



**KTH Industrial Engineering
and Management**

SEPTEMBER 25, 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 CuIn_{1-x}Ga_xSe₂ (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 CuIn_{1-x}Ga_xSe₂**
R. Chen and C. Persson, *J. Appl. Phys.* **112**, 103708 (2012).
- III **Dielectric function spectra at 40 K and critical-point energies for CuIn_{0.7}Ga_{0.3}Se₂**
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).

Comment on my own contribution

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:

- 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 CuInSe₂**
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 CuIn_{0.5}Ga_{0.5}Se₂ 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	9
1.2.3 Thin film materials	10
1.3 CIGS materials	10
1.3.1 Crystal structure	11
1.3.2 Optical properties and doping in the CIGS	12
1.3.3 CIGS solar cell structure	12
2 Theory	14
2.1 Electronic structure calculations	14
2.1.1 The quantum many-body problem	14
2.1.2 The Born-Oppenheimer approximation	15
2.1.3 Hartree, Hartree-Fock approximation and density functional theory	17
2.1.3.1 Hartree approximation	18
2.1.3.2 Hartree-Fock approximation	20
2.1.3.3 Density functional theory	22
2.1.3.4 Kohn-Sham equation	24
2.1.3.5 The exchange-correlation energy	26
2.1.4 Solving the secular equation	27
2.1.5 Eigenvalue problem	28
2.2 Full-potential linearised augmented plane wave method	29
2.2.1 Introduction	29
2.2.2 Wavefunction	30
2.2.2.1 Augmented plane wave method	30
2.2.2.2 Linearized augmented plane wave method	31

2.2.2.3	Linearized augmented plane wave method plus local orbitals	32
2.2.2.4	Augmented plane wave method plus local orbitals	32
2.2.3	Effective potential	32
2.3	Dielectric function	33
3	Results and discussion	34
3.1	Modeling and calculations	34
3.1.1	Copper indium gallium diselenide	34
3.1.1.1	Parameterization of energy bands	34
3.1.1.2	Dielectric function spectra	44
4	Concluding remarks and future work	48
	Acknowledgements	49

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 2012 (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 around 12000 million tonnes oil equivalent (METO), which is equivalent with around 15 terawatts. Normally lightbulbs in our homes consume around 100 watts of energy, 1 terawatt means is 10 billion of the lightbulbs is 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.

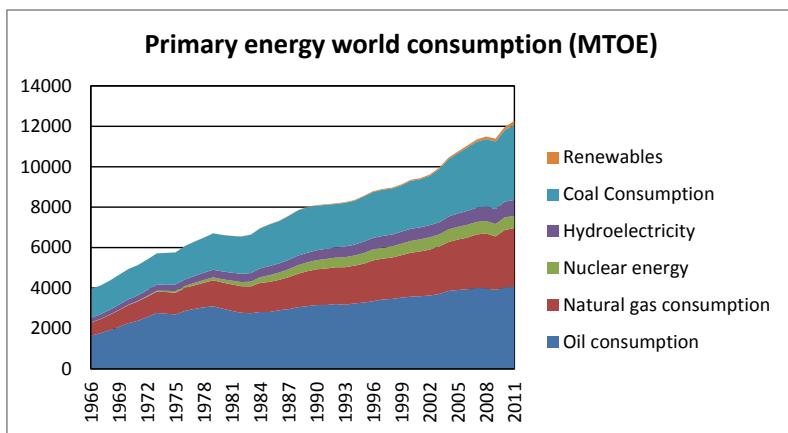


Figure 1.1. Data is from BP statistical review of world energy 2012, METO means million tonnes oil equivalent.

By the year of 2050, the total world energy consumption will around double. Therefore, it is urgent to explore more sustainable and healthy energy sources. From **Fig. 1.1**, one will notice that renewable energy (mainly solar energy, wind power, hydropower and geothermal energy) in 2012 (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 more or less a black body with a temperature of about 6000 K (**Fig. 1.2**). In photovoltaic field, solar spectrum is represented by Air Mass (AM), and 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.

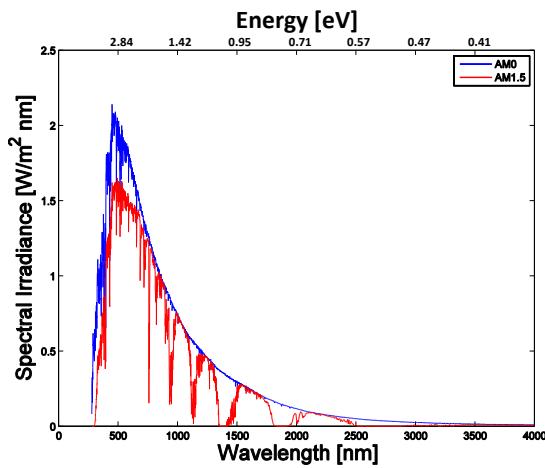


Figure 1.2. Solar irradiance spectrum (Y-axis ??).

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 businesses, homes, and pools. Therefore, the solar thermal collectors do not convert sunlight to electricity directly, but transfer the energy to the water instead. The advantage is the conversion efficiency is relatively higher, however, the price is higher as well and it needs huge amount of cherished water resources. The second one is solar chemical, which takes advantage of solar energy by absorbing sunlight in a chemical reaction, and the conversion efficiency is quite low. The last one is solar photovoltaics (solar cell), which is the way to use solar panels to convert sunlight into electricity. The price is relatively lower compared with solar thermal, and the conversion efficiency is higher than solar chemical. More importantly, it is environmentally friendly.

1.1 Solar cells

In the worldwide, many researchers are working on the field of solar cell. Therefore, 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 44.4%, 27.6%, 22.8% and 16.2%, respectively. Therefore, the solar cell is a very important and promising way to produce the renewable energy.

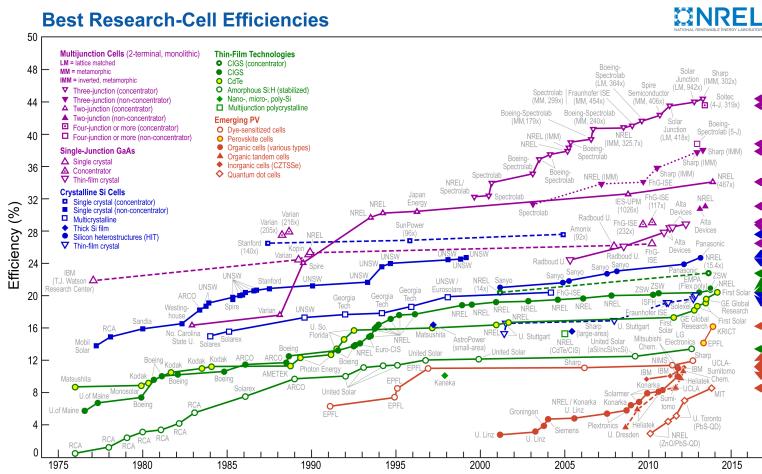


Figure 1.3. Best research-cell efficiencies. Figure is from National Renewable Energy Laboratory (NREL), Golden, Colorado.

For the multijunction cells, the conversion efficiency is relatively higher compared with others, however, the price is much more expensive as well due to more amount of materials used for production. Compared with multijunction cells, the crystalline silicon has lower conversion efficiency, but lower price as well. The price for thin film is cheaper than crystalline silicon, however, the efficiency is little lower than crystalline silicon. The new emerging cells are cheapest compared with all the others type of cells, however, the conversion efficiency is relatively lower.

1.1.1 Single-junction solar cells

The p-n junction is the fundamental building block of solar cells. The principle behind the p-n junction needs to be demonstrated in order to better understand the principle of solar cells. The single p-n homojunction will be explored in this section.

We will start from the separate n-type material and p-type material. In the **Fig. 1.4**, the left one is the n-type, where the material has many free negatively charged electrons which can move around the material, and there are numbers of positively charged immobile donor ions as well. Similarly, the right one is the p-type, and it has many free

positively charged holes which can move through 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.

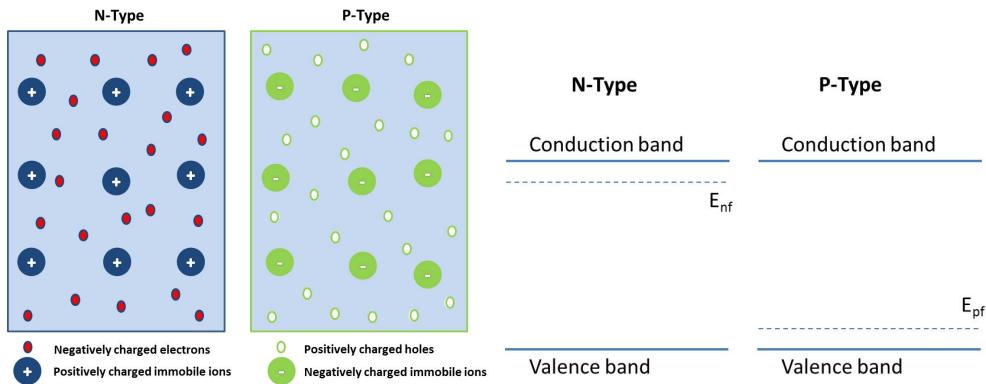


Figure 1.4. Left: Doped material (n-type and p-type). Right: Fermi level of n-type and p-type materials

If the n-type and p-type materials are joined, the free electrons (holes) in the n-type (p-type) material will diffuse into p-type (n-type) material due to the concentration difference (Fig. 1.5). In the region which is near the interface between n-type and p-type materials, the fixed donor and acceptor ions establish a "build in" electric field which points from n-type material to p-type material. 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. 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 establishes the "build in" electric field is also called space charge region (SCR).

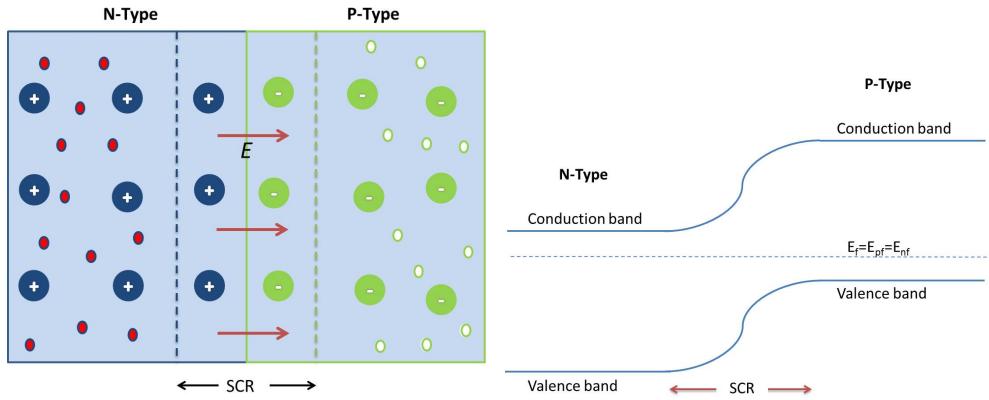


Figure 1.5. Left: The p-n homojunction. Right: Fermi level of the p-n homejunction at the equilibrium.

Before discussing further the p-n junction in the solar cells, the forward and reverse bias will be introduced in the **Fig. 1.6**. The forward bias means that a negative voltage is applied on the n-type side, and a positive voltage is applied on the p-type side. In this case, the SCR will be shrunk. With the increasing of the voltage, the SCR will rather small and the current will flow freely. Conversely, the reverse bias means that a negative voltage is applied on the p-type side, and a positive voltage is applied on the n-type side. The SCR will become thicker, if the voltage is large enough, then the whole junction will breakdown and current will flow.

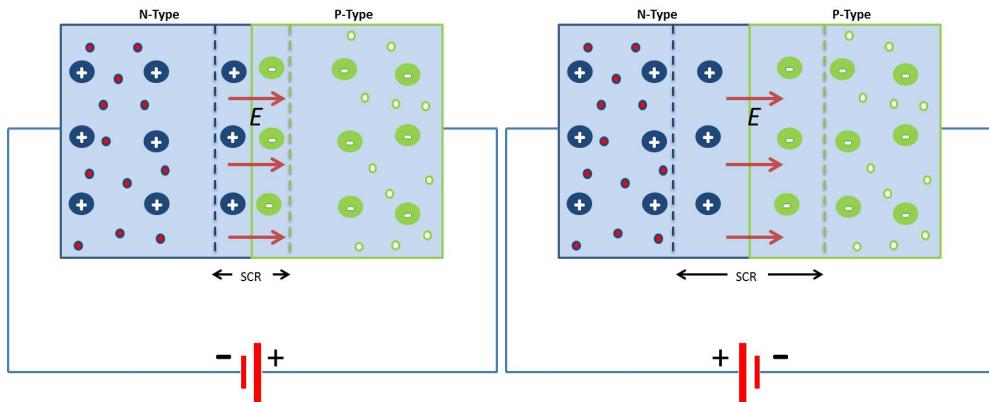


Figure 1.6. Forward and reverse bias

The p-n junction material with and without the illumination is discussed in the **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 no illumination. However, if the light shines

on the whole material, the current will be generated from the p-type to the n-type side (conventional current). Because the bond inside the material will be broken under the illumination, which will generate one pair of electron hole. Apparently, there are three regions in the whole junction material where the bonds are possible to be broken, the n-type region, the p-n junction, and the p-type region. In the either n-type or p-type region, the electron–hole pair will not remain long time, it is most probable that they will back into the bonded positions again since the pair electron-hole is rather close each other. However, the pair electron-hole pair 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 pair in the either n-type and p-type also have the chance to diffuse into the SCR, it will also contribute the generation of current in this case.

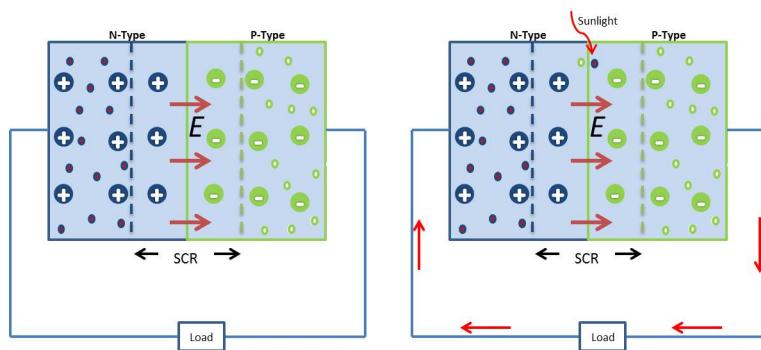


Figure 1.7. The p-n junction under illumination

The band-profile of ...

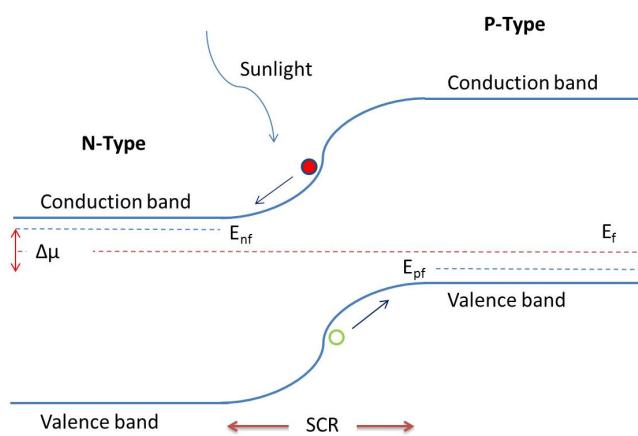


Figure 1.8. Band profile and fermi level of the p-n junction under illumination

The I-V characteristics is defined in the Fig. 1.9 with some important parameters in 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 current when the solar cells are with the illumination conditions. And 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 $P_{out} = V_{mp}I_{mp}$, that is the rectangle bounded by the dashed lines in the Fig. 1.9.

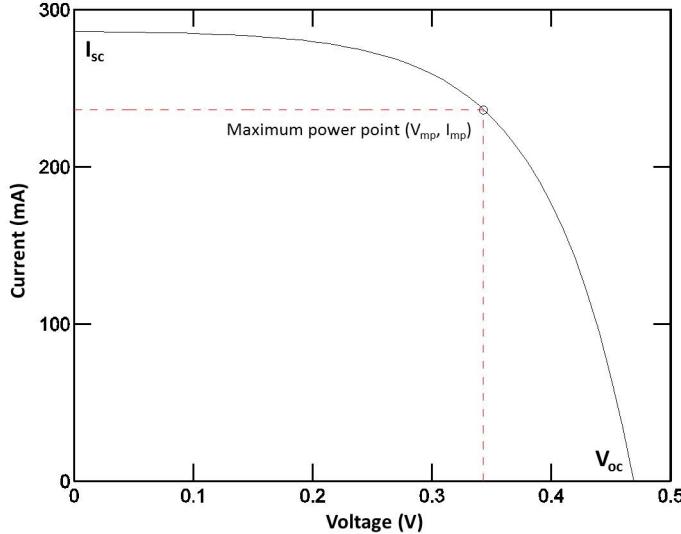


Figure 1.9. I-V characteristics

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

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

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

where P_{in} is the incident photon energy per second. The conversion efficiency of the solar cells is proportional to fill factor, 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. On the material surface, the photons are absorbed and recombined, this is severely reduce the efficiency. There is more detailed analysis of conversion efficiency in the Ref[?? Principles of solar energy conversion].

1.2 Solar cell materials

In 1839, French physicist A. E. Becquerel revealed the photovoltaic effect for the first time, later 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 is demonstrated. Until now, there are many different solar cell materials. The reason why best solar cell material do not come out yet is that it is expected to be not only high efficiency but also environmentally friendly and low cost. It means that it requires not only that the process of growth and manufacturing solar cell materials should be cheaper, but also that it has longer application life, and the raw material should be abundant and non-toxic as well. In this section, five main solar cell materials are discussed briefly, that is, silicon (Si), gallium arsenide (GaAs), cadmium telluride (CdTe), Copper indium gallium selenide ($\text{Cu}(\text{In}, \text{Ga})\text{Se}_2$).

Potential solar cell materials need to fullfil several properties, such as large absorption coefficient and the range of band gap is around 1 to 1.6 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, enviromental safety and many others. Thereby, only some of them is suitable to produce in reality. In the **Fig. 1.10**, evolution for several solar cell materials is illustrated.

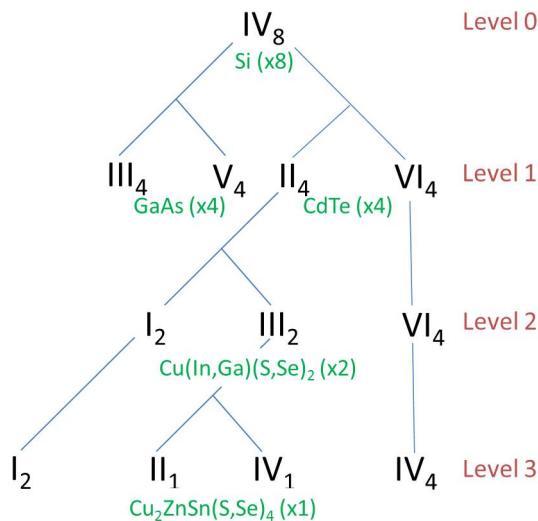


Figure 1.10. Tree of tetragonal bonded semiconductor, the roman numerals means the group numbers in the chemistry element periodic table, the subscript means the number of elements.

From above **Fig. 1.10**, based on group number IV element silicon (level 0) in the chemistry element periodic table, two categories semiconductor materials are deduced (level 1), such as gallium arsenide (GaAs) in group III-V and cadmium telluride (CdTe) in group II-VI. The element in group II in level 1 could be further divided into elements

in group I and III, such as CIGS and CIGSe (level 2). From the group III in level 2, the element is possible to be substituted from group II and IV in level 3, such as CZTS and CZTSe.

1.2.1 Crystalline silicon solar cells

The solar cell based on silicon dominates the solar power world currently, 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. In the crust of Earth, there are over 90% composed of silicate minerals, which also means 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 a thicker Si 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) is a III-V compound of the elements gallium and arsenic. It has a zinc blende crystal structure with a direct bandgap around 1.42 eV. Some electronic properties of GaAs are superior to Silicon, such as higher electron mobility, higher saturated electron velocity, absorb sunlight more efficiently due to the direct bandgap. The optimum bandgap for the single junction solar cell is suggested around 1.4 eV by theoretical calculations, 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 has higher price compared to 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 is considered as well.

Recently, the multi-junction solar cells are based on the III-V compounds, and the conversion efficiency record is reached around 37.7% for multi-junction device based on InGaP-GaAs-InGaAs lead by the group Sharp.

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 used 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. There are mainly four thin film materials considered in the solar cells: amorphous silicon (a-Si), cadmium telluride (CdTe), Cu(In, Ga)Se₂ (CIGS), and Cu₂ZnSn(S, Se)₄ (CZTSSe).

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 process of manufacturing is easy to control, which means 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 semiconductor with high optical absorption coefficients. It is seen as the most promising solar cell material for the near future. It is always used in a heterojunction structure, mainly it is with the thinner n-type cadmium sulfide layer. The conversion efficiency of CIGS reached up to 20% in the laboratory cell. The interesting part is that it can be alloyed with Gallium (Ga), and the band gap can be tuned along with that. The band gap is between 1 eV to 1.7 eV for this alloy. CIGS does not contain any toxic element. However, the cost and scarcity of Indium (In) will become the major problem for the CIGS.

1.3 CIGS materials

CIGS material is a chalcopyrite-type material, which is considered to be the most promising thin film solar cell material. It has a direct band gap between 1 eV and 1.7 eV, and the conversion efficiency in laboratory already surpassed 20%. CuInSe₂ was first synthesized by Hahn in 1953. It was first used 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 Ga.

1.3.1 Crystal structure

From the Fig. 1.10, the crystal structure of CIGS can be derived from the zinc blende crystal structure of ZnSe. In the Fig. 1.11, the crystal structure of ZnSe and CIGS is 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 different bond strength and lengths between Cu-Se and In-Se or Ga-Se, therefore, the lattice parameter c is not exact 2a normally.

Chalcopyrite CIGS (with $x = 0.0$ and 1.0) has the space group XYZ (XYZ; space group no. 23). The conventional unit cell has four copper atoms on the Wyckoff positions XYZ , four indium/gallium atoms on position XYZ , and eight selenium atoms on the XYZ position. The cation positions have all $\text{XYZ}xyz$ point-group symmetry, and Se have $\text{XYZ}xyz$ symmetry. The Se $\text{XYZ}xyz$ position is fully defined with the position (x, y, z) , and each anion Se-atom has two inequivalent bonds $\delta\text{X}-\text{Se}$ to the cations $\text{X} = \text{Cu}$ 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; this is the most probable local atomic configuration in a random alloy structure.

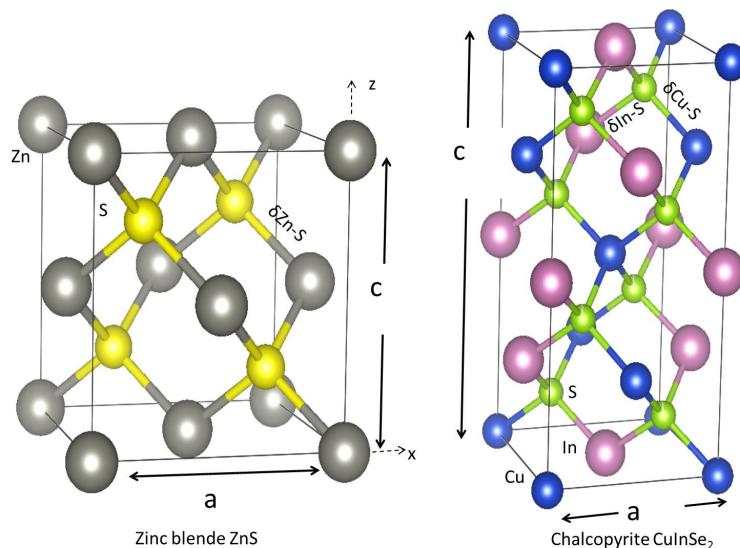


Figure 1.11

1.3.2 Optical properties and doping in the CIGS

CuInSe_2 has a direct band gap around 1 eV, and the absorption coefficient is above $1 \times 10^4 \text{ cm}^{-1}$. 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. 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. The reason will be discussed in the next paragraph. At last, the overview properties of CIGS material is described in the table ???

CIGS is a nonstoichiometric compound with the deviations from stoichiometry in several percentage range. The high quality thin film solar cells mainly use Cu-poor (Cu: 22.5-24.5%) high offstoichiometric CIGS absorber. The main native defects in CIGS include $\langle 2\text{V}_{\text{Cu}}, \text{In}_{\text{Cu}} \rangle$, $\langle \text{Cu}_{\text{In}}, \text{In}_{\text{Cu}} \rangle$, V_{Cu} , In_{Cu} , Cu_{In} , V_{Se} and Ga_{Cu} . V_{Cu} and III_{Cu} are 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}).

There are some extrinsic divalent cation donors as well, such as Zn_{Cu} , Cd_{Cu} and Cl_{Se} . And the formation energy is relatively low for CIS and CGS. Therefore, it is possible to grow n-type for them.

CIS can be grown n-type under the condition Se-poor or In-rich. 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 that the Ga-rich CIGS is not suitable for high efficiency solar cells.

1.3.3 CIGS solar cell structure

The solar cells device based on CIGS is a heterojunction device, which has several thin film layers with different functional properties. A schematic of a conventional device structure is shown in Fig. 1.12

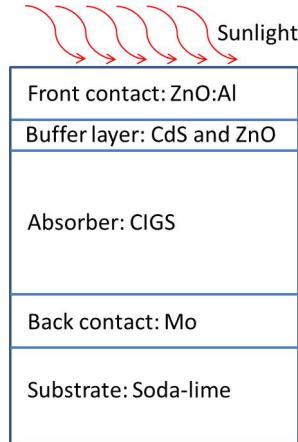


Figure 1.12

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 with thickness 1 mm to 3 mm. 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, and it is dopped by intrinsic defects. The n-type buffer layer cadmium sulfide (CdS) is on the top of CIGS. The intrinsic zinc oxide (i-ZnO) and n-type ZnO layer are followed, and it works as window layer. The n-type ZnO is dopped by the aluminum (Al). This CIGS/CdS/ZnO strucutre was found 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 the atom is constructed by nuclei and electrons. 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. Let us start from the time-independent many-body Schrödinger equation,

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

where $\Psi(\{\mathbf{r}_i, \mathbf{R}_I\})$ is the exact wavefunction for the above Schrödinger equation, \mathbf{r}_i and \mathbf{R}_I stands for electron and nucleus coordinators, respectively.

In the Eq. 2.1, E is the total energy of the system, and \hat{H} is the Hamiltonian which has the following form:

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

where the indices i, j are used for electron and I, J are used for atomic nuclei, Z_I means the charge of the I -th nucleus, M denotes the nuclear mass, m_e is the electron mass, ε_0 is vacuum permittivity.

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

$\hbar/(mc\alpha)$, where α is the fine structure constant ($\alpha = e^2/(4\pi\varepsilon_0 c\hbar)$), so the velocity of light in atomic units is $c = 1/\alpha$ and $e^2/(4\pi\varepsilon_0) = 1$. The Schrödinger Eq. (2.2) in atomic units has the following form:

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

In Eq. 2.3, the first and second terms are the kinetic energy operator of the electron and nuclei, respectively. The other terms in order are Coulomb interactions between electrons and nuclei, electrons and electrons and nuclei and nuclei.

Since there are so many atoms to calculate in reality, more importantly, the exactly form of the wavefunction is unknown, so one can not solve the Eq. 2.1 exactly at present. To approximate the exact solution, generally one can divide it into three different levels, the first level is the Born-Oppenheimer approximation and the second level is Hartree, Hartree-Fock (HF), density functional theory (DFT) and Kohn-Sham (KS) equation, and 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

In order to simplify the Eq. 2.1, the first attempt is to separate the wavefunction of electrons and nuclei, but since there is a couple term between the electron and nucleus in the Schrödinger Hamiltonian in Eq. (2.3), thereby one can not do that simply. On the other hand, let us observe the Eq. 2.3 again, there exists a small value $1/M_I$, 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, so the nuclei are treated as fixed. The result is that the electrons is seen as interacting under both the external potential caused by nuclei that are fixed in some positions and that of the other electrons. A more vividly description is seen from Fig. 2.1:

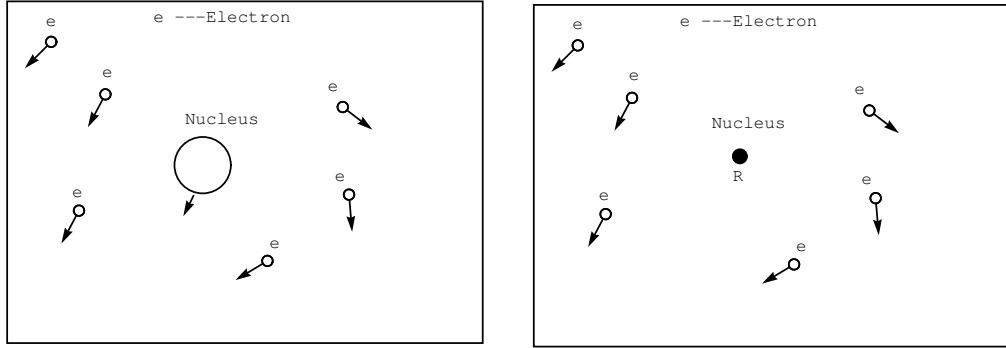


Figure 2.1. Left panel: normal interacting system. Right panel: the Born-Oppenheimer approximation. The arrows denotes the movement of the nucleus or electron.

The separation of motion between electrons and nuclei is called the Born-Oppenheimer approximation, since the position of nuclei is fixed, one can define:

$$\Psi(\{\mathbf{r}_i, \mathbf{R}_I\}) \approx \Psi_{BO}(\{\mathbf{r}_i, \mathbf{R}_I\}) = \theta(\{\mathbf{R}_I\}) \Psi'(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}), \quad (2.4)$$

where $\Psi'(\{\mathbf{r}_i\}; \{\mathbf{R}_I\})$ is the electrons wavefunction in the Born-Oppenheimer approximation, the semicolon means that \mathbf{R}_I is only discrete value belonging to the set of atomic positions.

The Eq. 2.3 can be defined as:

$$\hat{H} = - \sum_I \frac{\nabla_I^2}{2M_I} + \hat{H}_{BO}. \quad (2.5)$$

where

$$\begin{aligned} \hat{H}_{BO} &= - \sum_i \frac{\nabla_i^2}{2} - \sum_i V_{ext}(\mathbf{r}_i) + \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \\ &= \hat{T} + \hat{V}_{ext} + \hat{V}_{int} + N_H, \end{aligned} \quad (2.6)$$

and

$$V_{ext}(\mathbf{r}_i) = \sum_I \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|}, \quad (2.7)$$

where \hat{H}_{BO} is the Hamitonian for the electrons within the Born-Oppenheimer approximation. The subscript *ext* means *external* in Eq. 2.7, this term describes the external potentials interaction.

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

$$\left(-\sum_I \frac{\nabla_I^2}{2M_I} + \hat{H}_{BO}\right)\theta(\{\mathbf{R}_I\})\Psi'(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}) = E_m(\{\mathbf{R}_I\})\theta(\{\mathbf{R}_I\})\Psi'(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}). \quad (2.8)$$

Here, $E_m(\{\mathbf{R}_I\})$ is the total energy, after some steps of derivation, one can end up with the Eq. 2.9 :

$$\hat{H}_{BO}\Psi'(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}) = E_{BO}(\{\mathbf{R}_I\})\Psi'(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}), \quad (2.9)$$

where $E_{BO}(\{\mathbf{R}_I\})$ is the energy of this electronic system,

$$(\hat{H}_{I1} + \hat{H}_{I2} + \hat{H}_{I3} + E_{BO}(\{\mathbf{R}_I\}))\theta(\{\mathbf{R}_I\}) = E_m(\{\mathbf{R}_I\})\theta(\{\mathbf{R}_I\}), \quad (2.10)$$

where

$$\begin{aligned} \hat{H}_{I1} &= -\sum_I \frac{\nabla_I^2}{2M_I} \\ \hat{H}_{I2} &= -\sum_I \frac{1}{M_I} \int \Psi'(\{\mathbf{r}_i\}; \{\mathbf{R}_I\})^* \nabla_I \Psi'(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}) d\mathbf{r} \nabla_I \\ \hat{H}_{I3} &= -\sum_I \frac{1}{M_I} \int \Psi'(\{\mathbf{r}_i\}; \{\mathbf{R}_I\})^* \nabla_I^2 \Psi'(\{\mathbf{r}_i\}; \{\mathbf{R}_I\}) d\mathbf{r}. \end{aligned} \quad (2.11)$$

From Eq. 2.10, one observes that the lattice dynamical properties of certain system within the Born-Oppenheimer approximation could be obtained. To solve this equation, the ground state energy $E_{BO}(\{\mathbf{R}_I\})$ of electron system is needed. Here $\{\mathbf{R}_I\}$ are the parametrized values from the atom position.

In summary, the Schrödinger equation of electron and nucleus is derived separately within the Born-Oppenheimer approximation. When one refers to calculations of the ground state properties, which means to take use of the Schrödinger equation of electron only, e.g., Eq. 2.9, and Schrödinger equation of nucleus is used for the calculation of lattice dynamics.

One notices that the Eq. 2.6 is much simpler than Eq. 2.3, but still the equation is not solvable. Further approximations are needed to solve this many-body problem.

2.1.3 Hartree, Hartree-Fock approximation and density functional theory

From last 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 $10^{23}/cm^3$, it is a huge numerical problem, however, it is still possible to solve. The second one is that the Hamiltonian operator is applied to single electron, however, the wavefunction is not shown how it depends on the single-electron wavefunction. The latter problem could be solved by the following three methods:

The first method is to figure out a way to separate or approximate the wavefunction into single-electron function, like the Hartree and Hartree-Fock (HF) method. The second method is to find a Hamiltonian which is possible to act on wavefunction directly, like density functional theory (DFT). 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, but takes advantage of single-electron wavefunction. This will be demonstrated in the next section.

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 has the following expression:

$$\Psi_H(\{\mathbf{r}_i\}) = \phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2) \cdots \phi_N(\mathbf{r}_N), \quad (2.12)$$

where i goes through all the electrons, and means state of the i -th electron in the position of \mathbf{r}_i , from here and following the \mathbf{R} is suppressed in the wavefunction since they are in fixed position. The total energy of the system can be written down in the following way :

$$E_H = \langle \Psi_H(\{\mathbf{r}_i\}) | \hat{H}_{BO} | \Psi_H(\{\mathbf{r}_i\}) \rangle. \quad (2.13)$$

Therefore, making the substitution using Eq. 2.6 and 2.12 into Eq. 2.13, the total energy of system is:

$$\begin{aligned} E_H &= \sum_i \langle \phi_i(\mathbf{r}) | -\frac{\nabla_i^2}{2} + V_{ext}(\mathbf{r}) | \phi_i(\mathbf{r}) \rangle \\ &\quad + \frac{1}{2} \sum_{i \neq j} \langle \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') | \frac{1}{|\mathbf{r} - \mathbf{r}'|} | \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \rangle. \end{aligned} \quad (2.14)$$

In order to calculate the stationary state with the lowest energy of the system, the variation of the wavefunction should be zero variation in the energy, one can set up the following equation with Lagrange multiplier E_H^i ,

$$\delta [E_H - \sum_i E_H^i (\langle \phi_i(\mathbf{r}) | \phi_i(\mathbf{r}) \rangle - 1)] = 0. \quad (2.15)$$

where ($\langle \phi_i(\mathbf{r}) | \phi_i(\mathbf{r}) \rangle = 1$) means that the wavefunction is normalized.

In order to calculate the term $\delta(E_H)$, one has to derive

$$\begin{aligned} & \delta \left(\sum_i \langle \phi_i(\mathbf{r}) | -\frac{\nabla_i^2}{2} + V_{ext}(\mathbf{r}) | \phi_i(\mathbf{r}) \rangle \right) \\ &= \sum_i \left\{ \langle \delta \phi_i(\mathbf{r}) | -\frac{\nabla_i^2}{2} + V_{ext}(\mathbf{r}) | \phi_i(\mathbf{r}) \rangle \right. \\ & \quad \left. + \langle \delta \phi_i(\mathbf{r}) | -\frac{\nabla_i^2}{2} + V_{ext}(\mathbf{r}) | \phi_i(\mathbf{r}) \rangle^* \right\}. \end{aligned} \quad (2.16)$$

and

$$\begin{aligned} & \delta \left(\frac{1}{2} \sum_{i \neq j} \langle \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') | \frac{1}{|\mathbf{r} - \mathbf{r}'|} | \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \rangle \right) \\ &= \frac{1}{2} \sum_{i \neq j} \left\{ \langle \delta \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') | \frac{1}{|\mathbf{r} - \mathbf{r}'|} | \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \rangle \right. \\ & \quad \left. + \langle \phi_i(\mathbf{r}) \delta \phi_j(\mathbf{r}') | \frac{1}{|\mathbf{r} - \mathbf{r}'|} | \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \rangle + T_{rem} \right\}, \end{aligned} \quad (2.17)$$

where T_{rem} is the remaining terms, actually these remaining terms will not affect the derivation.

Before getting the final result, there is one more identity

$$\begin{aligned} & \langle \phi_i(\mathbf{r}) \delta \phi_j(\mathbf{r}') | \frac{1}{|\mathbf{r} - \mathbf{r}'|} | \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \rangle \\ &= \int d\mathbf{r} d\mathbf{r}' \phi_i^*(\mathbf{r}) \delta \phi_j^*(\mathbf{r}') \frac{1}{|\mathbf{r} - \mathbf{r}'|} \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \\ &= \int d\mathbf{r}' d\mathbf{r} \phi_i^*(\mathbf{r}') \delta \phi_j^*(\mathbf{r}) \frac{1}{|\mathbf{r}' - \mathbf{r}|} \phi_i(\mathbf{r}') \phi_j(\mathbf{r}) \\ &= \langle \delta \phi_j(\mathbf{r}) \phi_i(\mathbf{r}') | \frac{1}{|\mathbf{r} - \mathbf{r}'|} | \phi_j(\mathbf{r}) \phi_i(\mathbf{r}') \rangle. \end{aligned} \quad (2.18)$$

So the Eq. 2.17 will become

$$\begin{aligned} & \delta \left(\frac{1}{2} \sum_{i \neq j} \langle \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') | \frac{1}{|\mathbf{r} - \mathbf{r}'|} | \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \rangle \right) \\ &= \sum_{i \neq j} \left\{ \langle \delta \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') | \frac{1}{|\mathbf{r} - \mathbf{r}'|} | \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \rangle + \frac{T_{rem}}{2} \right\}. \end{aligned} \quad (2.19)$$

Taking use of the above two Eq. 2.16, 2.19 and variation in the term of $\delta\phi_i^*(\mathbf{r})$, the Eq. 2.15 becomes:

$$\begin{aligned}
 & \frac{1}{\delta\phi_i^*(\mathbf{r})} \delta [E_H - \sum_i E_H^i (\langle \phi_i(\mathbf{r}) | \phi_i(\mathbf{r}) \rangle - 1)] \\
 &= \frac{1}{\delta\phi_i^*(\mathbf{r})} \left\{ \sum_i \left\{ \langle \delta\phi_i(\mathbf{r}) | -\frac{\nabla_i^2}{2} + V_{ext}(\mathbf{r}) | \phi_i(\mathbf{r}) \rangle + \langle \delta\phi_i(\mathbf{r}) | -\frac{\nabla_i^2}{2} + V_{ext}(\mathbf{r}) | \phi_i(\mathbf{r}) \rangle^* \right\} \right. \\
 &\quad \left. + \sum_{i \neq j} \left\{ \langle \delta\phi_i(\mathbf{r}) \phi_j(\mathbf{r}') | \frac{1}{|\mathbf{r} - \mathbf{r}'|} | \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \rangle + \frac{T_{rem}}{2} \right\} - \sum_i E_H^i (\langle \phi_i(\mathbf{r}) | \phi_i(\mathbf{r}) \rangle - 1) \right\} \\
 &= \left\{ -\frac{\nabla_i^2}{2} + V_{ext}(\mathbf{r}) \right\} \phi_i(\mathbf{r}) + \sum_{j \neq i} \left\{ \langle \phi_j(\mathbf{r}') | \frac{1}{|\mathbf{r} - \mathbf{r}'|} | \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \rangle - E_H^i \phi_i(\mathbf{r}) \right\}.
 \end{aligned} \tag{2.20}$$

Therefore,

$$\left(-\frac{\nabla_i^2}{2} + V_{ext}(\mathbf{r}) + \sum_{j \neq i} \langle \phi_j(\mathbf{r}') | \frac{1}{|\mathbf{r} - \mathbf{r}'|} | \phi_j(\mathbf{r}') \rangle \right) \phi_i(\mathbf{r}) = E_H^i \phi_i(\mathbf{r}), \tag{2.21}$$

where E_H^i also can be treated as the energy. The Eq. 2.20 is a group of dependent single particle equations, and after checking it, one will find out this equation is self-consistent and can be solved using iteratively.

2.1.3.2 Hartree-Fock approximation

Hartree approximation is the simplest approximation, and Hartree-Fock approximation is the method which considers the antisymmetry of the wavefunction. It means that if the positions of two electrons (with same spin) are exchanged, the wavefunction should change the sign, like:

$$\Psi_{HF}(\dots \mathbf{r}_i \dots \mathbf{r}_j) = -\Psi_{HF}(\dots \mathbf{r}_j \dots \mathbf{r}_i). \tag{2.22}$$

Slater introduced an way to construct the wavefunction due to the Eq. 2.22 based on the Hartree approximation, the wavefunction of the many-electron Schrödinger equation is described in a matrix determinant for the N number of electrons (without spin):

$$\Psi_{HF}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{r}_1) & \phi_1(\mathbf{r}_2) & \cdots & \phi_1(\mathbf{r}_N) \\ \phi_2(\mathbf{r}_1) & \phi_1(\mathbf{r}_2) & \cdots & \phi_1(\mathbf{r}_N) \\ \vdots & \vdots & & \vdots \\ \phi_N(\mathbf{r}_1) & \phi_N(\mathbf{r}_2) & \cdots & \phi_N(\mathbf{r}_N) \end{vmatrix}, \tag{2.23}$$

where i goes through all the electron, and $\phi_i(\mathbf{r}_i)$ means state of the (i)-th electron in the position of \mathbf{r}_i , so if one exchanges two columns in the Eq. 2.23, the result is satisfied with the Eq. 2.22.

Repeating all the processes already done through the Hartree approximation, the total energy of Hartree-Fock is:

$$\begin{aligned} E_{HF} = & \sum_i \left\langle \phi_i(\mathbf{r}) \right| -\frac{\nabla_i^2}{2} + V_{ext}(\mathbf{r}) \left| \phi_i(\mathbf{r}) \right\rangle \\ & + \frac{1}{2} \sum_{i \neq j} \left\langle \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \right| \frac{1}{|\mathbf{r} - \mathbf{r}'|} \left| \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \right\rangle \\ & - \frac{1}{2} \sum_{i \neq j} \left\langle \phi_i(\mathbf{r}') \phi_j(\mathbf{r}) \right| \frac{1}{|\mathbf{r} - \mathbf{r}'|} \left| \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \right\rangle. \end{aligned} \quad (2.24)$$

In the same mathematical way as in previous section but somewhat more complicated, the single particle Hartree-Fock equation can be obtained:

$$\begin{aligned} \left\{ -\frac{\nabla_i^2}{2} + V_{ext}(\mathbf{r}) + \sum_{j \neq i} \left\langle \phi_j(\mathbf{r}') \right| \frac{1}{|\mathbf{r} - \mathbf{r}'|} \left| \phi_j(\mathbf{r}') \right\rangle \right\} \phi_i(\mathbf{r}) \\ - \sum_{i \neq j} \left\langle \phi_j(\mathbf{r}') \right| \frac{1}{|\mathbf{r} - \mathbf{r}'|} \left| \phi_i(\mathbf{r}') \right\rangle \phi_j(\mathbf{r}) = E_{HF}^i \phi_i(\mathbf{r}). \end{aligned} \quad (2.25)$$

Compared with Hartree equation, there is an extra term in the equation above. This is called exchange term. In order to organize the equation in a nice and clear way, finally it ends up with:

$$\left\{ -\frac{\nabla_i^2}{2} + V_{ext}(\mathbf{r}) + V_H(\mathbf{r}) \right\} \phi_i(\mathbf{r}) = E_{HF}^i \phi_i(\mathbf{r}), \quad (2.26)$$

where

$$V_H(\mathbf{r}) = \int \frac{\rho(\mathbf{r}') - \rho_i^{HF}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}'. \quad (2.27)$$

$$\begin{aligned} \rho_i^{HF}(\mathbf{r}, \mathbf{r}') &= \sum_j \frac{\phi_i(\mathbf{r}') \phi_i^*(\mathbf{r}) \phi_j(\mathbf{r}) \phi_j^*(\mathbf{r}')}{\phi_i(\mathbf{r}) \phi_i^*(\mathbf{r})} \\ \rho(\mathbf{r}) &= \sum_i |\phi_i(\mathbf{r})|^2. \end{aligned} \quad (2.28)$$

The Eq. 2.26 could be solved in the same way like Hartree approximation but plus one extra term.

2.1.3.3 Density functional theory

The Hartree and Hartree-Fock methods are very classic methods to solve the many-electron Schrödinger equation. However, the HF method only includes the exchange term, not the electron correlation term, they are not suitable in the case of electrons in solid. Apart from the two methods mentioned before, there is a modern method to deal with the more complicated calculation of electrons, namely, the density functional theory, which is introduced by Hohenberg and Kohn in 1964, Kohn and Pople was awarded Chemistry Nobel Prize in 1998.

The idea of the DFT is to treat the electron density of solid instead of using the many-particle wavefunction, so one can benefit that the degree of freedom reduces from $3N$ (N is the number of electrons) to 3, which is apparently less complicate than those of Hartree and Hartree-Fock during calculation.

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 of the system, and the second one is how to solve this problem. In order to know that 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 also indicate that all the ground state properties are determined by the true ground state density ρ , for example, 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 one can get the wavefunction, 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 following:

Let us assume that there exists two external potentials named $V_{ext}^1(\mathbf{r})$ and $V_{ext}^2(\mathbf{r})$ leading to the same ground state electron density ρ . Obviously, this will lead to two different Hamiltonians, that is, \hat{H}_1 and \hat{H}_2 , and as well as two different corresponding wavefunctions named Ψ_1 and Ψ_2 . Since Ψ_1 are not the ground state wavefunction of \hat{H}_2 , the same rules to Ψ_2 and \hat{H}_1 , two following inequality equations will be satisfied:

$$\begin{aligned} E_1 &= \langle \Psi_1 | \hat{H}_1 | \Psi_1 \rangle < \langle \Psi_2 | \hat{H}_1 | \Psi_2 \rangle. \\ E_2 &= \langle \Psi_2 | \hat{H}_2 | \Psi_2 \rangle < \langle \Psi_1 | \hat{H}_2 | \Psi_1 \rangle. \end{aligned} \tag{2.29}$$

Taking advantage of the form of Hamiltonian from Eq. 2.6, one can get:

$$\begin{aligned}
 & \langle \Psi_2 | \hat{H}_1 | \Psi_2 \rangle \\
 &= \langle \Psi_2 | \hat{H}_2 + \hat{H}_1 - \hat{H}_2 | \Psi_2 \rangle \\
 &= \langle \Psi_2 | \hat{H}_2 | \Psi_2 \rangle + \langle \Psi_2 | \hat{H}_1 - \hat{H}_2 | \Psi_2 \rangle \\
 &= E_2 + \int d\mathbf{r} (V_{ext}^1(\mathbf{r}) - V_{ext}^2(\mathbf{r})) \rho.
 \end{aligned} \tag{2.30}$$

Using Eq. 2.29 and Eq. 2.30, one can get:

$$E_1 < E_2 + \int d\mathbf{r} (V_{ext}^1(\mathbf{r}) - V_{ext}^2(\mathbf{r})) \rho. \tag{2.31}$$

Another similar inequality equation will be gained if one changes the form of equation $\langle \Psi_1 | \hat{H}_2 | \Psi_1 \rangle$ like Eq. 2.30.

$$E_2 < E_1 + \int d\mathbf{r} (V_{ext}^2(\mathbf{r}) - V_{ext}^1(\mathbf{r})) \rho. \tag{2.32}$$

Plus the left and right sides from Eq. 2.31 and 2.33, one will gain a contradictory result:

$$E_1 + E_2 < E_2 + E_1. \tag{2.33}$$

Thereby, The external potential $V_{ext}(\mathbf{r})$ is 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, $E[\rho'] > E[\rho]$, where ρ is the exact ground state density.*

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

$$\begin{aligned}
 E[\rho] &= \langle \Psi | \hat{T} + \hat{V}_{int} + \hat{V}_{ext} | \Psi \rangle \\
 &= \langle \Psi | \hat{V}_{ext} | \Psi \rangle + \underbrace{\langle \Psi | \hat{T} + \hat{V}_{int} | \Psi \rangle}_{\int \rho(\mathbf{r}) V_{ext}} \\
 &= \int \rho(\mathbf{r}) V_{ext} + F[\rho].
 \end{aligned} \tag{2.34}$$

In the above Eq. 2.34, the term of $F[\rho]$ is the universal functional for all the system of electrons.

The functional of total energy $E[\rho']$ reach the minimum at the exact ground state electron density ρ :

$$E[\rho'] = \int \rho'(\mathbf{r}) V_{ext} + F[\rho'] > E[\rho]. \tag{2.35}$$

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

From those two theorems, one definitely know how to solve this problem theoretically, but in practice, $E[\rho]$ is unknown. Therefore, one more method is needed to deal with it, this 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 wavefunction. The DFT only consider to take use of information from Hamiltonian, however not solvable. Therefore, is it possible to combine these two ideas together? the answer is yes, 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 $\Psi_{KS}(\{\mathbf{r}_i\}) = \phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2) \cdots \phi_M(\mathbf{r}_M)$, where M is a number which is less than 10^{23} a lot and $\phi_i(\mathbf{r}_i)$ is auxiliary independent single-electron wavefunction:

$$\rho(\mathbf{r}) = \sum_i \phi_i^*(\mathbf{r})\phi_i(\mathbf{r}). \quad (2.36)$$

If the density is exact, thereby, the total energy is defined exactly, which can be expressed by the following:

$$\begin{aligned} E[\rho] &= T[\rho] + V_{int}[\rho] + V_{ext}[\rho] \\ &= \underbrace{(T[\rho] - T_0[\rho])}_{(T[\rho] - T_0[\rho])} + \underbrace{(V_{int}[\rho] - V_{HF}[\rho])}_{(V_{int}[\rho] - V_{HF}[\rho])} + T_0[\rho] + V_{HF}[\rho] + V_{ext}[\rho] \\ &= T_0[\rho] + V_{HF}[\rho] + V_C[\rho] + V_{ext}[\rho], \end{aligned} \quad (2.37)$$

where $E[\rho]$ is the total energy, 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:

$$V_{HF}[\rho] = V_H[\rho] + V_X[\rho], \quad (2.38)$$

where $V_H[\rho]$ and $V_X[\rho]$ are the Hartree contribution and exchange contribution, respectively. Thereby, the Eq. 2.37 can be further defined like the following:

$$\begin{aligned} E[\rho] &= T_0[\rho] + V_H[\rho] + V_{ext}[\rho] + \underbrace{V_X[\rho] + V_C[\rho]} \\ &= T_0[\rho] + V_H[\rho] + V_{ext}[\rho] + V_{XC}[\rho], \end{aligned} \quad (2.39)$$

where $V_{XC}[\rho]$ is the exchange-correlation term. From Eq. 2.37, the explicit expression of $V_{XC}[\rho]$ one does not know. The first three terms one already know

$$T_0[\rho] = \langle \Psi_i(\mathbf{r}) | -\frac{\nabla_r^2}{2} | \Psi_i(\mathbf{r}) \rangle \quad (2.40)$$

$$V_H[\rho] = \frac{1}{2} \int \int d\mathbf{r} d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (2.41)$$

$$V_{ext}[\rho] = \int d\mathbf{r} \rho(\mathbf{r}) V_{ext}(\mathbf{r}). \quad (2.42)$$

In order to derive the ground state of the above system, one can view this problem as the process of minimizing the total energy by varying the wavefunction Ψ^* . Since the wavefunction can be constructed to electron density (Eq. 2.36), just like the derivation in the section of Hartree, finally one can get the Kohn-Sham equation:

$$(-\frac{\nabla_i^2}{2} + V_{KS}(\mathbf{r}))\phi_i(\mathbf{r}) = E_{KS}^i \phi_i(\mathbf{r}), \quad (2.43)$$

where

$$\begin{aligned} V_{KS}(\mathbf{r}) &= V_{ext}(\mathbf{r}) + \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta V_{XC}}{\delta \rho(\mathbf{r})} \\ &= V_{ext}(\mathbf{r}) + \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + v_{xc}(\mathbf{r}). \end{aligned} \quad (2.44)$$

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

There is no total energy equation expression yet, however, if one can change the Eq. 2.43 as

$$\sum_i \phi_i^*(\mathbf{r}) \left(-\frac{\nabla_i^2}{2} + V_{KS}(\mathbf{r}) \right) \phi_i(\mathbf{r}) = \sum_i \phi_i^* E_{KS}^i \phi_i(\mathbf{r}). \quad (2.45)$$

Based on Eq. 2.39 and 2.45, the total energy expression is:

$$\begin{aligned} E &= \sum_i E_{KS}^i - \frac{1}{2} \int \int d\mathbf{r} d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \\ &\quad + V_{XC}[\rho] - \int v_{xc}(\mathbf{r}) \rho(\mathbf{r}). \end{aligned} \quad (2.46)$$

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 energy

This part 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, like 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 energy in the very tiny small volume is equal to the homogeneous electrons with the same density in the volume, the explicit equation is:

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

where E_{xc}^{LDA} is the exchange-correlation energy functional for the LDA, $\rho(\mathbf{r})$ is the charge density in the position of \mathbf{r} , 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:

$$\rho(\mathbf{r}) = \rho = \frac{N}{V}, \quad (2.48)$$

where N is number of electrons within the solid, and V is the volume of solid, which also means that the ε_{xc} is the function of $\rho(\mathbf{r})$, 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:

$$E_{GGA}^{xc} = \int \rho(\mathbf{r}) \varepsilon_{xc}((\rho(\mathbf{r})), \nabla \rho(\mathbf{r})) d\mathbf{r}, \quad (2.49)$$

where E_{GGA}^{xc} 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} .

Of course, there are more modern methods to approximate the exchange-correlation energy, such as: the optimized effective potential (OEP) method and the hybrid functionals and others .

2.1.4 Solving the secular equation

The process for solving Kohn-Sham equation can be solved by iteration as well, but the difference between solving Kohn-Sham equation and Hartree or Hartree-Fock equation is that the wavefunction is replaced by the electron density, thereby first an initial electron density is defined by some way, and later on the equation is solved iteratively until the reasonable solution is obtained.

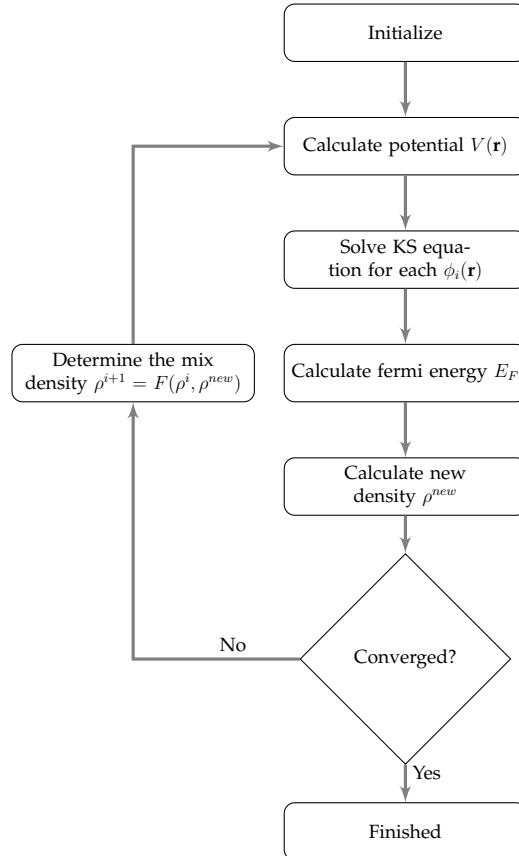


Figure 2.2. Flow chart of the $(i+1)$ th iterations for solving Kohn-Sham equation.

where 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 the eq (2.43), The Kohn-Sham equation will be transformed into the general eigenvalue problem. Now if the Kohn-Sham equation is defined in the following form:

$$\hat{H}\Psi(r) = E\Psi(r). \quad (2.50)$$

The wavefunction is defined as follows:

$$\Psi(r) = \sum_j^N C_j \Phi_j(r), \quad (2.51)$$

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

If the Eq. 2.51 is plugged into the Eq. 2.50, and then left multiply Φ_j in order, finally it will end up with a set of equations:

$$\begin{bmatrix} \Phi_1 \hat{H} \Phi_1 & \Phi_1 \hat{H} \Phi_2 & \cdots & \Phi_1 \hat{H} \Phi_N \\ \Phi_2 \hat{H} \Phi_1 & \Phi_2 \hat{H} \Phi_2 & \cdots & \Phi_2 \hat{H} \Phi_N \\ \vdots & \vdots & & \vdots \\ \Phi_N \hat{H} \Phi_1 & \Phi_N \hat{H} \Phi_2 & \cdots & \Phi_N \hat{H} \Phi_N \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_N \end{bmatrix} \quad (2.52)$$

$$= E \begin{bmatrix} \Phi_1 \Phi_1 & \Phi_1 \Phi_2 & \cdots & \Phi_1 \Phi_N \\ \Phi_2 \Phi_1 & \Phi_2 \Phi_2 & \cdots & \Phi_2 \Phi_N \\ \vdots & \vdots & & \vdots \\ \Phi_N \Phi_1 & \Phi_N \Phi_2 & \cdots & \Phi_N \Phi_N \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_N \end{bmatrix}.$$

If one defines $H_{ij} = \Phi_i \hat{H} \Phi_j$ and $S_{ij} = \Phi_i \Phi_j$, then the Eq. 2.52 becomes:

$$\begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1N} \\ H_{21} & H_{22} & \cdots & H_{2N} \\ \vdots & \vdots & & \vdots \\ H_{N1} & H_{N2} & \cdots & H_{NN} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_N \end{bmatrix} \quad (2.53)$$

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

After some manipulations,

$$\begin{bmatrix} H_{11} - ES_{11} & H_{12} - ES_{12} & \cdots & H_{1N} - ES_{1N} \\ H_{21} - ES_{21} & H_{22} - ES_{22} & \cdots & H_{2N} - ES_{2N} \\ \vdots & \vdots & & \vdots \\ H_{N1} - ES_{N1} & H_{N2} - ES_{N2} & \cdots & H_{NN} - ES_{NN} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_N \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (2.54)$$

Apparently, the Eq. 2.54 is an eigenvalue problem. In order to get the $C_i (i = 1 \dots N)$, one has to set:

$$\begin{vmatrix} H_{11} - ES_{11} & H_{12} - ES_{12} & \cdots & H_{1N} - ES_{1N} \\ H_{21} - ES_{21} & H_{22} - ES_{22} & \cdots & H_{2N} - ES_{2N} \\ \vdots & \vdots & & \vdots \\ H_{N1} - ES_{N1} & H_{N2} - ES_{N2} & \cdots & H_{NN} - ES_{NN} \end{vmatrix} = 0. \quad (2.55)$$

2.2 Full-potential linearised 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 calculation?

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

Slater re-considered the way to describe the wavefunction, 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, called muffin tin (MT) region, the remaining region is called interstitial region (IR) (see Fig. 2.3). An atomic-like function is defined as the wavefunction in the MT region, this is reason why the 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, but far away the atomic core, the electron behaves like free electrons, therefore plane wave is suitable (see Eq. 2.56). However, the drawback of APW method is the wavefunction is dependent with the energy, which leads to the nonlinear eigenvalue problem (see Eq. 2.57), in order to get the exact energy, the method has to decide repeatedly until certain condition is satisfied, 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 that, they noticed that the

taylor expansion of radial function (Eq. 2.60), and then make use of it to linearize the APW method, thereby the method is called linearized augmented plane wave (LAPW) method. However, the drawback is that this method 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. 2.62). 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. 2.63).

2.2.2 Wavefunction

2.2.2.1 Augmented plane wave method

The wavefunction is defined by Slater:

$$\phi_{\mathbf{k}+\mathbf{G}}^{APW}(\mathbf{r}) = \begin{cases} \frac{1}{\sqrt{\Omega}} e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}} & \text{if } \mathbf{r} \in I \\ \sum_{\alpha} \sum_{\ell m} f_{\ell m}(r_{\alpha}, \mathbf{k}+\mathbf{G}, E) Y_{\ell m}(\hat{\mathbf{r}}_{\alpha}) & \text{if } \mathbf{r} \in S_{\alpha}, \end{cases} \quad (2.56)$$

where $f_{\ell m}(r_{\alpha}, \mathbf{k}+\mathbf{G}, E) = A_{\ell m}^{\alpha}(\mathbf{k}+\mathbf{G}) u_{\ell}^{\alpha}(r_{\alpha}, E)$, and $A_{\ell m}^{\alpha}(\mathbf{k}+\mathbf{G})$ is the expansion coefficients, and $u_{\ell}^{\alpha}(r_{\alpha}, E)$ is the radial function, which is dependent with energy E , and the radial function could be decided by the following:

$$-\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{du_{\ell}}{dr} \right) + \left(\frac{\ell(\ell+1)}{r^2} + V(r_{\alpha}) \right) u_{\ell}(r_{\alpha}) = E u_{\ell}(r_{\alpha}), \quad (2.57)$$

where $V(r_{\alpha})$ is the spherical potential.

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

From the Fig. 2.3, one will notice that the unit cell is divided into muffin-tin spheres (α, β) and an interstitial region (IR), and $\mathbf{r} = \mathbf{R}^{\alpha} + \mathbf{r}_{\alpha}$ is guaranteed. One can take use of the Rayleigh expansion formula:

$$e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}} = e^{i(\mathbf{k}+\mathbf{G})\mathbf{R}^{\alpha}} 4\pi \sum_{\ell m} i^{\ell} j_{\ell}(|\mathbf{k}+\mathbf{G}|r_{\alpha}) Y_{\ell m}(\hat{\mathbf{r}}_{\alpha}) \widehat{Y_{\ell m}^*(\mathbf{k}+\mathbf{G})}. \quad (2.58)$$

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

$$A_{\ell m}^{\alpha}(\mathbf{k}+\mathbf{G}) = \frac{e^{i(\mathbf{k}+\mathbf{G})\mathbf{R}^{\alpha}} 4\pi \sum_{\ell m} i^{\ell} j_{\ell}(|\mathbf{k}+\mathbf{G}|r_{\alpha}) Y_{\ell m}^*(\widehat{\mathbf{k}+\mathbf{G}})}{\sqrt{\Omega} u_{\ell}^{\alpha}(r_{\alpha}, E)}. \quad (2.59)$$

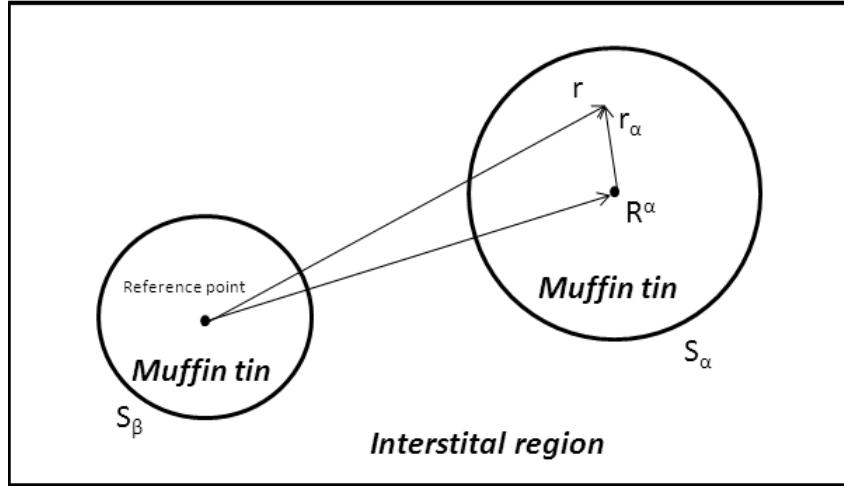


Figure 2.3. Partition of the unit cell.

There are one main drawbacks about the APW method. The wavefunction is energy dependent, which means that the code will 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 and 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 expressed as follows:

$$u_\ell^\alpha(r_\alpha, E) = u_\ell^\alpha(r_\alpha, E_\ell) + (E - E_\ell)\dot{u}_\ell^\alpha(r_\alpha, E_\ell) + O((E - E_\ell)^2). \quad (2.60)$$

Therefore they re-defined the wavefunction in the following way:

$$\phi_{\mathbf{k}+\mathbf{G}}^{LAPW}(\mathbf{r}) = \begin{cases} \frac{1}{\sqrt{\Omega}} e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}} & \text{if } \mathbf{r} \in I \\ \sum_{\alpha} \sum_{\ell m} f_{\ell m}(r_\alpha, \mathbf{k}+\mathbf{G}, E_\ell) Y_{\ell m}(\hat{\mathbf{r}}_\alpha) & \text{if } \mathbf{r} \in S_\alpha, \end{cases} \quad (2.61)$$

where $f_{\ell m}(r_\alpha, \mathbf{k}+\mathbf{G}, E_\ell) = A_{\ell m}^\alpha(\mathbf{k}+\mathbf{G})u_\ell^\alpha(r_\alpha, E_\ell) + B_{\ell m}^\alpha(\mathbf{k}+\mathbf{G})\dot{u}_\ell^\alpha(r_\alpha, E_\ell)$. $A_{\ell m}^\alpha(\mathbf{k}+\mathbf{G})$ and $B_{\ell m}^\alpha(\mathbf{k}+\mathbf{G})$ are the expansion coefficients, and $\dot{u}_\ell^\alpha(r_\alpha, E_\ell)$ is the derivative of the radial function.

Here energy E_ℓ is considered as pre-calculated parameter, 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, but it has to match for two parameters, fortunately, even though, it

still use less time comparing with APW method. However, there is one drawback about this method, what if energy difference is bigger enough in the same ℓ character, which E_ℓ is correct one? Therefore this situation could be a big error, these states are called as semi-core state, for example, it exists 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 the basis set, and add smaller number of basis set, which has the following format:

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

where $A_{\ell m}^\alpha$ and $B_{\ell m}^\alpha$ is matching value and derivative on the sphere boundary to zero, but not plane wave, and E'_ℓ is the chosen energy from semi-core state.

2.2.2.4 Augmented plane wave method plus local orbitals

Actually, there is one more method which will deal with the energy-dependent case, which is called as augmented plane wave method plus local orbitals (APW+lo), the basis function has two kinds, one is similar with APW method, but only without the derivative terms, e.g., $f_{\ell m}(r_\alpha, \mathbf{k}+\mathbf{G}, E_\ell) = A_{\ell m}^\alpha(k+G)u_\ell^\alpha(r_\alpha, E_\ell)$. Another basis function is:

$$\phi_{\mathbf{k}+\mathbf{G}}^{lo}(\mathbf{r}) = \begin{cases} 0 & \text{if } \mathbf{r} \in I \\ (A_{\ell m, lo}^\alpha u_\ell^\alpha(r_\alpha, E_\ell) + B_{\ell m, lo}^\alpha \dot{u}_\ell^\alpha(r_\alpha, E_\ell)) Y_{\ell m}(\hat{\mathbf{r}}_\alpha) & \text{if } \mathbf{r} \in S_\alpha. \end{cases} \quad (2.63)$$

The value of $A_{\ell m, lo}^\alpha$ and $B_{\ell m, lo}^\alpha$ are obtained by normalization and local orbital has zero value at the muffin tin boundary. This method can not deal with semicore states like LAPW+LO, however, it does increase the efficiency to linearize Slater's APW method.

2.2.3 Effective potential

The potential in the FP-LAPW method is also divided into two regions, the MT region and the interstitial region.

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

2.3 Dielectric function

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

$$\varepsilon = \varepsilon_1 + i\varepsilon_2, \quad (2.64)$$

where 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.

Here a summary of the basic derivation of dielectric function is presented using random-phase approximation (RPA).

Response to external electric field E in the linear approximation:

$$D = \varepsilon E, \quad (2.65)$$

where D is electric displacement, and E is the electric field.

After some derivations, the imaginary part of ε is:

$$\varepsilon_2 = \frac{4\pi e^2}{m^2\omega^2} \sum_{cv} \int d\mathbf{k} \langle \Psi_{ck}|p^\alpha| \Psi_{vk} \rangle \langle \Psi_{vk}|p^\beta| \Psi_{ck} \rangle \delta(\varepsilon_{ck} - \varepsilon_{vk} - \omega), \quad (2.66)$$

where c and v run over the conduction bands and valence bands, for example, one can find the figure which is calculated using this equation in the paper of ??.

The interband contribution of dielectric function is:

$$\varepsilon_2 = \frac{4\pi e^2}{m^2\omega^2} \sum_{cv} \int d\mathbf{k} \langle \Psi_{ck}|p^\alpha| \Psi_{vk} \rangle \langle \Psi_{vk}|p^\beta| \Psi_{ck} \rangle (f(\varepsilon_{ck}) - f(\varepsilon_{vk})) \delta(\varepsilon_{ck} - \varepsilon_{vk} - \omega), \quad (2.67)$$

which is often calculated in the field that is related to optical property, for example, it is calculated in the paper of ??, which is plotted in the Fig.?? of that paper.

Here only some very 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. ??.

Chapter 3

Results and discussion

In this chapter, the major result for the doctoral training is demonstrated. It has two branches, one is the result from the point of modeling and calculations for the thin film materials copper indium gallium (di)selenide (CIGS) and copper zinc tin sulfide/selenide (CZTS/Se), the other is methodology, which includes the result from scalar relativistic approximation (SRA) and spin-orbit coupling (SOC).

3.1 Modeling and calculations

3.1.1 Copper indium gallium diselenide

In this section, the parameterization of $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ ($x=0, 0.5$, and 1) energy bands is described firstly. Afterwards some results based on this parameterization are demonstrated. At last, calculation of the dielectric function spectra for the material $\text{CuIn}_{0.5}\text{Ga}_{0.5}\text{Se}_2$ is presented and analyzed. Moreover, it compared with experimental results as well.

3.1.1.1 Parameterization of energy bands

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 assumed normally:

$$E_j^{pb}(\mathbf{k}) = E_j(\mathbf{0}) \pm \left[\frac{\tilde{\mathbf{k}}_x^2 + \tilde{\mathbf{k}}_y^2}{m_j^\perp} + \frac{\tilde{\mathbf{k}}_z^2}{m_j^\parallel} \right], \quad (3.1)$$

where m_j^\perp is transverse electron masses and m_j^\parallel is longitudinal electron masses.

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 valences bands (VBs) and the lowest conduction band (CB) for the materials $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ where $x=0, 0.5$, and 1 , that is, CuInSe_2 , $\text{CuIn}_{0.5}\text{Ga}_{0.5}\text{Se}_2$, and CuGaSe_2 . This parameterization is based on the all electron and full-potential linearized augmented plane wave (FP-LAPW) calculation. Normally, the $\mathbf{k} \cdot \mathbf{p}$ method is utilized to parameterize the energy bands. However, due to the crystal-field interaction and the spin-orbit coupling, the energy dispersions of $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ are rather complex, 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. Finally, the following equation is used to parameterize the energy bands, which is valid around 0.5 eV below VBM and 0.5 eV above CBM.

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

where E_j^0 , $\Delta_{j,n}$, and $\delta_{j,m}$ are fitting parameters, 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.

Table 3.1. Parameters of equation 3.1**Table 3.2.** Parameters of equation 3.2

Firstly, 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 FP-LAPW method and the pba in Eq. 3.1.

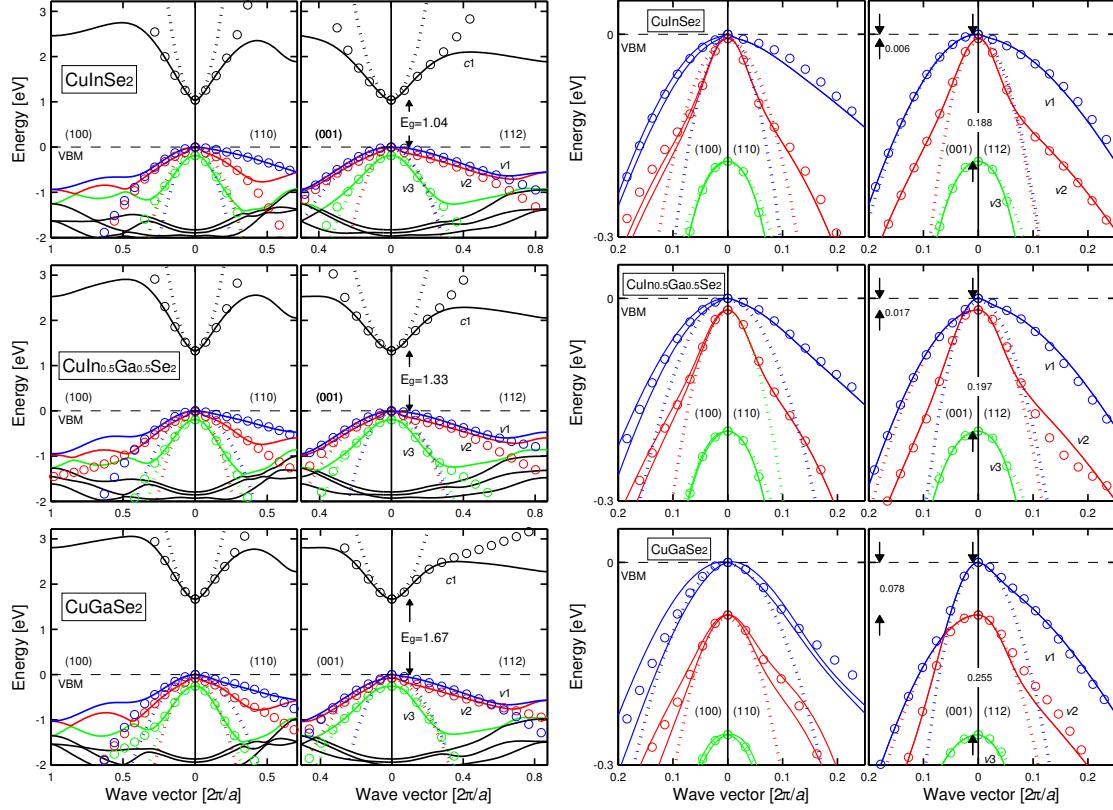


Figure 3.1. Left panel: Electronic band structure along four directions. the circles are the results of the full band parameterization (fbp), and the dotted lines represent the parabolic band approximation (pba). Righ panel: The close-up of left panel for the valence bands close to Γ point.

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

In order to demonstrate the anisotropic and non-parabolic of energy bands, some properties are investigated, such as effective electron and hole masses and constant energy surface.

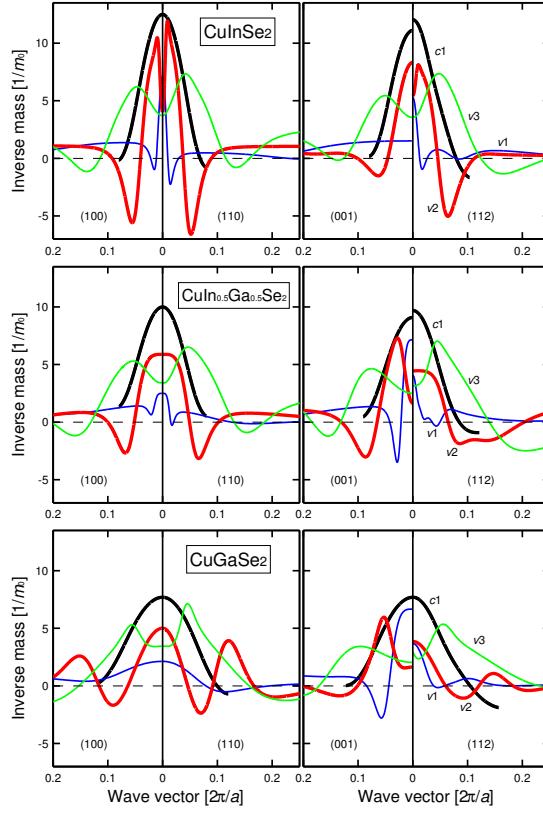


Figure 3.2. Inverse of the effective electron and hole masses in the four symmetry directions for the $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ ($x=0, 0.5$, and 1).

The effective masses are calculated numerically by:

$$m_j(\mathbf{k}) = \pm \hbar^2 / (\partial^2 E_j(\mathbf{k}) / \partial \mathbf{k}^2). \quad (3.3)$$

In order to better visibility, therefore, the inverse of the mass are presented. The result from Fig. 3.2 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 anisotropy close to the Γ point, however, the electron masses of conduction bands are rather isotropic ($m_{c1}^{100}(\mathbf{0}) \approx m_{c1}^{110}(\mathbf{0}) \approx m_{c1}^{001}(\mathbf{0}) \approx m_{c1}^{112}(\mathbf{0})$).

To further demonstrate the non-parabolic of energy bands, the constant energy surfaces for CuInSe_2 and CuGaSe_2 are plotted.

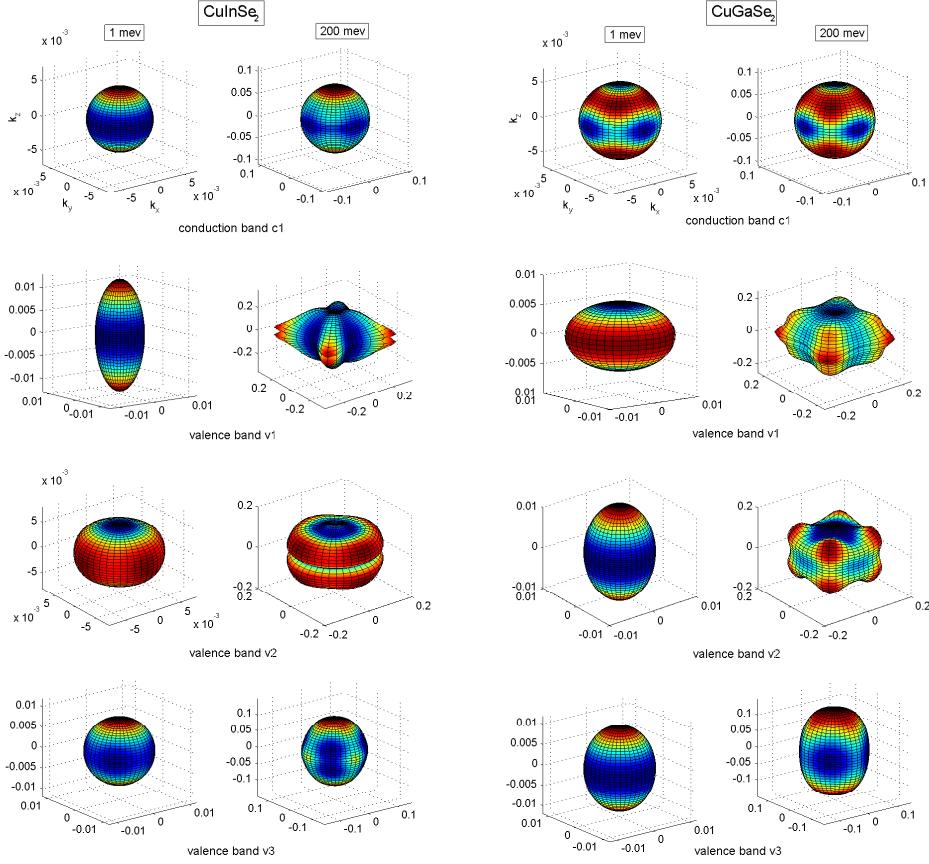


Figure 3.3. Constant energy surfaces for the three uppermost VBs and the lowest CB for the energies $E = 1$ meV (left column ellipsoidal) and $E = 200$ meV (right column). The $CuInSe_2$ and $CuGaSe_2$ are demonstrated in this figure.

In the **Fig. 3.3**, 1 meV represents that it is near to Γ point ; 200 meV means 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 topes 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 $m_{v1}^{\parallel}=0.66m_0$). 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 ellipsoidal shape at all when the energies goes up to 200 meV.

In order to analyze the impact of non-parabolicity and anisotropy of the energy bands, some properties are compared in the following text between fbp and pba, respectively, such as density-of-states (DOS), DOS mass, carrier concentration in intrinsic and p-type of $CuIn_{1-x}Ga_xSe_2$.

The DOS in j th band is defined as:

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

where $E_j(\mathbf{k}) = E$ 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 will be written:

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

where 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 ($m_{v/c}^{DOS}$) is defined:

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

where the DOS mass contains the non-parabolicity and anisotropy of the band dispersion.

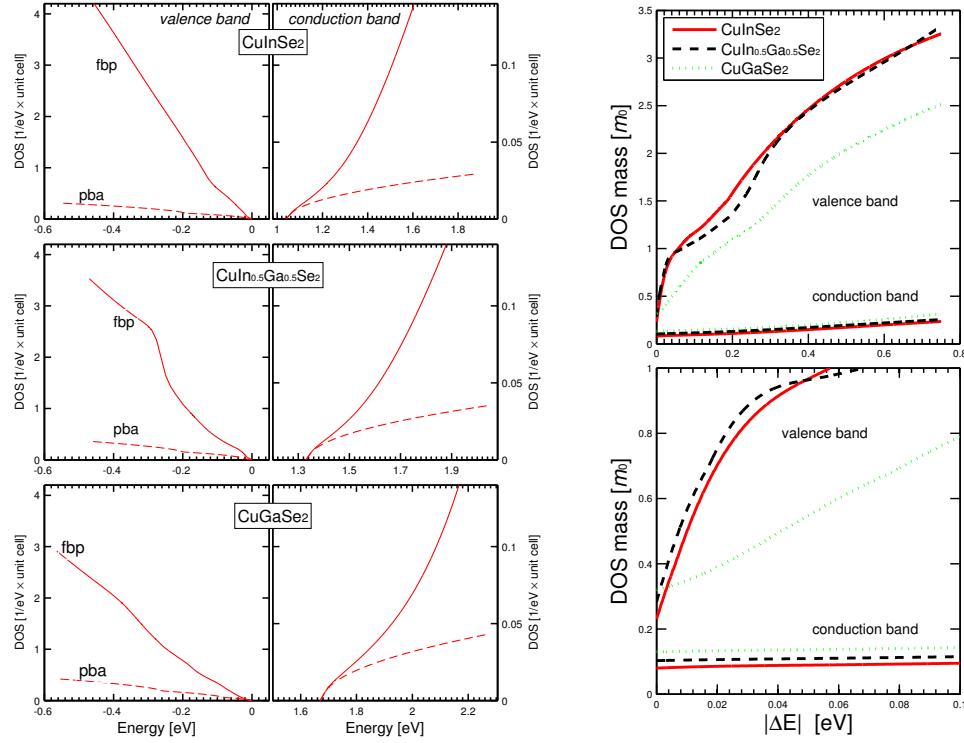


Figure 3.4. Left: Total DOS of the VBs and CB. The solid lines show the full band parameterization (fbp), and the dashed lines represent the parabolic band approximation (pba). Right: The DOS mass $m_{v/c}^{\text{DOS}}$ is calculated in the equation 3.6.

The **Fig. 3.4** (left) demonstrates 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 are more flat than parabolic bands. The **Fig. 3.4** (right) demonstrates that the DOS masses of $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ are strongly energy dependent 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 effective 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.

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

$$\begin{aligned} n_v(T) &= \int_{-\infty}^{E_{v1}(0)} g_v(E)(1 - f(E))dE, \\ n_c(T) &= \int_{E_{c1}(0)}^{\infty} g_c(E)f(E)dE, \end{aligned} \quad (3.7)$$

where $f(E) = 1/[1 + \exp(E - E_F)/k_B T]$ 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:

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

where $N_{A_{\alpha}}$ is the acceptor concentration of the α th defect, $\Delta_{A_{\alpha}}$ means 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 used 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.

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

where parameters a and b are used from experimental values.

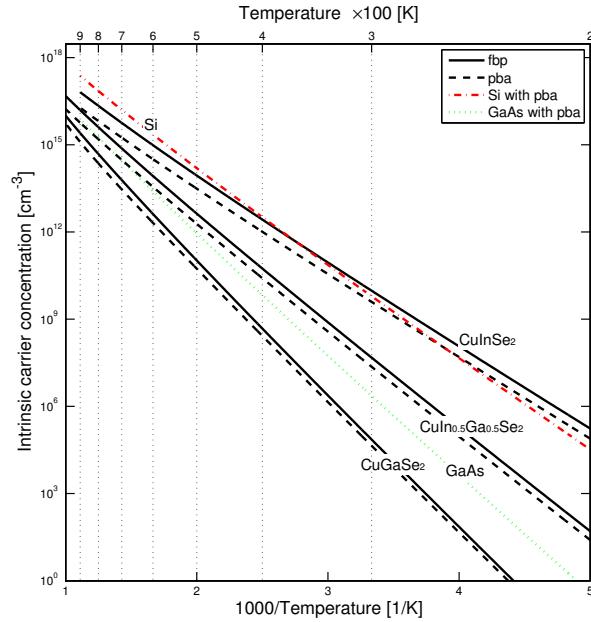


Figure 3.5. Intrinsic carrier concentration as function of temperature.

From Fig. 3.5, the free carrier concentration is increased dramatically with the increasing of temperature for the $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$. For example, in the case of CuInSe_2 , the carrier concentration is increased up to around 10^5 times higher from temperture 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.

The p-type carrier concentration (Fig. 3.6) for $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ is presented using Eq. 3.9.

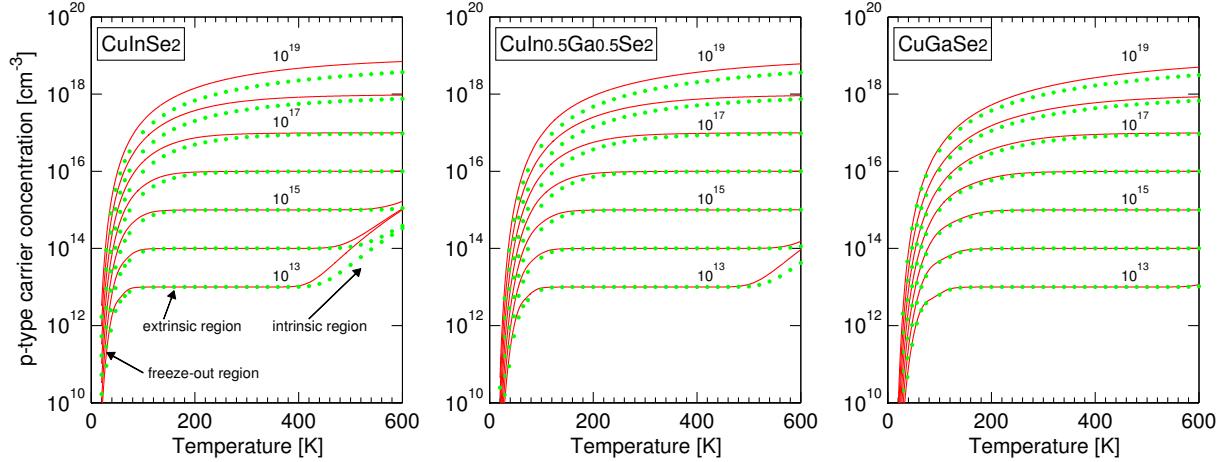


Figure 3.6. Free carrier concentration as function of the temperature in p-type for CuInSe₂, CuIn_{0.5}Ga_{0.5}Se₂, and CuGaSe₂. The effective doping concentration $N_A = 10^{13}, 10^{14}, 10^{15}, \dots, 10^{19}$ acceptors/cm³ are considered.

From **Fig. 3.6**, 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.

3.1.1.2 Dielectric function spectra

In this work, the dielectric function (ϵ) spectra of CuIn_{0.5}Ga_{0.5}Se₂ is calculated by the full-potential linearized augmented plane wave (FP-LAPW) method using the generalized gradient approximation (GGA) plus an onsite Coulomb interaction U of the Cu d states. Afterwards, the different contributions to $\epsilon_2(\text{Im}(\epsilon))$ 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 (CuIn_{0.7}Ga_{0.3}Se₂) at temperature of 40 K and 300 K, and they are in a good agreement.

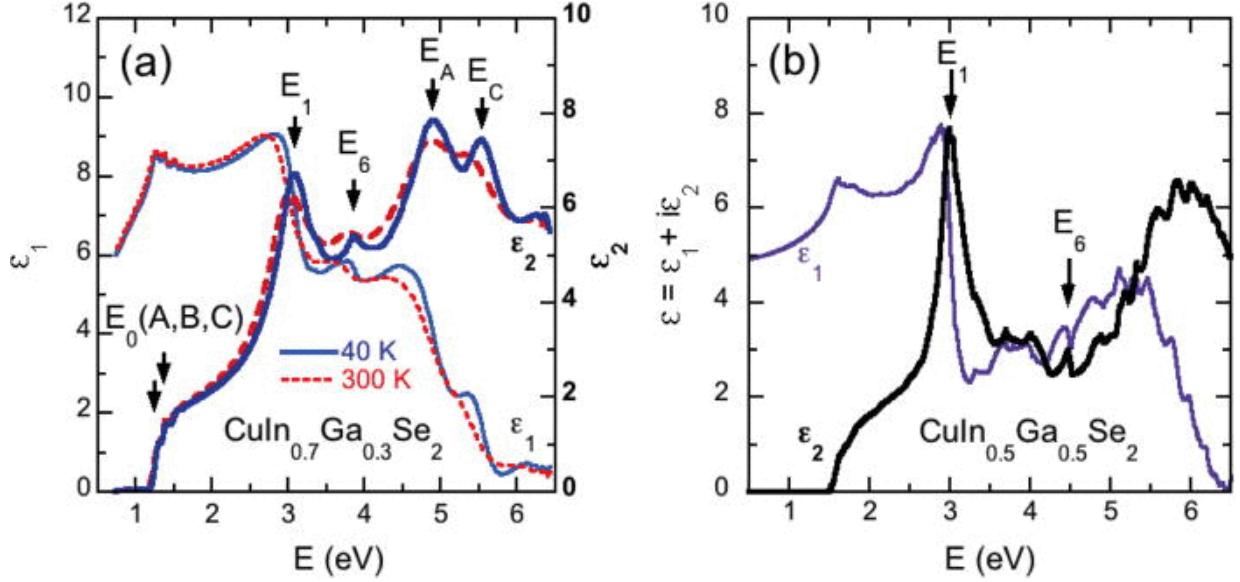


Figure 3.7. Left panel: The real (ϵ_1) and imaginary (ϵ_2) part of dielectric function spectra for $CuIn_{0.7}Ga_{0.3}Se_2$ at 40 K (solid blue line) and 300 K (dashed red lines). Four prominent CP features are shown. Right panel: the dielectric function spectra for $CuIn_{0.5}Ga_{0.5}Se_2$ calculated by FP-LAPW method at 0 K. The major CP features are identified.

From **Fig. 3.7**, 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.

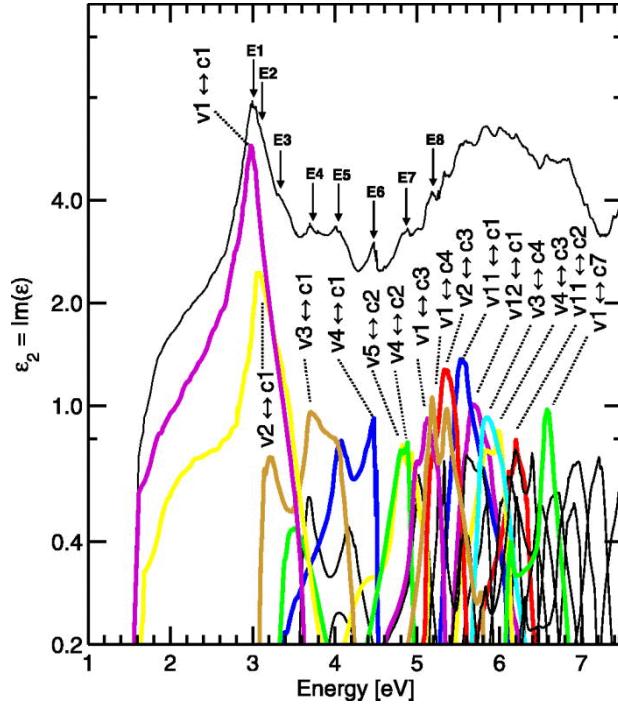


Figure 3.8. Band-to-band analysis of the contribution to the total ε_2 spectrum. The vertical axis is in the log scale.

where v_1 and c_1 means the topes valence band and lowest conduction band, respectively. From the **Fig. 3.8** and **Fig. 3.9**, 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. 3.8**. 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. 3.10**). 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.

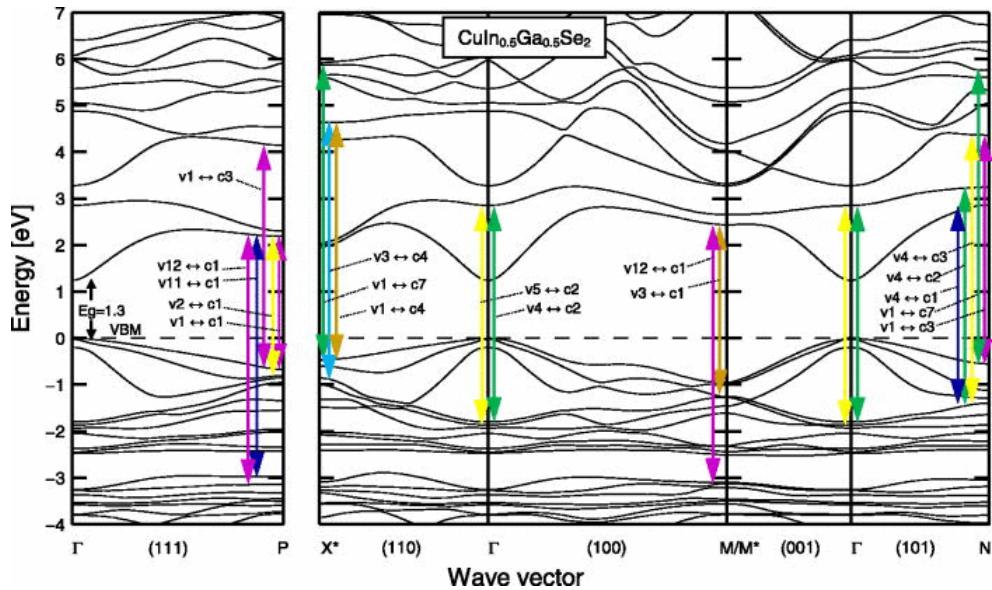


Figure 3.9. Band-to-band analysis of the contribution to the total ε_2 spectrum. The vertical axis is in the log scale.

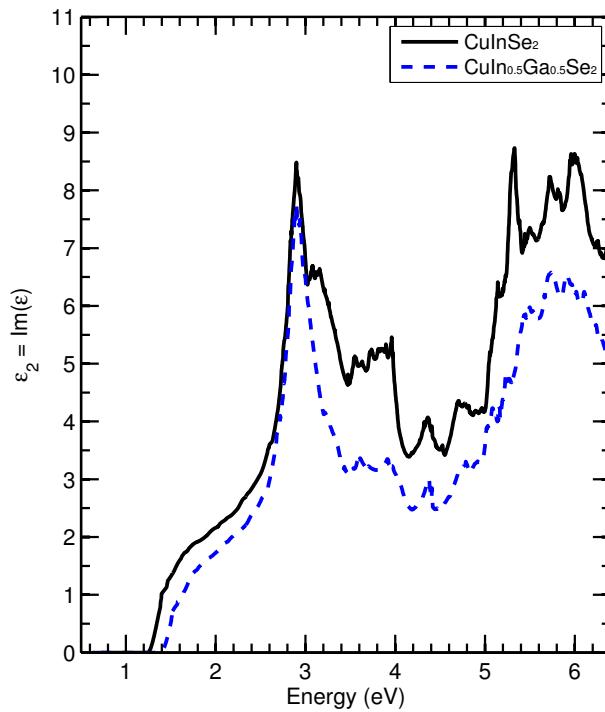


Figure 3.10. The ε_2 spectra for CuInSe₂ and CuIn₀·₅Ga₀·₅Se₂.

Chapter 4

Concluding remarks and future work

Acknowledgements