First-principles Calculations on the Electronic, Vibrational, and Optical Properties of Semiconductor Nanowires

A Thesis
Presented to
The Academic Faculty
by
Li Yang

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Physics
Georgia Institute of Technology
December 2006
First-principles Calculations on the Electronic, Vibrational, and Optical Properties of Semiconductor Nanowires

Approved by:

Professor Mei-Yin Chou, Adviser  
School of Physics  
*Georgia Institute of technology*

Professor Andrew Zangwill  
School of Physics  
*Georgia Institute of technology*

Professor Uzi Landman  
School of Physics  
*Georgia Institute of technology*

Professor Phillip First  
School of Physics  
*Georgia Institute of technology*

Professor David Sherrill  
School of Chemistry and Biochemistry  
*Georgia Institute of technology*

Date Approved: July, 27, 2006
To my father,

Jinyue Yang,

who shows me the way to science.
ACKNOWLEDGEMENTS

I take this opportunity to thank all the people who have supported and helped me both in this work and on my path to this point.

Foremost, I am sincerely grateful to my thesis advisor, Professor Mei-Yin Chou, for her kindness, advice, and excellent supervision throughout the course of this work. Her exceptional insight and broad view led me to the wonderful field of the computational solid physics and helped me avoid so many detours during the progress of my research. In addition, her impressive carefulness and diligence have shown me how to work as a fruitful physicist, which will be my greatest asset in the future.

Special thanks are due to Dr. Xingyuan Zhao and Dr. J.A. Alford. Dr. Zhao led me to my first research project and gave me his creative advice. His solid physics knowledge helped me overcome difficulties at the beginning of my research. Dr. Alford took the time to answer questions and offer generous guidance during my learning of the GW code.

I would like to thank the friendly people in our research group: Mr. Lang Zeng, Ms. Zhu Ma, Mr. Pei Lin, Dr. A. Peles, Dr. Jia-An Yan, Dr. Wenying Ruan and Dr. Wolfgang Geist. Their help and enthusiasm for computational physics made this group an exciting and enjoyable place during my research. Thanks to Mrs. Judy Melton for her careful proofreading and feedbacks in writing my thesis.

There are so many great people at Georgia Tech who have helped me in the past six years that I cannot list all of them. I thank the physics faculty who, through the years of my education, provided a stimulating atmosphere for leaning.

Finally, I give my heartfelt thanks to Jing Yu, my wife. I could not have finished my PhD work without her quiet but steady support. Her love, friendship, and encouragement have made the past six years at Georgia Tech full of pleasure. I also thank my family for their love and emotional encouragement.
TABLE OF CONTENTS

DEDICATION ............................................................... iii
ACKNOWLEDGEMENTS ..................................................... iv
LIST OF TABLES .......................................................... vii
LIST OF FIGURES ........................................................ vix
SUMMARY ................................................................. xv
I  INTRODUCTION ......................................................... 1
II  THEORETICAL METHOD ............................................... 9
  2.1 Density Functional Theory ........................................ 9
    2.1.1 Hohenberg-Kohn theorem .................................... 9
    2.1.2 Kohn-Sham scheme and Local Density Functional Approximation (LDA) ............................................. 11
  2.2 Pseudopotential ................................................... 16
  2.3 Many-body Perturbation Theory ................................... 21
    2.3.1 Hedin Equations ............................................ 23
    2.3.2 GW Approximation ........................................... 27
    2.3.3 Excitonic Effect (Bethe-Salpeter Equation) ............... 32
III LATTICE VIBRATIONS IN SILICON NANOWIRES ......................... 41
  3.1 Lattice Vibrational Modes in the Harmonic Crystal ............. 42
  3.2 Linear Response Theory ......................................... 46
  3.3 Structure of SiNWs ............................................... 49
  3.4 The Density of Vibrational Modes ................................ 52
  3.5 The Frequency Shift of Optical Modes and its Size Dependence ... 54
  3.6 Confined Modes and Their Size Dependence ........................ 57
  3.7 Raman Activities of SiNWs ....................................... 61
  3.8 Summary of this Chapter ......................................... 70
IV EXCITONIC EFFECT IN THE OPTICAL ABSORPTION OF SILICON NANOWIRES .............. 72
  4.1 The Structure of SiNWs and Calculational Parameters .......... 73
LIST OF TABLES

3.1 Three SiNWs of different sizes. Listed are the number of hydrogen atoms and silicon atoms per unit cell. The right column shows diameters of these three wires. ........................................................... 49

3.2 The fitting results of the frequency red shift in silicon NWs. The shift is fitted with the formula as \( \delta \omega = \frac{b}{r^n} \). ........................................................... 57

5.1 The core-shell structure NWs we calculated. The NWs are marked with an index to help the discussion. The core size is normalized to the NW diameter. Each of these five type has two kinds of components, silicon core and germanium core. ................................................... 101

5.2 The relaxed bond length of core-shell NWs calculated from three different starting bond-length values, showing a converged final structure. ................. 102

B.1 The core-shell NW with a silicon core and germanium shell, which is marked as type I structure in Table 5.1. The critical charge of the confined valence band is 0.92 inside the shell, and the critical charge of the confined conduction band is 0.26 inside the core. As a result, the band offset of valence bands is 0.19 eV, and that of conduction bands is 0.40 eV. ..................... 121

B.2 The core-shell NW with a silicon core and germanium shell, which is marked as type II structure in Table 5.1. The critical charge of the confined valence band is 0.89 inside the shell, and the critical charge of the confined conduction band is 0.33 inside the core. As a result, the band offset of valence bands is 0.21 eV, and that of conduction bands is 0.40 eV. ..................... 121

B.3 The core-shell NW with a silicon core and germanium shell, which is marked as type III structure in Table 5.1. The critical charge of the confined valence band is 0.82 inside the shell, and the critical charge of the confined conduction band is 0.47 inside the core. As a result, the band offset of valence bands is 0.19 eV, and that of conduction bands is 0.29 eV. ..................... 122

B.4 The core-shell NW with a silicon core and germanium shell, which is marked as type IV structure in Table 5.1. The critical charge of the confined valence band is 0.72 inside the shell, and the critical charge of the confined conduction band is 0.61 inside the core. As a result, the band offset of valence bands is 0.18 eV, and that of conduction bands is 0.22 eV. ..................... 123

B.5 The core-shell NW with a silicon core and germanium shell, which is marked as type V structure in Table 5.1. The critical charge of the confined valence band is 0.57 inside the shell, and the critical charge of the confined conduction band is 0.75 inside the core. As a result, the band offset of valence bands is 0.26 eV, and that of conduction bands is 0.16 eV. ..................... 124
B.6  The core-shell NW with a germanium core and silicon shell, which is marked as the type I structure in Table 5.1. The critical charge of the confined valence band is 0.26 inside the core, and the critical charge of the confined conduction band is 0.92 inside the shell. As a result, the band offset of valence bands is 0.52 eV, and that of conduction bands is 0.03 eV.

B.7  The core-shell NW with a germanium core and silicon shell, which is marked as the type II structure in Table 5.1. The critical charge of the confined valence band is 0.33 inside the core, and the critical charge of the confined conduction band is 0.89 inside the shell. As a result, the band offset of valence bands is 0.60 eV, and that of conduction bands is 0.02 eV.

B.8  The core-shell NW with a germanium core and silicon shell, which is marked as the type III structure in Table 5.1. The critical charge of the confined valence band is 0.47 inside the core, and the critical charge of the confined conduction band is 0.82 inside the shell. As a result, the band offset of valence bands is 0.63 eV, and that of conduction bands is 0.03 eV.

B.9  The core-shell NW with a germanium core and silicon shell, which is marked as the type IV structure in Table 5.1. The critical charge of the confined valence band is 0.61 inside the core, and the critical charge of the confined conduction band is 0.72 inside the shell. As a result, the band offset of valence bands is 0.60 eV, and that of conduction bands is 0.03 eV.

B.10 The core-shell NW with a germanium core and silicon shell, which is marked as the type V structure in Table 5.1. The critical charge of the confined valence band is 0.75 inside the core, and the critical charge of the confined conduction band is 0.57 inside the shell. As a result, the band offset of valence bands is 0.46 eV, and that of conduction bands is 0.02 eV.
## LIST OF FIGURES

2.1 Schematic representation of the self-consistent loop for solution of Kohn-Sham equation. .................................................. 12

2.2 Correlation energy of an unpolarized homogeneous electron gas as a function of the density parameter $r_s$ [32]. .................................................. 14

2.3 Comparison of pseudopotentials for carbon (dotted line for s and solid line for p) in real space and reciprocal space. In order from left to right generated using the procedures of: Troullier and Martins; kerker; Hamann, Schluter, and Chiang; Vanderbilt. .................................................. 20

2.4 The Feynman diagram of Eq. (2.101). It has the same formulation as Bethe-Salpeter equation, which is from relativity theory. As a result, we find the two-particle Green’s function obeys the Bethe-Salpeter equation. Here the interaction kernel is the interaction between the electron and hole, which is also called the exciton. .................................................. 34

2.5 The flow chart of the Green’s function method in calculating the electronic structure and optical absorption spectrum. The LDA results are used as the starting point. In real calculations of Bethe-Salpeter equation, we have to use a more denser k-grid than in GW calculations, so the calculations of the dielectric function may not overlap as in this figure. .................................................. 39

3.1 Ball-and-stick model of the hydrogen-passivated silicon NW under investigation: side view (left) and top view (right). The large atoms are silicon and small atoms are hydrogen. There are 16 silicon and 12 hydrogen atoms in each unit cell. Five unit cells are shown in the left figure. ...................... 49

3.2 Configurations of hydrogen atoms on the [100] surface of silicon crystal from reference [73]. We list three kinds of configurations. The left three figures are the top views, and the right three figures are the side views. In the SiNWs we study, the second type of configuration has the lowest energy and is used in our calculations. .......................... 50

3.3 Convergence of the total energy with respect to the cutoff energy and k-grid. The cutoff energy is set to be 18 Ry as marked in the left figure. The k-grid is set to be 1x1x8 as marked in the right figure. ...................... 52

3.4 Density of state of phonons at the Γ point for a silicon [110] wire with diameter $d = 1.2, 1.8$, and $2.2$ nm, respectively. The Gaussian broadening with a width of 0.05 THz is used in this figure. The corresponding quantity for the silicon bulk is shown at the top of the figure. The arrows mark positions of confined modes below 8 THz. “L” marks the position of LO mode, and “T” marks the position of TO modes. In the silicon bulk, these modes become degenerated. “A, B, C, and D” are the confined modes shown in Fig. 3.7. .......................... 53
3.5 Three optical modes in SiNWs are represented in three columns. Each column has three pictures. From the top to the bottom, the optical modes are tracked as the diameter of the NW increases. 

3.6 Red shift of the three optical modes in silicon NWs. The shifted frequency refers to the bulk optical mode frequency. $T_1$ and $T_2$ modes are the type I mode in Fig. 3.5, and L mode is the type II mode in Fig. 3.5. We use a $\frac{1}{\beta r}$ function to fit the points. The result shows that $\alpha$ is always around the 1.3-1.5 range. (The detail value is listed in Table 3.2) 

3.7 Confined modes in SiNWs. Figures A and B are radial breathing modes, with a different quantum number. Figure C is a confined mode that has only $\theta$ component. Figure D is another kind confined mode that has a $\frac{\pi}{2}$ phase difference between $\theta$ and $r$ directions. 

3.8 First-order Bessel function. The arrows mark the points with the zero derivative. In our discussion, the marked positions represent the boundary. By choosing a different boundary condition, we find different kinds of nodal structure in the radial breathing mode (RBM). 

3.9 Blue shift of the radial breathing modes (RBM) as the diameter of the wire decreases. We list the frequency of the RBM in three NWs, and use a $\frac{1}{a}$ function to fit the points. The result shows that $a$ is around 1.22. 

3.10 Raman scattering. A photon with energy $\omega_L$ and momentum $q_L$ is scattered by phonon with energy $\omega_k$ and momentum $k$ to the photon with $\omega_s$ and momentum $q_s$. Figure (a) shows phonon absorption, and the photon energy is shifted to a higher frequency, which is called the anti-Stokes process. Figure (b) shows phonon emission, and the photon energy is shifted to a lower frequency, which is called the Stokes process. 

3.11 Convergence of the dielectric tensor with respect to the cutoff energy and the k-grid. Here we plot the zz component of the dielectric tensor as an example. It is clear that the dielectric tensor is sensitive to the k-grid, which is shown in the left figure. We choose to use 1x1x8 k-grid. Because the cutoff energy is not an important factor in the calculation of the dielectric tensor, 18 Ry is enough for this Raman scattering calculation. 

3.12 Raman activity of the smallest SiNW (no space average). We fix the incoming polarization of the light in the first three graphs, and set the outgoing light direction in x, y, and z, respectively. The index of the direction is specified; (001) means the z direction, and (100) means the x direction. For example, the last figure shows the result for incoming light polarization in the x direction, and outgoing polarization in the y direction. In addition, the polarization along the z direction is much stronger than other configurations, which is preferred in Raman scattering measurement. 

3.13 Raman activity of the smallest SiNW averaged over the orientation but without including the depolarization effect. The breathing mode is highly active in the smallest wire, which can be used to estimate the wire size.
3.14 A typical Raman spectrum from experiment (reference [83]). The angle is between the incoming light polarization direction and the wire growing direction. It is clear that the Raman scattering peaks disappear when the above two directions are perpendicular to each other. This is called the depolarization effect.

3.15 Raman activities of the smallest SiNW with space average and depolarization effect. Compared with the previous space averaged Raman activity, the peak related to the RBM is more clear which helps people in experiments estimate the size of NWs easily from the Raman scattering spectrum.

4.1 The ball-stick model of the SiNW. The brown atoms are silicon, and the grey atoms are hydrogen. We define the x, y, and z axis in the figures. The origin of the coordinate is in the inversion center of the wire. The left panel shows the side view, and the right panel is the top view.

4.2 The Coulomb truncation between SiNWs. For simplicity, we use an orthogonal lattice as an example. The truncation radius $r_c$ is half of the lattice constant $d$ of the supercell.

4.3 The inverse dielectric function $\varepsilon^{-1}(q_z)$. This function sharply decreases when $q_z$ is small, which requires denser k-grid to be described.

4.4 The quasiparticle energy compared with the Kohn-Sham energy. The GW correction is fairly linear with the LDA results. We fit the correction to valence bands and conduction bands separately and extend the shift to the whole band structure.

4.5 One-particle absorption spectra of the SiNW with different number of bands included. The curves are smearing by a Gaussian of an energy width of 0.1 eV. (a) 12 valence bands and 12 conduction bands; (b) 15 valence bands and 15 conduction bands; (c) 20 valence bands and 20 conduction bands. The main structure of the above three curves is similar; only small changes in the relative amplitude of the peaks are found. The position of the band gap is marked with an arrow in the top panel. In addition, we mark the three main absorption peaks for the convenience of the discussion with a lower triangle, circle and upper triangle.

4.6 Band structure of the SiNW near the band gap. The solid dots represent the bands that are included in exciton calculations. The main contributions to the optical spectra are marked with I, II, and III. Transition III is not around the zone center; the conduction bands and valence bands are parallel to each other resulting in a large joint density of states (JDOS).

4.7 The imaginary part of the dielectric function for bulk silicon. The dash-dot line is the optical absorption without the electron-hole interaction. The solid line is the absorption including electron-hole interaction. Two absorption peaks $E_1$ and $E_2$ are marked. The excitonic effect enhances the $E_1$ peak and suppresses the $E_2$ peak. The joint densities of states are plotted at the lower panel with the same legend.
4.8 The optical absorption spectrum of an isolated silicon nanowire with a 1.2 nm diameter for polarization along the axis. The smearing width is 0.1 eV. The dash-dot line is the one-particle spectrum calculated with 12 valence bands and 12 conduction bands. The solid line is the optical spectrum including the electron-hole interaction. The strong e-h interaction contributes to the huge shift of the optical spectra. We mark the three main peaks of the excitonic spectrum with an upper triangle, a circle and a lower triangle, respectively. In addition, the peak marked by a circle is enhanced by the excitonic effect.

4.9 The density of states of excitons (DOSE) is shown by the solid line, and the joint density of states of band structure (JDOS) by the dot-dash line. The shift from excitonic effect is about 0.3 eV.

4.10 The density of the electron for several excitonic states with the hole fixed at the origin. From top to bottom, five representative states are shown with the energy values of 3.3 eV, 4.2 eV, 4.2 eV, 4.7 eV, and 6.7 eV, respectively. Corresponding cross section plots on the \(xz\) plane are shown on the right, while the three-dimensional plots for the same data are shown on the left. Because the bottom two states are delocalized, we have to use a large scale to plot them.

4.11 Charge density plots in the x-y plane of the nanowire. The hole is fixed in the origin (\(r_h = 0\)), and the density of the electron is plotted. The left figures are the 3-D plots, and the right figures are the corresponding contour plots. The energies of these two excitons are 3.2 eV and 3.5 eV, respectively.

4.12 The depolarization effect of an isolated SiNW. When the polarization direction is perpendicular to the nanowire, the optical absorption is extremely small as the solid line shows. This effect cannot be included in one-particle calculations [99].

4.13 The evolution of the bound excitons at the band edge. In figure (a), the transition \(A\) is marked in the band structure. In figure (b), the dash-dotted line is the one-particle result, and the solid line is the BSE result. Four excitonic states originated from transition \(A\) are identified in the BSE calculation: \(A_1, A_2, A_3,\) and \(A_4\). Figure (c) shows the binding energy of the four excitons, \(A_1, A_2, A_3,\) and \(A_4\). The dotted line is a fitting result. The binding energy as a function of quantum number \(n\) is \(E_0 n^{-1}\), which is close to the case of a hydrogen atom.

4.14 Electron charge distributions in the yz plane for exciton states \(A_1, A_2, A_3\) and \(A_4\), with the hole fixed at the origin. As the quantum number increases, the binding becomes weaker and weaker. With nodal structure develops, the exciton wave function extends more in real space. Note that the scale changes in the last two figures, \(A_3\) and \(A_4\).

5.1 The basic structure of a MOSFET, one of the most popular FETs. The voltage of the gate controls the width of the channel and tunes the transconductance between the source and the drain. The substrate can have a different doping type from the gates. The mobility of free carriers inside the conduction channel is important to determine the speed of this device.
5.2 The NW-based field effect transistor [115]. The drain and source are connected by the core-shell NW. The gate structure is under the NW, so this FET is called a back-gate FET. The advantage of the back-gate FET is that it is easily fabricated because of its simple structure and has a large linear working region. ................................................................. 96

5.3 The doping process in core-shell NWs. Assuming that core-shell NWs have a similar band lineup as Si/Ge superlattices, we expect that the free carriers will move to the lower-energy state and become separate from doped impurities. 97

5.4 Calculated band offsets of Si/Ge heterojunction [111]. The left panel is the case with germanium strained, and the right panel is the case with silicon strained. We only consider the solid line result, and the dash line is some weighted average offset that is beyond our consideration. In addition, the relative values of this figure are meaningful because of the choice of reference energy. ................................................................. 99

5.5 Calculated band offsets of Si/Ge superlattices [113]. The left panel is the case with germanium strained, and the right panel is the case with silicon strained. Only the offset values of valence bands are marked in this figure, and they are similar to those of Si/Ge bulk heterojunction. Because this result is about the long-period superlattices where the confinement is weak, it is obvious the correction induced by the confinement is small. ............. 100

5.6 Ball-and-stick model of the hydrogen-passivated Si/Ge core shell NWs along [110] direction. The diameter of the wire is about 4.1 nm. Although we calculate five core-shell structures, we only plot two of them (I and IV from Table 5.1). ................................................................. 101

5.7 The energy convergence with respect to the cutoff energy and k-grid. The cutoff energy is set to be 170 eV as marked in the left figure; the k-grid is set to be 1x1x4 shown in the right figure. Although the total energy convergence with the cutoff energy and k-grid is not perfect, we find they are enough for the accurate band structure calculations. ................................................................. 102

5.8 The band gap variation with different core ratios for NWs with a fixed diameter. The upper figure is the silicon-core case, and the lower figure is the germanium-core case. ................................................................. 103

5.9 An example of the band lineup of core-shell NWs. The left panel is the band lineup. When we consider the confinement, we focus on the states amplified in the right panel. The dash lines mark the states confined within the core and with a larger energy level spacing. The dot lines mark the states smearing over the whole wire and experience weaker confinement, which results in a smaller energy level spacing. ................................................................. 104

5.10 The band structure close to the fundamental band gap. The upper panel is a NW with a silicon core, and the lower panel is a NW with a germanium core. The discretization of the bands close to the band gap implies that there are band offsets between core and shell states. ................................................................. 105
5.11 The charge density of electronic states close to the fundamental band gap, and the NW is with a silicon core and a germanium shell along the \([111]\) direction. The left panel is the charge density of the top of valence bands, and the right panel is that of the bottom of conduction bands. We can see that the charge density in the right panel is not confined within the core or the shell. As a result, it is hard to see the offset in this structure. 

5.12 The charge distribution of the core-shell structure III in Table 5.1 (silicon core). We plot the states close to the band gap. The three figures in the upper panels are from conduction bands, and the three in the lower panels are from valence bands. The number above each plot is the index of the band. The beginning of valence bands is confined within the shell part, while the conduction bands are confined within the core part. When we go further away from the band gap, the non-confined bands will show up, such as VB547 and CB553, which give us the critical transition point from the confined bands to the non-confined bands.

5.13 The charge distribution of the core-shell structure I in Table 5.1 (germanium core). We plot the states close to the band gap at \(\Gamma\) point. The arrangement of figures and the meaning of numbers above each plot are the same as in Fig. 5.12. The beginning of valence bands is confined within the core part, while the conduction bands are confined within the shell part. 

5.14 The band offsets of core-shell NWs. The upper part shows the offsets in silicon-core NWs, while the lower parts show germanium-core NWs. 

5.15 Band gap evolution of SiNWs. We mark the LDA energy gap with arrows and the corresponding GW result with dotted lines. 

5.16 The band lineup of the core-shell structure. Figure (a) presents the silicon core case, which suggests doping electrons at the shell part. Figure (b) presents the germanium core case, which suggests doping holes at the shell part. 

5.17 An example of self-doping process. The Fermi surface is below the top of valence bands, which induce free holes without doping impurities. It is found that there is the self-doping process in Si/Ge core-shell NWs. 

A.1 The band structure evolution from the pure germanium wire to the pure silicon wire with the silicon core size increasing. The indices in the plots are same as in Table 5.1. 

A.2 The band structure evolution from the pure silicon wire to the pure germanium wire with the germanium core size increasing. The indices in the plots are same as in Table 5.1.
SUMMARY

The semiconductor nanowire (NW) has attracted significant attention as a new one-dimensional structure for fundamental studies and also as a potential building block for nanodevices. When the size of NWs becomes close to the order of the de Broglie wavelength of electrons, the quantum confinement effect significantly affects the electronic behaviors of NWs and raises expectations for many novel applications in microelectronics. As the technology advances, more and more fine semiconductor NWs are fabricated in the laboratories, which provide good platforms to study the special properties of low-dimensional systems. From these new nanostructures, we choose to study the single crystal silicon nanowire (SiNW). One advantage of this choice of SiNWs is that we have the solid and complete knowledge about the electronic structure of bulk silicon, which provides valuable information for us to understand the quantum confinement effect in SiNWs. The other reason for this choice is that the silicon-based technology is so highly developed that scientists prefer silicon as the elementary material for the cost consideration, even it may not be the highest performance candidate. Therefore, our calculations on SiNWs should be useful for the industrial development of today’s nanotechnology.

As a starting point, we begin by studying the ground electronic state of SiNWs and performing calculations on the following three physical properties. Based on thorough knowledge about the ground state of electrons, the first part of our work is to calculate the properties of lattice vibrations of SiNWs. The density functional theory (DFT) based linear-response method is used to obtain lattice vibrational modes. We plot the density of vibrational modes at the Γ point of the Brillouin zone and show a clear evolutionary from bulk silicon to the narrow SiNW. Two kinds of frequency shifts of lattice vibrational modes are found: One is the red shift of the optical modes, the other is the blue shift of the radial breathing modes (RBMs). We discover that the size dependence of the frequency
shifts of RBMs can be described well by the elastic model with the cylindrical boundary confinement. In order to characterize these confined modes in SiNWs, we calculate the first-order Raman activities of the smallest SiNW, and find that the RBM is strongly active in the scattering spectrum. Therefore, our calculated result about the size dependence of the frequency of RBM provides an easy way to estimate the size of the nanowires from the corresponding Raman spectrum.

The excited-state properties of nanostructures are of critical importance in the design of functional optical devices. The low dimensionality and reduced size tend to strengthen the effective Coulomb interaction in nanostructures. Quantitative evaluations of physical properties manifesting this effect are therefore timely and valuable to the nanotechnology research. In the second part of my work, we focus on the correlated electron-hole states in semiconductor NWs and their influence on the optical absorption spectrum. First-principles calculations are performed for an isolated hydrogen-passivated SiNW with a diameter of 1.2 nm. By using plane waves and pseudopotentials, the quasiparticle states are calculated within the many-body perturbation theory with the so-called GW approximation, and the electron-hole interaction is evaluated with the Bethe-Salpeter equation (BSE). The enhanced Coulomb interaction in this confined system results in an unusually large shift (1-2 eV) of the optical spectrum as well as a significant increase in certain absorption peak intensity. The current results predict anomalous excited-state properties in semiconductor NWs that may impact future applications of these nanostructures.

In the third part of my work, the electronic band structures of Si/Ge core-shell NWs are studied with first-principles calculations. The electronic states close to the band gap of NWs along the [110] direction are found to be confined within the core and the shell, respectively. Our calculated results show that the band offset between the core and shell electronic states are not only size dependent but also core-shell ratio dependent, which gives important corrections to those results from bulk heterojunctions and superlattices. The existence of band offsets makes it possible to dope impurities in the shell part of NWs while inject the carrier to the core part, or vise versa. This novel doping mechanism avoids the scattering process induced by doped impurities and gives hope to the high-speed NW
devices. Finally, based on our calculations, the optimal doping strategy is proposed.
CHAPTER I

INTRODUCTION

Most of the interesting properties of the condensed matter are related to electrons, such as the band structure, optical absorption spectra, even the lattice vibrations, etc. However, it is a real challenge to deal with the electronic structure in solids, because the number of electrons in materials is huge (the order of $10^{23}$), and they are correlated with each other and interact with ions. Since one cannot exactly solve the equations of electrons, many approximations have been developed to make this studies feasible. For example, electrons in metals are approximated by the free electron gas, which is surprisingly successful in describing the electronic behavior in Alkali metals. Another example is the Fermi liquid theory, which describes well the behavior of quasiparticles close to the Fermi surface. However, in semiconductors and insulators, the free electron gas model is not enough. A more accurate method including the interaction between the electron with the ions and other electrons is expected. The state-of-the-art method to study the electron structure of solids is the density functional theory (DFT).

The origin of DFT can be traced back to the late of 1920s [1]. Thomas and Fermi proposed a functional theory to describe the inhomogeneous electron gas. By neglecting the exchange and correlation between electrons, they established a total energy functional. They used the density of electrons as the variable instead of the coordinates of electrons, which extremely simplified the equation, reducing it from 3N dimensions to one dimension. Of course, their final result was not accurate enough because of the absence of the interaction between electrons, but this density-based functional suggested a new direction to solve many-body problem. In 1964, Hohenberg and Kohn published a paper to prove that the desired density functional existed [2]. The key point is that there is a universal density functional in the expression of the total energy, which is sufficient to determine the ground state of electrons in the presence of a given external potential. Later, this idea was expanded
to the spin density functional and finite temperature cases [3, 4].

The Hohenberg-Kohn theorem only proved that there existed a density functional, but did not give the details of the form of that functional. In 1965, Kohn and Sham proposed a self-consistent procedure to apply DFT to atomic and solid-state calculations [5]. They included all the many-body effect into the exchange-correlation energy functional and minimized the energy functional to derive the Kohn-Sham variational equations that were single-particle equations. The many-body effect created a new local exchange-correlation potential term $V_{xc}$ in addition to the Hartree term.

The main problem is to find a working form of the exchange-correlation term. There are two popular approximations, the local density approximation (LDA) and the generalized-gradient approximation (GGA). The exchange-correlation energy is an integral over all space of the exchange-correlation energy density. The LDA assumes that this exchange-correlation energy density at each point is the same as that of a homogeneous electron gas with the same density. The GGA form proposes a more accurate way to include the gradient term of the local density. A few forms of the expansion have been proposed, such as Becke (B88) [6], Perdew and Wang (PW91) [7], and Perdew, Burke and Enzerhof (PBE) [8].

The Kohn-Sham equation mentioned above needs to be solved in a self-consistent way. Under a specific approximation (LDA or GGA), the form of the equation is fixed. One can first guess an electronic charge density of the system and get the Kohn-Sham potential. With that potential, the Kohn-Sham equation can be solved. The resulting single-particle wavefunctions can be used to construct a new charge density, which determine the new Kohn-Sham potential, and so on. Finally, the converged result can be obtained by several iterations. This self-consistent procedure is highly efficient to be implemented into computer codes. As a result, it is nowadays a standard practice to perform the DFT calculations in first-principles calculations.

Despite its simplicity, the LDA and GGA successfully describe the ground-state properties of atoms, molecules, and solids. Structural properties of solids such as the lattice constant, the bulk modulus, and the cohesive energy are generally determined to within a
few percent of the experimental values [9, 10].

The availability of accurate DFT-based techniques to study the electronic structure of crystals has renewed the interest in the study of their response to external perturbations. One of the successful applications is studying the phonon in solids, which comprises of perturbations on ions from equilibrium positions. We will give the details about the first-principles phonon calculations in Chapter III, so only a simple introduction is presented here.

There are mainly two kinds of phonon calculation methods based on DFT. The first one is referred as the "direct" method [11, 12]. Through this method, the perturbed system is handled on the same footing as the unperturbed one, and the response is obtained by comparing the properties of these two systems, such as the total energy difference or the force difference acted on each ion. This "direct" method is rather straightforward computationally, and it can be implemented into the total energy code easily. In addition, this method allows us to study both linear and nonlinear effects directly. However, when we deal with the systems containing a large number of atoms per unit cell, the "direct" method is expensive in numerical simulations.

The other DFT-based phonon calculation method is the linear response method [13]. The idea of this method is simple. The response to the external perturbation can be calculated by quantities form the unperturbed system. In history, the first linear response method for phonon calculations is to find the inverse of the dielectric matrix (DM) by using the eigenfunctions and energy levels of the unperturbed system. As in its name, the linear response method only considers the linear effect of the perturbation, and it needs extra steps to study the nonlinear effect. But the merits of the this method is impressive too. It does not need a supercell for calculating phonons at any point of the Brillouin zone, which significantly increases the efficiency of the computation.

As the advance of the linear response method, several new developments have been proposed [14, 16, 17, 15]. Because the bottle neck for linear-response calculations is the inclusion of a large number of electronic states to get the DM instead of only the valence states, a Green’s function method has been proposed to avoid the conduction states. By
introducing this procedure, the efficiency of the linear-response method is greatly improved. Since the 1980’s, first-principles linear-response methods have been developed to successfully study phonons in many materials. They obtain satisfying results about lattice vibrations, and have been extended to consider the nonlinear effect. In the phonon part of my PhD work, we mainly use the linear response method to discuss the lattice vibrational modes in silicon nanowires (NWs). It has to be point out that although the phonon is related to the perturbation of ions from equilibrium, under the adiabatic approximation we still consider the ground state of electrons as the ions are displaced.

It has been known up to date that the density functional theory can describe well the ground state of electrons. However, people often expect to study properties of materials related to the excited electronic states. Unfortunately, DFT is not designed to describe excited states. As a result, several perturbation theories have been developed to solve this problem. In the following part of this introduction, the Green’s function method for the calculation of excited states is presented, which is widely used to correct the quasiparticle energy spectrum and optical absorption, and has obtained satisfying results for the past twenty years [18].

In the derivation of the Kohn-Sham theory, the eigenvalues enter as Lagrange parameters that ensure the orthogonality of the Kohn-Sham orbitals of fictitious particles. As a consequence, the eigenvalues and the Kohn-Sham orbitals must be considered as mathematical tools that contain no relevant physical information besides the fact that the square of the eigenfunctions sums up to the exact local ground state density. In particular, there is no formal justification that links the eigenvalues to the energy dispersion of electrons in a solid. On the other hand, if one compares the Kohn-Sham eigenvalues with the energy spectrum of electrons from experiments, there is a systematic derivation between them. The energy gap of Kohn-Sham eigenvalues seems to consistently underestimate the band gap of solids. This is the well-known band gap problem of LDA and GGA.

In practical, it is found that the LDA (or GGA) energy gap can be corrected by including the electron self-energy correction to the exchange-correlation potential. There are many methods developed to correct the band gap. One of the successful candidates is the so-called
GW approximation [19], which is proposed by Hedin and Lundquist with the one-particle Green’s function under the random phase approximation (RPA). This GW approximation supplies an \textit{ab initio} correction to the Kohn-Sham energy gap, and the results agree with experimental data well.

The GW approximation is a model for the determination of excited states based on the quasiparticle concept and the Green’s function method. Below is a picture about the quasiparticle in a many-body system. The Coulomb repulsion between electrons leads to a depletion of the negative charge around a given electron, and the ensemble of the electron and its surrounding positive screening charge form a quasiparticle. The mathematical description of quasiparticles is based on the single-particle Green’s function, whose exact determination requires the complete knowledge of the quasiparticle self-energy. In fact, the self-energy can be expanded linearly in terms of the dressed Green’s function ($G$) and the screened interaction, leading to the name of this approximation (GW).

In practice, one gets the quasiparticle energy by solving the quasiparticle equation, which includes the non-Hermisian self-energy operator. It looks similar to the single-particle equation in the Kohn-Sham theory, but the solutions of the quasiparticle equation, the eigenvalues and the eigenfunctions, are physically meaningful, in contrast to the mere mathematical tools as in the case of DFT. Although the Hamiltonian of the quasiparticle is not Hermisian, the real part of the energy is dominant, and the quasiparticle picture is generally good enough.

A lot of studies have been performed with the GW method. An early review of bulk and surface calculations using the GW method was given by Hybertsen and Louie [20]. Bechstedt discussed the physics of the GW method in relation to model approaches for the calculation of the dielectric response as well as the self-energy [21]. Godby reviewed quasiparticle calculations for jellium, simple metals, and semiconductors [22]. Mahan examined different GW approximations resulting from the inclusion of the self-energy and vertex diagrams beyond the random phase approximation [23]. Pollmann \textit{et al.} concentrated on GW calculations for semiconductor surfaces [24], and Farid considered mathematical aspects of quasiparticle calculations [25].
Going beyond the single-particle picture, when an electron is excited from the fully occupied states, this will create a vacant state below the Fermi level, which is called the hole. The electron-hole pair is known as the exciton, and is described by the Bethe-Salpeter equation (BSE). Although exciton calculations for real systems within a fully first-principles scheme have recently been performed, the importance of the excitonic effects has been known for a long time. For example, Hanke and Sham solved the BSE for the electron-hole response function of bulk silicon, using a linear combination of atomic orbital basis, with a semi-empirical band structure fitted to optical experiments [26].

The first-principles approaches used today to solve the BSE mostly follow the scheme introduced by Onida for the spectrum of the cluster $Na_4$ and that $Li_2O$ [27]. This procedure consists of (i) a ground-state DFT calculation; (ii) a GW calculation to correct the eigenvalues; (iii) the solution for the BSE using GW eigenvalues, Kohn-Sham orbitals, and the static RPA screening for the electron-hole interaction. This scheme has been applied to the calculation of the absorption spectra of many materials, and the results agree with the experimental data well.

From the brief review above, there is no doubt that first-principles calculation have performed well in the past calculations, and become a standard procedure to study the electronic properties of materials.

On the other hand, in the past ten years, as the size of devices becomes smaller and smaller, the quantum effect is a more and more important factor in the consideration of material properties. The size of future devices is in the nanometer scale, and a new branch of science, namely nanotechnology, has been formed. Not only the engineers are interested in nanodevices, the scientists are also interested in low-dimensional systems. Therefore, extensive efforts have been devoted to the field of nanotechnology, with the hope to discover more exciting properties and to identify the potential for next-generation devices. As a result, numerous kinds of nanomaterials have been fabricated in experiments, such as nanotubes, nanowires, nanobelts, nanoparticles, etc. So many new phenomena have been observed that these new but complicated nanomaterials present exciting properties for physicists to study.
In conjunction with the experimental investigations, there have been numerous theoretical studies about nanostructures. As we point out above, the electronic structure is significantly changed by quantum confinement, which makes it hard to apply empirical models from bulk materials to the study of nanostructures. For example, when we talk about the thermal conductivity in nanowires (NWs), one can apply the empirical model to this problem. But we have to know the dispersion of phonons and the sound velocity in NWs, which are not given by the empirical model. To these cases, first-principles calculations are valuable to study nanostructures because they do not need parameters in advance. In addition, the good transferability and accuracy of first-principles calculations make the calculated results reliable.

There have been many kinds of nanostructures manufactured in the laboratories. Since we are more interested in the one-dimensional system, the NW attracts our attention. We focus on two properties of single crystal silicon NWs (SiNW) [28, 29, 30, 31], one is the lattice vibrational modes (phonons); the other is the electronic structure and the optical properties of SiNWs.

The phonon property is tightly related to the structure of SiNW, and it is the foundation for further studies of the thermal conductivity, the electron-phonon interaction, etc. Despite the importance of the lattice vibrational modes in SiNWs, there are none first-principles calculations on this subject. As a result, when the experimentalists deal with the thermal and transport problems, they usually use the parameters such as sound velocity, Grunesen number, etc., based on bulk values, which is obviously not accurate. Therefore, it is urgent to find these fundamental phonon properties from the first-principles calculations. In the first part of my thesis, I will begin with the concept of phonons and the linear response method to address this question. A thoroughly study about the lattice vibrational modes is performed, and the quantum confinement effect is discovered in both optical modes and confined modes. In order to guide future experiment, the Raman spectrum of a SiNW is calculated by the linear response method.

The second topic of my PhD work is the excited states in SiNW. It is well known that the excited states determine many properties of the systems, such as the band gap, optical
absorption, and photo illuminance, etc. We use the Green’s function method to describe
the energy spectra of excited states. Because of the special low-dimensional geometry of
SiNWs, the screening effect is very different from the bulk materials, which gives rise to
many new phenomena, for example a very large GW correction to the Kohn-Sham band
gap. Furthermore, the strong interaction between electrons and holes is calculated. This
enhanced excitonic effect significantly affects the optical absorption spectrum of SiNWs. In
this calculation, we have investigated all the absorption peaks in SiNWs and identified their
connections with those in the bulk, which helps us to predict the excitonic effect in other
semiconductor NWs.

In the last part of my PhD works, we study the electronic structure of Si/Ge core-shell
structures. The lack of the band lineup information for heterogenous NWs, hampers the
further development of the NW-based devices. In this study, we use the LDA method to
figure out the band offset in radial Si/Ge core-shell structures. With the detail study of the
electronic charge distribution and the band offset, a novel doping mechanism is proposed.
One can introduce the donors in the shell region and obtain the carriers in core region, or
vise versa. This method can avoid the scattering problem induced by the doped impurities
and improve the mobility of carriers. We find the optimal strategy for doping impurities,
depending on the specific band offset of the core-shell structures, as demonstrated in the
recent experiment.
CHAPTER II

THEORETICAL METHOD

In this chapter, we will give a brief review of the theoretical efforts about first-principles calculations, which are the foundation of my PhD work. This Chapter is organized as the following. The first section is an introduction to density functional theory (DFT). The Hohenberg-Kohn theorem is proved, and the Kohn-Sham equation is derived under the local density functional approximation (LDA). The second section is a conceptual review of pseudopotentials. Because we mainly use norm-conserving pseudopotentials, we discuss a widely used type, the Troullier-Martins pseudopotentials. The last section is the many-body perturbation theory. In order to give the correct energy of excited states, we have to use the Green’s function method to calculated the self-energy of of the electron in a many-body system. By using the one-particle Green’s function, we expand the self-energy according to the dynamical screened interaction, which results in the GW approximation. In addition to the first order correction by the GW approximation, we continue to consider the electron-hole interaction, which is determined by the Bethe-Salpeter equation (BSE) with the help of the two-particle Green’s function.

2.1 Density Functional Theory

The density functional theory is a theory of correlated many-body systems. The key point of this theory is that any ground-state property of a system of many interacting particles can be viewed as a functional of the ground state density \( n_0(\vec{r}) \). That is, one scalar function of position \( n_0(\vec{r}) \), in principle, determines all the information in the many-body wave functions for the ground state.

2.1.1 Hohenberg-Kohn theorem

As we discussed in the Introduction Chapter, the origin of the density functional theory can be tracked back to the work done by Thomas and Fermi in 1920’s [1]. In 1964, Hohenberg
and Kohn proved that there existed a universal functional of the density, \( F[n_0(\vec{r})] \), independent of the external potential, such that the expression \( E = \int v(\vec{r})n_0(\vec{r})d\vec{r} + F[n_0(\vec{r})] \) has its minimal value of the correct ground state energy associated with external potential \( v(\vec{r}) \) [2]. We will prove this theorem in two steps.

(i) The first step is to prove that the potential \( v(\vec{r}) \) is a unique functional of the density \( n_0(\vec{r}) \). The Hamiltonian is

\[
H = T + U + V, \tag{2.1}
\]

where \( T \) is the kinetic energy, \( U \) is the Coulomb interaction term between electrons, and \( V \) is the external potential, such as the static electric field. With the specification of the external field \( V \), the Hamiltonian is fixed. Suppose the ground state is \( |\Psi\rangle \), the corresponding ground state charge density is \( n_0(\vec{r}) \). With a different external potential \( v'(\vec{r}) \), the corresponding Hamiltonian is \( H' \) and results in a different ground state \( |\Psi'\rangle \). Assume that \( |\Psi'\rangle \) gives the same density \( n_0(\vec{r}) \), we prove that there will be an inconsistency.

Because \( |\Psi'\rangle \) is the ground state, we have the minimum energy formula as

\[
E' = \langle \Psi' | H' | \Psi' \rangle < \langle \Psi | H' | \Psi \rangle = \langle \Psi | (H + V' - V) | \Psi \rangle, \tag{2.2}
\]

so that

\[
E' < E + \int [v'(\vec{r}) - v(\vec{r})]n_0(\vec{r})d\vec{r}. \tag{2.3}
\]

Interchanging primed and unprimed variables, we can get

\[
E < E' + \int [v(\vec{r}) - v'(\vec{r})]n_0(\vec{r})d\vec{r}. \tag{2.4}
\]

Add Eq. (2.3) and Eq. (2.4), the inconsistency comes out as

\[
E + E' < E + E'. \tag{2.5}
\]

The only solution is that \( |\Psi\rangle \) and \( |\Psi'\rangle \) contribute to different density. So the external potential \( v(\vec{r}) \) is a unique functional of \( n_0(\vec{r}) \).

(ii) The second step is to prove that there is a universal functional \( F[n_0] \), and the ground state of the system is the global minimum of the energy functional with the density \( n_0(\vec{r}) \) as the exact ground state density.
We define the functional has the form

$$F[n_0(\vec{r})] \equiv \langle \Psi | T + U | \Psi \rangle,$$  

(2.6)

where $F[n_0]$ is a universal functional, valid for any number of particles and any external potentials.

For a given external potential $v(\vec{r})$, the energy functional can be written as

$$E_v[n_0] = \int v(\vec{r})n_0(\vec{r})d\vec{r} + F[n_0].$$  

(2.7)

With the right density $n_0$, $E_v[n_0]$ equals to the ground state energy.

Assume that the number of electrons is fixed, $\int n_0(\vec{r})d\vec{r} = N$, the energy functional $E_v[n_0]$ should has the minimum value for the correct density $n_0(\vec{r})$. Consider any other kind of wave function $|\Psi'|$, we have

$$E_v[\Psi'] = \int v(\vec{r})n'_0(\vec{r})d\vec{r} + F[n'] > E_v[\Psi] = \int v(\vec{r})n_0(\vec{r})d\vec{r} + F[n].$$  

(2.8)

So the energy $E_v(n_0)$ has the minimum value with correct density $n_0(\vec{r})$, which is the ground state charge density.

From the Hohenberg-Kohn theorem, the properties of the many-body system is easily determined if we know the universal functional $F[n_0]$. However, in their work, they could not specify the exact form of the functional. So the major part of the complexities of the many-body problems is associated with how to find a practical form of the universal functional $F[n_0]$.

### 2.1.2 Kohn-Sham scheme and Local Density Functional Approximation (LDA)

In 1965, a year after the Hohenberg-Kohn’s paper, Kohn and Sham published their work in the development of density functional theory [5]. They replaced the many-body variational problem with the independent-particle calculation, including all the many-body effect in a potential, which is called exchange-correlation term. By providing a set of auxiliary equations, they use a self-consistent procedure to solve the ground state of the system. In the following part of this section, we will list the major steps of their work.
Generally, the energy of the system can be written as

\[ E = \int v(\vec{r})n(\vec{r})d\vec{r} + \frac{1}{2} \int \int \frac{n(\vec{r})n(\vec{r}')}{|\vec{r} - \vec{r}'|}d\vec{r}d\vec{r}' + G[n], \]  

(2.9)

where \( n(\vec{r}) \) is the density and \( G[n] \) is a universal functional of the density.

Furthermore, the functional part can be split into

\[ G[n] \equiv T[n] + E_{xc}[n], \]  

(2.10)

where the term \( T[n] \) is the kinetic energy of a fictitious non-interacting system with the same ground-state density, and \( E_{xc}[n] \) is the exchange-correlation energy. The exchange-correlation energy contains all the many-body effect except for the Hartree term, and it is impossible to know the exact form of it. For a slowly varying \( n(\vec{r}) \), the exchange-correlation energy can be assumed to has the localized form, which is written as

\[ E_{xc} = \int n(\vec{r})\varepsilon_{xc}(n(\vec{r}))d\vec{r}, \]  

(2.11)
\( \varepsilon_{xc}(n) \) is the exchange-correlation energy per electron of a uniform electron gas of density \( n \). This is called the local density functional approximation (LDA), in which the exchange-correlation energy is determined by the corresponding term \( \varepsilon_{xc}(n(\vec{r})) \) from the homogeneous electron gas.

Subject to the constraint of a fixed number of all electrons, one can get the variational form
\[
\int \delta n(\vec{r}) \left\{ \varphi(\vec{r}) + \frac{\delta T[n]}{\delta n(\vec{r})} + \mu_{xc}(n(\vec{r})) \right\} d\vec{r} = 0.
\] (2.12)

Here
\[
\varphi(\vec{r}) = v(\vec{r}) + \int \frac{n(\vec{r}^\prime)}{|\vec{r}^\prime - \vec{r}|} d\vec{r}^\prime.
\] (2.13)

and
\[
\mu_{xc}(n) = \frac{d(n \varepsilon_{xc}(n))}{dn}.
\] (2.14)

So the variation form of the Eq. (12) results in a single-particle Schrödinger-like equation as
\[
\{-\frac{1}{2} \nabla^2 + [\varphi(\vec{r}) + \mu_{xc}(n(\vec{r}))] \} \psi_i(\vec{r}) = \epsilon_i \psi_i(\vec{r}),
\] (2.15)

and the density is
\[
n(\vec{r}) = \sum_{i=1}^{N} |\psi_i(\vec{r})|^2.
\] (2.16)

Eq. (2.15) is also called the Kohn-Sham equation. This equation significantly simplifies the many-body problem and reduces it to an independent particle equation with the new potential \( \varphi(\vec{r}) + \mu_{xc}(n(\vec{r})) \).

The equations from Eq. (2.13) to Eq. (2.16) consist of the self-consistent procedure in practical calculations. One can first guess an initial density \( n(\vec{r}) \). With Eqs. (2.13) and (2.14), the potential \( \varphi(\vec{r}) \) and \( \mu_{xc}(n(\vec{r})) \) can be obtained. Using these potentials, one can solve Eq. (2.15) and get the eigenvalues and eigenfunctions. Finally the new density can be constructed from the eigenfunctions obtained by the last step, and one can begin a new round of self-consistent steps. As the solution of the Kohn-Sham equation converges, the total energy can be obtained as
\[
E = \sum_{i=1}^{N} \epsilon_i - \frac{1}{2} \int \int \frac{n(\vec{r})n(\vec{r}^\prime)}{|\vec{r} - \vec{r}^\prime|} d\vec{r} d\vec{r}^\prime + \int n(\vec{r})[\varepsilon_{xc}(n(\vec{r})) - \mu_{xc}(n(\vec{r}))] d\vec{r}.
\] (2.17)
Figure 2.2: Correlation energy of an unpolarized homogeneous electron gas as a function of the density parameter $r_s$ [32].
where $\epsilon_i$ is the eigenvalue of Kohn-Sham particles.

Finally, the form of the exchange-correlation term in Eq. (2.14) has to be specified. The exchange part of this term can be analytical calculated with the Hartree-Fork approximation, while the correlation energy of free electron gas is obtained with some other methods, such as the quantum Monte Carlo method (QMC). By fitting simulation results, several models are provided, such as the “Ceperley-Alder” form (Pedew-Zunger (PZ)) [33, 34]. In Fig. 2.2, the comparison of several simulations is shown. In practical calculations, one can choose one of these correlations depending on the detail of the system studied.

It has to be pointed out that the eigenvalue and eigenfunction of the Kohn-Shame equation, Eq. (2.15), only have a mathematical meaning. Although in some special cases they can be related to some quantities of a real physics system, they do not contain the relevant physical information generally.

For example, the eigenvalues cannot be thought as the eigenenergy of the many-body system. The only direct meaning is that the sum of the modula of eigenfunctions gives the physical density of the system. However, it is desirable to connect the Kohn-Sham state to the electronic state in a real system state. A popular case is that one calculates the band structure of materials with the Kohn-Sham method. Of course, it is not rigorously right, because the band structure under LDA corresponds to the states of the Kohn-Sham particle, a collection of fictitious particles, instead of the quasiparticle states of a many-body system. But this method is widely used in a lot of calculations, because to people’s surprise, in most cases, the band structure of the Kohn-Sham particles is close to the experimental results. As a result, despite of the above difference, it is a standard way to calculated the electronic band structure with the Kohn-Sham equation as a first estimation.

The main problem of the Kohn-Sham band structure is that it systematically underestimates the band gap. In fact, this is a problem of density functional theory, because the above two gaps are completely different. The Kohn-Sham gap is the gap of the Kohn-Sham particles, while the experimental gap is the electron (or quasiparticle) gap. However, there are a lot of people who believe that the Kohn-Sham equation can give the right band gap, and this problem can be solved by finding the right exchange-correlation form. As a result,
some models such as hybrid LDA, are developed. In real calculation these models usually improve the band gap values. However, we think it is not a right way to correct the band gap problem. To answer this question, a quasiparticle concept is needed, and a rigorous self-energy calculation is necessary. One of the candidates is the Green’s function method and is developed as the GW approximation, which is the topic of the third Section of this Chapter.

Despite the meaning of the Kohn-Sham equation, the Kohn-Sham theory is very successful in solving the many-body system. For instance, the structural properties of materials such as the lattice constant, the bulk modulus, and the cohesive energy are generally determined to within a few percent of the experiment values. Throughout my whole PhD work, the LDA works as the first step to all calculations.

2.2 Pseudopotential

When people numerically calculate the electronic structure in materials, a highly efficient and easily implemented algorithm is needed. The plane wave method is one of the most hopeful candidates, and the orthogonal plane wave method (OPW) is widely used in the early calculations [35, 36]. However, the deep and sharp nuclei attractive potential makes the number of plane waves in OPW very large and slows the calculation speed extremely. To avoid this difficulty, people suggest many models, and the pseudopotential is one of the successful solutions.

The key point of the pseudopotential is that the core electron states are not changed much under different chemical environments. The main contribution of the core electron states is to make the valence electron orthogonal to them. So we can model the this effect as a potential, which is repulsive because of the orthogonality. This repulsive potential can cancel part of the sharp and deep nuclei ionic potential and decreases the number of plane waves needed to expand the electronic states in solids.

The pseudopotential was first introduced by Fermi in 1934 [37], and was popular after the 1950’s. We start this section from the Philips-Kleinman cancellation theorem under the OPW method [38].
The general form of valence electron states can be written with

\[ |\Psi\rangle = |\Phi\rangle + \sum C_{\text{core}} |\Phi_{\text{core}}\rangle, \]  

(2.18)

where the part \( |\Phi_{\text{core}}\rangle \) is the core electron state and has the sharp oscillation. \( |\Phi\rangle \) is hoped to be the smooth part of the electron states. Since the valence electron states are orthogonal to the core states, we have

\[ C_{\text{core}} = -\langle \Phi_{\text{core}} | \Phi \rangle. \]  

(2.19)

Then substitute Eq. (2.18) into the static Schrodinger equation \( H |\Psi\rangle = E |\Psi\rangle \), we get

\[ H |\Phi\rangle - \sum E_{\text{core}} \langle \Phi_{\text{core}} | \Phi \rangle |\Phi_{\text{core}}\rangle = E (|\Phi\rangle - \sum \langle \Phi_{\text{core}} | \Phi \rangle |\Phi_{\text{core}}\rangle). \]  

(2.20)

We can rearrange the formula to

\[ (H + \sum_{\text{core}} (E - E_{\text{core}})|\Phi_{\text{core}}\rangle \langle \Phi_{\text{core}}|)|\Phi\rangle = E |\Phi\rangle. \]  

(2.21)

The new operator \( \sum_{\text{core}} (E - E_{\text{core}})|\Phi_{\text{core}}\rangle \langle \Phi_{\text{core}}| \) shown with the original Hamiltonian acts like a potential term, and is defined as the pseudopotential

\[ V_{ps} = \sum_{\text{core}} (E - E_{\text{core}})|\Phi_{\text{core}}\rangle \langle \Phi_{\text{core}}|. \]  

(2.22)

With the help of the pseudopotential, the Schrodinger equation can be rewritten as

\[ (H + V_{ps})|\Phi\rangle = E |\Phi\rangle. \]  

(2.23)

Because \( |\Phi\rangle \) is the smooth part of the wave function \( |\Psi\rangle \), the solution to Eq. (2.23) needs a much smaller number of plane waves and improves the efficiency of the calculation significantly. So the plane wave method in conjunction with the pseudopotential is very popular in numerical calculations.

In the point view of physics, the pseudopotential term \( V_{ps} \) acts as a repulsive potential, because the energy of valence electrons is always higher than that of core electrons. It is also short-ranged because the core electronic state is close to the nuclei. As a result, the repulsive and short-range pseudopotential cancels the sharp and deep part of the ionic nuclei potential. Furthermore, as shown in Eq. (2.21), the pseudopotential method gives
the same eigenvalue despite of the wave function is replaced with the pseudo wave function, which is favorable in numerical calculations.

There are mainly two types of pseudopotentials. One is the empirical pseudopotential \cite{39, 40}, and the other one is called \textit{ab initio} pseudopotential. The first one uses the parameterized pseudopotential to fit known experiment results and to calculate new properties of materials with this fitted pseudopotential. But this method needs those parameters, which are strongly depending on the fitting process and cannot be thought as a complete first-principles approach. Alternately one can directly calculate core electron states with density functional theory and construct the pseudopotential without fitting to experiments, which is called the \textit{ab initio} pseudopotential. This \textit{ab initio} pseudopotential has really good transferability and accuracy. It is widely used in today’s first-principles calculations. In my PhD work, we use two of the \textit{ab initio} pseudopotentials (the Trullier-Martin form and the ultrasoft form) \cite{41, 42}. In the following, the main framework of the Trullier-Martin pseudopotential is introduced.

The Troullier-Martin pseudopotential belongs to a type of \textit{ab initio} pseudopotentials, the norm-conserving pseudopotential. This kind of pseudopotentials provide a straightforward way to construct the \textit{ab initio} pseudopotential and have good transferability. Here I first list the four general conditions for norm-conserving pseudopotentials. With the separation of variables, the atomic Kohn-Sham equation along the radial direction can be written as

$$[-\frac{1}{2} \frac{d^2}{dr^2} + \frac{l(l+1)}{2r^2} + V(n; r)]rR_{nl}(r) = \varepsilon_{nl} rR_{nl}(r).$$

Then the four conditions are:

1. The valence pseudo-wave-functions generated from the pseudopotential contain no nodes. The reason of this requirement is that we would like to obtain the smooth pseudo-wave-functions which need small number of plane waves.

2. The normalized pseudo-wave-function (PP) is equal to the normalized all-electron wave function (AE) beyond the chosen cutoff radius \(r_{cl}\),

$$R^{PP}_l(r) = R^{AE}_l(r),$$

for \(r > r_{cl}\), or converges rapidly to that value.
(3) The charge enclosed inside \( r_{cl} \) for the above two wave functions must be equal,
\[
\int_0^{r_{cl}} |R_{l}^{PP}(r)|^2 r^2 dr = \int_0^{r_{cl}} |R_{l}^{AE}(r)|^2 r^2 dr.
\] (2.26)

(4) The valence all-electron and pseudopotential eigenvalues must be equal,
\[
\varepsilon_{l}^{PP} = \varepsilon_{l}^{AE}.
\] (2.27)

When the pseudopotential satisfies the above four conditions, it is called the norm-conserving pseudopotential. The early work is done by Kerker [43], who define the pseudo-wave-function to be have the form as
\[
R_{l}^{PP} = R_{l}^{AE},
\] (2.28)
when \( r > r_{cl} \), and
\[
R_{l}^{PP} = r^l \exp[p(r)],
\] (2.29)
when \( r < r_{cl} \), where \( p(r) \) is a polynomial of \( r \),
\[
p(r) = c_0 + \sum_{i=2}^{4} c_i r^i.
\] (2.30)

Troullier and Martins use a better form of Eq. (2.29). They use seven parameters to optimize the pseudo-wave-function instead of four in Kerker’s work. The polynomial used in Troullier-Martins pseudopotential is
\[
p(r) = c_0 + c_2 r^2 + c_4 r^4 + c_6 r^6 + c_8 r^8 + c_{10} r^{10} + c_{12} r^{12}.
\] (2.31)

As we can see in the general form of pseudopotentials, Eq. (2.22), the pseudopotential is different for wave functions with different angular momenta. This character of pseudopotential is called semi-local. As a result, the pseudopotential directly determined from the atomic configuration has both local and non-local components. If we project them into the spherical harmonic basis, it has the form as
\[
\hat{V}_{PS} = V_{local}(r) + \sum_{lm} |Y_{lm}\rangle \hat{V}_{non-local} |Y_{lm}\rangle.
\] (2.32)

This is not efficient for computations. It is proposed by Kleinman and Bylander (KB) that there is a projecting process which can transfer the above pseudopotential form into a
complete non-local form [44]. This scheme is widely used in today’s calculation and saves much of the computation time.

One parameter in the above introduction is very important, which is the cutoff radius $r_{cl}$. The choice of the cutoff radius determines the accuracy and the transferability of the pseudopotential. The smaller cutoff radius needs more plane waves to simulate the potential (hard core), but it results in a better transferable potential; the larger cutoff radius can save the plane waves (soft core), but the transferability is worse. So it is important to find a balanced cutoff radius in first-principles calculations.

Troullier-Martins pseudopotentials are very good in many cases. It can reach the accuracy with a smaller number of plane waves. The comparison with some general pseudopotential is shown in Fig. 2.3. From this figure, it is clear that the Troullier-Martins form is smoother in fourier space, which is favorable in numerical simulations.

In the last part of my PhD work, the ultrasoft pseudopotential is used in Vienna Ab-Initio Simulation Package (VASP). This kind of pseudopotential is not norm-conserved, because it relax the restriction so that the charge within the cutoff radius $c_{cl}$ is not equal to the all-electron result. The good new of the ultrasoft pseudopotential is that it can obtain a
smoother potential because of the lack of charge restriction, which speed up the Kohn-Sham calculation. The bad news is that it needs an extra procedure to compensate the charge difference and makes it hard to write computer programs for other calculations, such as the excited states calculation.

### 2.3 Many-body Perturbation Theory

It is well known that the density functional theory can describe well the ground state of electronic systems. On the other hand, a lot of interesting physics phenomena are not only the properties of electronic ground state, they are tightly related to the excited states of electrons, such as photoemission, electron loss spectrum, and optical absorption, etc. In not strongly correlated electronic systems, most of the above phenomena are related to the electrons close to Fermi level, which make it possible to describe the collective behavior of these excited states with the quasiparticle concept.

The basic idea of quasiparticle concept is to use weakly interacting particles to replace the strongly interacting many-body system. For example, the electrons in solids strongly interact via the Coulomb interaction. However, when we see the collective behavior of this system, it is found that the electrons repel each other, which results in a depletion of electrons and effectively a positive charge cloud around the bare electrons. This surrounding positive charge cloud will screening the strong bare Coulomb interaction. The sum of bare electrons and the positive charge cloud can be thought as quasiparticles, and this idea reduces the strongly interacting electrons to a weakly interacting quasiparticle system. Of course, the introduction of the quasiparticle concept gives a non-Hermisian Hamiltonian for the quasiparticle equation. However, if the excitation is close to the Fermi level, quasiparticle picture is good enough to describe the excitation, and the life time of quasiparticle should be long enough, which makes the approximation reasonable.

The Green’s function method provides a powerful way to solve the strongly interacting electronic systems. With the help of the Green’s function, the self-energy of quasiparticles, which includes all the exchange and correlating effects, can be expanded in terms of the interaction between the particles. However, it is very hard to get a converged result if
we expand the self-energy in terms of the bare interaction between electrons. In the 1960s, Hedin suggested a new expansion scheme, which uses the screened interaction to expand the self-energy [19]. This method is very successful when applied to semiconductor materials. The first order of this expansion is to describe the self-energy \( M \) by the product of the one-particle Green’s function \( G \) and the screened interaction \( W \), \( M = iGW \). Although this approximation should optimize the product with the right quasiparticle wave function, one typically optimizes the Green’s function and the screened interaction, separately, without self iteration. This practical way is not very rigorous, but this separated optimization is enough to give an accurate answer. The above scheme is called the GW approximation.

Through this quasiparticle calculation, the well-known problem of DFT, an underestimated band gap, is solved to a large extent.

When we consider the optical absorption of semiconductor materials, the above GW approximation is not enough. When the photon is absorbed, an electron is excited to the conduction band and leaves a vacancy in the valence band, which is a quasiparticle called the hole. The excited electron and hole interact with each other, and form a new kind of excitation, which is called the exciton or electron-hole pair. In this case, the single-particle Green’s function is not enough, and we have to construct the two-particle Green’s function to describe excitons. The corresponding equation of the two-particle Green’s function is Bethe-Salpeter equation (BSE) [26, 45, 46]. The derivation of Bethe-Salpeter equation can be thought as the second order expansion of the self-energy with the screened interaction.

Solving of the Bethe-Salpeter equation follows three steps in first-principles calculations. (i) Solve the Kohn-Sham equation, and use it as the ground state; (ii) Calculate the quasiparticle energy with the GW approximation, which gives the accurate band structure of electrons; (iii) with the wave function and eigenvalues from the previous steps, the Bethe-Salpeter equation is solved and the optical absorption is determined by the exciton solutions. A good review can be found in [27]. We will see that the whole calculations in chapter IV follow these three steps.

We ignore the details discussion of the Green’s function in the thesis, because readers can find many excellent textbooks on this topics. The main point we should remember is
that the form of the Green’s function makes it easy to calculate most of the interesting properties of the system, such as total energy, excitation spectrum, etc. As the beginning of the introduction, we would like to start with the quasiparticle (GW) approximation. In the following subsections, the derivation of the Bethe-Salpeter equation will be shown.

2.3.1 Hedin Equations

The motivation of considering quasiparticle is from the failure in the band gap calculation from density functional theory. Although the DFT band gap uses the eigenvalues from the Kohn-Sham equation, which does not represent the electron eigenenergy, it is convenient to interpret the Khon-Sham eigenvalues as the electronic band structure. In order to solve this problem, we have to consider the self-energy of electrons. The following derivation is mainly from the Hedin’s work [19] and is implemented into the plane wave method from 1980s.

The Hamiltonian of the electronic system can be written as

\[ H = H_0 + H_1, \]  

(2.33)

where the unpertubed part

\[ H_0 = \int \psi^\dagger(x)h(x)\psi(x)dx + \frac{1}{2}\int \psi^\dagger(x)\psi^\dagger(x')v(x, x')\psi(x')\psi(x)dx\,dx'. \]  

(2.34)

The potential term \( v(x, x') \) is the bare Coulomb potential. The perturbation term is

\[ H_1 = \int \rho(x)w(x, t)dx, \]  

(2.35)

where \( \rho(x) \) is the charge density

\[ \rho(x) = \psi^\dagger(x)\psi(x), \]  

(2.36)

and \( w(x, t) \) is the external time-dependent perturbation.

Without the perturbation, the time-evolve operator is defined as \( U(t, t') \). With the perturbation term \( H_1 \), the time-evolution operator is defined as \( V(t, t') \). In the starting and final states, the perturbation is turned off. In the intermediate state, we turn on the perturbation slowly, which is guaranteed by Gelmann-Low theorem that all eigenstates are not changed much from their original position.
From the Schrödinger equation, we have

$$V(t, t') = U(t, t') - \frac{i}{\hbar} \int_{t'}^{t} U(t, t'')H_1(t'')V(t'', t')dt''.$$ (2.37)

The field operator in Heisenberg representation is

$$\psi(x, t) = V(-T_0, t)\psi(x)V(t, -T_0).$$ (2.38)

The Schrödinger equation of the field operator is

$$i\hbar \frac{\partial \psi(x, t)}{\partial t} = V(-T_0, t)[\psi(x), H_0 + H_1]V(t, -T_0).$$ (2.39)

By including the details of Hamiltonian in Eq. (35) and (36), the integral equation can be derived as

$$[i\hbar \frac{\partial}{\partial t} - h(x) - w(x, t)]\psi(x, t) - \int v(x, x')\psi^*(x', t)\psi(x', t)dx'\psi(x, t) = 0.$$ (2.40)

Considering the step function satisfies $$\frac{d\theta}{dt} = \delta(t)$$, and the commutator of field operator as $$\psi(x)\psi^*(x') + \psi^*(x')\psi(x) = \delta(x, x')$$, the above equation can be written as

$$[i\hbar \frac{\partial}{\partial t} - h(x) - w(x, t)]T(\psi(x, t)\psi^*(x', t'))$$

$$- \int v(x, x')T(\psi^*(x'', t)\psi(x'', t)\psi(x, t)\psi^*(x', t'))dx'' = i\hbar\delta(x, x')\delta(t, t').$$ (2.41)

Here $T$ stands for the time-order operator, which is defined as

$$T(\psi(x, t)\psi^*(x', t')) = \psi(x, t)\psi^*(x', t')\theta(t - t') - \psi^*(x', t')\psi(x, t)\theta(t' - t).$$ (2.42)

Before we go to the next step, we have to specify one useful formula

$$\frac{\delta V(t, t')}{\delta w(x_2, t_2)} = -\frac{i}{\hbar} \text{sgn}(t - t')V(t, t_2)\rho(x_2)V(t_2, t').$$ (2.43)

With this formula, the time-order of field operators can be written as

$$\frac{\delta}{\delta w(3)}V(T_0, -T_0)T(\psi(1)\psi^*(2)) = -\frac{i}{\hbar}V(T_0, -T_0)T(\psi^*(3)\psi(3)\psi(1)\psi^*(2)),$$ (2.44)

by assuming $t_1$ to be between $T_0$ and $-T_0$. Here we use the simplified notations of the coordinates. For example, $\psi(x_1, t_1) = \psi(1)$. 24
Now we can define the one-particle Green’s function as

\[
G(1, 2) = -\frac{i}{\hbar} \frac{\langle N|U(-T_0, T_0)V(T_0, -T_0)T(\psi(1)\psi^\dagger(2))|N \rangle}{\langle N|U(-T_0, T_0)V(T_0, -T_0)|N \rangle},
\]  

(2.45)

where the \( |N \rangle \) is the ground state of the electronic system with \( w = 0 \). So this one-particle Green’s function is defined with the ground state including the interaction between particles.

With the help of above definitions, we can write down the evolution equation of this one-particle Green’s function as

\[
(i\hbar \frac{\partial}{\partial t_1} - h(1) - V(1))G(1, 2) - i\hbar \int v(1^+, 3) \frac{\delta}{\delta w(3)} G(1, 2)d(3) = \delta(1, 2).
\]  

(2.46)

where the \( V(1) \) is defined as

\[
V(1) = w(1) + \int v(1^+, 3) \frac{\langle N|U(-T_0, T_0)V(T_0, -T_0)T(\psi(1)\psi^\dagger(3))|N \rangle}{\langle N|U(-T_0, T_0)V(T_0, -T_0)|N \rangle} d(3).
\]  

(2.47)

Here \( 1^+ = (x_1, t_1 + \Delta) \) and \( v(1, 2) = v(x_1, x_2)\delta(t_1 - t_2) \).

With the definition of the one-particle Green’s function, we can simplify the formula of \( V(1) \) to be

\[
V(1) = w(1) - i\hbar \int v(1, 3)G(3, 3^+)d(3).
\]  

(2.48)

Now we can define the self-energy operator \( M \) by

\[
(i\hbar \frac{\partial}{\partial t_1} - h(1) - V(1))G(1, 2) - \int M(1, 3)G(3, 2)d(3) = \delta(1, 2).
\]  

(2.49)

It is clear that all the exchange and correlation effect, which are not included in Hartree term, are all summed into the self-energy operator.

Because we will use the inverse of the Green’s function later, so we also define the inverse as

\[
\int G(1, 3)G^{-1}(3, 2)d(3) = \delta(1, 2),
\]  

(2.50)

The response of the Green’s function to the external potential can be derived as

\[
\frac{\delta G(1, 2)}{\delta w(3)} = -\int G(1, 4) \frac{\delta G^{-1}(4, 5)}{\delta w(3)} G(5, 2)d(4)d(5)
\]  

(2.51)

With all the definitions above, we can write the self-energy operator in the Green’s function form as

\[
M(1, 2) = -i\hbar \int v(1^+, 3)G(1, 4) \frac{\delta G^{-1}(4, 2)}{\delta w(3)}d(3)d(4).
\]  

(2.52)
Because we want to use the screened interaction, it is defined here

$$W(1, 2) = \int v(1, 3) \frac{\delta V(2)}{\delta w(3)} d(3).$$  \hspace{1cm} (2.53)

By considering the definition of $V$, we can write the screened interaction as

$$W(1, 2) = v(1, 2) + i\hbar \int v(1, 3) v(2, 4) G(4, 5) \frac{\delta G^{-1}(5, 6)}{\delta w(3)} G(6, 4^+) d(3)d(4)d(5)d(6).$$  \hspace{1cm} (2.54)

Using the chain rule, the variation of external potential has the form as

$$\frac{\delta}{\delta w(1)} = \int \frac{\delta V(2)}{\delta w(1)} \frac{\delta}{\delta V(2)} d(2).$$  \hspace{1cm} (2.55)

Finally, the screened interaction can be written as

$$W(1, 2) = v(1, 2) + \int W(1, 3) P(3, 4) v(4, 2) d(3)d(4),$$  \hspace{1cm} (2.56)

where the polarizability operator is defined as

$$P(3, 4) = i\hbar \int G(4, 5) G(6, 4^+) \frac{\delta G^{-1}(5, 6)}{\delta V(3)} d(5)d(6).$$  \hspace{1cm} (2.57)

In order to simplify the formula, we introduce the vertex function $\Gamma$, which is defined as

$$\Gamma(1, 2; 3) = -\left(\delta G^{-1}(1, 2)/\delta v(3)\right).$$  \hspace{1cm} (2.58)

Because $G^{-1} = (i\hbar \frac{\partial}{\partial t} - h(1) - V(1)) - M$, so

$$\Gamma(1, 2; 3) = \delta(1, 2)\delta(1, 3) + (\delta M(1, 2)/\delta V(3)).$$  \hspace{1cm} (2.59)

Then we get the final version of four Hedin’s equations:

$$W(1, 2) = v(1, 2) + \int W(1, 3) P(3, 4) v(4, 2) d(3)d(4);$$  \hspace{1cm} (2.60)

$$M(1, 2) = i\hbar \int W(1^+, 3) G(1, 4) \Gamma(4, 2; 3) d(3)d(4);$$  \hspace{1cm} (2.61)

$$P(1, 2) = -i\hbar \int G(2, 3) G(4, 2^+) \Gamma(3, 4; 1) d(3)d(4);$$  \hspace{1cm} (2.62)

$$\Gamma(1, 2; 3) = \delta(1, 2)\delta(1, 3) + \int \frac{\delta M(1, 2)}{\delta G(4, 5)} G(4, 6) G(7, 5) \Gamma(6, 7; 3) d(4)d(5)d(6)d(7).$$  \hspace{1cm} (2.63)

It is clear that the above equations consist of a loop, which means that we have to truncate it with some approximation. If we truncate the vertex function in the Hedin’s equations, the GW approximation is the first order expansion result, and the Bethe-Salpeter equation is the second order expansion.

26
2.3.2 GW Approximation

The simplest approximation is to keep only the first term of the vertex function. With

\[ \Gamma(1, 2; 3) = \delta(1, 2)\delta(1, 3). \]  

(2.64)

This is a random phase approximation. Then the polarizability is given by

\[ P(1, 2) = -i\hbar G(1, 2^+)G(2, 1). \]  

(2.65)

As a result, the self-energy operator is simplified to

\[ M(1, 2) = i\hbar G(1, 2)W(1^+, 2). \]  

(2.66)

So the self-energy operator is the product of one-particle Green’s function and the screened Coulomb interaction term, this is also why this approximation is called the GW approximation.

The flowing part of this subsection is focused on how to implement the GW approximation in first-principles calculations. The next subsection will truncate at the higher order of the Hedin’s equations, which will give the Bethe-Salpeter equation concerning with the excitonic effect.

It is widely used to divide the electrons in solids into two kinds. One is the core electrons, which do not strongly affect the chemical properties of materials; the other is the valence electrons, which are active in the formation of chemical bonds. When considering most of the electronic properties of solids, we would like to focus on the valence electrons; the core electrons only contribute to the orthogonalization of the valence wave functions, which is also the origin of the pseudopotential idea. As a result, we divide the Green’s function and the screened interaction into two parts: the contributions from valence and core electrons. It is found that the self-energy operator can be written as

\[ M = i\hbar(G_c W + G_v W_c P_c W_v + G_v W_c) \]  

(2.67)

In real calculations, we find that the first two terms are much smaller than the last one in most semiconductors. So our first-principles calculations only consider the GW
correction from the valence electrons ignoring the corrections from core electrons and the interaction between core and valence electrons. The rest part of the GW calculation follows the framework in Ref. [20].

The Fourier transform of the self-energy term is

\[
M(r, r'; E) = i \int \frac{dE'}{2\pi} e^{-i\delta E'} G(r, r'; E - E') W(r, r'; E'),
\]

(2.68)

where \(\delta = 0^+\). We can also make a decomposition of the screened interaction as

\[
W(r, r'; E) = v(r, r') + \int dE' \frac{2E'B(r, r'; E)}{E^2 - (E' - i\eta)^2},
\]

(2.69)

where \(\eta = 0^+\), and \(B\) is the spectral function for the screened interaction.

In order to calculate the self-energy, we have to introduce some approximations. Here the Coulomb hole + Screened exchange (COHSEX) approximation is used to split the self-energy. The COHSEX approximation is a physically motivated to separate of the self-energy into a Coulomb-hole (COH) part and a screened-exchange (SEX) part whose static limit has been used extensively to correct the Hartree-Fock band structure. In GW calculations, if we assume the life time of the quasiparticle is long enough, then we only consider the real part of the self-energy, and break it into

\[
Re M = M_{SEX} + M_{COH},
\]

(2.70)

where the screened exchange part is from the pole of Green’s function, and the Coulomb hole part is from the pole of the screened interaction.

In first-principles calculations, we use the result from the local density approximation (LDA) as the ground states. Here we try to use these wave function to construct the Green’s function as

\[
G(r, r'; E) = \sum_{n,k} \frac{\phi_{nk}(r) \phi_{nk}^*(r')}{E - \varepsilon_{nk} - i\delta_{nk}},
\]

(2.71)

where the \(\phi_{nk}(r)\) and \(\phi_{nk}^*(r')\) are the wave functions from LDA results. To make the integral meaningful, \(\delta_{nk} = 0^+\) for \(\varepsilon_{nk} < \mu\), and \(\delta_{nk} = 0^-\) for \(\varepsilon_{nk} > \mu\), where \(\mu\) is the chemical potential.
Then the real part of the self-energy can be written as

\[ M_{SEX}(\mathbf{r}, \mathbf{r}'; E) = -\sum_{n,k}^{occ} \phi_{nk}(\mathbf{r})\phi_{nk}^{*}(\mathbf{r}')W(\mathbf{r}, \mathbf{r}'; E - \varepsilon_{nk}); \quad (2.72) \]

\[ M_{COH}(\mathbf{r}, \mathbf{r}'; E) = \sum_{n,k} \phi_{nk}(\mathbf{r})\phi_{nk}^{*}(\mathbf{r}')P\int_{0}^{\infty} dE' \frac{B(\mathbf{r}, \mathbf{r}'; E)}{E - \varepsilon_{nk} - E'}. \quad (2.73) \]

There is one thing we have to point out in the above formulations. The SEX term only sums up the valence wave function, because we want the loop integration to converge. In Eq. (69), the integration loop is under the real number axis to satisfy the convergence, because the imaginary of the exponent has to be negative. When combined with the Green’s function in Eq. (72), we have to require \( \delta_{nk} = 0^+ \), which results in \( \varepsilon_{nk} < \mu \). However, in the COH term, the integration is from zero to infinity, so it covers all the electronic states. This difference between the SEX and COH term, will be reflected later in GW calculations. The SEX term is relatively cheaper, while the COH term is extremely expensive. In fact, it is impossible to include all the bands in COH term, we have to choose a finite number of bands to get the converged result. Generally, it is safe to calculate 5-10 times more bands than the number of the valence bands in semiconductors to get the GW correction.

From Eq. (2.72) and Eq. (2.73), it is clear that the key step for the GW correction is to calculate the screened interaction. The screened interaction is determined by the screened dielectric constant. Our strategy is a two-step frame work. First we calculate the static dielectric constant; the second step is to extend the static result to a dynamical case with the general plasmon pole model (GPP) [47].

As the starting point, we use the definition of the screened interaction

\[ W(1, 2) = \int d(3) \epsilon^{-1}(1, 3)v(3, 2). \quad (2.74) \]

When transferred into the k space and frequency space, the convolution becomes a product as

\[ W_{GG'}(q, \omega) = \epsilon^{-1}_{GG'}(q, \omega)v(q + G'), \quad (2.75) \]

with the pane-wave basis.
In random phase approximation (RPA), the polarizability is related with the dielectric constant

$$\epsilon_{GG'}(q, \omega) = \delta_{GG'} - v(q + G)P_{GG'}(q, \omega). \quad (2.76)$$

First, we want to consider the static case. According to the work of Adler and Wiser, the static polarizability is

$$P_{0GG'}(q, \omega = 0) = \sum_{nn'k} \langle n, k | e^{-i(q+G) \cdot r'} | n', k + q \rangle \langle n' , k + q | e^{-i(q+G') \cdot r} | n,k \rangle f(\epsilon_{n',k+q}) - f(\epsilon_{n,k}) \epsilon_{n',k+q} - \epsilon_{n,k}, \quad (2.77)$$

where $\epsilon_{n,k}$ is the energy of the state with band number $n$ and wave vector $k$, and $f$ is the occupation number of the electronic state.

When we get the static polarizability, we also obtain the static dielectric constant. In fact, we mainly use the inverse of the dielectric constant. In the plane-wave method, the dielectric tensor is a huge matrix. In practical first-principles calculations, the inversion operation of the dielectric matrix will take very long time and is the bottle neck of the whole calculation.

Next we want to extend the static result to dynamical case. Under the general plasmon pole model, the imaginary part of the dynamical dielectric tensor is

$$\text{Im}\epsilon_{GG'}^{-1}(q, \omega) = A_{GG'}(q) \delta[\omega - \tilde{\omega}_{GG'}(q)] - \delta[\omega + \tilde{\omega}_{GG'}(q)], \quad (2.78)$$

and the real part is

$$\text{Re}\epsilon_{GG'}^{-1}(q, \omega) = 1 + \frac{\Omega^2_{GG'}(q)}{\omega^2 - \tilde{\omega}^2_{GG'}(q)}. \quad (2.79)$$

To calculate the above dynamical dielectric constant, we have to specify the values of $A_{GG'}(q)$, $\Omega^2_{GG'}(q)$ and $\tilde{\omega}^2_{GG'}(q)$. We list the main results instead of the detail derivations here.

With the help of the Kramers-Kronig relation of the real and imaginary parts of the dielectric constant, the mode frequency can be written as

$$\tilde{\omega}^2_{GG'}(q) = \frac{\Omega^2_{GG'}(q)}{\delta_{GG'} - \epsilon^{-1}_{GG'}(q, \omega = 0)} \cdot (2.80)$$

The plasmon frequency is calculated with the form as

$$\Omega^2_{GG'}(q) = \omega^2 P \cdot \frac{(q + G) \cdot (q + G')}{|q + G|^2} \rho(G - G') \rho(0). \quad (2.81)$$
Here $\rho$ is the charge density from the LDA result, and $\omega_P$ is the frequency of the plasmon mode.

The amplitude $A$ is determined by

$$A_{GG'}(q) = -\frac{\pi}{2} \frac{\Omega_{GG'}(q)}{\tilde{\omega}_{GG'}(q)}. \quad (2.82)$$

With all the results above, we can write down the self-energy formulation in our first-principles calculations. The SEX term is

$$\langle nk | M_{SEX}(r, r'; E) | n' k \rangle = - \sum_{n_1, q, G, G'} \langle nk | e^{i(q+G)\cdot r} | n_1, k-q \rangle \langle n_1, k-q | e^{-i(q+G)\cdot r'} | n' k \rangle \times [1 + \frac{\Omega_{GG'}^2(q)}{(E - \varepsilon_{n_1 k-q})^2 - \tilde{\omega}_{GG'}^2(q)}] v(q + G'). \quad (2.83)$$

The COH term is

$$\langle nk | M_{COH}(r, r'; E) | n' k \rangle = - \sum_{n_1, q, G, G'} \langle nk | e^{i(q+G)\cdot r} | n_1, k-q \rangle \langle n_1, k-q | e^{-i(q+G)\cdot r'} | n' k \rangle \times \frac{1}{2} \frac{\Omega_{GG'}^2(q)}{\tilde{\omega}_{GG'}(q)[E - \varepsilon_{n_1 k-q} - \tilde{\omega}_{GG'}(q)]} v(q + G'). \quad (2.84)$$

In the above formulas, the state $|nk\rangle$ is the one-particle state of the electronic system from the LDA calculations.

In practical calculations, it is found that the non-diagonal term of the self-energy of semiconductors has a small correction to the final results, so when we perform the GW calculations, we only consider the diagonal terms. Furthermore, the calculational results show that the wave function of the quasiparticles is very similar to the LDA results. This is surprising and saves us a lot of time in higher order calculations. Therefore, we can still use the wave functions from LDA results in later calculations dealing with the Bethe-Salpeter equation.

Because of these useful features of the self-energy operator, the quasiparticle energy can be written as

$$E_{nk}^{qp} = \epsilon_{nk}^{LDA} - \langle nk | V_{XC}^{LDA} | nk \rangle + \langle nk | M(E_{nk}^{qp}) | nk \rangle. \quad (2.85)$$

One important aspect in above formula is that the quasiparticle energy should be solved iteratively, because the self-energy correction in the above formula depends on the quasiparticle energy too. It is found that the result without iterations is acceptable. So our GW
calculation is not self-consistent. However, it should be remembered that the self-consistent calculation may be important for other systems. For example, when the quasiparticle energy is not far from the plasmon frequency, the self-consistent scheme may be needed.

In real calculations, the above self-energy correction can be simplified to the general first order perturbation form as

\[
E_{\text{qp}}^{nk} = E_0^{nk} + \frac{\Delta M_{nk}(\varepsilon_{nk})}{1 - \Delta M_{nk}(\varepsilon_{nk})/\Delta E} \left( E_0^{nk} - \varepsilon_{nk} \right),
\]

(2.86)

Where the ground state energy is

\[
E_0^{nk} = \varepsilon^{\text{LDA}}_{nk} - \langle nk|V^{\text{LDA}}_{XC}|nk\rangle + \langle nk|M(\varepsilon^{\text{LDA}}_{nk})|nk\rangle.
\]

(2.87)

The typical energy perturbation \(\Delta E\) is around 0.5 eV. Our calculation shows that the final GW correction is not sensitive to the choice of \(\Delta E\).

By now, we have set up the main frame work of the first-principles quasiparticle calculation. It starts from the random phase truncation in Hedin’s equation, and most of the attention is paid to the dynamical screening calculation. In practical calculations, the self-consistence procedure is found not necessary, which improves the efficiency of the GW calculation significantly. Finally, all the calculations are implemented into the first-principles calculation based on the LDA results.

Although, the above GW calculational procedure is not extremely solid on the rigorous theoretical ground, such as ignoring the core electrons and non-self-consistence, it is very successful in most band structure calculations and agrees with the experiment results well. Today, the GW correction has been a standard way to obtain the trustable band structure in many solids.

2.3.3 Excitonic Effect (Bethe-Salpeter Equation)

As we show in the introduction part, the single-particle Green’s function is not enough when we consider the electron-hole pair interaction. We have to consider the next order approximation of the self-energy beyond the quasiparticle calculations in last section. This evolution equation of the two-particle Green’s function is known as the Bethe-Salpeter equation (BSE). There are many ways to obtain the BSE. Most of Green’s function experts
may prefer the “simplest” way to derive BSE, such as the ladder approximation. To me, I would like to choose the similar analytical method to derive the BSE as I did in the GW calculations. So in the following, most of derivation is from the Refs. [46, 48, 49].

Before we derive the Bethe-Salpeter equation, there are some definitions which we have to specify. The first one is the two-particle Green’s function. Ignore the normalization constant, the two-particle Green’s function is defined as

\[ G_{2}(1, 1'; 2, 2') = T\langle \psi(1)\psi(1')\psi(2)\psi(2')\rangle. \] (2.88)

From Eq. (2.45), we can see the variation of one-particle Green’s function is related to this by

\[ \frac{\delta G(1, 1'; w)}{\delta w(2)}|_{w=0} = (-i)^{2}\langle T[\psi(1)\psi(1')n(2)]\rangle|_{w=0} - \langle T[\psi(1)\psi(1')]\rangle\langle n(2) \rangle \]

\[ = [G_{2}(1, 2; 1', 2^{+}) - G(1 - 1')G(2 - 2^{+})]|_{w=0}. \] (2.89)

Later, we use the correlation of the two-particle Green’s function in the Bethe-Salpeter equation, so we have to define the correlation function as following

\[ L(1, 2; 1', 2') = [G_{2}(1, 2; 1', 2^{+}) - G(1 - 1')G(2 - 2^{+})]|_{w=0}. \] (2.90)

In Eq. (2.89), we find that

\[ L(1, 2; 1', 2^{+}) = \frac{\delta G(1, 1')}{\delta w(2)}|_{w=0}. \] (2.91)

Now we can start to give a simple derivation of the Bethe-Salpeter equation. Combine Eq. (2.46) and Eq. (2.89), we get the self-energy operator as

\[ M(1, 1') = i\int G^{-1}(1, 2)G_{2}(2, 3^{-}; 1', 3^{+})V(3 - 1')d(2)d(3). \] (2.92)

It is well known that the Green’s function has the Dyson equation form as

\[ G = G_{0} + G_{0}(w + M)G. \] (2.93)

So we have

\[ G^{-1}(1, 1') = G_{0}^{-1}(1, 1') - w(1, 1') - M(1, 1'). \] (2.94)
Because of $GG^{-1} = 1$, from Eq. (2.52) and Eq. (2.94), we have
\[
\frac{\delta G(1,1')}{\delta w(2',2)} = \int d(3)d(4)G(1,3)[\delta(3,2')\delta(4,2) + \frac{\delta M(3,4)}{\delta w(2',2)}]G(4,1').
\] (2.95)

If combined with Eq. (2.89), the evolution of the correlation function of the two-particle's Green’s function is
\[
L(1,2;1',2') = G(1,2')G(2,1') + \int d(3)d(4)G(1,3)\frac{\delta M(3,4)}{\delta w(2',2)}G(4,1').
\] (2.96)

The self-energy operator only depends on $w$ through its dependence on $G(w)$, so by the chain rule, the variation of the self-energy operator with $w$ can be written as
\[
\frac{\delta M(3,4)}{\delta w(2',2)}|_{w=0} = \int d(5)d(6)\frac{\delta M(3,4;G)}{\delta G(6,5)}|_{w=0} \times \frac{\delta G(6,5)}{\delta w(2',2)}|_{w=0}.
\] (2.97)

We can define the effective interaction kernel as
\[
K(3,5;4,6) = \frac{\delta M(3,4)}{\delta G(6,5)}|_{w=0}.
\] (2.98)

Then combining with the definition of the correlation function of the two-particle Green’s function, Eq. (2.89), we obtain the integration equation of $L$ as
\[
L(1,2;1',2') = G(1-2')G(2-1') + \int G(1-3)G(4-1')K(3,5;4,6)L(6,2;5,2')d(3)d(4)d(5)d(6).
\] (2.99)
If we define

\[ L_0(1, 2; 1', 2') = G(1 - 2')G(2 - 1'), \]  

(2.100)

Eq. (2.99) can be simplified as

\[ L(1, 2; 1', 2') = L_0(1, 2; 1', 2') + \int d(3)d(4)d(5)d(6)L_0(1, 4; 1', 3)K(3, 5; 4, 6)L(6, 2; 5, 2'). \]  

(2.101)

If we use Feynman diagrams to express Eq. (2.101), the figure is shown in Fig. 2.4. This kind of Feynman diagrams was first found in relativity theory and called the Bethe-Salpeter equation. So in condensed matter physics, when we consider the two-particle Green’s function, the general evolution equation is also the Behste-Salpeter equation (BSE). It is found that the Bethe-Salpeter equation in solids describe the two-particle excitation well, and can be used to explain the optical absorption spectrum. One of the excitations in solids related to BSE is the electron-hole pair. If we can solve BSE, which means we know the solution of the two-particle Green’s function, the excitation energy related to the electron-hole pair can be evaluated. Since 1960s, the Bethe-Salpeter equation has been widely used to compute the optical absorption properties in solids. In the flowing, we try to list the simple framework of the BSE in first-principles calculations.

The basic idea of the calculation is to implement the Bethe-salpeter equation into the plane-wave based first-principles calculations. The BSE can be thought as the higher order approximation of ground state calculations. One of the hopeful candidates for the zeroth order approximation is the local density approximation (LDA). So we will try to use the single-particle wave functions to express the Bethe-Salpeter equation.

According to the definition of \( L_0 \), we can write its expression in the frequency domain as

\[ L_0(1, 2; 1', 2'; \omega) = i \sum_{v,c} \left[ \frac{\psi_v(x_1)\psi_v^*(x'_1)\psi_c(x_2)\psi_c^*(x'_2)}{\omega - (E_c - E_v)} - \frac{\psi_v(x_1)\psi_v^*(x'_1)\psi_c(x_2)\psi_c^*(x'_2)}{\omega + (E_c - E_v)} \right], \]  

(2.102)

where the indices, \( v \) and \( c \), stand for the valence bands and conduction bands in semiconductors, respectively. The wave function \( \psi(x) \) is the result from LDA calculations, because the wave function of quasiparticles is almost the same as that from LDA.
In the same way, the correlation of the two-particle Green’s function can be written as

$$L(1, 2; 1', 2'; \omega) = i \sum_S \left[ \chi_S(x_1, x_1') \chi_S^*(x_2, x_2) \frac{1}{\omega - \Omega_S} - \chi_S(x_2, x_2') \chi_S^*(x_1, x_1) \frac{1}{\omega + \Omega_S} \right],$$  \hspace{1cm} (2.103)

where the electron-hole pair wave function is

$$\chi_S(x, x') = -\langle N, 0 | \psi^\dagger(x') \psi(x) | N, S \rangle.$$  \hspace{1cm} (2.104)

In semiconductors and insulators, the electronic bands are occupied or empty, so the electron-hole pair wave function can be simplified as

$$\chi_S(x, x') = \sum_v \sum_c \left[ A_{vc}^S \psi_c(x) \psi_v^\dagger(x') + B_{vc}^S \psi_v(x) \psi_c^\dagger(x') \right].$$  \hspace{1cm} (2.105)

With the help of Eq. (2.105), the Bethe-Salpeter equation (Eq. (2.101)) is transferred to the general matrix form

$$(E_c - E_v) A_{vc}^S + \sum_{v'c'} K_{vc,v'c'}^{AA}(\Omega_S) A_{v'c'}^S + \sum_{v'c'} K_{vc,v'c'}^{AB}(\Omega_S) B_{v'c'}^S = \Omega_S A_{vc}^S,$$

$$(E_c - E_v) B_{vc}^S + \sum_{v'c'} K_{vc,v'c'}^{BB}(\Omega_S) B_{v'c'}^S + \sum_{v'c'} K_{vc,v'c'}^{BA}(\Omega_S) A_{v'c'}^S = \Omega_S B_{vc}^S,$$  \hspace{1cm} (2.106)

where the interaction kernel terms are

$$K_{vc,v'c'}^{AA} = i \int d(3) d(4) d(5) d(6) \psi_v(x_4) \psi_v^\dagger(x_3) K(35; 46; \Omega_S) \psi_c^\dagger(x_6) \psi_c(x_5);$$  \hspace{1cm} (2.107)

$$K_{vc,v'c'}^{AB} = i \int d(3) d(4) d(5) d(6) \psi_v(x_4) \psi_v^\dagger(x_3) K(35; 46; \Omega_S) \psi_c^\dagger(x_6) \psi_c(x_5).$$  \hspace{1cm} (2.108)

In real calculations, it is found that the values of off-diagonal blocks, such as $K^{AB}$, are much smaller than those of the diagonal blocks. So we can make approximation to the coupled matrix equations in Eq. (2.106) by ignoring the off-diagonal blocks. Then we get the simplified Bethe-Salpeter equation form

$$(E_c - E_v) A_{vc}^S + \sum_{v'c'} K_{vc,v'c'}^{AA}(\Omega_S) A_{v'c'}^S = \Omega_S A_{vc}^S.$$  \hspace{1cm} (2.109)

This approximation is also called in Tamm-Dancoff approximation.

Now, our problem becomes to find the details of Eq. (2.109). Here we have not given the form of the interaction kernel $K$. From Eq. (2.98), the kernel can be calculated by variations of the self-energy operator to the change of the one-particle Green’s function. In
order to use the same notion as in other people’s work, we separate the Coulomb interaction from the original self-energy definition

\[ M(3, 4) = V_{\text{coul}}(3)\delta(3, 4) + \Sigma(3, 4). \] (2.110)

Generally, people also call \( \Sigma(3, 4) \) as the self-energy, we should notice their difference.

Use the above definition, the interaction kernel can be expressed as

\[ K(35; 46) = \frac{\delta[V_{\text{coul}}(3)\delta(3, 4) + \Sigma(3, 4)]}{\delta G_1(6, 5)\delta G_1(6, 5)}. \] (2.111)

Considering the second quantization of Coulomb potential in the above formulation, which gives a Green’s function term, the first term of the derivative is

\[ K^x(35; 46) = -i\delta(3, 4)\delta(5^-, 6)v(3, 6). \] (2.112)

Then we apply the GW approximation to the second term of Eq. (2.111). We have

\[ K^d(35; 46) = i\delta(3, 6)\delta(4, 5)W(3^+, 4). \] (2.113)

So the electron-hole interaction kernel is

\[ K(35; 46) = K^x(35; 46) + K^d(35; 46) = -i\delta(3, 4)\delta(5^-, 6)v(3, 6) + i\delta(3, 6)\delta(4, 5)W(3^+, 4). \] (2.114)

The term \( K^x \) is from the exchange term of the bare Coulomb potential and is called the exchange term; \( K^d \) is from the direct interaction between the electron and hole by the screened Coulomb potential and is called the direct term.

In real calculations, these terms have the forms

\[ \langle vc|K^d(\Omega_S)|v'c'\rangle = \int dxdx'\psi^*_c(x)\psi_{c'}(x)\psi_{v}(x')\psi^*_{v'}(x') \frac{i}{2\pi} \int d\omega e^{-i\omega_0}W(x, x', \omega) \]

\[ \times \left[ \frac{1}{\Omega_S - \omega - (E^Q_{c'} - E^Q_{v}) + i0^+} + \frac{1}{\Omega_S + \omega - (E^Q_{c} - E^Q_{v}) + i0^+} \right]; \] (2.115)

and

\[ \langle vc|K^x(\Omega_S)|v'c'\rangle = \int dxdx'\psi^*_c(x)\psi_{c'}(x)\psi_{v}(x, x')\psi^*_{v'}(x'). \] (2.116)

Here the same problem in the GW approximation happens again, namely, how to evaluate the dynamically screened Coulomb interaction between the electron and the hole, and
how to calculate the dynamical dielectric function. As we do in the GW approximation, we
first calculate the static dielectric function, then extend it to the dynamical case with the
general plasmon pole model. So the screened interaction can be written as

\[ W(x, x', \omega) = \sum_l W_l(x, x') \frac{\omega_l}{2} \frac{1}{\omega - \omega_l + i0^+} - \frac{1}{\omega + \omega_l - i0^+}, \]

(2.117)

where the index \( l \) stands for the plasmon mode. So \( \omega_l \) is the plasmon mode frequency, and
the summation of the above formula includes all plasmon mode \( l \).

With the help of Eq. (2.117), the matrix element of the direct term \( K^d \) is

\[ \langle \nu c | K^d(\Omega_S) | \nu' c' \rangle = -\sum_l \int dxdx' \psi_\nu^*(x) \psi_\nu(x') \psi_c(x') \psi_c^*(x) W_l(x, x') \]

\[ \times \frac{\omega_l}{2} \frac{1}{\omega_l - (\Omega_S - (E_{QP}^{c'} - E_{QP}^c)) + \frac{1}{\omega_l - (\Omega_S - (E_{QP}^{c'} - E_{QP}^c))}. \]

(2.118)

The expression of the interaction kernel depends on the value of the electron-hole binding
energy \( \Omega_S \). This means that we have to assume one value at the beginning of the calculation
of the BSE and use the self-consistence method to get the right answer. Fortunately, in
most semiconductors, the plasmon frequency is much higher than the electron-hole binding
energy, \( \omega_l \gg \Omega_S - (E_{QP}^{c'} - E_{QP}^c) \), so we can eliminate the frequency dependency, and get

\[ \langle \nu c | K^d | \nu' c' \rangle = -\int dxdx' \psi_\nu^*(x) \psi_\nu(x') \psi_c(x') \psi_c^*(x') W_l(x, x', \omega = 0). \]

(2.119)

By now, we have specified all terms in Bethe-Salpeter equation calculations. When
considering the optical absorption, we have to include the spin in the Hamiltonian. As a
result, the states are clarified as either a singlet or a triplet. Because of the selection rule,
only the singlet state is relevant in the optical absorption. So for the optical absorption
spectrum in this thesis, we only consider the singlet states.

In summary, we draw a flow chart in Fig. 2.5. All the calculations start from the
local density functional result. In the quasiparticle calculation, the Kohn-Sham eigenvalues
will be corrected by including the self-energy, but the wave function from LDA will be
treated as an accurate result and used without correction. With the wave function from
LDA, the static screened dielectric function is calculated and extended to the dynamical
case by the general plasmon pole (GPP) model. This is a key step, because not only the
Figure 2.5: The flow chart of the Green’s function method in calculating the electronic structure and optical absorption spectrum. The LDA results are used as the starting point. In real calculations of Bethe-Salpeter equation, we have to use a more denser k-grid than in GW calculations, so the calculations of the dielectric function may not overlap as in this figure.
self-energy of the quasiparticle is based on the dynamical screening, but also the electron-hole interaction depends on this screening. After this step, the self-energy is calculated by the GW approximation, and a trustable band structure is obtained. On the other hand, the dynamical screening makes it possible to compute the direct term of the electron-hole interaction kernel. Combined with the exchange from LDA results, we are ready to solve the Bethe-Salpeter equation. However, it is extremely expensive to calculate the multi-integrations for the interaction kernel. Later, we will see that in the nanowire case, we have to solve the BSE on a coarse k-grid to save time and interpolate the result to a denser grid, which improves the efficiency significantly. Finally, the Bethe-Salpeter equation is solved under the Tamm-Dancoff approximation, and the binding energy of the electron-hole pair is specified. With the binding energy, eigenfunction of BSE, and the quasiparticle energy, the optical absorption can be calculated and compared with experiments.
CHAPTER III

LATTICE VIBRATIONS IN SILICON NANOWIRES

The mass of a nucleus is much bigger than that of an electron, and the relaxation time of the
nucleus is much longer than that of the electron. So when studying the motion of the nuclei,
we assume that electrons are at their equilibrium states for a given nuclear configuration.
This is called the adiabatic approximation or the Born-Oppenheimer approximation [50].
All of the electron’s contribution is contained in the forces acting on the nucleus, which
makes it possible to treat the motion of nuclei classically. By solving dynamical equations
of nuclei, the canonical form of lattice vibrations, also called the phonon, is studied. There
are many properties of materials tightly related to the phonons, such as the electron-phonon
interaction, thermal properties of materials, and heat transportation characters, etc. [51,
52, 53, 54, 55, 56]. So, the study of lattice vibrations is a fundamental and important step
to understand the properties of solids.

The nanowire (NW) is a one-dimensional nanostructure with confinement in the other
two (perpendicular) directions [28, 29, 30]. This confinement strongly affects the lattice
vibrational modes and induces many interesting features, such as confined modes, and
frequency shifts, etc. [57, 58]. Unfortunately, among numerous experimental investigations,
few of them provide clean and definitive results about phonons in NWs. For example, it
is very hard to produce NWs with a uniform size in experiments, and the features of NWs
are strongly affected by the size of the NW because of the confinement effect. As a result,
there are always large errors and discrepancies in measurements of NWs. Another difficulty
in experiment is the oxidized shell of NWs, which prevents the observation of intrinsic
phonon modes inside NWs. Besides experiments on NWs, many theoretical studies have
been performed [59, 60, 61, 62, 63, 64, 65, 66]. Most of them use empirical models from
bulk materials. However, the quantum confinement in NWs will change electronic states
significantly, so the empirical model is not enough for an accurate study of NWs.
As introduced in the previous chapter, first-principles calculations do not need parameters in advance and have good transferability, which is important in studying NWs. First-principles calculations have been very successful in studying phonons of bulk materials [12, 14, 67]. In this chapter, we discuss first-principles calculations of phonons on silicon NWs (SiNWs). A clean physical picture of lattice vibrational modes is expected from this calculation, and we expect to provide the useful direction for future experiments.

3.1 Lattice Vibrational Modes in the Harmonic Crystal

In this section, we start with a simple derivation of lattice vibrations in a harmonic crystal [68]. The position of the ion in the crystal can be defined as \( \mathbf{r}(\mathbf{R}) = \mathbf{R} + \mathbf{u}(\mathbf{R}) \). The equilibrium position of the ion is \( \mathbf{R} \), and the small deviation from equilibrium is \( \mathbf{u} \). Here we consider only the monoatomic lattices to make the notation clear. It is easy to generalize this definition to lattices that consist of more than one atom per unit cell, and the position of an ion is \( \mathbf{r}_j(\mathbf{R}) = \mathbf{R} + \mathbf{d}_j + \mathbf{u}_j(\mathbf{R}) \), where \( j \) is the index to specify the ions in one unit cell. The total static energy is the function of the positions of ions, so it has the form:

\[
E = E(\{\mathbf{R} + \mathbf{u}(\mathbf{R})\}).
\] (3.1)

Assuming the adiabatic approximation and keeping terms up to the second order, we write down the total static energy of the crystal under a small perturbation of ion positions from the equilibrium as

\[
E = E|_{u=0} + \sum_{\mathbf{R}} \mathbf{u}(\mathbf{R}) \cdot [\nabla_u(E)]|_{u=0} + \frac{1}{2} \sum_{\mathbf{R},\mathbf{R}'} \mathbf{u}(\mathbf{R}) \cdot \mathbf{D}(\mathbf{R}, \mathbf{R}') \cdot \mathbf{u}(\mathbf{R}').
\] (3.2)

Generally, the second order term of the total energy is also called the harmonic energy, whose detail form is written as

\[
E_{\text{harmonic}} = \frac{1}{2} \sum_{\mathbf{R},\mathbf{R}'} \sum_{\mu\nu} u_\mu(\mathbf{R}) D_{\mu\nu}(\mathbf{R}, \mathbf{R}') u_\nu(\mathbf{R}').
\] (3.3)

The tensor \( D_{\mu\nu}(\mathbf{R}, \mathbf{R}') \) is named as the force matrix. It is written mathematically as

\[
D_{\mu\nu}(\mathbf{R}, \mathbf{R}') = \frac{\partial^2 E}{\partial u_\mu(\mathbf{R}) \partial u_\nu(\mathbf{R}')}|_{u=0}.
\] (3.4)
Consider the crystal has the translational invariance symmetry, the tensor $D_{\mu\nu}(\mathbf{R}, \mathbf{R}')$ in Eq. (3.4) depends only on the difference between $\mathbf{R}$ and $\mathbf{R}'$, which is

$$D_{\mu\nu}(\mathbf{R}, \mathbf{R}') = \frac{\partial^2 E}{\partial u_\mu(\mathbf{R}) \partial u_\nu(\mathbf{R}')} \bigg|_{u=0} = \frac{\partial^2 E}{\partial u_\mu(\mathbf{R} - \mathbf{R}') \partial u_\nu(0)} \bigg|_{u=0} = D_{\mu\nu}(\mathbf{R} - \mathbf{R}'). \quad (3.5)$$

The force matrix has several symmetry rules. The first one is that the force matrix is invariant under inversion

$$D_{\mu\nu}(\mathbf{R} - \mathbf{R}') = D_{\mu\nu}(\mathbf{R}' - \mathbf{R}). \quad (3.6)$$

This symmetry is based on the fact that the Bravais lattice has the inversion symmetry.

The second symmetry of the force matrix is the sum rule. If we shift the whole lattice by the same displacement amount, there is no force acting on the ions, and we have

$$\sum_{\mathbf{R}} D_{\mu\nu}(\mathbf{R}) = 0. \quad (3.7)$$

With the definition of the force matrix, we have the dynamical equation for the ions with mass $M$:

$$M\ddot{u}_\mu(\mathbf{R}) = -\sum_{\mathbf{R}'} \sum_{\nu} D_{\mu\nu}(\mathbf{R} - \mathbf{R}') u_\nu(\mathbf{R}'). \quad (3.8)$$

The solution has the form

$$u(\mathbf{R}, t) = c e^{i(\mathbf{k} \cdot \mathbf{R} - \omega t)}, \quad (3.9)$$

where $c$ is the polarization unit vector, and $\mathbf{k}$ is a vector in the reciprocal space. Because $\mathbf{R} - \mathbf{R}'$ is also a lattice vector, we take the Fourier transform of the force matrix to obtain the dynamical matrix

$$D(\mathbf{k}) = \sum_{\mathbf{R}} D(\mathbf{R}) e^{i\mathbf{k} \cdot \mathbf{R}}. \quad (3.10)$$

Then the dynamical equation can be written in the reciprocal space as

$$M\omega^2 c = D(\mathbf{k}) \cdot c. \quad (3.11)$$

We diagonalize the above equation, and obtain eigenfrequencies and eigenmodes. The eigenmode is the confined motion of ions and is called the lattice vibration. The eigenfrequency $\omega(\mathbf{k})$ is the frequency of the corresponding lattice vibration.
With the knowledge from above derivations, we introduce the quantum theory of the harmonic crystal. The eigenmodes and eigenfrequencies are thought as those of the harmonic oscillator. Assume we have an eigenmode, whose frequency is $\omega(k)$ and polarization vector is $c(k)$, the annihilation operator is

$$\hat{a}_k = \frac{1}{\sqrt{N}} \sum_R e^{-ik\cdot R} c(k) \cdot \left[ \sqrt{\frac{M \omega(k)}{2\hbar}} u(R) + i \sqrt{\frac{1}{2\hbar M \omega(k)}} P(R) \right],$$  \hspace{1cm} (3.12)

and the creation operator is

$$\hat{a}^\dagger_k = \frac{1}{\sqrt{N}} \sum_R e^{-ik\cdot R} c(k) \cdot \left[ \sqrt{\frac{M \omega(k)}{2\hbar}} u(R) - i \sqrt{\frac{1}{2\hbar M \omega(k)}} P(R) \right],$$  \hspace{1cm} (3.13)

where $u(R)$ is the real space eigenmode operator, $M$ is the mass of the ion, and $P(R)$ is the momentum operator of the eigenmode. $u(R)$ and $P(R)$ have the following commutation relations:

$$[u_\mu(R), P_\nu(R')] = i\hbar \delta_{\mu\nu} \delta_{RR'},$$  \hspace{1cm} (3.14)

and

$$[u_\mu(R), u_\nu(R')] = [P_\mu(R), P_\nu(R')] = 0.$$  \hspace{1cm} (3.15)

Given the above commutators of $u(R)$ and $P(R)$, we get the commutation rules for $\hat{a}_k$ and $\hat{a}^\dagger_k$

$$[\hat{a}_k, \hat{a}^\dagger_{k'}] = \delta_{kk'},$$  \hspace{1cm} (3.16)

$$[\hat{a}_k, \hat{a}_{k'}] = [\hat{a}^\dagger_k, \hat{a}^\dagger_{k'}] = 0.$$  \hspace{1cm} (3.17)

As we do in the harmonic oscillator calculation, the Hamiltonian of the harmonic crystal is written with the creation and annihilation operators

$$H = \sum_k \hbar \omega(k)(\hat{a}^\dagger_k \hat{a}_k + \frac{1}{2}).$$  \hspace{1cm} (3.18)

All above derivations consider the monoatomic crystal because the index is simple and clear. It is not hard to expand these conclusions to a crystal that has more than one atom per unit cell. When we calculate the phonon in practice, the equations are similar to the ones above but with more complicated indices.
From the above introduction, we see that the critical step in studying the lattice vibration is to obtain the dynamical matrix or the force matrix. In first-principles calculations, there are mainly three methods to get the dynamical matrix.

The first one is the force matrix method that directly perturbs ions from their equilibrium positions [11, 69, 70]. Then the forces that act on the ions are calculated with the Feynman-Hellman theorem. With these forces, we construct the force matrix, which is defined in Eq. (3.4). With the force matrix, the eigenmode and eigenfrequency of lattices are obtained. In practice, we will find that the force acted on ions decays rapidly with increasing separation from the perturbed ion, so the cutoff distance of the force can be specified. The cutoff distance of the force simplifies the force matrix calculation, but it is necessary to use a force cutoff distance large enough to ensure the converged result.

The second one is the frozen phonon method [12, 71]. When we are interested in a special vibrational mode, the perturbation is proportional to this mode, and the frequency of this mode is obtained from the change of the total energy. The frozen phonon method is an easy and direct way to evaluate the frequency of lattice vibrations, and the physical picture of this method is clear. The other advantage is that the frozen phonon method can be used to study the unharmonic effect, because the calculated energy change includes all of the higher order terms. But the problem of the frozen phonon method is that when we consider the long-wavelength lattice vibrations, the calculation has to include a large super cell, which greatly increases the computational difficulty.

The last method, the linear response method, does not directly calculate the force matrix but considers the linear response of the system. With the response properties, the dynamical matrix can be derived. In the linear response calculations, there is not a big difference between calculations for a k point near the zone center and other k point in the Brillouin zone. But as in its name, the linear response method considers only the linear term, and the unharmonic effect is not included. As a result, we need extra steps if the unharmonic effect needs to be calculated. Because we mainly use the linear response method later, the next section focuses on the background of this method.
3.2 Linear Response Theory

There are several good works using the linear response method, such as the dielectric matrix (DM) method by R. Resta [14], direct perturbation theory by X. Gonze and C. Lee [15], and the self-consistent scheme by S. Gironcoli and S. Baroni, et al. [16, 17]. We use the package named Plane-Wave Self-Consistent Field (PWSCF) from Baroni’s group. The key ideas of this self-consistent method are summarized in this section.

The original linear response method calculates the dielectric function (DM) that is the linear response of the electronic charge density to the external perturbation. The dielectric matrix includes so many interesting properties that are far more than what is needed for phonon calculations. The negative aspect of the DM method is that it requires inverting the dielectric matrix for phonon calculations, which is very expensive in numerical computations. Baroni’s idea is to find directly the linear response of the charge density to the external perturbation without calculating the details of the dielectric matrix. Combined with the local density approximation (LDA), the linear perturbation term of the total energy is evaluated. To save the computation time, a self-consistent scheme is used to obtain the linear term of the change of the charge density.

We assume that the total energy $E_\lambda$ depends on a parameter $\lambda$ in the external potential. The Hellmann-Feynman theorem states that the “force” associated with the variation of the external parameters $\lambda$ is given by the ground-state expectation value of the derivative of the external potential $V_\lambda$

$$\frac{\partial E_\lambda}{\partial \lambda} = \int n_\lambda(r) \frac{\partial V_\lambda(r)}{\partial \lambda} \, dr.$$ \hspace{1cm} (3.19)

The charge density $n_\lambda$ can be expanded in terms of the parameter $\lambda$ as

$$n_\lambda(r) = n_0(r) + \frac{\partial n_\lambda}{\partial \lambda} \lambda + \ldots,$$ \hspace{1cm} (3.20)

where $n_0$ is the charge density when $\lambda = 0$.

Then, the energy in Eq. (3.19) can be written as

$$E_\lambda = E_0 + \lambda \int n_0(r) \frac{\partial V_\lambda(r)}{\partial \lambda} + \frac{1}{2} \lambda^2 \int \left[ \frac{\partial n_\lambda(r)}{\partial \lambda} \frac{\partial V_\lambda(r)}{\partial \lambda} + n_0(r) \frac{\partial^2 V_\lambda(r)}{\partial \lambda^2} \right] \, dr + \ldots.$$ \hspace{1cm} (3.21)

Consider the case of lattice vibrations, the parameter $\lambda$ is regarded as the displacement of
ions from the equilibrium position; then, the harmonic term of Eq. (3.21) can be related to
the harmonic energy part in the expression of Eq. (3.2). We define the force matrix as

\[ C_{\alpha\beta}^{\mu\nu}(R - R') = C_{\mu\nu}^{\alpha\beta,\text{ion}}(R - R') + C_{\mu\nu}^{\alpha\beta,\text{elec}}(R - R'), \tag{3.22} \]

where the ion part \( C_{\mu\nu}^{\alpha\beta,\text{ion}} \) is the second derivative of the Ewald energy \[72\], which is well
known. The form of the force matrix here is a little different from our definition in Eq. (3.4).
In Eq. (3.22), we generalize the previous definition of the force matrix in Eq. (3.4) to the
situation where the crystal has more than one atom per unit cell. As a result, \( \alpha \) and \( \beta \) are
needed as indexes to specify the atoms inside a unit cell.

The electronic part is the most difficult part to be evaluated in the formula above, which
has the form

\[ C_{\mu\nu}^{\alpha\beta,\text{elec}}(R - R') = \int \left[ \frac{\partial n(r)}{\partial u_\mu^\alpha(q)} \frac{\partial V_{\text{ion}}(r)}{\partial u_\nu^\beta(q)} + n_0(r) \frac{\partial^2 V_{\text{ion}}(r)}{\partial u_\mu^\alpha(R) \partial u_\nu^\beta(R')} \right] dr, \tag{3.23} \]

where \( V_{\text{ion}} \) is the ionic potential. In our calculation, the ionic potential is the sum of
pseudopotentials acting on electrons.

In reciprocal space, the matrix in the above equation is represented as

\[ C_{\mu\nu}^{\alpha\beta,\text{elec}}(q) = \int \left[ \frac{\partial n(r)}{\partial u_\mu^\alpha(q)} \frac{\partial V_{\text{ion}}(r)}{\partial u_\nu^\beta(q)} + n_0(r) \frac{\partial^2 V_{\text{ion}}(r)}{\partial u_\mu^\alpha(q) \partial u_\nu^\beta(\mathbf{q})} \right] dr, \tag{3.24} \]

where \( u_\mu^\alpha(q) \) is the Fourier transform of the displacement of ions. In practical calculations,
we choose the displacement of ions according to the vibrational modes.

Finally, the dynamical matrix is defined as

\[ D_{\mu\nu}^{\alpha\beta}(q) = \frac{C_{\mu\nu}^{\alpha\beta}(q)}{\sqrt{M_\alpha M_\beta}}. \tag{3.25} \]

Here, the point is to find the value of \( \frac{\partial n(r)}{\partial u_\mu^\alpha(q)} \) in Eq. (3.24), which is the linear response of
the charge density to lattice perturbations. Because we only care about the linear term, it is
not necessary to solve for the whole charge density. Combined with the density functional
theory method, an efficient self-consistent way was developed by Baroni, et al \[16, 17\].

Within the single-particle picture, the external potential of the particle has the form as

\[ V_{\text{SCF}}(r) = V_{\text{ion}}(r) + e^2 \int \frac{n(r')}{|r - r'|} dr' + v_{\text{XC}}(n(r)). \tag{3.26} \]
As discussed in the introduction part of the DFT, all of the many-body effects are included in the exchange-correlation part \( v_{XC} \), which is now evaluated within the local density approximation (LDA).

With a perturbation of ions \( \Delta V_{ion} \) imposed into the potential, \( V_{SCF} \) becomes \( V_{SCF} + \Delta V_{SCF} \). Under the local density approximation (LDA), the Kohn-Sham equation has a form similar to the Schrodinger equation. The above perturbation can be treated by the same procedure as in one-particle quantum mechanics. The linear term of the change of the density can be derived as

\[
\Delta n(q + G) = \frac{4}{N\Omega} \sum_{\mathbf{k}} \sum_{c,v} \langle \psi_{v,\mathbf{k}} | e^{-i(q+G) \cdot \mathbf{r}} | \psi_{c,\mathbf{k}+q} \rangle \langle \psi_{c,\mathbf{k}+q} | \Delta V_{SCF} | \psi_{v,\mathbf{k}} \rangle, \tag{3.27}
\]

where \( \Delta n(q + G) \) is the Fourier transform of \( \Delta n(r) \), and \( G \) is in the reciprocal lattice space. \( |\psi\rangle \) and \( \varepsilon \) are the wave function and eigenvalue obtained from LDA calculations without the perturbation. The indices \( v \) and \( c \) stand for the valence and the conduction bands, respectively. The vector \( \mathbf{k} \) is sampled from a k-grid in the first Brilluoin zone. \( \Omega \) is the volume of one unit cell.

Replacing the density in Eq. (3.26) and keeping the linear term, we find

\[
\Delta V_{SCF} = \Delta V_{ion}(r) + e^2 \int \frac{\Delta n(r')}{|r - r'|} dr' + \Delta n(r) \frac{d v_{XC}}{dn} \tag{3.28}
\]

Eqs. (3.27) and (3.28) consist of a self-consistent loop. The efficiency is improved by the self-consistent procedure outlined above that contains only the charge density and the potential. As a result, solving the Kohn-Sham equation is not required, which saves computing time.

In practical calculations, we perturb the atomic positions with the lattice vibrational mode, and the self-consistency will produce the final change of the charge density. The matrix in Eq. (3.24) and the dynamical matrix can be calculated with this linear response method. In addition, the first-order perturbation in Eq. (3.27) requires the calculation of a large number of conduction bands, which is computationally inefficient. Baroni, \textit{et al.}, developed a set of Green’s function approach to project the state \( V_{SCF} |\psi_c\rangle \) into the valence band space, which is more efficient than the original form in Eq. (3.27). The details of this Green’s function procedure can be found in references [16, 17].

48
Table 3.1: Three SiNWs of different sizes. Listed are the number of hydrogen atoms and silicon atoms per unit cell. The right column shows diameters of these three wires.

<table>
<thead>
<tr>
<th>Index</th>
<th>Number of H</th>
<th>Number of Si</th>
<th>Diameter (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>16</td>
<td>1.2</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>42</td>
<td>1.8</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>76</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Figure 3.1: Ball-and-stick model of the hydrogen-passivated silicon NW under investigation: side view (left) and top view (right). The large atoms are silicon and small atoms are hydrogen. There are 16 silicon and 12 hydrogen atoms in each unit cell. Five unit cells are shown in the left figure.

3.3 Structure of SiNWs

In this work, we study hydrogen-passivated SiNWs oriented along the [110] direction using the density functional theory (DFT) in the local density approximation (LDA). We consider three different sized NWs with diameters of 1.2 nm, 1.8 nm, and 2.2 nm, respectively. Table 3.1 lists the detailed structure of these three wires.

We construct SiNWs from the bulk structure. The surface dangling bonds are passivated by hydrogen atoms. Because of the small atomic mass and simple electronic structure of the hydrogen atom, this surface passivation should hardly affect the electronic properties of NWs. Particularly, the Si-H bond is much stronger than the Si-Si bond, so we expect
that the lattice vibrational modes related to hydrogen atoms motion should have much higher frequencies. In other words, we expect that the lattice vibration from surface hydrogen atoms is easily identified from those intrinsic lattice vibrations of NWs in the energy spectrum.

The original structure of the smallest wire is shown in Fig. 3.1. In order to make our following description clear, we set the cartesian coordinate in Fig. 3.1. The wire’s growing direction, the [110] direction, is the z axis. The x axis is along the [001] direction, and the y axis is along the [1 \bar{1} 0] direction.

The computational package is PWSCF [74], a software using the linear-response phonon calculation with normal-conserving pseudopotentials. In fact, PWSCF can also handle ultrasoft pseudopotentials, which require much smaller plane-wave cutoff energy. But we will calculate the Raman scattering spectrum of NWs later, and PWSCF can only accept the normal conserving pseudopoten
tial when calculating the Raman activity. To make the whole work consistent, we choose the normal-conserving pseudopotential from the beginning.

In first-principles calculations, we have to choose computational parameters to ensure
that our results are converged. Because PWSCF uses the plane-wave method, we arrange SiNWs periodically and keep the intermediate vacuum large enough to avoid any interaction between wires. In our calculation, the vacuum distance between wires is about 0.5 nm. The cutoff energy is set at 18 Ry to satisfy the total energy convergency. For the Brillouin zone integration, we use a 1x1x8 k-grid in the Monkhorst-Pack special point scheme. The total energy convergence with respect to the cutoff plane-wave energy and the k-grid is tested, and the result is shown in Fig. 3.3. The total energy is not completely converged at 18 Ry, but we find that the linear response of the dielectric tensor is not sensitive to the cutoff energy. Therefore, the energy cutoff of 18 Ry is acceptable for this calculation. The 1x1x4 k-grid is sufficient for the density functional theory calculation. But later when we calculate the Raman spectrum, it is necessary to use more k-points in that step, so we choose 1x1x8 k-grid from the beginning. After fixing the energy cutoff and k-grid, we set the criterion for the potential self-consistency in the Kohn-Sham equation as $10^{-8}$ Ry, and 0.05 eV/Å is the force criterion for the relaxation process.

In first-principles calculations of lattice vibrations, the stability of the structure is important. The structure giving the global minimum of the total energy is difficult to reach, because the relaxation process is often trapped by the local minima of the total energy. In our relaxation process, if we completely depend on the PWSCF package, the final energy spectrum of lattice vibrations has several negative-frequency modes, which means that our structure is not at the globally minimal energy position. By analyzing the negative frequency (unstable) modes, we find that those modes are tightly related to the motion of hydrogen atoms on the [001] surface. So the instability should be removed by finding the minimal energy position for those hydrogen atoms. Thanks to the previous work [73], it has been found that the hydrogen atoms on the [001] surface will relax from the ideal positions. There are several kinds of relaxations as shown in Fig. 3.2. By comparing them in our NW structure, we find the second type of the relaxation gives us the minimal total energy. Using this configuration, we get rid of the negative-frequency modes. After this relaxation procedure, the final structure of the smallest SiNW is shown in Fig. 3.1. In the larger wires, similar results are obtained, and the hydrogen atoms on the [001] surface are rearranged as
Figure 3.3: Convergence of the total energy with respect to the cutoff energy and k-grid. The cutoff energy is set to be 18 Ry as marked in the left figure. The k-grid is set to be 1x1x8 as marked in the right figure.

3.4 The Density of Vibrational Modes

The density of vibrational modes is detected by inelastic neutron scattering experiments. Therefore, the theoretical calculation of the density of vibrational modes provides an easy way to explain the experimental results. In addition, it is expected that the quantum confinement effect will induce a change in the lattice vibrational modes of NWs, which should be reflected in the change of the density of vibrational modes. Beyond these considerations, we hope that our calculated result is helpful to the Raman scattering experiment because the Raman scattering is a powerful experimental tool to detect the structure of materials. Since the momentum of the photon involved in the Raman scattering is much smaller than the size of the first Brillouin zone, we begin with the calculation of the density of vibrational modes at the Γ point and expect to see the size dependence of it.

With the relaxed structure, we calculate the dynamical matrix at the Γ point. The eigenvalues and eigenvectors of lattice vibrations are obtained by diagonalizing this matrix. Because of the big difference between the mass of hydrogen and silicon atoms, the vibrational modes are classified into two types. One consists of the motions of silicon atoms mainly,
Figure 3.4: Density of state of phonons at the Γ point for a silicon [110] wire with diameter d = 1.2, 1.8, and 2.2 nm, respectively. The Gaussian broadening with a width of 0.05 THz is used in this figure. The corresponding quantity for the silicon bulk is shown at the top of the figure. The arrows mark positions of confined modes below 8 THz. “L” marks the position of LO mode, and “T” marks the position of TO modes. In the silicon bulk, these modes become degenerated. “A, B, C, and D” are the confined modes shown in Fig. 3.7.

called NW modes; the other consists of the motions of hydrogen atoms mainly, called surface modes. In our calculation, we find that there is an energy gap between these two types of modes. All the modes below 15.5 THz are NW modes, while all the surface modes have higher frequencies than 15.5 THz. This is very helpful to our later analysis, because we are interested in the quantum confinement effect instead of the surface effect. In the following calculation, we focus on the NW modes below 15.5 THz. As explained above, we are interested in the density of vibrational modes at the Γ point, which is shown in Fig. 3.4. The density in the silicon bulk projected to Γ along the [110] direction is also plotted in order to illustrate the quantum confinement.

The main structure of the density of vibrational modes in SiNWs is similar to that of bulk silicon, for example, the high peaks near 13-15 THz and several humps in the 2-5 THz range. However, there are two special features found in SiNWs.

Compared with bulk silicon, the high frequency peaks in the density of vibrational modes become wider in NWs. One effect of the quantum confinement in NWs is to fold more bulk
silicon phonon modes back to the zone center. Obviously the smaller the diameter of the wire, the wider the peaks.

In the high frequency range of the spectrum, the peaks consist of optical modes, which exhibit slight shifts to the low frequency as the diameter of the wire decreases. In fact, this feature has been found in experiment [60, 75, 76, 77, 78]. The shift is explained as a result induced by the experimental detection, such as the conservation of the momentum in the interaction between the phonon and photon. However, our results show that the shift is an intrinsic property of a NW. In order to characterize this frequency shift, it is important to study the details of optical modes. Considering that there are three optical modes in bulk silicon, we focus on these three optical modes to see the evolution from bulk silicon to SiNWs.

3.5 The Frequency Shift of Optical Modes and its Size Dependence

After examining hundreds of vibrational modes, we identify the three optical modes in SiNWs that originate from bulk silicon. We mark them in Fig. 3.4. The three modes can be classified into two categories: one has a large frequency shift, which we mark as the “L” mode; the other one includes two almost degenerated modes and has a smaller frequency shift, which we mark as the “T” mode. The different frequency shift between optical modes in Fig. 3.4 is interesting. Although the boundary condition is same for all optical modes, the same confinement induces different frequency shifts for different optical modes. Obviously, a careful study of atomic motions in these optical modes will help us understand this effect.

In the silicon bulk, these three modes are degenerated at the Γ point. The degeneracy is lifted in SiNWs because of the change of symmetry. In Fig. 3.5, we plot the corresponding atomic motions of these three optical modes. In this figure, there are three kinds of lattice vibrations labeled with type $T_1$, $T_2$, and $L$, which correspond to the three optical modes in bulk silicon. Because the figure is two-dimensional, the atomic motions along the z direction in the $L$ mode are projected to the y direction instead. $T_1$ and $T_2$ are the transverse modes, and $L$ is the longitudinal mode. From the top to the bottom, the size of the wire increases,
and the atomic motions of these modes are approaching the bulk limit.

Before discussing the size dependence of optical modes in SiNWs, we first comment on the boundary condition for the optical modes of NWs. A long standing question is which boundary condition should be chosen in calculating the optical modes of NWs, a free boundary condition or a clumped boundary condition [66]. Because the choice of the boundary condition determines the form of the vibrational modes, a considerable amount work has been done to answer this question, and there is no final conclusion up to this point. From Fig. 3.5, we find that as the size of the NW increases, atoms inside the NW move more intensively than those on the surface. So at least for the three optical modes that we study here, the clamped boundary condition seems more appropriate.
Figure 3.6: Red shift of the three optical modes in silicon NWs. The shifted frequency refers to the bulk optical mode frequency. $T_1$ and $T_2$ modes are the type I mode in Fig. 3.5, and L mode is the type II mode in Fig. 3.5. We use a $\frac{1}{d^\alpha}$ function to fit the points. The result shows that $\alpha$ is always around the 1.3-1.5 range. (The detail value is listed in Table 3.2)

The size dependence of these optical modes is of central interest, and it is plotted in Fig. 3.6. The fitting result shows that the frequency red shift, $\Delta \omega$, is related to the diameter of the wires, $d$, in such a way that

$$\Delta \omega \propto \frac{1}{d^\alpha}, \quad (3.29)$$

where the value of $\alpha$ is around 1.3-1.5, as listed in Table 3.2.

The frequency red shift of optical modes in SiNWs arises from the loss of Si-Si bonds on the surface. Because the ratio of the number of lost bonds to the number of total bonds scales as $1/d$, we expect the amount of the red shift to be proportional to $1/d$. This is close to the fitting results shown in Table 3.2.

Furthermore, we find that the optical mode (LO) along the z direction, namely the wire growing direction, has a much larger red shift than the other two optical modes (TO). Clearly, the bond-bending force is much smaller than the bond-stretching force. The longitudinal optical mode (LO) loses more stretch bonds, resulting in a larger red shift than the other two optical modes.
Table 3.2: The fitting results of the frequency red shift in silicon NWs. The shift is fitted with the formula as $\delta\omega = \frac{b}{r^{\alpha}}$.

<table>
<thead>
<tr>
<th>Index</th>
<th>value of b</th>
<th>error of b</th>
<th>value of $\alpha$</th>
<th>error of $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_1$</td>
<td>0.63</td>
<td>0.01</td>
<td>1.34</td>
<td>0.05</td>
</tr>
<tr>
<td>$T_2$</td>
<td>0.85</td>
<td>0.06</td>
<td>1.53</td>
<td>0.17</td>
</tr>
<tr>
<td>$L$</td>
<td>1.66</td>
<td>0.07</td>
<td>1.44</td>
<td>0.10</td>
</tr>
</tbody>
</table>

3.6 Confined Modes and Their Size Dependence

The confined vibrational modes are unique results generated by quantum confinement in nanostructures [79]. They are sensitive to the size and the boundary condition of NWs. Therefore, by studying the confined modes, we learn a lot of information about the structure of NWs and provide some useful ideas for the future experiment based on our calculational results.

In Fig. 3.7, we plot a few representative confined modes of SiNWs. These confined modes have special symmetry, and are different from lattice vibrational modes in bulk silicon. We show two kinds of confined modes here. One is the radial breathing mode (RBM), where the atomic movement in this kind of mode is only along the radial direction. Because of the boundary condition, there is an amplitude modulation along the radial direction. In Fig. 3.7(B), we see the clean nodal structure because of this modulation. Another kind of confined mode is shown in Fig. 3.7(C) and (D). Fig. 3.7(C) has only angular components, and is called the rotational mode, which is a special mode in NWs. Fig. 3.7(D) has not only the radial component but also the angular component, and each of these two components has its own space modulation. So the mode in Fig. 3.7(D) looks more complicated than the radial breathing mode.

Because the geometry of SiNWs under consideration is very close to that of a cylinder, an elastic media model under the cylindrical boundary condition is used to describe these confined modes. We do not consider the variation along the z direction, because the wire...
Figure 3.7: Confined modes in SiNWs. Figures A and B are radial breathing modes, with a different quantum number. Figure C is a confined mode that has only $\hat{\theta}$ component. Figure D is another kind confined mode that has a $\frac{\pi}{2}$ phase difference between $\hat{\theta}$ and $\hat{r}$ directions.

has an infinite length along the z direction. Therefore, the system is described by a two dimensional wave equation

$$\frac{\omega^2}{c^2} \tilde{\phi}(r, \theta) = k^2 \tilde{\phi}(r, \theta) = -\nabla^2 \tilde{\phi}(r, \theta), \quad (3.30)$$

where $\omega$ is the frequency, $c$ is the sound velocity, and $k$ is the wave number of the elastic wave.

Assume a general form of the wave is

$$\tilde{\phi}(r, \theta) = R(r, \theta)\hat{r} + \Theta(r, \theta)\hat{\theta}. \quad (3.31)$$

We obtain a set of equations for the radial and tangential directions, respectively,

$$-k^2 R(r, \theta) = \frac{\partial^2}{\partial r^2} R + \frac{1}{r} \frac{\partial}{\partial r} R + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} R - \frac{1}{r^2} R - \frac{2}{r^2} \frac{\partial}{\partial \theta} \Theta; \quad (3.32)$$

$$-k^2 \Theta(r, \theta) = \frac{\partial^2}{\partial r^2} \Theta + \frac{1}{r} \frac{\partial}{\partial r} \Theta + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} \Theta - \frac{1}{r^2} \Theta + \frac{1}{r^2} \frac{\partial}{\partial \theta} R. \quad (3.33)$$

58
The angular component of the radial breathing mode (RBM) is zero, $\Theta(r, \theta) = 0$. Let $s = kr$, and we reach the standard first-order Bessel equation in the radial direction

$$\frac{d^2 R}{ds^2} + \frac{1}{s} \frac{dR}{ds} + (1 - \frac{1}{s^2})R = 0.$$ (3.34)

The solution is the first-order Bessel function $J_1(s)$, which is plotted in Fig. 3.8. Because a free boundary condition is expected, the solution is chosen such that the boundary is at the place where the curvature is zero. The first zero curvature position gives the mode in Fig. 3.7(A). As the energy increases, the boundary moves to the second zero curvature position, which gives rise to the mode in Fig. 3.7(B).

For the confined mode in Fig. 3.7(D), it is noted that there is a $\pi/2$ phase difference between $R(r, \theta)$ and $\Theta(r, \theta)$. Let $R(r, \theta) = A(r) \cos(\theta)$ and $\Theta(r, \theta) = A(r) \sin(\theta)$, the new equation for $A(r)$ is still a Bessel equation. So the motions of vibrational modes in Fig. 3.7(D) along the radial and angular directions, respectively, are still Bessel functions.

Since confined modes are generated by confinement conditions, as the size of the SiNW
Figure 3.9: Blue shift of the radial breathing modes (RBM) as the diameter of the wire decreases. We list the frequency of the RBM in three NWs, and use a $\frac{1}{r^a}$ function to fit the points. The result shows that $a$ is around 1.22.

increases, the confined mode will decay. The equations above also show how the confined modes decay with the increasing diameter. For example, we focus on the radial breathing mode (RBM). From the characteristics of the Bessel function, we see the boundary is fixed at $s = 1.8$ with the first arrow marked in Fig. 3.8. That means $kr = \frac{\omega}{c}r = 1.8$, so the relation between the frequencies of breathing modes and the size of the wire is

$$\omega_{\text{breathing}} \propto \frac{1}{r}.$$  \hspace{1cm} (3.35)

In Fig. 3.9, we show the frequencies of the RBM as a function of the diameter of SiNWs. The solid points are the results of first-principles calculations, and they are fitted by a dashed line in the form of $\omega_{\text{breathing}} \propto \frac{1}{r^a}$, where $a$ is 1.2, close to 1.0. The size dependence of the RBM is important in practical applications. If the frequency of the RBM is measured, we can easily estimate the size of NWs because of this size dependence. In the next section, we will find that the RBM can be investigated using the Raman scattering spectrum.
Figure 3.10: Raman scattering. A photon with energy $\omega_L$ and momentum $\mathbf{q}_L$ is scattered by phonon with energy $\omega_k$ and momentum $\mathbf{k}$ to the photon with $\omega_s$ and momentum $\mathbf{q}_s$. Figure (a) shows phonon absorption, and the photon energy is shifted to a higher frequency, which is called the anti-Stokes process. Figure (b) shows phonon emission, and the photon energy is shifted to a lower frequency, which is called the Stokes process.

3.7 Raman Activities of SiNWs

The Raman scattering is the most widely used method to detect vibrational modes in solids [80]. When light goes through the solids, there will be scattering processes between the photon and the phonon, which are very useful to detect the lattice vibration (phonon) [81]. There are two kinds of scattering processes based on how the phonon is involved. The first one is called the Brillouin scattering if the involved phonon is an acoustic one; the other is called the Raman scattering if the involved phonon is an optical one.

We classify these scattering processes according to other criterion. If the phonon is emitted in the scattering process, a spectral analysis of the scattered light will reveal spectral satellite lines below the elastic (Rayleigh) scattering peak at the incident frequency. Such lines are called Stokes lines, and this scattering process is the Stokes process. If a phonon is absorbed in the scattering process, it is also possible to observe satellite lines at frequencies above the scattered frequency as the phonon energy is added to the incident photon energy. These lines, generally weaker, are called anti-Stokes lines, and this scattering process is the anti-Stokes process. These two kinds of scattering are shown in Fig. 3.10.
The conservation of energy and momentum holds in these scatterings

$$\hbar \omega_s = \hbar \omega_L \pm \hbar \omega_k,$$

(3.36)

and

$$\hbar q_s = \hbar q_L \pm \hbar k,$$

(3.37)

where $\omega_L$ and $q_L$ are the frequency and wave number of the incoming photon, $\omega_s$ and $q_s$ are those of the outgoing photon, and $\omega_k$ and $k$ are those of the involved phonon.

Here we consider a simple case of momentum conservation. If the difference between the phonon momentum is a reciprocal lattice vector, the corresponding phonon has the same reduced momentum in the first Brillouin zone. So the momentum conservation in Eq. (3.37) can have an extra term on the left side, which is a reciprocal lattice vector (the Umklapp term). However, in the Raman scattering process, the extra reciprocal lattice vector is neglected because the momentum of the photon is very small and approximated as $q_L \approx q_s = 0$. From Eq. (3.37), the momentum of the phonon has to be zero. That is why we only need the knowledge of phonons at the $\Gamma$ point of the Brillouin zone under the first order Raman scattering process. Finally, it is necessary to point out that it is not always right to neglect the Umklap process that is important in many other phenomena, such as the phonon-phonon interaction or electron-phonon interaction.

In first-principles calculations of the Raman scattering intensity, the fundamental problem is to calculate the dielectric constant (tensor) and find the response of the NWs to external electric field. Because the frequency of the photon involved is much higher than that of the phonon, we assume that the dielectric constant used in the Raman scattering calculation has the form of the high frequency limit, where the frequency is taken as $\omega = \infty$.

In the linear response calculation with the plane-wave method, the external field is ill-defined, because a uniform electric field destroys the periodical boundary condition. A method is developed, which replaces the original problem with a boundary insensitive form. All of the detailed derivation can be found in reference [17]. After all, the final dielectric tensor is written as

$$\epsilon_{\alpha\beta}^{\infty} = \delta_{\alpha\beta} + \frac{16\pi e^2}{N\Omega} \sum_k \sum_v \langle \phi_{v,k}^\alpha \frac{\partial \psi_{v,k}}{\partial E_{\beta}} \rangle.$$  

(3.38)
where $E_\beta$ is the external electric field, and the auxiliary function is defined with the commutator between the Hamiltonian and position vector as

$$|\phi_{\nu,k}^\alpha\rangle = \sum_c |\psi_{c,k}\rangle \langle \psi_{c,k} | [H, r_\alpha] |\psi_{\nu,k}\rangle \frac{\epsilon_{c,k} - \epsilon_{\nu,k}}{\epsilon_{c,k}}.$$  \hspace{1cm} (3.39)

In Eq. (3.38), the value of dielectric tensor strongly depends on the choice of the k-grid, and the convergence of the dielectric tensor is slower than the total energy in general LDA calculations with increasing k points. So first we have to test the dielectric tensor convergence with respect to the k-grid choice. The results are shown in Fig. 3.11.

Comparing Fig. 3.3 and Fig. 3.11, we find the different convergence behavior between the total energy and the dielectric tensor. In Fig. 3.3, the total energy convergence is sensitive to the plane wave cutoff energy, and not sensitive to the choice of the k-grid. However, in Fig. 3.11, the dielectric tensor is sensitive to the choice of the k-grid instead of the plane wave cutoff energy. As a balance, we choose the cutoff energy to be 18 Ry, and the k-grid to be 1x1x8. Although they are not perfectly converged, the error is acceptable without paying the high price of expensive computations.

Once we know the dielectric tensor of NWs, the formula for Raman activities is derived...
as the following. An incoming photon of frequency $\omega_L$ and polarization $e_L$ is scattered into an outgoing photon of frequency $\omega_s$ and polarization $e_s$, either creating (stokes process) or annihilating (anti-Stokes process) a photon of frequency $\omega_k$. We consider here a non-resonant Raman scattering process, which can be described in the Placzek approximation [80, 82]. The power cross section of the Stokes process involving a phonon of eigenmode $k$ reads (in esu units):

$$\frac{d\sigma}{d\Omega} = \frac{\omega_s^4 V}{c^4} I_{\text{raman}}^k \frac{\hbar}{2\omega_k} (n_k + 1).$$

(3.40)

Where $V$ is the volume of the scattering sample, $c$ is the speed of light, and $n_k$ is the Boson factor,

$$n_k = \frac{1}{\exp\left(\frac{\hbar \omega_k}{k_B T}\right) - 1}. \quad (3.41)$$

Here $I_{\text{raman}}^k$ is referred to as the Raman activity associated to the vibrational eigenmode $k$. We have

$$I_{\text{raman}}^k = |e_s \cdot \alpha^k \cdot e_L|, \quad (3.42)$$

and the Raman tensor $\alpha^k$ is defined as

$$\alpha_{ij}^k = \sqrt{\frac{2}{\Omega}} \sum_{al} \frac{\partial \chi_{ij}}{\partial r_{al}} u_{al}^k, \quad (3.43)$$

where $a$ is the index to specify the atom in one unit cell, and $l$ is the index of coordinates $x$, $y$, and $z$. $\chi_{ij}$ is the electric polarizability tensor

$$\chi_{ij} = \frac{\epsilon_{ij} - \delta_{ij}}{4\pi}. \quad (3.44)$$

Before we go to the next step, a symmetry analysis of the Raman tensor in Eq. (3.43) is in order. The point group of SiNWs along the [110] direction is 2/m. There are four symmetry operations including inversion symmetry in this group. Any phonon mode with the inversion symmetry is Raman scattering active. As a result, half of vibrational modes fall into this category. Furthermore, the Raman tensors assume a well-defined form using the notation of Loudon, which has two kinds of form $A_g$ and $B_g$:

$$\begin{pmatrix} 0 & a & b \\ a & 0 & 0 \\ b & 0 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} c & 0 & 0 \\ 0 & d & f \\ 0 & f & e \end{pmatrix}$$
Figure 3.12: Raman activity of the smallest SiNW (no space average). We fix the incoming polarization of the light in the first three graphs, and set the outgoing light direction in x, y, and z, respectively. The index of the direction is specified; (001) means the z direction, and (100) means the x direction. For example, the last figure shows the result for incoming light polarization in the x direction, and outgoing polarization in the y direction. In addition, the polarization along the z direction is much stronger than other configurations, which is preferred in Raman scattering measurement.
In our calculations of the Raman tensor, PWSCF does not impose the symmetry into the Raman tensor automatically. But we do find that the calculated Raman tensors exhibit the same form as $A_g$ and $B_g$ shown above. This match is a proof of the validity of our first-principles calculations.

The Raman activity defined in Eq. (3.42) and Eq. (3.43) requires the derivatives of the polarizability tensor with respect to the atomic displacements. In our calculation, we try to perturb all the atoms in a unit cell according to the vibrational mode. Since we only consider the linear response term, this perturbation is equivalent to the general method in Eq. (3.43). The advantage of our method is that the Raman tensor obtained through this way maintains the perfect symmetry form. In practice, we perturb the atoms with a small displacement (about 2 percent of the silicon bulk lattice constant) to maintain the linearity.

In addition, from the formula of the Raman activity, Eq. (3.42), the intensity of the

\[ \text{Figure 3.13: Raman activity of the smallest SiNW averaged over the orientation but without including the depolarization effect. The breathing mode is highly active in the smallest wire, which can be used to estimate the wire size.} \]
Raman peak depends on the polarization directions of the incoming and outgoing light. With different combinations of the incoming and outgoing polarization directions, one can detect different phonon modes, which provides us a practical way to pick up interesting modes from the Raman spectrum. In order to see this effect, we calculate the Raman activity with combinations of different incoming and outgoing polarization directions. In Fig. 3.12, we see that the Raman activities of these vibrational modes are very sensitive to the polarization direction of the light. If the NW is arranged well along one direction, the selectivity of the mode provides a good way to distinguish some special modes from others. For example, if we are interested in the breathing mode around 5.3 THz, choosing incoming and outcoming polarization to be along the z direction is a good way to detect it. With this combination, the Raman activity of this breathing mode is the strongest.

In experiments, NWs may be oriented randomly. The above expression has to be averaged properly over the orientation. The averaged result depends on the relative orientations of the direction and polarization of the incident and scattered beams. In most experiments, a plane-polarized incident laser beam is used. Furthermore, the direction of the incident beam, the polarization direction of this beam, and the direction of the observation are perpendicular to each other. Under these circumstances,

$$I_{\text{raman}}^k = 5(\alpha_{xx}^k + \alpha_{yy}^k + \alpha_{zz}^k)^2 + \frac{7}{4}[\frac{1}{4}(\alpha_{xx}^k - \alpha_{yy}^k)^2 + (\alpha_{xx}^k - \alpha_{zz}^k)^2 + (\alpha_{yy}^k - \alpha_{zz}^k)^2 + 6(\alpha_{xy}^k)^2 + (\alpha_{xz}^k)^2 + (\alpha_{yz}^k)^2]]$$

(3.45)

In Fig. 3.13, we show the orientation-averaged Raman activity for the smallest wire. There are many low frequency modes, some of them are confined modes. For example, there is a high peak around 5.3 THz, which is a radial breathing mode (RBM). These confined modes contain much useful information about SiNWs, such as the size of the wire, the boundary condition, etc. In the high frequency range, a lot of optical modes show up resulting from the folding of the Brillouin zone. They give rise to a group of high peaks at the high frequency end. As the diameter of the SiNW increases, the peaks evolve to form the Raman spectrum of bulk silicon.

In nanostructures, the response to the external electric field has a strong directional dependence. Researchers find that the induced electric field perpendicular to the wire axial
Figure 3.14: A typical Raman spectrum from experiment (reference [83]). The angle is between the incoming light polarization direction and the wire growing direction. It is clear that the Raman scattering peaks disappear when the above two directions are perpendicular to each other. This is called the depolarization effect. Direction will balance most of the external electric field, and results in a much weaker total electric field. For example, in the Raman scattering experiment, the amplitude of peaks related to the Stokes and anti-Stokes processes strongly depends on the incoming light polarization direction. In Fig. 3.14, the Raman scattering peaks disappear when the polarization direction of the incoming light is perpendicular to the nanotube growing direction. This effect is a reflection of the many-body effect of the electronic structure in NWs and is called the depolarization effect. In chapter IV, the depolarization effect will be discussed in details with the help of the two-body Green’s function. Here, we just give the corrected Raman spectrum by including the depolarization effect.

The Raman spectrum including the depolarization effect is shown in Fig. 3.15. We have
eliminated the contribution from the polarization along the vertical direction of SiNWs in Eq. (3.45). Fortunately, the RBM is still strongly active even when the depolarization effect is included. Because we have found the size dependence of the RBM in Eq. (3.35), it serves as a fast way to estimate the size of NWs by measuring the frequency of the RBM in their Raman spectrum.

There are some experimental results of the Raman spectrum of the RBM in nanotubes and nanoparticles, but few results of the Raman scattering are available for NWs. One of the reasons is that the difficulty in manufacturing NWs with a uniform size distribution makes it hard to find clean confined modes based on those spectra. In addition, most of experimental works are study the oxidized silicon NWs, which are covered by a hard
silicon dioxide shell. This hard shell changes the boundary condition and modifies the
Raman activity of these confined modes significantly, which prevents us from observing
the intrinsic confined vibrational modes. However, more high-quality NW samples will be
produced in the future, and the detection of the RBM in NWs will be accomplished some
day.

3.8 Summary of this Chapter

In this Chapter, we begin with the concepts of lattice vibration and dynamical matrix. Two
fundamental methods in the first-principles calculations of lattice dynamics are introduced;
one is the direct force matrix method; the other one is the linear response method. Because
most of our calculations are performed with PWSCF, a package based on the linear response
method, we focus on the introduction of the linear response method. Compared with the
original inverse dielectric matrix (DM) method, the efficiency of PWSCF is significantly
improved by the self-consistent scheme.

The first step of the first-principles calculation of SiNWs is to obtain the stable structure.
In light of previous work on the surface of crystal silicon, we reach the globally stable
structure, and the inner structure of SiNWs keeps the diamond structure, which makes it
easy for us to compare our calculated result with that of bulk silicon.

The lattice vibrational modes at the zone center are calculated, and the density of
phonon modes is shown. The clear evolution from the smallest SiNW to the silicon bulk is
shown from the plots of densities of modes. The quantum confinement induces two effects
in the lattice vibrational modes of SiNWs: One is the small red shift of the optical modes;
the other is the creation of confined modes because of the cylindrical boundary.

The red shift of the optical modes has the size dependence of $\frac{1}{r}$. By studying the detail
motion of these modes, we suggest that this effect is related to the loss of Si-Si covalent
bonds on the surface. On the other hand, the confined modes are studied too. The radial
breathing mode (RBM) is found. The nodal structure and angular modulation are found
in these confined modes, and we use an elastic model to explain these confined modes
successfully.
As a powerful experimental measurement, the Raman scattering is widely used in the study of lattice vibrational modes. The first-order Raman spectrum is obtained by the linear term in the change of dielectric tensor due to the perturbation. Including the direction average and the depolarization effect, the Raman activities of vibrational modes are calculated. Combined with the size-dependence of RBM, a easy way to estimate the size of NWs is proposed.
CHAPTER IV

EXCITONIC EFFECT IN THE OPTICAL ABSORPTION OF SILICON NANOWIRES

For more than two decades, first-principles calculations of materials have had fruitful developments. The density functional theory (DFT) is proved to be a trustworthy method to understand the electronic structure of materials (especially the ground state of electrons). However, the DFT result does not describe excited states well because it generally underestimates the electronic band gap in semiconductors [84]. Considerable effort has been devoted to solving this band gap problem. One of the most successful methods is the so-called GW approximation for the self-energy in terms of a product of the single-particle Green’s function $G$ and the dynamical screened interaction $W$. By including the GW approximation [19, 20, 85], reasonably good band gaps can be obtained in most calculations.

Furthermore, when an electron is promoted to an excited state, a corresponding hole is created. The two-body effect associated with the electron-hole pair or excitons, introduces an important correction to the optical absorption spectra [45, 46]. In order to include the electron-hole interaction in the \textit{ab initio} calculation, the two-body Green’s function $G_2$ needs to be evaluated by solving the equation for the motion of $G_2$ (known as Bethe-Salpeter equation). The optical transition can be corrected by including the electron-hole interaction [26]. The above method is very successful and gives us accurate optical spectra in many cases [48].

As a novel material for building future electronic devices, the optical properties of nanostructures attract a lot of attention. Because of the special geometry and quantum confinement effect, it is found in experiments and theoretical calculations that the exciton has an extremely large binding energy in nanotubes and quantum dots [86, 87, 88, 89]. However, most of the studies are based on empirical models, in which parameters are obtained from
4.1 The Structure of SiNWs and Calculational Parameters

The structure of the SiNW is shown in Fig. 5.6, which is the same structure shown in Fig. 3.1 in the third chapter. The diameter of the wire is close to 1.2 nm with 16 silicon atoms and 12 hydrogen atoms per unit cell. The surface dangling bonds are passivated by hydrogen atoms. In the later computations, we do not find any surface state close to the band gap.
Figure 4.2: The Coulomb truncation between SiNWs. For simplicity, we use an orthogonal lattice as an example. The truncation radius $r_c$ is half of the lattice constant $d$ of the supercell.

We use DFT with the local density approximation (LDA) to obtain the ground electronic state of SiNWs and choose the plane-wave basis to expand the single-particle wave functions. SiNWs are arranged into a hexagonal lattice to save the number of plane waves. The Truller-Martin pseudopotentials are used. In order to satisfy the convergence of the total energy, the energy cutoff is set to be 12 Ry, while the Monkhorst k-grid scheme of 1x1x4 [95] is used in the Brillouin zone integration.

When we include the excitonic effect, the electron-hole pair creates a dipole field that induces an interaction between different NWs. Even in the same wire, the periodical boundary condition is artificial for the exciton state. To avoid the above image effect, we have to truncate the Coulomb field by cutoff radiuses as

$$v_c(r) = \frac{1}{r} \theta(r - r_c) \theta(|z| - z_c).$$

Since this truncation cannot affect the electron-hole interactions inside the same wire, we choose the truncation distance $r_c$ to be a little bit larger than the diameter of the wire $d$. As a result, we have to set the lattice constant of the hexagonal supercell to be more than
Figure 4.3: The inverse dielectric function $\varepsilon^{-1}_{\infty}(q_z)$. This function sharply decreases when $q_z$ is small, which requires denser k-grid to be described.

double that of the diameter of SiNWs. This truncation scheme is shown in Fig. 4.2, which introduces a large vacuum between NWs and increases the number of plane waves needed in the calculation. In the future, it may be advantageous to use localized orbits to simulate isolated NWs, which saves the computations cost for the vacuum. For the truncation along the wire-growing direction, a large cutoff length is necessary because some excitonic states are significantly more extended along the wire direction. So the value of $z_c$ is set to be around 420 atomic units, which includes more than one hundred unit cells.

Because of the Coulomb truncation, the behavior of the dielectric constant is different from the untruncated case. As we show in Fig. 4.3, the homogenous part of the static dielectric constant has a very sharp slope in the long-wave limit. To describe this feature, a dense k-point grid is needed. In LDA calculations, a 1x1x4 k-point grid is enough. However, in the quasiparticle case, we cannot reach the converging result until the grid reaches 1x1x32. When we solve the Bethe-Salpeter equation (BSE), a much more denser k-point grid is needed. The 1x1x192 grid is used in the BSE and optical absorption calculations.
Figure 4.4: The quasiparticle energy compared with the Kohn-Sham energy. The GW correction is fairly linear with the LDA results. We fit the correction to valence bands and conduction bands separately and extend the shift to the whole band structure.

4.2 LDA and GW Band Gaps

The first step of our calculation is to obtain an accurate band structure for SiNWs. The quasiparticle energy versus the LDA Kohn-Sham energy is plotted in Fig. 4.4. It is clear that the values of the GW correction are almost linear with the Kohn-Sham energy. We fit the results in Fig. 4.4 with Eq. (2.88) and Eq. (2.89). The scissor rule is applicable in SiNWs, and we can extend GW corrections by these fitted coefficients to the electronic bands that are not included in Fig. 4.4.

One of the interesting features in Fig. 4.4 is that the GW correction is extremely large for SiNWs. The original LDA Kohn-Sham energy gap is around 1.6 eV, and the GW correction opens it up to 3.2 eV, which is double of the LDA result. In the second chapter when we derive the formula of the GW approximation, the self-energy correction is divided into two parts, the screened-exchange term and the Coulomb-hole term. Because of the special geometry of SiNWs, the screened dielectric constant is much smaller than that of bulk materials, so the screened exchange term will give a large correction. At the same
Figure 4.5: One-particle absorption spectra of the SiNW with different number of bands included. The curves are smearing by a Gaussian of an energy width of 0.1 eV. (a) 12 valence bands and 12 conduction bands; (b) 15 valence bands and 15 conduction bands; (c) 20 valence bands and 20 conduction bands. The main structure of the above three curves is similar; only small changes in the relative amplitude of the peaks are found. The position of the band gap is marked with an arrow in the top panel. In addition, we mark the three main absorption peaks for the convenience of the discussion with a lower triangle, circle and upper triangle.
time, the overlap between wave functions is enhanced by the low dimensionality of SiNWs, so the Coulomb-hole term is also amplified. That is why in most low-dimensional systems, such as nanotubes, clusters, molecules, etc., similar large GW corrections are found [90].

### 4.3 Optical Absorption Formulas

The optical absorption of materials is described by the imaginary part of the frequency-dependent dielectric function [96, 97]

\[
\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega). \tag{4.2}
\]

Without the electron-hole interaction, the one-particle optical absorption is calculated by the Fermi’s golden rule as the sum of transitions between the independent single electron and hole states.

\[
\varepsilon_2(\omega) = \frac{16\pi e^2}{\omega^2} \sum_{v,c} |\vec{\lambda} \cdot \langle v|\vec{\nu}|c\rangle|^2 \delta(\omega - (E_c - E_v)), \tag{4.3}
\]

where \(\vec{\lambda}\) is the polarization unit vector, \(|v\rangle\) is the valence state, \(|c\rangle\) is the conduction state, and \(E_{v,c}\) is the corresponding eigenenergy.

When the electron-hole interaction is included, the above formula should be expressed as

\[
\varepsilon_2(\omega) = \frac{16\pi e^2}{\omega^2} \sum_{S} |\vec{\lambda} \cdot \langle 0|\vec{\nu}|S\rangle|^2 \delta(\omega - \Omega_S). \tag{4.4}
\]

Considering the expression for excitons

\[
|S\rangle = \sum_{v,c} A_{v,c}^S |vc\rangle, \tag{4.5}
\]

Eq. (4.4) can be written as

\[
\varepsilon_2(\omega) = \frac{16\pi e^2}{\omega^2} \sum_{S} \sum_{v,c} |\vec{\lambda} \cdot A_{v,c}^S \langle v|\vec{\nu}|c\rangle|^2 \delta(\omega - \Omega_S). \tag{4.6}
\]
Since the transition matrix $\langle v|\vec{v}|c \rangle$ and the exciton amplitude $A^{S,v,c}$ are complex numbers, the relative phase between them is important. The interference effect enhances some transitions, depresses other transitions, and induces a significant change of the whole optical spectra.

4.4 The One-particle Optical Spectra

After getting the band structure of SiNWs, we calculate the one-particle optical absorption using Eq. (4.3). The result is shown in Fig. 4.5 with a 0.1 eV smearing. In the one-particle calculation, we include many electronic bands, and the computational cost scales as $N_v N_c$, where $N_c$ and $N_v$ are the numbers of conduction and valence bands included in optical transitions. However, when we consider the exciton calculation, the interaction kernel of BSE is proportional to $N_v^2 N_c^2$. The computation time will increase rapidly with an increasing number of bands. In order to make consistent comparisons, we have to limit the number of bands in the optical absorption in the one-particle calculation.

In Fig. 4.5, we show the dependence of the one-particle optical absorption spectra on the number of bands included. The perfectly converged optical spectrum is obtained by including 20 valence bands and 20 conduction bands. After including 12 valence bands and 12 conduction bands, the main structure of the spectrum does not change any more. In other words, the positions of absorption peaks and their relative intensity are almost converged with 12x12 bands. Furthermore, we assume that 12 valence bands and 12 conduction bands are enough to show the main features of the optical absorption even when the excitonic effect is considered. By limiting the number of bands included, the calculation becomes feasible without sacrificing the convergence of the optical absorption.

There are two features in the one-particle optical spectra shown in Fig. 4.5 that we should consider:

(i) There is a large energy region between the quasiparticle energy gap (marked with an arrow in Fig. 4.5) and the optical absorption edge where there is no, or very little, optical absorption. It indicates that the transitions between the bands close to the direct band gap are not optical active. This is expected because bulk silicon has an indirect band
Figure 4.6: Band structure of the SiNW near the band gap. The solid dots represent the bands that are included in exciton calculations. The main contributions to the optical spectra are marked with I, II, and III. Transition III is not around the zone center; the conduction bands and valence bands are parallel to each other resulting in a large joint density of states (JDOS).
Figure 4.7: The imaginary part of the dielectric function for bulk silicon. The dash-dot line is the optical absorption without the electron-hole interaction. The solid line is the absorption including electron-hole interaction. Two absorption peaks $E_1$ and $E_2$ are marked. The excitonic effect enhances the $E_1$ peak and suppresses the $E_2$ peak. The joint densities of states are plotted at the lower panel with the same legend.

gap. Although the quantum confinement folds the bottom of conduction bands to the zone center, the contribution from those conduction bands to optical transitions with the top of valence bands is very small, if any. In other words, there should be a big difference between the energy gap measured by the Scanning Tunneling Microscopy (STM) and the optical absorption gap [98].

(ii) The main structure of the optical absorption spectrum consists of three peaks. To better understand these peaks, we show the main optical transitions in Fig. 4.6. The transitions marked by I cover the optical absorption from the band gap up to the first peak.
These transitions are clearly the result of quantum confinement. As the size of the wire gets larger, the quantum confinement will be weaker, so the first peak will disappear in the limit of bulk silicon. The second high peak, marked with a solid circle, has two groups of transitions. One is transition II; the other is the transition III, as marked in Fig. 4.6. When we study the origin of the bands involved in transitions II and III, we find that they are from folded bands of bulk silicon along the $\Gamma - L$ direction, corresponding to the $E_1$ peak in bulk silicon. As shown in a previous study [99], the one-particle result concludes that the quantum confinement effect in the SiNW enhances the $E_1$ peak and moves it to a higher energy. The last peak marked with a inverse triangle originates from the $E_2$ peak of the bulk spectrum and will dominate the optical spectrum in the bulk limit.

4.5 The Optical Spectrum Including the Excitonic Effect

The optical spectrum with the excitonic effect is calculated by Eq. (4.6). The optical absorption in the silicon bulk can provide a good reference point to understand what happens in SiNWs. So at the beginning of this section, we give a short discussion about the excitonic effect in bulk silicon. The optical absorption spectra of the silicon bulk with and without the excitonic effect are shown in Fig. 4.7. The main absorption peaks $E_1$ and $E_2$ are marked. Previous studies have thoroughly discussed the origin of these peaks and the excitonic effect in the silicon bulk. One of the most interesting features is that the excitonic effect redistributes the strengths of $E_1$ and $E_2$ peaks, increasing the $E_1$ intensity while decreasing the $E_2$ intensity. As shown in Fig. 4.7, the density of states (DOS) of the exciton in bulk silicon changes little, while the optical spectra shows a significant excitonic effect. As we will see in the following, the SiNW also shows the similar character in its absorption spectra.

The optical spectrum of the SiNW including the excitonic effect are shown in Fig. 4.8. As discussed in the last section, 12 valence bands and 12 conduction bands are included in this calculation. The smearing with a width of 0.1 eV is used to obtain a smooth spectrum. There are three features in this spectra.

(i) The absorption edge is shifted significantly in Fig. 4.8. We mark the position of the
Figure 4.8: The optical absorption spectrum of an isolated silicon nanowire with a 1.2 nm diameter for polarization along the axis. The smearing width is 0.1 eV. The dash-dot line is the one-particle spectrum calculated with 12 valence bands and 12 conduction bands. The solid line is the optical spectrum including the electron-hole interaction. The strong e-h interaction contributes to the huge shift of the optical spectra. We mark the three main peaks of the excitonic spectrum with an upper triangle, a circle and a lower triangle, respectively. In addition, the peak marked by a circle is enhanced by the excitonic effect.

The first exciton with an arrow in Fig. 4.8. Compared with quasiparticle results, the exciton binding energy is around 1.1 eV. While the GW correction to the self-energy opens the band gap from 1.6 eV to 3.2 eV, the excitonic effect pulls down the first exciton state to 2.1 eV. The above two effects compensate for each other. Recently this cancellation between these two corrections is also found in a graphite sheet.

The exciton in bulk silicon has a binding energy of 14.7 meV, which is two orders less than that of SiNWs. So the exciton binding energy is significantly enhanced in SiNWs. However, when we plot the joint density of states (JDOS) of single particles and the DOS of excitons in Fig. 4.9, the excitonic effect only shifts the DOS by about 0.3 eV. This means most of excitons do not have a large binding energy. The small change of the DOS and the significant modification of the optical absorption of SiNWs seem to contradict each other. To answer this question, we have to reexamine the formula of the optical absorption calculation, because there is a tight relationship between the exciton wave function and their optical activity. In order to make this relation clear in SiNWs, we try to plot the wave
Figure 4.9: The density of states of excitons (DOSE) is shown by the solid line, and the joint density of states of band structure (JDOS) by the dot-dash line. The shift from excitonic effect is about 0.3 eV.

From Eq. (4.5), we can write the exciton wave function in real space as

$$\chi_s(x_e, x_h) = \sum_{k}^{\text{hole}} \sum_{v}^{\text{elec}} \sum_{c}^{\text{A}} A_{vck}^s \psi_{c,k}(x_e) \psi_{v,k}^*(x_h),$$

(4.7)

where $A_{vck}^s$ is the coefficient of this expansion, $\psi_{c,k}(x_e)$ is the electron wave function, and $\psi_{v,k}^*(x_h)$ is the complex conjugate of the hole wave function. Here we fix the hole at the center of one unit cell, and plot the electron distribution around it.

In Fig. 4.10, we plot five typical exciton states. The first three are optical active excitons, while the other two are non-optical active excitons (dark excitons). One of the interesting features of this figure is that the optical active excitons are localized, while the size of the dark excitons is much larger. It is easy to explain the absorption formula, Eq. (4.4). The localized exciton has a larger overlap between the hole and electron, which contributes to the larger integration in Eq. (4.6). So in general, the strongly optical active exciton has a large overlap of the electron and the hole. (Of course, the overlap is not the only factor to
Figure 4.10: The density of the electron for several excitonic states with the hole fixed at the origin. From top to bottom, five representative states are shown with the energy values of 3.3 eV, 4.2 eV, 4.2 eV, 4.7 eV, and 6.7 eV, respectively. Corresponding cross section plots on the \(xz\) plane are shown on the right, while the three-dimensional plots for the same data are shown on the left. Because the bottom two states are delocalized, we have to use a large scale to plot them.
determine the optical activity of excitons. The symmetry and details of wave functions are also important in calculating the optical absorption spectrum.) At the same time, the larger overlap between the electron and hole gives rise to a larger interaction between them, which means a larger binding energy. For those dark excitons, most of them are non-localized and have a small overlap between the electron and the hole, resulting in a small optical activity and weak interaction.

An interesting finding is that most of excitons are dark, and they dominate the DOS of excitons. Because of the delocalization of the wave function of dark excitons, the electron-hole interaction is weak in these excitons, which contributes to the similarity between the JDOS of single particles and the DOS of excitons. On the other hand, a small number of optical active excitons dominate the optical absorption spectrum; their large binding energy results in a large apparent shift of the optical absorption.

(ii) The second peak, marked by a solid circle, is enhanced by the excitonic effect. As we can see in the optical absorption formula, Eq. (4.6), the coefficients of an exciton, \(A^S_{v,c,k}\), are very important. It can be thought that the \(A^S_{v,c,k}\) takes the oscillator strengths of single-particle results and creates a new combination (an exciton). Thus as we see in bulk silicon, coefficients \(A^S_{v,c,k}\) redistribute the \(E_1\) and \(E_2\) peaks. In addition, the phase interference between \(A^S_{v,c,k}\) and the transition matrix will considerably affect the absorption amplitude. In the SiNW, we find most of the components in this huge peak are from the solid circle marked transitions of single-particle results in Fig. 4.5, which are redistributed by coefficients \(A^S_{v,c,k}\). In the previous section, we found that these transitions originate from the \(E_1\) peak in the silicon bulk. The third marked peaks in Fig. 4.8 is related to the \(E_2\) peak, which is suppressed by the electron-hole interaction in the SiNW. Therefore, the excitonic effect enhances the \(E_1\) peak and suppresses the \(E_2\) peak in the SiNW.

As we discussed above, the quantum confinement increases the overlap between the electron and the hole and induces a strong excitonic effect. To get a more detail understanding of the confinement, we plot the wave function of excitons in real space again. As we do in Fig. 4.10, we still fix the hole at the center of one unit cell and plot the electron charge distribution. Because the strong confinement is from the surface boundary, we plot wave
Figure 4.11: Charge density plots in the x-y plane of the nanowire. The hole is fixed in the origin ($r_h = 0$), and the density of the electron is plotted. The left figures are the 3-D plots, and the right figures are the corresponding contour plots. The energies of these two excitons are 3.2 eV and 3.5 eV, respectively.
Figure 4.12: The depolarization effect of an isolated SiNW. When the polarization direction is perpendicular to the nanowire, the optical absorption is extremely small as the solid line shows. This effect cannot be included in one-particle calculations [99].

functions in the cross section in Fig. 4.11. The plots on the left side are 3-dimensional pictures, and the right side ones are the corresponding contour plots. It is clear to see the different nodal structure characteristics of a cylindrical confinement.

4.6 The Local Field Effect

Another interesting topic in nanostructures is the direction dependence of the optical absorption spectra. The depolarization effect [100, 101, 102] plays an important role in the optical absorption spectra of nanostructures, which suppresses the optical absorption if the polarization vector is perpendicular to the NW growing direction. Within the framework of BSE, we have also calculated the depolarization effect in SiNWs. Instead of including only the local field (such as the random phase approximation), we also take into account the attraction of the electron and the hole, expecting to obtain this directional dependence of the absorption spectrum. The result is shown in Fig. 4.12. Clearly the depolarization effect
almost completely suppresses the absorption intensity. The strong directional dependence of the optical absorption shown in our calculation gives rise to broad applications in many fields, such as micro antenna arrays and the laser cavity [103, 104].

4.7 The Hydrogen-like Model in SiNWs

The hydrogen model is often used to describe excitons in semiconductor materials because of its simplicity. However, it has not been proved that the hydrogen model is correct in nanostructures. Particularly, some important excitons are not the electron-hole pairs at the zone center in the SiNW, so it is hard to apply the effective mass approximation. In the following, we examine band edge excitons in SiNWs to see how well the hydrogen model describes excitonic states in SiNWs.

The simplest excitons in our calculation are derived from the band edge states responsible for the transition A marked in Fig. 4.13(a). This transition in the one-particle optical absorption spectrum involves the bottom of the conduction bands and the top of the valence bands. Four bound excitons are identified and marked in Fig. 4.13(b) as $A_1$, $A_2$, $A_3$, and $A_4$. Since these excitons are at the zone center, we expect the hydrogen model to describe their properties. We also plot the binding energy of these excitons in Fig. 4.13(c), which is fitted well by the energy levels of a hydrogen atom. The energy levels of a hydrogen atom scales as $\frac{1}{n^2}$. Here our fitting result in SiNWs is $\frac{1}{n^{1.8}}$, not far away from the ideal scaling. In Fig. 4.14, we plot excitonic wave functions in the real space in the xy plane for these four states. The hole is fixed at the center of a unit cell, and the electronic charge density is shown. As the quantum number increases, more and more nodal structures develop along the wire direction. This set of excitons can be described by a hydrogen-like model, although the corresponding “hydrogen” states are very squeezed in the SiNW.

Compared with the typical binding energy in the silicon bulk, which is about 14.7 meV, this set of excitons have an extremely large binding energy. There are two explanations for the huge electron-hole binding energy in SiNWs [105]. One is the quantum confinement by the special geometry of SiNWs. The quasi-one-dimension SiNWs create a significant overlap between the electron and the hole distributions, resulting in a large binding energy.
Figure 4.13: The evolution of the bound excitons at the band edge. In figure (a), the transition A is marked in the band structure. In figure (b), the dash-dotted line is the one-particle result, and the solid line is the BSE result. Four excitonic states originated from transition A are identified in the BSE calculation: $A_1$, $A_2$, $A_3$, and $A_4$. Figure (c) shows the binding energy of the four excitons, $A_1$, $A_2$, $A_3$, and $A_4$. The dotted line is a fitting result. The binding energy as a function of quantum number $n$ is $\frac{E_0}{n^{1.81}}$, which is close to the case of a hydrogen atom.
Figure 4.14: Electron charge distributions in the yz plane for exciton states $A_1$, $A_2$, $A_3$ and $A_4$, with the hole fixed at the origin. As the quantum number increases, the binding becomes weaker and weaker. With nodal structure develops, the exciton wave function extends more in real space. Note that the scale changes in the last two figures, $A_3$ and $A_4$. 
The other explanation concerns the weak screening in NWs. Because of the vast vacuum around the wire, the effective screening is much reduced. Since the electron and the hole interact with each other through the screened Coulomb field, the smaller dielectric constant further increases the binding energy.

As discussed above, a larger binding energy means a larger overlap between the electron and the hole, which is also likely to have a large oscillator strength. Figs. 4.13 and 4.14 demonstrate this point. From $A_1$ to $A_4$, the size of the exciton increases, which reduces the overlap between the electron and the hole, and the binding energy decreases. In addition, the optical absorption also decays rapidly from $A_1$ to $A_4$.

4.8 Summary

We have conducted a detailed study on the excitonic effect in an isolated SiNW. The LDA and quasiparticle energy gaps are calculated. By solving the Bethe-Salpeter equation, the optical absorption spectrum is obtained, and the excitonic states are studied. Compared with the silicon bulk, many new features appear in the optical absorption spectrum of SiNWs. The exciton binding energy is much larger than that in bulk materials. The weakly screening Coulomb interaction and one-dimensional geometry contribute to this large electron-hole binding energy. The new high peaks in the optical absorption spectrum arise from the quantum confinement. We analyze the origin of this high absorption peak in detail. The depolarization effect, also called the local effect, is reproduced by our calculation with the electron-hole interaction included. The excitonic states in real space are plotted. The squeezed hydrogen atom model fits excitonic states in SiNWs very well.
CHAPTER V

ELECTRONIC STRUCTURE IN SI/GE CORE-SHELL NANOWIRES

As the characteristic size of microelectronics approaches the nanometer size, nanostructure-based devices attract more and more attention in the nanomaterial research. The nanowire (NW), as one of widely studied one-dimensional nanostructures, exhibits remarkable properties because of the quantum confinement effect. Potential applications can be found in fabricating novel high-performance devices, such as the p-n junction diode [106], the p-channel or n-channel coaxial gated field effect transistor (FET) [107, 108], and complimentary FET [109], etc. These devices show outstanding properties compared with traditional ones. For example, free carriers in NWs have impressively high mobility, which gives hope to terahertz circuits. In addition, the NW transistor has a better transconductance that is necessary for high gain devices [110].

When fabricating most of these devices, researchers have to use the doping process to supply free carriers. However, doped impurities play the role of defects at the same time, which scatter free carriers and decrease the mobility. To avoid these difficulties, a novel doping mechanism has been proposed in the core-shell NW, a radial heterojunction nanostructure. It is suggested that one can dope impurities in the shell and inject free carriers to the core or vise versa. Since doped impurities and free carriers are separated in the core and shell regions, respectively, the scattering rate between them is reduced. Therefore, the mobility of free carriers is improved, which is preferred for high-speed devices. Recently, this idea has been realized in experiment. The corresponding core-shell NW-based device shows very competitive electronic performance, and becomes a hopeful candidate for elementary devices of the future microelectronics.

Despite the attractive points mentioned above, there are several puzzles in the doping process. For example, when we discuss the band offset in core-shell NWs, it implies that
localized electrons can exist in the core and shell regions, respectively. But the existence of these separated states in the thin NW is an open question without a clear up-to-date answer. Even if there exists a band offset between core and shell states, other questions have not been answered yet, such as what the kind of band offsets is in NWs, how large the band offset is, how the band offset evolves with the size of NWs, etc. Although previous knowledge on bulk Si/Ge heterojunctions [111, 112] and superlattices [113, 114] are instructive, they are insufficient to provide information on the cylindrical core-shell NWs with quantum confinement. In order to guide the future fabrication of the NW-based high-speed devices, a clear and accurate understanding about the heterojunction under strong cylindrical confinement is urgent and important.

In order to answer the above questions, we have performed first-principles calculations in Si/Ge core-shell NWs. The electronic band structures of Si/Ge core-shell NWs are studied. States close to the band gap in NWs along the [110] direction are resolved and separated into the core and shell states, respectively. Band offsets between the core and shell electronic states are found to depend not only on the diameter of the NWs but also on the core-shell size ratio, which includes significant corrections to the results from the Si/Ge heterojunctions and superlattices. The existence of the band offset makes it possible to dope impurities in the shell part of NWs and inject free carriers to the core region, or vice versa. Finally, an optimal doping strategy is concluded based on our calculation.

5.1 Motivation for Studying Core-shell NWs

The motivation of our calculation is to obtain the knowledge about the electronic structure of Si/Ge core-shell NWs in order to guide the applications. One of the most important applications is the field effect transistor (FET), which is the fundamental element of today’s microelectronics because of its simplicity and easy integration into the integrated circuits (IC). Therefore, we introduce the core-shell NW FET and discuss how the core-shell NW based FET works and its advantages in this section.

There are many types of FETs; the most widely used one is the metal-oxide-semiconductor FET (MOSFET), whose structure is shown in Fig. 5.1. The key part of the FET is the
The basic structure of a MOSFET, one of the most popular FETs. The voltage of the gate controls the width of the channel and tunes the transconductance between the source and the drain. The substrate can have a different doping type from the gates. The mobility of free carriers inside the conduction channel is important to determine the speed of this device.

Figure 5.1: The basic structure of a MOSFET, one of the most popular FETs. The voltage of the gate controls the width of the channel and tunes the transconductance between the source and the drain. The substrate can have a different doping type from the gates. The mobility of free carriers inside the conduction channel is important to determine the speed of this device.

gate structure that controls the conduction channel between the source and the drain. The function of the gate varies under different working conditions. In analog circuits, the voltage of the gate controls the current going through the channel and works as a simple amplifier. In digital circuits, the gate works as a switch. By turning on or off the conduction channel, the “0” and “1” signals are realized.

In the process of device miniaturization, NWs are considered a good element for one-dimensional transportation. Many NW-based devices have been produced in labs. Among them, the NW-based FET attracts attention because of the importance of FETs in microelectronics. By measuring the key parameters of NW-based FETs, scientists find that the saturation current density and transconductance of these FETs are very large, which is great news for high gain circuits. Furthermore, with the measured higher mobility of carriers, the NW-based devices can work under the higher frequency condition, making it possible to build high-speed circuits. Although no implementation at today’s industrial level has
Figure 5.2: The NW-based field effect transistor [115]. The drain and source are connected by the core-shell NW. The gate structure is under the NW, so this FET is called a back-gate FET. The advantage of the back-gate FET is that it is easily fabricated because of its simple structure and has a large linear working region.

taken place, the NW-based FET is valuable as a promising choice. Fig. 5.2 presents one of the common NW-based FETs.

In Fig. 5.2, the NW works as a platform to supply the conduction channel between the source and the drain. The current going thorough the NW can be controlled by changing the voltage of the gate. In practical, there are mainly two kinds of NW based FETs clarified by the gate structure: one is the back-gate FET that is shown in Fig. 5.2; the other is the top-gate FET that has the better high-frequency performance. Because our main topic is the role of core-shell NWs, we focus on the property of the NW implemented inside FETs.

The property of free carriers is an essential factor in determining the quality of microelectronic devices. The most popular way to get free carriers in semiconductors is by impurities doping, which is also true in NW devices discussed above. However, impurities worsen the mobility of carriers as scattering centers. In light of the band lineup of semiconductor superlattices, a novel doping mechanism is proposed in core-shell NWs, which is
shown in Fig. 5.3. The band structure of Si/Ge core-shell NWs is expected to be close to that in superlattices, which is called the type II band offset. As shown in Fig. 5.3, free carriers will jump into the lower energy state automatically and are separated from impurities by the band offset. Therefore, the mobility of free carriers is improved by decreasing the scattering rate with impurities.

However, there are several problems in this process as we pointed out at the beginning of this chapter. Because the characteristic size of the electron wave function is the order of the de Broglie wave length, the electron is not localized at one point in real space. In the thin core-shell NWs, the typical diameter of the wires is around 3 to 6 nm. In such thin wires, it is an open question if the electronic states can be separated into the core part and shell part, respectively. If it is impossible to distinguish the shell state and the core state, the concept of the band offset makes no sense. Furthermore, there is no doubt that the band offset depends on the size of core-shell NWs because of the quantum confinement effect. But, how it depends on the confinement is still unknown.

By now, many experiments have been done regarding the doping process in core-shell NWs [116, 118, 119, 117, 120], but there are few theoretical studies to answer questions

Figure 5.3: The doping process in core-shell NWs. Assuming that core-shell NWs have a similar band lineup as Si/Ge superlattices, we expect that the free carriers will move to the lower-energy state and become separate from doped impurities.
posed above. To interpret experimental results, people usually use results from bulk Si/Ge heterojunctions or superlattices, which are not accurate in core-shell NWs because they lack of the consideration of the cylindrical quantum confinement. As the technology advances, new experiments even show free carriers in the core-shell NW-based transistor without doped impurities. They find that there is a self-doping process in Si/Ge core-shell NWs because of the temperature dependence of the Fermi surface. However, the band offset between the core and shell of NWs is still an important factor in doping efficiency even for the self-doping process. In other words, a clear and accurate understanding of the band structure in the radial heterojunction of core-shell NWs is timely and important.

5.2 Review of Band Offsets in Bulk Heterostructures and Superlattice

Because of the importance of the band offset in determining the electronic properties of devices, there are several methods for calculating the band offset in semiconductors, such as the tight-binding method, the effective mass model, etc. Compared with the above methods, first-principles calculations are more powerful in heterojunction studies, because they do not need the reference energy. The internal field induced by the heterojunction, which is very hard to deal with, can be considered automatically in first-principles calculations. The first \textit{ab initio} result of the Si/Ge heterojunction was reported in the middle of 1980s [111], which is shown in Fig. 5.4. From this figure, we can see that there are several important characteristics for the band offset of the Si/Ge heterojunction.

(i) Because of the mismatch of the lattice constants between silicon (5.43 Å) and germanium (5.65 Å), the strained lattice will affect the band offset strongly. In the left panel of Fig. 5.4, the germanium part is strained, the offset of conduction bands is 0.28 eV, and that of valence bands is 0.54 eV. However, when the silicon side is strained as shown in the right panel, the offset of conduction bands is 0.55 eV, and that of valence bands is 0.31 eV. So the structure of the heterojunction is an important factor to influence values of band offsets in a heterojunction.

(ii) The strain of the Si/Ge heterojunction modifies the energies of the bands. In the
Figure 5.4: Calculated band offsets of Si/Ge heterojunction [111]. The left panel is the case with germanium strained, and the right panel is the case with silicon strained. We only consider the solid line result, and the dash line is some weighted average offset that is beyond our consideration. In addition, the relative values of this figure are meaningful because of the choice of reference energy.

germanium crystal, the minimum of the conduction bands is at the $\Delta$ point. However, in the strained Si/Ge heterojunction, the minimum changes to the $L$ point shown in the left panel of Fig. 5.4.

In superlattices, the band offset has been a topic for years. Recently, more results are discovered in Si/Ge superlattices. We pick out one of the results in Fig. 5.5. The offset of valence bands changes to 0.85 eV or 0.37 eV depending on which side is strained. The difference of the band offset is the result from both the strain condition and quantum confinement.

The confinement is one-dimensional in superlattices. In core-shell NWs, the confinement is cylindrically two-dimensional, which should introduce a larger correction to the band offset. As pointed out in the introduction part, we perform first-principles calculations to get trustable band offset values in core-shell NWs in the following part of this chapter.
Figure 5.5: Calculated band offsets of Si/Ge superlattices [113]. The left panel is the case with germanium strained, and the right panel is the case with silicon strained. Only the offset values of valence bands are marked in this figure, and they are similar to those of Si/Ge bulk heterojunction. Because this result is about the long-period superlattices where the confinement is weak, it is obvious the correction induced by the confinement is small.

5.3 The Structure of Core-shell NWs

In addition to the confinement of the wire boundary, the core-shell radial heterojunction adds an additional confinement. To study this extra confinement, we focus on core-shell NWs with a fixed diameter but with different core-shell size ratios. Because of the surface energy consideration and experiment results, we choose NWs with the cylindrical shape. In addition, all NWs have a same diameter, around 4.1 nm, throughout the calculation, and the total number of atoms per unit cell is fixed to be 314 (including 52 hydrogen atoms on the surface). We calculate five different core/shell size-ratio NWs along the [110] direction as listed in Table 5.1. There are two types of core-shell NWs in Table 5.1: one has a silicon shell and a germanium core; the other has a germanium shell and a silicon core. The size of the core in this Table is the fractional core diameter with respect to the diameter of the whole NW. In addition, we calculate the same size of the pure silicon NW and germanium NW as references. The structures of core-shell NWs are shown in Fig. 5.6. The outer
Figure 5.6: Ball-and-stick model of the hydrogen-passivated Si/Ge core shell NWs along [110] direction. The diameter of the wire is about 4.1 nm. Although we calculate five core-shell structures, we only plot two of them (I and IV from Table 5.1).

hydrogen atoms passivate dangling bonds to eliminate surface states close to the band gap.

We use the Vienna Ab-initio Simulation Package (VASP) for our first-principles calculations. The plane-wave method is used, combined with ultrasoft pseudopotentials [42]. The non-local pseudopotential is projected into the real space for computational efficiency. To keep the periodical boundary conditions, Si/Ge core-shell NWs are arranged into a hexagonal lattice with enough vacuum distance between them (more than 6 Å).

The first step for this calculation is to test the convergence of the total energy with respect to the k-grid and plane-wave cutoff energy. We use the Monk-Horst scheme as the k-grid sampling [95]. The convergence is shown in Fig. 5.7. We choose the 1x1x4 k-grid and 170 eV cutoff energy, which are marked with arrows. Although the cutoff energy is

Table 5.1: The core-shell structure NWs we calculated. The NWs are marked with an index to help the discussion. The core size is normalized to the NW diameter. Each of these five type has two kinds of components, silicon core and germanium core.

<table>
<thead>
<tr>
<th>Index of NWs</th>
<th>wire I</th>
<th>wire II</th>
<th>wire III</th>
<th>wire IV</th>
<th>wire V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atoms at core</td>
<td>30</td>
<td>42</td>
<td>62</td>
<td>82</td>
<td>122</td>
</tr>
<tr>
<td>Atoms at shell</td>
<td>232</td>
<td>220</td>
<td>200</td>
<td>180</td>
<td>140</td>
</tr>
<tr>
<td>Core size</td>
<td>0.34</td>
<td>0.40</td>
<td>0.49</td>
<td>0.56</td>
<td>0.68</td>
</tr>
</tbody>
</table>
Figure 5.7: The energy convergence with respect to the cutoff energy and k-grid. The cutoff energy is set to be 170 eV as marked in the left figure; the k-grid is set to be 1x1x4 shown in the right figure. Although the total energy convergence with the cutoff energy and k-grid is not perfect, we find they are enough for the accurate band structure calculations. The self-consistence converge of the total energy is set to be less than $10^{-5}$ eV. The force is relaxed to be less than 0.05 eV/Å, and the stress along the z direction is reduced to be small.

As we show in previous calculations about lattice vibrational modes in SiNWs, it is very often that the relaxation process is trapped by some local minimal energy, which prevents the relaxed structure from reaching the global minimum of the total energy. Therefore, it requires the extra attention to determine the relaxed structure in core-shell NWs. We start from three different bond lengths to construct the core-shell structure in our relaxation process. The first one is the Si-Si bond length; the second one is the Ge-Ge bond length.

Table 5.2: The relaxed bond length of core-shell NWs calculated from three different starting bond-length values, showing a converged final structure.

<table>
<thead>
<tr>
<th>Index</th>
<th>Starting from Si</th>
<th>Starting from Ge</th>
<th>Starting from average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si-Si bond length</td>
<td>2.332</td>
<td>2.336</td>
<td>2.332</td>
</tr>
<tr>
<td>Ge-Ge bond length</td>
<td>2.380</td>
<td>2.388</td>
<td>2.383</td>
</tr>
</tbody>
</table>
the third one is the average of the Si-Si and Ge-Ge bond lengths. After relaxation, it is found that all of the above three starting points yield almost the same final structure, which is shown in Table 5.2. Therefore, the relaxed result is not sensitive to the starting structure.

The last question about the structure is the stress inside the core-shell NWs. There is a mismatch of the silicon and germanium lattice constant (4 percents difference), which introduces the strain in the interface. In bulk Si/Ge heterojunctions, the electronic band structure will be quite different depending on which side is strained. In our calculation, the problem is even more complicated because the core and the shell play different roles in the distribution of the strain. To simplify our consideration, we only calculate the final structure obtained by the above relaxation process without considering other different strained structures.

### 5.4 The Band Structure of Si/Ge Core-shell NWs

The electronic band structure is calculated using the local density functional approximation (LDA) [5]. All calculated core-shell NWs are found to have a direct band gap, although in some germanium-core NWs, the direct and indirect band gaps are close. In Fig. 5.8, we
plot the variation of the band gap with the change of the core ratio in core-shell NWs with a fixed diameter. The upper panel is for the silicon-core NWs, and the lower panel for the germanium-core NWs. Both kinds of core-shell NWs show similar features. For example, in silicon core NWs, when the core size is zero, the band gap is exactly the value of the pure germanium wire. As the core size increases, the band gap decreases. When the core size is about half of the wire diameter, the band gap reaches its minimum. After that point, the band gap increases with the core size. Finally, when the core ratio is one, the band gap is exactly the result of a pure silicon NW. In fact, the fundamental band gap is the energy gap between a core and a shell state. When we change the core-shell ratio, the change of the confinement for the core and shell states will introduce the change of the energy gap between these states. So quantum confinement in core-shell NWs plays an important role in determining the variation of the fundamental band gap in Fig. 5.8.

The core-shell NW can be thought of as a radial Si/Ge heterojunction and acts as a quantum well inside the NW. We expect that there is a band offset between electronic states inside and outside the well. An example of a possible band lineup is shown in Fig. 5.9, where the core part acts as a quantum well. Fig. 5.9 shows that there are two kinds of electronic

\[ \text{Figure 5.9: An example of the band lineup of core-shell NWs.} \]
Figure 5.10: The band structure close to the fundamental band gap. The upper panel is a NW with a silicon core, and the lower panel is a NW with a germanium core. The discretization of the bands close to the band gap implies that there are band offsets between core and shell states.

states. The states, marked with the dash line, are confined within the core of the NW; the states, marked with the dot line, extend over the whole wire. The dash-line states are confined more strongly than the dot-line states, so energy levels between dash-line states are more discrete. As a result, if there exists a band offset between states inside and outside the quantum well, we expect to see more discrete bands close to the fundamental band gap.

To prove the above idea, we plot the calculated band structure of core-shell NWs in Fig. 5.10. Since there are too many structures, we just pick out two typical ones in this figure, which are the type II wire in Table 5.1. The detail band structures of all calculated NWs are shown in Appendix I. As we expect, the bands close to the gap in these two plots in Fig. 5.10 clearly show large discretization. So these bands are probably confined in the core or shell part as shown in Fig. 5.9. Furthermore, when we exchange the core-shell NWs from silicon core to germanium core, the overall shape of the band structure close to the band gap is also exchanged, which further supports that there is a band offset in core-shell NWs. We will examine the charge density of those bands close the gap in order to determine the value of band offsets.
Figure 5.11: The charge density of electronic states close to the fundamental band gap, and the NW is with a silicon core and a germanium shell along the [111] direction. The left panel is the charge density of the top of valence bands, and the right panel is that of the bottom of conduction bands. We can see that the charge density in the right panel is not confined within the core or the shell. As a result, it is hard to see the offset in this structure.

5.5 Band Offset of Core-shell NWs

Because the band edge is at the zone center, which is associated with the electron transport, only the charge density and band offset at the Γ point are considered. Therefore, all of the following charge densities and band offsets are those of electrons close to the Γ point. In addition, we plot the charge densities that are integrated over the axial direction in order to show them in two-dimensional figures.

It must be pointed out that confined states may not exist in all wire orientations. In the thin [111] NW with a diameter around 3 nm, we do not clearly find a set of confined states at the bottom of conduction bands. In Fig. 5.11, the left panel is the integrated charge density of the state at the top of valence bands, which is confined within the core. The right panel is the integrated charge density of the state at the bottom of conduction bands, which extends over the whole wire. The smearing of the state at the bottom of conduction bands implies that there is no potential well deep enough to create a band offset for conduction bands in NWs along the [111] direction. As a result, the [111] Si/Ge core-shell NW may not be a good choice for n-doping when its diameter is small.
Figure 5.12: The charge distribution of the core-shell structure III in Table 5.1 (silicon core). We plot the states close to the band gap. The three figures in the upper panels are from conduction bands, and the three in the lower panels are from valence bands. The number above each plot is the index of the band. The beginning of valence bands is confined within the shell part, while the conduction bands are confined within the core part. When we go further away from the band gap, the non-confined bands will show up, such as VB547 and CB553, which give us the critical transition point from the confined bands to the non-confined bands.
In core-shell NWs along the [110] direction, examining the electronic charge density confirms that electronic states close to the top of the valence bands and the bottom of the conduction bands are confined in either the core or the shell part. First, we focus on core-shell NWs with a silicon core. The plots of the electronic charge density in the real space for a few represented states are shown in Fig. 5.12. The index of these bands is marked in the figure. The upper panels show the plots of conduction bands, and the lower panels show the plots of valence bands. It is clear that electronic states in the upper panel are confined within the core, and states in the lower panel are confined within the shell. Similar results are found in the germanium-core NWs, but the confinement in the core and the shell is inverted. In Fig. 5.13, the electronic states close to the top of valence bands are confined within the core, and the states close to the bottom of the conduction bands are confined within the shell.

When we further check the distribution of those states in the above figures, it is found that the confinement in the core-shell structure is anisotropic. For example, in Fig. 5.12, the electronic state CB551 shows a weaker confinement along the [110] direction than other directions, so the charge density leaks along that direction, resulting in an electronic charge density that is not perfectly cylindrical in shape. This is understandable if we consider the result of bulk Si/Ge heterojunctions. In bulk Si/Ge heterojunctions, the band offset of conduction bands of the [110] direction is around 0.05 eV, which is much less than the value of the [100] direction (0.5 eV). So even in Si/Ge core-shell NWs, where the band offset values are strongly affected by the cylindrical confinement, the results from the bulk can still explain the behavior of the confined states.

We track electronic states close to the band gap. For example, in NWs with a silicon core, the bottom of the conduction bands, index (CB551), is confined within the core part. We go through the higher bands until we find the conduction band (CB553), whose main distribution is not confined within the core part, then we take this as the transition point where the electronic state changes from being confined in the core to extending over the whole wire. The band offset value is determined by taking the energy difference between the states (CB551) and (CB553). In the same way, the band offset of valence bands can be
calculated as well.

The above method is clear in theory, but sometimes difficulties arise in practice in assigning an electronic state as a core state or shell state. An example is the valence band (VB) in Fig. 5.12. The main part of charge of this state extends to both the core and shell regions. It is necessary to find a criterion to identify the core and shell states. We choose to use the average charge density inside the core and shell part, respectively. For example, if the average charge density inside the core is twice as large as that inside the shell, we assign this state as a core state. With the help of this criterion, we specify electronic states close to the band gap into confined states and non-confined states.

In our calculations, there are two types of core-shell NWs (silicon core and germanium core), and each of them has five different sizes as we show in Table 5.1. So, totally we have investigated ten NWs. It is impossible to list all these average charge densities here, so we show the details in Appendix II, which contains the comparison of average charge densities.

![Figure 5.13:](image)

**Figure 5.13:** The charge distribution of the core-shell structure I in Table 5.1 (germanium core). We plot the states close to the band gap at \( \Gamma \) point. The arrangement of figures and the meaning of numbers above each plot are the same as in Fig. 5.12. The beginning of valence bands is confined within the core part, while the conduction bands are confined within the shell part.
and how we identify the core and shell states.

The above procedure gives us the band offset values in Fig. 5.14. As we can see that the size dependence of band offsets is complicated. In the silicon-core case, the conduction band offset shows larger and stronger size dependent. In the germanium-core case, the valence band offset is larger and almost constant (around 500-700 meV). A large band offset is preferred in the doping process, because a small band offset will make it easy for carriers to jump back to the doping region when the temperature is raised, thus reducing the efficiency of the doping.

These band offsets are different from bulk Si/Ge heterojunction results. In bulk heterojunctions, the band offset of valence bands is around 500 meV. However, the Si/Ge core-shell structure shows a very different band lineup depending on whether the core is made of silicon or germanium. The reason is that the planar symmetry is broken by the cylindrical geometry of core-shell structures, where the core and the shell have different quantum confinement.

Finally, there is one question about the values of the band offset in core-shell NWs. The local density functional approximation (LDA) underestimates the band gap systematically. In the above results, we use LDA to obtain the electronic structure. The question is whether LDA results are accurate enough for band offset results or not. Our research shows that the calculated band gap may be far from the experimental gap, but the offset is close to the accurate answer as explained below.

The most widely used method to correct the electronic band structure is the quasiparticle calculation using the GW approximation [20]. The electron self-energy will introduce a large correction to the LDA band gap, which almost doubles the band gap in the smallest SiNWs (1.2 nm). However, the GW correction mainly opens up the band gap in both germanium and silicon NWs, with a very small change of the relative positions between these valence bands or conduction bands. As a result, if we suppose that the GW approximation gives the similar correction to both core-shell NWs and pure silicon NWs, the band offset of conduction bands or valence bands of core-shell NWs will not be changed much even if we include the quasiparticle correction.
Figure 5.14: The band offsets of core-shell NWs. The upper part shows the offsets in silicon-core NWs, while the lower parts show germanium-core NWs.
The electronic states confined within the core or the shell will have different quantum confinement as shown in Fig. 5.9. Therefore, the quasiparticle correction to the core states and shell states is not uniform. For example, if the state we are interested in is the state confined within the core part, then after the GW correction, we expect to have a larger correction to this state than other states that extend out of the core. It is difficult to give the accurate answer here, because calculated core-shell NWs are too large for the GW calculation. However, we can give an estimation of the GW correction based on the previous quasiparticle results of pure NWs [99].

As an example, we choose the silicon core NWs with the core ratio close to 0.5, which is the type III structure in Table 5.1. We then estimate the GW correction to the confined state and the non-confined state. The difference between the GW corrections is the maximum error of the LDA band offset.

It is noted that the band gap of germanium wire is similar to that of silicon wire. As shown above, for a silicon-core NW, the bottom of the conduction bands is the first core

Figure 5.15: Band gap evolution of SiNWs. We mark the LDA energy gap with arrows and the corresponding GW result with dotted lines.
state within conduction bands. We have to find the first core state in valence bands to obtain the energy gap of core states. As noted in previous discussions, the top of the valence bands is VB550, and we find the first core state within valence bands is VB532. So the energy gap is the energy difference between CB551 and VB532, which gives the value of 0.76 eV. If we look at the value in Fig. 5.15, we find that the corresponding diameter for the band gap is about 3 nm, which is larger than the real core size (2 nm). This is likely due to the core-shell interface having a weaker confinement effect than the wire-vacuum interface, so the electronic state actually extends further. From Fig. 5.15, we can find the corresponding GW corrected gap is 1.50 eV, so the correction from the GW calculation is about 0.74 eV. In the same way, we can get the GW correction of the non-confined state, which is 0.67 eV. So the GW correction difference between the confined and non-confined state is 0.74 - 0.67 = 0.07 eV, a small error for the size of the band offset.

In conclusion, the difference of these two corrections is less than 70 meV. The above estimation is not an accurate calculation, but it shows that the order of the GW correction will not change much of the band offset and our conclusion. Thus the values of the band offset from LDA calculation are acceptable for our consideration.

5.6 The Optimal Doping Strategy of Core-shell NWs

A motivation for this work is to find the optimal doping strategy from the calculated result. Without the band offset data in core-shell NWs, one usually uses the offset values from bulk Si/Ge heterojunctions.

Based on the above results, we can roughly draw the band offset of core-shell NWs, shown in Fig. 5.16. This kind of band structure makes it possible to dope impurities in the core (shell) and get free carriers in the shell (core). The overall band lineup here is similar to that in the bulk Si/Ge heterojunction and superlattices, such as the bands in silicon are lower than the bands in germanium. So the quantum confinement in core-shell NWs does not completely change the basic feature of the band lineup. But the confinement significantly changes values of the band offset as we show in the last section, which is in fact more important in practice for the application of core-shell NWs.
Figure 5.16: The band lineup of the core-shell structure. Figure (a) presents the silicon core case, which suggests doping electrons at the shell part. Figure (b) presents the germanium core case, which suggests doping holes at the shell part.

When studying more details about the band offset, we find that the difference between core and shell states is clear in valence bands, while it is fuzzy in conduction bands. Because the bottom of the conduction bands are the ground states related with the quantum confinement, and the band offset only modifies the charge distribution more or less, these states generally tend to smear over the whole wire, as shown in the CB552 of Fig. 5.13. For example, CB552 is a shell state, but has significant smearing to the core. This smearing means a significant interaction between shell states and core states and increases the scattering rate, which does not lead to the high mobility.

Then, what is the optimal doping strategy in experiment? The band offset needs to be large enough compared with the thermal energy. Choice I: in silicon-core and germanium-shell NWs, the impurity should be a donor and located in the shell part. Choice II: in germanium-core and silicon-shell NWs, the impurity should be an acceptor and located in the shell part. However, when we examine that the wire states in Choice I, they tend to extend to the whole wire because of the weak confinement. The optimal doping should be Choice II, which agree with the recently published case.

Also, another interesting doping method in core-shell NWs is suggested. Because of the
Figure 5.17: An example of self-doping process. The Fermi surface is below the top of valence bands, which induce free holes without doping impurities. It is found that there is the self-doping process in Si/Ge core-shell NWs.

special band offset structure, self doping is possible in Si/Ge core-shell NWs. The Fermi surface depends on the temperature. Under the right temperature, the Fermi surface can cross conduction bands or valence bands, which supplies free carriers in intrinsic semiconductor materials. A simplified explanation of this self-doping process is shown in Fig. 5.17 for a germanium-core/silicon-shell NW. Because the Fermi surface is below the top of valence bands, there are free holes generated without doping impurities. Although the self-doping process does not need impurities, the band offset is still important to determine the doping efficiency.

5.7 Summary

Because free carriers in NWs have high transconductance and mobility, it is promising to fabricate high-speed devices with NWs. Doping by impurities is one way to obtain free carriers in NWs, but impurities will worsen the mobility of carriers because of the scattering between impurities and carriers. In order to avoid this scattering, a novel doping method in core-shell NWs is proposed. The idea is that the scattering rate can be suppressed if free carriers and impurities are spatially separated. The core-shell NW provides an ideal platform to realize this doping process. If there is a band offset between the core and shell
electronic states, free carriers can move into the lower available states. As a result, one can create the donor or acceptor inside the core or the shell, and get free carriers in a separate region, giving rise to the spacial separation of impurities and carriers.

We perform first-principles calculations on Si/Ge core-shell NWs. The electronic band structures are obtained for different core-shell ratio NWs. Our results show that the band offset does exist in core-shell NWs along the [110] direction. The band offset is significantly affected by the quantum confinement from the cylindrical geometry. Comparing with the band offset data of bulk Si/Ge heterojunctions, we find that the quantum confinement effect cannot be ignored. The confinement is strong enough to modify the value of the band offset by a factor of two or three.

Finally, the question is how to get the optimal doping configuration from our results. A larger band offset is preferred for the higher doping efficiency, and better spacial separation of the core and shell states is also preferred because it helps to reduce the scattering rate. As a result, a suggestion that acceptors should be doped at the shell and induce a p-type doping in germanium core NWs is suggested based on our calculated band offset data, because this choice obtains a better spacial separation of impurities and carriers and a larger band offset.
APPENDIX A

THE BAND STRUCTURE OF SI/GE CORE-SHELL NWS

In this Appendix, the whole set of band structures of Si/Ge core-shell NWs are shown. Because we are interested in the energy discretization of the bands induced by confinement, only the bands close to the band gap are plotted.

As shown in Fig. A.1 and Fig. A.2, the bands close to the gap in core-shell NWs clearly show large discretization, while those bands in the pure NWs do not have that clear energy discretization. So these discrete bands are probably confined by the core or shell as shown in Fig. 5.9. In addition, when we exchange the core-shell NWs from silicon core to germanium core, the rough shape of the band structure close to the band gap is also exchanged, which further supports that there is a band offset in core-shell NWs.

There is one more feature in these two figures. We find that the discretization of the conduction bands is more noticeable than that of the valence bands in silicon core NWs. This suggests that the band offset of the conduction bands is larger than that of the valence bands in silicon core NWs. In the same way, the band offset of the valence bands is larger than that of the conduction bands in germanium core NWs. In fact, these two conclusions are confirmed by the later charge density plots in Chapter V.
Figure A.1: The band structure evolution from the pure germanium wire to the pure silicon wire with the silicon core size increasing. The indices in the plots are same as in Table 5.1.
Figure A.2: The band structure evolution from the pure silicon wire to the pure germanium wire with the germanium core size increasing. The indices in the plots are same as in Table 5.1.
APPENDIX B

THE CHARGE DENSITY IN CORE-SHELL NANOWIRES

In this Appendix, we show the details of calculating the band offset in Si/Ge core-shell NWs. Because it is hard to identify the confined and non-confined states by eyes, we have to use a criterion to separate these states. Considering that most of the charge should be within the confined region for a confined state, we apply the following rule to identify the confined state.

In order to simplify the notation, we set the total volume of a unit cell to be 1, and the total charge of a unit cell to be 1. Suppose the charge inside the confined region, whose volume is $V_a$, is $q_a$, then the charge outside the confined region, whose volume is $1 - V_a$, is $1 - q_a$. So the average charge density inside the confined region is

$$n_c = \frac{q_a}{V_a}, \quad \text{(B.1)}$$

and the average charge density outside the confined region is

$$n_{nc} = \frac{1 - q_a}{1 - V_a}. \quad \text{(B.2)}$$

If $n_c$ is more than twice of $n_{nc}$, we believe that the corresponding state is a confined state. Otherwise, the state is not confined. The energy difference between the first confined state and the first non-confined state is defined as the value of the band offset. The idea for this identification is not a strict theoretical justification but based on the consideration of the application of the band offset. Because the weakly confined states have significant extension out of the confined region, which helps to increasing the scattering and worsens the mobility.

All the details are listed in the following tables. The first five tables are of core-shell NWs with the silicon core and germanium shell, and the last five tables are of core-shell NWs with germanium core and silicon shell.
Table B.1: The core-shell NW with a silicon core and germanium shell, which is marked as type I structure in Table 5.1. The critical charge of the confined valence band is 0.92 inside the shell, and the critical charge of the confined conduction band is 0.26 inside the core. As a result, the band offset of valence bands is 0.19 eV, and that of conduction bands is 0.40 eV.

<table>
<thead>
<tr>
<th>band index</th>
<th>charge inside the core</th>
<th>charge inside the shell</th>
<th>transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>545</td>
<td>0.19</td>
<td>0.81</td>
<td>non-confined</td>
</tr>
<tr>
<td>546</td>
<td>0.07</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>547</td>
<td>0.03</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>548</td>
<td>0.04</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>549</td>
<td>0.06</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>550</td>
<td>0.07</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>551</td>
<td>0.59</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>552</td>
<td>0.55</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>553</td>
<td>0.27</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>554</td>
<td>0.21</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>555</td>
<td>0.09</td>
<td>0.91</td>
<td></td>
</tr>
</tbody>
</table>

Table B.2: The core-shell NW with a silicon core and germanium shell, which is marked as type II structure in Table 5.1. The critical charge of the confined valence band is 0.89 inside the shell, and the critical charge of the confined conduction band is 0.33 inside the core. As a result, the band offset of valence bands is 0.21 eV, and that of conduction bands is 0.40 eV.

<table>
<thead>
<tr>
<th>band index</th>
<th>charge inside the core</th>
<th>charge inside the shell</th>
<th>transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>544</td>
<td>0.23</td>
<td>0.77</td>
<td>non-confined</td>
</tr>
<tr>
<td>545</td>
<td>0.07</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>546</td>
<td>0.07</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>547</td>
<td>0.04</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>548</td>
<td>0.03</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>549</td>
<td>0.10</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>550</td>
<td>0.06</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>551</td>
<td>0.66</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>552</td>
<td>0.62</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>553</td>
<td>0.45</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>554</td>
<td>0.29</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>555</td>
<td>0.18</td>
<td>0.82</td>
<td></td>
</tr>
</tbody>
</table>
Table B.3: The core-shell NW with a silicon core and germanium shell, which is marked as type III structure in Table 5.1. The critical charge of the confined valence band is 0.82 inside the shell, and the critical charge of the confined conduction band is 0.47 inside the core. As a result, the band offset of valence bands is 0.19 eV, and that of conduction bands is 0.29 eV.

<table>
<thead>
<tr>
<th>band index</th>
<th>charge inside the core</th>
<th>charge inside the shell</th>
<th>transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>543</td>
<td>0.17</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>544</td>
<td>0.40</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>545</td>
<td>0.07</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>546</td>
<td>0.17</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>547</td>
<td>0.20</td>
<td>0.80</td>
<td>non-confined</td>
</tr>
<tr>
<td>548</td>
<td>0.08</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>549</td>
<td>0.09</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>550</td>
<td>0.12</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>551</td>
<td>0.75</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>552</td>
<td>0.71</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>553</td>
<td>0.65</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>554</td>
<td>0.46</td>
<td>0.54</td>
<td>non-confined</td>
</tr>
<tr>
<td>555</td>
<td>0.65</td>
<td>0.35</td>
<td></td>
</tr>
</tbody>
</table>
Table B.4: The core-shell NW with a silicon core and germanium shell, which is marked as type IV structure in Table 5.1. The critical charge of the confined valence band is 0.72 inside the shell, and the critical charge of the confined conduction band is 0.61 inside the core. As a result, the band offset of valence bands is 0.18 eV, and that of conduction bands is 0.22 eV.

<table>
<thead>
<tr>
<th>band index</th>
<th>charge inside the core</th>
<th>charge inside the shell</th>
<th>transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>543</td>
<td>0.31</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>544</td>
<td>0.48</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>545</td>
<td>0.14</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>546</td>
<td>0.21</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>547</td>
<td>0.28</td>
<td>0.72</td>
<td>non-confined</td>
</tr>
<tr>
<td>548</td>
<td>0.14</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>549</td>
<td>0.14</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>550</td>
<td>0.17</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>551</td>
<td>0.83</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>552</td>
<td>0.82</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>553</td>
<td>0.77</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>554</td>
<td>0.60</td>
<td>0.40</td>
<td>non-confined</td>
</tr>
<tr>
<td>555</td>
<td>0.72</td>
<td>0.28</td>
<td></td>
</tr>
</tbody>
</table>
Table B.5: The core-shell NW with a silicon core and germanium shell, which is marked as type V structure in Table 5.1. The critical charge of the confined valence band is 0.57 inside the shell, and the critical charge of the confined conduction band is 0.75 inside the core. As a result, the band offset of valence bands is 0.26 eV, and that of conduction bands is 0.16 eV.

<table>
<thead>
<tr>
<th>band index</th>
<th>charge inside the core</th>
<th>charge inside the shell</th>
<th>transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>543</td>
<td>0.74</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>544</td>
<td>0.45</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>545</td>
<td>0.59</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>546</td>
<td>0.53</td>
<td>0.17</td>
<td>non-confined</td>
</tr>
<tr>
<td>547</td>
<td>0.14</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>548</td>
<td>0.18</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>549</td>
<td>0.19</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>550</td>
<td>0.23</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>551</td>
<td>0.89</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>552</td>
<td>0.88</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>553</td>
<td>0.86</td>
<td>0.14</td>
<td>non-confined</td>
</tr>
<tr>
<td>554</td>
<td>0.71</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>555</td>
<td>0.82</td>
<td>0.18</td>
<td></td>
</tr>
</tbody>
</table>
Table B.6: The core-shell NW with a germanium core and silicon shell, which is marked as the type I structure in Table 5.1. The critical charge of the confined valence band is 0.26 inside the core, and the critical charge of the confined conduction band is 0.92 inside the shell. As a result, the band offset of valence bands is 0.52 eV, and that of conduction bands is 0.03 eV.

<table>
<thead>
<tr>
<th>band index</th>
<th>charge inside the core</th>
<th>charge inside the shell</th>
<th>transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>543</td>
<td>0.14</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>544</td>
<td>0.22</td>
<td>0.78</td>
<td>non-confined</td>
</tr>
<tr>
<td>545</td>
<td>0.31</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>546</td>
<td>0.30</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td>547</td>
<td>0.48</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>548</td>
<td>0.65</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>549</td>
<td>0.60</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>550</td>
<td>0.55</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>551</td>
<td>0.09</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>552</td>
<td>0.08</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>553</td>
<td>0.21</td>
<td>0.79</td>
<td>non-confined</td>
</tr>
<tr>
<td>554</td>
<td>0.14</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>555</td>
<td>0.19</td>
<td>0.81</td>
<td></td>
</tr>
</tbody>
</table>
Table B.7: The core-shell NW with a germanium core and silicon shell, which is marked as the type II structure in Table 5.1. The critical charge of the confined valence band is 0.33 inside the core, and the critical charge of the confined conduction band is 0.89 inside the shell. As a result, the band offset of valence bands is 0.60 eV, and that of conduction bands is 0.02 eV.

<table>
<thead>
<tr>
<th>band index</th>
<th>charge inside the core</th>
<th>charge inside the shell</th>
<th>transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>543</td>
<td>0.22</td>
<td>0.78</td>
<td>non-confined</td>
</tr>
<tr>
<td>544</td>
<td>0.35</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>545</td>
<td>0.42</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>546</td>
<td>0.48</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>547</td>
<td>0.57</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>548</td>
<td>0.72</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>549</td>
<td>0.64</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>550</td>
<td>0.60</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>551</td>
<td>0.08</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>552</td>
<td>0.07</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>553</td>
<td>0.29</td>
<td>0.71</td>
<td>non-confined</td>
</tr>
<tr>
<td>554</td>
<td>0.17</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>555</td>
<td>0.21</td>
<td>0.79</td>
<td></td>
</tr>
</tbody>
</table>
Table B.8: The core-shell NW with a germanium core and silicon shell, which is marked as the type III structure in Table 5.1. The critical charge of the confined valence band is 0.47 inside the core, and the critical charge of the confined conduction band is 0.82 inside the shell. As a result, the band offset of valence bands is 0.63 eV, and that of conduction bands is 0.03 eV.

<table>
<thead>
<tr>
<th>band index</th>
<th>charge inside the core</th>
<th>charge inside the shell</th>
<th>transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>540</td>
<td>0.34</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>541</td>
<td>0.44</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>542</td>
<td>0.44</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>543</td>
<td>0.49</td>
<td>0.51</td>
<td>non-confined</td>
</tr>
<tr>
<td>544</td>
<td>0.55</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>545</td>
<td>0.59</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>546</td>
<td>0.70</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>547</td>
<td>0.73</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>548</td>
<td>0.75</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>549</td>
<td>0.80</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>550</td>
<td>0.70</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>551</td>
<td>0.13</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>552</td>
<td>0.12</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>553</td>
<td>0.37</td>
<td>0.71</td>
<td>non-confined</td>
</tr>
</tbody>
</table>
Table B.9: The core-shell NW with a germanium core and silicon shell, which is marked as the type IV structure in Table 5.1. The critical charge of the confined valence band is 0.61 inside the core, and the critical charge of the confined conduction band is 0.72 inside the shell. As a result, the band offset of valence bands is 0.60 eV, and that of conduction bands is 0.03 eV.

<table>
<thead>
<tr>
<th>band index</th>
<th>charge inside the core</th>
<th>charge inside the shell</th>
<th>transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>540</td>
<td>0.50</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>541</td>
<td>0.59</td>
<td>0.41</td>
<td>non-confined</td>
</tr>
<tr>
<td>542</td>
<td>0.64</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>543</td>
<td>0.63</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>544</td>
<td>0.64</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>545</td>
<td>0.70</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>546</td>
<td>0.77</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>547</td>
<td>0.78</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>548</td>
<td>0.79</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>549</td>
<td>0.85</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>550</td>
<td>0.76</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>551</td>
<td>0.15</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>552</td>
<td>0.15</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>553</td>
<td>0.52</td>
<td>0.48</td>
<td>non-confined</td>
</tr>
</tbody>
</table>
Table B.10: The core-shell NW with a germanium core and silicon shell, which is marked as the type V structure in Table 5.1. The critical charge of the confined valence band is 0.75 inside the core, and the critical charge of the confined conduction band is 0.57 inside the shell. As a result, the band offset of valence bands is 0.46 eV, and that of conduction bands is 0.02 eV.

<table>
<thead>
<tr>
<th>band index</th>
<th>charge inside the core</th>
<th>charge inside the shell</th>
<th>transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>540</td>
<td>0.74</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>541</td>
<td>0.73</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>542</td>
<td>0.71</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>543</td>
<td>0.75</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>544</td>
<td>0.75</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>545</td>
<td>0.79</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>546</td>
<td>0.85</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>547</td>
<td>0.84</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>548</td>
<td>0.85</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>549</td>
<td>0.89</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>550</td>
<td>0.82</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>551</td>
<td>0.18</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>552</td>
<td>0.18</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>553</td>
<td>0.72</td>
<td>0.28</td>
<td>non-confined</td>
</tr>
</tbody>
</table>
REFERENCES


[40] I.V. Abarenkov and V. Heine, Phil. Mag. 11, 387 (1965).


VITA

Li Yang was born in Shimian county, Sichuan, China on November 9th, 1974. He received his B.S. in 1997 and M.S. from Beijing Normal University, China. He came to U.S.A. in 2000, studying at Georgia Institute of Technology. Since May 2002, he has been a research assistant at the school of physics. His work has concentrated on the first-principles calculations on the electronic structure of nanowires. His immediate postgraduate plan involves the postdoc work in the Lawrence Berkeley National Laboratory (Berkeley Lab) and as a visiting postdoc in the physics department of UC Berkeley. His scientific publications during the PhD study include:


