



**KTH Industrial Engineering
and Management**

OCTOBER 29, 2014

RONGZHEN CHEN

Doctoral Thesis

School of Industrial Engineering and Management, Department of
Materials Science and Engineering, KTH, Sweden, 2014

ISRN KTH/MSE-12/09-SE+AMFY/AVH
ISBN 978-91-7501-313-8

Materialvetenskap
KTH
SE-100 44 Stockholm
Sweden

Akademisk avhandling som med tillstånd av Kungliga Tekniska Högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen in materialvetenskap ??? 2014 kl ???? i ??, Materialvetenskap, Kungliga Tekniska Högskolan, Brinellvägen 23, Stockholm.

© Rongzhen Chen, ???, 2014

Tryck: Universitetsservice US AB

Abstract

Preface

List of included publications:

- I Parameterization of $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ ($x = 0, 0.5$, and 1) energy bands
R. Chen and C. Persson, *Thin Solid Films* **519**, 7503 (2011).
- II Band-edge density-of-states and carrier concentrations in intrinsic and *p*-type $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$
R. Chen and C. Persson, *J. Appl. Phys.* **112**, 103708 (2012).
- III Dielectric function spectra at 40 K and critical-point energies for $\text{CuIn}_{0.7}\text{Ga}_{0.3}\text{Se}_2$
S.G. Choi, R. Chen, C. Persson, T.J. Kim, S.Y. Hwang, Y. D. Kim, and L. M. Mansfield, *Appl. Phys. Lett.* **101**, 261903 (2012).

My contribution to the publications:

- Paper I:** modeling, analysis of result, literature survey; the manuscript was written jointly.
- Paper II:** modeling, analysis of result, literature survey; main part of the manuscript was written.
- Paper III:** all calculations, analysis of the theoretical part, part of literature survey; the manuscript was written jointly.

Publications not included in the thesis:

Book chapter:

- IV Electronic structure and optical properties from first-principles modeling
C. Persson, R. Chen, H. Zhao, M. Kumar, and D. Huang, Chapter in "Copper zinc tin sulphide-based thin film solar cells", edited by K. Ito (John Wiley & Sons, ??CITY??, 2014??); in preparation.

International conference contributions:

- V Band structure and optical properties of CuInSe_2
R. Chen and C. Persson, *Advanced Materials Research Journal* **894**, 254 (2014).
4th Int. Conf. on Adv. Mater. Res (ICAMR-4), Macao, China, 23–24 Jan. 2014.
- VI Electronic modeling and optical properties of $\text{CuIn}_{0.5}\text{Ga}_{0.5}\text{Se}_2$ thin film solar cell
R. Chen and C. Persson, *J. Appl. Math. & Phys.* **2**, 41 (2014).
Conf. on New Adv. Cond. Matter Phys. (NACMP 2014), Shenzhen, 14–16 Jan 2014.

Contents

Preface	v
Contents	vii
1 Introduction	1
1.1 Solar cells	3
1.1.1 Single-junction solar cells	3
1.2 Solar cell materials	8
1.2.1 Crystalline silicon solar cells	9
1.2.2 Gallium arsenide	10
1.2.3 Thin film materials	10
1.3 CuIn _{1-x} Ga _x Se ₂ (CIGS) materials	11
1.3.1 Crystal structure	11
1.3.2 Optical properties and defects in the CIGS	12
1.3.3 CIGS solar cell structure	13
2 Theory	15
2.1 Electronic structure calculations	15
2.1.1 The quantum many-body problem	15
2.1.2 The Born-Oppenheimer approximation	16
2.1.3 Hartree, Hartree-Fock approximation and density functional theory	18
2.1.3.1 Hartree approximation	19
2.1.3.2 Hartree-Fock approximation	21
2.1.3.3 Density functional theory	23
2.1.3.4 Kohn-Sham equation	25
2.1.3.5 The exchange-correlation potential	27
2.1.4 Solving the secular equation	28
2.1.5 Eigenvalue problem	30
2.2 Full-potential linearized augmented plane wave method	31
2.2.1 Introduction	31
2.2.2 Wavefunction	32
2.2.2.1 Augmented plane wave method	32
2.2.2.2 Linearized augmented plane wave method	33

2.2.2.3	Linearized augmented plane wave method plus local orbitals	34
2.2.2.4	Augmented plane wave method plus local orbitals	34
2.2.3	Effective potential	35
2.3	Dielectric function	35
2.4	Spin-orbit coupling	36
2.4.1	Dirac equation	36
2.4.2	Derivation of spin-orbit coupling	36
3	Results and discussion	39
3.1	Parameterization of energy bands for CIGS	39
3.1.1	Parameterization method	39
3.1.2	Non-parabolicity properties	41
3.1.3	Application of the full band parameterization	45
3.2	Calculations of dielectric function for CIGS	53
4	Summary of the publications	57
5	Concluding remarks and future work	58
Acknowledgements		59
6	Appendices	60

Chapter 1

Introduction

With the increasing of energy consumption, more and more energy or power is needed. According to the statistical review of world energy on 2014 (**Fig. 1.1**), the required energy is mainly satisfied by the fossil fuels (mainly coal, petroleum and natural gas), with a market share of around 87%. The total energy consumption is between 12000 and 13000 million tonnes oil equivalent (MTOE), which is equivalent to around 15 terawatts. Normally lightbulbs in our homes consume around between 50 to 100 watts of energy, 1 terawatt implies 10 billion of the 100 watts lightbulbs are lighted at the same time. Unfortunately, the fossil fuels is very limited energy and non-renewable resources, one day which is not far from now it will be dissipated due to the energy consumption growth.

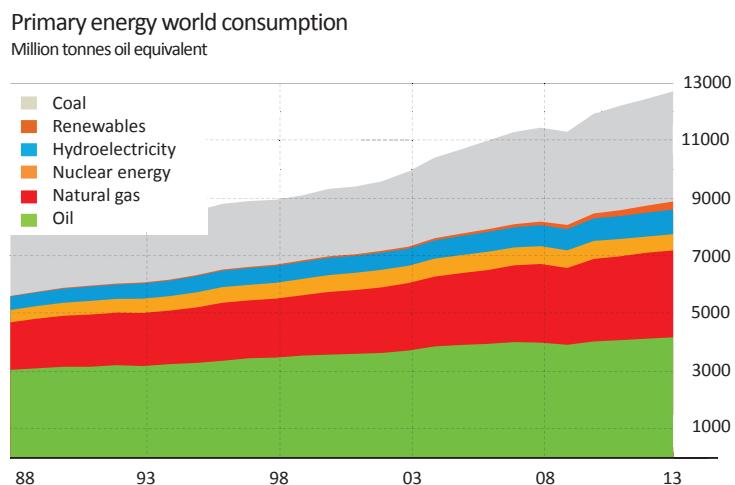


Figure 1.1. Data is from BP statistical review of world energy 2014.

By the year of 2050, the total world energy consumption will double. Therefore, it is urgent to explore more sustainable and environmentally friendly energy sources. From **Fig. 1.1**, one will notice that renewable energy (mainly solar energy, wind power, and

geothermal energy) in 2014 accounts for around 2% of energy consumption globally. It is important to focus on the renewable energy research from a long term point of view. The solar energy technologies are one of the hot topic among the renewable energy research considering the point of CO₂ free, reliable energy supply, no cooling water requirement and operation in silence.

The solar energy technologies are the way to produce electricity from the light of the sun. Sunlight is a portion of the radiation by the sun, like infrared, visible, and ultraviolet light. The spectrum of the sun is close to the spectrum of black body with a temperature of about 6000 K (Fig. 1.2). In the field of photovoltaics (PV), solar spectrum is represented by air mass (AM) which defines the direct optical path length through the Earth's atmosphere. The AM1.5 and AM0 are important: AM1.5 is the air mass at a solar zenith angle of 48.19 degree, and AM0 mean the solar spectrum outside of the atmosphere. Generally, the AM1.5 is seen as the reference spectrum in PV field. In Fig. 1.2, the absorption in the atmosphere is quite strong by gasses, dust and aerosols, as well as the scattering of light from air molecules.



Figure 1.2. Solar irradiance spectrum

There are mainly three kinds of solar energy technologies. The first one is solar thermal, which utilizes the flat sunlight collector plates to harness the energy from sunlight to heat water for use in industries, homes, and pools. Therefore, the solar thermal collectors do not convert sunlight to electricity directly, but transfer the energy to heat up the water instead. The advantage is the conversion efficiency is relatively higher. The second one is solar chemical, which takes advantage of solar energy by absorbing sunlight in a chemical reaction. However, the conversion efficiency is quite low. The last one is solar photovoltaics (solar cell), which is the way to utilize solar panels to convert sunlight into electricity. The installation is easier, occupy less space and less maintenance compared with solar thermal. The conversion efficiency is higher than solar chemical. However, all the three solar energy technologies are environmentally friendly.

1.1 Solar cells

In the worldwide, the conversion efficiency in all different types of solar cell is improved remarkably. From **Fig. 1.3**, one can notice that the highest efficiency for multijunction cells, crystalline silicon cells, thin-film technologies and new emerging cells are around 44.7%, 27.6%, 23.3% and 19.3%(??by Yang Yang not this figure so far??), respectively. Therefore, the solar cell is a very important and promising way to produce the renewable energy.

Multijunction cells are the cells which contain multi *p-n* junctions (or subcells) which have different band gap for each *p-n* junction. Therefore, different wavelengths of light from the Sun are absorbed for each of junction. For example, wider band gap junction is at the front of the cell, which can absorb the photons with high energy; the junction with low band gap can absorb the photos with lower energy. Therefore, the conversion efficiency is higher than single *p-n* junction, for example, the maximum conversion efficiency is 44.7% by Soitec using four-junction or more in **Fig. 1.3**. Crystalline silicon cells are the most widely utilized in the photovoltaic industries, which built the solar cells using crystalline silicon (c-Si). It has two types in the crystalline silicon photovoltaics: mono-crystalline silicon and multi-crystalline silicon. The crystalline silicon cells have high efficiency, for example, the maximum conversion efficiency is 27.6% with concentrator by Amonix and maximum 25.6% without concentrator by Panasonic in the **Fig. 1.3**. Thin-Film solar cells are the cells which are made by depositing one or several thin layers, which allows the cells to be rather flexible and resulting in lower weight. The maximum conversion efficiency is lower than crystalline silicon today, which has the maximum conversion efficiency 23.3% using CIGS with concentrator by NREL and 21.7% without concentrator by the center for solar energy and hydrogen research (ZSW) in Stuttgart in **Fig. 1.3**. The Emerging PV in **Fig. 1.3** represents the newest ways to create electricity from sunlight and potentially with higher conversion efficiency, such as perovskite cells, for which the maximum conversion efficiency is 19.3% by the group Yang Yang at the University of California, Los Angeles. (update the above figure until May, 2014) Perovskite cells jump into the world of solar cells only in 2009, and the conversion efficiency is improved remarkable within 5 years. Certainly, the search and optimisation of alternatively solar cell materials is still an ongoing and active area today.

1.1.1 Single-junction solar cells

The *p-n* junction is the fundamental building block of solar cells. The single *p-n* homojunction will be explored in this section.

We start from the separate *n*-type material and *p*-type material at room temperature (assume that all the donor (acceptor) atoms are positively (negatively) ionized at room temperature). In **Fig. 1.4**, the left one shows the *n*-type and *p*-type materials. The *n*-type material has many free negatively charged electrons which can move freely in the material, and there are numbers of positively charged immobile donor ions as well.

Best Research-Cell Efficiencies



Figure 1.3. Best research-cell efficiencies. Figure is from National Renewable Energy Laboratory (NREL), Golden, Colorado.

Similarly, the *p*-type material has many free positively charged holes which can move freely in the material, and there are numbers of negatively charged immobile acceptor ions as well. However, the material is still neutral in both *n*-type and *p*-type. On the right in Fig. 1.4, the corresponding Fermi levels are shown. The Fermi level (E_{nf}) is closer to conduction band minimum for *n*-type material due to the many free negatively charged electrons. Conversely, the Fermi level of *p*-type material (E_{pf}) is closer to the valence band maximum to the free positively charged holes.

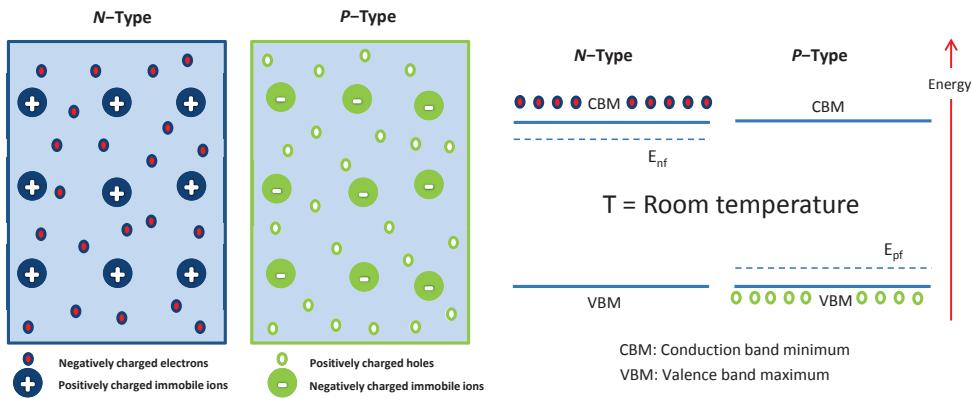


Figure 1.4. Left: Doped (*n*-type and *p*-type) materials in dark at room temperature. Right: Energy band diagram of separated *p*-*n* homojunction in dark at room temperature for two-level model. Assume that all the donor (acceptor) atoms are positively (negatively) ionized.

If the *n*-type and *p*-type materials are joined, the free electrons (holes) in *n*-type (*p*-type) material will diffuse into *p*-type (*n*-type) material due to the lower concentrations of electrons (holes) (Fig. 1.5). In the region which is near the interface between *n*-type and *p*-type materials, the ionized donor and acceptor ions create a "build in" electric field which points from the *n*-type material to the *p*-type material. This can cause the drift of carriers in the opposite direction. The "build-in" electric field will force the electrons (holes) back into the *n*-type (*p*-type). At certain point, the whole material will reach a stable equilibrium due to the achieved balance between diffusion and drift. Formation of the "build-in" electric field is rather important for the solar cells, even though there is no current in the material so far. In the following text, the region which forms the "build-in" electric field is also called space charge region (SCR). The different Fermi levels for *n*-type and *p*-type materials are equal at the stable equilibrium. Therefore, the energy bands bend over and create a potential barrier near the junction (Right in Fig. 1.5). Finally, there is an internal potential V_{bi} in the junction, which will block the diffusion.

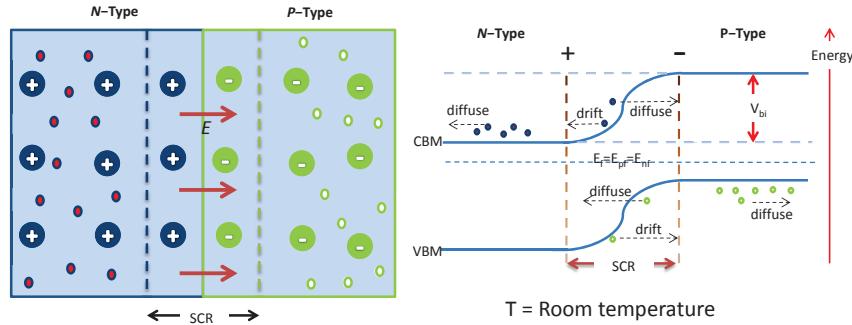


Figure 1.5. Left: The p - n homojunction in dark at room temperature. Right: Energy band diagram of a p - n homejunction at the equilibrium in dark at room temperature for two-level model.

The p - n junction cell with and without the illumination is discussed in **Fig. 1.6** and **Fig. 1.7**. If there was a wire with certain resistance connecting the n -type and p -type, there is no current in the wire under the condition of dark (no illumination). However, if the light shines on the cell or component, a current will be generated from the p -type to the n -type side (conventional current). Because the electrons from valence bands (VBs) goes to conduction bands (CBs), which can generate pairs of electron-hole. At the same time, the recombination of paired electron-hole occurs. The rate of generation is faster than that of recombination, therefore, net generation occurs. Apparently, there are three regions in the whole junction cell where the electrons goes from VBs to CBs, the n -type region, the p - n junction, and the p -type region. In the either n -type or p -type region (especially, region which is far away SCR), the electron-hole pairs can not remain long time, it is most probable that electrons will jump down from CBs to VBs again. However, the electron-hole pairs will be separated in the p - n junction region due to the "build-in" electric field, therefore, the current will be generated. Actually, the electron-hole pairs in the either n-type and p-type (especially, for them which are near SCR) also have the chance to diffuse into the SCR, it will contribute the generation of current or reduce the current.

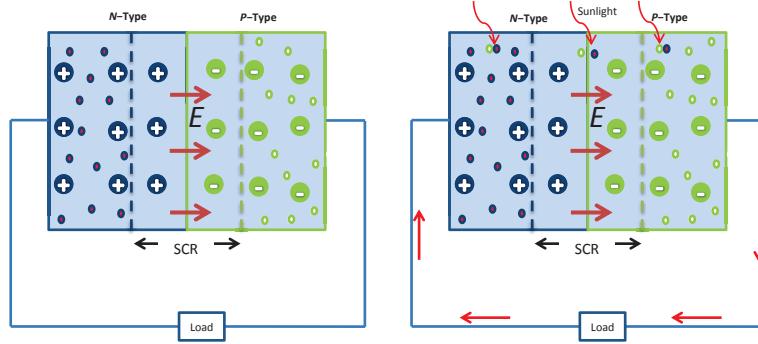


Figure 1.6. The p - n homojunction under illumination with load at room temperature for two-level model.

In Fig. 1.7, the SCR becomes more "smooth" due to the extra load, such as lightbulb, which is equivalent to apply external potential. The stabilized Fermi level at the stable equilibrium splits under illumination. The chemical potential $\nabla\mu = E_{nf} - E_{vf}$ is created, which is considered as the electron charge times the voltage across the device. The generation and recombination by impuriteis are not analyzed in here.

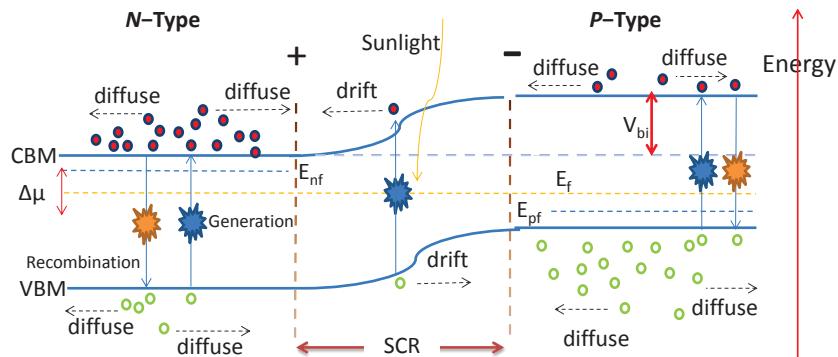


Figure 1.7. Energy band diagram of a p - n homojunction under illumination with load.

The I-V characteristics is defined in Fig. 1.8 with some important parameters of the solar cells. The V_{oc} and I_{sc} are the open circuit voltage and short circuit current, respectively. They are the maximum voltage and maximum current from the solar cells. The V_{mp} and I_{mp} are the voltage and current which will yield the maximum power. The maximum power generated by the solar cells is $P_{out} = V_{mp} \times I_{mp}$, that is the rectangle bounded by the dashed lines in the Fig. 1.8.

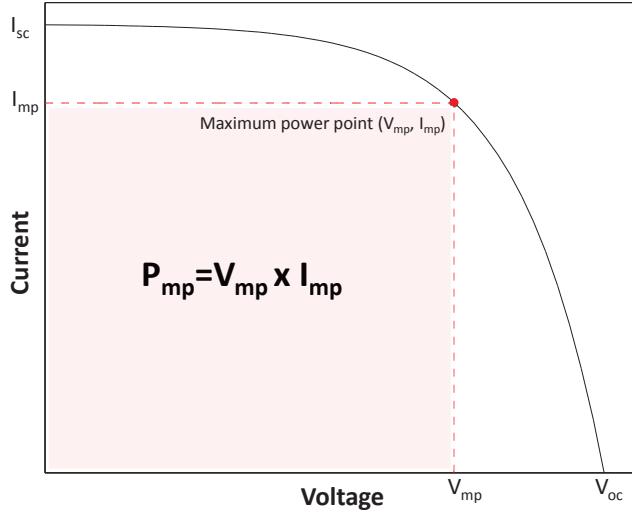


Figure 1.8. I-V characteristics

The fill factor (FF) and the power conversion efficiency (η) are often represented the solar cell performance:

$$FF = \frac{P_{out}}{V_{oc} \cdot I_{sc}} = \frac{V_{mp} \cdot I_{mp}}{V_{oc} \cdot I_{sc}} \quad (1.1)$$

$$\eta = \frac{P_{out}}{P_{in}} = \frac{FF \cdot V_{oc} \cdot I_{sc}}{P_{in}}. \quad (1.2)$$

Here, P_{in} is the incident photon power per second. The conversion efficiency of the solar cells is proportional to FF, V_{oc} , and I_{sc} . There are several aspects which will affect the conversion efficiency. V_{oc} is directly proportional to the band gap of the material, I_{sc} is proportional to the number of absorbed photons. When the band gap is decreased, the more of the spectrum is absorbed. However, the V_{oc} will be reduced in this case, more importantly, the excess energy of photons is lost due to the thermalization in the cells. When the band gap is increased, transparency losses from the photons with energy lower than band gap. There is more detailed analysis of conversion efficiency in the Ref[?? Principles of solar energy conversion].

1.2 Solar cell materials

In 1839, the French physicist A. E. Becquerel revealed the photovoltaic effect for the first time. Charles Fritts built the first solid state photovoltaic (PV) cell using semiconductor selenium in 1883. It is not until 1941 that the first silicon-based solar cell was demonstrated. Today, there are many different types of solar cell materials. The reason why the best solar cell material is not realized yet is that it is expected to be not only

high efficiency but also environmentally friendly and low cost. It requires not only that the growth and manufacturing process of solar cell materials shall be cheaper, but also that the devices shall have longer application life, moreover the raw material should be abundant and non-toxic as well. In this section, four main solar cell materials are discussed briefly: silicon (Si), gallium arsenide (GaAs), cadmium telluride (CdTe), copper indium gallium diselenide ($\text{Cu}(\text{In}, \text{Ga})\text{Se}_2$).

Potential solar cell materials need to fulfill several properties, such as large absorption coefficient and a band gap energy E_g between 0.7 to 2.0 eV. Under these conditions, there are quite many materials satisfying the requirements. However, some other properties are needed to be considered as well, such as cost and environmental safety. Thereby, only part of them are suitable to produce in reality. In Fig. 1.9, several solar cell materials are illustrated.

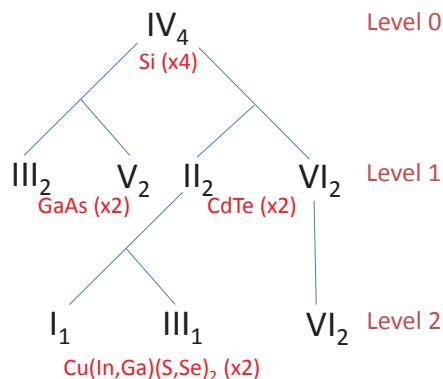


Figure 1.9. Tree of tetragonal bonded semiconductor, the roman numerals mean the group numbers in the chemical element periodic table, and the subscript implies the number of elements.

In Fig. 1.9, the formation of tetragonal semiconductors is considered as a series of cation mutations where the total number valence electrons is the same and keep the charge neutral. For example, group number IV element silicon (level 0) with four 4⁺ ions is equivalent to two 3⁺ ions and two 5⁺ ions, such as GaAs (level 1). It is also equivalent to two 2⁺ ions and two 6⁺ ions, such as CdTe (level 1). The $\text{Cu}(\text{In}, \text{Ga})(\text{S}, \text{Se})_2$ can be generated applying the same process on II element on the level 1. This method was suggested by Goodman and Pamplin.

1.2.1 Crystalline silicon solar cells

The solar cell based on silicon dominates the solar power world today, which accounts for more than 90% of the total PV market. This kind of solar cell takes advantage of different forms of silicon, that is, monocrystalline silicon and polycrystalline silicon. The success of Si is due to a number of reasons. Over 90% is composed of silicate minerals in the crust of earth, which also yields huge available amounts of Si. Moreover, it

has higher conversion efficiency, and it is also proved that it has excellent stability and reliability under the outdoor condition. However, Si also have drawbacks. It has an indirect band gap and hence it has a lower optical absorption coefficient. In order to absorb the incident sunlight fully, it requires to thicker Si (around 40 μm) to absorb the sunlight. Crystalline Silicon have to be high quality and defect free in order to avoid losing the carriers before collection. Last but not least, it is costly to purify the Si from silicate minerals, which really limits the cost reduction potential of wafer-based silicon technology.

However, the solar cells based crystalline silicon technology is still leading the market of solar cell since many companies are trying to lower the cost of the whole process.

1.2.2 Gallium arsenide

Gallium arsenide (GaAs) has a zinc blende crystal structure with a direct band gap around 1.5 eV at room temperature. Some electronic properties of GaAs are superior to Si, such as higher electron mobility, higher saturated electron velocity, absorb sunlight more efficiently due to the direct band gap. The optimum band gap for the single junction solar cell is suggested around 1.3 eV by theoretical calculation from Henry (1980) who modified the original Shockley-Queisser limit. Therefore one of the most applications of GaAs is solar cells. GaAs has been extensively researched since the 1950s, and the first GaAs solar cells were established in 1970 by the Zhores Alferov's team. Today, the conversion efficiency for single function solar cell based on GaAs is around 28.8%. However, it is more difficult to grow and the solar cell component has higher price in comparison with Si. Many researches are focusing on how to reduce the price, and the main application solar cell based on GaAs is in the space application. At last, the arsenic toxicity should be considered as well.

The conversion efficiency for four-junction GaInP/GaAs//GaInAsP/GaInAs concentrator solar cells is reached 44.7% by Soitec on March 2014.

1.2.3 Thin film materials

Thin film solar cells have several thin films with the total thickness less than 10 μm . The cost can potentially be lower since the less materials are utilized to make thin film solar cells. The development of thin film solar cell was started since 1970s. Currently, the conversion efficiency for the single junction already surpassed 20%, even though it is still not as high as the solar cells based on crystalline silicon. Three different thin film materials are discussed in this section: amorphous silicon (a-Si), cadmium telluride (CdTe), and copper indium gallium diselenide (CIGS).

Amorphous silicon solar cells are the first thin film solar cell material which reach the large-scale production. It has higher absorption coefficient than crystalline silicon, therefore, the thickness can be less than 1 μm . The main disadvantages a-Si solar cells is the lower efficiency, the actual conversion efficiency for the commercial single junction

solar cells based on a-Si is between 4% to 8%. This limits the development of a-Si thin film solar cells. A-Si solar cells are suited to the situation which requires low cost over high efficiency.

CdTe was first reported in the 1960s. However, it is not developed rapidly until in the early 1990s. CdTe has a number of advantages as an absorber. It has higher absorption coefficient. The band gap is around 1.45 eV, which is very near the optimum value for single-junction solar cells. Moreover, the manufacturing process is easier to control, which results in the cost of manufacturing is low. Moreover, the commercial modules already reach the efficiency of 16%. However, an important question is needed to be considered in order to large-scale CdTe manufacturing: cadmium toxicity and tellurium availability.

CIGS are direct band gap semiconductors with high optical absorption coefficients. It is seen as the most promising solar cell material for the near future. It is always employed in a heterojunction structure, mainly it is with the thinner *n*-type CdS layer. The conversion efficiency of CIGS reached up to 20% in the laboratory cell. The interesting part is that it can be alloyed by the ratio of Ga/(Ga+In), and the band gap can be tuned along with that. The band gap is between 1.0 eV to 1.7 eV for this alloy. CIGS does not contain any toxic element.

1.3 CuIn_{1-x}Ga_xSe₂(CIGS) materials

CIGS material is a chalcopyrite-type material, which is considered to be one of the most promising thin film solar cell material. The direct band gap is from around 1.0 eV to 1.7 eV by alloying Ga in the CuInSe₂, and the conversion efficiency in laboratory already surpassed 20%. CuInSe₂ was first synthesized by Hahn in 1953. It was first exploited as an absorber material in a single crystal solar cell in 1974, which is based on CuInSe₂ and CdS. The conversion efficiency is around 5%. The first thin film solar cells based on CuInSe₂ and CdS was invented by Kazmerski. During 1980s, Boeing Corporation did much research on the thin film polycrystalline CIGS solar cells. To date, the highest conversion efficiency in lab situation for the solar cells based on CIGS is between 20% and 21% by alloying with Ga.

1.3.1 Crystal structure

The crystal structure of CIGS can be derived from the zinc blende crystal structure of zinc selenide (ZnSe). In the Fig. 1.10, the crystal structures of ZnSe and CIGS are presented. The elements Zn are replaced by Cu and In or Ga elements in the zinc blende of ZnSe. It requires to double the unit cell in the z-direction. Because the bond strength and lengths between Cu-Se and In-Se or Ga-Se are different, therefore, the lattice parameter c is not exact 2a normally.

Chalcopyrite CIGS (with $x = 0.0$ and 1.0) has the space group D_{2d}^{12} ($I\bar{4}2d$; space group no. 122). The conventional unit cell has four copper atoms on the Wyckoff positions $4a$, four indium/gallium atoms on position $4b$, and eight selenium atoms on the $8d$ position. The cation positions have all S_4 point-group symmetry, and Se have C_2 symmetry. The Se $8d$ position is fully defined with the position (x, y, z) , and each anion Se-atom has two inequivalent bonds $\delta X-\text{Se}$ to the cations $X = \text{Cu}$ and In/Ga . For the alloy of CIGS (in this work, $x = 0.5$ with 50% In and 50% Ga), the structure is chosen so that each Se atom bonds to two Cu atoms, one In and one Ga atom. The space group is S_4^2 ($I\bar{4}$; space group no. 82).

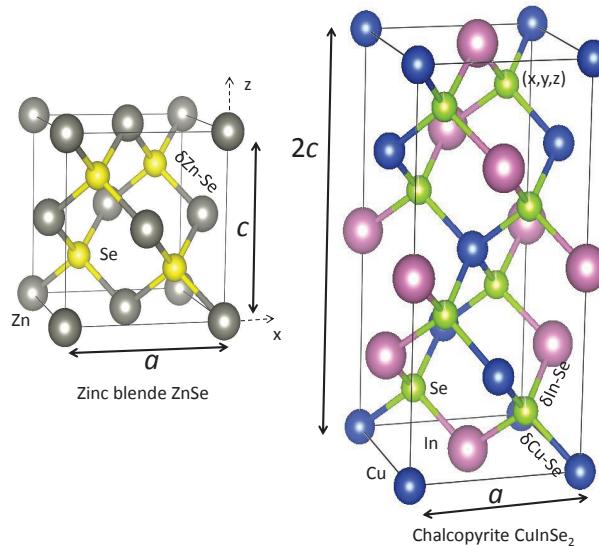


Figure 1.10

1.3.2 Optical properties and defects in the CIGS

CuInSe_2 has a direct band gap around 1.0 eV, and the absorption coefficient is relatively higher than Si due to direct band gap. The quaternary CIGS alloy will be available by alloying Ga element, while the band gap is tuned as well from 1.0 eV to 1.7 eV. The high absorption coefficient makes the CIGS material possible to be an absorber for the thin film solar cells. The band gap can be approximated by the function of Ga content (x):

$$E_g(x) = 1.0 + 0.564x + 0.116x^2 \quad (1.3)$$

Alloying the Ga element will decrease the electron affinity of CIGS, which will make the conduction band upward shift, however, the valence band remain the same position. This also explains the reason why the band gap increases with more Ga element in the CIGS material. An overview properties of CIGS material is described in **Table. 1.1**

Properties of CuInSe ₂ and CuGaSe ₂		
Properties	CuInSe ₂	CuGaSe ₂
Space group	D_{2d}^{12} (I-42d), no. 122	D_{2d}^{12} (I-42d), no. 122
Lattice constants (Å)	$a = b = 5.78, c = 11.55$ $a = b = 5.78^d, c = 11.64^d$	$a = b = 5.61, c = 11.00$ $a = b = 5.61^e, c = 11.02^e$
Wyckoff positions	Cu:4a, In:4b, Se:8d	Cu:4a, Ga:4b, Se:8d
Direct band gap (eV)	Eg = 1.04	Eg = 1.68
Effective masses on Γ point (m_0)	Electrons: 0.08 Holes(heavy): 0.72 Electrons: 0.08 ^a Holes(heavy): 0.23 ^a	Electrons: 0.14 ^b Holes(heavy): 1.2 ^c Electrons: 0.13 ^a Holes(heavy): 0.32 ^a
Main intrinsic defects	<i>p</i> -type: V _{Se} ; In _{Cu} <i>p</i> -type: V _{Cu} ; Cu _{In}	<i>n</i> -type: V _{Se} ; Ga _{Cu} <i>p</i> -type: V _{Cu} ; Cu _{Ga}
Crystal field splitting Δ_{cf} (eV) [300K]	0.006, 0.007 ^a	-0.09, -0.104 ^a
Spin-orbit splitting Δ_{so} (eV) [77K]	0.23, 0.193 ^a	0.231, 0.210 ^a
Dielectric constants $\varepsilon(0)$	13.6 ^d , 15.7 ^d	11.0 ^d , 8.5 ^d
Melting temperature (K)	1259	1310 – 1340
Thermal expansion coefficients [300K] 1/K	a axis: 11.23×10^{-6} c axis: 7.90×10^{-6}	a axis: 13.1×10^{-6} (landolt) c axis: 5.2×10^{-6} (landolt)
Thermal conductivity [273K] W/(cm*K)	0.086	0.129

Table 1.1. most of the data from Semiconductor Basic data [?],^aClas Persson [4]

CIGS is a nonstoichiometric compound with the deviations from stoichiometry in several percentage range. The high quality thin film solar cells mainly employ Cu-poor (Cu: 22.5–24.5%) high offstoichiometric CIGS absorber. The main native defects in CIGS include <2V_{Cu}, In_{Cu}>, <Cu_{In}, In_{Cu}>, V_{Cu}, In_{Cu}, Cu_{In} V_{Se} and Ga_{Cu}. V_{Cu} is the most important native defects in CIGS due to their low formation energies. Therefore, CIGS can be grown *p*-type easily with the condition of Cu-poor (V_{Cu}), which can be grown *n*-type under the condition Se-poor or In-rich. There are some extrinsic divalent cation donors as well, such as Zn_{Cu}, Cd_{Cu} and Cl_{Se}. The formation energy for them is relatively low for CIS and CGS. Therefore, CIS is possbile to be *n*-type material. However, CGS is not possible to be *n*-type under equilibrium conditions, because the low formation energy of V_{Cu} limits the possibility of achieving electronic *n*-type character, especially in Ga-rich CIGS. It is maybe also explained why the best solar cell is with the Ga content of 30% (x = 0.3), however, the band gap energy of the CIGS suggests that the optimum solar cell conversion efficiency is obtained with between x = 0.5 and x = 0.7.

1.3.3 CIGS solar cell structure

The solar cells device based on CIGS is a heterojunction device, which normally has five thin film layers with different functional properties. A schematic of a conventional device structure is shown in Fig. 1.11

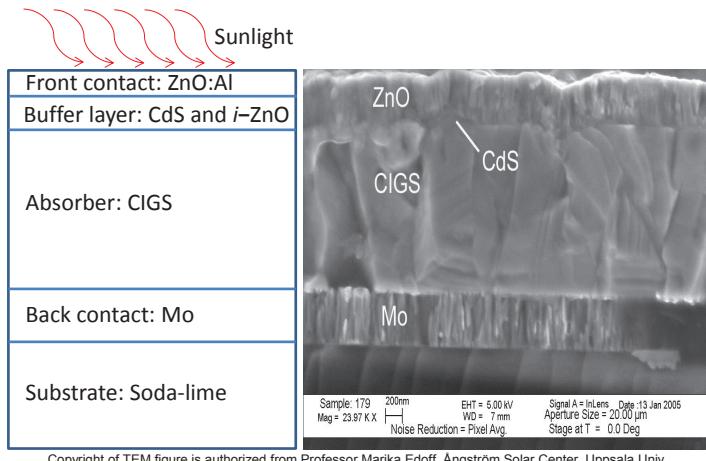


Figure 1.11. Structure of CIGS solar cell device.

Substrat is on the bottom, and there are mainly three kinds of substrat: soda-lime glass, metal and polyimide. The most common substrate is the one based on soda-lime glass containing sodium (Na) with thickness 1 mm to 3 mm. The Na will improve the efficiency and reliablity of the solar cells as well as process tolerance. The molybdenum works as back contact due to its low resistivity and stability at high temperature with thickness around 500 nm. The most important part of the device is the *p*-type absorber layer: CIGS, which is dopped by intrinsic defects and around 1500 -- 2000 nm. The *n*-type buffer layer CdS is on the top of CIGS, which is around 60 nm. The intrinsic zinc oxide (*i*-ZnO) and *n*-type ZnO layer are followed, and it works as window layer. The *i*-ZnO is to avoid the damage of the CIGS and CdS from sputtering damage when de-poziting the ZnO:Al window layer, and the *n*-type ZnO is dopped by the aluminum (Al) in order to get higher conductivity. This CIGS/CdS/ZnO strucutre was optimized to improve the cell performance.

Chapter 2

Theory

2.1 Electronic structure calculations

2.1.1 The quantum many-body problem

A solid material contains a huge number of atoms (around 10^{23} cm^{-3}), and each atom is constructed by a nucleus and electron(s). According to the quantum mechanics principles, all the properties of any system are known if one can figure out a way to solve the quantum many-body Schrödinger equation exactly. In this thesis, the time-independent many-body Schrödinger equation is only considered, which is given as

$$H^{en}\Psi^{en}(\{\mathbf{r}_i, \mathbf{R}_I\}) = E^{en}\Psi^{en}(\{\mathbf{r}_i, \mathbf{R}_I\}). \quad (2.1)$$

Here, term with the superscript "en" implies that this term is related with electrons and nuclei. $\Psi^{en}(\{\mathbf{r}_i, \mathbf{R}_I\})$ is the exact wavefunction, \mathbf{r}_i and \mathbf{R}_I stands for coordinators of electron and nucleus, respectively. E^{en} is the system total energy. H^{en} is the Hamiltonian which is defined as

$$\begin{aligned} H^{en} = & - \sum_i^{Ne} \frac{\hbar^2}{2m_e} \nabla_i^2 - \sum_I^{Nn} \frac{\hbar^2}{2m_I} \nabla_I^2 - \sum_i^{Ne} \sum_I^{Nn} \frac{Z_I e^2}{4\pi\varepsilon_0 |\mathbf{r}_i - \mathbf{R}_I|} \\ & + \frac{1}{2} \sum_i^{Ne} \sum_{j \neq i}^{Ne} \frac{e^2}{4\pi\varepsilon_0 |\mathbf{r}_i - \mathbf{r}_j|} + \frac{1}{2} \sum_I^{Nn} \sum_{J \neq I}^{Nn} \frac{Z_I Z_J e^2}{4\pi\varepsilon_0 |\mathbf{R}_I - \mathbf{R}_J|}. \end{aligned} \quad (2.2)$$

Here, the indices i, j are used for the electrons, and I, J are used for atomic nuclei. Z_I implies the charge of the I :th nucleus. m_I denotes the nuclear mass, m_e is the electron mass, and ε_0 is vacuum permittivity.

In atomic units, the reduced Planck constant \hbar , the electron mass m_e , the Bohr radius a_0 , and the electron charge e are equal to 1. The Bohr radius is given by the formula $a_0 =$

$\hbar/(mc\alpha)$. Here, α is the fine structure constant ($\alpha = e^2/(4\pi\varepsilon_0 c\hbar)$). Therefore, the velocity of light in atomic units is $c = 1/\alpha$ and $e^2/(4\pi\varepsilon_0) = 1$. In atomic units, **Eq. 2.2** is given as

$$\begin{aligned} H = & - \sum_i^{N_e} \frac{\nabla_i^2}{2} - \sum_I^{N_n} \frac{\nabla_I^2}{2M_I} - \sum_i^{N_e} \sum_I^{N_n} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} \\ & + \frac{1}{2} \sum_i^{N_e} \sum_{j \neq i}^{N_e} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \frac{1}{2} \sum_I^{N_n} \sum_{J \neq I}^{N_n} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}. \end{aligned} \quad (2.3)$$

Here, M_I is the mass of the i :th nucleus in atomic units. The first and second terms are the kinetic energy operator of the electrons and nuclei, respectively. The other terms are Coulomb interactions between electrons and nuclei, electrons and electrons and nuclei and nuclei in sequence.

One can not solve **Eq. 2.1** exactly at present since there are enormous amount of atoms to calculate in reality. More importantly, the exactly form of the wavefunction is unknown. To approximate the exact solution, one divides it into three different levels generally: the first level is the Born-Oppenheimer approximation; the second level is Hartree, Hartree-Fock (HF), density functional theory (DFT) and Kohn-Sham (KS) equation; the last level is to solve the secular equation which is an equation that is solved to find the eigenvalue of matrix.

2.1.2 The Born-Oppenheimer approximation

Eq. 2.1 can be approximated in order to solve it. The first attempt is to separate the wavefunction of electron and nucleus. Because the Schrödinger Hamiltonian in **Eq. 2.3** has a couple term between the electron and nucleus, thereby one can not do that simply. On the other hand, there exists a small value $1/M_I$ in **Eq. 2.3**, which is part of the nucleus kinetic energy operator term. The reason is that the mass of nucleus is much larger than that of electron, therefore the nuclei are treated as fixed. This indicates that the electrons are seen as interacting under both the external potential caused by nuclei that are fixed in some positions and that of the other electrons.

The separation of motion between electrons and nuclei is called the Born-Oppenheimer approximation. Since the positions of nuclei are fixed, wavefunction can be written as

$$\Psi^{en}(\{\mathbf{r}_i, \mathbf{R}_I\}) \approx \theta(\{\mathbf{R}_I\}) \Psi(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}). \quad (2.4)$$

Here, $\Psi(\{\mathbf{r}_i\}; \{\mathbf{R}_I\})$ is the electrons wavefunction in the Born-Oppenheimer approximation. Since the electrons are under the potential of nuclei, and thus the wavefunction of electrons is related with the nucleus positions.

Eq. 2.3 can be rewritten as

$$\begin{aligned}
H^{en} &= H + H^n \\
H &= T + U_{ext} + U_{int} \\
H^n &= - \sum_I \frac{\nabla_I^2}{2M_I} + H_{zz}
\end{aligned} \tag{2.5}$$

Furthermore, all the unknown terms in Eq. 2.5 are defined as

$$\begin{aligned}
T &= - \sum_i^{Ne} \frac{\nabla_i^2}{2} \\
U_{ext} &= \sum_i^{Ne} \sum_J^{Nn} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} \\
U_{int} &= \frac{1}{2} \sum_i^{Ne} \sum_{j \neq i}^{Ne} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \\
H_{zz} &= \frac{1}{2} \sum_I^{Nn} \sum_{J \neq I}^{Nn} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}.
\end{aligned} \tag{2.6}$$

Here, H is the Hamiltonian for the electronic system within the Born-Oppenheimer approximation. The subscript *ext* implies *external* in Eq. 2.5, and U_{ext} describes the external potentials interaction.

Thereby, the new Schrödinger equation combined with Eq. 2.4 and Eq. 2.5 is

$$\left(- \sum_I^{Nn} \frac{\nabla_I^2}{2M_I} + H_{zz} + H \right) \theta(\{\mathbf{R}_I\}) \Psi(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}) = E^{en}(\{\mathbf{R}_I\}) \theta(\{\mathbf{R}_I\}) \Psi(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}). \tag{2.7}$$

Here, $E^{en}(\{\mathbf{R}_I\})$ is the system total energy, which is \mathbf{R}_I -dependent because system wavefunction depends on nuclei positions. One ends up with the following equation using Eq. 2.6

$$\begin{aligned}
H\Psi(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}) &= E(\{\mathbf{R}_I\})\Psi(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}) \\
\left(H_{zz1} + H_{zz2} + H_{zz3} + H_{zz} + E(\{\mathbf{R}_I\}) \right) \theta(\{\mathbf{R}_I\}) &= E^{en}(\{\mathbf{R}_I\}) \theta(\{\mathbf{R}_I\}).
\end{aligned} \tag{2.8}$$

Here, $E(\{\mathbf{R}_I\})$ is the total energy of electronic system, which is also \mathbf{R}_I -dependent because electrons wavefunction indirectly depends on nuclei positions. The H_{zz1} , H_{zz2} , and H_{zz3} in Eq. 2.8 are derived as

$$\begin{aligned}
H_{zz1} &= - \sum_I^{Nn} \frac{\nabla_I^2}{2M_I} \\
H_{zz2} &= - \sum_I^{Nn} \frac{1}{M_I} \int \Psi'(\{\mathbf{r}_i\}; \{\mathbf{R}_I\})^* \nabla_I \Psi'(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}) d\mathbf{r} \nabla_I \\
H_{zz3} &= - \sum_I^{Nn} \frac{1}{M_I} \int \Psi'(\{\mathbf{r}_i\}; \{\mathbf{R}_I\})^* \nabla_I^2 \Psi'(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}) d\mathbf{r}.
\end{aligned} \tag{2.9}$$

From [Eq. 2.8](#), one observes that the lattice dynamical properties of certain system within the Born-Oppenheimer approximation can be obtained. To solve this equation, the ground state energy $E(\{\mathbf{R}_I\})$ of electronic system is needed. Here, $\{\mathbf{R}_I\}$ are the parameterized values from the atom position.

In summary, the Schrödinger equations of the electrons and nuclei are derived separately within the Born-Oppenheimer approximation. When one refers to calculations of the ground state properties, the Schrödinger equation of the electrons is applied (the first line in [Eq. 2.8](#)). The Schrödinger equation of nuclei is employed for the calculation of lattice dynamics (H_{zz2} and H_{zz3} are ignored in the second line in [Eq. 2.8](#) normally).

The [Eq. 2.8](#) (the first line) is much simpler than [Eq. 2.1](#). However, it is still not solvable. Further approximations are needed to solve this many-body problem.

2.1.3 Hartree, Hartree-Fock approximation and density functional theory

In previous section, the separation of wavefunction is given within Born-Oppenheimer approximation. The quantum many-body Schrödinger problem becomes the many-electron Schrödinger problem. There are two major problems from the Born-Oppenheimer approximation: the first problem is that the number of electron is around 10^{23} cm^{-3} , which is a huge numerical problem. However, it is still possible to solve; the second one is that the Hamiltonian includes operators which apply to the single electrons. However, how the wavefunction depends on the single-electron wavefunctions is unknown. The latter problem can be solved by one of the following three methods: the first method is to figure out a way to separate or approximate the wavefunction into single-electron function, such as the Hartree and Hartree-Fock (HF) method; the second method is to find the explicit relation between total energy and wavefunction, such as density functional theory (DFT). Within DFT, the system total energy is a functional of electron density which can be constructed by wavefunction. Either of these two methods has "pros and cons"; the third one is called Kohn-Sham equation, which is a combination of above two methods. It starts from DFT, however takes advantage of single-electron wavefunctions.

2.1.3.1 Hartree approximation

The simplest approximation of the wavefunction for the many-electron Schrödinger equation is the form of acting like independent electrons. The wavefunction with N independent electrons is given as

$$\Psi^H(\{\mathbf{r}_i\}) = \phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2) \cdots \phi_N(\mathbf{r}_N). \quad (2.10)$$

Here, i goes through all the electrons. $\phi_i(\mathbf{r}_i)$ implies state of the i :th electron, where the different states of electrons are orthonormalized. From here and following text, the \mathbf{R} is suppressed in the wavefunction since they are in fixed positions. The variables $\{\mathbf{r}_i\}$ include the coordinates of space and spin. The total energy of the electronic system can be written as

$$E^H = \langle \Psi^H(\{\mathbf{r}_i\}) | H | \Psi^H(\{\mathbf{r}_i\}) \rangle. \quad (2.11)$$

H can be rewritten as

$$\begin{aligned} H &= \sum_i^{Ne} \left(\frac{\nabla_i^2}{2} + \sum_I^{Nn} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} \right) + \frac{1}{2} \sum_i^{Ne} \sum_{j \neq i}^{Ne} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \\ &= \sum_i^{Ne} h_1(\mathbf{r}_i) + \frac{1}{2} \sum_i^{Ne} \sum_{j \neq i}^{Ne} h_2(\mathbf{r}_i, \mathbf{r}_j). \end{aligned} \quad (2.12)$$

Therefore, the total energy of system is given by making the substitution using Eq. 2.5 and Eq. 2.10 into Eq. 2.11

$$\begin{aligned} E^H &= \sum_i^{Ne} \langle \phi_i(\mathbf{r}_i) | h_1(\mathbf{r}_i) | \phi_i(\mathbf{r}_i) \rangle \\ &\quad + \frac{1}{2} \sum_i^{Ne} \sum_{j \neq i}^{Ne} \langle \phi_i(\mathbf{r}_i) \phi_j(\mathbf{r}_j) | h_2(\mathbf{r}_i, \mathbf{r}_j) | \phi_i(\mathbf{r}_i) \phi_j(\mathbf{r}_j) \rangle. \end{aligned} \quad (2.13)$$

In order to calculate the stationary state with the lowest energy of the system, the functional differentiation with respect to infinitesimal change of any wavefunction $\phi_k(\mathbf{r})$ should be zero. That is,

$$\phi_k(\mathbf{r}) \rightarrow \phi_k(\mathbf{r}) + \delta\phi_k(\mathbf{r}) \Rightarrow \delta E^H = 0. \quad (2.14)$$

Furthermore, the constraint is that the different states of electrons are orthonormalized. According to method of Lagrange multipliers, the variation with respect to any wavefunction $\phi_k^*(\mathbf{r})$ and Lagrange multiplier $E_{i,j}^H$ is satisfied

$$\frac{\delta \left(E^H - \sum_i^{Ne} \sum_j^{Ne} E_{i,j}^H (\langle \phi_i(\mathbf{r}_i) | \phi_j(\mathbf{r}_j) \rangle - \delta_{ij}) \right)}{\delta \phi_k^*(\mathbf{r})} = 0. \quad (2.15)$$

Here, $\phi_k(\mathbf{r})$ also can be utilized, however, variation with respect to $\phi_k^*(\mathbf{r})$ and $\phi_k(\mathbf{r})$ are equivalent. The result is more organized using $\phi_k^*(\mathbf{r})$.

δE^H in Eq. 2.15 can be calculated by two parts. The first part is

$$\begin{aligned} & \delta \left(\sum_i^{Ne} \langle \phi_i(\mathbf{r}_i) | h_1(\mathbf{r}_i) | \phi_i(\mathbf{r}_i) \rangle \right) \\ &= \langle \delta \phi_k(\mathbf{r}) | h_1(\mathbf{r}) | \phi_k(\mathbf{r}) \rangle + \langle \phi_k(\mathbf{r}) | h_1(\mathbf{r}) | \delta \phi_k(\mathbf{r}) \rangle \end{aligned} \quad (2.16)$$

The second part is

$$\begin{aligned} & \frac{1}{2} \delta \sum_i^{Ne} \sum_{j \neq i}^{Ne} \langle \phi_i(\mathbf{r}_i) \phi_j(\mathbf{r}_j) | h_2(\mathbf{r}_i, \mathbf{r}_j) | \phi_i(\mathbf{r}_i) \phi_j(\mathbf{r}_j) \rangle \\ &= \frac{1}{2} \sum_{i \neq k}^{Ne} \left(\langle \phi_i(\mathbf{r}_i) \delta \phi_k(\mathbf{r}) | h_2(\mathbf{r}_i, \mathbf{r}) | \phi_i(\mathbf{r}_i) \phi_k(\mathbf{r}) \rangle \right. \\ & \quad \left. + \langle \phi_i(\mathbf{r}_i) \phi_k(\mathbf{r}) | h_2(\mathbf{r}_i, \mathbf{r}) | \phi_i(\mathbf{r}_i) \delta \phi_k(\mathbf{r}) \rangle \right) \\ &+ \frac{1}{2} \sum_{j \neq k}^{Ne} \left(\langle \delta \phi_k(\mathbf{r}) \delta \phi_j(\mathbf{r}_j) | h_2(\mathbf{r}, \mathbf{r}_j) | \phi_k(\mathbf{r}) \phi_j(\mathbf{r}_j) \rangle \right. \\ & \quad \left. + \langle \phi_k(\mathbf{r}) \phi_j(\mathbf{r}_j) | h_2(\mathbf{r}, \mathbf{r}_j) | \delta \phi_k(\mathbf{r}) \phi_j(\mathbf{r}_j) \rangle \right) \\ &= \sum_{i \neq k}^{Ne} \left(\langle \phi_i(\mathbf{r}_i) \delta \phi_k(\mathbf{r}) | h_2(\mathbf{r}_i, \mathbf{r}) | \phi_i(\mathbf{r}_i) \phi_k(\mathbf{r}) \rangle \right. \\ & \quad \left. + \langle \phi_i(\mathbf{r}_i) \phi_k(\mathbf{r}) | h_2(\mathbf{r}_i, \mathbf{r}) | \phi_i(\mathbf{r}_i) \delta \phi_k(\mathbf{r}) \rangle \right) \end{aligned} \quad (2.17)$$

Here, the factor of $\frac{1}{2}$ cancels because the 2nd (3rd) line is the same with 4th (5th) line in Eq. 2.17 due to the exchangeable of the index of i and j .

To get the final solution, one more calculation is given as

$$\begin{aligned}
& \delta \sum_i^{Ne} \sum_j^{Ne} E_{i,j}^H \left(<\phi_i(\mathbf{r}_i)|\phi_j(\mathbf{r}_j)> - \delta_{i,j} \right) \\
&= \sum_j^{Ne} E_{k,j}^H \left(<\delta\phi_k(\mathbf{r})|\phi_j(\mathbf{r}_j)> \right) + \sum_i^{Ne} E_{i,k}^H \left(<\phi_i(\mathbf{r}_i)|\delta\phi_k(\mathbf{r})> \right) \\
&= \sum_i^{Ne} E_{k,i}^H \left(<\delta\phi_k(\mathbf{r})|\phi_i(\mathbf{r}_i)> + <\phi_i(\mathbf{r}_i)|\delta\phi_k(\mathbf{r})> \right)
\end{aligned} \tag{2.18}$$

Therefore, **Eq. 2.15** can be derived as

$$\left(-\frac{\nabla_k^2}{2} + \sum_I^{Nn} \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} + \sum_{j \neq k}^{Ne} <\phi_j(\mathbf{r}')|\frac{1}{|\mathbf{r} - \mathbf{r}'|}|\phi_j(\mathbf{r}')> \right) \phi_k(\mathbf{r}) = \sum_i^{Ne} E_{k,i}^H \phi_i(\mathbf{r}) \tag{2.19}$$

There are many solutions to **Eq. 2.19**, each corresponding to a different set of $E_{k,i}^H$. One can choose to $E_{k,i}^H$ which satisfies $E_{k,i}^H = \delta_{k,i} \epsilon_k^H$. The **Eq. 2.19** can be rewriteen as

$$\begin{aligned}
& \left(-\frac{\nabla_k^2}{2} + \sum_I^{Nn} \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} + \sum_{j \neq k}^{Ne} <\phi_j(\mathbf{r}')|\frac{1}{|\mathbf{r} - \mathbf{r}'|}|\phi_j(\mathbf{r}')> \right) \phi_k(\mathbf{r}) = \epsilon_k^H \phi_k(\mathbf{r}) \\
& \quad \Downarrow \\
& \left(-\frac{\nabla_i^2}{2} + \sum_I^{Nn} \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} + \sum_{j \neq i}^{Ne} <\phi_j(\mathbf{r}')|\frac{1}{|\mathbf{r} - \mathbf{r}'|}|\phi_j(\mathbf{r}')> \right) \phi_i(\mathbf{r}) = \epsilon_i^H \phi_i(\mathbf{r})
\end{aligned} \tag{2.20}$$

Here, **Eq. 2.20** is a group of dependent single particle equations. ϵ_k^H is identified as the eigenvalue for this single-electron Hartree equation.

2.1.3.2 Hartree-Fock approximation

Hartree approximation is the simplest approximation, while Hartree-Fock approximation is the method which considers the antisymmetry of the wavefunction. It is shown as

$$\Psi^{HF}(\dots \mathbf{r}_i \dots \mathbf{r}_j \dots) = -\Psi^{HF}(\dots \mathbf{r}_j \dots \mathbf{r}_i \dots). \tag{2.21}$$

Here, each variable \mathbf{r}_i includes the coordinates of space and spin. Slater introduced an way to construct the wavefunction subject to **Eq. 2.21**. The wavefunction of the many-electron Schrödinger equation is described in a matrix determinant for the N number

of electrons (for aesthetical reason, N implies the number of electrons and not Ne in the section)

$$\Psi^{HF}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{r}_1) & \phi_2(\mathbf{r}_1) & \cdots & \phi_N(\mathbf{r}_1) \\ \phi_1(\mathbf{r}_2) & \phi_2(\mathbf{r}_2) & \cdots & \phi_N(\mathbf{r}_2) \\ \vdots & \vdots & & \vdots \\ \phi_1(\mathbf{r}_N) & \phi_2(\mathbf{r}_N) & \cdots & \phi_N(\mathbf{r}_N) \end{vmatrix}. \quad (2.22)$$

Here, the factor in front ensures normalization, i goes through all the electrons, and $\phi_i(\mathbf{r}_i)$ implies state of the (i):th electron. If two rows in Eq. 2.22 are exchanged, the result is agreed with Eq. 2.21.

The total energy of Hartree-Fock approximation, which can be determined similarly as in the previous section of Hartree approximation, is given as

$$\begin{aligned} E^{HF} = & \sum_i^N \left\langle \phi_i(\mathbf{r}) \right| - \frac{\nabla_i^2}{2} + \sum_I^{Nn} \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} \left| \phi_i(\mathbf{r}) \right\rangle \\ & + \frac{1}{2} \sum_i^N \sum_j^N \left\langle \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \right| \frac{1}{|\mathbf{r} - \mathbf{r}'|} \left| \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \right\rangle \\ & - \frac{1}{2} \sum_i^N \sum_j^N \left\langle \phi_i(\mathbf{r}') \phi_j(\mathbf{r}) \right| \frac{1}{|\mathbf{r} - \mathbf{r}'|} \left| \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \right\rangle. \end{aligned} \quad (2.23)$$

Here, there is no contribution to the sum when $i = j$. In the same mathematical way as in previous section but somewhat more complicated, the single particle Hartree-Fock equation can be obtained

$$\begin{aligned} & \left\{ -\frac{\nabla_i^2}{2} + \sum_I^{Nn} \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} + \sum_j^N \left\langle \phi_j(\mathbf{r}') \right| \frac{1}{|\mathbf{r} - \mathbf{r}'|} \left| \phi_j(\mathbf{r}') \right\rangle \right\} \phi_i(\mathbf{r}) \\ & - \sum_j^N \left\langle \phi_j(\mathbf{r}') \right| \frac{1}{|\mathbf{r} - \mathbf{r}'|} \left| \phi_i(\mathbf{r}') \right\rangle \phi_j(\mathbf{r}) = \epsilon_i^{HF} \phi_i(\mathbf{r}). \end{aligned} \quad (2.24)$$

In comparison with Hartree equation in Eq. 2.20, there is an extra term in the Eq. 2.24, which is so called exchange term. In order to show the equation in a more organized way, Eq. 2.24 is equivalent to

$$\left\{ -\frac{\nabla_i^2}{2} + \sum_I^{Nn} \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} + V_{HF}(\mathbf{r}) \right\} \phi_i(\mathbf{r}) = \epsilon_i^{HF} \phi_i(\mathbf{r}). \quad (2.25)$$

$$\begin{aligned}
V_{HF}(\mathbf{r}) &= \int \frac{\rho(\mathbf{r}') - \rho_i^{HF}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \\
\rho_i^{HF}(\mathbf{r}, \mathbf{r}') &= \sum_j^N \frac{\phi_i(\mathbf{r}')\phi_i^*(\mathbf{r})\phi_j(\mathbf{r})\phi_j^*(\mathbf{r}')}{\phi_i(\mathbf{r})\phi_i^*(\mathbf{r})} \\
\rho(\mathbf{r}) &= \sum_i^N |\phi_i(\mathbf{r})|^2.
\end{aligned} \tag{2.26}$$

2.1.3.3 Density functional theory

The Hartree and Hartree-Fock (HF) methods are very classic methods to solve the many-electron Schrödinger equation. However, the HF method only includes the exchange term and not the electron correlation term. It is not suitable for the solid materials. Apart from the Hartree and HF method, there is a modern method to solve the more complicated calculations of electron system, namely density functional theory (DFT). It is introduced by Hohenberg and Kohn in 1964, moreover, Kohn and Pople was awarded Chemistry Nobel Prize in 1998.

The idea of the DFT is to treat the electron density in solid materials instead of using the many-particle wavefunction. One can benefit that the degree of freedom reduces from $3N$ (N is the number of electrons) to 3. It is apparently less complicate than those of Hartree and HF methods.

The density as basic variable

There are two questions coming out if considering the electron density as the role of wavefunction. The first one is whether it is the equivalence relation between the electron density and wavefunction in the electron system. The second one is how to solve the problem if considering the electron density instead of the wavefunction. In order to explain above two questions, there are two very basic theorems introduced by Hohenberg and Kohn:

Theorem 1 *The first theorem states that the external potential $V_{ext}(\mathbf{r})$ is determined uniquely for any electron system by the ground-state electron density ρ .*

The above theorem indicates that all the ground state properties are determined by the true ground-state density ρ as well, such as, the total energy $E = E[\rho]$.

The above theorem also explains the equivalence relation between the electron density and wavefunction. Because Hamiltonian is obtained from external potential, then the wavefunction is obtained. Therefore the corresponding electron density is determined. Moreover, from the theorem, the external potential is unique decided by electron density. Therefore the electron density contains the same information as the wavefunction.

The proof of the theorem is given. Let us assume that there exists two external potentials named $V_{ext}(\mathbf{r})$ and $V'_{ext}(\mathbf{r})$ leading to the same ground state electron density ρ . Obviously, this will lead to two different Hamiltonians, that is, H and H' , as well as two different corresponding wavefunctions named Ψ and Ψ' . Since Ψ are not the ground state wavefunction of H' , the same rules to Ψ' and H , two following inequality equations are satisfied

$$\begin{aligned} E &= \langle \Psi | H | \Psi \rangle < \langle \Psi' | H | \Psi' \rangle, \\ E' &= \langle \Psi' | H' | \Psi' \rangle < \langle \Psi | H' | \Psi \rangle. \end{aligned} \quad (2.27)$$

Taking advantage of the Hamiltonian from **Eq. 2.5** ($V_{ext}(\mathbf{r})$ instead of U_{ext}), the following equation is derived

$$\begin{aligned} &\langle \Psi' | H | \Psi' \rangle \\ &= \langle \Psi' | H' + H - H' | \Psi' \rangle \\ &= \langle \Psi' | H' | \Psi' \rangle + \langle \Psi' | H - H' | \Psi' \rangle \\ &= E' + \int d\mathbf{r} \left(V_{ext}(\mathbf{r}) - V'_{ext}(\mathbf{r}) \right) \rho(\mathbf{r}). \end{aligned} \quad (2.28)$$

Using **Eq. 2.27** and **Eq. 2.28**, the following relation is given by

$$E < E' + \int d\mathbf{r} \left(V_{ext}(\mathbf{r}) - V'_{ext}(\mathbf{r}) \right) \rho(\mathbf{r}). \quad (2.29)$$

Another similar inequality equation can be gained if one changes the equation $\langle \Psi | H' | \Psi \rangle$ like **Eq. 2.28**

$$E' < E + \int d\mathbf{r} \left(V'_{ext}(\mathbf{r}) - V_{ext}(\mathbf{r}) \right) \rho(\mathbf{r}). \quad (2.30)$$

Plus the left side and right side from both **Eq. 2.29** and **Eq. 2.30**, a contradictory result is given as

$$E + E' < E' + E. \quad (2.31)$$

Since this is an incorrect relation, the external potential $V_{ext}(\mathbf{r})$ has to be unique.

Theorem 2 *The second theorem states that there is a universal functional $F[\rho]$ for the total energy in the terms of the electron density ρ with any external potential $V_{ext}(\mathbf{r})$, and the exact ground-state density is gained when the ground state total energy functional reaches its minimal value, that is, $E[\rho'] > E[\rho]$. Here, ρ is the exact ground-state density.*

The proof of theorem is given. Because of the first theorem, the total energy can be expressed in the following way (ignoring the interaction between nuclei)

$$\begin{aligned} E[\rho] &= \langle \Psi | T + V_{int} + V_{ext} | \Psi \rangle \\ &= \langle \Psi | V_{ext} | \Psi \rangle + \underbrace{\langle \Psi | T + V_{int} | \Psi \rangle}_{\text{}} \\ &= \int \rho(\mathbf{r}) V_{ext} + F[\rho]. \end{aligned} \quad (2.32)$$

In Eq. 2.32, the term of $F[\rho]$ is the universal functional for all the many-electron system. The functional of total energy $E[\rho']$ reach the minimum at the exact ground-state electron density ρ

$$E[\rho'] = \int \rho'(\mathbf{r}) V_{ext} + F[\rho'] > E[\rho]. \quad (2.33)$$

From above equation, one knows the total energy for the case of exact ground-state electron density is lower than any other cases. It also generates that one can achieve the exact ground-state electron density by minimizing the total energy.

From those two theorems, one knows how to solve this problem theoretically. However in practice, $E[\rho]$ is unknown. Therefore, one more method is needed to solve it, which is so called Kohn-Sham (KS) equation.

2.1.3.4 Kohn-Sham equation

The Hartree and Hartree-Fock methods are introduced to solve the many-body problem, both of which are based on the idea of transforming complex many-electron problem to single-electron problem by using different wavefunctions. The DFT only consider to take use of information from Hamiltonian, however not solvable. Therefore, it is best to combine these two ideas together. The DFT is solved by Kohn-Sham equation introduced by Kohn and Sham in 1965, which is constructed in the following text (ignoring the interaction between nuclei).

Assume that the exact ground-state density is obtained by the wavefunction $\Psi^{KS}(\{\mathbf{r}_i\}) = \Psi_1^{KS}(\mathbf{r}_1)\Psi_2^{KS}(\mathbf{r}_2)\dots\Psi_M^{KS}(\mathbf{r}_M)$. Here, M is a number which is less than 10^{23} a lot and $\Psi_i^{KS}(\mathbf{r}_i)$ is auxiliary independent single-electron wavefunction, and electron density is defined as

$$\rho(\mathbf{r}) = \sum_i \Psi_i^{KS*}(\mathbf{r}) \Psi_i^{KS}(\mathbf{r}). \quad (2.34)$$

If the density is exact, thereby the total energy is defined exactly. It is given as

$$\begin{aligned}
E[\rho] &= T[\rho] + V_{int}[\rho] + V_{ext}[\rho] \\
&= T_0[\rho] + V_{HF}[\rho] + V_{ext}[\rho] + \underbrace{(T[\rho] - T_0[\rho])}_{(V_{int}[\rho] - V_{HF}[\rho])} \\
&= T_0[\rho] + V_{HF}[\rho] + V_{ext}[\rho] + V_c[\rho].
\end{aligned} \tag{2.35}$$

Here, $E[\rho]$ is the total energy, and ρ is the ground-state density. $T[\rho]$, $V_{int}[\rho]$ and $V_{ext}[\rho]$ in order are the energy from external potential, the exact kinetic, and the exact electron-electron potential energy. $T_0[\rho]$ is the kinetic energy of independent electrons, $V_{HF}[\rho]$ is the potential from the Hartree-Fock approximation, and $V_c[\rho]$ denotes correlation energy.

From the Hartree-Fock approximation, one can further know that

$$V_{HF}[\rho] = V_H[\rho] + V_x[\rho]. \tag{2.36}$$

Here, $V_H[\rho]$ and $V_x[\rho]$ are the Hartree contribution and exchange contribution, respectively. Thereby, the Eq. 2.35 can be further defined as

$$\begin{aligned}
E[\rho] &= T_0[\rho] + V_H[\rho] + V_{ext}[\rho] + \underbrace{V_x[\rho] + V_c[\rho]}_{V_{xc}[\rho]} \\
&= T_0[\rho] + V_H[\rho] + V_{ext}[\rho] + V_{xc}[\rho].
\end{aligned} \tag{2.37}$$

Here, $V_{xc}[\rho]$ is the exchange-correlation term. From Eq. 2.35, one does not know the explicit expression of $V_{xc}[\rho]$. However, the first three terms are give as

$$T_0[\rho] = \sum_i \langle \Psi_i^{KS}(\mathbf{r}) | -\frac{\nabla_i^2}{2} | \Psi_i^{KS}(\mathbf{r}) \rangle \tag{2.38}$$

$$V_H[\rho] = \frac{1}{2} \int \int d\mathbf{r} d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \tag{2.39}$$

$$V_{ext}[\rho] = \int d\mathbf{r} \rho(\mathbf{r}) V_{ext}(\mathbf{r}). \tag{2.40}$$

In order to derive the ground state properties in the many-electron system, one can view this problem as the process of minimizing the total energy by varying the wavefunction $\Psi_i^{KS*}(\mathbf{r})$. Since the wavefunction can be constructed to electron density (Eq. 2.34), just like the derivation in the section of Hartree approximation, finally one can get the Kohn-Sham equation

$$\left(-\frac{\nabla_i^2}{2} + V^{KS}(\mathbf{r}) \right) \Psi_i^{KS}(\mathbf{r}) = E_i^{KS} \Psi_i^{KS}(\mathbf{r}). \tag{2.41}$$

Here,

$$\begin{aligned} V^{KS}(\mathbf{r}) &= V_{ext}(\mathbf{r}) + \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta V_{xc}}{\delta \rho(\mathbf{r})} \\ &= V_{ext}(\mathbf{r}) + \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + v_{xc}(\mathbf{r}). \end{aligned} \quad (2.42)$$

One can think of the Hamiltonian in another point of view, that is, single particle system with three different potentials.

There is no total energy equation expression yet. However, if one changes the Eq. 2.41 as

$$\sum_i \Psi_i^{KS*}(\mathbf{r}) \left(-\frac{\nabla_i^2}{2} + V^{KS}(\mathbf{r}) \right) \Psi_i^{KS}(\mathbf{r}) = \sum_i \Psi_i^{KS*}(\mathbf{r}) E_i^{KS} \Psi_i^{KS}(\mathbf{r}). \quad (2.43)$$

Based on Eq. 2.37 and Eq. 2.43, the total energy expression is

$$\begin{aligned} E &= \sum_i E_i^{KS} - \frac{1}{2} \int \int d\mathbf{r} d\mathbf{r}' \frac{\rho(\mathbf{r}) \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \\ &\quad + V_{xc}[\rho] - \int v_{xc}(\mathbf{r}) \rho(\mathbf{r}). \end{aligned} \quad (2.44)$$

So far there are two problems still in the air: one is the exact format of $v_{xc}(\mathbf{r})$; the other one is how to solve the Kohn-Sham equation.

2.1.3.5 The exchange-correlation potential

The exchange-correlation potential is the most difficult part during the process of solving the Kohn-Sham equation, because it is still unknown yet. Therefore, there are varies of approximations about it, such as, the local density approximation (LDA), generalized-gradient approximation (GGA) and many others.

The local density approximation

The local density approximation is the simplest way to approximate the exchange-correlation part. The idea is that the value of the exchange-correlation potential in the very tiny small volume is equal to the homogeneous electrons with the same density in the volume, the explicit equation is

$$E_{xc}^{LDA} = \int \rho(\mathbf{r}) \varepsilon_{xc}(\rho(\mathbf{r})) d\mathbf{r}. \quad (2.45)$$

Here, E_{xc}^{LDA} is the exchange-correlation energy functional for the LDA, $\rho(\mathbf{r})$ is the charge density, ε_{xc} is homogeneous electrons gas with variable of $\rho(\mathbf{r})$.

The ε_{xc} is an ideal state within a solid, which assumes that the charge is homogeneously all over the space

$$\rho(\mathbf{r}) = \rho = \frac{N}{V}. \quad (2.46)$$

Here, N is number of electrons within the solid, and V is the volume of solid. not functional.

The generalized gradient approximation

The GGA is the approximation beyond the LDA, which incorporates not only the density within the tiny volume, but also the gradient of the density. The explicit equation is

$$E_{xc}^{GGA} = \int \rho(\mathbf{r}) \varepsilon_{xc}(\rho(\mathbf{r}), \nabla \rho(\mathbf{r})) d\mathbf{r}. \quad (2.47)$$

Here, E_{xc}^{GGA} is the exchange-correlation energy functional for the GGA, $\nabla \rho(\mathbf{r})$ is the gradient of charge density in the position of \mathbf{r} .

Today, there exists hundreds of different potentials. It is still ongoing development. The advantage of Kohn-Sham is that one new potential is implemented easily. In the same time, the number of potential is huge due to the simplicity of implementation.

2.1.4 Solving the secular equation

The process for solving Kohn-Sham equation can be achieved by iteration as well. However, the difference between solving Kohn-Sham equation and Hartree or Hartree-Fock equation is that the wavefunction is replaced by the electron density in Kohn-Sham equation. Thereby, an initial electron density is defined at the beginning of the calculation, then the equation is solved iteratively until the reasonable solution is obtained.

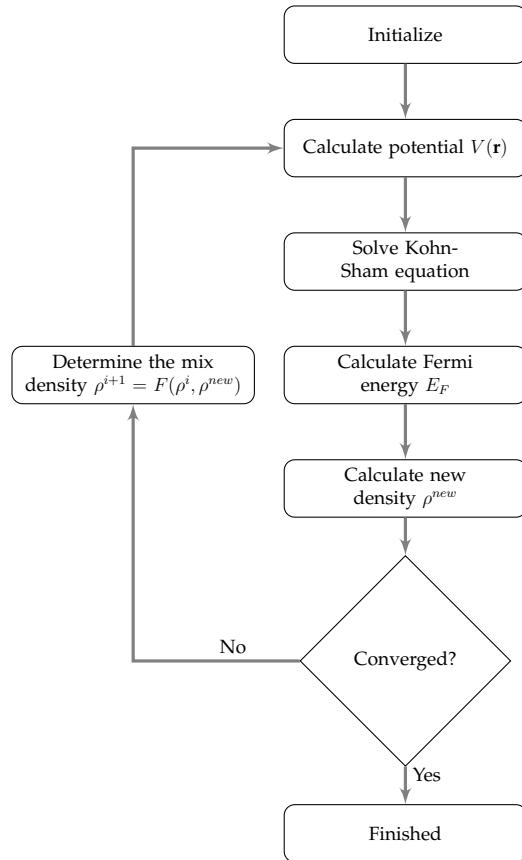


Figure 2.1. Flow chart of the $(i+1)$:th iterations for solving Kohn-Sham equation.

Here, ρ^i and ρ^{i+1} are the charge density of the (i) :th and $(i+1)$:th iteration solving Kohn-Sham equation, respectively.

2.1.5 Eigenvalue problem

In order to solve [Eq. 2.41](#), the Kohn-Sham equation can be transformed into the general eigenvalue problem. If the Kohn-Sham equation is given as (ignore the index in wavefunction compared with [Eq. 2.41](#))

$$H\Psi^{KS}(r) = E\Psi^{KS}(r). \quad (2.48)$$

The wavefunction is defined as

$$\Psi^{KS}(r) = \sum_j^N C_j \Phi_j(r). \quad (2.49)$$

Here, C_j is a complex number, and $\Phi_j(r)$ is the basis function of wavefunction.

If [Eq. 2.49](#) is plugged into [Eq. 2.48](#), and left multiply $\Phi_1^*, \Phi_2^*, \dots, \Phi_N^*$ in order, finally it will end up with a set of equations

$$\begin{aligned} & \begin{bmatrix} \Phi_1^* H \Phi_1 & \Phi_1^* H \Phi_2 & \cdots & \Phi_1^* H \Phi_N \\ \Phi_2^* H \Phi_1 & \Phi_2^* H \Phi_2 & \cdots & \Phi_2^* H \Phi_N \\ \vdots & \vdots & & \vdots \\ \Phi_N^* H \Phi_1 & \Phi_N^* H \Phi_2 & \cdots & \Phi_N^* H \Phi_N \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_N \end{bmatrix} \\ &= E \begin{bmatrix} \Phi_1^* \Phi_1 & \Phi_1^* \Phi_2 & \cdots & \Phi_1^* \Phi_N \\ \Phi_2^* \Phi_1 & \Phi_2^* \Phi_2 & \cdots & \Phi_2^* \Phi_N \\ \vdots & \vdots & & \vdots \\ \Phi_N^* \Phi_1 & \Phi_N^* \Phi_2 & \cdots & \Phi_N^* \Phi_N \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_N \end{bmatrix}. \end{aligned} \quad (2.50)$$

If $H_{ij} = \Phi_i^* H \Phi_j$ and $S_{ij} = \Phi_i^* \Phi_j$ are defined, then the [Eq. 2.50](#) becomes

$$\begin{aligned} & \begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1N} \\ H_{21} & H_{22} & \cdots & H_{2N} \\ \vdots & \vdots & & \vdots \\ H_{N1} & H_{N2} & \cdots & H_{NN} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_N \end{bmatrix} \\ &= E \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1N} \\ S_{21} & S_{22} & \cdots & S_{2N} \\ \vdots & \vdots & & \vdots \\ S_{N1} & S_{N2} & \cdots & S_{NN} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_N \end{bmatrix}. \end{aligned} \quad (2.51)$$

There are left part and right part in [Eq. 2.51](#). If the right part moves to the left part, and do the subtraction. The new equation is given as

$$\begin{bmatrix} H_{11} - ES_{11} & H_{12} - ES_{12} & \cdots & H_{1N} - ES_{1N} \\ H_{21} - ES_{21} & H_{22} - ES_{22} & \cdots & H_{2N} - ES_{2N} \\ \vdots & \vdots & & \vdots \\ H_{N1} - ES_{N1} & H_{N2} - ES_{N2} & \cdots & H_{NN} - ES_{NN} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_N \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (2.52)$$

Apparently, Eq. 2.52 is an eigenvalue problem. In order to get the $C_i (i = 1 \cdots N)$, one needs to solve the eigenvalue problem

$$\begin{vmatrix} H_{11} - ES_{11} & H_{12} - ES_{12} & \cdots & H_{1N} - ES_{1N} \\ H_{21} - ES_{21} & H_{22} - ES_{22} & \cdots & H_{2N} - ES_{2N} \\ \vdots & \vdots & & \vdots \\ H_{N1} - ES_{N1} & H_{N2} - ES_{N2} & \cdots & H_{NN} - ES_{NN} \end{vmatrix} = 0. \quad (2.53)$$

2.2 Full-potential linearized augmented plane wave method

2.2.1 Introduction

So far, one already know how to solve the Kohn-Sham equation. However, there are still two more questions, what is the exact form of wavefunction and potential in the realistic calculations?

One maybe naturally choose a set of plane waves as the wavefunction because of Bloch theory. There is a drawback about the plane waves as wavefunction when describing the atomic core region. Because the wavefunction change dramatically in that region, therefore one needs to choose relatively more plane waves to approximate it. It implies it will take more time to calculate.

Slater re-considered the way to describe the wavefunction (Eq. 2.55). The unit cell is splitted into two regions: one is the sphere region which is defined by the center of atom, but non-overlap each sphere, is called muffin tin (MT) region; the remaining region is called interstitial (I) region (Fig. 2.2). An atomic-like function is defined as the wavefunction in the MT region, this is reason why this method is called augmented plane wave (APW). The dual representation of the wavefunction is reasonable, because the wavefunction approaching atomic core is somehow like inside atom; however far away the atomic core, the electron behaves like free electrons, therefore plane waves are suitable (Eq. 2.55). However, the drawback of APW method is the wavefunction is dependent with the energy, which leads to the nonlinear eigenvalue problem (Eq. 2.56). This method has to decide repeatedly until certain condition is satisfied to achieve the exact energy, which is really time-consuming.

In order to find a way out, is it possible to let the wavefunction energy-independent? Andesen, Koelling and Arbman proposed a way to describe it, they noticed that the Taylor expansion of radial function (Eq. 2.59). This method is called linearized augmented plane wave (LAPW) method due to make use of the linearized energies in the radial functions. However, the drawback is that it does not describe the semicore state well, it is corrected by a method named linearized augmented plane wave plus local orbitals (LAPW+LO) which is proposed by Singh (Eq. 2.61). Sjöstedt, Nordström and Singh also give an efficient way to linearize Slater's APW method, named augmented plane wave plus local orbitals (APW+lo) (Eq. 2.62). Both of above methods consider local orbitals, it is customary to write in different ways as LO and lo, respectively. One can read more information about FPLAPW method from [?, ?, ?].

2.2.2 Wavefunction

2.2.2.1 Augmented plane wave method

The Kohn-Sham wavefunction is defined as

$$\Psi_{i\mathbf{k}}^{KS}(\mathbf{r}) = \sum_{\mathbf{G}} C_{i,\mathbf{k}+\mathbf{G}} \phi_{\mathbf{k}+\mathbf{G}}(\mathbf{r}). \quad (2.54)$$

Here, $\phi_{\mathbf{k}+\mathbf{G}}(\mathbf{r})$ is the basis function of wavefunction. It is written in slightly different format in order to distinct the different methods in the following text, for example, $\phi_{\mathbf{k}+\mathbf{G}}^{APW}(\mathbf{r})$ represents the basis function for the augmented plane wave (APW) method.

The basis set for APW is defined by Slater

$$\phi_{\mathbf{k}+\mathbf{G}}^{APW}(\mathbf{r}) = \begin{cases} \frac{1}{\sqrt{\Omega}} e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}} & \text{if } \mathbf{r} \in I \\ \sum_{\alpha} \sum_{\ell m} f_{\ell m}(r_{\alpha}, \mathbf{k} + \mathbf{G}, E) Y_{\ell m}(\hat{\mathbf{r}}_{\alpha}) & \text{if } r_{\alpha} \in MT. \end{cases} \quad (2.55)$$

Here, $f_{\ell m}(r_{\alpha}, \mathbf{k} + \mathbf{G}, E) = A_{\ell m}^{\alpha}(\mathbf{k} + \mathbf{G}) u_{\ell}(r_{\alpha}, E)$, and $A_{\ell m}^{\alpha}(\mathbf{k} + \mathbf{G})$ is the expansion coefficients. Radial function $u_{\ell}(r_{\alpha}, E)$ is dependent with energy E , and it can be decided by the following equation

$$-\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{du_{\ell}}{dr} \right) + \left(\frac{\ell(\ell+1)}{r^2} + V(\mathbf{r}) \right) u_{\ell}(r_{\alpha}) = Eu_{\ell}(r_{\alpha}). \quad (2.56)$$

Here, $V(\mathbf{r})$ is the spherical potential.

Because the wavefunction has dual representation, one has to make sure the continuity on the sphere boundary, which is solved by matching each ℓm of the dual representation.

From Fig. 2.2, one notices that unit cell is divided into MT spheres (α, β) and an (I) region, where $\mathbf{r} = \mathbf{R}_{\alpha} + \mathbf{r}_{\alpha}$ is guaranteed. The Rayleigh expansion formula yields

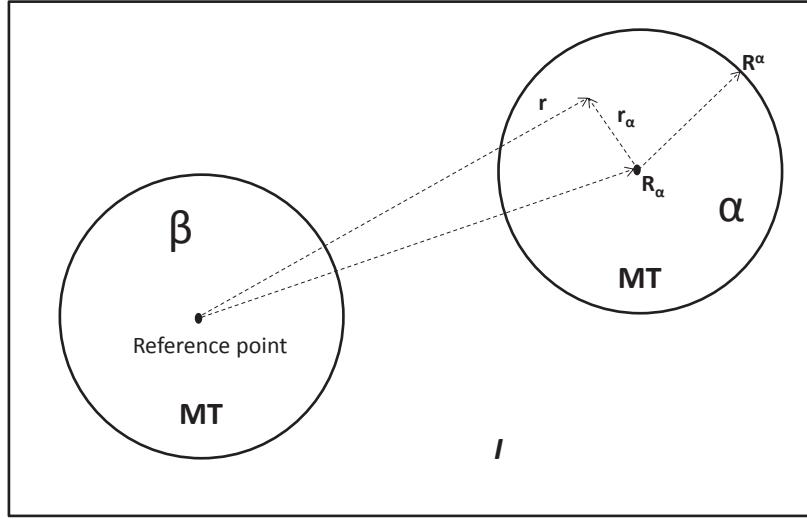


Figure 2.2. Partition of the unit cell.

$$e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}} = e^{i(\mathbf{k}+\mathbf{G})\mathbf{R}_\alpha} 4\pi \sum_{\ell m} i^\ell j_\ell(|\mathbf{k} + \mathbf{G}|r_\alpha) Y_{\ell m}(\hat{\mathbf{r}}_\alpha) Y_{\ell m}^*(\widehat{\mathbf{k} + \mathbf{G}}). \quad (2.57)$$

After matching those two representations, the following equation is satisfied for each ℓm

$$A_{\ell m}^\alpha(\mathbf{k} + \mathbf{G}) = \frac{e^{i(\mathbf{k}+\mathbf{G})\mathbf{R}_\alpha} 4\pi \sum_{\ell m} i^\ell j_\ell(|\mathbf{k} + \mathbf{G}|r_\alpha) Y_{\ell m}^*(\widehat{\mathbf{k} + \mathbf{G}})}{\sqrt{\Omega u_\ell^\alpha(r_\alpha, E)}}. \quad (2.58)$$

The main drawbacks about the APW method is that the wavefunction is energy dependent. It generates that the code needs to search for the energy in order to calculate the exact energy. Therefore it is really time-consuming.

2.2.2.2 Linearized augmented plane wave method

In order to decouple the energy from the wavefunction, Andesen, Koelling and Arbman found out a way to separate them, they noticed that the Taylor expansion of the radial function on certain energy, which can be given as

$$u_\ell(r_\alpha, E) = u_\ell(r_\alpha, E_\ell) + (E - E_\ell)\dot{u}_\ell(r_\alpha, E_\ell) + O((E - E_\ell)^2). \quad (2.59)$$

Therefore the basis function is re-defined as

$$\phi_{\mathbf{k}+\mathbf{G}}^{LAPW}(\mathbf{r}) = \begin{cases} \frac{1}{\sqrt{\Omega}} e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}} & \text{if } \mathbf{r} \in I \\ \sum_{\alpha} \sum_{\ell m} f_{\ell m}(r_{\alpha}, \mathbf{k} + \mathbf{G}, E_{\ell}) Y_{\ell m}(\hat{\mathbf{r}}_{\alpha}) & \text{if } r_{\alpha} \in MT. \end{cases} \quad (2.60)$$

Here, $f_{\ell m}(r_{\alpha}, \mathbf{k} + \mathbf{G}, E_{\ell}) = A_{\ell m}^{\alpha}(\mathbf{k} + \mathbf{G}) u_{\ell}(r_{\alpha}, E_{\ell}) + B_{\ell m}^{\alpha}(\mathbf{k} + \mathbf{G}) \dot{u}_{\ell}(r_{\alpha}, E_{\ell})$. $A_{\ell m}^{\alpha}(\mathbf{k} + \mathbf{G})$ and $B_{\ell m}^{\alpha}(\mathbf{k} + \mathbf{G})$ are the expansion coefficients, and $\dot{u}_{\ell}(r_{\alpha}, E_{\ell})$ is the derivative of the radial function. Energy E_{ℓ} is considered as pre-calculated parameter in Eq. 2.60. Actually, it is chosen by the middle of each ℓ -character band, therefore this method is called linearized augmented plane wave (LAPW) method.

Apparently, LAPW method is more suitable in reality, because the wavefunction is decoupled with energy. However it has to match for two parameters. Fortunately, it still take less time comparing with APW method. However, there is one drawback about this method, what if energy in the same ℓ character is different enough, which E_{ℓ} is correct? These states are called as semi-core states which exist in the actinides and the rare earths.

2.2.2.3 Linearized augmented plane wave method plus local orbitals

Comparing with LAPW method, linearized augmented plane wave method plus local orbitals (LAPW+LO) method extend and add smaller number of basis set, which is defined as

$$\phi_{\mathbf{k}+\mathbf{G}}^{LO}(\mathbf{r}) = \begin{cases} 0 & \text{if } \mathbf{r} \in I \\ (A_{\ell m}^{\alpha, LO} u_{\ell}(r_{\alpha}, E_{\ell}) + B_{\ell m}^{\alpha, LO} \dot{u}_{\ell}(r_{\alpha}, E_{\ell}) + C_{\ell m}^{\alpha, LO} u_{\ell}(r_{\alpha}, E'_{\ell})) Y_{\ell m}(\hat{\mathbf{r}}_{\alpha}) & \text{if } r_{\alpha} \in MT. \end{cases} \quad (2.61)$$

Here, $A_{\ell m}^{\alpha, LO}$, $B_{\ell m}^{\alpha, LO}$, and $C_{\ell m}^{\alpha, LO}$ can be obtained by normalization, as well as value and derivation on the sphere boundary to zero. The E'_{ℓ} is the chosen energy from semi-core state.

2.2.2.4 Augmented plane wave method plus local orbitals

There is one method which can solve the APW method efficiently, which is called as augmented plane wave method plus local orbitals (APW+lo). The basis function has two types: one is similar with APW method, but only without the derivative terms, that is, $f_{\ell m}(r_{\alpha}, \mathbf{k} + \mathbf{G}, E_{\ell}) = A_{\ell m}^{\alpha}(\mathbf{k} + \mathbf{G}) u_{\ell}(r_{\alpha}, E_{\ell})$; the other basis function is

$$\phi_{\mathbf{k}+\mathbf{G}}^{lo}(\mathbf{r}) = \begin{cases} 0 & \text{if } \mathbf{r} \in I \\ (A_{\ell m}^{\alpha, lo} u_{\ell}(r_{\alpha}, E_{\ell}) + B_{\ell m}^{\alpha, lo} \dot{u}_{\ell}(r_{\alpha}, E_{\ell})) Y_{\ell m}(\hat{\mathbf{r}}_{\alpha}) & \text{if } r_{\alpha} \in MT. \end{cases} \quad (2.62)$$

The value of $A_{\ell m}^{\alpha, lo}$ and $B_{\ell m}^{\alpha, lo}$ are obtained by normalization and local orbital has zero value at the muffin tin boundary. This method can not solve semicore states like LAPW+LO.

However, it does increase the efficiency to linearize Slater's APW method. Certainly, it exists some types of basis function which can mix the advantages from mentioned methods.

2.2.3 Effective potential

The potential in the FPLAPW method is also divided into two regions, the MT region and the I region.

$$V(\mathbf{r}) = \begin{cases} \sum_{\mathbf{G}} V_{\mathbf{G}} e^{i\mathbf{Gr}} & \text{if } \mathbf{r} \in I \\ \sum_{\ell m} V_{\ell m}^{\alpha}(r_{\alpha}) Y_{\ell m}(\hat{\mathbf{r}}_{\alpha}) & \text{if } r_{\alpha} \in MT. \end{cases}$$

2.3 Dielectric function

The dielectric function describes the optical property of materials. Normally, it is written as ε , which has two parts

$$\varepsilon = \varepsilon_1 + i\varepsilon_2. \quad (2.63)$$

Here, ε_1 denotes how much the material is polarized when an electric field is applied, and ε_2 is related with absorption of the material. The imaginary part of interband contribution to the dielectric function is defined as

$$\varepsilon_2^{\alpha\beta} = \frac{\hbar e^2}{\pi m^2 \omega^2} \sum_{cv} \int d\mathbf{k} \langle \Psi_{c\mathbf{k}} | p^{\alpha} | \Psi_{v\mathbf{k}} \rangle \langle \Psi_{v\mathbf{k}} | p^{\beta} | \Psi_{c\mathbf{k}} \rangle (f(\varepsilon_{c\mathbf{k}}) - f(\varepsilon_{v\mathbf{k}})) \delta(\varepsilon_{c\mathbf{k}} - \varepsilon_{v\mathbf{k}} - \omega). \quad (2.64)$$

Here, c and v are the index of conduction band and valence band, respectively. The total imaginary part of dielectric function can be calculated when c and v run over all the conduction bands and valence bands. Similarly, the interband contribution can be achieved by calculating single conduction band and single valence band.

The real part of dielectric function can be calculated by Kramers-Kronig relations

$$\varepsilon_1^{\alpha\beta} = \delta^{\alpha\beta} + \frac{2}{\pi} P \int_0^{\infty} \frac{\omega' \varepsilon_2^{\alpha\beta}(\omega')}{\omega'^2 - \omega^2} d\omega'. \quad (2.65)$$

The absorption coefficient can be obtained by the real part and imaginary part of dielectric function

$$\alpha(\omega) = \frac{\sqrt{2}\omega}{c} \left[\sqrt{\varepsilon_1(\omega)^2 + \varepsilon_2(\omega)^2} - \varepsilon_1(\omega) \right]^{1/2}. \quad (2.66)$$

In this section, only some basic equations are covered when it is related to calculate the dielectric function. If someone is interested in this topic, one can find more detailed in Ref. ??.

2.4 Spin-orbit coupling

2.4.1 Dirac equation

Non-relativistic quantum mechanics has broad application. However the non-relativistic quantum is not suitable to describe the system where the speed of electron is near the speed of light c . Therefore, Dirac introduced an equation which is called Dirac equation applying for relativistic case.

Dirac defined the Hamiltonian as

$$H_{dirac} = c\boldsymbol{\alpha}\mathbf{P} + \beta mc^2 + V. \quad (2.67)$$

Here, $\mathbf{P} = -i\hbar\nabla$ is the momentum operator, V is the general potential, and m is the mass of electron. $\boldsymbol{\alpha}$ and β are 4×4 matrices, which are defined as

$$\boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix}, \beta = \begin{pmatrix} \mathbf{I} & 0 \\ 0 & -\mathbf{I} \end{pmatrix}. \quad (2.68)$$

Here, \mathbf{I} is unit matrix. $\boldsymbol{\sigma}$ is Pauli matrix, which is given as

$$\boldsymbol{\sigma} = (\boldsymbol{\sigma}_x \ \boldsymbol{\sigma}_y \ \boldsymbol{\sigma}_z) \quad (2.69)$$

$$\boldsymbol{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \boldsymbol{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \boldsymbol{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (2.70)$$

2.4.2 Derivation of spin-orbit coupling

Assume that Ψ is the wavefunction of Hamiltonian in Eq. 2.67 which has four components. However, it can be written with only two terms

$$\Psi = \begin{pmatrix} \phi^\uparrow \\ \phi^\downarrow \\ \chi^\uparrow \\ \chi^\downarrow \end{pmatrix}, \Psi = \begin{pmatrix} \phi \\ \chi \end{pmatrix}. \quad (2.71)$$

Here, ϕ includes the two terms of ϕ^\uparrow and ϕ^\downarrow , and χ contains χ^\uparrow and χ^\downarrow . Under the nonrelativistic limit, ϕ is bigger than χ by the ratio of v/c . Here, v and c are speed of

particle and light, respectively. Therefore, the ϕ is considered as the large term and χ is the small one.

In order to derive the spin-orbit coupling term, we need to take use of the nonrelativistic limit approximation ($v^2/c^2 \ll 1$). The time-independent Dirac equation is give as

$$E\Psi = (c\boldsymbol{\alpha}\mathbf{P} + \beta mc^2 + V)\Psi. \quad (2.72)$$

For convenience, the following equation is defined

$$E' = E - mc^2. \quad (2.73)$$

Here, E is the total energy, mc^2 and E' are the rest mass energy and the remaining energy excluding the rest mass energy, respectively. Under the nonrelativistic limit, E' is far smaller than mc^2 . The following equation is given when Eq. 2.71 and Eq. 2.73 are put into Eq. 2.72

$$\begin{aligned} (E' - V)\Phi - c\boldsymbol{\sigma}\mathbf{P}\chi &= 0 \\ -c\boldsymbol{\sigma}\mathbf{P}\Phi + (E' + 2mc^2 - V)\chi &= 0. \end{aligned} \quad (2.74)$$

To eliminate the χ (otherwise, it is the antiparticle problem), it ends up with the equation

$$(V + \frac{1}{2m}(\boldsymbol{\sigma}\mathbf{P})(1 + \frac{E' - V}{2mc^2})^{-1}(\boldsymbol{\sigma}\mathbf{P}))\phi = E'\phi, \quad (2.75)$$

where $E' - V$ is far smaller than $2mc^2$, therefore taking advantage of the taylor expansion of it, as well as the following identites

$$\begin{aligned} [\mathbf{P}, V] &= -i\hbar\nabla V \\ (\boldsymbol{\sigma}\mathbf{A})(\boldsymbol{\sigma}\mathbf{B}) &= \mathbf{AB} + i\boldsymbol{\sigma}[\mathbf{A} \times \mathbf{B}]. \end{aligned} \quad (2.76)$$

The final equation is obtained under the nonrelativistic limit

$$E'\phi = (\frac{\mathbf{P}^2}{2m} + V - \frac{\mathbf{P}^4}{8m^3c^2} - \frac{i\hbar}{4m^2c^2}(\nabla V)\mathbf{P} + \frac{\hbar}{4m^2c^2}\boldsymbol{\sigma}[\nabla V \times \mathbf{P}])\phi. \quad (2.77)$$

Furthermore, we can approximate the above equation to the simpler expression under spherical symmetry potential

$$E'\phi = (\frac{\mathbf{P}^2}{2m} + V - \frac{\mathbf{P}^4}{8m^3c^2} - \frac{i\hbar}{4m^2c^2}(\nabla V)\mathbf{P} + \frac{1}{2m^2c^2}\frac{1}{R}\frac{dV}{dR}\mathbf{SL})\phi. \quad (2.78)$$

Here, $\mathbf{S} = \hbar\boldsymbol{\sigma}/2$ is the Pauli spinor, and $\mathbf{L} = \mathbf{R} \times \mathbf{P}$ is the orbital angular momentum operator. The terms of $\mathbf{P}^2/(2m) + V$ is Schrödinger term, $\mathbf{P}^4/(8m^3c^2)$ and $i\hbar(\nabla V)\mathbf{P}/(4m^2c^2)$ are the mass enhancement and Darwin term, respectively, both of them together is called the scalar relativistic approximation (SRA). The last term is the spin-orbit coupling (SOC) term. One has to notice that the Darwin term is not hermitian operator, alternatively, $\nabla^2V/(8m^2c^2)$ is suggested in the realistic implementation.

Chapter 3

Results and discussion

In this section, the major results for the licentiate thesis are demonstrated. The first one is parameterization of energy bands for $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ (CIGS) where $x = 0, 0.5$, and 1 , as well as indirect results based on the parameterization. The second one is the calculation of the dielectric function spectra for the $\text{CuIn}_{0.5}\text{Ga}_{0.5}\text{Se}_2$ by all electron and full-potential linearized augmented plane wave (FPLAPW) method, which is compared with experimental results as well. While the probable electronic origins of the critical point (CP) features are discussed.

3.1 Parameterization of energy bands for CIGS

3.1.1 Parameterization method

The curvature of energy bands is often demonstrated by the effective electron and hole masses, therefore the parabolic energy dispersion of energy bands, also named parabolic band approximation (pba), is given as

$$E_j^{pb}(\mathbf{k}) = E_j(\mathbf{0}) \pm \left[\frac{\tilde{\mathbf{k}}_x^2 + \tilde{\mathbf{k}}_y^2}{m_j^\perp} + \frac{\tilde{\mathbf{k}}_z^2}{m_j^\parallel} \right] \quad (3.1)$$
$$\tilde{\mathbf{k}}_\alpha^2 = \frac{\hbar \mathbf{k}_\alpha^2}{2e}, \text{ where } \alpha = x, y, \text{ and } z.$$

Here, m_j^\perp and m_j^\parallel are transverse electron masses and longitudinal electron masses, respectively.

Parameters	CuInSe ₂				CuIn _{0.5} Ga _{0.5} Se ₂				CuGaSe ₂			
	c1	v1	v2	v3	c1	v1	v2	v3	c1	v1	v2	v3
$E_j(\mathbf{0})$	0.97	0	0.01	0.19	1.2	0	0.02	0.2	1.47	0	0.08	0.26
m_j^\perp	0.08	0.14	0.25	0.27	0.1	0.4	0.17	0.29	0.13	0.47	0.2	0.29
m_j^{\parallel}	0.09	0.66	0.12	0.28	0.11	0.14	0.61	0.4	0.13	0.15	0.61	0.49

Table 3.1. Parameters of Eq. 3.1 to describe the parabolic energy dispersions of the lowest conduction band (CB) and the three uppermost valence bands (VBs) in the vicinity of the Γ -point. $E_{v1}(\mathbf{0})$ is the valence band maximum (VBM) and $E_{c1}(\mathbf{0})$ is the fundamental band-gap energy Eg.

Unfortunately, these effective masses are only valid around the considered \mathbf{k} point, which is not suitable to describe the non-parabolic away from this \mathbf{k} point. However, the accurate shape of energy bands is important when one simulates and analyzes the electron transport or band filling for some materials.

In this work, we have parameterized the three uppermost VBs and the lowest CB for the materials CuIn_{1-x}Ga_xSe₂ where $x = 0, 0.5$, and 1 , that is, CuInSe₂, CuIn_{0.5}Ga_{0.5}Se₂, and CuGaSe₂. This parameterization is based on the all electron and full-potential linearized augmented plane wave (FPLAPW) calculation. Normally, the $\mathbf{k} \cdot \mathbf{p}$ method is utilized to parameterize the energy bands. However, the energy dispersions of CuIn_{1-x}Ga_xSe₂ are rather complex due to the crystal-field interaction and the spin-orbit coupling. Therefore, the regular $\mathbf{k} \cdot \mathbf{p}$ method is not sufficient to describe the energy bands. We manage to extend the $\mathbf{k} \cdot \mathbf{p}$ expression to higher orders, which is called full band parameterization (fbp) in the following text. Thb fbp is valid around 0.5 eV below VBM and 0.5 eV above conduction band minimum (CBM). The explicit equation is given as

$$\begin{aligned}
E_j(\mathbf{k}) = & E_j^{pb}(\mathbf{k}) + E_j^0 + \Delta_{j,1} \left(\delta_{j,1}^2 \left(\frac{\tilde{\mathbf{k}}_x^4 + \tilde{\mathbf{k}}_y^4}{m_0^2} \right) + \delta_{j,2}^2 \left(\frac{\tilde{\mathbf{k}}_x^2 \tilde{\mathbf{k}}_y^2}{m_0^2} \right) + 1 \right)^{1/2} \\
& + \Delta_{j,2} \left(\delta_{j,3}^3 \left(\frac{\tilde{\mathbf{k}}_x^6 + \tilde{\mathbf{k}}_y^6}{m_0^3} \right) + \delta_{j,4}^3 \left(\frac{\tilde{\mathbf{k}}_x^2 \tilde{\mathbf{k}}_y^4 + \tilde{\mathbf{k}}_x^4 \tilde{\mathbf{k}}_y^2}{m_0^3} \right) + 1 \right)^{1/3} \\
& + \Delta_{j,3} \left(\delta_{j,5}^2 \left(\frac{\tilde{\mathbf{k}}_z^4}{m_0^2} \right) + 1 \right)^{1/2} + \Delta_{j,4} \left(\delta_{j,6}^3 \left(\frac{\tilde{\mathbf{k}}_z^6}{m_0^3} \right) + 1 \right)^{1/3} \\
& + \Delta_{j,5} \left(\delta_{j,7}^2 \left(\frac{\tilde{\mathbf{k}}_x^2 \tilde{\mathbf{k}}_z^2 + \tilde{\mathbf{k}}_y^2 \tilde{\mathbf{k}}_z^2}{m_0^3} \right) + 1 \right)^{1/2} \\
& + \Delta_{j,6} \left(\delta_{j,8}^3 \left(\frac{\tilde{\mathbf{k}}_x^4 \tilde{\mathbf{k}}_z^2 + \tilde{\mathbf{k}}_y^4 \tilde{\mathbf{k}}_z^2}{m_0^3} \right) + \delta_{j,9}^3 \left(\frac{\tilde{\mathbf{k}}_x^2 \tilde{\mathbf{k}}_z^4 + \tilde{\mathbf{k}}_y^2 \tilde{\mathbf{k}}_z^4}{m_0^3} \right) + \delta_{j,10}^3 \left(\frac{\tilde{\mathbf{k}}_x^2 \tilde{\mathbf{k}}_y^2 \tilde{\mathbf{k}}_z^2}{m_0^3} \right) + 1 \right)^{1/3}. \tag{3.2}
\end{aligned}$$

Here, E_j^0 , $\Delta_{j,n}$, and $\delta_{j,m}$ are fitting parameters (Table. 3.2), m_0 is the electron rest mass.

In the Eq. 3.2, each term represents one parabolic dispersion, the higher order terms describe the larger wave vectors away from Γ point. Unfortunately, the complex energy dispersions of $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ requires many fitting parameters. However, the conduction band need less fitting parameters.

3.1.2 Non-parabolicity properties

The fbp of $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ ($x = 0, 0.5$, and 1) are plotted compared with the calculation based on the FPLAPW method and the pba in Eq. 3.1.

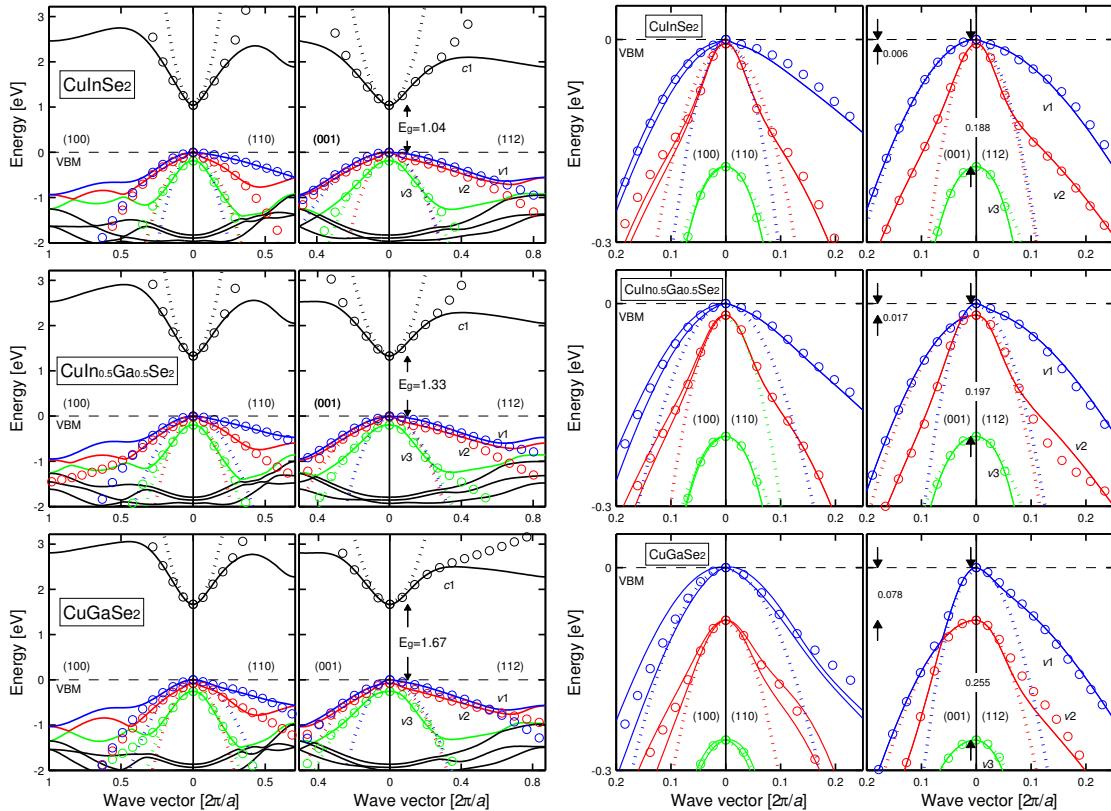


Figure 3.1. Left panel: Electronic band structure along four directions. The circles are the results of the full band parameterization (fbp), and the dotted lines represent the parabolic band approximation (pba). Right panel: The close-up of left panel for the valence bands close to Γ point.

From Fig. 3.1, one will observe that the parameterized energy bands can describe the energy below VBM around 0.5 eV accurately, and around 0.5 eV above the CB minimum (CBM) as well. However, the pba is only valid below VBM around -4 , -10 , and -40 meV for CuInSe_2 , $\text{CuIn}_{0.5}\text{Ga}_{0.5}\text{Se}_2$, and CuGaSe_2 , respectively. Since the lowest conduc-

Parameters	CuInSe ₂			CuIn _{0.5} Ga _{0.5} Se ₂			CuGaSe ₂					
	c1	v1	v2	v3	c1	v1	v2	v3	c1	v1	v2	v3
E_j^0	-4.010	-5.311	-5.242	-5.620	-4.146	-6.210	-6.426	-5.835	-3.927	5.284	-5.789	-2.783
$\triangle_{j,1}$	-0.295	0.006	-0.002	-0.021	-0.230	0.106	1.321	-0.026	-0.454	0.116	1.284	-1.194
$\triangle_{j,2}$	0	0.098	0.104	0.308	0	0.002	0.096	0.386	0	-10.771	-0.837	-0.024
$\triangle_{j,3}$	-0.242	0.018	0.124	-0.025	-0.293	0.937	-0.017	-0.838	-0.419	0.088	0.347	-0.303
$\triangle_{j,4}$	0	0.188	0.076	0.238	0	0.021	0.163	0.789	0	0.076	-0.051	0.608
$\triangle_{j,5}$	-0.016	-0.048	0.001	-0.009	-0.046	0.001	0.011	0.37	-0.047	0.022	0.274	0.374
$\triangle_{j,6}$	0	0.037	-0.073	0.117	0	-0.022	-0.313	-0.012	0	-0.111	-0.525	-4.362
$\delta_{j,1}$	30.669	952	2304.147	94.139	27.157	5.517	1.029	57.007	11.865	10.526	9.269	3.839
$\delta_{j,2}$	47.374	1754.386	4587.156	220.556	44.506	13.61	0.413	153.435	20.262	28.545	21.582	6.232
$\delta_{j,3}$	0	3.97	72.296	11.746	0	487.805	46.95	8.489	0	0.131	9.116	59.625
$\delta_{j,4}$	0	3.688	123.274	18.078	0	1128.668	72.844	13.509	0	0.126	21.328	148.516
$\delta_{j,5}$	31.852	6.041	56.004	64.137	21.124	1.314	247.158	7.921	12.978	7.709	12.141	8.014
$\delta_{j,6}$	0	3.134	6.1	12.031	0	267.523	29.076	8.743	0	64.185	66.808	5.309
$\delta_{j,7}$	222.641	37.004	3846.154	206.148	79.879	3322.259	212.902	16.836	76.319	236.742	16.24	5.092
$\delta_{j,8}$	0	12.647	16.269	6.982	0	92.954	4.885	57.89	0	31.947	6.71	1.209
$\delta_{j,9}$	0	61.565	33.169	34.114	0	118.064	0	273.4	0	34.784	4.831	1.394
$\delta_{j,10}$	0	46.679	31.275	32.237	0	110.327	6.074	153.523	0	40.765	5.243	2.124

Table 3.2. Parameters of Eq. 3.2 to describe the non-parabolic contribution to the energy dispersions $E_j(\mathbf{k})$ of the lowest CB and the three uppermost VBs. The notation of the energy bands ($j = c1, v1, v2$, and $v3$) refers to a spin-independent band indexing, where $c1$ represents the bottommost CB and $v1$ represents the topmost valence band (VB) (see Fig. 3.1).

tion band is more parabolic, therefore, the less fitting parameters are expected (Table 3.2).

To further utilize this parameterization method, the energy dispersions of the lowest CB and the three uppermost VBs are parameterized for Kesterite and Stannite $\text{Cu}_2\text{ZnSn}(\text{S},\text{Se})_4$ as well. It works as good as CIGS, therefore, this parameterization method can be utilized to other materials potentially.

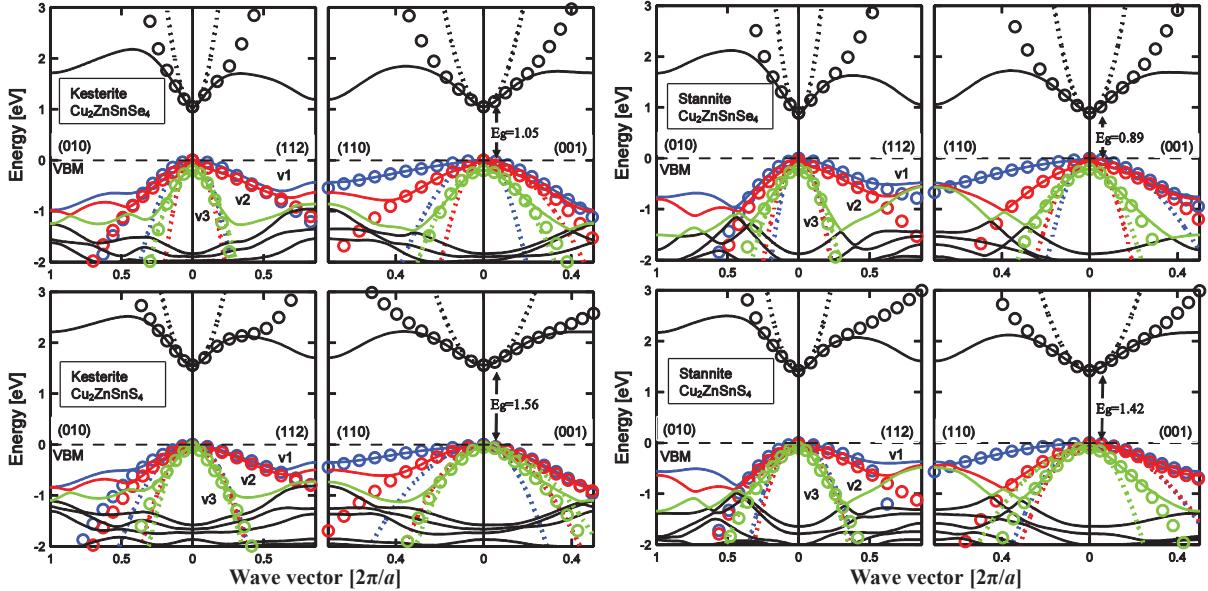


Figure 3.2. Electronic band structure of the kesterite and stannite structures of $\text{Cu}_2\text{ZnSnS}_4$ and $\text{Cu}_2\text{ZnSnSe}_4$ along the symmetry directions ($k_x, k_y, k_z = (010), (112), (110)$, and (001)). The circles are the results of the full band parameterization (fbp), and the dotted lines represent the parabolic band approximation (pba).

In order to demonstrate the anisotropic and non-parabolic of energy bands further, the constant energy surface $S_j(E)$ is determined for CuInSe_2 and CuGaSe_2 .

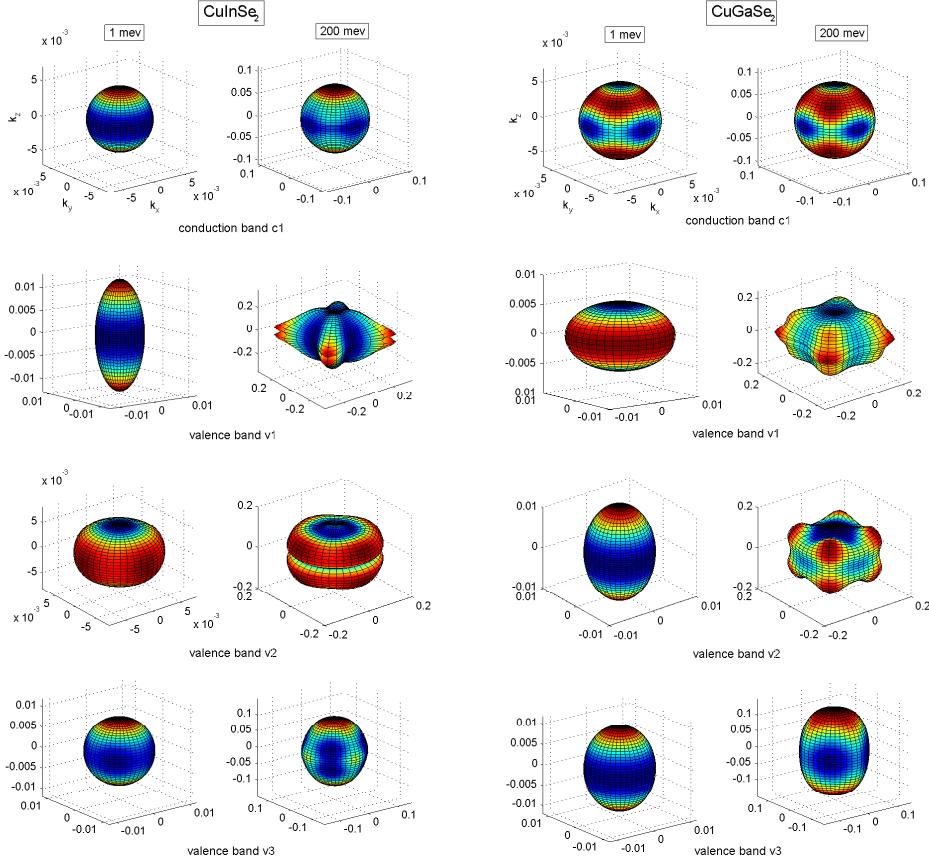


Figure 3.3. Constant energy surfaces for the three uppermost VBs and the lowest CB for the energies $E = 1$ meV (left column ellipsoidal) and $E = 200$ meV (right column). The CuInSe_2 and CuGaSe_2 are demonstrated in this figure.

In the **Fig. 3.3**, 1 meV represents that it is near to Γ point ; 200 meV implies that it is far away the Γ point. One will notice that the pba is proper to describe the energy bands close to the Γ point, and it is ellipsoidal shaped sphere. For example, for the topmost VB (v_1) of CuInSe_2 , the constant energy surface is ellipsoidal in the vicinity of the Γ point since the effective masses are anisotropic ($m_{v1}^{\perp} = 0.14m_0$ and $m_{v1}^{\parallel} = 0.66m_0$ in **Table 3.1**). However, the constant energy surface becomes non-ellipsoidal when the energies is far away from Γ point. For example, for the same band, the constant energy surface is not ellipoidal shape at all when the energies goes up to 200 meV. It also implies that it is not correct to consider the constant effective mass for simulation. For the CB, the change is relatively small.

3.1.3 Application of the full band parameterization

The parameterized energy bands can be exploited to reveal the detailed information near the VB and CB edged. In this section, the effective masses, density-of-states (DOS), ... are calculated based on this fbp. In comparison with the results based on pba, most of the results are calculated with the pba as well.

The effective masses are \mathbf{k} -independent for the pba, which is not fully correct due to the non-parabolicity and anisotropy of the energy bands. The effective masses tensors are calculated numerically by

$$m_j(\mathbf{k}) = \pm \hbar^2 / (\partial^2 E_j(\mathbf{k}) / \partial \mathbf{k}^2) \text{ where } j = c1, v1, v2, \text{ and } v3. \quad (3.3)$$

The result from **Fig. 3.4** demonstrates that the energy bands of $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ are strong non-parabolic, since it should be constant in the pba along each symmetry direction. The effective hole masses of the two topmost VBs are very anisotropic close to the Γ point, however, the electron masses of conduction bands are rather isotropic ($m_{c1}^{100}(\mathbf{0}) \approx m_{c1}^{110}(\mathbf{0}) \approx m_{c1}^{001}(\mathbf{0}) \approx m_{c1}^{112}(\mathbf{0})$). Hole masses of CuGaSe_2 vary somewhat less compared with those of CuInSe_2 due to CuGaSe_2 has larger split between the VBs. In order to better visibility, therefore, the inverse of the mass are presented.

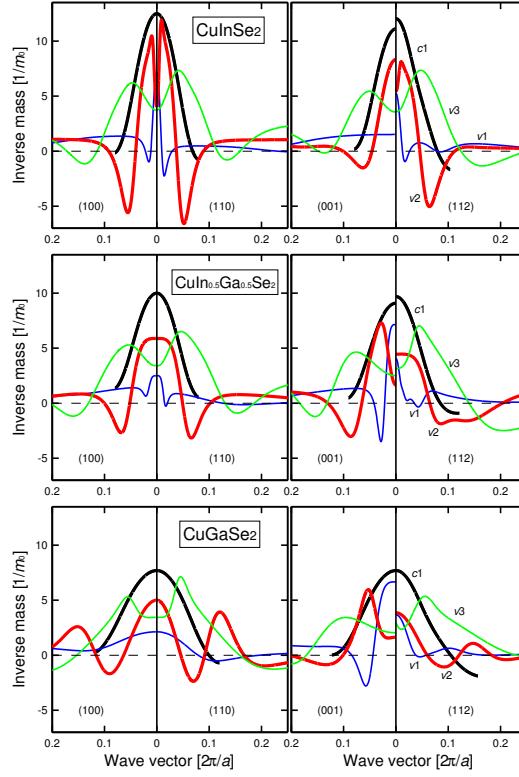


Figure 3.4. Inverse of the effective electron and hole masses in the four symmetry directions for the CuIn_{1-x}Ga_xSe₂ (x=0, 0.5, and 1).

In order to further analyze the impact of non-parabolicity and anisotropy of the energy bands, the density-of-states (DOS) is calculated based on fba and pba. The DOS in j th band is defined as

$$g_j(E) = \frac{1}{\Omega} \sum_{\mathbf{k}} 2\delta(E - E_j(\mathbf{k})) = \frac{1}{4\pi^3} \int_{E_j(\mathbf{k})=E} \frac{dS(\mathbf{k})}{|\nabla_{\mathbf{k}} E_j(\mathbf{k})|}. \quad (3.4)$$

Here, $E_j(\mathbf{k}) = E$ is the \mathbf{k} space surface with constant energy E , and the $\nabla_{\mathbf{k}} E_j(\mathbf{k})$ is the gradient of the energy dispersion. In the case of the pba, the Eq. 3.4 can be written as

$$g_j^{pba}(E) = \frac{1}{2\pi^2} \left(\frac{2m_j^{DOS}}{\hbar^2} \right)^{3/2} \sqrt{|E - E_j(\mathbf{0})|}. \quad (3.5)$$

Here, the DOS mass m_j^{DOS} is equal to $(m_j^\perp m_j^\perp m_j^{\parallel})^{1/3}$, which represents the extent of filling the specific band with free carriers to certain energy. In order to take advantage of the simple Eq. 3.5 for the non-parabolic energy bands, the energy-dependent DOS mass

$(m_{v/c}^{DOS})$, which contains the non-parabolicity and anisotropy of the band dispersion, is defined as

$$g_{v/c}(E) = \sum_j g_j(E) = \frac{1}{2\pi^2} \left(\frac{2m_{v/c}^{DOS}(E)}{\hbar^2} \right)^{3/2} \sqrt{|E - E_{v1/c1}(\mathbf{0})|}. \quad (3.6)$$

Here, the energy-dependent DOS mass $m_{v/c}^{DOS}(E)$ contains the non-parabolicity and anisotropy of the band dispersion.

The Fig. 3.5 (left) demonstartes that the non-parabolicity of the bands strongly affect the DOS dispersions. The difference is remarkable. The fbp always generates larger DOS, the reason is that the non-parabolic energy bands is more flat than parabolic bands. The Fig. 3.5 (right) demonstrates that the DOS masses of $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ is strong energy dependence with the VB DOS mass, which proves further the importance of considering non-parabolicity and anisotropy of the energy bands, especially for the VBs in the case of $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$. For example, the effecive mass of topmost VB for CuInSe_2 is around $0.23 m_0$ close to the Γ point, which goes up to around $1.00 m_0$ when E is around 0.1 eV. However, the change in the CB DOS mass is subtle, but also goes up to 2–3 times with respect to the value around Γ point.

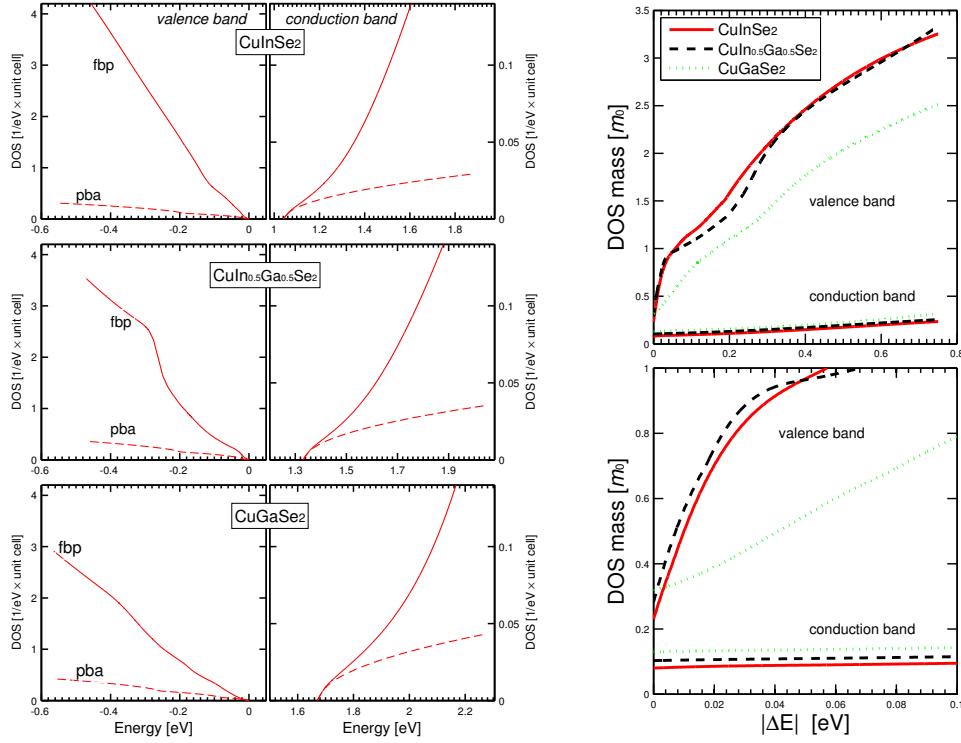


Figure 3.5. Left: Total DOS of the VBs and CB. The solid lines show the full band parameterization (fbp), and the dashed lines represent the parabolic band approximation (pba). Right: The DOS mass $m_{v/c}^{DOS}$ is calculated in the Eq. 3.6.

The number of states with energies up to E is obtained from the total DOS and the constant energy surfaces in Fig. . It shows that the VBs (CB) in Ga rich compounds will be less (more) populated by holes (electrons) for a given quasi-Fermi energy. It also demonstrates that the pba underestimates the band filling in the VBs and the CB strongly. For example, the number of VB states of CuInSe₂ is increased by a factor of around 18 at the positive energy $|\Delta E| = E_{v1}(\mathbf{0}) - E_{F,v}^* = 0.1\text{eV}$ when the fba is included. For the CB, the number is increased by a factor of around 3. At $|\Delta E| = 0.5\text{eV}$, the increase is as much as around 41 and 8 times, respectively.

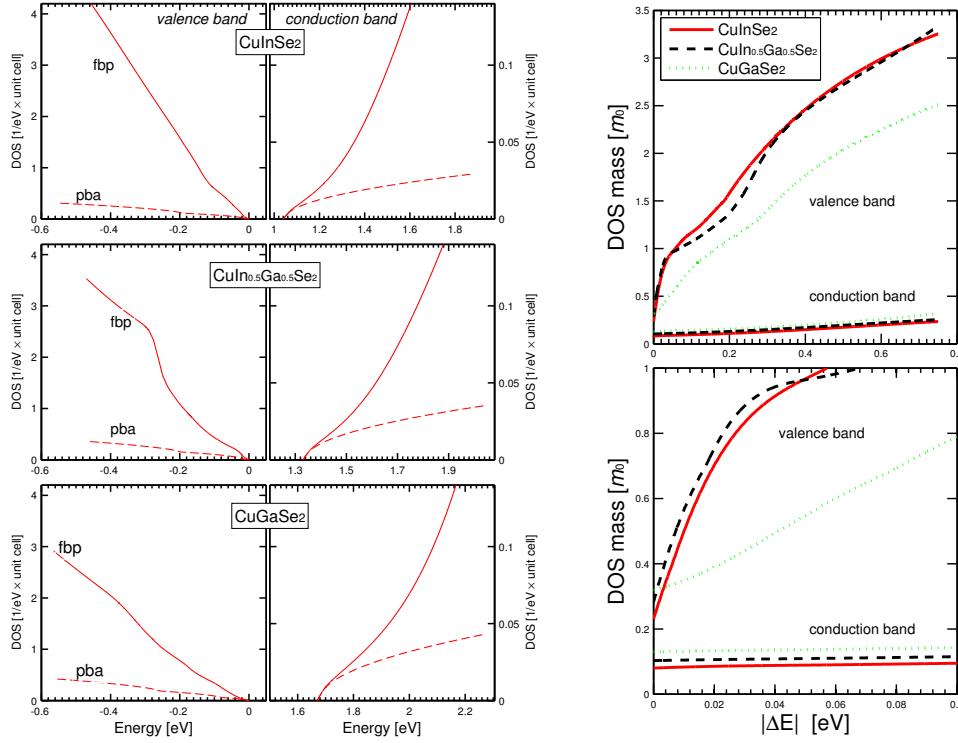


Figure 3.6. Carrier concentration p or n as functions of the quasi-Fermi energy E_F^* of the VBs $E_{F,v}^*$ and of the CB $E_{F,c}^*$. Left column shows the results for large energy scale up to 0.5 eV, and right column displays a close-up for small Fermi energies. In the figure, $|\Delta E|$ is the positive energy difference $E_{v1}(\mathbf{0}) - E_{F,v}^*$ for the VBs and $E_{F,c}^* - E_{c1}(\mathbf{0})$ for the CB. The carrier concentrations consider external band filling in intrinsic materials at $T = 0$ K. The results demonstrate that the parabolic band approximation strongly underestimates the band filling of both the VBs and CB.

The concentration of free holes $n_v(T)$ and free electrons $n_c(T)$ is calculated from the DOS by

$$n_v(T) = \int_{-\infty}^{E_{v1}(\mathbf{0})} g_v(E)(1 - f(E))dE,$$

$$n_c(T) = \int_{E_{c1}(\mathbf{0})}^{\infty} g_c(E)f(E)dE. \quad (3.7)$$

Here, $f(E) = 1/[1 + \exp(E - E_F)/k_B T]$ is the Fermi distribution function. The intrinsic carrier concentration can be expressed as

$$n_i(T) = \sqrt{n_c(T) \cdot n_v(T)}. \quad (3.8)$$

The extrinsic carrier concentration for *p*-type materials can be derived as:

$$n_v(T) = \frac{n_i^2(T)}{n_v(T)} + \sum_{\alpha} \frac{N_{A_{\alpha}}}{1 + g_{A_{\alpha}} e^{(\Delta_{A_{\alpha}} - E_F)/k_B T}}. \quad (3.9)$$

Here, $N_{A_{\alpha}}$ is the acceptor concentration of the α th defect, $\Delta_{A_{\alpha}}$ implies the energy level of the acceptor state, and the $g_{A_{\alpha}}$ is the spin degeneracy factor. The measured ionization energies for V_{Cu} are utilized from Ref. ??.

The free carrier concentration in intrinsic for $CuIn_{1-x}Ga_xSe_2$ ($x = 0, 0.5$, and 1) is obtained considering the temperature dependency of the band gaps (Eq. 3.10). The results from the pba for silicon (Si) and GaAs are compared with our simulation as well.

$$E_g(T) = E_g(0) - \frac{a \cdot T^2}{b + T}. \quad (3.10)$$

Here, parameters a and b are exploited from experimental values.

In Fig. 3.7, the temperature dependent band gap and Fermi level (left panel) and intrinsic carrier concentration (right panel) are shown. In the left panel of Fig. 3.7, the band gap is 1.04, 1.33, and 1.67 eV for $CuIn_{1-x}Ga_xSe_2$ where $x = 0, 0.5$, and 1 at temperature around 0 K, respectively. Fermi level is exactly the mid-gap energy. The Fermi level changes only slightly with temperature. However, for $CuInSe_2$ and $CuIn_{0.5}Ga_{0.5}Se_2$, the Fermi levels increase somewhat more compared with $CuGaSe_2$. This is primarily due to the CB of DOS mass for $CuGaSe_2$ is almost the same, but the VB of DOS mass for $CuGaSe_2$ change smaller compared with $CuInSe_2$ and $CuIn_{0.5}Ga_{0.5}Se_2$ which will affect the DOS. As a consequence, the Fermi level is closer to the CB minimum in the In rich compounds. At $T = 300$ K and 600 K, the band-gap energies and Fermi energy are $E_g(300) = 1.02, 1.29$, and 1.62 eV, $E_g(600) = 0.98, 1.24$, and 1.55 eV, $E_F(300) = 0.55, 0.69$, and 0.84 eV, and $E_F(600) = 0.59, 0.71$, and 0.84 eV for $x = 0, 0.5$, and 1 respectively. In the right panel of Fig. 3.7, the intrinsic carrier concentration is increased dramatically with the increasing of temperature for the $CuIn_{1-x}Ga_xSe_2$. For example, in the case of $CuInSe_2$, the carrier concentration is increased up to around 10^5 times higher from temperature 300 K to 600 K. One also notice that the intrinsic carrier contribution for the Si and $CuInSe_2$ are very comparable. The free carrier concentration is increased goes up to 2 – 3 times by taking into account the non-parabolicity of the energy bands.

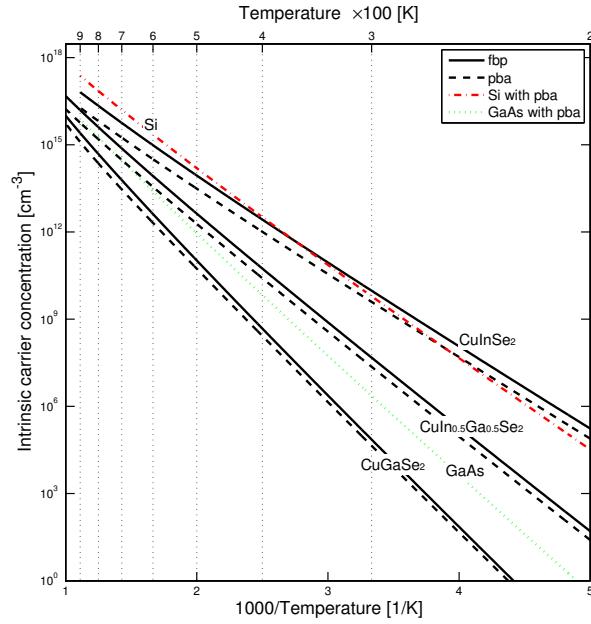


Figure 3.7. Intrinsic carrier concentration as function of temperature.

Since as-grown $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ typically has *p*-type character, the Fermi level and *p*-type carrier concentration of free holes $n_v(T)$ and free electrons $n_c(T)$ is calculated by assuming the presences of the native Cu vacancies as acceptors.

The calculated Fermi level E_F^P in *p*-type materials is presented in **Fig. 3.8**, referred to the Fermi level of the intrinsic materials E_F from **Fig. 3.7**. Only at very high temperatures ($T > 400K$) and for low acceptor concentrations, the Fermi level of *p*-type $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ will reach the Fermi level of corresponding intrinsic compounds. Moreover, although the different compounds have comparable acceptor ionization energies, the Ga rich alloy has lower relative Fermi level; this is a direct consequence of the larger band gap of the Ga rich alloy. By comparing the calculations with the parabolic band approximation (dotted lines in **Fig. 3.8**) and the full band approach (solid lines) for each of the three $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ alloys, one can notice that the Femi level is similar for the two models only at low and at very high temperatures. In the mid-temperature region the difference is however apparent, especially for the high acceptor concentrations. Assuming parabolic bands yields always a lower Fermi level. The reason for this effect is that the energy dependent effective mass $m_{v/c}^{DOS}(E)$ in the full band parameterization is always larger than the corresponding Γ -point mass. Although the effect on the absolute value of the Fermi level seems to be small, it has an impact on the free carrier concentrations for highly doped materials.

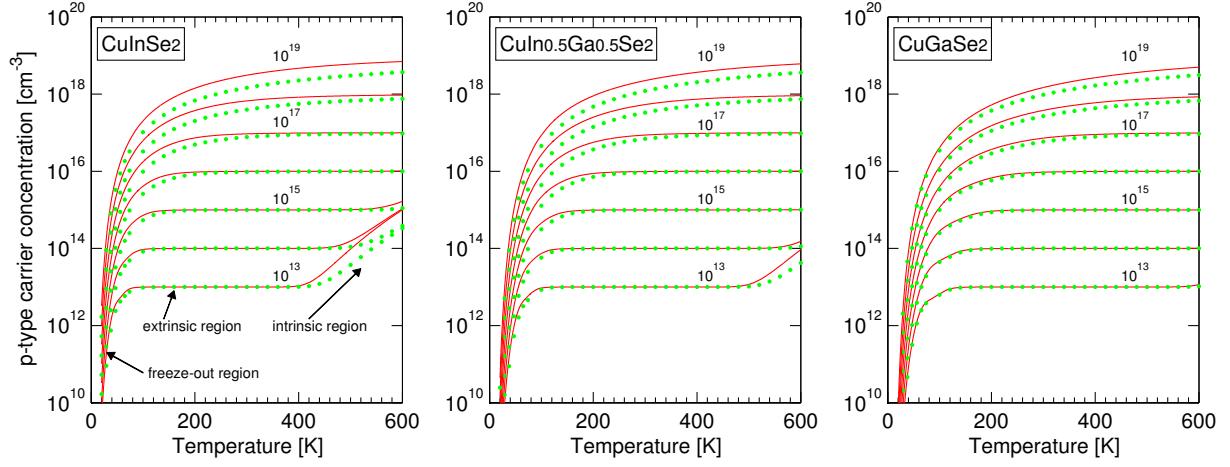


Figure 3.8. Fermi level as function of the temperature $20 \leq T \leq 600$ of *p*-type CuInSe₂, CuIn_{0.5}Ga_{0.5}Se₂, and CuGaSe₂ for the effective doping concentration $N_A = 10^{13}, 10^{14}, 10^{15}, \dots, 10^{19}$ acceptors/cm³. The energy scale $E_F^p - E_F$ describes the Fermi energy with respect to the intrinsic E_F (**Fig. 3.7**). Dashed lines represent the VBM with respect to the intrinsic Fermi level. Solid and dotted lines represents the full band parameterization and the parabolic band approximation, respectively.

The p-type carrier concentration (**Fig. 3.9**) for CuIn_{1-x}Ga_xSe₂ is presented using Eq. 3.9

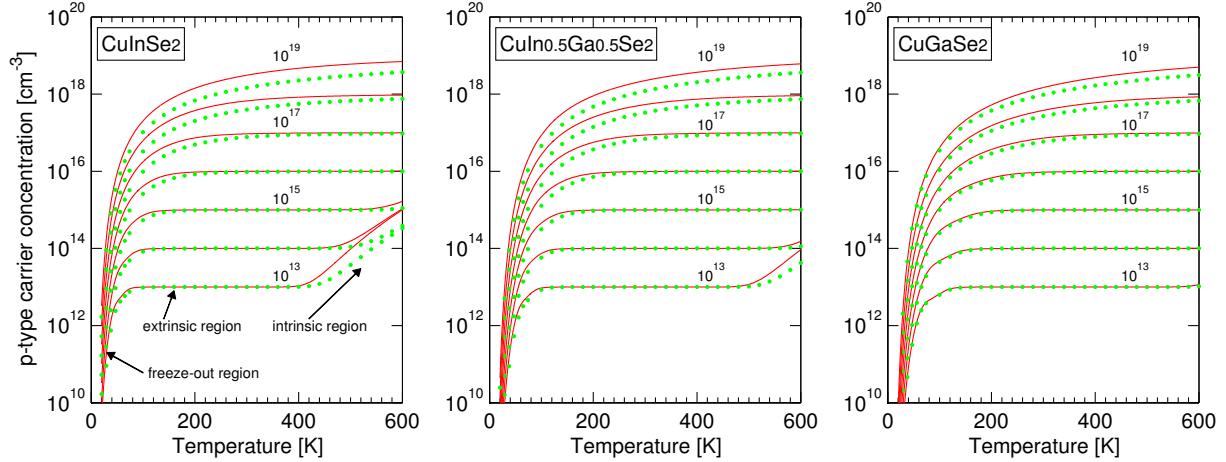


Figure 3.9. Free carrier concentration as function of the temperature in *p*-type for CuInSe_2 , $\text{CuIn}_{0.5}\text{Ga}_{0.5}\text{Se}_2$, and CuGaSe_2 . The effective doping concentration $N_A = 10^{13}, 10^{14}, 10^{15}, \dots, \text{and } 10^{19} \text{ acceptors}/\text{cm}^3$ are considered.

From **Fig. 3.9**, the carrier concentration is recognized by three different regions: the freeze-out region, the extrinsic region and the intrinsic region. The transition from the freeze-out region to the extrinsic region happens below the room temperature except that the uncompensated acceptor concentration is above around 10^{18} cm^{-3} . The transition from the extrinsic ergion to the intrinsic region for In rich compounds occurs at the lower temperature since they have smaller band gaps. The result based on the pba underestimates the carrier concentration around by the factor of 2 in the both freeze-out and intrinsic regions. Therefore, the non-parabolic energy bands is required in order to describe the carrier concentration more accurately.

3.2 Calculations of dielectric function for CIGS

In this work, the dielectric function (ε) spectra of $\text{CuIn}_{0.5}\text{Ga}_{0.5}\text{Se}_2$ is calculated by the full-potential linearized augmented plane wave (FPLAPW) method using the generalized gradient approximation (GGA) plus an onsite Coulomb interaction U of the Cu *d* states. Afterwards, the different contributions to $\varepsilon_2(\text{Im}(\varepsilon))$ in terms of the transitions between the valence bands and the conduction bands are identified. At last, the \mathbf{k} -dependence of the interband critical points (CPs) along the main symmetry directions is analyzed. The result is compared with experimental work ($\text{CuIn}_{0.7}\text{Ga}_{0.3}\text{Se}_2$) at temperature of 40 K and 300 K, and they are in a good agreement.

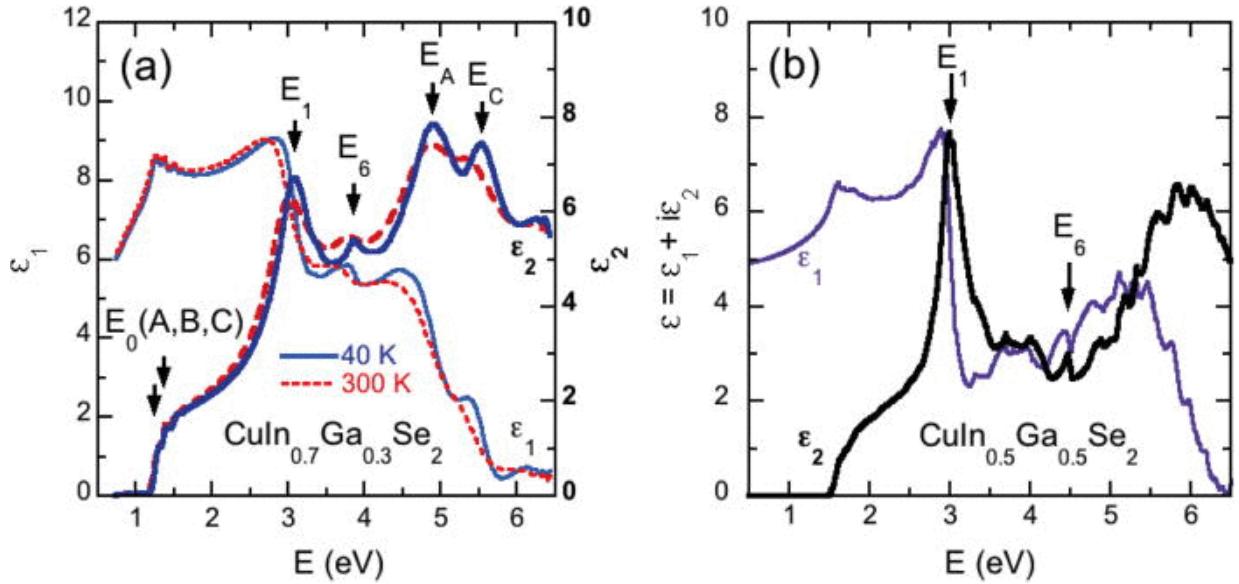


Figure 3.10. Left panel: The real (ϵ_1) and imaginary (ϵ_2) part of dielectric function spectra for $CuIn_{0.7}Ga_{0.3}Se_2$ at 40 K (solid blue line) and 300 K (dashed red lines). Four prominent CP features are shown. Right panel: the dielectric function spectra for $CuIn_{0.5}Ga_{0.5}Se_2$ calculated by FPLAPW method at 0 K. The major CP features are identified.

From Fig. 3.10, one will notice that the general shape between experimental and calculated result is similar. The calculation indicates that there is no big difference in the optical properties for those two materials, except the shift of CP energies.

The analysis based on experimental work indicates that there are twelve CPs from 2.5 eV to 6.4 eV. The electronic origin for each CP is analyzed based on the calculated result. First, the contribution to dielectric function between the valence bands and conduction bands is presented.

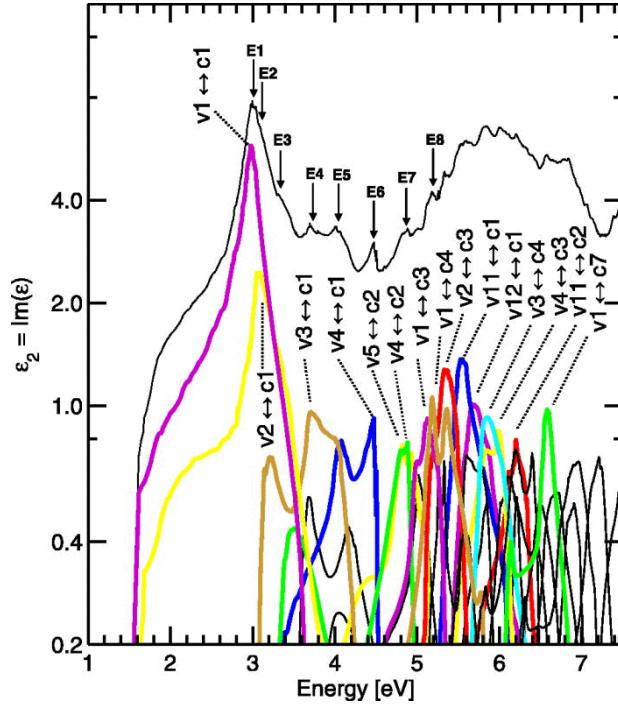


Figure 3.11. Band-to-band analysis of the contribution to the total ε_2 spectrum. The vertical axis is in the log scale.

Here, v_1 and c_1 represent the topmost valence band and lowest conduction band, respectively. From the Fig. 3.11 and Fig. 3.12, one will notice that the E_1 CP comes from the $v_1 \rightarrow c_1$ transition near the $P(1/2, 1/2, 1/2)$ point of the Brillouin zone (BZ). The E_2 and E_3 CPs are corresponding to transition of $v_2 \rightarrow c_1$ in the P point as well in the BZ. The E_2 and E_3 CPs are small peak in the Fig. 3.11. However, the calculation of $CuInSe_2$ indicates that they happens 0.1–0.2 eV higher than the E_1 CP, which is distinct spectral features (Fig. 3.13). The E_4 CP comes from the transition of $v_3 \rightarrow c_1$ at the $M(1, 0, 0) = M^*(0, 0, 1)$ point. The E_5 CP is contributed by the transitions $v_4 \rightarrow c_1$ at the $N(1/2, 0, 1/2)$ point and $v_3 \rightarrow c_1$ at the M/M^* point. The E_6 CP feature corresponding to the $v_4 \rightarrow c_1$ at the N point. The E_7 is from the transitions $v_4 \rightarrow c_2$ at the $\Gamma(0, 0, 0)$ and N point, the $v_5 \rightarrow c_2$ at the Γ point is also contributed.

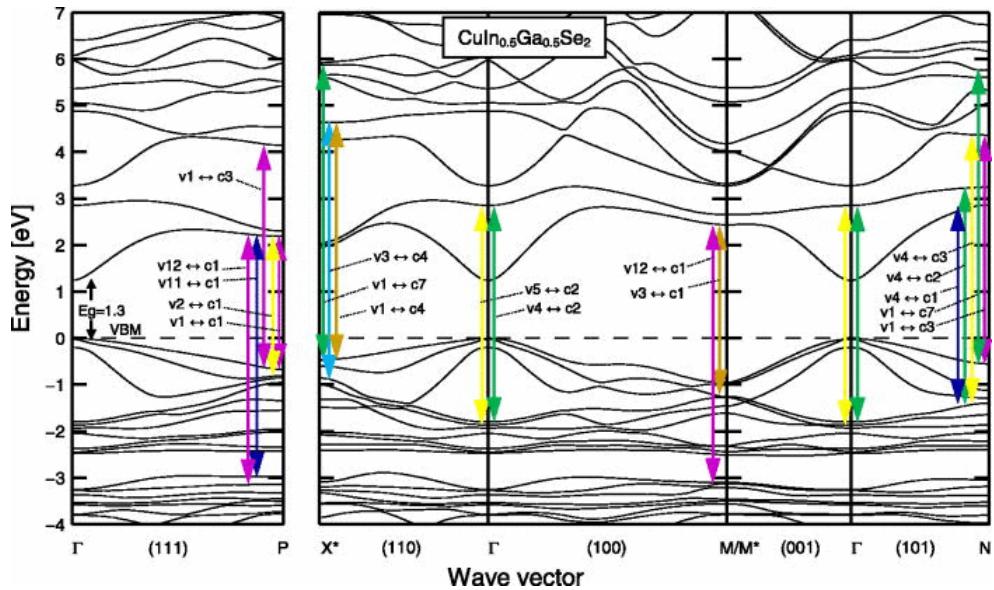


Figure 3.12. Band-to-band analysis of the contribution to the total ε_2 spectrum. The vertical axis is in the log scale.

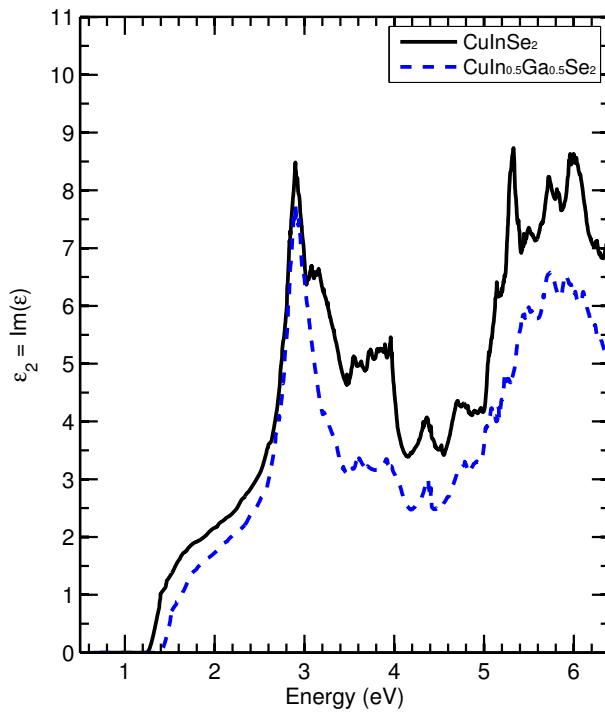


Figure 3.13. The ε_2 spectra for CuInSe₂ and CuIn₀·₅Ga₀·₅Se₂.

Chapter 4

Summary of the publications

Chapter 5

Concluding remarks and future work

Acknowledgements

Chapter 6

Appendices

In order to solve the Hartree, Hartree-Fock and Kohn-Sham equation, some mathematical knowledge are need to be introduction in this section, that is, Lagrange multiplier and functionals.

Lagrange multiplier

Lagrange multiplier is a method which can maximize or minimize functions with the constraint. In order to maximize or minimize a function $f(x, y, z)$ with constraint $g(x, y, z) = c$, the generalized format of lagrange multiplier method is given as

$$\begin{aligned}\frac{\partial (f(x, y, z) - \lambda g(x, y, z))}{\partial x} &= 0 \\ \frac{\partial (f(x, y, z) - \lambda g(x, y, z))}{\partial y} &= 0 \\ \frac{\partial (f(x, y, z) - \lambda g(x, y, z))}{\partial z} &= 0 \\ \frac{\partial (f(x, y, z) - \lambda g(x, y, z))}{\partial \lambda} &= 0.\end{aligned}\tag{6.1}$$

For example, assume that the ground state wavefunction is Ψ_0 , and ground state total energy E_0 is give as

$$E_0 = \min(E) = \min \langle \Psi(\mathbf{r}) | \widehat{H} | \Psi(\mathbf{r}) \rangle, \text{ and } \Psi(\mathbf{r}) \longrightarrow \Psi_0(\mathbf{0}).\tag{6.2}$$

The constraint of wavefunction is normalization

$$\langle \Psi(\mathbf{r}) | \Psi(\mathbf{r}) \rangle = 1.\tag{6.3}$$

Therefore, problem of the ground state total energy can be solved by Lagrange multi-

plier method

$$\begin{aligned}\frac{\partial(E - \lambda g)}{\partial\Psi(\mathbf{r})} &= \frac{\partial\left(<\Psi(\mathbf{r})|\widehat{H}|\Psi(\mathbf{r})> - \lambda(<\Psi(\mathbf{r})|\Psi(\mathbf{r})> - 1)\right)}{\partial\Psi(\mathbf{r})} = 0, \\ \frac{\partial(E - \lambda g)}{\partial\lambda} &= \frac{\partial\left(<\Psi(\mathbf{r})|\widehat{H}|\Psi(\mathbf{r})> - \lambda(<\Psi(\mathbf{r})|\Psi(\mathbf{r})> - 1)\right)}{\partial\lambda} = 0.\end{aligned}\quad (6.4)$$

Functional

A functional takes a function as an input, and output is a number. In previous section, the total energy E is the function of wavefunction $\Psi(\mathbf{r})$, which is a functional. In order to solve the Eq. 6.4, basic properties of functionals and their derivatives are needed to introduced.

Assume that a set of points are defined as $y^0 = (y_1^0, y_2^0, \dots, y_N^0)$. A function with multivariables is given as $F(y_1, y_2, \dots, y_N)$, then a small change away from y^0 leads to the change of function $F(y_1, y_2, \dots, y_N)$ by

$$\delta F(y_1, y_2, \dots, y_N) = \sum_{i=1}^N \frac{\partial F(y_1, y_2, \dots, y_N)}{\partial y_i} \Big|_{y_i^0} \delta y_i. \quad (6.5)$$

A one-to-one map $y_n = g(x_n)$ is defned where $g(x_n)$ has N finite number of variables. In order to be utilized in physics, the $g(x_n)$ can be allowed all the real numbers in certain interval, such as $[x_0, x_N]$, that is, N is allowed to be infinite number (∞).

Assume that x_n is defined within interval $[x_0, x_N]$, and there are N points in this interval, as well as $x_{n+1} - x_n = \varepsilon$. Therefore, $x_N - x_0 = N\varepsilon$, that is, $x_n = n\varepsilon + x_0$, and $y_n = g(x_n) = g(x_0 + n\varepsilon)$. Therefore, $g(x_n) = g(x)$ with the limit $\varepsilon \rightarrow 0$ and $N \rightarrow \infty$, where x can be any real number in the interval. That is, $F(g(x_1), g(x_2), \dots, g(x_N)) = F(g(x))$ under the limit of $N \rightarrow \infty$, which is a functional.

The derivative of function can lead to the derivative of functionals. If the y_n is substituted for $g(x_n)$ in Eq. 6.5, the new equation is given as

$$\delta F(g(x_1), g(x_2), \dots, g(x_N)) = \sum_{i=1}^N \frac{\partial F(g(x_1), g(x_2), \dots, g(x_N))}{\partial g(x_i)} \Big|_{g^0(x_i)} \delta g(x_i). \quad (6.6)$$

Recall the definition of integral for function

$$\int_{x_0}^{x_N} f(x) dx = \lim_{\varepsilon_i \rightarrow 0} \sum_{i=1}^N \varepsilon_i f(x_i), \text{ where } N \rightarrow \infty. \quad (6.7)$$

Therefore, the Eq. 6.6 is equivalence to

$$\begin{aligned}
& \delta F(g(x_1), g(x_2), \dots, g(x_N)) \\
&= \sum_{i=1}^N \frac{\partial F(g(x_1), g(x_2), \dots, g(x_N))}{\partial g(x_i)} \Big|_{g^0(x_i)} \delta g(x_i) \\
&= \sum_{i=1}^N \varepsilon_i \frac{\partial F(g(x_1), g(x_2), \dots, g(x_N))}{\varepsilon_i \partial g(x_i)} \Big|_{g^0(x_i)} \delta g(x_i) \\
&= \int_{x_0}^{x_N} \frac{\partial F(g(x))}{\partial g(x)} \Big|_{g^0(x)} \delta g(x) dx \\
&= \int \frac{\partial F(g(x))}{\partial g(x)} \delta g(x) dx \\
&= F[g(x) + \delta g(x)] - F[g(x)]
\end{aligned} \tag{6.8}$$

In order to calculate the derivative of a functional in reality, first, $F[g(x) + \delta g(x)] - F[g(x)]$ is solved using Taylor expansion; Second, the result from first step is compared with $\int \frac{\partial F(g(x))}{\partial g(x)} \delta g(x) dx$. For example, a funcional $F[g(x)] = \int f(g(x)) dx$ is defined, and

$$\begin{aligned}
& F[g(x) + \delta g(x)] - F[g(x)] \\
&= \int f(g(x) + \delta g(x)) dy - \int f(g(x)) dy \\
&= \int (f(g(x)) + \frac{\partial f(g(x))}{\partial g(x)} \delta g(x) + O(\delta g(x))^2) dx - \int f(g(x)) dx \\
&= \int \frac{\partial f(g(x))}{\partial g(x)} \delta g(x) dx
\end{aligned} \tag{6.9}$$

Therefore,

$$\frac{\partial F(g(x))}{\partial g(x)} \delta g(x) = \frac{\partial f(g(x))}{\partial g(x)} \delta g(x) \tag{6.10}$$

That is,

$$\frac{\partial F(g(x))}{\partial g(x)} = \frac{\partial f(g(x))}{\partial g(x)} \tag{6.11}$$

The result in Eq. 6.11 is useful to derive the solution in Hartree, Hartree Fock and Kohn-Sham equation.