32):7990–8004, AU-GUST 2009.
- [56] Christian Tuma, A. Daniel Boese, and Nicholas C. Handy. Predicting the binding energies of h-bonded complexes: A comparative dft study. *Phys. Chem. Chem. Phys.*, 1(17):3939–3947, 1999.
- [57] Jean-Francois Paul and Edmond Payen. Vacancy formation on mos2 hydrodesulfurization catalyst: Dft study of the mechanism. J. Phys. Chem. B, 107(17):4057–4064, 2003.
- [58] G. Giovannetti, P. A. Khomyakov, G. Brocks, V. M. Karpan, J. van den Brink, and P. J. Kelly. Doping graphene with metal contacts. *Phys. Rev. Lett.*, 101(2):026803, 2008.
- [59] Laurie J. Bulter. Chemical reaction dynamics beyond the born-oppenheimer approximation. Annu. Rev. Phys. Chem., 49:125–171, 1998.
- [60] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. Phys. Rev., 136(3B):B864, 1964.
- [61] W. Kohn and L. J. Sham. Self-cnosistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133, 1965.
- [62] R. G. Parr and W. Yang. Density-Functional Theory of Atoms and Molecules. Oxford University Press, New York, 1989.
- [63] Robert O Jones and Olle Gunnarsson. The density functional formalism, its applications and prospects. *Reviews of Modern Physics*, 61(3):689, 1989.

- [64] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, 45(7):566–569, AUGUST 1980.
- [65] S. Goedecker, M. Teter, and J. Hutter. Separable dual-space gaussian pseudopotentials. *Phys. Rev. B*, 54(3):1703–1710, JULY 1996.
- [66] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Separable dual-space gaussian pseudopotentials. *Phys. Rev. Lett.*, 77(18):3865–3868, OCTOBER 1996.
- [67] J. P. Perdew. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23(10):5048–5079, 1981.
- [68] J. P. Perdew and Y. Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B*, 45(23):13244, 1992.
- [69] Paul AM Dirac. Note on exchange phenomena in the thomas atom. In Mathematical Proceedings of the Cambridge Philosophical Society, volume 26, pages 376–385. Cambridge University Press, 1930.
- [70] David M Ceperley and BJ Alder. Ground state of the electron gas by a stochastic method. *Physical Review Letters*, 45(7):566, 1980.
- [71] Seymour H Vosko, Leslie Wilk, and Marwan Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of physics*, 58(8):1200–1211, 1980.
- [72] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.
- [73] Fabien Tran and Peter Blaha. Accurate band gaps of semiconductors and insulators with a semilocal exchange-correlation potential. *Phys. Rev. Lett.*, 102(22):226401, 2009.

- [74] Jochen Heyd, Gustavo E Scuseria, and Matthias Ernzerhof. Hybrid functionals based on a screened coulomb potential. The Journal of chemical physics, 118(18):8207–8215, 2003.
- [75] Axel D Becke. Density-functional thermochemistry. iii. the role of exact exchange. The Journal of chemical physics, 98(7):5648–5652, 1993.
- [76] Ying-Chih Chen, Jing-Zhe Chen, Vincent Michaud-Rioux, Qing Shi, and Hong Guo. Efficient evaluation of nonlocal operators in density functional theory. *Physical Review B*, 97(7):075139, 2018.
- [77] John P Perdew, Matthias Ernzerhof, and Kieron Burke. Rationale for mixing exact exchange with density functional approximations. *The Journal of chemical physics*, 105(22):9982–9985, 1996.
- [78] Vincent Michaud-Rioux, Lei Zhang, and Hong Guo. Rescu: A real space electronic structure method. *Journal of Computational Physics*, 307:593–613, 2016.
- [79] Jochen Heyd, Juan E Peralta, Gustavo E Scuseria, and Richard L Martin. Energy band gaps and lattice parameters evaluated with the heydscuseria-ernzerhof screened hybrid functional. *The Journal of chemical physics*, 123(17):174101, 2005.
- [80] José M Soler, Emilio Artacho, Julian D Gale, Alberto García, Javier Junquera, Pablo Ordejón, and Daniel Sánchez-Portal. The siesta method for ab initio order-n materials simulation. *Journal of Physics: Condensed Matter*, 14(11):2745, 2002.
- [81] Jeremy Taylor. ab initio modeling of Transport in Atomic Scale Devices. PhD thesis, McGill University, Montreal, Canada, 2000.
- [82] DR Hamann. Optimized norm-conserving vanderbilt pseudopotentials. *Physical Review B*, 88(8):085117, 2013.

- [83] Martin Schlipf and François Gygi. Optimization algorithm for the generation of oncv pseudopotentials. *Computer Physics Communications*, 196:36–44, 2015.
- [84] Georg Kresse and D Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59(3):1758, 1999.
- [85] Peter E Blöchl. Projector augmented-wave method. Physical review B, 50(24):17953, 1994.
- [86] N. W. Ashcroft and N. D. Mermin. Solid State Physics. Thomson Learning, Toronto, 1976.
- [87] O. Krogh Anderson. Linear methods in band theory. Phys. Rev. B, 12(8):3060– 3083, 1975.
- [88] O. K. Andersen. Simple approach to the band-structure problem. Solid State Communications, 13(2):133–136, 1973.
- [89] H. L. Skriver. The LMTO method. Springer, Berlin, 1984.
- [90] Jesse Maassen. First principles simulations of nanoelectronic devices. PhD thesis, McGill University, Montreal, Canada, 2011.
- [91] O. K. Andersen, Z. Pawlowska, and O. Jepsen. Illustration of the linearmuffintin-orbital tight-binding representation: Compact orbitals and charge density in si. *Phys. Rev. B*, 34(8):5253–5269, October 1986.
- [92] Georg Kresse and Jürgen Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical review B*, 54(16):11169, 1996.
- [93] Junqiao Wu. When group-iii nitrides go infrared: New properties and perspectives. Journal of Applied Physics, 106(1):5, 2009.
- [94] Yajie Dong, Bozhi Tian, Thomas J Kempa, and Charles M Lieber. Coaxial group iii- nitride nanowire photovoltaics. Nano letters, 9(5):2183–2187, 2009.

- [95] Yat Li, Jie Xiang, Fang Qian, Silvija Gradecak, Yue Wu, Hao Yan, Douglas A Blom, and Charles M Lieber. Dopant-free gan/aln/algan radial nanowire heterostructures as high electron mobility transistors. *Nano letters*, 6(7):1468–1473, 2006.
- [96] Md Golam Kibria, Ruimin Qiao, Wanli Yang, Idris Boukahil, Xianghua Kong, Faqrul Alam Chowdhury, Michel L Trudeau, Wei Ji, Hong Guo, FJ Himpsel, et al. Atomic-scale origin of long-term stability and high performance of pgan nanowire arrays for photocatalytic overall pure water splitting. Advanced Materials, 28(38):8388–8397, 2016.
- [97] Kazuhiko Maeda, Kentaro Teramura, Daling Lu, Tsuyoshi Takata, Nobuo Saito, Yasunobu Inoue, and Kazunari Domen. Photocatalyst releasing hydrogen from water. *Nature*, 440(7082):295, 2006.
- [98] Shizhao Fan, Bandar AlOtaibi, Steffi Y Woo, Yongjie Wang, Gianluigi A Botton, and Zetian Mi. High efficiency solar-to-hydrogen conversion on a monolithically integrated ingan/gan/si adaptive tunnel junction photocathode. *Nano letters*, 15(4):2721–2726, 2015.
- [99] B AlOtaibi, HPT Nguyen, S Zhao, MG Kibria, S Fan, and Z Mi. Highly stable photoelectrochemical water splitting and hydrogen generation using a doubleband ingan/gan core/shell nanowire photoanode. Nano letters, 13(9):4356– 4361, 2013.
- [100] Chong Liu, Jinyao Tang, Hao Ming Chen, Bin Liu, and Peidong Yang. A fully integrated nanosystem of semiconductor nanowires for direct solar water splitting. *Nano letters*, 13(6):2989–2992, 2013.
- [101] Bandar AlOtaibi, Shizhao Fan, Defa Wang, Jinhua Ye, and Zetian Mi. Waferlevel artificial photosynthesis for co2 reduction into ch4 and co using gan nanowires. ACS Catalysis, 5(9):5342–5348, 2015.

- [102] Sergio Trasatti. The absolute electrode potential: an explanatory note (recommendations 1986). Pure and Applied Chemistry, 58(7):955–966, 1986.
- [103] Michael G Walter, Emily L Warren, James R McKone, Shannon W Boettcher, Qixi Mi, Elizabeth A Santori, and Nathan S Lewis. Solar water splitting cells. *Chemical reviews*, 110(11):6446–6473, 2010.
- [104] S Strite and Hy Morkoç. Gan, aln, and inn: a review. Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena, 10(4):1237–1266, 1992.
- [105] Kwiseon Kim, Walter RL Lambrecht, and Benjamin Segall. Elastic constants and related properties of tetrahedrally bonded bn, aln, gan, and inn. *Physical Review B*, 53(24):16310, 1996.
- [106] AF Wright. Elastic properties of zinc-blende and wurtzite aln, gan, and inn. Journal of Applied physics, 82(6):2833–2839, 1997.
- [107] Fabio Bernardini and Vincenzo Fiorentini. Spontaneous versus piezoelectric polarization in iii–v nitrides: conceptual aspects and practical consequences. *physica status solidi* (b), 216(1):391–398, 1999.
- [108] AE Romanov, TJ Baker, S Nakamura, JS Speck, and ERATO/JST UCSB Group. Strain-induced polarization in wurtzite iii-nitride semipolar layers. *Jour*nal of Applied Physics, 100(2):023522, 2006.
- [109] KM Yu, WL Sarney, SV Novikov, D Detert, R Zhao, JD Denlinger, SP Svensson, OD Dubon, W Walukiewicz, and CT Foxon. Highly mismatched n-rich gan1- xsbx films grown by low temperature molecular beam epitaxy. Applied Physics Letters, 102(10):102104, 2013.
- [110] Swathi Sunkara, Venkat Kalyan Vendra, Jacek Bogdan Jasinski, Todd Deutsch, Antonis N Andriotis, Krishna Rajan, Madhu Menon, and Mahendra Sunkara.

New visible light absorbing materials for solar fuels, ga (sbx) n1- x. Advanced Materials, 26(18):2878–2882, 2014.

- [111] J Wu, W Shan, and W Walukiewicz. Band anticrossing in highly mismatched iii–v semiconductor alloys. Semiconductor Science and Technology, 17(8):860, 2002.
- [112] Website: https://www.brown.edu/departments/engineering/labs/avdw/atat/.
- [113] Alfonso Franciosi and Chris G Van de Walle. Heterojunction band offset engineering. Surface Science Reports, 25(1-4):1–140, 1996.
- [114] Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. Journal of computational physics, 117(1):1–19, 1995.
- [115] J Tersoff. Modeling solid-state chemistry: Interatomic potentials for multicomponent systems. *Physical Review B*, 39(8):5566, 1989.
- [116] D Powell, MA Migliorato, and AG Cullis. Optimized tersoff potential parameters for tetrahedrally bonded iii-v semiconductors. *Physical Review B*, 75(11):115202, 2007.
- [117] I Vurgaftman and JR n Meyer. Band parameters for nitrogen-containing semiconductors. Journal of Applied Physics, 94(6):3675–3696, 2003.
- [118] Amrit De and Craig E Pryor. Predicted band structures of iii-v semiconductors in the wurtzite phase. *Physical Review B*, 81(15):155210, 2010.
- [119] MG Kibria, FA Chowdhury, S Zhao, B AlOtaibi, ML Trudeau, H Guo, and Z Mi. Visible light-driven efficient overall water splitting using p-type metalnitride nanowire arrays. *Nature communications*, 6:6797, 2015.
- [120] N Segercrantz, KM Yu, M Ting, WL Sarney, SP Svensson, SV Novikov, CT Foxon, and W Walukiewicz. Electronic band structure of highly mismatched gan1- xsbx alloys in a broad composition range. Applied Physics Letters, 107(14):142104, 2015.

- [121] John P Perdew, Adrienn Ruzsinszky, Gábor I Csonka, Oleg A Vydrov, Gustavo E Scuseria, Lucian A Constantin, Xiaolan Zhou, and Kieron Burke. Restoring the density-gradient expansion for exchange in solids and surfaces. *Physical Review Letters*, 100(13):136406, 2008.
- [122] Supriyo Datta. Electronic Transport in Mesoscopic Systems. Cambridge University Press, New York, 1995.
- [123] Hartmut Haug and Antti-Pekka Jauho. Quantum Kinetics in Transport and Optics of Semiconductors. Springer-Verlag, Berlin, 1998.
- [124] M. P. Lopez Sancho, J. M. Lopez Sancho, and J. Rubio. Highly convergent schemes for the calculation of bulk and surface green's functions. J. Phys. F: Met. Phys., 15(4):851–858, 1985.
- [125] Sergey V. Faleev, Francois Leonard, Derek A. Stewart, and Mark van Schilfgaarde. Ab initio tight-binding lmto method for nonequilibrium electron transport in nanosystems. Phys. Rev. B, 71(19):195422, May 2005.
- [126] Youqi Ke. Theory of Non-equilibrium Vertex Correction. PhD thesis, McGill University, Montreal, Canada, 2010.
- [127] Asen Asenov. Random dopant induced threshold voltage lowering and fluctuations in sub-0.1/spl mu/m mosfet's: A 3-d" atomistic" simulation study. *IEEE Transactions on Electron Devices*, 45(12):2505–2513, 1998.
- [128] Yu Zhu, Lei Liu, and Hong Guo. Green's function theory for predicting deviceto-device variability. *Phys. Rev. B*, 88(8):085420, 2013.
- [129] P. Soven. Coherent-potential model of substitutional disordered alloys. Phys. Rev., 156(3):809–813, 1967.
- [130] A Bansil, S Kaprzyk, PE Mijnarends, and J Toboła. Electronic structure and magnetism of fe 3- x v x x (x= si, ga, and al) alloys by the kkr-cpa method. *Physical Review B*, 60(19):13396, 1999.
- [131] Qing Shi. A first principles investigation of device-to-device variability. Master's thesis, McGill University, Montreal, Canada, 2013.
- [132] Robert W Dutton and Zhiping Yu. Technology CAD Computer Simulation of IC Processes and Devices, volume 243. Springer Science & Business Media, 2012.
- [133] Qingkai Yu, Luis A Jauregui, Wei Wu, Robert Colby, Jifa Tian, Zhihua Su, Helin Cao, Zhihong Liu, Deepak Pandey, Dongguang Wei, et al. Control and characterization of individual grains and grain boundaries in graphene grown by chemical vapour deposition. *Nature materials*, 10(6):443, 2011.
- [134] Hui Fang, Mahmut Tosun, Gyungseon Seol, Ting Chia Chang, Kuniharu Takei, Jing Guo, and Ali Javey. Degenerate n-doping of few-layer transition metal dichalcogenides by potassium. *Nano letters*, 13(5):1991–1995, 2013.
- [135] Nazila Harati Pour, Yoska Anugrah, Sangfeng Wu, Xiaodong Xu, and Steven J Koester. Chemical doping for threshold control and contact resistance reduction in graphene and mos 2 field effect transistors. In *Device Research Conference* (DRC), 2013 71st Annual, pages 101–102. IEEE, 2013.
- [136] D Basu, MJ Gilbert, LF Register, Sanjay K Banerjee, and Allan H MacDonald. Effect of edge roughness on electronic transport in graphene nanoribbon channel metal-oxide-semiconductor field-effect transistors. *Applied Physics Letters*, 92(4):042114, 2008.
- [137] Asen Asenov, Fikru Adamu-Lema, Xingsheng Wang, and Salvatore Maria Amoroso. Problems with the continuous doping tcad simulations of decananometer cmos transistors. *IEEE Transactions on Electron Devices*, 61(8):2745–2751, 2014.
- [138] Timothy B Boykin, Mathieu Luisier, Andreas Schenk, Neerav Kharche, and Gerhard Klimeck. The electronic structure and transmission characteristics of

disordered algaas nanowires. *IEEE transactions on nanotechnology*, 6(1):43–47, 2007.

- [139] Andre K Geim and Konstantin S Novoselov. The rise of graphene. In Nanoscience and Technology: A Collection of Reviews from Nature Journals, pages 11–19. World Scientific, 2010.
- [140] Qimin Yan, Bing Huang, Jie Yu, Fawei Zheng, Ji Zang, Jian Wu, Bing-Lin Gu, Feng Liu, and Wenhui Duan. Intrinsic current- voltage characteristics of graphene nanoribbon transistors and effect of edge doping. *Nano letters*, 7(6):1469–1473, 2007.
- [141] Melinda Y Han, Barbaros Özyilmaz, Yuanbo Zhang, and Philip Kim. Energy band-gap engineering of graphene nanoribbons. *Physical review letters*, 98(20):206805, 2007.
- [142] Xiaolin Li, Xinran Wang, Li Zhang, Sangwon Lee, and Hongjie Dai. Chemically derived, ultrasmooth graphene nanoribbon semiconductors. *science*, 319(5867):1229–1232, 2008.
- [143] Xinran Wang, Yijian Ouyang, Xiaolin Li, Hailiang Wang, Jing Guo, and Hongjie Dai. Room-temperature all-semiconducting sub-10-nm graphene nanoribbon field-effect transistors. *Physical review letters*, 100(20):206803, 2008.
- [144] Alessandro Betti, Gianluca Fiori, and Giuseppe Iannaccone. Strong mobility degradation in ideal graphene nanoribbons due to phonon scattering. Applied Physics Letters, 98(21):212111, 2011.
- [145] Bing Huang, Qiang Xu, and Su-Huai Wei. Theoretical study of corundum as an ideal gate dielectric material for graphene transistors. *Physical Review B*, 84(15):155406, 2011.
- [146] Gianluca Giovannetti, Petr A Khomyakov, Geert Brocks, Paul J Kelly, and Jeroen Van Den Brink. Substrate-induced band gap in graphene on hexagonal

boron nitride: Ab initio density functional calculations. *Physical Review B*, 76(7):073103, 2007.

- [147] Cory R Dean, Andrea F Young, Inanc Meric, Chris Lee, Lei Wang, Sebastian Sorgenfrei, Kenji Watanabe, Takashi Taniguchi, Phillip Kim, Kenneth L Shepard, et al. Boron nitride substrates for high-quality graphene electronics. *Nature nanotechnology*, 5(10):722, 2010.
- [148] Regiane Nascimento, Jonathan da Rocha Martins, Ronaldo JC Batista, and Helio Chacham. Band gaps of bn-doped graphene: Fluctuations, trends, and bounds. *The Journal of Physical Chemistry C*, 119(9):5055–5061, 2015.
- [149] Ondrej L Krivanek, Matthew F Chisholm, Valeria Nicolosi, Timothy J Pennycook, George J Corbin, Niklas Dellby, Matthew F Murfitt, Christopher S Own, Zoltan S Szilagyi, Mark P Oxley, et al. Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy. *Nature*, 464(7288):571, 2010.
- [150] Sun-Min Jung, Eun Kwang Lee, Min Choi, Dongbin Shin, In-Yup Jeon, Jeong-Min Seo, Hu Young Jeong, Noejung Park, Joon Hak Oh, and Jong-Beom Baek. Direct solvothermal synthesis of b/n-doped graphene. Angewandte Chemie, 126(9):2430-2433, 2014.
- [151] Adrian M Ionescu and Heike Riel. Tunnel field-effect transistors as energyefficient electronic switches. *nature*, 479(7373):329, 2011.
- [152] Website: http://www.nanoacademic.ca/index.jsp.
- [153] S. M. Sze and Kwok K. Ng. Physics of Semiconductor Devices. Wiley-Interscience, New York, 2006.
- [154] L. Nordheim. Zur elektronentheorie der metalle. i. Ann. Phys. (Leipzig), 401(5):607–640, 1931.
- [155] Chenming Hu. Modern semiconductor devices for integrated circuits, volume 2. Prentice Hall Upper Saddle River, NJ, 2010.

- [156] Haiyan Jiang, Sihong Shao, Wei Cai, and Pingwen Zhang. Boundary treatments in non-equilibrium green's function (negf) methods for quantum transport in nano-mosfets. *Journal of Computational Physics*, 227(13):6553–6573, 2008.
- [157] S Bohloul, L Zhang, K Gong, and H Guo. Theoretical impurity-limited carrier mobility of monolayer black phosphorus. *Applied Physics Letters*, 108(3):033508, 2016.
- [158] Mathieu César, Dongping Liu, Daniel Gall, and Hong Guo. Calculated resistances of single grain boundaries in copper. *Physical Review Applied*, 2(4):044007, 2014.
- [159] Mathieu César, Daniel Gall, and Hong Guo. Reducing grain-boundary resistivity of copper nanowires by doping. *Physical Review Applied*, 5(5):054018, 2016.
- [160] A Svizhenko, MP Anantram, TR Govindan, B Biegel, and R Venugopal. Twodimensional quantum mechanical modeling of nanotransistors. *Journal of Applied Physics*, 91(4):2343–2354, 2002.
- [161] Donald G Anderson. Iterative procedures for nonlinear integral equations. Journal of the ACM (JACM), 12(4):547–560, 1965.
- [162] V Eyert. A comparative study on methods for convergence acceleration of iterative vector sequences. Journal of Computational Physics, 124(2):271–285, 1996.
- [163] N.W. Ashcroft and N.D. Mermin. Solid State Physics. Saunders College Publishing, Fort Worth, 1976.
- [164] Zi Wang, Youqi Ke, Dongping Liu, Hong Guo, and Kirk H Bevan. Low bias short channel impurity mobility in graphene from first principles. Applied Physics Letters, 101(9):093102, 2012.

- [165] SW Han, Hyuksang Kwon, Seong Keun Kim, Sunmin Ryu, Won Seok Yun, DH Kim, JH Hwang, J-S Kang, J Baik, HJ Shin, et al. Band-gap transition induced by interlayer van der waals interaction in mos 2. *Physical Review B*, 84(4):045409, 2011.
- [166] Kin Fai Mak, Changgu Lee, James Hone, Jie Shan, and Tony F Heinz. Atomically thin mos 2: a new direct-gap semiconductor. *Physical review letters*, 105(13):136805, 2010.
- [167] Branimir Radisavljevic, Aleksandra Radenovic, Jacopo Brivio, Valentina Giacometti, and Andras Kis. Single-layer mos 2 transistors. *Nature nanotechnology*, 6(3):147, 2011.
- [168] Kapildeb Dolui, Ivan Rungger, Chaitanya Das Pemmaraju, and Stefano Sanvito. Possible doping strategies for mos 2 monolayers: An ab initio study. *Physical Review B*, 88(7):075420, 2013.
- [169] Yung-Chang Lin, Dumitru O Dumcenco, Hannu-Pekka Komsa, Yoshiko Niimi, Arkady V Krasheninnikov, Ying-Sheng Huang, and Kazu Suenaga. Properties of individual dopant atoms in single-layer mos2: atomic structure, migration, and enhanced reactivity. *Advanced materials*, 26(18):2857–2861, 2014.
- [170] Lingming Yang, Kausik Majumdar, Han Liu, Yuchen Du, Heng Wu, Michael Hatzistergos, PY Hung, Robert Tieckelmann, Wilman Tsai, Chris Hobbs, et al. Chloride molecular doping technique on 2d materials: Ws2 and mos2. Nano letters, 14(11):6275–6280, 2014.
- [171] Lingming Yang, Kausik Majumdar, Yuchen Du, Han Liu, Heng Wu, Michael Hatzistergos, PY Hung, Robert Tieckelmann, Wilman Tsai, Chris Hobbs, et al. High-performance mos 2 field-effect transistors enabled by chloride doping: Record low contact resistance (0.5 kω· μm) and record high drain current (460 μa/μm). In 2014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers, pages 1–2. IEEE, 2014.

- [172] Levente Vitos. Total-energy method based on the exact muffin-tin orbitals theory. *Physical Review B*, 64(1):014107, 2001.
- [173] Levente Vitos, IA Abrikosov, and Börje Johansson. Anisotropic lattice distortions in random alloys from first-principles theory. *Physical review letters*, 87(15):156401, 2001.
- [174] Qingyun Zhang, Jiawei Yan, Yu Zhang, and Youqi Ke. Exact muffin tin orbital based first-principles method for electronic-structure and electron-transport simulation of device materials. *Physical Review B*, 100(7):075134, 2019.
- [175] Chenyi Zhou and Hong Guo. Nonequilibrium dual-fermion approach to electronic transport in disordered nanostructures. *Physical Review B*, 99(7):075414, 2019.
- [176] D. Zwillinger. Handbook of Differential Equations. Academic Press, New York, 1989.