# Ab initio Calculations of the Electronic Properties of CuIn(S,Se)2 and other Materials for Photovoltaic Applications

Julien Vidal

Thèse présentée pour obtenir le grade de

DOCTEUR DE L'ÉCOLE POLYTECHNIQUE

par

Julien VIDAL

---

# *Ab initio* Calculations of the Electronic Properties of $CuIn(S,Se)_2$ and other Materials for Photovoltaic Applications

---

Soutenue le 21 Mai 2010 devant le jury composé de :

| | | |
|---|---|---|
| John J. | REHR | Rapporteur |
| Mark | VAN SCHILFGAARDE | Rapporteur |
| Susanne | SIEBENTRITT | Examinateur |
| Silke | BIERMANN | Examinateur |
| Frédéric | JOUVE | Examinateur |
| Silvana | BOTTI | Directrice |
| Jean-François | GUILLEMOLES | Directeur |
| Lucia | REINING | Directrice |

# Contents

# Chapter 1

# Motivations and introduction

The recent worldwide rise of important issues like global warming and energy supply has shone some light on solar energy and most notably photovoltaics (PV). Even though the physical principle of this technology was discovered more than 170 years ago by Alexandre-Edmond Becquerel who observed an increase of conductivity of his sample under light irradiation, the manufacture of photovoltaic solar cells has only advanced recently. Indeed, PV presents several advantages compared to other electrical power generation technologies:

- the potential source of energy is huge : 174 PW received by the earth from the sun at the upper atmosphere corresponding to a solar irradition of $\sim 1300\mathrm{W/m^2}$.

- it is a *green* source of energy (life-cycle greenhouse gas emission ranging between nuclear and wind and an order of magnitude smaller than fossil energy sources.)

- there is still room for improvement: from the point of view of mass production, there are enormous possibilities of cutting cost and improving the efficiency. From the point of view of research, the record efficiency of a photovoltaic solar cell (40.7%) is still far from the theoretical limit for energy conversion of solar radiation (Carnot efficiency $\eta \approx 85.0\%$ considering the average temperature on earth).

While the first practical devices were built more than 50 years ago, PV is still a wide (and wild) field of discoveries for physicists, chemists or engineers. New solar cell designs and principles arise at a very fast rate, keeping the PV research area extremely challenging and interesting. One particular subject of interest is the search for new materials. Indeed, PV solar cells

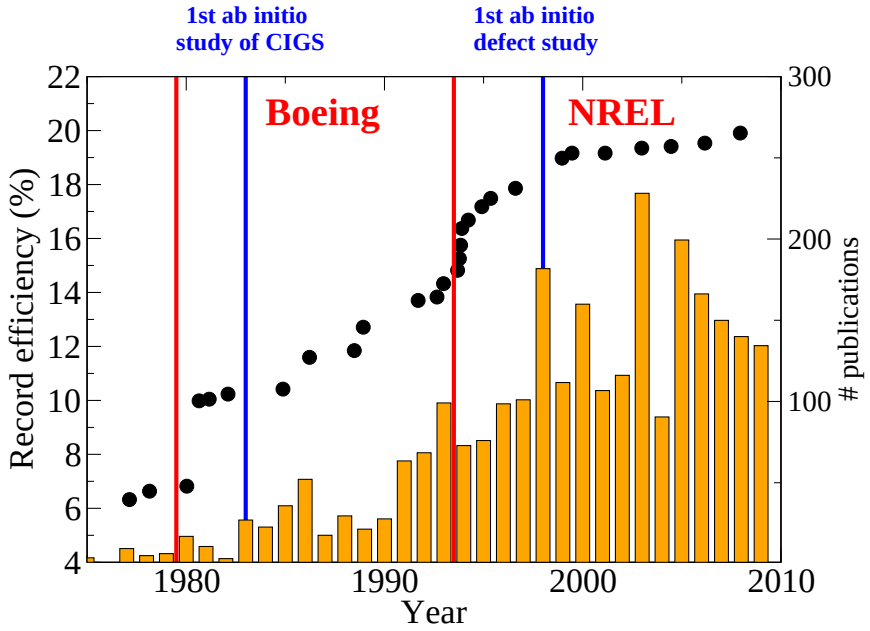


**Figure 1.1:** Bar diagram: Number of publications on Cu(In,Ga)$Se_2$ versus year. Curve: record efficiency of Cu(In,Ga)$Se_2$ solar cells versus year. Two blue line refers to the two most cited articles based on *ab initio* methods. [1, 2]

are stackings of layers made of different materials with specific properties. These properties were chosen in order to achieve the precise goal of a solar cell: convert light into electricity. If only one of these materials fail in its task, then the solar cell will display poor properties. Therefore, a profound understanding of the physics and chemistry of materials is needed in order to build high efficiency device.

This thesis has been written in this prospect: applying advanced methods in order to gain deeper insight in the physics of the materials. While it is a very exciting challenge, it is also an extremely difficult and paradoxical task. On one hand, the theoretical explanation of the PV effect waited for more than 100 years between Becquerel's discovery and the Shockley Queisser paper [3]: it involved quantum effects that require to treat the material at the nanoscale level. On the other hand, scientists had built solar cells for years since Becquerel's discovery, dealing with macroscopic quantities like current density and voltage. Microscopic and macroscopic worlds usually do not get along very well: one can find one striking example in this thesis as microscopic changes of the crystalline structure induce large macroscopic variations of the bandgap which are experimentally not observed. We study CuIn(S,Se)$_2$ (CIS), belonging to the family of the Cu-based chalcopyrite which is believed

to be the most promising class of material of the thin film technologies. It is a very good light absorbing material which displays intrinsic doping properties, electrical and optical innocuity to large stoichiometric variations and other remarkable properties for PV applications. Figure 1.1 shows the evolution of efficiency of $CuInGaSe_2$-based solar cell versus time. For the last 30 years, the research conducted on this material has achieved a steady increase of solar cell performance: one should notice the two abrupt *steps* in the evolution related to synthesis improvements (the so-called Boeing process and 3-stage process), and the two main theoretical contributions to the field based on first principle calculations [4, 2]. Each of the latter has permitted a better understanding of the physics of the material and most notably of the physics of defects. It explained the capability of doping of the material and the innocuity of large defect concentrations on the optical properties of CIS. However these two contributions were using *ab initio* methods based only on Density Functional Theory (DFT). While being the state of the art method when dealing with ground state properties, this method is known to fail considerably for excited states properties (for example it underestimates the bandgap by 50% to 100%). In the case of CIS, the underestimation of the bandgap is so large that DFT predicts CIS to be a semimetal despite the experimentally observed insulating nature of CIS. Therefore, we use the $GW$ method which is a state-of-the-art method when dealing with bandgaps. For the last few years, important efforts have been carried out in order to make this method computationally tractable [5]. $GW$ is now able to treat systems with typically up to 100 atoms. In the case of CIS, it requires however particular care because of the presence of $d$ electron. Therefore as a test case ZnO was studied first in order to get some technicalities of $GW$ tested on a simple model. Self-consistent schemes for the $GW$ calculation turned out to be necessary in the case of CIS and ZnO, which makes calculations more cumbersome. In the end, however, $GW$ not only succeeds in describing correctly the bandgap of CIS but also reveals the existence of a *hidden* dependence of the bandgap. While the issue had been asserted long time ago in one of the first papers using *ab initio* methods on CIS [1], it has been discarded for more than 20 years in most theoretical and experimental studies. The dependence of the bandgap on an atomic displacement is very surprising as thin films or polycrystalline CIS samples show an extremely stable bandgap despite large dispersion of some structural parameters. At the same time, the samples are usually not ideal crystals. The question of defects in CIS had been already adressed theoretically [2]. It was shown that some of the intrinsic defects have a very low formation energy that can explain the observed deviation from stoichiometry. In this thesis, we propose a mechanism including a feedback loop that relates changes in the lattice structure to the

changes in the defect concentration.
In the first chapter of this thesis, we will present the principle of PV solar cells with a special emphasis on the CIS absorber. In the second and third chapter, we will describe the methods we used to treat the many body problem. Finally, in the last chapter, we will apply methods presented in chapter 2 and 3 to CIS and pay a particular attention to the dependence of the bandgap on the anion displacement and the concentration of defects.

# Chapter 2

# Materials for solar cells

The photovoltaic (PV) effect was discovered in 1839 by Alexandre-Edmond Becquerel who observed an increase of conductivity under light irradiation. The effect was finally understood by Einstein in 1905, by introducing the concept of light quanta [6].
Let us consider the band structure of a semiconductor: the valence states are separated from the conduction states by a forbidden gap where no state exists (see Fig. 2.1). This forbidden gap is called the bandgap $E_g$. By shining light on a sample, the photons promote the transition of the electrons from the valence to the conduction, leaving a hole in the valence. Clearly, this can only happen if the energy of the impinging photon $\hbar\nu$ should be larger than the bandgap $E_g$. The electron-hole pair can either recombine, by reemitting a photon, loose its energy to the lattice or be fully separated and participate in the current flow. Due to this latter possibility, the photovoltaic effect can be used in order to provide electrical current and power. The increase of conductivity observed by Becquerel happens for photon energies $\hbar\nu > E_g$: therefore it is natural to consider the bandgap as the keystone of the photovoltaic effect. It should closely correspond to the solar spectrum in order to obtain maximum photovoltaic conversion. However, the simple picture describing the photovoltaic effect is rather far from the reality of a photovoltaic device. In order to extract electrical power, one should not only favor light absorption by the material but also prevent the recombination of electron-hole pairs and enable a correct charge collection. Figure 2.1 displays one of the many recombination processes that may occur in a solar cell where the presence of a localized state inside the bandgap can act as a recombination center. The electrons populating the conduction bands can relax to such state, emitting a phonon. Once the electron is localized, it does not participate anymore in the conductivity and moreover it can recombine further with a hole in the valence. Such an effect is one of the main shortcomings of the actual devices.

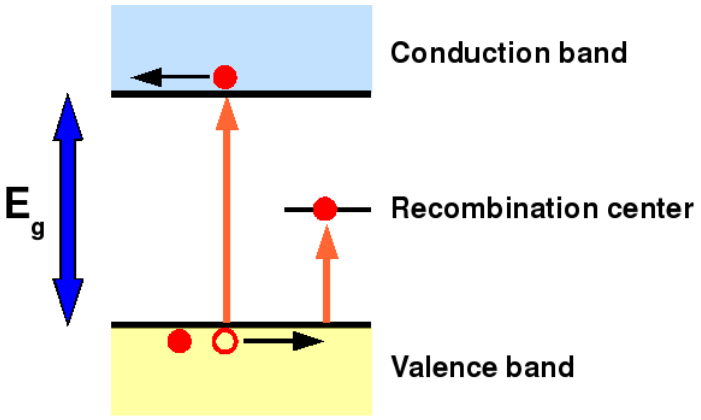


**Figure 2.1:** Schematic of the photovoltaic effect. Also one of the major source of loss in a photovoltaic solar cell recombination due to trap located inside the bandgap is displayed on the right.

## 2.1 Photovoltaic solar cells

### 2.1.1 History and technology

The first practical solar cell was built at Bell laboratories in the 50s [7]. It showed a conversion efficiency of about 6% which was orders of magnitude larger than for the selenium photocells. The device developed by Chapin and co-workers used silicon grown in a *special manner* [8], *i.e.*, high quality crystalline silicon. Furthermore, they also introduced a device design that would be used in the following 50 years to build photovoltaic solar cells: the p-n junction, formed by two different materials, one p-doped and the other one n-doped. The association of two materials with different doping properties creates an electric field across the junction which can be used to separate the electron-hole pairs and to conduct the charges toward the electrodes. However most of solid state PV solar cell still rely on the p-n junction. The developpement of possible applications followed shortly after Chapin's breakthrough with the first autonomous electrical system installed in Americus, Georgia in 1955 [8]. These first solar cells were exclusively based on silicon technology and consequently benefited from all the advances in other silicon-based devices. The choice of silicon was certainly dictated by technological limitations and by theoretical works that predicted the maximum photovoltaic conversion for semiconductor to occur for materials whose

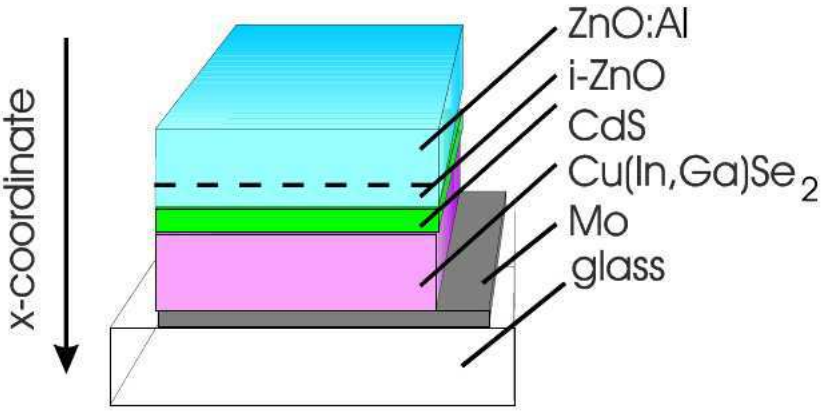


**Figure 2.2:** Representation of the solar cell displaying the layer stacking ZnO/CdS/Cu(In,Ga)$Se_2$. [9]

bandgaps $E_g \approx 1.1$ eV [3]. However, silicon is a rather bad energy harvesting material as it has an indirect bandgap and low absorption coefficient over the range of interest of the solar spectrum.
During 20 years, PV cells remained too expensive when compared to other sources of energy. Only one domain departed from this trend: the use of PV panels for space applications. High efficiency multiple junctions based on GaAs and other semiconducting alloys were developed to selectively absorb the solar spectrum outside Earth's atmosphere and thereby improving the conversion. Nevertheless, their complexity made their industrial production extremely expensive but they remained the only viable option for energy production in space. The oil crisis in the 70s, the constant decrease of PV prices and the emergence of new technologies in the 80s and 90s made the PV solar cells more and more attractive. One of these emerging technologies is the thin-film solar cell. At first sight, it has many advantages: the use of smaller quantities of material results in a significant reduction of the cost, the possibility of using different substrates opens the way to flexible solar cells, and lower cost of better absorbing material when compared to crystalline PV solar cells. Considering the low absorption capability of silicon, the latter advantage is of primary interest and resulted in the advent of three materials for thin film solar cells:

- amorphous silicon (a-Si): It has a direct bandgap of 1.7 eV but suffers from the Staebler-Wronski effect [10]. While not completely understood, this light induced metastability is most likely caused by defects

or mediated by hydrogen atoms. It results in a decrease of the efficiency during the first 6 months of use. a-Si is currently the leading technology of the thin film family but due to the above mentioned limitation it is thought that it will leave its spot to one of the following technologies.

- CdTe: it has an optimal bandgap for photovoltaic conversion (1.4 eV). However, the major drawback of CdTe is the health risk of Cd, which impacts unfavorably the economical and technical growth of this technology. Moreover, it suffers from severe losses due to collection issues, junction recombination and low carrier concentration [11].

- Cu(In,Ga)$Se_2$ (CIGS) and related alloys: the most promising as it holds the highest record efficiency for thin film solar cells. CIGS presents the advantage of being electrically and optically tunable due to its mixture of In and Ga. Moreover, it exhibits a large absorption coefficient (two or three orders of magnitude larger than the one of silicon at $\omega = 1.5$ eV) which permits to obtain rather large currents with very thin films.

### 2.1.2 CIGS solar cell

Semiconductor-based solar cells are formed by a complex stacking of layers that are designed to achieve three very specific tasks: (i) transport of light to the absorber, (ii) generation of electron-hole pairs, (iii) separation and transport of the charges to the contacts. Therefore, the skeleton of the device is usually formed by a p-n junction. Extra transparent layers are deposited on top of the p-n junction in order to realize the contacts and to allow current extraction.
On a substrate[1], a layer of molybdenum is grown to realize the back contact of the device. Molybdenum presents all the qualities required for a potential back contact material: it displays good electric conductivity and sustains very high temperatures due to its high melting point which enables the use of any techniques to grow the CIGS absorber layer on top of it. A direct consequence of the use of Mo as back contact is the formation of Mo(S,$Se)_2$ at the interface between CIGS and Mo: the size of this layer is significant and impacts the performance of the device. This back contact is in most cases found to be an ohmic contact [13].
The absorbing layer is then grown and can be made p-type or n-type depending on the condition of growth. As the transparent conducting layer

[1] The development of many types of substrates in the past years has favored practical applications of thin films technologies: the original glass substrate has been replaced by flexible substrates such as titanium foil or polyimide [12].

is usually doped n, the most natural doping for CIGS is with holes p-type. Moreover, p-type thin films of CIS present usually much better electrical properties than their n-type counterparts. Then follows a layer,the so-called buffer layer, that is formed by a n-type transparent semiconductor. CdS has been selected despite its small bandgap. Thereby, some electron-hole pairs are also created in the buffer layer but do not contribute significantly to the photogenerated current. There have been some promising attempts to replace CdS by ZnS [14] or $In_2S_3$ [15] for instance that leads to an increase of the efficiency of the solar cell [16].
The last layer to deposit is a transparent conducting oxide (TCO) that plays the role of the front contact. Some oxides display both transparent and conducting properties through a complex interplay between doping and optical properties. ZnO heavily doped with Al (ZnO:Al) is the usual TCO deposited on top of the buffer layer. A large amount of research has been carried out trying to improve this stage as losses at the front contact are believed to plague the efficiency of the device. One of the conclusions is that the deposition of undoped ZnO (i-ZnO) prior to the deposition of the doped one can greatly improve the solar cell. We will see in the final section of this chapter that this is due to the particularly complex defect physics of CIGS.

**In conclusion**, we describe the structure of a CIGS-based solar cell. The number of interfaces together with the variety of materials used make the working of such device extremely complex. A single *defective* layer might hinder the efficiency of the whole device. In the next section, we will present the physics behind the solar cell, most notably the origin of the current-voltage dependence of the solar cell.

## 2.2 Physics of the solar cell

The semiconductor-based solar cells generally rely on the physics of the p-n junction as this is the central component of the device. When two semiconductors respectively n- and p-doped are assembled together, diffusion of charges between the two materials occurs creating two distinct regions in each doped semiconductor: the space charge region, where all dopants are ionized creating a uniform background charge, and the quasi-neutral region, one in each doped semiconductor. When applying a voltage to the device, the width of the space charge region can be modulated, allowing carriers to drift from one material to the other. As the materials have different type of majority carriers due to their concurrent doping, a recombination current appears in each of the quasi-neutral regions in order to regulate the excess of minority carriers that cross the space charge region. The current voltage

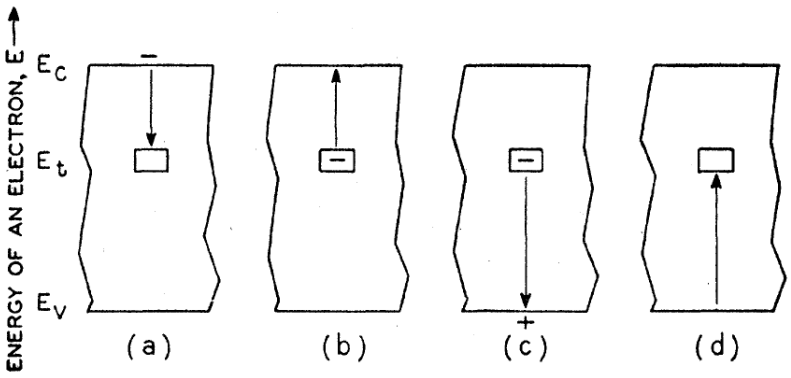


**Figure 2.3:** Recombination and generation processes consider in the Shockley-Hall-Read theory based on a trap states in the middle of the gap: (a) capture of an electron, (b) emission of an electron, (c) capture of a hole, (d) emission of a hole. [17]

curve can often be expressed as

$$I = I_0(e^{\frac{qV}{k_B T}} - 1)\,, \tag{2.1}$$

where $I_0$ is the saturation current, $k_B$ the Boltzmann constant and $T$ the temperature. This equation is the so-called Shockley diode equation. It relies on many assumptions about the recombination processes in the different regions of the p-n junction (there is no recombination in the space charge region for example) or the amount of minority carriers present in the device (the so-called *low injection* approximation). However, most of these assumptions do not usually hold for solar cells:

- due to the high number of electron-hole pairs generated in the different regions of the device, the low injection approximation can fail.
- the recombination in the space charge region cannot be discarded.

Let us consider the continuity equation in the steady state regime

$$0 = G_n - R_n + \frac{1}{q}\nabla J_n\,, \tag{2.2}$$

$$0 = G_p - R_p + \frac{1}{q}\nabla J_p\,, \tag{2.3}$$

where $G_n$ ($G_p$) is the generation rate for electrons (holes), $R_n$ ($R_p$) is the recombination rate for electrons (holes) and $J_n$ ($J_p$) is the current density of

electrons (holes). By performing an integration over the whole volume of the device

$$\int d^3\mathbf{r}\frac{1}{q}\nabla J_n = \int d^3\mathbf{r}\,(R_n - G_n)\,, \tag{2.4}$$

$$\int d^3\mathbf{r}\frac{1}{q}\nabla J_p = \int d^3\mathbf{r}\,(R_p - G_p)\,. \tag{2.5}$$

The divergence theorem allows us to transform the integration over the volume to a surface integration and then obtain directly the total current as

$$I = \int d\mathbf{r}^3 q\,(R_n - G_n)\,. \tag{2.6}$$

The I-V characteristic depends on the recombination processes that take place in the device. If one would like to obtain the diode equation, one should:

- restrict the integration domain to the quasi-neutral region only as in the diode model no recombination is assumed
- consider the net recombination rate $U$ originating from a standard conduction-to-valence recombination process:

$$U_{n,p} = R_{n,p} - G_{n,p} = \delta(n,p)/\tau_{n,p}\,, \tag{2.7}$$

  where $\tau_{n,p}$ is an effective lifetime of electrons (n) and holes (p).

Going back to the description of the solar cell, we consider the recombination and generation of carriers beyond the radiative conduction-to-valence recombination model [17], the so-called Shockley-Read-Hall (SRH) recombination model. Figure 2.3 shows the original picture of Shockley and Read which displays the four processes encompassed by their analysis. It is based on the presence of a state inside the bandgap that acts as a trap for electrons and holes. Before going further and for the sake of clarity, we may decompose the generation term between the equilibrium term $G_e$ and the photogenerated term $G_h$. We can therefore still define a net recombination rate in Eq. (2.3) and the illumination gives rise to an independent current $I_h$ which depends only on the illumination conditions. Equation (2.6) then becomes

$$I = I_h + \int d\mathbf{r}^3 q\,(U_{n,p})\,. \tag{2.8}$$

The voltage dependent term is still to be evaluated and involves only net recombination rates $U_{n,p}$. Shockley and Read derived two simplified expressions for their recombination rate: one assuming arbitrary charge modulation

$\delta n$ or $\delta p$ and small number of traps, another assuming low injection and an arbitrary number of traps [17]. We will consider the first simplified model as solar cells work under high injection conditions and one tries to limit the concentration of traps as much as possible when building the device. The net recombination rate $U$ reads

$$U = \frac{(p_0 + \delta p)(n_0 + \delta n) - n_1 p_1}{\tau_{p0}(p_0 + \delta p + p_1) + \tau_{n0}(n_0 + \delta n + n_1)}\,, \tag{2.9}$$

where $n_0$ ($p_0$) is the equilibrium electron (hole) concentration, $\delta n$ ($\delta p$) is the excess of electrons (holes) and $\tau_{n0}$ ($\tau_{p0}$) is the effective life time for electrons (holes) that can be expressed in term of the concentration of recombination centers N, its capture cross section $\sigma_n$ ($\sigma_p$) and the thermal velocity of electrons and holes $v_{th}$

$$\tau_{n0,p0} = \frac{1}{N\sigma_{n,p}v_{th}}\,. \tag{2.10}$$

The two last terms $n_1$ and $p_1$ to be defined are the equilibrium concentrations of electrons and holes when the Fermi level $E_F$ coincides with the energy of the trap $E_t$. Considering a Boltzmann distribution for the carrier, they read

$$n_1 = N_c e^{\frac{E_t - E_c}{k_B T}}\,, \tag{2.11}$$

$$p_1 = N_v e^{\frac{E_v - E_t}{k_B T}}\,, \tag{2.12}$$

where $N_c$ and $N_v$ are the densities of state at the valence (conduction) edge obtained within an effective mass approximation. If one recasts expression (2.9) in order to obtain the effective lifetime such as in equation (2.7), then the lifetime of carriers is dependent on the position of the recombination center in the bandgap through the terms $n_1$ and $p_1$. If, for example, one considers a shallow acceptor, then the recombination process will be more or less equivalent to a conduction-to-valence recombination. However, for midgap recombination center, the lifetime is significantly reduced and as a consequence also the current provided by the solar cell. The recombination rate is also proportional to the carrier concentration. Illumination might therefore modify the recombination and, in some cases, saturate some limiting recombination processes and improve the efficiency of the solar cell. In practice, devices achieving light concentration are coupled to the solar cell to increase the flux of photons and therefore the carrier concentration.

Another important assumption about the Schokley diode equation is the presence of recombination only in the quasi-neutral regions. However, let us consider the case where recombination occurs in the space charge region through a midgap recombination center. For the sake of simplicity, we consider that the Fermi level is constant along the device (*i.e.* thermodynamic

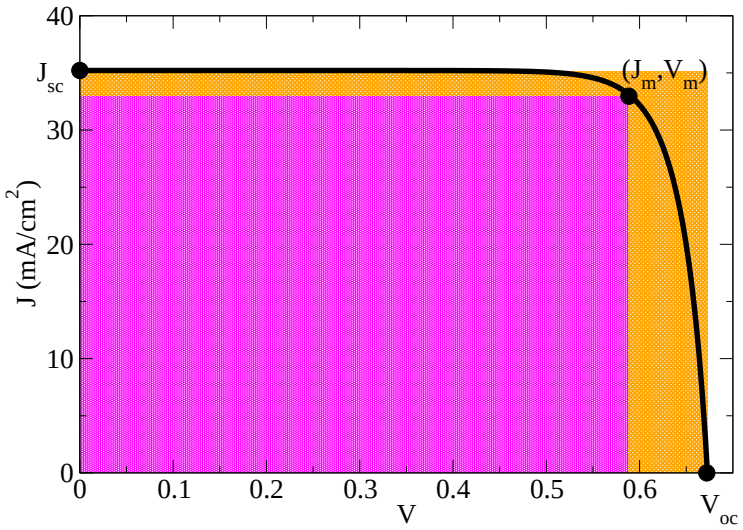


**Figure 2.4:** I-V characteristic of a CIGS record cell [18]. Magenta area is equal to the maximum power point ($J_m$,$V_m$) while orange area correspond to P=$V_{oc} \times J_{sc}$

equilibrium) while a built-in potential bends the bands in the space charge region. Consequently, in the space charge regions, the Fermi level can get closer to the energy level of the recombination center. Then, two recombination processes may occur: conduction-to-midgap and midgap-to-valence recombination. The corresponding Schockley diode equation reads hence

$$I = I_0(e^{\frac{qV}{2k_BT}} - 1)\,, \tag{2.13}$$

where $I_0$ is the saturation current. The main difference between equation (2.1) and equation (2.13) is the presence in the denominator of the exponential of a factor 2. Such factor is the specific to the case of a midgap recombination center. One can generalize the diode equation

$$I = I_0(e^{\frac{qV}{nk_BT}} - 1)\,, \tag{2.14}$$

where $n$ is called the diode ideality factor. The ideality factor can be interpreted as the fingerprints of the recombination process at work in the device: if one obtains through some fitting procedure an ideality factor close to 2, the most important recombination processes occur in the space charge region, while for values closer to 1, the recombination takes place in the quasi-neutral region. Usually, for CIGS-based solar cells, $n \approx 1.6/1.7$. Record cell displays however a much smaller $n \approx 1.3$ [18], which confirms the harmful character of recombination in the space charge regions.

The ideality factor is experimentally extracted through a non-linear fitting of the I-V curve. However, under certain circonstances like low temperature or low illumination, the form (2.14) does not have sufficient flexibility of fit the experimental curve. In those cases, the *two diode model* is preferred as it increases the number of degrees of freedom.

$$I = I_{01} e^{\frac{qV}{n_1 k_B T}} + I_{02} e^{\frac{qV}{n_2 k_B T}} . \tag{2.15}$$

In order to be complete, we should mention that the solar cell suffers from exactly the same problems as any other semiconductor device: parallel and series resistances. Series resistances $R_s$ is of particular interest in cases of high photogenerated current (for example in the situation of light concentration as the Joule effect scales as $R_s I^2$). On the other hand, shunt resistances $\mathrm{R_{shunt}}$ account for the leakage of current through the cell [2]. The I-V characteristic is then recast as

$$I = I_h - I_0 \left( e^{\frac{q(V+R_s \times I)}{n k_B T}} - 1 \right) - \frac{V + R_s I}{R_{\text{shunt}}} . \tag{2.16}$$

Figure 2.4 displays the I-V characteristic with the parameter $I_h, I_0, n, R_s$ and $R_{\text{shunt}}$ established by a fitting procedure for a state-of-the-art ZnO/CdS/CIGS solar cell [18].

**In conclusion**, we presented the equation governing the voltage-current characteristic of solar cell. There is a constant tradeoff between the photogenerated current and the saturation current that depends on the load to which the solar cell is connected. Technically, the I-V characteristics are an extremely rich source of information: one can make the connection between microscopic phenomena such as recombination processes and macroscopic quantities such as the voltage or the current. In the next section, we will look at these macroscopic quantities in detail.

## 2.3 I-V characteristics

### Open-circuit voltage

The open circuit voltage is the voltage corresponding to no current flow in the device. From Eq. (2.16) it is expressed as

$$V_{oc} = \frac{n k_B T}{q} \ln \left( \frac{J_h}{J_0} + 1 \right) \approx \frac{n k_B T}{q} \ln \left( \frac{J_h}{J_0} \right) . \tag{2.17}$$

[2] One could also notice the analogy between the shunt resistance and the midgap level as both can be viewed as parallel path for current.

It is interesting to express the saturation current of the diode depending on the dominant recombination process.

- **Recombination in the quasi-neutral regions** then $n = 1$

$$V_{\text{oc}} = \frac{E_g}{q} - \frac{k_B T}{q} \ln\left(\frac{q D_n N_c N_v}{J_h N_a L_n}\right) . \tag{2.18}$$

  The open circuit voltage is therefore always bounded by the bandgap of the absorber. Besides, it is impacted by bulk recombination losses in the neutral region depending on the concentration of dopant $N_a$, the diffusion length of electrons $L_n = \sqrt{D_n \tau_n}$ with $D_p$ the diffusion coefficient of holes, and $\tau_p$ the lifetime of holes calculated with the previously presented SHR theory. The lower the dopant concentration is , the smaller $V_{\text{oc}}$ becomes. This effect can however be counterbalanced by a longer lifetime of the carrier. It could be tempting to increase the doping of the absorber. However, it might result in the appearance of other recombination processes such as Auger recombination [13] or tunneling enhanced recombination [19, 20].

- **Recombination in the space charge regions** $n = 2$, thereby loss will affect twice more $V_{\text{oc}}$ than loss in the quasi-neutral regions. Sah *et al* [21] expressed the recombination current arising from the total space charge region by

$$J_r \approx \frac{k_B T \pi}{2 \tau_n E} \sqrt{N_c N_v} \exp\left(\frac{qV}{2 k_B T}\right) , \tag{2.19}$$

  where $E$ is the electric field at the position of maximum recombination in the space charge region. While, this expression is only valid under certains conditions (*ie..*, hole and electron lifetimes and mobilities should be equal and the recombination center energy is situated at the Fermi level), it is helpful to understand the phenomena at work in the space charge region. Indeed, there are two competing phenomena in the space charge region: the electric field creates a rapid drift of the carriers to the edge of the space charge region while recombination centers may trap these carriers. The characteristic length of the first process is given by $k_B T/qE$ and appear in the expression of the recombination current in order to replace the diffusion length $L_n$. The open circuit current then reads

$$V_{\text{oc}} = \frac{E_g}{q} - \frac{2 k_B T}{q} \ln\left(\frac{k_B T D_n \pi \sqrt{N_c N_v}}{2 J_h E L_n^2}\right) , \tag{2.20}$$

The dependence of expression (2.20) and expression (2.18) on the concentration of dopants $N_a$ is the same and therefore all the considerations about modulating the doping in order to lower the loss due to recombination still hold in the case of recombination in the space charge region.

- **Recombination at the interface CdS/CIGS interface**

  One of the main sources of loss is the recombination occuring at interfaces. Usually, heterojunctions are the seat of important lattice distortions that can create defects or various traps for carriers. It is legitimate to investigate the effect of surface recombination on the open-circuit voltage

$$V_{\mathrm{oc}} = \frac{\Phi_p}{q} - \frac{k_B T}{q} \ln\left(\frac{q S_p N_v}{J_h}\right) , \tag{2.21}$$

  where $S_p$ is the interface recombination velocity for holes and $\Phi_p$ is the hole barrier formed by the band offset between CIGS and CdS. The use of a very thin layer of CdS and highly doped ZnO permits to increase the band bending and reduce the recombination velocity of the holes at the interface. Therefore, the main loss mechanism in high-efficiency CIGS solar cell is believed to originate from bulk recombination in the absorber.

In addition to the three recombination processes mentioned, other processes might impact $V_{oc}$ like the recombination at the back contact or due to grain boundaries [22]. The typical values of $V_{oc}$ for the chalcopyrite family ranges between 515 meV for $CuInSe_2$ to 861 meV for $CuGaSe_2$. The bandgap hierarchy for the chalcopyrite structure is preserved: indeed $CuGaSe_2$ has the largest bandgap and $CuInSe_2$ the smallest for this family of material.

### Short circuit current $I_{sc}$

In the case of high quality solar cells (where $I_0$ and $R_s$ are low and $R_{shunt}$ high), the short circuit current is approximatively equal to the photogenerated current. Therefore it gives a good indication on the electron-hole pair generation, and how the charges are separated and collected. Instinctively, we can forsee that large bandgap materials will absorb less photons than small bandgap materials resulting in smaller short circuit currents. For $CuInSe_2$, a typical value of $J_{sc}$ is 41 $mA/cm^2$ while it is reduced to 14.2 $mA/cm^2$ for $CuGaSe_2$.

### Maximum Power Point

The maximum power point corresponds to the operation point (V, I) that maximizes the electrical power $P = V \times I$. It can be represented graphically by the largest rectangle that can be encompassed under the I-V characteristics. This power varies with the irradiation and therefore, under operation, the load to which the solar cell is connected should be adjusted instaneously in order to obtain maximum power transfer.

### Efficiency

Efficiency is one of the most popular indicators of solar cell performance. Indeed it directly asserts the macroscopic energy conversion. Energy efficiency is a common quantity in describing thermodynamical system: the energy efficiency of the Carnot cycle is simply

$$\eta_{\mathrm{C}} = 1 - \frac{T_{\mathrm{c}}}{T_{\mathrm{h}}}\,, \tag{2.22}$$

where $T_{\mathrm{c}}$ is the temperature of the cold reservoir and $T_{\mathrm{h}}$ is the temperature of the hot reservoir. In the case of a solar cell, the hot reservoir is simply the sun whose spectra can be described by a black body with $T_{\mathrm{h}} = 6000\mathrm{K}$ and the solar cell is the cold reservoir at room temperature $T_{\mathrm{c}} = 300\mathrm{K}$. Then, the Carnot efficiency limits the energy conversion to 95%. However, the entropy generation associated with input radiative flux and output work should also be taken into account: the subsequent efficiency is called the Landsberg limit [23] $\eta_L = 93.3\%$. It is considered as the upper bound for solar energy conversion efficiency. In 1961, Shockley and Queisser present a derivation of the efficiency of a solar cell based on the detailed balance principle [3] .In this context, detail balance means that the rate of radiative recombination at thermal equilibrium for a given frequency interval is equal to the corresponding rate of generation of electron hole pairs at thermal equilibrium. The maximum efficiency is obtained when the recombination processes considered are entirely radiative ($\eta = 31\%$) for a bandgap $\mathrm{E}_g \approx 1.4$ eV. The Shockley-Queisser limit is considered as the most realistic limit for solar cells based on a p-n junction. The highest efficient recorded for a single junction solar cell is 24.5% [24], very close to the theoretical limit. The analysis of Shockley *et al.* is however plagued by several rough approximations:

- all the light above the bandgap is absorbed in the active region.
- one photon creates a pair of electron-holes which are all extracted if they do not recombine.

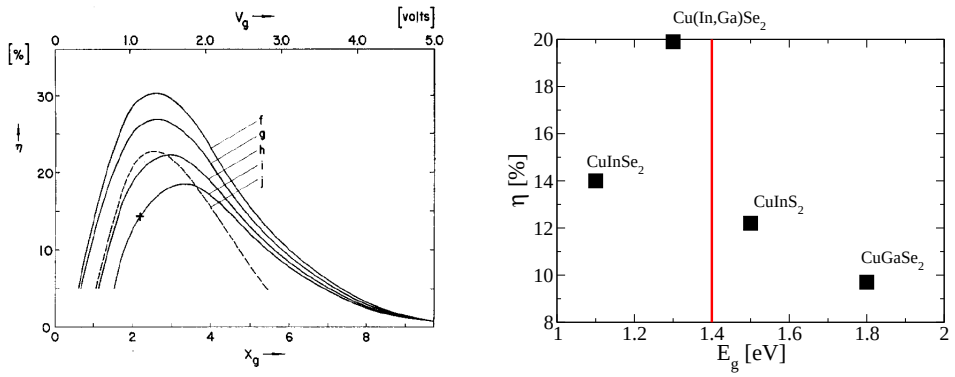


**Figure 2.5:** Left panel: Efficiency $\eta$ of solar cell versus the bandgap $E_g$ [3]. Different curves $f, g, h, i, j$ refers to variation of the irradiation solid angles, the ratio of non-radiative/radiative recombination and the fraction of reflected radiation. Numerical details are given in Ref. [3]. Right panel: Efficiency of the record CIGS cell with different absorbers [26] versus the bandgap of the absorber. The red line refers to the optimal value for the bandgap $E_g \approx 1.4$ eV.

- recombination occurs only in the quasi-neutral region.

The full account of all losses in the efficiency is a rather difficult task [25]. In practice, the efficiency is expressed as the ratio between the maximum power point MPP and the irradiation power received by the device $W \times A$ where W is the light irradiance and A the surface of the solar cell

$$\eta = \frac{P_m}{W \times A} \,. \tag{2.23}$$

Figure 2.5 shows the record efficiency achieved for different type of Cu-based chalcopyrite absorbers versus the bandgap value. One can see that the family encloses the optimal bandgap and that the highest efficiency is achieved for the material with the bandgap closest to the optimal value. The general trend of the curve follows the theoretical prediction displayed in the left panel of Fig. 2.5 *i.e.* the assymmetry around the maximum efficiency point, which makes the material with a smaller bandgap than the *optimal* always better than material above the optimal value of $E_g$. Chalcopyrite family exhibits the higher efficiency for the thin film family 19.9 % for a solar cell made of $Cu(In,Ga)Se_2$.

### Fill Factor

The fill factor is another parameter that is defined as the ratio between the MPP and $V_{oc} \times I_{sc}$. Again the graphical analysis of this quantity is relatively simple as it is the ratio between the surface of the largest rectangle encompassed under the I-V characteristics and the surface of the smallest rectangle that encomposses the I-V curve.

$$FF = \frac{P_m}{V_{oc} \times I_{sc}} \,. \tag{2.24}$$

The fill factor is typically impacted by series resistance and shunt resistances. Typical fill factors in the Cu-based chalcopyrite family range between 68% [$CuGaSe_2$] and 77% [$Cu(In,Ga)Se_2$]

## 2.4 CuIn(Se,S)$_2$, the *perfect* photovoltaic material ?

$CuIn(S,Se)_2$ crystallizes in the chalcopyrite phase, with space group $I\bar{4}2d$ which forms an isoelectrical analog to the III-V binary semiconductor phase, zinc blende. Experimental observations of phases CuAu [27] or CuPt [28] have been reported in the past but can be considered marginal due to the scarcity of the experimental evidences in the literature. In the chalcopyrite phase, each anion is coordinated to two indium and two copper atoms while each cation is tetrahedrally coordinated to four anions. The unit cell contains two chemical formula, *i.e.* 8 atoms. The corresponding Wyckoff positions are Cu (0,0,0), In (0,0,1/2) and (S,Se) ($u$,1/4,1/8). The chalcopyrite structure is depicted in Fig. 2.6.

If we replace indium atoms by copper atoms, we recover the zinc blende phase. However there are significant differences between the chalcopyrite phase and the zinc-blende. The existence of two different cations (In and Cu) results in two different bonding lengths $R_{In-(S,Se)}$ and $R_{Cu-(S,Se)}$ distorting the crystallographic structure in two ways. First the cell is tetragonally distorted, *i.e.* $\eta \equiv c/2a \neq 1$. For In-based chalcopyrite, the tetragonal distortion is very small $\eta \approx 1.006$. Second, the ideal zinc-blende site of the anion is disturbed leading to a parameter $u$ called the anion displacement. If the two zinc-blende sublattices forming the chalcopyrite lattice were identical, $u = \frac{1}{4}$ whereas in the case of chalcopyrites

$$u = \frac{1}{4} + \frac{R^2_{Cu-(S,Se)} - R^2_{In-(S,Se)}}{a^2} \,, \tag{2.25}$$

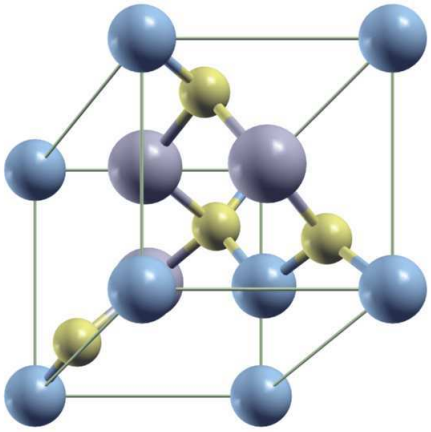

**Figure 2.6:** Atomic structure of $CuInS_2$. Blue middle-size spheres represent copper atoms, large grey spheres represent indium atoms and small yellowish spheres represent sulfur atoms.

where $a$ is the lattice parameter. The value of $u$ lies between 0.21 and 0.235 for CIS.

Now let us turn to the composition of CIS thin films: they present significant deviations from stoichiometry depending on the conditions of growth. It is not surprising as CIS materials display a very rich phase diagram. Owing to its ternary nature, many possible binary or ternary phases can form depending on temperature and the compositional ranges. Figure 2.7 shows the phase diagram of $CuInSe_2$ as a function of the content of Cu and the temperature. First, the pure chalcopyrite phase $\alpha$ is restricted to a rather small compositional range. Its domain of stability decreases significantly at low temperature. Besides, it is bounded for high fraction of Cu by the segregation of $Cu_{2-y}Se$ phase and for lower fraction of Cu by the appearance of secondary phase $\beta$ and $\delta$. The $\beta$-phase is an order defect phase corresponding to $CuIn_3Se_5$ while $\delta$ is the high temperature sphalerite phase. More surprising is the fact that the chalcopyrite phase is not stable for perfect stoichiometry. Usually, PV grade CIGS shows a significant Cu-poor character and therefore lies within the stability domain of the $\alpha$ phase. All such deviations from the perfect stoichiometry can therefore be viewed as the fingerprint of the presence of intrinsic defects. Understanding the physics of the defects in CIS allows us to better understand the phase diagram and also the mechanism

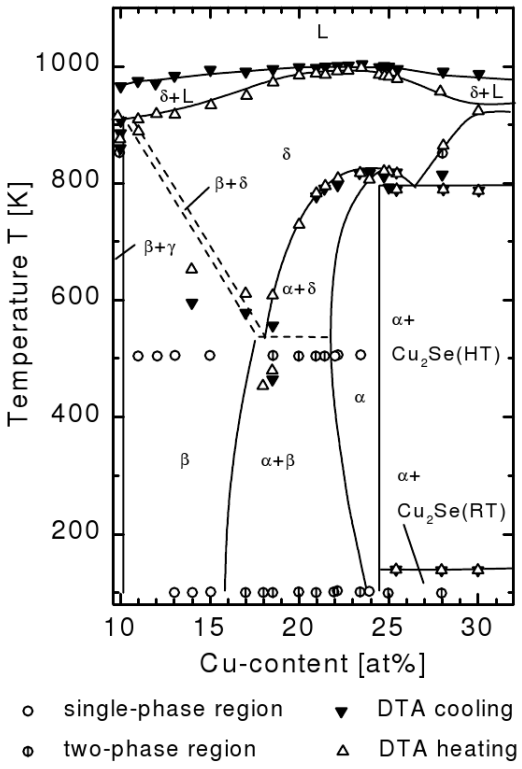


**Figure 2.7:** Phase diagram of $CuInSe_2$ along the quasi-binary cut $In_2Se_3$-$Cu_2Se$ in the range of 10-32 % of Cu and in the temperature range 100-1000° C [29]

of doping at work in CIS.

Zhang *et al.* [2] published in 1998 a seminal work on the physics of defects in $CuInSe_2$ using *ab initio* methods. They drew five important conclusions:

- Defect formation energy in CIS depends strongly on the growth condition through the chemical potential of the atomic species
- The copper vacancy $V_{Cu}$ has an extremely low formation energy.
- $V_{Cu}$ forms shallow acceptor level (see Fig. 2.8) which explains the self-doping of CIS.
- Neutral defect pairs with charge transfer involving two copper vacancies $V_{Cu}^{-}$ and one indium copper antisite $In_{Cu}^{2+}$ are abundant, and electrically benign due to the neutralization process of the midgap $In_{Cu}^{2+}$ states by two $V_{Cu}^{-}$ (see Fig. 2.8).

The first conclusion is a rather general consideration of the *ab initio* calculation of the defect formation energies. However it allows us to analyse the benefit of the intrinsic ZnO layer on the device performance: $V_{Cu}$ and $In_{Cu}$ are believed to stabilize the polar surface (112) of CIS and CIS grain boundaries. It is experimentally observed that CIS surfaces usually present a strong Cu depletion that can be associated with doping inversion: while the bulk is p-type, the surface turns out to be n-doped. The fact that the formation energies of $V_{Cu}$ and $In_{Cu}$ are $E_f$-dependent allows one to design the band structure in order to control the position of the Fermi level at the surface. For instance, by introducing a layer of intrinsic ZnO, the Fermi level is pushed away from the conduction band. Consequently, the $V_{Cu}$ formation energy significantly increases and counterbalances the type inversion that occurs at the heterojunctions [30].
The two previous conclusions agree well with experimental observations. The p-type character of CIS under Cu-poor growth condition originates from $V_{Cu}$, seen experimentally as a depletion of Cu atoms. The existence of an ordered defect compound (ODC) phase $\beta$ results from the repetition of the neutral complex defect $2V_{Cu}^{-} + In_{Cu}^{2+}$[3]. The two latter defects explain the usual depletion of Cu atom experimentally observed in thin films. In thin-films, the $\beta$ phase is rarely observed due to: (i) diffusion of Na from the glass substrate into CIS [31] and replacement of indium by gallium preventing the formation of such phase, (ii) high growth temperature extends the stability domain of the $\alpha$ phase as seen in Fig. 2.7, (iii) nucleation of the $\beta$ phase is forbidden by the important interfacial energy originating from electrostatic [32] or mismatch effects [33]. Following this study on defects involving Cu and In, Lany *et al.* [34, 35] have introduced two defects that can explain the experimentally observed metastable electrical behavior in bulk CIGS [36, 37]. While $V_{Se}$ can explain the *p*-type persistant photoconductivity, $V_{Se}$-$V_{Cu}$ may account for the appearance of a deep hole trap that appears at 260 meV above the valence band [36].

The discovery of the morphology enhancement action of Na is one of the main reasons for the second technological step observed in the evolution of the record efficiency. However, the mechanism behind this improvement is still not known exactly and may proceed by multiple effects, two of the many being the effect of Na on thin film growth and the incorporation of Na into the lattice. During growth, incorporation of Na leads to the formation of $NaSe_x$ that can mediate the inclusion of Se into the film, preventing formation of the $V_{Se}$ donor [38]. Moreover, Na can delay the growth of $CuInSe_2$ at

[3] ODC usually displays a *n*-type doping character which would make the use of an *p*-type TCO layer necessary. Therefore, p-type absorber are preferred.

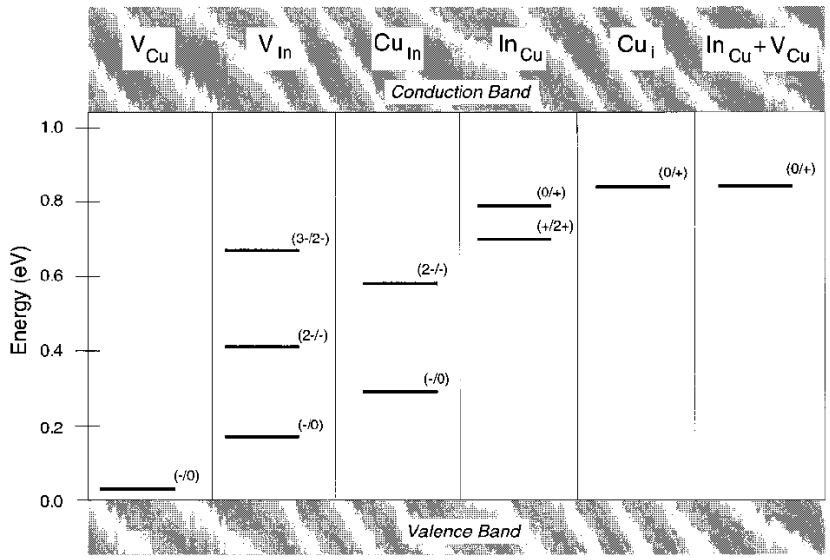


**Figure 2.8:** Defect transition energy for $CuInSe_2$ calculated by means of *ab initio* methods. [2]

temperature below 380°, requiring the growth temperature to be increased and consequently increasing the stability domain of the $\alpha$ phase. The incorporation of Na is however believed to occur at the film surface and grain boundary [39]. Another interpretation has been proposed as Na-Se bonds were observed experimentally [40]: Na may replace In or Ga so that $Na_{In,Ga}$ is formed and acts as a shallow acceptor, improving the *p*-type character of CIS. Replacement of Cu by Na has been proposed and may prevent the formation of the deep defect $In_{Cu}$ [41]. Both hypothesis can be valid, but the large amount of Na required to grow device quality CIGS is of the order of 0.1 % and such amounts of Na is not detected in the sample [42]. Thus, one may favour the assumption that the action of Na is more predominant during film growth than by an actual incorporation into the lattice.
The second hypothesis about the predominance of the $\alpha$ phase was the beneficial effect of Ga. Similarly to Na, Ga addition acts on the growth by slowing down the process [38] and also on the film itself due to its incorporation into the lattice. However, there is absolutely no doubt about the presence of Ga in the lattice. Wei *et al.* [43] explained the improvement of the film properties by the slightly different defect physics between $CuGaSe_2$ and $CuInSe_2$: $Ga_{Cu}$ formation energy is higher than $In_{Cu}$ so this compensating donor will be much less likely to occur in the film. Moreover, the complex defect $2V_{Cu}^{-}+In_{Cu}^{2+}$ is less likely in $CuGaSe_2$ than its counterpart in $CuInSe_2$. Therefore, the $\beta$

phase is less probable in $CuGaSe_2$ and addition of Ga in $CuInSe_2$ makes the domain of stability $\beta$ phase shrink. It was noticed empirically that there exists an optimum of Ga-content around Ga/Ga+In $\approx$ 0.2-0.3 that makes the solar cells the most efficient [44]. Record cells [45] together with commercial products [46] use such a ratio for their $Cu(In,Ga)Se_2$ absorbers. Other factors are known to impact favorably the growth of CIGS or the properties of the film like addition of sulfur or oxygen [13].

Grain boundaries play an intriguing role in the performance of CIGS solar cells. Indeed, while Si- and GaAs-based polycrystalline solar cells suffer from large recombination at grain boundaries [47, 48], CIS polycrystalline solar cells outperform significantly their single-crystalline counterparts: today, record efficiency of 20% [45] is achieved for polycrystalline to be compared with the lower efficiency of the record single-crystal solar cell 13% [49, 50]. Usually, the grain size in CIS is rather small and does not exceed the thickness of thin films *i.e.* 1-2$\mu$m. As a comparison, the grain size in polycrystalline Si can be larger than 5 mm. Once a grain boundary is formed, states may appear in the bandgap and they can act as recombination centers. Such states result in a charge that is cancelled by the formation of a depletion layer around the grain boundary. Such space charge region induces a band bending of the valence and conduction band: consequently, the grain is thought to act as a hole barrier whose barrier height is defined as $\Phi_b$.[4]. Grain boundaries in CIS are believed to originate from the polar (112) free surface passivated with $V_{Cu}$, $In_{Cu}$ and $2V_{Cu}^{-}$+$In_{Cu}^{2+}$ defects. As in that case the *p-d* repulsion that governs the valence band maximum behavior is relaxed, removing Cu atoms will result in the downshift of the valence band maximum: such internal valence band offset $\Delta E_v$ is often called the neutral barrier as it does not depend on the space charge region. The total barrier is then $\Phi_b^* = \Phi_b + \Delta E_v$ and is typically of the order of 0.2-0.3 eV. Under conditions of illumination, the majority carriers might get trapped in the grain and therefore diminish the grain charge and consequently $\Phi_b$. Therefore, $\Delta E_v$ must be sufficient to prevent majority carriers from reaching the recombination center at the grain boundaries: $\Delta E_v > 0.2$ eV is believed to be enough to eclipse such recombination process. Grain boundaries improve the performance of the polycrystalline CIGS solar cell by extending the space charge region of the $p-n$ junction [52]. They can also serve as a guide for minority carrier current through the junction. Such view can only be possible if the grain boundaries

[4] It has to be noticed that the real action of grain boundaries is not entirely understood at an experimental and theoretical level [51]. Therefore, the following discussion will only be made at the light of the recent reports found in the literature.

are orientated perpendicularly to the p-n junction. When grain boundaries are oriented parallel to the heterojunction, space charge region results in a *dead layer* that prevents the current flow [51].

**In conclusion**, CIGS is a rather versatile material. Experimental evidences necessary to validate theoretical predictions are extremely difficult to obtain due to the intricated nature of the optical and electronic properties. Moreover, defects seem to play an important role but their theoretical study require the use of large supercells. Therefore, up to now, all *ab initio* calculations performed in this material have used density functional theory. In the next section, we will review DFT, its advantages and deficiencies and we will particularly show that it is not suitable for prediction of the bandgap.

# Chapter 3

# The Many Body problem and Density Functional theory

Understanding the behavior of interacting electrons in a solid is still a tremendous task and can only be adressed partially. This is not due to purely theoretical reasons but also to numerical reasons. Computers have became more and more powerful allowing physicists to solve more and more complex problems. However, the interacting-electron problem has been known for more than 80 years and despite the advent of computers and supercomputers, this particular task is still out of reach. Consider the Hamiltonian for a system of electrons with mass $\mathrm{m}_e$ and nuclei with mass $M_I$

$$\mathcal{H} = \sum_i \frac{-\hbar^2}{2m_e}\nabla_i^2 + \sum_I \frac{-\hbar^2}{2M_I}\nabla_I^2 + \frac{1}{2}\sum_{i\neq j}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,I}\frac{Z_I e^2}{|\mathbf{R}_I - \mathbf{r}_i|} + \frac{1}{2}\sum_{I\neq J}\frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|} \quad (3.1)$$

$$= T_e + T_N + U_{e-e} + U_{e-N} + U_{N-N}\,, \quad (3.2)$$

where indices $i$ and $j$ run over electrons and $I$ and $J$ over nuclei. The first two terms represent the kinetic energy of electrons ($\mathrm{T}_e$) and nuclei ($\mathrm{T}_N$) respectively while the three last terms represent the Coulomb interaction between, respectively, electron-electron ($\mathrm{U}_{e-e}$), electron-nucleus ($\mathrm{U}_{e-N}$), and nucleus-nucleus ($\mathrm{U}_{N-N}$). The time-independent non-relativistic Schrödinger equation (SE) then reads

$$\mathcal{H}\Psi = E\Psi\,, \quad (3.3)$$

where $\Psi$ is the many-body wave function and $E$ is the total energy. The time independent expression of any observable $\mathcal{O}$ is given by its expectation

value $\langle \mathcal{O} \rangle$ such as

$$\langle \mathcal{O} \rangle = \frac{\langle \Psi | \mathcal{O} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \,. \tag{3.4}$$

For example, the expectation value of the Hamiltonian $\mathcal{H}$ is the total energy $E$. The solution of Eq. (3.3) $\Psi$ is a stationary point of the total energy. In particular, the state with the lowest energy $E_0$ is called the ground state $\Psi_0$. It can be determined by minimizing the total energy expression. If one were able to solve Eq. (3.3), one would have to deal with very large quantities as $\Psi$ depends on the variables $\{\mathbf{r}_1..\mathbf{r}_N, \mathbf{R}_1...\mathbf{R}_M\}$ where $N$ and $M$ represent respectively, the number of electrons and the number of nuclei in the system. One can notice immediately that the resolution of the many-body equation depends on the internal degrees of freedom of a system of $N$ electrons and a system of $M$ nuclei. Values of $N$ and $M$ in a solid are of the order of the Avogadro's number, that is $10^{23}$. Solving the many body Schrödinger equation appears to be impossible without any further approximation due to the enormous amount of information contained in $\Psi$. Several approximations have been developed to simplify the many-body problem and eventually open a path to the numerical solution of the Schrödinger equation.

## 3.1 Born-Oppenheimer approximation

By noticing that the masses of nuclei are about 3-4 orders of magnitude larger than those of the electrons, one can decouple the slow dynamics of nuclei from the fast one of electrons. This is the so-called Born Oppenheimer (BO) approximation or adiabatic approximation. It is an extremely important approximation as it allows us to simplify significantly the many-body problem. Let us consider the following separation of the eigenfunction of $\mathcal{H}$

$$\Psi = \phi(\mathbf{r}_i : \mathbf{R}_I)\chi(\mathbf{R}_I) \,, \tag{3.5}$$

where $\phi$ satisfies the Schrödinger-like equation in a static lattice

$$(T_e + U_{e-e} + U_{e-N})\phi(\mathbf{r}_i : \mathbf{R}_I) = \mathcal{E}_e(\mathbf{R}_I)\phi(\mathbf{r}_i : \mathbf{R}_I) \,. \tag{3.6}$$

By applying the operator $\mathcal{H}$ to $\Psi$

$$\begin{aligned} \mathcal{H}\Psi = \phi(\mathbf{r}_i : \mathbf{R}_I)\{T_N + U_{N-N} + \mathcal{E}_e(\mathbf{R}_I)\}\chi(\mathbf{R}_I) \\ - \sum_I \frac{\hbar^2}{2M_I} \left\{ 2\nabla_I \chi \cdot \nabla_I \phi + \chi \nabla_I^2 \phi \right\} \,, \end{aligned} \tag{3.7}$$

we can make $\chi(\mathbf{R}_I)$ statisfy the following equation:

$$\{T_N + \mathcal{E}_e + U_{N-N}\}\chi = \mathcal{E}\chi \,, \tag{3.8}$$

where the total energy of the electron system $\mathcal{E}_e$ is added to the lattice energy as a *potential*. This is the so-called adiabatic contribution of the electrons to the lattice motion. The second line of Eq. (3.7) is called the non adiabatic terms. We can ask ourselves how much the diagonal non-adiabatic terms contribute to the total energy. The diagonal matrix element of the first non adiabatic term vanishes due to the normalization of the eigenfunction. Following Ziman [53], the second term can be roughly estimated to proportional to $\mathrm{m}_e/\mathrm{M}_I$ which is of the order of $10^{-3}$ or $10^{-5}$. Note that we only consider the diagonal element but we should keep in mind that off-diagonal matrix elements of the first non-adiabatic term give rise to the electron-phonon interaction , *i.e.*, energy is transfered from the lattice to the electrons resulting in an electronic excitation [54].

The Born-Oppenheimer approximation [55] consists in neglecting the second line of Eq. (3.7). Within this approximation, all the electrons remain in a given electronic state $\phi_i$ when the atoms move. Moreover, no energy transfer occurs between the lattice and the electron. The decoupling nature of the BO approximation permits to treat separately the nuclei and the electrons. Therefore, as we are interested mainly in phenomena involving electrons only, we can simplify the many-body problem to an electronic problem only. All the first principles calculations that will follow in this thesis have been made using the BO approximation *i.e.* positions of atoms have been kept fixed throughout the calculation. Nevertheless, atomic positions have to be known in order to define our system: either we have access to them from experimental data or we can determine them using a very general formalism called the *force theorem*.

## 3.2 Hellmann-Feynman theorem

Feynman proposed in 1939 to use the more *classical* concept of force in order to determine the structure of molecules. He derived the force theorem[56] that allows one to express forces in molecules using concepts of classical electrostatics once the electronic density is obtained by solving the Schrödinger equation. Then the forces acting on nuclei $i$ can be written as

$$F_i = -\frac{\partial E}{\partial \mathbf{R}_i}\,. \tag{3.9}$$

The total energy is expressed as the expectation value of the Hamiltonian $E = \langle\Psi|\mathcal{H}|\Psi\rangle$ where $\Psi$ is the eigenfunction of the Hamiltonian.

$$F_i = -\langle\Psi|\frac{\partial \mathcal{H}}{\partial \mathbf{R}_i}|\Psi\rangle - \langle\frac{\partial \Psi}{\partial \mathbf{R}_i}|\mathcal{H}|\Psi\rangle - \langle\Psi|\mathcal{H}|\frac{\partial \Psi}{\partial \mathbf{R}_i}\rangle\,. \tag{3.10}$$

The two last terms of expression (3.10) vanish because of the stationarity of the total energy with respect to variations of the wavefunctions. One can notice that the only terms of $\mathcal{H}$ dependent of $\mathbf{R}_i$ are $U_{e-I}$ and $U_{I-I}$. They do not dependent directly on the electron-electron interaction $U_{e-e}$ which is the most difficult term to model. Thereby, this theorem allows one to calculate the forces even though the internal interactions change when nuclei are moving. For the case of a local potential $v_{ext}(\mathbf{r})$, the force is expressed as

$$F_i = -\int d^3\mathbf{r} n(\mathbf{r}) \frac{\partial v_{ext}(\mathbf{r})}{\partial \mathbf{R}_I} - \frac{\partial E_{II}}{\partial \mathbf{R}_I} . \tag{3.11}$$

The force depends only the electronic density. For the case of non-local potentials, the above expression is not valid anymore and one should go back to Eq. (3.10) in order to derive the relation for the forces. Moreover, an extra component of the force should be added. We will go back to that point in section 3.8.

## 3.3 Density Functional Theory

One of the possible paths to solve the many body problem is based on the assumption that all the necessary information contained in the many-body wavefunction $\Psi_0$ in order to describe the ground state properties can be cast into a functional of the ground state electronic density $n_0(\mathbf{r})$ defined as

$$n_0(\mathbf{r}) = N \int d\mathbf{r}_2 ... d\mathbf{r}_N |\Psi_0(\mathbf{r}, \mathbf{r}_2 .. \mathbf{r}_N)|^2 . \tag{3.12}$$

The use of the *integrated* variable $n_0$ decreases considerably the workload and could permit to treat realistic system. It was introduced by Thomas [57] and Fermi [58]: while they neglected the electron-electron interaction beyond the Hartree potential, they approximated the kinetic energy as an explicit functional of the density. The kinetic energy of the system was calculated considering the kinetic energy of a non interacting homogeneous electron gas with density $n = n_0(\mathbf{r})$. However, due to the lack of electron-electron interactions, the Thomas-Fermi method failed to reproduce the physics of simple systems [59].
In 1964, Hohenberg and Kohn demonstrated formally that all properties of the system are unique functionals of the ground state electronic density [60]. They introduced the two following theorems:

- there is one to one mapping up to a constant between the space of the external potential $V_{ext}$ and the ground state electronic density $n_0$.

- there is a *universal* functional for the energy E[n], which is valid for any external potential $V_{ext}$. For a given $V_{ext}$, the ground state density $n_0(\mathbf{r})$ is the density that minimizes the functional E[n].

Then it follows that the total energy functional is expressed as

$$E[n] = T[n] + \int d\mathbf{r} n(\mathbf{r}) V_{ext}(\mathbf{r}) + V_{e-e}[n] \,, \tag{3.13}$$

where $T[n]$ represents the internal kinetic energy and $V_{e-e}[n]$ the internal potential energy *i.e.* representing the electron-electron interaction $\langle U_{e-e} \rangle$. One of the main difficulties is to express the kinetic term as a functional of the density. One can show by means of the virial theorem that the kinetic term is of the order of magnitude of the total energy. Therefore, particular attention should be paid when evaluating the kinetic term.
While Hohenberg and Kohn proved the existence of such functionals, no analytical functional is known for more than one electron. Kohn and Sham introduced a fictitious sytem that allows one to treat the non interacting kinetic energy contribution exactly and therefore to solve the many-body problem in practice (within some approximations).

## 3.4 Kohn-Sham system

In 1965, Kohn and Sham proposed a practical framework for the Density Functional Theory [61]: the fictitious Kohn-Sham system. The idea is simple: *replace the original many-body problem by an auxiliary independent particle problem.* The ground state density of the interacting system is required to be equal to the ground state density of the non interacting system. The system of non interacting particles has the usual kinetic operator and an effective potential $v_{\text{eff}}$ which is tuned such that the above requirement is fulfilled. We start from the Hohenberg and Kohn definition of the energy as a functional of the density (3.13)

$$E_{HK}[n] = T[n] + \int d\mathbf{r} v_{\text{ext}}(\mathbf{r}) n(\mathbf{r}) + V_{e-e}[n] \,, \tag{3.14}$$

where T[n] represents the internal kinetic energy and $E_{\text{int}}[n]$ the internal potential energy. We write the energy functional for the Kohn Sham system

$$E_{KS}[n] = T_s[n] + \int d\mathbf{r} v_{\text{ext}}(\mathbf{r}) n(\mathbf{r}) + E_{\text{Hartree}}[n] + E_{xc}[n] \,, \tag{3.15}$$

where $T_s[n]$ is the independent particle kinetic energy, $E_{\text{Hartree}}$ is the Hartree energy describing the self-interaction of the electronic density treated classically. $E_{xc}$ is defined as the functional of the density that ensures the condition $E_{KS}[n] = E_{HK}[n]$ and is expressed as:

$$E_{xc}[n] = T[n] - T_s[n] + V_{\text{e-e}}[n] - E_{\text{Hartree}}[n]\,. \tag{3.16}$$

It is the so-called exchange and correlation energy. Since the right hand side of Eq. (3.16) is formed by a sum of functionals of the density, $E_{xc}[n]$ is defined as a functional of the density. Unfortunately this functional is unknown.

In practice, from the Kohn-Sham energy functional, one can derive a set of equations called Kohn-Sham equations. As the HK theorem ensures that the KS functional is variational with respect to the electronic density, it follows the Euler equation

$$\frac{\delta}{\delta n}\left\{E_{KS}[n] - \mu\left(\sum_i n_i - N\right) - \sum_{i,j}\epsilon_{ij}\int d\mathbf{r}\left(\phi_i(\mathbf{r})\phi_j^*(\mathbf{r}) - \delta_{i,j}\right)\right\} = 0\,, \tag{3.17}$$

where $n_i$ are the occupation of wave functions $i$, $\mu$ is the Lagrange multiplier associated with the constraint $\int d\mathbf{r} n(\mathbf{r}) = N$ *i.e.* conservation of the number of particles and $\epsilon_{ij}$ are the Lagrange multipliers associated to the orthonormality constraint of the one-particle wave functions $\phi_i$. The left hand side of equation (3.17) can be decomposed further such as

$$\frac{\delta E_{KS}[n]}{\delta n(\mathbf{r})} = \frac{\delta T_s[n]}{\delta n(\mathbf{r})} + v_{\text{ext}}(\mathbf{r}) + \int d\mathbf{r}'\frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta E_{\text{xc}}}{\delta n(\mathbf{r})}[n]\,. \tag{3.18}$$

As the KS system is non interacting, Eq. (3.17) is solvable: the many body wave function is a Slater determinant of the single-particle wavefunctions veryfing the single-particle Schrödinger equation

$$\left(-\frac{\nabla^2}{2} + v_{KS}(\mathbf{r})\right)\phi_i(\mathbf{r}) = \epsilon_i\phi_i\,, \tag{3.19}$$

where $v_{\text{KS}}$ is a Kohn-Sham potential defined as

$$v_{KS}(\mathbf{r}) = v_{\text{ext}}(\mathbf{r}) + \int d\mathbf{r}'\frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta E_{xc}}{\delta n(\mathbf{r})}\,. \tag{3.20}$$

It follows from the fact that the system is non-interacting that the density is expressed as

$$n(\mathbf{r}) = \sum_{i=1}^{N}|\phi_i(\mathbf{r})|^2\,. \tag{3.21}$$

The exchange and correlation potential $v_{xc} = \delta E_{\rm xc}[n]/\delta n$ is an *a priori* complicated quantity. For the time being, we will consider it to be exact but it certainly needs approximations in order to apply the KS formalism in practice.

**In this section**, we recast the fully interacting particle problem into a non interacting particle problem. The mapping between the two systems was achieved by the introduction of an unknown quantity, $v_{\rm xc}$. We have obtained a practical scheme to solve the many-body problem though the Kohn Sham equations. In the following section, we will discuss the meaning of the quantities introduced by the Kohn Sham equations and more particularly the eigenenergies $\epsilon_i$.

## 3.5 Physicality of the Kohn-Sham eigenvalues

One of the most important quantities for photovoltaic application is the bandgap. *How do we define the bandgap in the context of the many-body problem ?*. Let us consider the experimental point of view first. The bandgap is defined experimentally as

$$E_g = I - A\,, \tag{3.22}$$

where $I$ is the first ionization energy and $A$ is the affinity energy. $I$ and $A$ are well defined within photomoemission and inverse photoemission experiments. In photoemission experiments, one photon of a given energy $\hbar\omega$ impinges the surface of the sample, travels from the surface to the bulk, the photon might be absorbed provoking the excitation of an electron. If the energy of the photon is sufficient *i.e.* larger than the vacuum level $E_{\rm vac} = E_f + \phi$ where $E_f$ is the Fermi level and $\phi$ the work function of the interface between the solid and the vacuum, the electron will be extracted from the sample with a kinetic energy $\rm E_{kin}$, leaving a hole in the system in a many-body state $i$ [1]. The conservation of energy yields

$$\hbar\omega + E(N,0) = (E_{\rm kin} + \phi) + E(N-1,i)\,, \tag{3.23}$$

[1] Photoemission is a surface sensitive technique. It can be explained by the concept of inelastic mean free path of the electron which is the distance that an electron covers before losing energy. Once the electron absorbs the energy of the photon, it shall then reach the surface before escaping the sample. The inelastic mean free path determines the depth of the sample that is actually probed. Moreover, the inelastic mean free path is energy dependent: generally, it presents a minimum at energy of the order of magnitude 100 eV and increases at high energy [62]. Therefore, hard X-Ray photoemission has recently encountered a great success in order to probe bulk-like properties.

where $E(N,0)$ is the ground state total energy of the system with N particles and $E(N-1,i)$ is the total energy of the system with N-1 particles in the excited state $i$. The minimum energy required to eject an electron from the sample is called the first ionization energy. Experimentally, it is expressed as

$$I = \min\left\{E_{\text{kin}} + \phi - \hbar\omega\right\}\,, \tag{3.24}$$

Reordering the terms of Eq. (3.23), expression (3.24) can also be written as

$$I = \min_i\left\{E(N-1,i) - E(N,0) = E(N-1,0) - E(N,0)\right\}\,. \tag{3.25}$$

Let us turn to the second constituent of the bandgap expression, the electronic affinity $A$. The latter is explained by considering the opposite process of the photoemission, the inverse photoemission. An electron with energy $\text{E}_{\text{kin}}$ is injected into the system with N electrons in its ground state. The electron has to overcome the work function of the surface and then might relax to one of the eigenstates of the system $i$, resulting in the emission of a photon with energy $\hbar\omega$. Invoking the conservation of energy one more time, the inverse photoemission process yields the following energy balance equation

$$E_{\text{kin}} + E(N,0) = \phi + \hbar\omega + E(N+1,i)\,. \tag{3.26}$$

Therefore, the electron affinity energy is given by

$$-A = \min\left\{E_{\text{kin}} - \phi - \hbar\omega\right\} = \min_i\left\{E(N+1,i) - E(N,0)\right\}\,, \tag{3.27}$$

and the bandgap $E_g$ is expressed as

$$E_g = I - A = E(N+1,0) + E(N-1,0) - 2E(N)\,. \tag{3.28}$$

The bandgap is defined as difference of ground state total energy for system with $N$, $N+1$ and $N-1$ electrons. One would like to cast this into a one electron picture. In the Hartree-Fock approximation [2], Koopmans' theorem [63] allows one to express the total energy difference through an eigenvalue difference. However, strict assumptions were made in order to obtain such results:

- electronic correlation: Hartree-Fock only contains exact exchange beyond Hartree approximation. Exchange is a purely quantum mechanical effects which originates from the Pauli exclusion principle and the indistinguishability of the particles. All the effects beyond exchange are called *Correlation*. Then, the many-body wave function is a single Slater determinant.

[2] See section 3 for more details about Hartree-Fock approximation.

- orbital relaxation: when removing one electron, no orbital relaxation has been performed. In practice, it means to work on the Slater determinant by removing rows or columns.

While the initial proof of the theorem was made for the first ionization, Koopmans' theorem is easily generalizable to the electron affinity or any unoccupied states following the same arguments. However due to the lack of electronic correlation and the omission of the orbital relaxation, HF estimates poorly the bandgap. In diamond, for instance, the experimental bandgap is 7.3 eV but HF predicts $E_g = 15$ eV [64]. One may think of using the same theorem considering KS equation as it incorporates correlation which is missing in the HF method. Unfortunately, Koopmans' theorem does not hold for the KS equations. The eigenvalues $\epsilon_i$ of (3.19) have been introduced in the derivation of the KS equations as Lagrange multipliers in order to ensure the condition of orthonormality of the single-electron wave functions. As a result, $\{\epsilon_i\}$ are just by-products of the resolution of the KS equations. There is a unique exception to the previous statement: the highest occupied state $\epsilon_N$ which is equal to the ionization energy of the system $\epsilon_N = -I$ [65]. In the context of finite systems, Levy *et al* suggested that as the KS electronic density is exact then considering the long range limit $r \to \infty$, $n_0(r) \to |\psi_{\mathrm{N}}|^2$, where $\psi_{\mathrm{N}}$ is the wave function of the highest occupied state. Intuitively, such state should be the most extended occupied state of the system, and therefore should form the long range part of the KS electronic density. Consequently, $\psi_{\mathrm{N}}$ is exact together with its eigenvalue $\epsilon_N$. Therefore, it is possible to express the affinity of a system of N electrons as the ionization of a system with N+1 electrons which is exactly $\epsilon^{KS}_{N+1}(N+1)$ and the expression for the bandgap becomes

$$E_g = \epsilon^{KS}_{N+1}(N+1) - \epsilon^{KS}_{N}(N)\,, \tag{3.29}$$

where the argument of $\epsilon^{KS}$ refers to the number of electrons of the system and the subscript to the index of the states. We have to deal now with two different systems: one with $N$ electrons and another one with $N+1$. However, we want to know the bandgap $E_g^{\mathrm{KS}}$ from a single calculation *i.e.* from the difference of the eigenvalues of the last occupied states and the first unoccupied states of the system with $N$ electrons. The bandgap expression then becomes

$$E_g = \left(\epsilon^{KS}_{N+1}(N) - \epsilon^{KS}_{N}(N)\right) + \left(\epsilon^{KS}_{N+1}(N+1) - \epsilon^{KS}_{N+1}(N)\right) = E_g^{\mathrm{KS}} + \Delta_{xc}\,, \tag{3.30}$$

where $\Delta_{xc} = \epsilon^{KS}_{N+1}(N+1) - \epsilon^{KS}_{N+1}(N)$ is the so called exchange and correlation discontinuity. One can notice that the KS bandgap will never equal the real bandgap unless $\Delta_{xc} = 0$. This is called the bandgap problem [66, 67]; it is the

reason why KS equations should not be used in order to evaluate bandgaps. Let us consider a solid where N is very large, the addition of an electron will then only induce infinitisemal changes of the electronic density. Therefore, $v_{\mathrm{xc}}$ must be highly non-analytical: infinitesimal changes in the density must induce important variations of the xc potential. $\Delta_{xc}$ may be considered as a measure of the degree of non-analytical behavior of the exchange and correlation potential. With the definition of the xc energy $E_{xc}$

$$E_{xc}[n] = \int d\mathbf{r} n(\mathbf{r}) \epsilon_{xc}([n], \mathbf{r}) \,, \tag{3.31}$$

where $\epsilon_{xc}$ is the exchange and correlation energy density, one obtains the xc potential

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}}{\delta n(\mathbf{r})} = \epsilon_{xc}([n], \mathbf{r}) + \int d\mathbf{r}' n(\mathbf{r}') \frac{\delta \epsilon_{xc}([n], \mathbf{r}')}{\delta n(\mathbf{r})} \,. \tag{3.32}$$

The second term of expression (3.32) ressembles a response function *i.e.* variations of the exchange and correlation energy density with respect to variations of the density. In the context of an extended system *i.e.* infinitisemal change of the electronic density $\delta$, the exact xc discontinuity reads

$$\Delta_{xc}(\mathbf{r}) = \lim_{\delta \to 0} \left( \frac{\delta E_{xc}}{\delta n(\mathbf{r})}\bigg|_{N+\delta} - \frac{\delta E_{xc}}{\delta n(\mathbf{r})}\bigg|_{N-\delta} \right) . \tag{3.33}$$

As already pointed out in expression (3.32), the response part of the exchange and correlation potential can vary discontinuously between different states (in the case of an insulator, in the bandgap region), giving rise to jump in the eigenvalues [54].

**In this section**, we introduced the concept of bandgaps and showed why the eigenvalues of the KS equation cannot be used to evaluate bandgaps. We also discussed the difficulty of obtaining practical exchange and correlation potentials. In the next section, we will introduce local and semi-local approximations to $v_{\mathrm{xc}}$ that are used in practice to study the electronic properties of solids.

## 3.6 Local and semilocal approximations to $V_{\mathrm{xc}}$

### 3.6.1 Local density approximation

The simplest approximation to the exchange and correlation energy is the Local Density Approximation (LDA). One uses the exchange and correlation

energy density of the homogeneous electron gas $\epsilon_{xc}^{\mathrm{hom}}$ [68], that yields the following exchange and correlation energy function

$$E_{xc}[n] = \int n(\mathbf{r})\epsilon_{xc}^{\mathrm{hom}}(n(\mathbf{r}))d^3\mathbf{r}\,. \tag{3.34}$$

One can immediately see that this approximation can be expected to work only for slowly varying electronic density ressembling the one of the homogeneous electron gas. However, it has proved to work rather well even for systems which have very inhomegeneous electron densities. In order to investigate this apparent paradox in more detail, we introduce the electron pair density $\rho_2(\mathbf{r},\mathbf{r'})$, that is the probability of finding one electron at position $\mathbf{r}$ and another at position $\mathbf{r'}$.

$$\rho_2(\mathbf{r},\mathbf{r}') = N(N-1)\int d\mathbf{r}_3...d\mathbf{r}_N\Psi^*(\mathbf{r},\mathbf{r}',\mathbf{r}_3...\mathbf{r}_N)\Psi(\mathbf{r},\mathbf{r}',\mathbf{r}_3...\mathbf{r}_N)\,, \tag{3.35}$$

where $\Psi$ is the many-body wave function. Due to correlation effects between electrons, $\rho_2(\mathbf{r},\mathbf{r'}) \neq \rho(\mathbf{r})\rho(\mathbf{r'})$. Therefore, one can define the exchange-correlation (xc) hole $\rho_{xc}(\mathbf{r},\mathbf{r'})$ as

$$\rho_2(\mathbf{r},\mathbf{r}') = \rho(\mathbf{r})\rho(\mathbf{r}') + \rho(\mathbf{r})\rho_{xc}(\mathbf{r},\mathbf{r}')\,. \tag{3.36}$$

From the point of view of probability theory, the xc-hole ressembles a conditional probability *i.e.* probability of finding an electron at $\mathbf{r}$' given there is one at $\mathbf{r}$

$$\rho_{xc}(\mathbf{r},\mathbf{r}') = \frac{\rho_2(\mathbf{r},\mathbf{r}')}{\rho(\mathbf{r})} - \rho(\mathbf{r}')\,. \tag{3.37}$$

Furthermore, the xc-hole has a physical meaning: because of the Pauli exclusion principle, one electron is surrounded by a deficit of charge adding up to exactly one electronic charge. This condition is expressed as a sum rule:

$$\int \rho_{xc}(\mathbf{r},\mathbf{r}')d\mathbf{r}' = -1\,. \tag{3.38}$$

Figure 3.1 shows the xc-hole calculated within Variational Monte Carlo (VMC) and within LDA for the case of silicon [69]. The LDA xc-hole is spherically symmetric which is not true for the xc-hole calculated in VMC, but the overall shape one is well reproduced by LDA. LDA xc-hole fulfills sum rule (3.38). It implies that while in some regions of space, LDA xc-hole may have positive errors, it must have negative errors in orther part of space leading to a systematic error cancellation. This point is even more evident

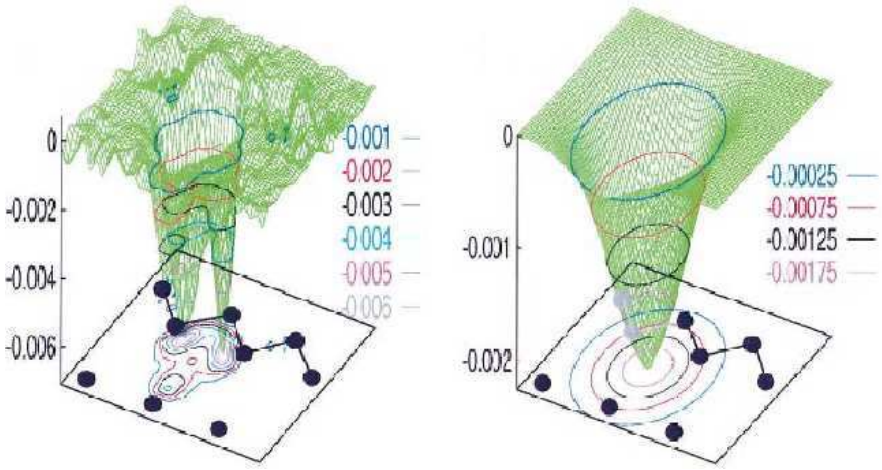


**Figure 3.1:** The exchange-correlation hole $\rho_{xc}(\mathbf{r}, \mathbf{r'})$, with one electron fixed at the tetrahedral interstitial site in the (110) plane of bulk silicon. Left panel Quantum Monte Carlo calculation. Right panel: LDA calculation [69].

if one expresses the exchange correlation energy functional as a function of the xc-hole

$$E_{xc} = \int d\mathbf{r} n(\mathbf{r}) \int d\mathbf{r}' \frac{\rho_{xc}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \,. \tag{3.39}$$

By observing that the Coulomb interaction is isotropic [70], substituting the variable $\mathbf{R}$=$\mathbf{r}$'-$\mathbf{r}$ in expression (3.39), the exchange correlation energy then reads

$$E_{xc} = \int d\mathbf{r} n(\mathbf{r}) \int_0^\infty dR R^2 \frac{1}{R} \int d\Omega \rho_{xc}(\mathbf{r}, \mathbf{R}) \,. \tag{3.40}$$

The exchange and correlation energy depends only on the spherical average of $\rho_{xc}$. Therefore, even if the description of the nonspherical part of $\rho_{xc}$ is inaccurate, the most important information is contained in the spherical part that should be correctly described by the electron gas.
However, LDA of course is not perfect: in particular it tends to favor homogeneous electronic densities. Therefore, it overestimates the binding energy of molecule and cohesive energy of solids. As a result, bond lengths and lattice parameters are usually underestimated. Besides, it tends to delocalize *localized* states such as $3d$ or $4f$ states. DFT-LDA bandgaps are usually strongly underestimated, which should however not be related only to the deficiency of the LDA itself but rather to the problem of KS formalism.

**In summary**, despite the apparent simplicity of the local density approximation, it is an extremely powerful approximation. Indeed, it contains most of the information necessary to evaluate $E_{xc}$ and fulfills exactly sum rules.

However, there is still room for improvement when treating non homogenous materials. A natural way to improve LDA is to expand $\mathrm{E_{xc}}[n]$ as a series of the density $n$. The expansion will be described in the next section as it gives rise to the generalized gradient approximation.

### 3.6.2 Generalized Gradient Approximation

A straightforward improvement over the LDA is to introduce a further dependence of the xc energy on the gradient of the density in order to take into account the inhomogeneity of the system. This has been proposed in the original paper of Kohn and Sham under the name of "gradient expansion approximation" (GEA).

$$E_{xc}^{\mathrm{GGA}}[n] = \int d^3\mathbf{r} n(\mathbf{r}) \epsilon_{xc}(n(\mathbf{r}), \nabla n(\mathbf{r}))\,. \tag{3.41}$$

However, the GEA provides no consistent improvement over LDA [71, 72]. An analysis of the xc-hole of GEA shows that while the short range part of the former is improved, the long range part is significantly worsened [73]. In order to correct this behavior, one should enforce the fulfillment of sum rules like Eq. (3.38). The resulting functionals are known under the name of Generalized Gradient Approximation (GGA) functionals. Usually, the fulfillment of the exact behavior in some limit is enforced: for example the exact hole condition [74], the asymptotic behavior of the exchange potential 1/r [75], the slowly varying limit [76], or the correct scaling law [77]. Compared to LDA, GGA tends to improve the total energy and atomization energy. It has a clear tendency to expand bonds, which is the reason why GGA usually overestimates lattice parameters. One of the famous successes of GGA compared to LDA is the correct prediction of the bcc ferromagnetic ground state of iron [78, 79, 80] due to its better estimation of the lattice parameter. Interestingly, GGA does not improve considerably the bandgap with respect to the LDA. In fact, GGA and LDA bandgaps are very similar, which is the a strong indication that the bandgap problem originates from the xc discontinuity $\Delta_{xc}$ rather than the exchange and correlation functional itself. Furthermore, even for ground state problems, there is no systematic improvement of GGA over LDA: for example, while GGA is particularly accurate for first and second row constituents, it totally fails to predict correct bond lengths and cohesive energy for heavy constituents such as metal $5d$ [81] or fifth row elements [82, 83] while LDA results are very close to experiments. However, comparing GGA with LDA can give some idea about the physics at work in the system of interest.

**In this section**, we presented the Generalized Gradient Approximation and how the expansion as a series of the density reveals the importance of sum rules fulfilled by LDA. The improvement compared to LDA is generally not spectacular in particular for bandgaps. Therefore, the main deficiency of LDA or GGA in the treatement of bandgaps lies within the KS formalism rather than the functionals itself. In the next section, we will present one of the many errors present within both GGA or LDA which is particularly important for the system we want to study.

## 3.7 Self interaction problem

### 3.7.1 Self Interaction correction

One of the main drawbacks of the previously introduced xc functionals is the self-interaction (SI) problem. Let us introduce this problem with an example. In 1934, Fermi *et al.* represented the expectation value of $U_{e-e}$ considering only the Hartree contribution [84]

$$U[n] = \frac{1}{2}\int\int d\mathbf{r}d\mathbf{r}'\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,. \tag{3.42}$$

In the limit of one electron where $n(\mathbf{r}) = |\phi(\mathbf{r})|^2$, expression (3.42) does not vanish: hence the electron interacts with itself, which is called the self-interaction problem. Hartree-Fock theory on the contrary is self-interaction free as the self-interaction in the Hartree term is cancelled *exactly* by the self-interaction term in the Fock exchange term

$$E_x = -\frac{1}{2}\sum_{i,j\in\mathrm{occ}}\int d\mathbf{r}\int d\mathbf{r}'\frac{\phi_i^*(\mathbf{r})\phi_j^*(\mathbf{r}')\phi_j(\mathbf{r})\phi_i(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,, \tag{3.43}$$

where $\phi_i$ are one-electron wave functions. Let us consider again the limit of one electron. $E_x$ becomes

$$E_x = -\frac{1}{2}\int d\mathbf{r}\int d\mathbf{r}'\frac{\phi^*(\mathbf{r})\phi(\mathbf{r})\phi(\mathbf{r}')\phi^*(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} = -U[n]\,. \tag{3.44}$$

Due to the exact formulation of $E_x$, the self-interaction vanishes. However, for a KS functional such as LDA or GGA, the self-interaction occuring in the Hartree term is only *partially* cancelled by the density dependent exchange potential. For real systems, localized states such as $d$ or $f$ electrons are particularly sensitive to the self-interation problem. Therefore, in this thesis

work it is extremely important to cure this particular drawback of approximated functionals as Cu $3d$ states are present in the vicinity of the bandgap region for CIS. For example, the self-interaction contribution to the Hartree potential is of the order of 15 eV for Cu $3d$ [85].

Fermi *et al.* [84] proposed a simple scheme to remove this spurious effect, namely

$$U_{e-e}[n] = U[n] - NU\left[\frac{n}{N}\right] , \tag{3.45}$$

where $N$ is the number of electrons of the system. $n/N$ is a rudimentary one-electron density. The self-interaction is therefore removed for all $N$ electrons of the system which explains the factor N in front of the second term of expression (3.45). Following Fermi's spirit, Perdew and coworkers [85, 86] developed the idea of a scheme fulfilling the exact condition in order to cancel the self interaction problem

$$\forall i \in \text{ occ}, E_H[|\phi_i|^2] + E_{xc}[|\phi_i|^2] = 0\,. \tag{3.46}$$

As the exchange term cancels completely the Hartree one, we obtain two relations

$$\forall i \in \text{ occ} \begin{cases} E_H[|\phi_i|^2] + E_x[|\phi_i|^2] = 0\,. \\ E_c[|\phi_i|^2] = 0\,. \end{cases} \tag{3.47}$$

Considering an approximate exchange and correlation functional $\mathrm{E}_{xc}$, the self-interaction corrected (SIC) functional reads

$$E_{xc}^{SIC}[n] = E_{xc}^{\text{app}}[n] - \sum_{i \in \text{ occ}} \left\{E_{xc}^{\text{app}}[|\phi_i|^2] + E_H^{\text{app}}[|\phi_i|^2]\right\} . \tag{3.48}$$

In order to get a practical scheme, one applies the same procedure as in the case of the original KS equation. One obtains a KS-like equation with an orbital dependent SIC potential (corresponding to the KS potential)

$$v_{\text{SIC}}^i(\mathbf{r}) = v_{\text{ext}}(\mathbf{r}) + v_H[n](\mathbf{r}) - v_H[n^i](\mathbf{r}) + v_{\text{xc}}[n](\mathbf{r}) - v_{\text{xc}}[n^i](\mathbf{r})\,, \tag{3.49}$$

where $n^i = |\phi_i|^2$.
There are two main advantages of the SIC potential: first it corrects the self-interaction problem as it was designed to do so. Second, the orbital dependence of the SIC potential introduces the strong non-analycity that was missing in LDA or GGA. Indeed, one can realize that the orbital dependence of $V_{\text{SIC}}$ follows the prescriptions proposed in section 3.5 such as the orbital dependence of xc potential.

Perdew *et al.* considered the SIC scheme in the light of Koopmans' theorem [86] and predicted the improvement of SIC bandgaps compared to DFT-KS ones. Indeed, SIC has shown great improvement of binding energies or atomization energies, making it a good candidate for calculating excitation energies. Moreover, the highest occupied states are believed to be improved compared to LDA due to the correct asymptotic behavior of the SIC potential [86]. Furthermore, the nice properties of the LDA xc-hole are not *destroyed* by the use of the SIC scheme [86]. However, while succesfully applied to free atoms [86], the SIC potential vanishes in the case of bulk wavefunctions which extend over the whole space. Therefore, any implementation of SIC in a solid requires a description in terms of localized states, like Wannier functions [87] or muffin-tin spheres [88]

**In conclusion**, we present one of the main pathologies of DFT for localized states, the so-called self interaction problem. It is of primer importance in this work as Cu $3d$ states play a major role in the physics of CIS. We introduce the self-interaction correction scheme that not only improves total energies but also eigenvalues.In subsequent sections, we will present other orbital dependent schemes that partially cure the self-interaction problem and might be appropriate for the treatment of localized states.

### 3.7.2 Meta-GGA

Meta-GGA is a functional very similar to GGA, derived from the same reasoning: the gradient expansion. But first, let us consider the expression of the ground state total energy. One can show by means of the virial theorem that the kinetic energy term is of the order of magnitude of the total energy. It is natural to try to describe it to the best of our knowledge and capabilities. Thomas and Fermi [57, 58] proposed a density dependent expression of the kinetic energy of the interacting system. The KS formalism follows a smarter path: as the expression of the kinetic energy of the non interacting system is trivial, the KS formalism only requires the evaluation of the difference between the KS kinetic energy and the true kinetic energy. The latter difference is believed to be much smaller than the true kinetic energy and therefore less dependent on approximations. The meta-GGA tries to improve the description of the kinetic energy by considering the kinetic energy density $\tau$

$$\tau = \frac{1}{2}\sum_i |\nabla\phi_i|^2 ., \tag{3.50}$$

then the meta-GGA functional $E_{xc}$ is expressed as

$$E_{xc}^{\text{meta-GGA}}[n] = \int d^3\mathbf{r} n(\mathbf{r}) \epsilon_{xc}(n(\mathbf{r}), \nabla n(\mathbf{r}), \tau) \,. \tag{3.51}$$

In the spirit of the self-interaction correction scheme, the meta-GGA can be made self-correlation free. This corresponds to the second condition given by Eq. (3.47) *i.e.* the correlation energy should vanish for any one-electron density. van Weizsäcker had introduced a kinetic energy $\tau^W$, which is a more performant indicator of one-electron region than Fermi's one:

$$\tau^W = \frac{|\nabla n|^2}{8n} \,. \tag{3.52}$$

Then one can build the ratio $\tau/\tau^W$ which is 1 for one-electron region [89] *ie* for regions of the solid where the electronic density is so low that it corresponds to one single state. With the help of this operator, one can construct a functional that nullifies the correlation energy in one-electron region of the space. The problem of self-interaction is however only slightly reduced as only the self-correlation is cured. In the case of Cu $3d$, self-correlation only accounts for less than 1 eV [85] *i.e.* it is an order of magnitude smaller than the self-interaction in the exchange term.

### 3.7.3 LDA+U

A different approach towards the correct treatment of localized states has been proposed, relying on a non-*ab initio* model hamitonian. Let us first introduce the Hubbard model Hamiltonian [90]

$$H = -t \sum_{(i,j) i \neq j, \sigma} (c_{i,\sigma}^\dagger c_{j,\sigma} + h.c.) + U \sum_i n_{i,\uparrow} n_{i,\downarrow} \,, \tag{3.53}$$

where $c_{i,\sigma}^\dagger$ and $c_{j,\sigma}$ are creation and annihilation operators and $n_{i,\sigma} = c_{i,\sigma}^\dagger c_{i,\sigma}$. The first term describes the kinetic energy in a tight-binding way, electrons hopping from one lattice site $i$ to another $j$. The sum over $\{i, j\}$ is restricted to nearest neighbors. The kinetic term has a tendency to delocalize states and is at the origin of the dispersion of the bands. On the other hand, the second term describes the on-site interaction: $U$ is the energy cost for two electrons to sit at the same site. Contrary to the kinetic term, the on-site term tends to localize states. The Hubbard Hamiltonian is particularly suited to study localized states such as $d$ or $f$ electrons. The fact that only opposite spins on the same site interact makes the Hubbard model self-interaction free. Therefore, it is appealing to combine it with LDA in order to obtain a

consistent treatment of both localized and delocalized states.

Following this spirit, Anisimov *et al.* [91] proposed a LDA+U total energy functional

$$E_{LDA+U} = E_{LDA} + E_U - E_{dc} = E_{LDA} + \frac{1}{2}U\sum_{i\neq j} n_i n_j - U\frac{N_d(N_d-1)}{2}\,, \quad (3.54)$$

where $n_i$ are the occupations of the $d$ orbitals and $N_d = N_d = \sum_{i\in} n_i$. The second term $E_U$ describes $d-d$ interaction only and the last term $E_{dc}$ is a *double counting* correction. Let us now move to the real implementation of LDA+U. One builds then an orbital dependent exchange and correlation potential $V_{LDA+U}$

$$v_{LDA+U}(\mathbf{r}) = \sum_i v_i(\mathbf{r})|i\rangle\langle i|\,. \quad (3.55)$$

where $i$ runs over all the possible principal $n$ and azimuthal $l$ quantum numbers and atomic sites. Note the presence of the projector operator that makes the potential act differently on delocalized ($s$ and $p$) and on the localized ($d$ and $f$) orbitals, making the potential orbital dependent hence non local. $v_i(\mathbf{r})$ is computed through the derivative of the total energy with respect to the partial electronic density $\rho_i(\mathbf{r})$. For $s$ and $p$ electrons *i.e.* $i = s$ and $p$, the potential $v_i(\mathbf{r})$

$$v_i(\mathbf{r}) = \frac{dE_{LDA+U}}{d\rho_i} = \frac{dE_{LDA}}{d\rho_i} = v_{LDA}(\mathbf{r})\,. \quad (3.56)$$

yielding $\epsilon_i = \epsilon_i^{\mathrm{LDA}}$. However, in the case of $d$ electrons *i.e.* $i = d$, the potential $V_i(\mathbf{r})$ is

$$v_i(\mathbf{r}) = v_{LDA}(\mathbf{r}) + U(\frac{1}{2} - n_i)\,, \quad (3.57)$$

which leads to the eigenvalue

$$\epsilon_i = \epsilon_{i,LDA} + U(\frac{1}{2} - n_i)\,. \quad (3.58)$$

When the $d$ orbital is occupied ($n_i$=1), the LDA+U energy is shifted by -U/2 with respect to the LDA eigenvalues. However when orbitals are unoccupied, their level is upshifted by +U/2. Then, full and empty bands are separated by an energy gap $\Delta = U$. LDA+U tends to localize electrons on $d$ orbitals where the occupation is larger than 1/2. LDA+U partially solves the self-interaction issue encountered in LDA. There is however still room for improvement, through the double counting term. [92].

More physics is contained in the U parameter than just a parameter from a model Hamiltonian. The above discussion only treats the one-band Hubbard model. However it can be generalized to more general cases [93, 94] *i.e.* taking into account exchange interaction between $d$ orbitals and the polarization of $d$ orbitals. Then, the parameter U can be interpreted as the energy cost to move one $d$ electron from one atom to another, renormalized by the screening due to the $s$ and $p$ orbitals [95]. Usually, an $s$ or $p$ orbital moves in the opposite direction in order to reduce the energetic cost.
Despite its clear physical meaning there exists several ways to determine U, ranging from fully *ab initio* to empirical approaches:

- U can be evaluated by means of constrained-density-functional calculation [96, 97, 98, 99, 100]: one $d$ electron is moved to one site, then the electronic structure is relaxed to take into account the screening of $s$ and $p$ states.

- U can be evaluated through fitting of experimental spectra: this is usually done comparing the density of states computed in LDA+U and photoemission spectra, possibly energy resolved only to retain the orbitals $d$ and $f$ affected by U.

**In conclusion**, LDA+U is one of the methods of choice in order to treat localized states and cures mostly the self-interaction correction indirectly. It shares with SIC the orbital dependence of its potential. Despite its clear physical meaning, the evaluation of U often involves simple fit procedures, which makes LDA+U a parameter-dependent method. We will detail in the result chapter the connection between U and the electronic screening in a solid. In the next section, we will present some technical details about DFT: *how is it implemented in practice, which approximations are made in order to treat real systems ?*

## 3.8 Practical Implementation of DFT

As solid-state physicists traditionally deal with perfect crystals, they take benefit of the periodic boundary condition using Bloch's theorem. The use of plane waves as a basis set follows naturally. Let us consider a crystal with a periodicity $\mathbf{R}$ such that the potential $v(\mathbf{r}) = v(\mathbf{r}+\mathbf{R})$. Then the wave function $\phi_{n,\mathbf{k}}$, eigenstate of the periodic Hamiltonian $H^{KS}$ can be written as :

$$\phi_{n,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n,\mathbf{k}}(\mathbf{r})\,, \tag{3.59}$$

where $\mathbf{k}$ is the wave vector and $u_{n,\mathbf{k}}$ is a crystal periodic function. This is Bloch's theorem. It implies that plane waves can be used as basis set in the context of periodic crystals. Thus, the wave functions can be expressed as a plane wave expansion using the reciprocal lattice space $\{\mathbf{G}\}$

$$\phi_{n,\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{N_{\mathbf{k}}\Omega_c}} \sum_{\mathbf{G}} u_{n,\mathbf{k}}(\mathbf{G}) e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}\,, \tag{3.60}$$

where $N_{\mathbf{k}}$ is the number of $\mathbf{k}$-points and $\Omega_c$ the volume of the primitive cell. Then, the KS equations take the form of a secular equation:

$$\sum_{\mathbf{G}'} H^{KS}(\mathbf{k}+\mathbf{G}, \mathbf{k}+\mathbf{G}') u_{n,\mathbf{k}}(\mathbf{G}') = \epsilon_{n,\mathbf{k}} u_{n,\mathbf{k}}(\mathbf{G})\,. \tag{3.61}$$

The matrix form of the KS Hamiltonian reads

$$H^{KS}(\mathbf{k}+\mathbf{G}, \mathbf{k}+\mathbf{G}') = \frac{(\mathbf{k}+\mathbf{G})^2}{2}\delta_{\mathbf{G},\mathbf{G}'} + v_{KS}(\mathbf{k}+\mathbf{G}, \mathbf{k}+\mathbf{G}')\,, \tag{3.62}$$

The first term of expression (3.62) is the kinetic term and the second term is the Fourier transform of the KS potential. A priori, the KS potential is *local*. Therefore, its Fourier representation can be simplified even further with a dependence on the difference $\mathbf{G}-\mathbf{G}'$ only. However, in practice the KS potential often has a non local component (from a nonlocal pseudopotential and for orbital dependent potential such as in LDA+U) whose Fourier transform should be evaluated as

$$V_{\mathrm{NL}}(\mathbf{k}+\mathbf{G}, \mathbf{k}+\mathbf{G}') = \frac{1}{\Omega_c}\int d\mathbf{r} d\mathbf{r}' e^{-i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} V_{\mathrm{NL}}(\mathbf{r}, \mathbf{r}') e^{i(\mathbf{k}+\mathbf{G}')\cdot\mathbf{r}'}\,. \tag{3.63}$$

The summation over $\mathbf{G}$ is truncated and in most cases the limit is defined by a cut-off energy $E_{\mathrm{cut}}$

$$\frac{|\mathbf{k}+\mathbf{G}|^2}{2} \leq E_{\mathrm{cut}}\,. \tag{3.64}$$

*i.e.*, the basis set is defined as a sphere whose radius is defined by a kinetic cut-off energy.

In principle, any basis can be selected in order to expand the wave functions. Plane waves are a set of choice, because of their mathematical simplicity, systematic convergence, periodic character that makes them suitable for reproducing the periodicity of the lattice.

Other basis sets exist though, one of the most common being atom-centered basis sets. Such basis may reduce in many cases significantly the number of basis functions in order to reproduce the true wave functions particularly

for the localized states. On the other hand, they may suffer from non systematic convergence issues like in the case of gaussian basis sets, and may have difficulties to reproduce delocalized states like conduction states. In the calculation of forces, they can also cause the appearance of spurious contributions an extra stress called when turning to force calculation. Overall, localized basis set are used extensively by the chemistry community since they are more suited for finite systems.
For solid-state problems, when plane waves are to be used, one has to adopt a smart approximation in order to get rid of the localized states. Most notably, core states are known to have little influence on the chemical bonding. One can hence separate the valence electrons from the core electrons using a pseudopotential or a projected augmented wave approach. They are discussed in more details in the appendix 1 of this thesis. Briefly, the so-called frozen core approximation keeps the core electrons in their atomic states. Their effect on the valence electrons can then be simulated by an effective potential, a so called pseudopotential. This reduces considerably the number of states to be considered in the calculation. Furthermore, it allows one to employ efficiently a plane wave basis set because the wave functions of the valence states are rather smooth compared to the ones of core states once the wiggles of the valence wave functions are smoothed out in the procedure of pseudization. Knowing the coefficients $u_{n,\mathbf{k}}(\mathbf{G})$ of the wave functions, we can express the density in the same basis

$$\rho(\mathbf{r}) = \sum_{n} \sum_{\mathbf{k}} \sum_{\mathbf{G}} \sum_{\mathbf{G'}} w_{\mathbf{k}} \theta(\epsilon_f - \epsilon_{n,\mathbf{k}}) u^*_{n,\mathbf{k}}(\mathbf{G}) u_{n,\mathbf{k}}(\mathbf{G'}) e^{i(\mathbf{G}-\mathbf{G'}).\mathbf{r}} , \qquad (3.65)$$

where $w_{\mathbf{k}}$ are the weights of the **k**-points. It can be seen that while the one-electron wave function $\phi_n$ depends only on **G**, the electronic density depends on **G**-**G'** which may lie outside of the sphere defined by the cut-off energy (3.64). Therefore, the expression of the electronic density requires to double the radius of the sphere. In practice, this is not a real issue as a careful convergence test should be performed, increasing the radius of the sphere until the quantity of interest is converged. The second summation present in the expression of the density runs over the **k**-points. The symmetry of the crystal permits to reduce integrations over the irreducible Brillouin zone only. Then different numerical schemes are available to minimize the number of **k**-points used: *mean value point* of Baldereschi *et al* [101] where one special **k**-point of the Brillouin zone is supposed to reproduce the average of any arbitraty periodic function. A generalized procedure was proposed by Chadi and Cohen [102] and obtains the best sampling relaxing the constraint about the single **k**-point. A better sampling was achieved by Monkhorst and Pack [103] who developed a uniform minimal grid of **k**-points. For the time

being, Monkhorst-Pack grid is the most efficient and popular in the solid-state community.

The Hellmann-Feynman theorem applies straightforwardly to DFT. The energy is at a variational minimum with respect to the electronic density and therefore any change of the density when a nucleus is moved will not contribute to $\mathbf{F}_i$. However, the use of non-local pseudopotentials urges the inclusion of the Kohn-Sham wave functions as the force is not dependent on the electronic density only as in Eq. (3.11). Additional force components may arise from the lack of completeness of the basis set or from the atomic dependent basis set such as the Pulay stress [104]. In this thesis, only plane waves were used so that they ensure the Pulay stress to be zero.
The numerical resolution of the Kohn-Sham system of equations is straightforward once the electronic density is calculated. Figure 3.2 describes the self-consistent procedure usually used in DFT. First, one should pick an arbitrary potential. The determination of a good starting point is a relatively intuitive task: the use of overlapping potentials or overlapping spherical atomic densities are one of the possibilities [105]. Then, Kohn-Sham equations are solved by matrix diagonalization or using conjugate gradient algorithms. Summing over all the occupied states $\phi_i$, the electronic density is obtained and permits to construct the Kohn-Sham potential $v_{KS}(\mathbf{r}) = v_{II}(\mathbf{r}) + v_H([n], \mathbf{r}) + v_{xc}([n], \mathbf{r})$. This potential is compared to the potential obtained from the previous iteration, by defining a distance for instance such as

$$D[v_i, v_{i+1}] = \sqrt{\frac{1}{\Omega_c} \int_{\Omega_c} d\mathbf{r} \left[v_i(\mathbf{r}) - v_{i+1}(\mathbf{r})\right]^2}, \tag{3.66}$$

where $v_i$ and $v_{i+1}$ are the potentials at iteration $i$ and $i+1$. If $D[i, i+1]$ lies below the convergence parameter, then the self-consistence cycle is stopped. Otherwise, a new potential is generated by mixing the potential with potentials from previous iterations. Various mixing schemes exist [106, 107, 108] in order to accelerate the convergence and to prevent any numerical instabilities. The HK theorem ensures that the potential can be replaced by the electronic density as the quantity to be determined self-consistently. Thus, the previous discussion holds similarly in the case of the electronic density.

**In conclusion**, we have presented the general form of a DFT numerical scheme as it is implemented in modern *ab initio* software. In the next section, we will introduce another approach which is based on a different energy functional and, most importantly, on a generalized KS formalism.

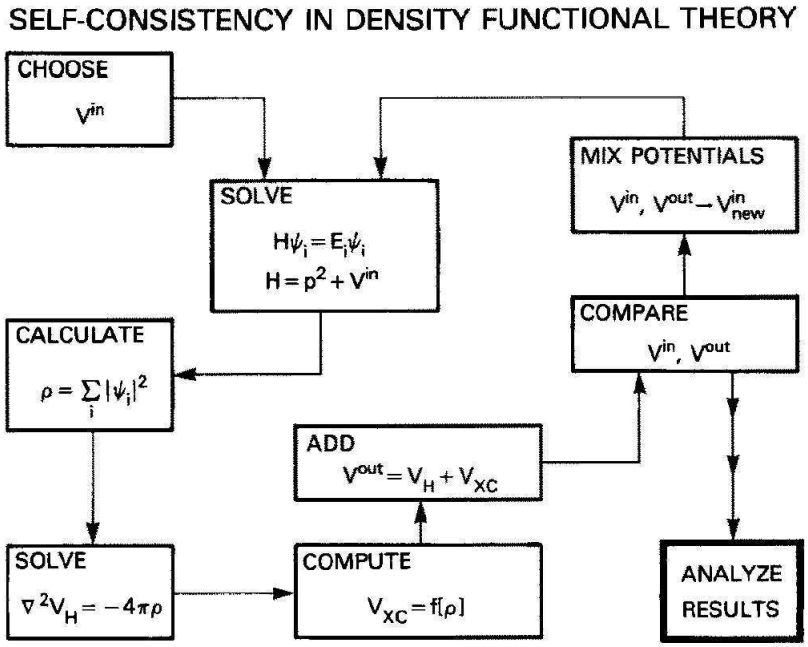


**Figure 3.2:** Flow chart describing the self-consistent procedure implementing DFT [109].

## 3.9 Hybrid Functionals

Hybrid functionals have encountered a growing success for the last 15 years. Hybrids partially cure the self-interaction problem in the ground state and yield improved bandgap compared to standard KS. We will show that the enhancement of bandgaps is mostly inherent to the use of a generalized KS formalism. In order to derive hybrid energy functionals, we introduce a rather general theorem which is often used to establish exact conditions or limits for exchange and correlation potentials: the adiabatic connection theorem [110, 111, 70]

### 3.9.1 Adiabatic Connection theorem

In sec. 3.4, we have seen that we can map the interacting electron system to a non interacting system with the help of an external potential $U_{\text{ext}}$. Let us consider an adiabatic path along this mapping by turning on the electron interaction. Let us consider an Hamiltonian with a similar form as the one of Eq. (3.6)

$$\mathcal{H}_\lambda = T_e + U_{\text{ext},\lambda} + \lambda U_{e-e}\,, \tag{3.67}$$

where $0 \leq \lambda \leq 1$. For $\lambda = 1$, we choose $U_{\text{ext},1} = U_{N-e}$ in order to recover the many body Hamiltonian of chapter 1. For any $\lambda \neq 1$, $U_{\text{ext},\lambda}$ defined to be the potential that yields the exact ground state electronic density. As a consequence for $\lambda = 0$, $U_{\text{ext},0} = U_{KS}$. Using the Hohenberg and Kohn theorem, we can write the total energy as a functional of the electronic density

$$E_{\text{tot},\lambda} = \int v_{\text{ext},\lambda} n(\mathbf{r}) d\mathbf{r} + F_\lambda[n] \,, \tag{3.68}$$

where the functional $F_\lambda[n]$ is defined such as

$$F_\lambda[n] = \langle \Psi_\lambda | T + \lambda v_{e-e} | \Psi_\lambda \rangle \,. \tag{3.69}$$

From this definition, we can rewrite Eq. (3.16) as

$$E_{xc}[n] = F_1[n] - F_0[n] - E_{\text{Hartree}}[n] \,, \tag{3.70}$$

$$\text{hence} \quad E_{xc}[n] = \int_0^1 \frac{\partial F_\lambda[n]}{\partial \lambda} d\lambda - E_{\text{Hartree}}[n] \,, \tag{3.71}$$

Using the Hellmann-Feynman theorem [56] the derivative of $F_\lambda$ with respect to $\lambda$ is

$$\frac{\partial F_\lambda[n]}{\partial \lambda} = \langle \Psi_\lambda | v_{e-e} | \Psi_\lambda \rangle \,, \tag{3.72}$$

which suggests the definition of a $\lambda$ exchange-correlation energy $\mathrm{E}_{\text{xc},\lambda}$

$$E_{xc,\lambda}[n] = \langle \Psi_\lambda | v_{e-e} | \Psi_\lambda \rangle - E_{\text{Hartree}}[n] \,. \tag{3.73}$$

The kinetic term of the exchange correlation energy disappears. Still, one has to integrate between the Kohn Sham system ($\lambda = 0$) and and the real many body system ($\lambda = 1$).
Many properties of $\mathrm{E}_{xc,\lambda}$ have been derived. The most important ones for the derivation of Hybrid functional are listed below

- for $\lambda = 0$, $E_{\text{xc},0} = E_{\text{x},0} = \langle \Psi_0 | v_{e-e} | \Psi_0 \rangle - E_{\text{Hartree}}[n]$ where $\Psi_0$ are Slater determinants of one electron Kohn-Sham wave functions. This expression shows very nicely that one of the bounds of the integral can be approximated by Fock exchange energy constructed with KS wave functions.

- $E_{\text{xc},\lambda}$ is a monotically decreasing function of $\lambda$ [112]

### 3.9.2 Hybrid functional

In 1993, Becke derived the first hybrid functional by approximating the integral of Eq. (3.71)with the trapezoidal rule integration [113]

$$E_{\mathrm{xc}} \approx \frac{1}{2}(E_{\mathrm{xc},0} + E_{\mathrm{xc},1}) \approx \frac{1}{2}(E_{\mathrm{x},0} + E_{\mathrm{xc,LDA}})\,. \tag{3.74}$$

Despite its poor improvement compared to modern semilocal functionals, this hybrid draw a particular attention to the concept of mixing a fraction of exact exchange with a fraction of local or semilocal exchange and correlation. A better approximation can be made by considering the monotonic behavior of $E_{\mathrm{xc},\lambda}$; mean-value theorem guarantees the existence of a mixing parameter $b$ such that

$$\begin{aligned} E_{\mathrm{xc}} &= bE_{\mathrm{x},0} + (1-b)E_{\mathrm{xc},1} = bE_{\mathrm{x},0} + (1-b)E_{\mathrm{x},1} + (1-b)E_{\mathrm{c},1}\,, & (3.75)\\ E_{\mathrm{xc}} &= bE_{\mathrm{x},0} + (1-b)E_{\mathrm{x,app}} + E_{\mathrm{c,app}}\,. & (3.76) \end{aligned}$$

where $\mathrm{E}_{x,0}$ is of the form the Fock exchange energy but evaluated with KS wave functions, and functionals $E_{\mathrm{x,c,app}}$ are approximated form of the exchange and correlation energy usually based on DFT such as LDA or GGA. Eq. (3.76) is the *common* form of an hybrid functional [3]. Correlation is accounted for through a semilocal functional only while the exchange energy results from the mixing of the exact exchange and an approximation of exchange. One question still remains: *how to determine the mixing parameter b ?*. Most of the attempts to determine $b$ rely on empirical mean square fitting of the atomization energies, ionization potential, proton affinities of a canonical set of organic and inorganic molecules known as the G2 set [114]. This procedure yields values of $b$ comprised between 0.16 and 0.28 [115]. The value b=0.25 emerges from work trying to rationalize the mixing parameter [116] but it appears to be extremely dependent on the type of GGA used [117, 118, 119]. The direct application of (3.76) with b=0.25 yields the PBE0 functional [119, 120]. This functional improves over the already very nice agreement with experiments of standard GGA for the prediction of structural parameters. Besides, the introduction of a portion of Hartree-Fock exchange partially solves the self-interaction problem, making the use of PBE0 particularly suitable for materials where $d$ or $f$ electrons play a major role.

Let us turn now to the bandgap problem. As we deal with an explicitly orbital dependent quantity *i.e.* the Fock exchange term, one may wonder:

[3] one should notice the simplification made $(1-b)E_{\mathrm{c},1} \approx E_{\mathrm{c,app}}$

*how can we apply the KS formalism to the Fock term ?* In order to resolve this apparent issue, we introduce the optimized effective potential (OEP) method [121]. In the OEP framework, the energy functional is thought to be an explicit functional of the single particle wave functions and energies $\{\phi_i\}$ and $\{\epsilon_i\}$, E$[\{\phi_i, \epsilon_i\}]$ which is an implicit functional of the density. Owing to HK theorems, the total energy is variational with respect to the KS potential. Therefore applying chain rule

$$\frac{\delta E}{\delta v_{KS}} = \sum_i \int d\mathbf{r}' \left[ \frac{\delta E}{\delta \phi_i(\mathbf{r}')} \frac{\delta \phi_i(\mathbf{r}')}{\delta v_{KS}(\mathbf{r})} + h.c. \right] + \frac{\partial E}{\partial \epsilon_i} \frac{\delta \epsilon_i}{\delta v_{KS}(\mathbf{r})} = 0 \,. \tag{3.77}$$

The sum runs over all states $i$ is not restricted to occupied states only. Using the condition that the orbitals $\phi_i$ are solution of the KS equation with a local potential, one finds the OEP integral equation for $V_{\text{xc}}$ [122], a Fredholm equation of first kind,

$$\int d\mathbf{r}' \frac{\delta \rho(\mathbf{r})}{\delta v_{KS}(\mathbf{r}')} v_{xc}(\mathbf{r}') = \Lambda_{xc}(\mathbf{r}) \,, \tag{3.78}$$

where the inhomegeneity $\Lambda_{xc}$ reads

$$\Lambda_{xc}(\mathbf{r}) = \sum_{i \neq j} \left[ |\phi_i(\mathbf{r})|^2 \frac{\partial E_{xc}}{\partial \epsilon_i} - \int d\mathbf{r}' \phi_i(\mathbf{r}) \frac{\phi_j(\mathbf{r}) \phi_j^*(\mathbf{r}')}{\epsilon_j - \epsilon_i} \frac{\delta E_{xc}}{\delta \phi_i(\mathbf{r}')} + c.c. \right] . \tag{3.79}$$

where $\phi_i$ are the KS orbitals. Compared to standard DFT, where the xc potential is calculated with the help of the electronic density, in OEP, one should solve Eq. (3.78) with the orbital dependent inhomegeneity (3.79) in order to obtain $v_{xc}(\mathbf{r})$ that can be inserted in the KS equations. The two sums including unoccupied states make the calculation extremely cumbersome compared to the DFT ones. The OEP method has been applied to the Fock exchange term $E_x[\{\phi_i\}]$: it is the so-called exact exchange (EXX) [123]. The exchange term is an explicit functional of the orbitals only. Therefore, one can drop the first term in expression (3.79). Besides, $E_x$ is a function of the occupied states only. Still, a sum over unoccupied states is still present: the solution of the EXX equation is much more involving than the resolution of HF[4]. EXX total energies are usually higher than HF ones: despite using the HF energy functional, EXX adds the constraint of the orbital satisfying the KS equation with a local potential. However, the OEP procedure does

[4] A further approximation allows one to replace the sum over unoccupied states by a sum over occupied states: the Krieger-Li-Iafrate (KLI) approximation. It considers that the eigenvalue difference in the denominator of $\Lambda_{xc}$ can be approximated by some average $\Delta\epsilon$. Then the numerator of the fraction can be estimated easily using closure relation. [124]

not destroy the self-interaction free properties of the HF energy functional, therefore EXX cancels exactly the self-interaction and might be a suitable method for the case of systems involving localized states.

As EXX is based on KS formalism, it should not give the correct bandgap. Nevertheless, good agreement of the Kohn-Sham bandgap calculated within EXX with respect to experiments has been reported in the literature for *sp* semiconductors [125, 126]. When calculations were performed on other materials than *sp* semiconductors such as noble gas [127], one found $E_g^{EXX} \neq E_g$. This trend was confirmed by calculations adding random phase approximation correlation to EXX [128, 129], which results in even smaller bandgaps closed to LDA bandgaps. We just remind that the difference between the experimental bandgap and the KS bandgap lies in the exchange and correlation discontinuity. If one adds $\Delta_{xc}^{EXX}$ to $E_g^{EXX}$, one obtains very large bandgaps, comparable to the HF ones [128]. Therefore, the agreement between the EXX KS bandgap and the experimental bandgaps was only incidental and strongly system dependent. In EXX, there is an important cancellation of error between the neglect of the exchange and correlation discontinuity and the lack of correlation.

In conclusion, the introduction of exact exchange in the energy functional results in the exact cancellation of the self-interaction. However, it yields no significant improvement in $E_g$ compared to explicit functionals of the density. This limitation stems from the locality of the KS formalism: the exchange and correlation potential is local (*i.e.* it depends only on $\mathbf{r}$). In order to go beyond the KS formalism, the Generalized Kohn Sham (GKS) [130] formalism has been introduced. It relies on the one-particle density matrix

$$\gamma(\mathbf{r}, \mathbf{r}') = \sum_i \phi_i(\mathbf{r})\phi_i^*(\mathbf{r}')\,. \tag{3.80}$$

Variation of the total energy is then realized over the one-particle density matrix $\delta E/\delta\gamma(\mathbf{r}, \mathbf{r}')$, yielding non local xc potential $v_{xc}(\mathbf{r}, \mathbf{r}')$. The requirement of high non-locality of the exchange and correlation potential may have been foreseen in the previous discussion for example by considering the form of the electron-electron interaction in Eq. (3.1). However it deserves some clarification: changes in the density at position $\mathbf{r}$ should induce significant variations of the xc potential at position $\mathbf{r}$', very far from initial point $\mathbf{r}$. The GKS improves the KS formalism in that respect. In the original paper by Seidl *et al.*, the GKS formalism was applied to screened exchange sX. The non-local potential corresponding to the Fock exchange term was then

derived as[5]

$$v_x^{GKS}(\mathbf{r},\mathbf{r}') = -\sum_i \frac{\phi_i(\mathbf{r})\phi_i^*(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,. \tag{3.81}$$

Later, it was generalized to the Hybrids [131]. The nonlocal hybrid potential reads hence:

$$v_{xc}(\mathbf{r},\mathbf{r}') = b \times v_x^{GKS}(\mathbf{r},\mathbf{r}') + (1-b) \times v_x^{KS}(\mathbf{r}) + v_c^{KS}(\mathbf{r}) \tag{3.82}$$

The improvement of the bandgaps is remarkable. This is due to the inclusion of some part of the xc discontinuity $\Delta_{xc}$ inside $E_g^{EXX}$. Let us consider first an expression of the eigenvalue $\epsilon_i^{GKS}$, using first order perturbation theory with respect to $\epsilon_i^{KS}$:

$$\epsilon_i^{GKS} = \epsilon_i^{KS} + \langle\psi_i^{KS}|\Delta\hat{v}_x|\psi_i^{KS}\rangle\,, \tag{3.83}$$

where $\psi_i^{KS}$ denotes the $i$th KS orbitals and $\Delta\hat{v}_x = v_x^{GKS}[\{\phi_i\}] - v_x^{KS}[n]$. The GKS bandgap $E_g^{GKS}$ is expressed as

$$E_g^{GKS} = E_g^{KS} + \langle\psi_{N+1}^{KS}|\Delta\hat{v}_x|\psi_{N+1}^{KS}\rangle - \langle\psi_N^{KS}|\Delta\hat{v}_x|\psi_N^{KS}\rangle\,, \tag{3.84}$$

where the second and third term can be viewed as the exchange discontinuity $\Delta_x$. Therefore, $E_g^{GKS}$ incorporates a part of $\Delta_{xc}$, which explains why the bandgap within GKS formalism is better. For instance, applying the formalism to PBE0 cures partially the usual DFT underestimation of the bandgap: for small- to medium-bandgap, PBE0 overestimates the bandgap while it underestimates it for large bandgap. The agreement with experiments is however much better compared to LDA [132].

**In conclusion**, we have presented the general concept of a hybrid energy functional. The admixture of HF not only corrects the self-interaction problem but also with the critical interplay of the GKS formalism improves the bandgap. However, while the addition of bare Coulomb interaction (i.e. the unscreened Fock term) may seem suitable for finite systems where the long range interaction is not well described by the short range LDA or GGA, for solids it is more arguable since the Coulomb interaction is strongly screened by the large number of electrons. In the next section, we present the screened hybrid functionals that have encountered great success in the last years in order to describe bandgaps in solids.

[5]Note that this is not equal to the Fock term of HF since the $\phi_i$ are GKS orbitals.

### 3.9.3 The concept of Range separation

The concept of range separation which leads to screened hybrids, was first applied to finite systems. It originates from the observation that the main deficiency of the local or semi local functionals is their bad description of long range behavior which is of outmost importance in finite systems. For example, $\lim_{\mathbf{r}\to\infty} v_{xc}^{LDA}(\mathbf{r}) \propto e^{-\alpha r}$ in the asymptotic regime of a finite systems while the real exchange and correlation potential should go like the bare Coulomb interaction -1/r. The same issue stands for the long range correlation. In order to remedy the situation, Savin and co workers [133, 134] initiated the idea of range separation. The description of the electron-electron interaction potential is an extremely exhausting task due to the singularity of the Coulomb potential at $r \to 0$. On the other hand, the long range exchange and dynamical correlation *i.e.* at $\mathbf{r} \to \infty$ might be necessary to describe certain properties of the system e.g. Rydberg states. The idea of Savin and co workers lies in splitting the long range and short range problems, and choose the *best* methods to treat either problem depending on their accuracy and their computational load. Such splitting strategy was particularly successful on finite systems [135, 136]: since standard DFT functionals are unable to describe long range interactions that are of primer importance in finite systems, more advanced methods such as configuration interaction were used to describe the long range problem while standard DFT methods took care of the description of the short range problem.
The previous analysis breaks down for extended systems: in extended systems, electron-electron interaction is screened. An unscreened long range component leads to pathologies (e.g. the density of states of the homogeneous electron gas at the Fermi level vanishes in Hartree-Fock) whereas middle and short range components are essential. In this spirit, Heyd, Scuseria and Ernzerhof presented a new functional called HSE03 based on the range separation but considering that long range components should be evaluated through a local KS potential [137]. By evaluating only the short range contribution to the exact exchange and thus implicitly screening the Coulomb potential, they have literally opened the way to the use of hybrid functionals to solid-state physics. The exchange and correlation energy functional reads

$$E_{\text{xc,HSE}} = \frac{1}{4}E_{\text{x,HF}}^{sr} + \frac{3}{4}E_{\text{x,GGA}}^{sr} + E_{\text{x,GGA}}^{lr} + E_{\text{c,GGA}} \,. \tag{3.85}$$

In this case, the partition of the Coulomb interaction is obtained using to the error function (erf) and its complementary (erfc) that was introduced by

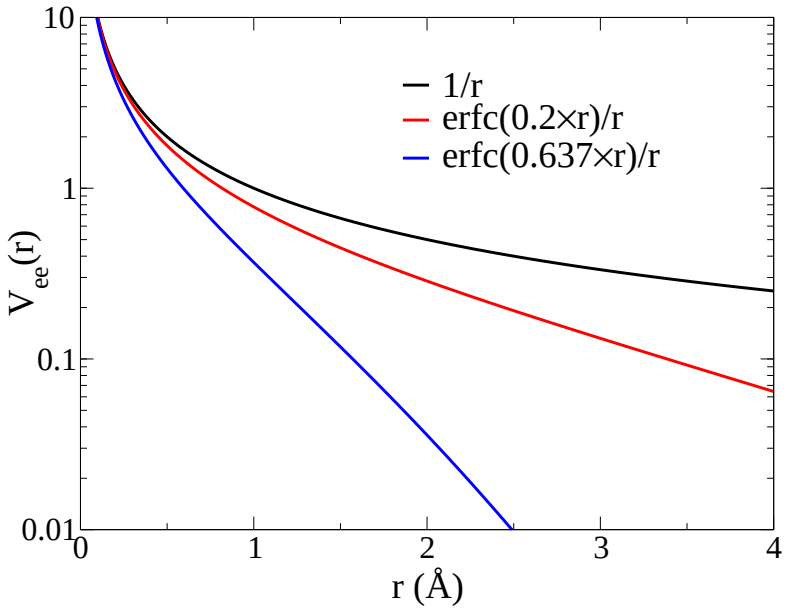


**Figure 3.3:** Comparison between the bare coulomb potential and the short range ersatz used in HSE06.

Leininger *et al* [134]:

$$\frac{1}{r} = f^{LR}(r) + f^{SR}(r) = \frac{\mathrm{erf}(\omega \mathrm{r})}{r} + \frac{\mathrm{erfc}(\omega \mathrm{r})}{r}\,, \tag{3.86}$$

where $\omega$ is the splitting parameter. The choice of the error function may look totally arbitrary. However, we know that the Fourier transform of an error function is a gaussian. Therefore it is not surprising that chemists first introduced this scheme for computational convenience [134]. Nevertheless, another function might be more suitable in the case of solid state physics. Values of $\omega$ found in the literature usually are $\omega = 0.3$ Å$^{-1}$ (HSE03) and $\omega = 0.2$ Å$^{-1}$ (HSE06). HSE hybrids show important improvement compared to PBE0 particularly for the prediction of small- to medium-gap [132]. Most of HSE06 results are independent of the choice of the splitting parameter to the only exception of the bandgap [138]. This important observation is one of the main drawbacks of HSE06 in order to estimate bandgaps: there is no systematic behavior of HSE06 compared to experiments.

One should notice that HSE06 only mixes the short range part of exact exchange. But one may wonder: *how short is the short range ?*. Let us consider an *optimal partitioning* of the Coulomb potential. Optimal refers to the case where the second term of expression (3.86) becomes physically insignificant. For example, if the second term is a slowly enough varying

potential (*i.e.* it has a rapidly decaying Fourier transform), it will do as little as shift the energy origin. Thereby, Lee *et al.* [139] defined a measure of the optimal partitioning

$$Z[\mathrm{erf}] = \int \mathrm{erf}(\omega r) d\mathbf{r} + \int k^2 \left| \mathcal{F}\left( \frac{\mathrm{erfc}(\omega r)}{r} \right) \right|^2 d\mathbf{k}\,, \tag{3.87}$$

where $\mathcal{F}$ stands for the Fourier Transform. Minimizing expression (3.87), Lee and co workers obtained the optimal value of $\omega = 0.637$ Å$^{-1}$, which differs significantly from the one obtained from the fitting procedure. Figure 3.3 compares $f^{SR}(\omega = 0.2)$ and $f^{SR}(\omega = 0.637)$ to the coulombian $1/r$. While the blue curve should retain only the short range part of the Coulomb potential, one can realize that the short-range component of HSE06 (red curve) ressembles much more the true 1/r (black curve) than the *true* short range. HSE06 does not contain only short range of exact exchange but also some long range components up to 5 Å. This characteristic length is of the order of the second nearest neighbour distance in most solids and compares well with the typical range of interaction in a solid [140].

**In this section**, we presented the state of the art of screened hybrid functionals. We explained the physical approximations that make the screened hybrids so successful in treating solids. The overall ground state properties are improved within HSE due to the fact that it cancels partially the self-interaction problem. Besides, HSE yields improved bandgap compared to KS due to the short range exchange interaction and most importantly to the GKS formalism. However, bandgaps are still not reliable within both formalisms and one needs a reference method in order to calculate accurate bandgaps. For this purpose, we will turn to Many Body Perturbation Theory (MBPT).

# Chapter 4

# Many Body Perturbation Theory

## 4.1 Introduction to Green's Functions

Let us consider a linear differential operator $\mathcal{L}$ such that the inhomegeneous equation $\mathcal{L}\phi(\mathbf{r}) = V(\mathbf{r})$ is to be solved. One defines Green's function as the solution of the equation

$$\mathcal{L}G(\mathbf{r}, \mathbf{r'}) = \delta(\mathbf{r} - \mathbf{r'})\,. \tag{4.1}$$

Green's function $G$ is a two-variable function and must have a discontinuity according to $\mathbf{r}$ in order to obtain the $\delta$ function on the right hand side of Eq. (4.1). Solutions of Eq. (4.1) replacing the delta function with any arbitrary potential $V$ can be expressed easily as a function of $G$

$$\phi(\mathbf{r}) = \int d\mathbf{r'} G(\mathbf{r}, \mathbf{r'}) V(\mathbf{r'})\,. \tag{4.2}$$

There is a clear practical advantage of using Green's functions: by solving once Eq. (4.1) one can obtain the solution of $\mathcal{L}\phi = V$ for any perturbation V by simply integrating Green's function with the inhomogeneity $V$. As a two-point quantity, the amount of information contained in $G$ is much more than in $\phi$. But at this cost, one can solve the differential equation for any inhomogeneity.

Another popular representation of a Green's function is using the resolvent formalism that allows one to use powerful complex variable analysis tool. The Green's function is expressed as

$$[z - \mathcal{L}]G(\mathbf{r}, \mathbf{r'}, z) = \delta(\mathbf{r} - \mathbf{r'})\,, \tag{4.3}$$

where $z$ is a complex variable. Assuming that $\mathcal{L}$ is self-adjoint, it admits a complete set of eigenfunctions $\{|\phi_n\rangle\}$ with the following eigenvalues $\{\lambda_n\}$. G

can be written as a function of the resolvent

$$\forall z \notin \{\lambda_n\}, G(z) = \frac{1}{z-\mathcal{L}} = (z-\mathcal{L})^{-1}\,. \tag{4.4}$$

Inserting the identity as a sum of projector $|\phi_n\rangle\langle\phi_n|$, one can obtain the following expression for $G(z)$

$$G(z) = \sum_n \frac{|\phi_n\rangle\langle\phi_n|}{z-\lambda_n}\,. \tag{4.5}$$

The polar structure of $G$ displays the discrete eigenvalue spectrum of its corresponding operator $\mathcal{L}$.

**In conclusion**, we present the mathematical definition of Green's function. Green's functions we will introduce and deal with in the subsequent section are slightly different from the ones we present in this section: besides their mathematical definition, they also have a physical meaning.

## 4.2 Green's function and Many Body Theory

Within the framework of second quantization and the Heisenberg picture, the one particle time ordered equilibrium Green's function is defined as

$$G(1,2) = -i\langle N|T\left[\hat{\psi}(1)\hat{\psi}^\dagger(2)\right]|N\rangle\,, \tag{4.6}$$

where $|N\rangle$ is the N-electron ground state vector and $\hat{\psi}$ ($\hat{\psi}^\dagger$) is the annihilation (creation) field operator. Index 1 stands for the compact form of $(\mathbf{r}_1, t_1, \sigma_1)$. $T$ is the Wick time ordering operator which orders all the operators with increasing time starting from the left.

$$G(1,2) = \begin{cases} -i\langle N|\hat{\psi}(1)\hat{\psi}^\dagger(2)|N\rangle & \text{if } \mathrm{t}_1 > \mathrm{t}_2\,. \\ i\langle N|\hat{\psi}^\dagger(2)\hat{\psi}(1)|N\rangle & \text{if } \mathrm{t}_1 < \mathrm{t}_2\,. \end{cases} \tag{4.7}$$

Let us consider the physical meaning of $G$. The field operators create and annihilate electrons. In the case $t_1 > t_2$, the Green's function represents therefore the probability amplitude to find an electron at time $t_1$, position $\mathbf{r}_1$ and with spin $\sigma_1$ when an electron is put at time $t_2$, position $\mathbf{r}_2$ and with spin $\sigma_2$. Therefore, one often refers to Green's function as the operator which describes the propagation an electron from $(\mathbf{r}_2, t_2, \sigma_2)$ to $(\mathbf{r}_1, t_1, \sigma_1)$ in the system. Alternatively, if $t_1 < t_2$, it describes the propagation of a hole from $(\mathbf{r}_1, t_1, \sigma_1)$ to $(\mathbf{r}_2, t_2, \sigma_2)$. Because $G$ is formed by two field operators, any

operator that can be expressed in terms of two field operators is a function of Green's function. For instance, let us consider a single particle operator that will be expressed within the second quantization framework as

$$\mathcal{J} = \int d12 \hat{\psi}^\dagger(1) J(1,2) \hat{\psi}(2) \,. \tag{4.8}$$

The expectation value of such operator is then

$$\langle N|\mathcal{J}|N\rangle = \int d12 J(1,2) G(1,2) \tag{4.9}$$

One could notice the similarity of this expression with the definition of Eq. (4.2). Indeed, mathematically speaking, $G$ defined in Eq. (4.6) is Green's function of a Hamiltonian that can be expressed in quadratic term of $\hat{\psi}$ and $\hat{\psi}^\dagger$[1]. One possible example is the definition of the density operator

$$\hat{\rho}(1) = \hat{\psi}^\dagger(1)\hat{\psi}(1) \,. \tag{4.10}$$

Its expectation value is simply given by the diagonal element in space and time of G(1,2)

$$\langle N|\hat{\rho}(\mathbf{r},\sigma)|N\rangle = G(\mathbf{r},\mathbf{r},t,t^+,\sigma,\sigma) \,. \tag{4.11}$$

We will now introduce the so-called Lehmann representation which is an analog of the resolvent formalism presented in sec. 4.1 . For sake of clarity, we will restrict ourselves to the case of a time-independent external potential (therefore the Green's function only depends on $\tau = t_1 - t_2$) and drop the spin variable $\sigma$ (we are not interested in magnetic systems in this thesis). Eq. (4.7) can be written as

$$\begin{aligned} iG(\mathbf{r}_1,\mathbf{r}_2,\tau) =& \theta(\tau)\langle N|\hat{\psi}(t_1,\mathbf{r}_1)\hat{\psi}^\dagger(t_2,\mathbf{r}_2)|N\rangle & (4.12)\\ & - \theta(-\tau)\langle N|\hat{\psi}^\dagger(\mathbf{r}_2,t_2)\hat{\psi}(\mathbf{r}_1,t_1)|N\rangle \,. & (4.13) \end{aligned}$$

We can introduce two identity relations involving two different Hilbert spaces $\{|N+1,i\rangle\}$ and $\{|N-1,i\rangle\}$. Thus, the identity will be expressed as

$$\mathcal{I} = \sum_i |N+1,i\rangle\langle N+1,i| = \sum_i |N-1,i\rangle\langle N-1,i| \,. \tag{4.14}$$

Before inserting the identity, one should remember that

$$e^{i\mathcal{H}t}|N+1,i\rangle = e^{iE_{N+1,i}t}|N+1,i\rangle \,. \tag{4.15}$$

[1] It stems from the fact that the single-particle Green's function is composed of only two field operators.

Then, expression (4.13) becomes

$$iG(1,2) = \theta(\tau) \sum_i f_i(\mathbf{r}_1) f_i^*(\mathbf{r}_2) e^{i(E_N - E_{N+1,i})\times\tau} \tag{4.16}$$

$$- \theta(-\tau) \sum_i g_i^*(\mathbf{r}_2) g_i(\mathbf{r}_1) e^{i(E_N - E_{N-1,i})\times\tau} \,, \tag{4.17}$$

where $f_i$ and $g_i$ are called Lehmann's amplitudes

$$f_i(\mathbf{r}) = \langle N|\hat{\psi}(\mathbf{r})|N+1,i\rangle \,, \tag{4.18}$$

$$g_i(\mathbf{r}) = \langle N-1,i|\hat{\psi}(\mathbf{r})|N\rangle \,, \tag{4.19}$$

Differences of total energies like $E_{N+1,i} - E_N$ or $E_N - E_{N-1,i}$ correspond to excitation energies $\epsilon_i$, namely

- $\epsilon_i = E_{N+1,i} - E_N$ is the energy required to add an electron to the ground-state; the system will then be in a state $i$ (ground or excited state). This addition excitation energy corresponds to the electron affinity $A$ when $i$ is the N+1 electron ground state.

- $\epsilon_i = E_N - E_{N-1,i}$ is the energy required to remove one electron; and leave the system in a state $i$: this removal excitation energy corresponds to the ionization energy $I$ when $i$ is the N-1 electron ground state.

For a metal, both excitation energies are equal and define the chemical potential $\mu$. For the case of an insulator, the mismatch between A and I defines the bandgap $E_g = I - A$ and there is a different chemical potential for electrons and holes. The energy differences appearing in the phase of the Green's function are hence electron addiation and removal energies, with a well defined physical meaning. They will in particular allow us to calculate bandgaps. This derivation gives also a hint why the Koopmans' theorem fails for Kohn Sham DFT. Excitation energies appear upon inclusion of excited states $|N \pm 1\rangle$ which corresponds to the removal or addition of an electron. Energies obtained from DFT are however Lagrange multipliers that constrain the conservation of number of particles in the system. One can immediately see the impossibility for Kohn-Sham DFT to fulfill Koopman's theorem.

We can further simplify expression (4.17) by generalizing the excitation energy as a function of the chemical potential $\mu$

$$\begin{array}{lll} \text{if} & \epsilon_i > \mu & \epsilon_i = E_{N+1,i} - E_N \,. \\ \text{if} & \epsilon_i < \mu & \epsilon_i = E_N - E_{N-1,i} \,. \end{array} \tag{4.20}$$

Consequently, it is useful to define a generalized form of Lehmann's amplitude $\Psi_i(\mathbf{r})$ as

$$\Psi_i(\mathbf{r}) = \begin{cases} f_i(\mathbf{r}) & \text{if } \epsilon_\mathrm{i} > \mu\,, \\ g_i(\mathbf{r}) & \text{if } \epsilon_\mathrm{i} < \mu\,. \end{cases} \tag{4.21}$$

One-particle Green's function is then expressed as

$$iG(\mathbf{r}_1, \mathbf{r}_2, \tau) = \sum_i \left[\theta(\tau)\theta(\epsilon_i - \mu) - \theta(-\tau)\theta(\mu - \epsilon_i)\right] \Psi_i(\mathbf{r}_1)\Psi_i^*(\mathbf{r}_2) \times e^{-i\epsilon_i \times \tau}\,. \tag{4.22}$$

The Fourier transform of such a quantity can be performed only if a small imaginary part $\eta$ is added to the exponential. Depending on the sign of $\tau$, the sign of $\eta$ will change: for removal energy $\eta$ is negative and for addition energy it is positive [2]. Finally, Lehmann's representation of $G$ after Fourier transform reads

$$G(\mathbf{r}_1, \mathbf{r}_2, \omega) = \sum_i \frac{\Psi_i(\mathbf{r}_1)\Psi_i^*(\mathbf{r}_2)}{\omega - \epsilon_i + i\eta \mathrm{sign}(\epsilon_i - \mu)}\,. \tag{4.23}$$

The pole structure of $G$ gives hence access to removal and addition energies. One should note that both excitation energies lie along the real axis, but the sign of $\eta$ (defined by the sign of $\epsilon_i - \mu$) positions them below (for addition energies) or above (for removal energies) the real axis. There is no pole lying in the energy range defined by the bandgap.
From the frequency dependent form of G, we can define the so-called spectral function $\mathrm{A}(\mathbf{r}_1, \mathbf{r}_2, \omega)$:

$$A(\mathbf{r}_1, \mathbf{r}_2, \omega) = \frac{1}{\pi}\mathrm{sign}(\mu - \omega)\mathrm{Im}G(\mathbf{r}_1, \mathbf{r}_2, \omega)\,. \tag{4.24}$$

Working with spectral functions has advantages compared to working with the full Green's function: one has to deal with a real-valued function instead of the complex-valued $G$. Because of Kramers-Kronig relations for $G$ (see Eq. (3.62)) $A$ contains the full amount of information of G. For example, one can access the ground-state electronic density using the relation:

$$\int_{-\infty}^{\mu} d\omega A(\mathbf{r}_1, \mathbf{r}_1, \omega) = \rho(\mathbf{r}_1)\,, \tag{4.25}$$

or the density of states $N(\omega) = Tr(A(\mathbf{r}_1, \mathbf{r}_2, \omega))$, or the ground-state total energy via the Galitskii Migdal [141] formula:

$$E = \frac{1}{2}\int d\mathbf{r}_1 \lim_{\mathbf{r}_2 \to \mathbf{r}_2} \int_{-\infty}^{\mu} d\omega \left[\omega + h(\mathbf{r}_1)\right] A(\mathbf{r}_1, \mathbf{r}_2, \omega)\,. \tag{4.26}$$

[2] This is due to the product of the two step functions. The time range of integration of the Fourier transform will be different depending on the position of $\epsilon_i$ according to $\mu$.

Experimentally speaking, Angle Resolved Photoemission Electron Spectra (ARPES) are closely related to the spectral function $A(\mathbf{k},\omega)$. Therefore, approximations made to construct $G$ can be tested by comparing the spectral function to ARPES data.

In conclusion, we have derived the spectral representation of $G$, which gives access to removal and addition energies and consequently to the bandgap. We have also introduced the spectral function which will be very useful when it comes to explain the concept of *quasiparticle*. In the next section, we will introduce an equation of motion in order to determine $G$.

## 4.3 Towards the Quasi-particle picture

Equation (4.6) for the Green's functions, is well defined, but one would need the N electron ground state which is in general not known. Therefore, one needs another equation to solve for $G$. It is common practice in quantum mechanics in order to evaluate an operator to use the Heisenberg equation also called equation of motion (EOM) as it describes the time-dependence of an operator by means of the commutator of the operator with the Hamiltonian of the system. For instance, the equation of motion of the field operators within the Heisenberg representation is given by

$$i\frac{\partial\hat{\psi}}{\partial t} = [\hat{\psi},\mathcal{H}]\,, \tag{4.27}$$

$$i\frac{\partial\hat{\psi}^\dagger}{\partial t} = [\hat{\psi}^\dagger,\mathcal{H}]\,. \tag{4.28}$$

Once again, the form of the $\mathcal{H}$ in terms of field operators will be of outmost importance. Assuming that $\mathcal{H} = H_0$, where $H_0$ is a non-interacting fermion Hamiltonian, its second quantization form is simply quadratic in $\hat{\psi}$ and $\hat{\psi}^\dagger$

$$H_0 = \int d\mathbf{r}_1 \hat{\psi}^\dagger(\mathbf{r}_1) h_0(\mathbf{r}_1)\hat{\psi}(\mathbf{r}_1) \tag{4.29}$$

Working around Eq. (4.28), the equation of motion for G is then simply

$$\left[i\frac{\partial}{\partial t_1} - h_0(\mathbf{r}_1)\right] G_0(\mathbf{r}_1,t_1,\mathbf{r}_2,t_2) = \delta(\mathbf{r}_1-\mathbf{r}_2)\delta(t_1-t_2)\,. \tag{4.30}$$

The subscript 0 has been added to Green's function symbol as it is a non-interacting Green's function. The spectral function of $G_0$ simply reads

$$A(\mathbf{r}_1,\mathbf{r}_2,\omega) = \sum_i \phi_i(\mathbf{r}_1)\phi_i^*(\mathbf{r}_2)\delta(\omega-\epsilon_i)\,, \tag{4.31}$$

where Lehmann's amplitudes $\phi_i$ are the independent-particle wavefunctions of $h_0$ and $\epsilon_i$ the corresponding eigenenergies. Its expression simplifies further the matrix element $\langle i|A|i\rangle$ of the spectral function for a given state $|i\rangle$:

$$\langle i|A|i\rangle = \delta(\omega - \epsilon_i)\,. \tag{4.32}$$

The spectral function of non-interacting Green's function is then a simple delta-peak centered at $\omega = \epsilon_i$. We will see that the interacting case will be more complicated and results in different features of the spectral function. When the electron-electron interaction is switched on, the equation of motion of G is significantly different. The second quantization representation of the electron-electron interaction involves four field operators:

$$\hat{U}_{ee} = \frac{1}{2}\int d\mathbf{r}_1 d\mathbf{r}_2 \hat{\psi}^\dagger(\mathbf{r}_1)\hat{\psi}^\dagger(\mathbf{r}_2)v_c(\mathbf{r}_1,\mathbf{r}_2)\hat{\psi}(\mathbf{r}_2)\hat{\psi}(\mathbf{r}_1)\,, \tag{4.33}$$

Thus, it is natural to introduce two-particle Green's function $G_2$ describing two particles at the same time as the electron-electron interaction involves two particles. The two-particle Green's function $G_2$ defined in (4.34) contains four field operators:

$$i^2 G_2(1,2,1',2') = \langle N|T\left[\hat{\psi}(1)\hat{\psi}(2)\hat{\psi}^\dagger(2')\hat{\psi}^\dagger(1')\right]|N\rangle\,, \tag{4.34}$$

With (4.28), (4.29) and (4.34), the equation of motion of $G$ depends on $G_2$ as

$$\left[i\frac{\partial}{\partial t_1} - h_0(\mathbf{r}_1)\right]G(1,2) + i\int d3 v_c(1,3)G_2(1,3^+,2,3^{++}) = \delta(1,2)\,, \tag{4.35}$$

where $+$ and $++$ indicate positive infinitesimalS added to the time variable $t$. The interacting Green's function is hence not true a Green's function mathematically speaking. Indeed, its equation of motion presents an inhomogeneity involving the two-particle Green's function. The two-particle Green's function depends through a similar equation of motion on the three-particle one $G_3$ and so on. Thus, the equation obviously would require lot of effort to be solved without being reformulated. From a mathematical perspective, the inhomogeneity in Eq. (4.35) should be transformed in order to make $G$ appear. The situation is not as bad as one may think. Indeed, not the full two-particle Green's function is required to solve the EOM for $G$: only a three point contraction of $G_2$ enters Eq. (4.35). Before doing any further approximation, it is necessary to describe the three-point $G_2$ and more particularly its polar structure. The Wick operator present in $G_2$ yields multiple contributions depending on the time ordering [142]. The three-point $G_2$ is

constrained such as $t_{3+} \leq t_{3++}$. Moreover, the Coulomb potential is time independent so it has a time delta-function making $t_1$ and $t_3$ equal. Then, the time constraint applied to $G_2$ in the homogeneity of Eq. (4.35) reads as

$$t_1 = t_3 \leq t_{3+} \leq t_{3++}\,. \tag{4.36}$$

Three out of the four field operators are then ordered. As a result, only two terms will arise from the two possible time ordering of $t_2$ with respect to the other

$$\begin{aligned} iG_2(1,3^+,2,3^{++}) =& \langle N|\hat{\psi}^\dagger(2)\hat{\psi}(1)\hat{\psi}(3^+)\hat{\psi}^\dagger(3^{++})|N\rangle\theta(t_2-t_1) \\ &+ \langle N|\hat{\psi}(1)\hat{\psi}(3^+)\hat{\psi}^\dagger(3^{++})\hat{\psi}^\dagger(2)|N\rangle\theta(t_3-t_2)\,. \end{aligned} \tag{4.37}$$

Similarly to what has been done for the derivation of Lehmann's representation of $G$, one should introduce the identity of Hilbert spaces $\{|N-1,i\rangle\}$ and $\{|N+1,i\rangle\}$ . However, here we are only interested by the pole structure of $G$. The poles originate from the time-dependent exponential using the reduced time parameter $\tau = t_i - t_j$ with $i \neq j$. The latter argument will result in many of the time exponentials being 1 because $t_1 = t_3 \simeq t_{3+} \simeq t_{3++}$ up to an infinitesimal. Therefore, one needs to insert $\mathcal{I}$ only between $\hat{\psi}^\dagger(2)$ and its adjacent field operator. Setting $\tau = t_2 - t_3$, expression (4.37) becomes

$$\begin{aligned} iG_2(1,3^+,2,3^{++}) =& \sum_i \langle N|\hat{\psi}^\dagger(2)|N-1,i\rangle\langle N-1,i|\hat{\psi}(1)\hat{\psi}(3^+)\hat{\psi}^\dagger(3^{++})|N\rangle\theta(\tau) \\ &+ \sum_j \langle N|\hat{\psi}(1)\hat{\psi}(3^+)\hat{\psi}^\dagger(3^{++})|N+1,j\rangle\langle N+1,j|\hat{\psi}^\dagger(2)|N\rangle\theta(-\tau)\,. \end{aligned} \tag{4.38}$$

Using the Heisenberg representation, one gets

$$\begin{aligned} iG_2(1,3^+,2,3^{++}) =& \sum_i \langle N|\hat{\psi}^\dagger(2)|N-1,i\rangle\langle N-1,i|\hat{\psi}(1)\hat{\psi}(3^+)\hat{\psi}^\dagger(3^{++})|N\rangle e^{-i(E_N-E_{N-1,i})\tau}\theta(\tau) \\ &+ \sum_j \langle N|\hat{\psi}(1)\hat{\psi}(3^+)\hat{\psi}^\dagger(3^{++})|N+1,j\rangle\langle N+1,j|\hat{\psi}^\dagger(2)|N\rangle e^{i(E_N-E_{N+1,i})\tau}\theta(-\tau)\,. \end{aligned} \tag{4.39}$$

The inhomogeneity term in Eq. (4.35) has hence the same polar structure as $G$, which is a further argument in favor of putting $G$ into evidence in the expression of $G_2$. In the next section, we will present an exact formulation involving a new operator $\Sigma$ called self-energy.

### 4.3.1 $\Sigma$ and Dyson equation

As previously demonstrated, the inhomogeneity in Eq. (4.35) has a polar structure close to the one of $G$. It is then natural to define a formal operator $\Sigma$ such that

$$i\int d3v_c(1,3)G(1,3^+,2,3^{++}) = \int d3\Sigma(1,3)G(3,2)\,, \tag{4.40}$$

where $\Sigma$ is a formal operator called the self energy. The definition yields a new equation of motion for $G$

$$\left[i\frac{\partial}{\partial t_1} - h_0(\mathbf{r}_1)\right]G(1,2) - \int d3\Sigma(1,3)G(3,2) = \delta(1,2)\,. \tag{4.41}$$

This equation can be related to the equation of motion for non-interacting Green's function by using the resolvent formalism presented in the introduction of this chapter. The frequency fourier transform of $G_0$ can be written as $G_0 = (\omega - h_0)^{-1}$ and consequently $G_0^{-1}(\omega) = \omega - h_0$. The Eq. (4.41) can be rewritten as

$$G = G_0 + G_0\Sigma G\,. \tag{4.42}$$

Eq. (4.42) is called a Dyson equation [142]. The knowledge of $G_0$ and $\Sigma$ allows one to access the pole structure of interacting Green's function. All the interaction effects are gathered into $\Sigma$, and for a given $\Sigma$, interactions to all orders are included in $G$. The main difficulty is to find a physically sound approximation to the self-energy. But first, let us have a closer look at the pole structure of the problem.

### 4.3.2 Quasiparticle picture

We will first derive mathematically the quasiparticle equation and then introduce the physical concept of quasiparticle. Expression (4.22) is Lehmann's representation of Green's function for a *finite system*: the excitation energies are discrete quantities. If one considers the *thermodynamic limit* corresponding to an infinite system [3], then the excitation energies will form a continuum of excitations and the discrete sum $\sum_i$ will be replaced by an integral $\int d\epsilon$. $G(\omega)$ is ill defined if $\omega$ is a real frequency, prompting to the introduction of complex variable $z$. The imaginary part of the excitation energies as we will see later are hence not artifacts introduced to reproduce experimental spectra but are formally required to define $G(\omega)$ based on its complex analytic continuation $G(z)$.

[3] The number of particles and the volume $\to \infty$, keeping the ratio of the two (*i.e.* the concentration) constant.

Following Farid [143], let us introduce the function $f(\epsilon)$ with the following form

$$f(\omega) = \sum_i \frac{\theta(\epsilon_i - e_0) - \theta(\epsilon_i - e_1)}{\epsilon_i - \omega} \,. \tag{4.43}$$

The function $f(\omega)$ presents isolated singularities at $\epsilon_i$ which are the excitations energy of the system. It ressembles closely Lehmann's representation of Green's function but also many of the quantities that will be introduced later in this thesis. In the thermodynamic limit, $f(\omega)$ should be replaced by the complex function $\tilde{f}(z)$ with $Im(z) \neq 0$. The analytical continuity holds if $f(\omega) \to f(z)$ when $Im(z) \to 0$. In this limit, the singularities condense into a branchcut. For the sake of clarity, we consider only one branch-cut delimitated by branch points $e_0$ and $e_1$ situated along the real axis. For instance, $G(\omega)$ presents two semi-infinite branch-cuts separated by the bandgap: one infinitesimally above (below) the real axis for removal (addition) energies. As stated previously, $\sum_i$ becomes an integral and the analytical continuation of $f$, $\tilde{f}(z)$ reads

$$\tilde{f}(z) = \ln\left(\frac{z - e_1}{z - e_0}\right) \,. \tag{4.44}$$

The analytical continuation of a logarithm is a standard textbook problem. The function (4.44) is a multi valued function[4] [144]. The n-valued complex function $f(z)$ can be considered as a one-valued complex function $f_n(z)$ over n complex planes called *Riemann sheets* $\mathcal{R}_n$. Riemann sheets are connected along the branch cut of $\{f_i(z)\}$ (here the branch cut is $[e_0, e_1]$). From the physical point of view, multi-valued functions are of limited interest: if one is interested in some specific physical properties then one should determine $\mathcal{R}_i$ such that $f_i(z) \to f(\omega)$ when $z \to \omega$. Thus, $\mathcal{R}_i$ is called the physical Riemann sheet and $f_i(z) = \tilde{f}(z)$ the analytical continuation of f($\omega$) in the complex plane. If one considers the limit of $\tilde{f}(\omega \mp i\eta)$ when $\eta \to 0$, one obtains using the Cauchy theorem to perform the integration $\int d\epsilon$

$$\tilde{f}(\omega \mp i\eta) = ln\left(\frac{|e_1 - \omega|}{|e_0 - \omega|}\right) \mp i\pi\theta(\omega - e_0)\theta(e_1 - \omega) \,. \tag{4.45}$$

Depending on where $\omega$ lies, the analytical continuation of $f(\omega)$ will be real valued (for $\omega > e_1$ or $\omega < e_0$) or complex valued when $e_0 < \omega < e_1$. One can extend this discussion to the case of $\tilde{G}(z)$: when $\epsilon$ lies in the bandgap, $G(z)$ is real valued while it becomes complex valued everywhere else. However,

[4] Let us consider $y = ln(z)$, $y$ is the solution of equation $z = e^y$. Since $e^{2i\pi} = 1$, $y$ can only be determined up to a constant $y_i = ln(|z|) + 2in\pi$.

along the real axis, $\tilde{G}(z)$ is always analytical except at some singularity points. Therefore, we can define a Cauchy relation to express $G(\mathbf{r}, \mathbf{r}', \omega)$

$$G(\mathbf{r}, \mathbf{r}', \omega) = \int_{-\infty}^{\mu} d\omega' \frac{A(\mathbf{r}, \mathbf{r}', \omega')}{\omega - \omega' - i\eta} + \int_{\mu}^{\infty} d\omega' \frac{A(\mathbf{r}, \mathbf{r}', \omega')}{\omega - \omega' + i\eta} \,. \tag{4.46}$$

where $\mu$ is the Fermi level and $A$ is the spectral function. Such decomposition can be particularly useful as one has to deal with real quantities like $A$, instead of complex ones, like $G$. Such analysis holds for all the quantities that could be expanded analytically to the whole complex plane.
In conclusion, we have demonstrated the existence of the analytical continuation of $G(\omega)$ into the complex plane. It permits to work in the thermodynamic limit *i.e.* infinite systems. We draw a special attention on the necessity of working in the physical Riemann sheet: it should be selected by considering the only solution that approaches the real-valued function uniformely. The latter point can be of great importance when one wants to perform the integration to obtain $G$ in the complex plane. For instance, the analytical continuation of $f(\omega)$ into a non physical Riemann sheet would give

$$\tilde{\tilde{f}}(z) = ln\left(\frac{z - e_0}{z - e_1}\right) - 2i\pi \,. \tag{4.47}$$

Up to now, we have defined a mathematically correct framework to treat the case of *infinite systems*. Let us consider the solution of the Dyson equation (4.42) which has been analytically continuated in the physical Riemann sheet:

$$\tilde{G}(z) = \left[\mathcal{I} - \tilde{G}_0(z)\tilde{\Sigma}(z)\right]^{-1} \tilde{G}_0(z) \,. \tag{4.48}$$

The latter expression can be expanded as a perturbation series in terms of $G_0$, $\tilde{\Sigma}$ being the perturbation. However, such series will not converge when $z$ is equal to a pole of $\tilde{G}(z)$, therefore the condition of divergence is given by [5]

$$\det\left[\mathcal{I} - \tilde{G}_0(z)\tilde{\Sigma}(z)\right] = 0 \,. \tag{4.49}$$

Using the properties $\det\left[\tilde{G}_0^{-1}\right] \neq 0$ and that det(A)×det(B)=det(AB), then Eq. (4.49) becomes

$$\det\left[z\mathcal{I} - H_0 + \tilde{\Sigma}(z)\right] = 0 \,. \tag{4.50}$$

Solutions of Eq. (4.50) of the form $z = z_s$ are simply the eigenvalues of the Hamiltonian $H_0 + \tilde{\Sigma}(z)$ which is generally non hermitian. Therefore, following

[5] There is an interesting property of $\tilde{G}(z)$ and $\tilde{G}_0(z)$: they are inversabile due to the absence of zero [143]

Morse and Feshbach [145], we can state that $H_0 + \tilde{\Sigma}(z)$ admits *left* and *right* eigenfunctions $\overline{\Psi}_i(z)$ and $\Psi_i(z)$ such that

$$\left[H_0 + \tilde{\Sigma}(z)\right] \Psi_i(z) = \tilde{E}_i(z)\Psi_i(z)\,, \tag{4.51}$$

$$\left[H_0 + \tilde{\Sigma}^\dagger(z)\right] \overline{\Psi}_i(z) = \tilde{E}_i^*(z)\overline{\Psi}_i(z)\,. \tag{4.52}$$

Eq. (4.52) are called the quasiparticle equation. A biorthonormal condition holds for $\overline{\Psi}$ and $\Psi$: for $\tilde{E}_i(z) \neq \tilde{E}_{i'}(z)$, $\langle \overline{\Psi}_i(z), \Psi_{i'}(z)\rangle = \delta_{i,i'}$. Then, the Green's function reads

$$\tilde{G}(\mathbf{r}_1, \mathbf{r}_2, z) = \sum_i \frac{\Psi_i(\mathbf{r}_1, z)\overline{\Psi}_i^*(\mathbf{r}_2, z)}{z - \tilde{E}_i(z)}\,. \tag{4.53}$$

Eq. (4.53) is called the biorthonormal representation of $\tilde{G}$. Such representation of Green's function differs significantly from the Lehmann representation (4.22): first, the numerator in the biorthonormal form is frequency depedent together with the poles themselves. Besides, the resolution of equation (4.52) should be performed for every value of $z$, which results in a complete set of eigenfunctions and eigenvalues for each $z$. However, noticing that the numerator in the Lehmann representation is frequency independent, one only needs one set of eigenfunctions for all real frequency [146]. Consequently, one can use the Mittag-Leffler theorem to represent $\tilde{G}(z)$ as a function of its residues and poles. The poles of $\tilde{G}(z)$, $z_i$ are obtained by solving

$$z_i = \tilde{E}_i(z_i)\,. \tag{4.54}$$

The function $\tilde{E}_i(z)$ is analytic, following the same argument used for $\tilde{G}(z)$ and $\tilde{\Sigma}(z)$. Besides, $\tilde{E}_i(z)$ is the analytical continuation in the complex plane of

$$E_i(\omega) = \lim_{\eta \to 0} \tilde{E}_i(\omega \mp i\eta)\,, \tag{4.55}$$

where $\mp$ depends on the position of $\epsilon$ compared to the Fermi level $\mu$. Using mathematical considerations (for more details see [143]), Farid proved that equation $E_i(\omega) = \omega$ does not have a solution. Therefore, solution of Eq. (4.54) should be sought in the complex plane. In order to respect the time-reversal symmetry of the system, the reflection condition $\tilde{E}_i(z^*) = \tilde{E}_i^*(z)$ holds. If one assumes that there exists a solution $z_0$ of Eq. (4.54) with $\mathrm{Im}(z_0) \neq 0$ then $z_0^*$ is also a solution of Eq. (4.54). Consequently, such solution violates the principle of causality. Therefore, the excitation energies should be sought in the non physical Riemann sheets.
The fact that these energies have an imaginary part stems not only from the

mathematical arguments developed previously but also has a profound physical meaning. In a system where there exists a continuum of excitation, any excitation will *decay* due to the presence of an infinite number of excitations infinitisemally close to the initial excitation. Thus, one-particle excitation in a infinite system *must* have a finite lifetime.
Going back to the expression of $\tilde{G}(z)$, one would like to calculate the residue of $\tilde{G}(z)$ so that to use Mittag-Leffler theorem [147]. Let us consider expansion of the denominator $z - \tilde{E}(z)$ in a region closed to $z_s$ such that

$$\frac{1}{z - \tilde{E}_i(z)} \approx \frac{Z_i}{z - z_i}, \text{ for sufficiently small } |z - z_i| \tag{4.56}$$

where $Z_s$ is called the renormalization factor

$$Z_i = \left(1 - \left.\frac{\partial \tilde{E}}{\partial z}\right|_{z=z_i}\right)^{-1} \tag{4.57}$$

The biorthonormal expression of Green's function (4.53) becomes then

$$\tilde{G}(\mathbf{r}_1, \mathbf{r}_2, z) = \sum_i Z_i \frac{\Psi_i(\mathbf{r}_1, z_i)\overline{\Psi}_i^*(\mathbf{r}_2, z_i)}{z - z_i}\,. \tag{4.58}$$

In practice, one needs to know $E_i(z)$ in the neighborhood of the poles $z_s$ and compute $\frac{\partial E_i(z)}{\partial z}$. Any static approximation to the self energy*i.e.* will results in $Z_i = 1$, simplifying further Eq. (4.58). As a consequence of the imaginary part of the QP energies, the spectral function of interacting Green's function is not a delta-function anymore but a lorentzian as it includes the finite lifetime of excitation $i$.

$$\langle i|A|i\rangle = \frac{1}{\pi}\mathrm{Im}\left[\frac{\mathrm{Im}(\langle i|\Sigma|i\rangle)}{(\omega - \epsilon_i - \mathrm{Re}(\langle i|\Sigma|i\rangle)^2 + \mathrm{Im}(\langle i|\Sigma|i\rangle)^2}\right]. \tag{4.59}$$

The non interacting excitation energy $\epsilon_i$ is shifted by $\mathrm{Re}(\langle i|\Sigma|i\rangle)$ while the imaginary part of $\langle i|\Sigma|i\rangle$ gives a finite lifetime to the excitation. The corresponding excitation is called *quasiparticle*: it corresponds to an electron together with its interaction with the environment. It can be considered to be a new particle often refered as a *dressed* particle due to the inclusion of some interactions within the quasiparticle formalism. Quasiparticles interact much more weakly between each other than bare electrons, which may reveal extremely useful when doing some expansion.
The pronounced peak in the spectral function corresponds to a *quasiparticle*

excitation. The width of the peaks is given by $|\mathrm{Im}(\langle i|\Sigma|i\rangle)|$: the longer is the lifetime of the excitation, the more the quasiparticle ressembles the bare electron. Thus, the imaginary part of the self-energy plays the role of an inverse lifetime. However, the spectral function of interacting Green's function can present other features apart from the quasiparticle peak called the incoherent part of the spectrum: usually broaden peaks appear away from the quasiparticle peak which correspond to collective excitations. They are called *satellites*. As the spectral function of interacting Green's function fulfills a sum-rule (4.25), the overall integration of $A$ over the whole range of energy must be constant. As a consequence, the presence of satellites will naturally reduces the strength of the quasiparticle peak: one says that the peak has been renormalized. Such renormalization corresponds to transfer of energy from the quasiparticle excitation to the collective modes represented by the satellites. In Fig. 4.1, the real part of $Z_i$ $\mathrm{Re}\{Z_i\}$ represents the integral of $A$ over a region centered around the quasiparticle peak. If one deals with an independent particle picture, $Z = 1$. When the incoherent part of the spectrum becomes predominant *i.e.* 1-Re$Z_i$ becomes large, then the excitation cannot be seen in a particle-like picture but for example as a collective excitation[6]. $Z_i$ is then a useful measure of the one-particle character of the excitation.

In conclusion, we have presented an equation whose eigenergies are the removal and addition energies. Then, solving the QP equation, we have access to the correct excitations energies. However, a last task is necessary, namely we have to find the self-energy.

## 4.4 Hedin's equations and the GW approximation

In order to derive a closed set of equations defining $G$ and $\Sigma$ the Schwinger's functional derivative approach [149] can be used. We generalize the Green's function of a system to the case of the presence of a fully non-local external potential $U(\mathbf{r}_1, \mathbf{r}_2, t)$ [142], the interacting one-particle Green's function can

[6] It is usually said that this case corresponds to a break down of the band-structure picture: indeed, spectral function $A$ should be able to describe the electronic band structure with its quasiparticle peaks. However, when such peaks are particularly renormalized due to the coupling to other excitations, the band-structure tends to disappear and only the spectral function can described the system.

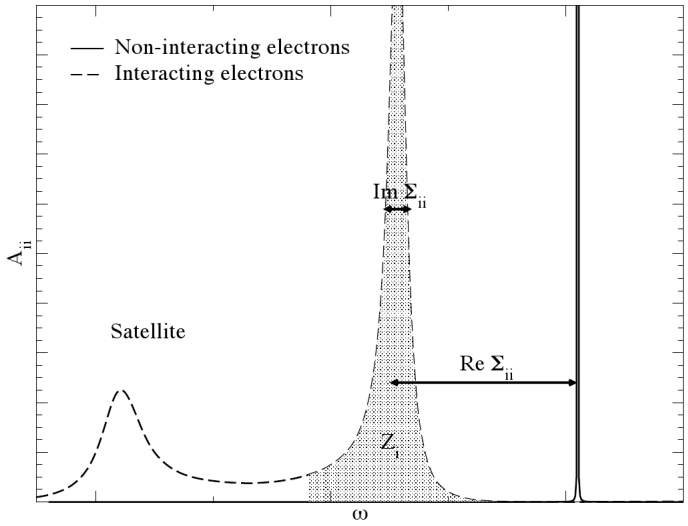


**Figure 4.1:** Matrix elements of the spectral functions of interacting (dashed line) and non-interacting (solid line) electron systems. [148]

then be written in the interaction picture as

$$G(1,2) = -i\frac{\langle N|T[\hat{S}\hat{\psi}(1)\hat{\psi}^{\dagger}(2)]|N\rangle}{\langle N|T[\hat{S}]|N\rangle}\,. \tag{4.60}$$

where $\hat{S}$ is the time evolution operator that is dependent on $U$ such that

$$T[\hat{S}] = T\left[\exp\left(-i\int_{-\infty}^{\infty}dt\int d\mathbf{r}_1 d\mathbf{r}_2\hat{\psi}^{\dagger}(\mathbf{r}_1)U(\mathbf{r}_1,\mathbf{r}_2,t)\hat{\psi}(\mathbf{r}_2)\right)\right]\,. \tag{4.61}$$

The *mysterious* potential introduced by Schwinger may be interpreted in the following way: expressing the equation of motion for $G$, we have noticed that we need to introduce the two-particle Green's function and so on. The physical implication of such observation is that particles moving in the system will interact with other particles of the system and create and destroy two-particles, three-particles excitation and so on along its way. This can be described introducing the concept of polarization of the medium. However, the fact that the propagation of one particle requires the sum over an infinite number of excitations makes a brute force application impossible. One can simulate these processes by introducing the fictitious extrernal potential that will be made vanishing at the end of the calculation.
Indeed by varying the propagator $G$ (4.60) with respect to the external po-

tential $U$ and using classical derivation rule, one gets

$$\delta G = \frac{\langle N|T[\delta\hat{S}\hat{\psi}(1)\hat{\psi}^{\dagger}(2)]|N\rangle}{\langle N|T[\hat{S}]|N\rangle} - G(1,2)\frac{\langle N|T[\delta\hat{S}]|N\rangle}{\langle N|T[\hat{S}]|N\rangle}\,. \tag{4.62}$$

The derivative of the exponential will add two extra field operators to the two already present in the definition of the one-particle Green's function. Therefore, using commutation relations between field operators, one can make $G_2$ appear and obtain the relatively simple relation [7]

$$\frac{\delta G(1,2)}{\delta U(3,4)} = G(1,2)G(4,3) - G_2(1,4,2,3)\,. \tag{4.63}$$

This relation is very useful as it allows us to express $G_2$ in terms of $G$. Therefore, it is now possible to express the self-energy by means of $G$ and its derivatives only. It is also worth noting that the derivative of $G$ with respect to $U$ is a correlation function as it is the difference between the propagations of two interacting particles and the propagations of two independent particles. At first sight, the expression provides much more information than needed. In the QP equation, the two-particle Green's function is only a three point function. In practice, inserting expression (4.63) in Eq. (4.41) results in considering a local potential $U$ in our derivation. We can consider now the exact quantity present in the quasiparticle equation

$$G_2(1,3^+,2,3^{++}) = G(1,2)G(3^+,3^{++}) - \frac{\delta G(1,2)}{\delta U(3)}\,, \tag{4.64}$$

where $U(3) = U(1,3)\delta(1,3)$. Inserting the aforementioned expression (4.64) in Eq. (4.41), one obtains two terms from the integral: one is a classical Hartree potential $-i\int d3v(1,3)G(3,3^+)$ which stems from the first term of the right hand side of Eq. (4.41) and the other term describes all the effects beyond the classical Hartree description and originates from the correlation function $\delta G/\delta U$. We introduce the derivative of the closure relation $GG^{-1} = \mathcal{I}$ as it permits to express $G^{-1}$ and its derivative as a function of $G$ and its derivative.

$$0 = \frac{\delta G}{\delta U}G^{-1} + G\frac{\delta G^1}{\delta U}\,. \tag{4.65}$$

Using relation (4.65), the self-energy $\Sigma$ can be expressed as

$$\Sigma(1,2) = -i\int d34v(1^+,3)G(1,4)\frac{\delta G^{-1}(4,2)}{\delta U(3)}\,. \tag{4.66}$$

[7] We use the compact notation but one should remember that $U$ is local in time, *i.e.*, $t_4 = t_3 + \delta$ where $\delta$ is an infinitesimal.

Schwinger's trick allows one to express the self-energy in terms of the one particle Green's function and its derivative at vanishing local potential $U$. The latter three-point quantity is still unknown but can be approximate in terms of perturbative approaches.

### 4.4.1 Perturbative approaches

There are several ways to solve the problem of finding a *practical* expression for $\Sigma$. We will first investigate two of them: the $G_0$ perturbation method and the $\Sigma$ perturbation method [146, 150].

#### $G_0$ perturbation method

The perturbative expansion is a rather popular method to solve many-body problems: it relies on an expansion in terms of some quantity (such as the Coulomb potential or $G_0$). The expansion is expected to require a limited number of terms to describe the physics of the system. For example, starting from Dyson's equation we obtain the following equation

$$\begin{aligned} G(1,2) = G_0(1,2) - i \int d3d4 G_0(1,3) v(3,4) \\ \times \left[ G_0(4,4+) - \frac{\delta}{\delta U(4)} \right] G(3,2)\,. \end{aligned} \tag{4.67}$$

One can replace $G(3,2)$ in the integrand by the unperturbed Green's function $G_0(3,2)$ as at first order $G(1,2) = G_0(1,2)$. In the presence of an external local potential $U$, $G_0$ satisfies

$$G_0(1,2) = \left[ i\frac{\partial}{\partial t_1} - h_0(\mathbf{r}_1) - U(\mathbf{r}_1) \right]^{-1} \delta(1,2) \tag{4.68}$$

Then Eq. (4.67) reads

$$\begin{aligned} G(1,2) =& G_0(1,2) - i \int d3d4 G_0(1,3) v(3,4) \\ & \times \left[ G_0(4,4+) - \frac{\delta}{\delta U(4)} \right] G_0(3,2) + \text{higher order terms}\,. \end{aligned} \tag{4.69}$$

The two first terms can be expressed exactly as a function of $G_0$, which yields

$$\begin{aligned} G(1,2) =& G_0(1,2) - i \int d3d4 G_0(1,3) v(3,4) \\ & \times \{ G_0(4,4+) G_0(3,2) - G_0(3,4) G_0(3,2) \} + \text{higher order terms}\,. \end{aligned} \tag{4.70}$$

The first term between curly brakets represents the Hartree interaction while the second term represents the exchange interaction. If one restricts to these two terms and neglect higher order terms, the approximation is simply non-self-consistent and perturbative Hartree-Fock. One can start over to second order and so on. However, the convergence of such series expansions is believed to be poor. Indeed, the poles of $G$ lies reasonably off the real axis due to their finite lifetime, which is not the case for $G_0$ which has an infinite lifetime. Therefore, one tries to construct an operator $G$ that describes damped excitation energies with undamped excitations through $G_0$. However, one can take benefit of Dyson's equation like in the following $\Sigma$ perturbation method.

### $\Sigma$ perturbation method

Instead of expanding G in a perturbation series, one can perform a perturbation series of $\Sigma$ and then solve for $G$. Starting from Eq. (4.66) and replacing $G$ by $G_0$ yields

$$\Sigma(1,2) = -i\delta(1,2)\int d3v(1,3)G_0(3,3+) + iv(1,2)G_0(1,2)\,. \tag{4.71}$$

In order to iterate to higher orders, one should make $\Sigma$ appear in the right hand side of Eq. (4.66), by replacing the term $G^{-1}$ by $G^{-1} = G_0^{-1} - U - \Sigma$. Then one can construct the second order term

$$\begin{aligned}\Sigma^{(2)}(1,2) = &\int d34v(1,3)G_0(1,2)v(2,4)G_0(4,3)G_0(3,4+) \\ &- \int d34v(1,3)G_0(1,4)v(4,2)G_0(4,3)G_0(3,2)\,.\end{aligned} \tag{4.72}$$

$\Sigma^{(2)}$ is a functional of $v$ and $G_0$, of second order in $v$. A third iteration of Eq. (4.66) can be realized to obtain the third order term and so on to obtain any order. Then the $\Sigma$ perturbation method consists in solving $\Sigma$ to the desired order and calculate $G$ using Dyson's equation. The method looks much heavier than the $G_0$ perturbation method. However, the approximation made for $\Sigma$ will be summed to all orders in $G$ owing to Dyson's equation. Consequently, it is believed that less perturbation terms are needed to describe $\Sigma$[8].

Starting from Eq. (4.71), one obtains using Dyson's equation, the Hartree-Fock (HF) approximation. Expression (4.71) is composed of two terms: the

[8] this statement may be hindered by the loss of cancellation between perturbation terms because of the truncated perturbation series.

first one describes the Hartree potential $v_H$ and the second one the exchange interaction. As the Hartree term is local, one can recast the one-particle Hamiltonian h(1) and the term of self-energy usually designates $\Sigma - v_H$. Therefore, the HF self-energy reads

$$\Sigma_x(1,2) = iv(\mathbf{r}_1, \mathbf{r}_2, t_1^+, t_2)\delta(t_1^+ - t_2)G_0(\mathbf{r}_1, \mathbf{r}_2, t_1^+, t_2)\,. \tag{4.73}$$

The HF self-energy is static and consequently Lehmann's amplitudes can be considered as one-particle wavefunctions. Considering Lehmann's representation of $G(\mathbf{r}_1, \mathbf{r}_2, \omega)$ and in frequency space, the HF self-energy reads

$$\Sigma_x(\mathbf{r}_1, \mathbf{r}_2, \omega) = 2i\pi \sum_i \theta(\mu - \epsilon_{\mathbf{k}_i})\Psi_{\mathbf{k}_i}(\mathbf{r}_1)\Psi^*_{\mathbf{k}_i}(\mathbf{r}_2) \times v(\mathbf{r}_1, \mathbf{r}_2)\,. \tag{4.74}$$

The step function depending on the Fermi level $\mu$ restricts the sum over all states to the occupied states only. This particularity originates from the interaction potential used (*i.e.*, the bare Coulomb interaction) which is static. [9] Thus, for periodic systems, the matrix elements of $\Sigma_x$ read

$$\langle i|\Sigma_x|i\rangle = -\frac{4\pi}{\Sigma_c}\sum_{\mathbf{q}}\sum_j \theta(\mu - \epsilon_{\mathbf{k}_j})\sum_{\mathbf{G}} \frac{\tilde{\rho}^*_{ij}(\mathbf{q}+\mathbf{G})\tilde{\rho}_{ij}(\mathbf{q}+\mathbf{G})}{|\mathbf{q}+\mathbf{G}|^2}\,, \tag{4.75}$$

where $\tilde{\rho}_{i,j}$ are matrix elements of plane waves

$$\tilde{\rho}_{i,j}(\mathbf{q}+\mathbf{G}) = \int d\mathbf{r}\Psi^*_{\mathbf{k}_i}(\mathbf{r})e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}\Psi_{\mathbf{k}_j}(\mathbf{r})\,. \tag{4.76}$$

This elemental brick will be very useful in the following discussion in order to obtain the Fourier transform of complicated quantities. The scaling of the computational time of $\Sigma_x$ is $N_v \times N_{\mathbf{k}} \times N_{\mathbf{G}}$ where $N_v$ is the number of occupied states, $N_{\mathbf{k}}$ the number of $\mathbf{k}$-points and $N_{\mathbf{G}}$ the number of $\mathbf{G}$-vectors.

We have presented several attempts to extend $\Sigma$ (or $G$) in terms of $G_0$ and $v$. However, the former quantity can be far from $G$ in an interacting system while the latter suffers from its lack of physicality in the context of a solid: the effective interaction between electrons is much weaker than the bare one in a solid as it is reduced by the dielectric constant of the medium. The dielectric constant is a *mean-field* quantity that sums over all possible polarization effect in the medium. It seems therefore natural to choose the fully interacting Green's function $G$ and the screened Coulomb interaction $W$ to express our self-energy. $W(1,2)$ first introduced by Hubbard [151], is the potential at point 1 when a test charge sits at point 2, including the polarization effects from all the electrons. This approach was the one of Hedin when he derived its celebrated set of 5 equations.

[9] Any static approximation to the interaction potential with the appropriate time ordering $\delta(t_2 - t_1^+)$ will result in self-energies that only depend on the occupied states only.

### 4.4.2 Hedin's equations

Following Hedin [152], let us first introduce the total classical potential V(1).

$$V(1) = U(1) - i \int d2v(1,2)G(2,2^+)\,, \tag{4.77}$$

which is the sum of the external local potential $U$ and the Hartree potential. Using chain rules for the derivation, Eq. (4.66) is recast in the form:

$$\Sigma(1,2) = -i \int d345v(1^+,3)G(1,4)\frac{\delta G^{-1}(4,2)}{\delta V(5)}\frac{\delta V(5)}{\delta U(3)}\,, \tag{4.78}$$

where the following quantity is introduced

$$\varepsilon^{-1}(1,2) = \frac{\delta V(1)}{\delta U(2)}\,, \tag{4.79}$$

Eq. (4.79) is the inverse of the dielectric function of the material. We also introduce the irreducible vertex function

$$\tilde{\Gamma}(1,2,3) = -\frac{\delta G^{-1}(1,2)}{\delta V(3)}\,. \tag{4.80}$$

*Irreducible* means that Hartree contributions are included in $V$. By introducing the screened Coulomb interaction

$$W(1,2) = \int d3\varepsilon^{-1}(3,2)v(1,3)\,. \tag{4.81}$$

$\Sigma$ finally reads

$$\Sigma = i \int d34G(1,4)W(3,1^+)\tilde{\Gamma}(4,2,3)\,. \tag{4.82}$$

This expression is the seed of the $GW$ method. However, some quantities still remain to be expressed more explicity, notably $\varepsilon^{-1}$ and $\tilde{\Gamma}$. Starting with the inverse dielectric function, we insert expression (4.77) in Eq. (4.79) and obtain

$$\varepsilon^{-1}(1,2) = \delta(1,2) - i \int d3v(1,3)\frac{\delta G(3,3^+)}{\delta U(2)}\,. \tag{4.83}$$

One can introduce the reducible polarizability of the system $\chi(1,2) = -i\delta\rho(1)/\delta U(2)$. As opposed to the definition of irreducibility, reducible means that the quantity $\rho$ is only sensitive to the bare external potential. The link between

irreducible $\tilde{\chi}$ and reducible $\chi$ representation of the polarizability is given using the chain rule for the derivation

$$\chi(1,2) = -i \int d3 \frac{\delta G(1,1^+)}{\delta V(3)} \frac{\delta V(3)}{\delta U(2)} \tag{4.84}$$

We can insert the expression of $\varepsilon^{-1}$ (4.83) back into

$$\chi(1,2) = \tilde{\chi}(1,2) + \int d34 \tilde{\chi}(1,3) v(3,4) \chi(4,2)\,. \tag{4.85}$$

The irreducible polarizability can be expressed as a function of $G$ and $\tilde{\Gamma}$. Using expression (4.65), one can show that the irreducible polarizability reads

$$\tilde{\chi}(1,2) = -i \int d34 G(1,3) \tilde{\Gamma}(3,4,2) G(4,1)\,. \tag{4.86}$$

Let us now turn to the irreducible vertex $\tilde{\Gamma}$. By differentiating the Dyson equation $G^{-1} = G_0^{-1} - V - \Sigma$, one gets

$$\tilde{\Gamma}(1,2,3) = \delta(1,2)\delta(1,3) + \frac{\delta \Sigma(1,2)}{\delta V(3)} \tag{4.87}$$

$$\tilde{\Gamma}(1,2,3) = \delta(1,2)\delta(1,3) + \int d45 \frac{\delta \Sigma(1,2)}{\delta G(4,5)} \frac{\delta G(4,5)}{\delta V(3)} \tag{4.88}$$

$$\tag{4.89}$$

Expression (4.65) can be used in order to calculate $\delta G / \delta V$. The irreducible vertex then reads

$$\tilde{\Gamma}(1,2,3) = \delta(1,2)\delta(1,3) + \int d4567 \frac{\delta \Sigma(1,2)}{\delta G(4,5)} G(4,6) G(7,5) \tilde{\Gamma}(6,7,3) \tag{4.90}$$

One can see that $\Sigma$ depends (not exclusively) on $\tilde{\Gamma}$ and $W$ as well as $\tilde{\Gamma}$ depends on $\Sigma$ and $W$ depends on $\tilde{\Gamma}$ through $\tilde{chi}$. The intrication of the equations require that a closed set is defined to make the resolution possible. Such a set was introduced by Hedin [152] and counts five equations of five

quantities

$$G(1,2) = G_0(1,2) + \int d34 G_0(1,3)\Sigma(3,4)G(4,2) \tag{4.91}$$

$$\tilde{\Gamma}(1,2,3) = \delta(1,2)\delta(1,3) + \int d4567 \frac{\delta\Sigma(1,2)}{\delta G(4,5)} G(4,6)G(7,5)\tilde{\Gamma}(6,7,3) \tag{4.92}$$

$$\tilde{\chi}(1,2) = -i \int d34 G(2,3)G(4,2)\tilde{\Gamma}(3,4,1) \tag{4.93}$$

$$W(1,2) = v(1,2) + \int d34 v(1,3)\tilde{\chi}(3,4)W(4,2) \tag{4.94}$$

$$\Sigma(1,2) = i \int d34 G(1,4)W(3,1^+)\tilde{\Gamma}(4,2,3) \tag{4.95}$$

The resolution of this system of equations is extremely difficult. However, one can think of an iterative process: starting from a simple $\Sigma$ for example, one can build the interacting Green's function $G$, the vertex function $\tilde{\Gamma}$ the irreducible polarizability $\tilde{\chi}$, the screened coulomb interaction $W$ and finally construct a new self-energy that can be considered as a new starting point for a second iteration. That was the proposition of Hedin in order to solve such system. However, the derivation of the second order vertex function shows that already at the second iteration, the vertex has an excessively complicated form. So one should hope that few iterations are actually required to reach a solution. One can notice that the choice of $W$ instead of $v$ was already motivated by the same consideration: expansion in powers of $W$ was supposed to converge faster than in powers of $v$.

**In conclusion**, we have presented two methods relying on perturbation theory and series expansion. While there are important differences between the two schemes, both relies on a fast convergence of the series in order to truncate. However, the previous assumption is at least for extended systems an hopeless dream due to the choice of $G_0$ and $v$ as expansion parameters. Hedin presented a closed set of 5 equations that rely on the screened Coulomb interaction $W$ instead of the bare Coulomb interaction $v$. Few cycles around Hedin's pentagon are believed to achieve convergence and yield good results. However, after one single iteration some of the quantities (particularly the vertex $\Gamma$) becomes excessively complicated. Therefore, it is necessary to introduce some approximation in order to solve Hedin's system of equations. We will see in the next section what kind of approximation should be made to Hedin's pentagon in order to calculate the QP excitation energies.

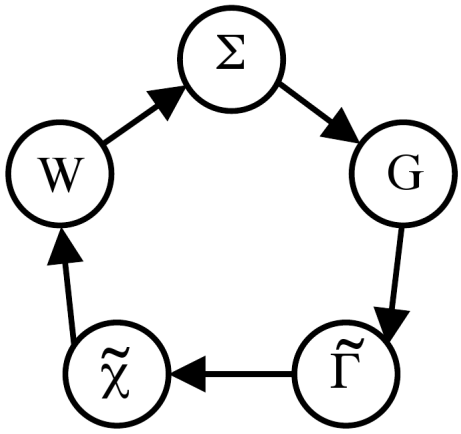


**Figure 4.2:** Symbolical representation of the practical solution of the Hedin's equations. [148]

### 4.4.3 GW approximation

In order to solve the set of equations, a symbolical representation might be useful (Figure 4.2). One can see that any approximation, at any node of the pentagon will propagate to the other 4 variables. One possibility is to approximate the vertex function is $\tilde{\Gamma}(1,2,3) = \delta(1,2)\delta(1,3)$ as it is the first term of $\Gamma$ in Eq. (4.90). The self-energy is then expressed as

$$\Sigma(1,2) = iG(1,2)W(2,1^+)\,, \tag{4.96}$$

and $\tilde{\chi}$ becomes $\tilde{\chi} = -iGG$. With this ansatz, $G$, $\Sigma$ and $\tilde{\chi}$ have to be calculated self-consistently from the Dyson equations. Alternatively, one can start from the Hartree picture assuming $\Sigma = 0$. It follows that $G = G_H$ where $G_H$ is the Hartree Green's function and $\tilde{\Gamma}(1,2,3) = \delta(1,2)\delta(1,3)$. The irreducible polarizability reduces simply to the independent particle one $\chi_0$. As a consequence, the dielectric function and $W$ are expressed within the random phase approximation (RPA). All effects beyond these approximations are considered small with respect to the first-order. The self-energy is expressed through the simple form

$$\Sigma(1,2) = iG_H(1,2)W_0(2,1^+)\,. \tag{4.97}$$

One should notice the use of Hartree Green's function $G_H$. A further cycle around Hedin's pentagon will result in the appearance of the interacting Green's function and then higher order in $G_0$ and $W$. However, the practical implementation of the $GW$ is rather different from this procedure and

is rather a compromise between using a fully interacting G as in (4.96) or a Hartree $G_H$ as in (4.97): most often the "best available" non interacting $G_0$ is used.

In the next section, we will present some of the approximations that are usually made in order to make GW numerically tractable. In order to obtain more physical insight, we look here at a simple approximation introduced by Hedin in his seminal paper of 1965 [152], the so-called COHSEX approximation [152]. Following Hedin's notations [153], we can express the screening as

$$W(\mathbf{r},\mathbf{r}',\omega) = \int d\mathbf{r}'' \frac{\varepsilon^{-1}(\mathbf{r},\mathbf{r}'',\omega)}{|\mathbf{r}''-\mathbf{r}'|} \tag{4.98}$$

$$= \frac{1}{|\mathbf{r}-\mathbf{r}'|} + \sum_i \frac{2\omega_i V^i(\mathbf{r})V^i(\mathbf{r}')}{\omega^2-\omega_i^2} \tag{4.99}$$

where $\omega_i$ are excitation energies such as $\omega_i = E(N,i) - E(N,0) - i\delta$ with E(N,i) total energy and $\delta$ being an infinitesimal and $V^i(\mathbf{r})$ is called the fluctuation potential

$$V^i(\mathbf{r}) = \int d\mathbf{r}' \frac{\langle N,i|\psi^\dagger(\mathbf{r}')\psi(\mathbf{r}')|N\rangle}{|\mathbf{r}-\mathbf{r}'|}\,. \tag{4.100}$$

The first term of the right hand side of expression (4.99) is simply the bare Coulomb interaction giving rise to $\Sigma_x$. The second term $W_{\rm pol} = W - v$ contains the full frequency dependent screening. We will focus on its contribution to the self energy $\Sigma_{\rm pol}$ obtained from the frequency Fourier transformation of (4.97)

$$\langle i|\Sigma_{\rm pol}(\omega)|i\rangle = \sum_{k\neq 0,l} \frac{|V_{il}^k|^2}{\omega + \omega_k {\rm sgn}(\mu-\epsilon_l) - \epsilon_l}\,, \tag{4.101}$$

where $\epsilon_i$ are the one particle removal or addition energies. One is interested in the diagonal matrix element of the polarization self-energy $\langle i|\Sigma_{\rm pol}|i\rangle$ evaluated at $\omega = \epsilon_i$[10]. In the denominator of expression (4.101), two energies compete : the difference of single-particle excitation $\epsilon_i - \epsilon_l$ and $\omega_k$. In order to compare the order of magnitude of each of these quantities, let us consider the example of Na for whom the smallest excitation enery $\epsilon_1 = -0.378$ Ry and the smallest $\omega_1 = 2.414$ Ry. Therefore, $\omega_k \gg |\epsilon_i - \epsilon_l|$. Besides, $|V_{il}^k|^2$ approaches zero when $|\epsilon_i - \epsilon_l|$ increases. Therefore, the matrix element of

[10] We will see later that we should in fact consider $\omega = E_k$ where $E_k$ is the QP energy.

the polarization self-energy might be approximated as

$$\langle i|\Sigma_{\rm pol}(\epsilon_i)|i\rangle = \sum_{k\neq 0,l} {\rm sgn}(\mu-\epsilon_l)\frac{|V^k_{il}|^2}{\omega_k}\,, \tag{4.102}$$

$$= 2\sum_{k\neq 0, l\in \text{ occ}} \frac{|V^k_{il}|^2}{\omega_k} - \sum_{k\neq 0,l}\frac{|V^k_{il}|^2}{\omega_k}\,. \tag{4.103}$$

Note that matrix elements of the static screening read

$$\langle ij|W_{\rm pol}(0)|i'j'\rangle = -2\sum_k \frac{V^k_{ii'}V^k_{jj'}}{\omega_k}\,. \tag{4.104}$$

The polarization self-energy can hence be expressed as a function of $\langle ij|W_{\rm pol}(0)|i'j'\rangle$

$$\langle i|\Sigma_{\rm pol}(\epsilon_i)|i\rangle = -\sum_{l\in \text{ occ}}\langle il|W_{\rm pol}(0)|li\rangle + \frac{1}{2}\sum_l \langle il|W_{\rm pol}(0)|li\rangle \tag{4.105}$$

The first term of expression (4.105) gives rise to the statically screened exchange (SEX) when associated with the Fock term $\Sigma_x$. The second term is called the Coulomb hole (COH) contribution and using the closure relation can be simplified further to $1/2\ \langle i|W_{\rm pol}|i\rangle$. The Coulomb hole contribution can be viewed in a classical picture: $1/2W_{\rm pol}$ represents the energy required to add a point charge $\mp\delta(\mathbf{r}-\mathbf{r}_0)$ from infinity to $\mathbf{r}_0$. The factor $\frac{1}{2}$ corresponds to the adiabadicity of the process. Thus, the matrix element $\langle i|W_{\rm pol}|i\rangle$ corresponds to the energy corresponding to the charge distribution $|\phi_i(\mathbf{r})|^2$. The introduction of an instantaneous screening $W$ changes the sum over all states into sum over occupied states only. The COHSEX method requires therefore relatively little effort to be implemented compared to the exact $GW$ approximation. The COHSEX approximation is known to overestimate bandgap in semiconductors by 20% [154]. Such difference is believed to originate from the absence of dynamical effects. However, it has the merit of highlighting the physics behind the GW approximation and the central role of the screened Coulomb interaction $W$. COHSEX is generally not employed to calculate excitation energies. Nevertheless, it is particularly useful in the practical implementation of GW when one wants to iterate $G$ and $\Sigma$ to approximate self-consistency [155]. We will present in the next section the technical details of the $GW$ approximation.

## 4.5 GW approximation in practice

### 4.5.1 All electron or pseudopotentials

Since the early practical implementation of $GW$ in the 80's [154], a controversy appeared about how to treat electron and core interaction [156, 154, 157]. Results obtained in all electron methods [158] questioned the use of pseudopotential to represent the core valence interaction [11]. Therefore, one may ask *is the standard core-valence partitioning still valid in GW ?* . To answer this question, let us consider the partitioning of Green's function $G = G_v + G_c$ [12] and the polarizability $P = P_v + P_c$ [13] where $c$ and $v$ refer to core and valence electrons [160] yielding a $GW$ self-energy

$$\Sigma \approx iG_cW + iG_vW_v + iG_vW_vP_cW_v \tag{4.106}$$

where $W_v$ is the valence polarizability such as $W_v = v + vP_vW_v$ [14] The first term refers to the screened exchange of core electrons. It is approximatively the bare exchange term as the screening is ineffective for small distance. The second term is what is usually calculated within the $GW$ approximation: the self-energy of the valence electrons. The third term is the screened polarization due to the core electrons acting on the valence electrons. The first and third term are most often considered at DFT level via the pseudopotential formalism. Within all-electron formalism, it can be calculated exactly, though. Calculation of these contributions validate the use of pseudopotentials: indeed the contribution of both terms is believed to be rather small. Such approximation has been first validated on atoms [161, 162], where the contribution of both terms was estimated $\sim 1$ eV. For $s-p$ semiconductors, with shallow $d$ electrons such as GaAs or Ge, the approximation breaks down and contributions of the core-valence interaction to the bandgap as high as 0.3 eV arise [154]. The case of $d$ electrons is of particular interest: the first term of expression (4.106) depends mostly on the overlap between core and valence wave functions. In the case of $d$ electrons, the spatial overlap with $s$ and $p$ electrons of the same shell is rather large: then the use of pseudopotentials becomes problematic. However, it is possible to incorporate all states

[11] While the motivation was praiseworthy, the underestimated $GW$ bandgaps they attributed to the use of pseudopotential was in fact due to the rather limited number of unoccupied states they used in their calculation. [159]

[12] The partitioning of $G$ is obtained by restricted the sum over states in Eq.(4.22) to core and valence states.

[13] The partitioning of $P$ is realized in a similar manner as the one for $G$, by restricting sum over transitions to transitions from core or valence to unoccupied states.

[14] As a consequence, one can express the screening as $W = W_v + W_vP_cW$.

within the same shell $n$ in the pseudopotential formalism and get closer to the all-electron results [163]: such procedure has been applied with success to the case of Cu $3d$ [164] and CdS [165]. The third term (4.106) is generally small and DFT is enough to obtain an estimate of it.
In conclusion, pseudopotential and all-electron $GW$ calculations are equivalent as long as the valence-core partitioning is chosen properly.

### 4.5.2 Perturbative, non perturbative or self-consistent?

#### Perturbative $G_0W_0$

The standard approximation is the "best $G$, best $W$" strategy, which overcomes the (possibly heavy) iterative procedure. The idea is to calculate $G$ and $W$ from the *best* mean-field theory that is available. In this context, *best* refers not only to the quality of the wave functions and energies but also to the computational cost one has to pay to obtain those quantities. Then one can perform a perturbative correction to the existing excitation energies. Let us first stress the analogy between the QP equation and the Kohn-Sham ones. If one considers the exchange and correlation potential $v_{\text{xc}}$ to be a simplified self-energy

$$\Sigma(\mathbf{r}, \mathbf{r'}, t, t') = v_{\text{xc}}(\mathbf{r})\delta(\mathbf{r} - \mathbf{r'})\delta(t - t')\,, \tag{4.107}$$

then the QP equation reduces simply to the KS equation. Consequently, the solution of the Kohn-Sham equation in terms of wave functions $\phi_i$ can be interpreted as approximate Lehmann's amplitudes of Green's function $G$, the KS energies $\epsilon_i$ being the excitation energies. Therefore it appears clearly that Kohn-Sham orbitals can be used as a starting point for the construction of Green's function and the self-energy. Furthermore, the KS energies and wave functions can be expected to be close enough to the quasiparticle ones to use of a perturbative scheme. This supposes that the wave functions are well described within DFT-LDA [154, 155] *i.e.* $\phi_i \approx \Psi_i$, $\Psi_i$ being the quasiparticle wave function. The first order corrections to the KS energies $\epsilon_i$ read then

$$\langle\phi_i|\Sigma(E_i) - v_{\text{xc}}|\phi_i\rangle = E_i - \epsilon_i \tag{4.108}$$

where $E_i$ is the quasiparticle energy. One should note that the self-energy should be evaluated at $\omega = E_i$. Usually, a Taylor expansion around $\epsilon_i$ is performed and only the linear term is conserved

$$\langle\phi_i|\Sigma(E_i)|\phi_i\rangle = \langle\phi_i|\Sigma(\epsilon_i)|\phi_i\rangle + (E_i - \epsilon_i)\langle\phi_i| \left.\frac{\partial\Sigma(\omega)}{\partial\omega}\right|_{\omega=\epsilon_i} |\phi_i\rangle + O((E_i - \epsilon_i)^2) \tag{4.109}$$

Therefore, the quasiparticle energy $E_i$ can be expressed as

$$E_i = \epsilon_i + Z_i \langle \psi_i | \Sigma(\epsilon_i) - v_{\mathrm{xc}} | \psi_i \rangle \tag{4.110}$$

where $Z_i$ is the renormalization factor of Eq. (4.57). The calculation of $Z_i$ involves the linearization of $\Sigma$ which is only fully efficient in the proximity of the Fermi energy. In fact, the frequency dependence of $GW$ being contained entirely in $\Sigma_c$, the linearization procedure should not be difficult as $\Sigma_c$ behaves particularly well in the region of the Fermi level due to the absence of poles.

$G_0W_0$ usually improves considerably the bandgap compared to DFT-KS within LDA (see Fig. 4.3). However, the quality of the final results depend strongly on the starting point: if the DFT-KS starting point is too far from the experimental band structure as it is the case for CuBr or ZnO (see Fig. 4.3), the resulting $G_0W_0$ will suffer from it. An underestimation of the bandgap leads to an overestimation of the screening and therefore, the LDA underestimation of the $G_0W_0$ correction to the bandgap. In practice, this shortcoming could be corrected by applying a simple scissor operator to the bandgap. However, the deficiencies of DFT-KS are multiple and depend strongly on the type of material. For example, ZnO is famous for the underestimation of the $p-d$ repulsion which tends to close its bandgap: hence, in order to obtain the correct bandgap it is necessary first to treat *correctly* the shallow $3d$ states of Zn. The perturbative approach being computationally favorable, the starting point may be improved in order to get closer to the experimental results. Several schemes have been proposed to fulfill this task which are based on some of the methods presented in chapter 2

- Exact exchange (EXX) [166]: This scheme is still within DFT with a local Kohn-Sham potential. Therefore, the application of the perturbative scheme requires that DFT-KS wave functions per se are *good*. Besides, EXX involves summation over unoccupied states which makes the numerical application as heavy as the $G_0W_0$ itself. It gives improved results for systems particularlly difficult to describe such as ZnO: it is interesting to notice that EXX succeeds in opening the bandgap, however it fails in positioning correctly the Zn $3d$ bands. EXX is self-interaction free and as a consequence, Rinke *et al.* postulated that the poor description of these localized states within GW was due to omission of electron-hole excitations or vertex [166].

- Hybrid functionals through generalized Kohn-Sham [131]: HSE03 or PBE0 functionals gave single particle energies very close to the QP

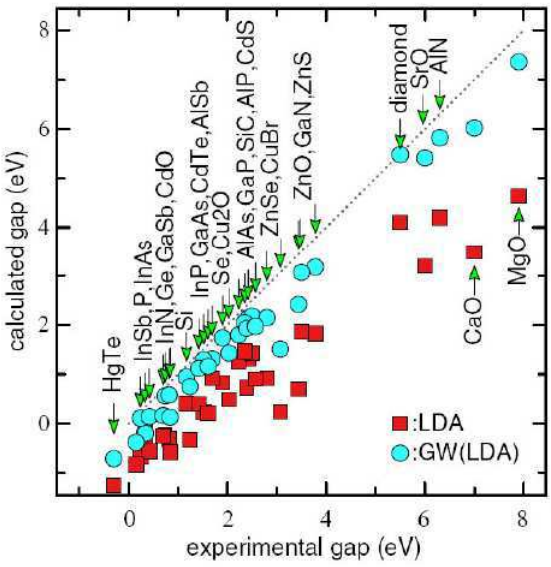


**Figure 4.3:** Comparison between the calculated and the experimental band gaps for a certain number of materials from Ref. [170].

ones, resulting in small GW corrections. The great advantage of that scheme with respect to EXX is the improvement of the wave functions due to the use of a non-local potential.

- LDA+U. Kioupakis *et al.* [167] were the first to use LDA+U as a starting point of *GW* in order to study bcc hydrogen. They showed that the bandgap was stable over a large range of values of $U$. Moreover, the metal insulator transition was perfectly described by $G_0W_0@LDA+U$. LDA+U corrects significantly the lack of localization for the electron $1s$ of Hydrogen, which allows a perturbative many-body treatment only. Jiang and co-workers [168] used LDA+U as an enhanced starting point for the study of Cerium oxide. One of the points raised in their work was the difficulty to treat simultaneously itinerant states that are rather well described in KS-DFT or *GW* and extremely localized states like $d$ or $f$ electrons which may be well described within GW, or require the incorporation of higher order correlation effects like in Dynamical Mean Field Theory (DMFT) [169].

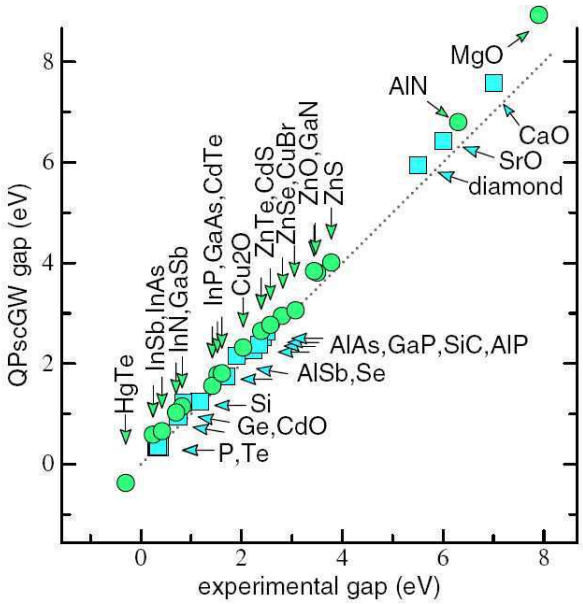


**Figure 4.4:** Comparison between the calculated and the experimental band gaps for a certain number of materials from Ref. [170].

### Self-consistency

The GW approximation restricts Hedin's pentagon to a simple triangle: starting from $G$, one obtains $W$ through the polarizability $\chi$ then $\Sigma$ and owing to Dyson's equation $G$. Cycle around the GW triangle is called full self-consistency. The efficiency of such iterative scheme has only been studied for models or simple systems: electron gas [171], metals [172] or simple semiconductors [173]. The results for spectroscopy were not encouraging at all: a systematic deterioration of the description of bandwidths, quasiparticle excitations and lifetimes was found. On the contrary, the ground state total energies were found to be strikingly closed to the ones of QMC [174]. Such agreement is related to the fact that fully self-consistent $GW$ fulfills conservation laws [149, 175] such as the number of particle, energy and total momentum. The solution to most of the problems seems to be the inclusion of the vertex $\Gamma$ in the self-consistent scheme: self-consistency and vertex corrections have shown opposite effects that cancel each other [176, 172, 177]. However, the *ab initio* calculation of realistic $G$ realistic vertex corrections is today still out of reach. In view of that, fully self-consistent schemes seem to fail to predict the excitation energies.

Holm *et al.* showed that the dynamical part of $GW$ was responsible for

the poor performance of self-consistency [171]. Therefore, self-consistent *GW* has only been considered within a static approximation [178, 155, 170, 179]. Such approaches known as restricted self-consistent *GW* present other advantages besides removing the dynamical part of $\Sigma$: the quasiparticle equation can be simplified considerably as the self-energy is frequency independent and can be made hermitian. Therefore, one only needs to solve the QP equation: doing so, one obtains Lehmann's amplitudes or QP wavefunctions[15] and excitation energies which permits to construct the self-energy operator. Thus, the self-consistence is achieved straighforwardly by solving the QP equation iteratively. Two main types of self-consistent schemes have been explored: the self-consistent COHSEX [155] and the Quasiparticle Self-consistent *GW* (QSGW) [170].

- Self-consistent COHSEX takes advantage of the hermitian nature of the COHSEX self-energy. Most importantly, it does not require the inclusion of unoccupied states in the calculation of the self-energy and screening is evaluated only statically. The bandgap predicted is usually too large. However, the *recipe* involves the use of the sc-COHSEX eigenstates as a starting point for a perturbative $G_0W_0$ step. Sc-COHSEX+$G_0W_0$ yields particularly good results for standard $s-p$ semiconductors and insulators [155], but also for transition metal oxides [180, 181].

- QSGW partially includes dynamical effects in a very subtle way. Let us follow the derivation made by Kotani *et al.* [179]. We consider the *bare* QP equation with the following Hamiltonian:

$$H_0 = -\frac{\nabla^2}{2} + v^{\text{eff}}(\mathbf{r}, \mathbf{r}') \tag{4.111}$$

  where $v^{\text{eff}}(\mathbf{r}, \mathbf{r}')$ is a non-local frequency-independent effective potential. The *real* or dressed QP equation is similar to the effective potential being replaced by the non-local frequency-dependent self-energy. The idea behind QSGW is to map the frequency dependent self-energy into an effective potential the $v^{\text{eff}}$. Such mapping involves two important steps: get rid of the frequency dependence of $\Sigma$ and make $\Sigma$ hermitian. Thus, the self-consistent scheme consists in constructing $\Sigma(\mathbf{r}, \mathbf{r}', \omega)$, and then the corresponding effective potential $v^{\text{eff}}(\mathbf{r}, \mathbf{r}')$. The convergence is attained when the solution of the bare QP equation $\{\epsilon_i, \psi_i\} \approx \{E_i, \Psi_i\}$ around the Fermi level, where $\{E_i, \Psi_i\}$ is the solution of the dressed

[15] The hermitianity of the QP Hamiltonian implies that Lehmann's amplitudes are considered as QP wave functions.

QP equation. The latter condition allows one to construct a relatively simple effective potential such as

$$v^{\text{eff}} = \frac{1}{2}\sum_{ij} |\psi_i\rangle \left\{ \text{Re}\left[\Sigma(\epsilon_i)\right]_{ij} + \text{Re}\left[\Sigma(\epsilon_j)\right]_{ji} \right\} \langle\psi_j| \tag{4.112}$$

QSGW gives similar results compared to sc-COHSEX+$G_0W_0$ and yields particularly good bandgaps (see Fig. 4.4). Moreover, it is a less crude approximation compared to the static COHSEX that tends to drastically overestimate some bandgaps. QSGW has proven its ability to treat a broad range of materials [182, 183, 178, 184]. However, in this thesis, we prefered the COHSEX approach as it does not involve sums over empty states for each iteration.

**In conclusion**, we have presented two approaches used for the practical implementation of $GW$: one based on first order perturbation theory called $G_0W_0$ and one based on the iteration of the $GW$ *triangle*. Depending on the system one wants to study, one is more relevant than the other. However, the perturbation being usually performed on top of DFT-LDA, the two mains reasons that prevent the use of $G_0W_0$ are the lack of localization of the wave functions and the severe underestimation of the bandgap which results in the overestimation of the screening.
In the next section, we will describe the calculation of $\Sigma$ within a plane wave basis set and describe the self-consistent procedure used for the calculations presented in this thesis.

### 4.5.3 Practical implementation

#### Calculation of $\chi_0$

Let us assume that we know $G_0$. Therefore, the independent particle polarizability simply reads then

$$\chi_0(\mathbf{r}_1, \mathbf{r}_2, \omega) = -\frac{i}{\pi}\int d\omega' G_0(\mathbf{r}_1, \mathbf{r}_2, \omega+\omega')G_0(\mathbf{r}_1.\mathbf{r}_2, \omega') \tag{4.113}$$

If one introduces Lehmann's representation of non-interacting Green's function (4.22), expression (4.113) becomes

$$\begin{aligned}\chi_0(\mathbf{r}_1, \mathbf{r}_2, \omega) = -\frac{2i}{2\pi}\sum_{\mathbf{k}_i,\mathbf{k}_j}\int d\omega' &\frac{\Psi_{\mathbf{k}_i}(\mathbf{r}_1)\Psi^*_{\mathbf{k}_i}(\mathbf{r}_2)}{\omega+\omega'-\epsilon_{\mathbf{k}_i}+i\eta\text{sign}(\epsilon_{\mathbf{k}_i}-\mu)} \\ &\times \frac{\Psi_{\mathbf{k}_j}(\mathbf{r}_1)\Psi^*_{\mathbf{k}_j}(\mathbf{r}_2)}{\omega'-\epsilon_{\mathbf{k}_j}+i\eta\text{sign}(\epsilon_{\mathbf{k}_j}-\mu)}\end{aligned} \tag{4.114}$$

The integration appears to be particularly cumbersome due to the presence of poles. However, one can take advantage of the residue theorem. By considering a closed path of integration, the integral of an analytical function of a complex variable $z$, $f(z)$ is given by

$$\oint f(z)dz = 2\pi i \sum_p Res_{z=z_p} f(z) \tag{4.115}$$

where $Res_{z=z_p}$ stands for the residues of the pole $z_p$. By choosing an adequate integration path, *i.e.* either using an arc in the upper or lower imaginary plan, one can reduce the integration over frequency to a sum over the poles of the function of expression (4.114). First Green's function has poles at $\omega' = \epsilon_{\mathbf{k}_i} - \omega - i\eta \text{sign}(\epsilon_{\mathbf{k}_i} - \mu)$ and the second at $\omega' = \epsilon_{\mathbf{k}_j} - i\eta \text{sign}(\epsilon_{\mathbf{k}_j} - \mu)$. Hence, $\chi_0$ simplifies to

$$\begin{aligned}\chi_0(\mathbf{r}_1, \mathbf{r}_2, \omega) = \frac{2}{N_{\mathbf{k}}\Omega_c} \sum_{\mathbf{k}_i, \mathbf{k}_j} \Psi_{\mathbf{k}_i}(\mathbf{r}_1)\Psi^*_{\mathbf{k}_j}(\mathbf{r}_1)\Psi^*_{\mathbf{k}_i}(\mathbf{r}_2)\Psi_{\mathbf{k}_j}(\mathbf{r}_2) \\ \times \left[\frac{\theta(\epsilon_{\mathbf{k}_i} - \mu)\theta(\mu - \epsilon_{\mathbf{k}_j})}{\omega - (\epsilon_{\mathbf{k}_i} - \epsilon_{\mathbf{k}_j} + i\eta)} - \frac{\theta(\mu - \epsilon_{\mathbf{k}_i})\theta(\epsilon_{\mathbf{k}_j} - \mu)}{\omega - (\epsilon_{\mathbf{k}_i} - \epsilon_{\mathbf{k}_j} - i\eta)}\right]\end{aligned} \tag{4.116}$$

The poles of $\chi_0$ lie in the lower plane when the transition energies are positive ($\epsilon_{\mathbf{k}_i} < \epsilon_{\mathbf{k}_j}$) and in the upper plane when the transition energies are negative ($\epsilon_{\mathbf{k}_i} > \epsilon_{\mathbf{k}_j}$). In a region centered on $\mu$ whose width is exactly the bandgap, $\chi_0$ has no pole. The Fourier transform of the polarizability reads in a compact form

$$\chi_{0,\mathbf{G},\mathbf{G}'}(\mathbf{q}, \omega) = \sum_{i,j} (f_i - f_j) \times \frac{\tilde{\rho}_{ij}(\mathbf{q} + \mathbf{G})\tilde{\rho}^*_{ij}(\mathbf{q} + \mathbf{G}')}{\omega - (\epsilon_i - \epsilon_j) + i\eta \text{sign}(\epsilon_i - \epsilon_j)} \tag{4.117}$$

where $i$ stands for $\mathbf{k}_i$, $f$ the occupation number of states $i$. The evaluation of $W_{\mathbf{G},\mathbf{G}'} = \varepsilon^{-1}_{\mathbf{G},\mathbf{G}'} v(\mathbf{q} + \mathbf{G}')$ is made with the help of the random phase approximation polarizability

$$\varepsilon_{\mathbf{G},\mathbf{G}'}(\mathbf{q}, \omega) = \delta_{\mathbf{G},\mathbf{G}'} - v(\mathbf{q} + \mathbf{G})\chi_{0,\mathbf{G},\mathbf{G}'}(\mathbf{q}, \omega) \tag{4.118}$$

where $\chi$ is the random phase approximation (RPA) polarizability $\chi = (1 - v\chi_0)^{-1}\chi_0$. Then, we have to construct the operator $\Sigma$ in order to solve the QP equation (4.52). First, we partition the self-energy as $\Sigma = \Sigma_x + \Sigma_c$ where $\Sigma_x$ is simply the Fock self-energy $\Sigma_x = iGv$ and $\Sigma_c$ is the correlation part of the self-energy defined as $\Sigma_c = \Sigma - \Sigma_x$[16]. The correlation part of

[16] One can notice the analogy with the definition of the correlation part of the exchange and correlation potential. Correlation being extremely difficult to define, its definition usually involves everything beyond exchange interaction.

the self-energy is frequency dependent since it involves the energy-dependent interaction $W_p = W - v$:

$$\Sigma_c(\mathbf{r}_1, \mathbf{r}_2, \omega) = \frac{i}{2} \int d\omega' e^{i\omega'\delta} G(\mathbf{r}_1, \mathbf{r}_2, \omega + \omega') W_p(\mathbf{r}_1, \mathbf{r}_2, \omega') \tag{4.119}$$

The most time-consuming task is the evaluation of the frequency integral. We have previously shown that $G$ and $W$ have poles along the real axis in the upper and lower imaginary half plane. The integration over the complicated pole structure of both $G$ and $W$ can be avoided by using a physical argument: observing the imaginary part of $\varepsilon^{-1}$, which can be measured in an electron loss experiments, one can realize that the spectra of $\varepsilon^{-1}(\omega)$ are dominated by plasmon peaks. Plasmons are quanta of the collective oscillations of the electron density. The plasmon frequency of the free electron gas is

$$\omega_p = \hbar \sqrt{\frac{ne^2}{m^* \varepsilon_0}} \tag{4.120}$$

where $n$ is the electronic density, $e$ the elemental charge, $m^*$ the effective mass of the electron and $\varepsilon_0$ the dielectric constant of the material. Usually, semiconductor plasmon peaks lie high in energy $\approx 20$ eV. Therefore, it appeared natural to model $W$ by a single plasmon-pole model as it was proposed in the original paper of L. Hedin [152]. Within this approximation, $\varepsilon^{-1}$ reads

$$\varepsilon^{-1}_{\mathbf{G},\mathbf{G}'}(\mathbf{q}, \omega) = \delta_{\mathbf{G},\mathbf{G}'} + \frac{\Omega^2_{\mathbf{G},\mathbf{G}'}}{\omega^2 - (\tilde{\omega}_{\mathbf{G},\mathbf{G}'} - i\eta)^2} \tag{4.121}$$

where $\Omega^2_{\mathbf{G},\mathbf{G}'}$ and $\tilde{\omega}_{\mathbf{G},\mathbf{G}'}$ are two parameters to determine and $\eta$ is an infinitesimal ensuring the right time-ordering. One needs two constraints for each $(\mathbf{q}, \mathbf{G}, \mathbf{G}')$ to evalute the two parameters of the model. Several procedures have been proposed which give similar results:

- Godby and Needs [185]: one has to calculate the dielectric function at $\omega = 0$ and at some frequency on the imaginary axis, typically of the order of the plasmon frequency $\omega_p$. The dielectric constant behaves smoothly on the imaginary axis, which makes the fitting procedure easier to obtain $\Omega_{\mathbf{G},\mathbf{G}'}$ and $\tilde{\omega}_{\mathbf{G},\mathbf{G}'}$.

- Hybertsen and Louie [140]: one calculates $\varepsilon^{-1}$ at $\omega = 0$ and enforces the f-sum rule

$$\int d\omega \omega \mathrm{Im}\varepsilon^{-1}_{\mathbf{G},\mathbf{G}'}(\mathbf{q}, \omega) = -\frac{\pi}{2}\omega_p^2 \frac{(\mathbf{q}+\mathbf{G}) \cdot (\mathbf{q}+\mathbf{G}')}{|\mathbf{q}+\mathbf{G}|^2} \frac{\rho(\mathbf{G}-\mathbf{G}')}{\rho(0)} \tag{4.122}$$

where $\rho(\mathbf{G})$ is the crystalline charge density. The parameters $\Omega^2_{\mathbf{G},\mathbf{G}'}$ and $\tilde{\omega}_{\mathbf{G},\mathbf{G}'}$ are straightfowardly derived as

$$\Omega^2_{\mathbf{G},\mathbf{G}'}(\mathbf{q}) = \omega_p^2 \frac{(\mathbf{q}+\mathbf{G})\cdot(\mathbf{q}+\mathbf{G}')}{|\mathbf{q}+\mathbf{G}|^2} \frac{\rho(\mathbf{G}-\mathbf{G}')}{\rho(0)}, \tag{4.123}$$

$$\tilde{\omega}_{\mathbf{G},\mathbf{G}'} = \frac{\Omega^2_{\mathbf{G},\mathbf{G}'}}{\delta_{\mathbf{G},\mathbf{G}'} - \varepsilon^{-1}_{\mathbf{G},\mathbf{G}'}(\mathbf{q},\omega=0)}. \tag{4.124}$$

While the Godby and Needs model privileges the region of interest, *i.e.*, the low energy region, the sum rule in the Hybertsen and Louie model does not favor the region of interest.

- In order to correct the shortcomings of HL plasmon pole, von der Linden and Horsch [186] proposed a generalized plasmon pole model based on the fulfillment of a generalized f-sum rule but only for diagonal elements of $\varepsilon_{\mathbf{G},\mathbf{G}'}$. Engel and Farid [187] improved further the von der Linden and Horsch model, proposing a scheme that yields plasmon poles independent of $\mathbf{G},\mathbf{G}'$.

All these models give very similar results considering the bandgap value. However, Shaltaf *et al.* [188] have recently shown that if one is interested the absolute position positions of the quasiparticle energy, the choice of the plasmon pole model is of primer importance. They studied the band offset at the interface between $Si/SiO_2$ and found important differences in the band edge shifts $\Delta E_{v,c}$ depending on the plasmon pole model used. The results were compared to the full frequency integration, which is the reference method. The authors concluded that the Godby and Needs model outperforms all the other plasmon pole models and yields very good agreement compared to fully frequency integration. In this thesis, we will use exclusively the Godby and Needs plasmon pole model. If one performs the frequency integration of Eq. (3.62) using the single plasmon pole model, the diagonal matrix element for $\Sigma_c$ can be evaluated by means of Cauchy's residue theorem

$$\langle i|\Sigma_c|i\rangle = \frac{2\pi}{V}\sum_j \sum_{\mathbf{G},\mathbf{G}'} \frac{\tilde{\rho}^*_{ij}(\mathbf{q}+\mathbf{G}')\tilde{\rho}_{ij}(\mathbf{q}+\mathbf{G}')}{|\mathbf{q}+\mathbf{G}'|^2} \times \frac{\Omega^2_{\mathbf{G},\mathbf{G}'}}{\tilde{\omega}_{\mathbf{G},\mathbf{G}'}} \left[\frac{\theta(\mu-\epsilon_j)}{\omega+\tilde{\omega}_{\mathbf{G},\mathbf{G}'}-\epsilon_j-i\eta} + \frac{\theta(\epsilon_j-\mu)}{\omega-\tilde{\omega}_{\mathbf{G},\mathbf{G}'}-\epsilon_j+i\eta}\right] \tag{4.125}$$

The matrix elements of $\Sigma_c$ have poles above ($\omega<\mu$) and below the real axis ($\omega>\mu$) apart from a region centered around the Fermi level and whose width

is $2\tilde{\omega}$. Their evaluation scales as $N_b \times N_{\mathbf{k}} \times N_{\mathbf{G}}^2$ where $N_b$ is the number of bands, $N_{\mathbf{k}}$ is the number of $\mathbf{k}$-points and $N_{\mathbf{G}}$ the number of $\mathbf{G}$ vectors. The calculation of $\Sigma_c$ not only involves the occupied states but also unoccupied states. The convergence with respect to the number of bands is actually very slow as demonstrated by recent reports [159]: it is clearly the bottleneck of $GW$ calculations and large efforts are currently undertaken to reduce the computational load [189, 5].

**In conclusion**, the plasmon pole model is a standard method that allows one to reduce considerably the computational cost of $GW$ as it allows the analytical integration over frequency in Eq. (4.119). While dictated by the observation of electron loss spectrum, the choice of the plasmon pole model is not relevant when the physics of the material requires the precise details of the frequency dependence of $W$ [190]. The single plasmon pole model can give good results even in presence of complicated structures in the electron energy loss spectrum [180].

The alternative to the plasmon-pole model is to perform the full frequency integration. Several ways exist [160, 191] and present advantages and drawbacks. We will focus on one of them which appears to be the most stable one: the contour integration method [191]. The integration over the real axis is replaced by an integration along the closed path displayed in Fig. 4.5. Using the residue theorem, we can express $\Sigma_c$ as

$$\Sigma_c(\omega) = \frac{i}{2\pi}\left[2i\pi \sum_{p \in \text{poles of G or W}} \lim_{z \to z_p} G(z)W_p(z)(z - z_p) - \int_{-i\infty}^{+i\infty} d\omega' G(\omega + \omega')W_p(\omega')\right] \tag{4.126}$$

The second term on the right hand side of Eq. (4.126) is the contribution from the imaginary axis. It is removed from the integration on the total contour in order to retain only the contribution along the real axis. As shown previously, $G$ and $W$ behave well along the imaginary axis due to the absence of poles. Hence, it is believed that the integration along the imaginary axis is numerically more affordable and stable than along the real axis. Therefore, let us concentrate on the first term, the sum over the poles of both $G$ and $W$. The choice of the contour path allows one to reduce the number of poles inside the contour. Only some poles of $G$ are enclosed in the closed integration path:

- in the upper-right quadrant, only when $\epsilon_i > \omega$, the poles of occupied states (*i.e.* $\mu > \epsilon_i$) which lie above the real axis.

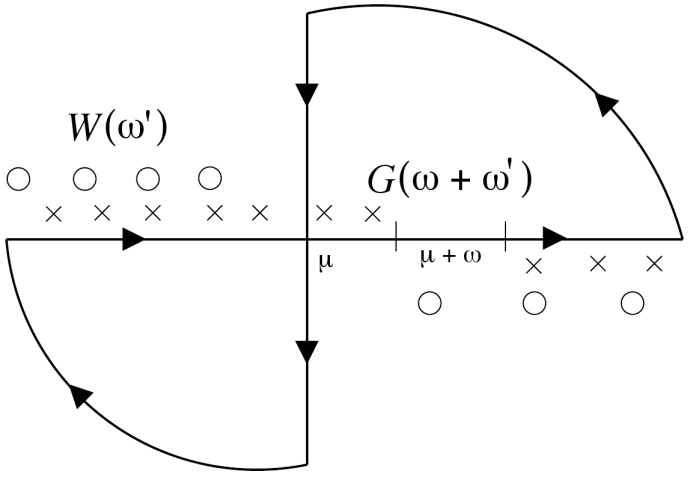


**Figure 4.5:** Location of the poles of Green's function $G$ (crosses) and of the screened Coulomb interaction $W$ (open circles) and path of the contour integral (arrows) used in the method. [148]

- in the lower-left quadrant, only when $\epsilon_i < \omega$, the poles of unoccupied states (*i.e.* $\mu < \epsilon_i$) which lie below the real axis.

If one translates all these conditions in terms of $\theta$-functions, one obtains for the first term of Eq. (4.126)

$$\sum_p = \sum_i \left[\theta(\mu - \epsilon_i)\ \theta(\epsilon_i - \omega) + \theta(\omega - \epsilon_i)\theta(\epsilon_i - \mu)\right] \times W_p(\epsilon_i - \omega + i\eta \mathrm{sign}(\epsilon_i - \mu))\,. \tag{4.127}$$

The evaluation of the second term of Eq. (4.126) can be carried out fully numerically [191] or semi-analytically [148]. The latter is performed taking advantage of the fact that the form of the frequency dependence of $G$ is known and that the screening $W$ is rather smooth along the imaginary axis. This solution permits to reduce the sampling of $\varepsilon^{-1}$ along the imaginary axis. This is of primer importance as the contour integration methods scales as $N_{\mathbf{G}}^2 \times N_b \times N_{\mathbf{k}} \times N_{\omega_l}$, where $N_{\omega_l}$ is the number of frequencies along the imaginary axis. The contour-deformation method is therefore much more cumbersome than the simple plasmon-pole model.

### Self-consistency

In the case of self consistent calculations, the screening and self-energy have to be recalculated typically ten times. This could in principle b done using $\Sigma_{\mathbf{G},\mathbf{G}'}$ in a planewave basis. However, a different path has been followed in order to reduce the computational load. Let us consider the case of an hermitian approximation to the self-energy $\Sigma$ and a basis set of LDA wavefunctions. As the DFT-LDA Hamiltonian is hermitian, then the LDA wavefunctions form a complete set of orthonormal basis functions. Thus, the quasiparticle wavefunctions $|\Psi^{\mathrm{QP}}_{\mathbf{k},i}\rangle$ can be written as

$$|\Psi^{\mathrm{QP}}_{\mathbf{k},i}\rangle = \sum_j c_{\mathbf{k},ij}|\Psi^{\mathrm{LDA}}_{\mathbf{k},j}\rangle \tag{4.128}$$

where $c_{\mathbf{k},ij}$ is the projection coefficient $c_{\mathbf{k},ij} = \langle\Psi^{\mathrm{QP}}_{\mathbf{k},i}|\Psi^{\mathrm{LDA}}_{\mathbf{k},j}\rangle$. Thus, we can build the full matrix $\langle i|\Sigma|j\rangle$ in the LDA Hilbert space and diagonalize the Hamiltonian. As $\Sigma$ is hermitian, the calculation of its matrix elements can be restricted to one half of the matrix. Moreover, the projection matrix $c$ will be unitary such as ${}^t c \cdot c = 1$. The self-consistent procedure will use a similar basis at each cycle in order to connect wave functions from iteration $l$ to wave functions of iteration $l+1$

$$|\Psi^{\mathrm{QP(l+1)}}_{\mathbf{k},i}\rangle = \sum_j c^l_{\mathbf{k},ij}|\Psi^{\mathrm{QP(l)}}_{\mathbf{k},j}\rangle \tag{4.129}$$

Then, the expansion of the QP wave functions in the LDA wave functions basis set is recovered by means of matrix multiplication $c = c^n \cdot c^{n-1}...c^1$. A question may arises: *how good is the LDA basis ?* Bruneval *et al.* studied extensively the size of such basis for different materials [148]. He proposed a simple test in order to determine the number of DFT wave functions to consider in the basis: his argument relies on the expected difference of Hartree-Fock wave functions with respect to LDA ones. QP wavefunctions are believed to be closer to the LDA wave functions. Therefore, if one can expand HF wave functions on the LDA basis sets, then it will be possible to expand QP ones too. A relatively small number of LDA basis functions was required to reproduce HF wave functions [155, 148]. The use of such basis set is particularly interesting for localized states that demand several thousands of basis functions in the plane wave basis set.

In the case of important changes of the wave functions, the QP electronic density might be significantly different from the LDA one. Consequently, important oscillations of the the matrix element of the Hartree potential have

been reported, which result in degeneracy breaking of the states due to small errors blowing up or non convergence of the self-consistent *GW* calculations. Therefore, a damping procedure is realized through the mixing of the density in order to prevent any oscillations. Thus, the electronic density at iteration $l$ $\rho_{\text{in}}^{l}$ depends on the previous iteration $l-1$

$$\rho_{\text{in}}^{l} = \alpha\rho_{\text{out}}^{l-1} + (1-\alpha)\rho_{\text{in}}^{l-1} \tag{4.130}$$

where $\rho_{\text{in}}$ is the density used as input *ie* the artificially damped density and $\rho_{\text{out}}$ is the *real* output density calculated from the wave functions. In practice, a value of $\alpha$ below 0.6 usually yields a sufficient damping to achieve convergence [148]. In some extreme cases, one has to decrease $\alpha$ to 0.2 when the changes of the wave functions are too important [180].

**In conclusion**, we have presented Many-Body Perturbation Theory and showed how we can calculate bandgaps within this theoretical framework. Then we have detailed the practical implementation of the *GW*. The most time- and ressource-consuming part is the calculation of the dynamical part of the self-energy. However, it can be significantly simplified by the use of a plasmon-pole model. We have also introduced the LDA basis sets which allow us to considerably lower the computational workload. In the next section, we will present GW results for ZnO and consider the different approximations of *GW* presented in this chapter. Then, we will investigate the In-based chalcopyrite compounds using DFT, Hybrids and *GW* technics.

# Chapter 5

# Results

## 5.1 Effect of self-consistency on *d* electrons

Since the early days of $G_0W_0$ calculations, the quality of KS wavefunctions has been questioned. Hybertsen *et al.* [154] found small differences between LDA and QP wave functions for Si. Bruneval *et al.* studied the effect of self-consistency in GW and in particular the quality of the LDA wave functions when compared to their GW counterparts [155] for Si, Ar and Al. Their analysis showed that away from high symmetry **k**-points, the difference between DFT and QP wavefunctions can be more significant even for simple materials such as silicon. Unfortunately, they only considered *sp* materials and left the question *how bad is the description of d electrons in DFT ?* unanswered for more complex systems. Around the same time, Van Schilfgaarde *et al.* stressed the importance of off-diagonal elements of $\Sigma$ in the calculation of $CeO_2$,ScN or $SrTiO_3$ [192, 193].
Following the spirit of Bruneval *et al*, we applied the different methods presented in Chapter 2 and 3 to zinc-blende (zb) phase of ZnO. This direct bandgap semiconductor has attracted great interest in recent years due to its application as a transparent conducting oxide (TCO) (see Chapter 1 for example). Under ambient conditions, ZnO crystallizes in the wurtzite (wz) structure and consequently only experimental results for the wz phase are available. Nevertheless, zb-ZnO represents a good toy model to investigate the approximations we will use for CIS compound. Indeed, calculations of the zinc blende phase are computationally tractable due to the small number of electrons in the primitive cell. Moreover, ZnO presents striking similarity with CIS: presence of shallow $3d$ electrons and underestimation of the *pd* repulsion resulting in a underestimated bandgap in DFT.

First, a discussion about the experimental data is necessary. As stated

before, the zinc-blende phase of ZnO has not been synthetized experimentally. Therefore, all experimental data originate from wurtzite samples. At the DFT level, we have compared the binding energy of Zn $3d$ states and the bandgap of wurtzite and zinc-blende phases. It results that the zinc blende phase has a smaller bandgap and deeper $3d$ states than the wurtzite phase. Therefore, one can consider that the zinc blende represents the upper bound for the $3d$ binding energy and the lower bound for the bandgap. As these two quantities are predicted respectively too high and too low in DFT, the comparison with experimental data will be meaningful. Besides, binding energies of $3d$ states observed experimentally suffer from a rather large range of values. The popular reference [194] explored the valence band structure of ZnO by means of x-ray photoemission spectroscopy: the binding energies of the Zn $3d$ was measured at 8.8 eV below the top of the valence band. Results compared well with previous UV experiments [195]. However, one recent experiment has shown that the binding energy of $3d$ electrons might be considerably smaller than what was predicted before ($\approx -7.5$ eV) [196, 197, 198]. The reasons for such situation are many:

- Sample quality and surface effect: photoemission is a surface sensitive techniques.
- Charge effects: Good samples can get charged in a photoemission experiment and this charge can affect the band positions.
- Alignment of the Fermi level: Even though the surface and the bulk Fermi levels normally agree, a band bending may occur which also can introduce uncertainties of about 0.5 eV.

In the following discussion, we choose therefore to consider the whole range of experimental data.

Figure 5.1 shows the band structure and density of states calculated in DFT-LDA. The valence band maximum is threefold degenerated, a characteristic of the zinc blende structure. It is situated at $\Gamma$ and is formed by O $2p$ (67%) and Zn $3d$ (32%) states. The conduction band minimum also occurs at $\Gamma$ and is formed by $s$ orbitals centered on the anionic and cationic sites. The DFT bandgap (0.5 eV) is severely underestimated with respect to experiments but compares well with previous DFT calculations [199]. One should notice that the use of pseudopotentials including semicores might decrease slightly the bandgaps, which explains the small discrepancy ($\sim 0.1$ eV) between our results and previous pseudopotential calculations. Let us now turn to the density of states which shows two pronounced peaks at 4 eV and 6 eV below the VBM. They originate from shallow core level of Zn

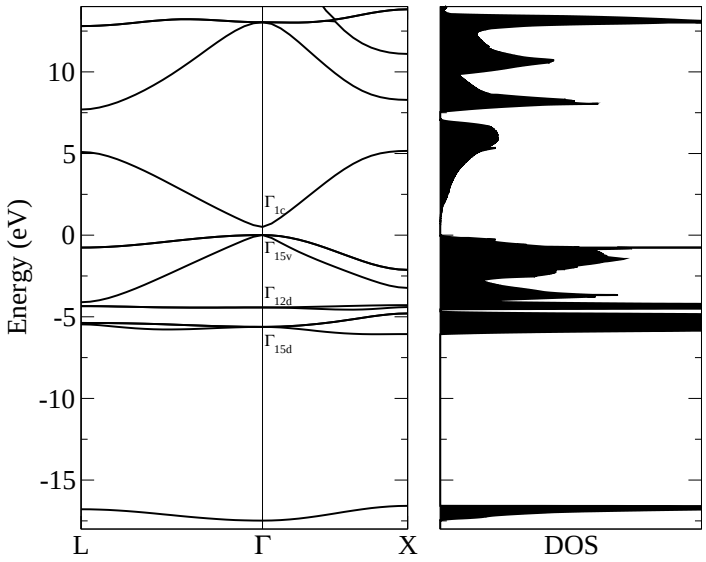


**Figure 5.1:** Left panel: ZnO bandstructure within LDA at the experimental lattice parameter. Right panel: the corresponding density of states.

$3d$. $\Gamma_{12d}$ is doubly degenerate and composed of purely $e_g$ states while $\Gamma_{15d}$ is triply degenerate and is mainly composed of $t_{2g}$ (83%) states with a significant contribution coming from the O $2p$ states (17%). The binding energy of this structure calculated in DFT present large disagreement with experiments (see Tab. 5.1).

In this thesis, I have performed $G_0W_0$ which succeeds in (i) opening the bandgap by 1.5 eV and in (ii) correcting the position of Zn $3d$ states. Still, both bandgaps and binding energies disagree significantly with experiments.

| Symmetry | DFT-LDA | $G_0W_0$ | sc-COHSEX+$G_0W_0$ | QSGW | Expt. |
|---|---|---|---|---|---|
| $\Gamma_{1c}$ | 0.5 | 2.09 | 3.99 | 3.8 | 3.6 |
| $\Gamma_{15v}$ | 0.0 | 0.0 | 0.0 | 0.0 | |
| $\Gamma_{12d}$ | -4.4 | -5.8 | -6.9 | -6.6 | -7.5 |
| $\Gamma_{15d}$ | -5.6 | -6.6 | -7.8 | -7.4 | |
| $\Gamma_{1v}$ | -17.5 | -17.7 | -19.2 | -19.0 | -20.9 |

**Table 5.1:** Features of the band structures of zb-ZnO within different approximations: DFT-LDA, $G_0W_0$, sc-COHSEX+$G_0W_0$ and QSGW calculated in this thesis. Experimental results for wurtzite structure are taken from ref [194, 196]

Because the states involved in the bandgap have a significant $3d$ component, the inclusion of non diagonal elements of $\Sigma$ might be necessary to correct the discrepancy with experiments [200].
The next step in our analysis will be to perform sc-COHSEX+$G_0W_0$. The bandgaps calculated with this method agree well with experiments. The binding energies of Zn $3d$ states are considerably improved compared to $G_0W_0$. Furthermore, sc-COHSEX+$G_0W_0$ and QSGW yield similar results and are in agreement with previous calculations [179]. Besides, we can compare our pseudopotential results against all-electron calculations [192] and demonstrate that the core-valence interaction are rather well described within the pseudopotential formalism for ZnO. In order to establish how much the change of wavefunctions affect our result, we also perform a self-consistent calculation only on energies, *i.e.*, updating the energies in Green's function $G$ and the screening $W$ but keeping the wave functions fixed (Tab. 5.2). One can see that self-consistency on energy predicts a bandgap too large compared to experiments while the position of $3d$ states are higher than for the full self-consistency. Besides, the bandwidth of Zn $3d$ states increased by 0.3 eV between the two flavors of self-consistency. While the discrepancy is important for the bandgap, the difference in the binding energies of $3d$ states is rather small.
We will now look at the matrix elements of the different quantities which enter the QP equation within the COHSEX approximation: Hartree hamiltonian $\langle H_{\text{Hartree}} \rangle$, Hartree-Fock self-energy $\langle \Sigma_x \rangle$ and the correlation part of the COHSEX self-energy $\langle \Sigma_c \rangle$. The two former terms depend only on the wavefunctions and not on the energies while the latter depends on both. A similar analysis has been carried out in Bruneval's work [148] for $Cu_2O$ and the author concluded that large variations of the matrix elements can occur under change of the wavefunctions but the variations of $\langle H_{\text{Hartree}} \rangle$ and $\langle \Sigma_x \rangle$ self energy mostly cancel each other resulting in the overall innocuity of the change in wavefunctions. First let us consider the top valence band $\Gamma_{12v}$.

| Symmetry | sc-COHSEX+$G_0W_0$ (E) | sc-COHSEX+$G_0W_0$ (WF and E) |
|---|---|---|
| $\Gamma_{1c}$ | 4.55 | 3.99 |
| $\Gamma_{15v}$ | 0.0 | 0.0 |
| $\Gamma_{12d}$ | -6.7 | -6.9 |
| $\Gamma_{15d}$ | -7.3 | -7.8 |

**Table 5.2:** Band structure features within different flavors of self-consistency: on energies only and on both energies and wavefunctions.

The use of sc-COHSEX wavefunctions to construct both operators results in the decrease of the absolute value of their matrix elements. However, the decrease of $\langle \Sigma_x \rangle$ is in absolute value more pronounced than the one of $\langle H_{\text{Hartree}} \rangle$: in this particular case, the conclusion of Bruneval *et al.* breaks down, change in wavefunctions results in an upward shift of 0.8 eV that explains mainly the difference in the bandgap between the two types of self-consistency. Despite the relativaly small amount of $d$ character of $\Gamma_{12v}$, changes may be important even at high symmetry **k**-points.
In conclusion, we can expect a large change of the DFT wavefunctions in CIS where the VBM is composed of a higher percentage of $d$ states. Besides, the remaining difference between $G_0W_0$ and the self consistent scheme for the bandgap might come from the strong overestimation of the screening in $W_0$.
Let us now turn to the effect on Zn $3d$ states which is believed to be even stronger: Tab. 5.2 shows that the matrix elements change by more than 4 eV in some cases. Symmetry of the states seems to play an important role as the effects for $t_{2g}$ states is significantly enhanced compared to $e_{2g}$. The $d$ states are pushed downwards with the use of *improved* QP-wavefunctions. One might wonder about the fact that self consistency on energy seems to contradict this conclusion. However, the result is just enhanced by an artifact of the iterative procedure: at each iteration,$\langle H_{\text{Hartree}} \rangle + \langle \Sigma_x \rangle$ is added to the QP energies. The table shows that this term is of the order of several eV and cannot be cancelled to a large extent by $\langle \Sigma_c \rangle$. Therefore, the states are naturally shifted down and improvement over $G_0W_0$ is obtained. Nevertheless, one might question the relevance of doing a calculation updating energies only in the case of systems involving $d$ electrons. Figure 5.2 displays the difference between the partial electronic density corresponding to the state

| | scCOHSEX+$G_0W_0$ (E) | | | scCOHSEX+$G_0W_0$ (WF&E) | | |
|---|---|---|---|---|---|---|
| Symmetry | $\langle H_{\text{Hartree}} \rangle$ | $\langle \Sigma_x \rangle$ | $\langle \Sigma_c \rangle$ | $\langle H_{\text{Hartree}} \rangle$ | $\langle \Sigma_x \rangle$ | $\langle \Sigma_c \rangle$ |
| $\Gamma_{1c}$ | 19.3 | -9.3 | -4.1 | 18.9 | -7.6 | -4.4 |
| $\Gamma_{15v}$ | 31.5 | -32.0 | -0.25 | 30.5 | -29.2 | -0.25 |
| $\Gamma_{12d}$ | 37.7 | -48.0 | 2.2 | 40.5 | -49.0 | 1.9 |
| $\Gamma_{15d}$ | 31.0 | -41.3 | 2.0 | 36.5 | -45.8 | 2.0 |

**Table 5.3:** Matrix elements of COHSEX constructed with self-consistent COHSEX wavefunctions and energies or LDA wave functions and self-consistent COHSEX energies

$\Gamma_{15v}$ in sc-COHSEX and in DFT-LDA expressed as

$$\Delta\rho = \sum_{\alpha=(\mathbf{k},i)} \left( \left|\psi_\alpha^{GW}\right|^2 - \left|\psi_\alpha^{DFT}\right|^2 \right) \tag{5.1}$$

Compared to DFT-LDA, sc-COHSEX removes electronic weight from the region centered on the zinc site to add it in region centered on the oxygen site. Therefore, the hybridized state $\Gamma_{15v}$ has a stronger O $2p$ character in sc-COHSEX than in DFT-LDA. Considering the projection of the VBM wave function on the LDA-basis set, one can immediately see that deep Zn $3d$ DFT-LDA states contribute significantly to the QP wave functions. The admixture of Zn $3d$ states in the VBM originates from the energetic proximity of O $2p$ states and $\Gamma_{15d}$, inherited from the underestimation of the binding energies of $3d$ electrons by DFT. On the other hand, sc-COHSEX also cures the $\Gamma_{15d}$ and $\Gamma_{12d}$ states calculated within DFT-LDA [1]: removing electronic weight around the O site, sc-COHSEX predicts an almost pure Zn $3d$ character for $\Gamma_{15d}$.

**In summary**, our conclusions are

- sc-COHSEX+$G_0W_0$ and QSGW methods predict binding energies and bandgaps in good agreement with experiments.
- When the usual DFT starting point is bad, self-consistency becomes necessary.
- Large changes in the wavefunctions might occur even at high symmetry **k**-points and impact significantly the bandgap.

## 5.2 Theoretical prediction of structural parameters of CIS

The chalcopyrite structure has been introduced in Chapter 1 and the techniques used to investigate its structure via *ab initio* methods in Chapter 2. The Hellmann-Feynman theorem allows one to calculate forces acting on atoms and therefore to predict the phase of the materials by means of energy minimization. The main limitation of the Hellmann-Feynman theorem is the choice of the total energy functional to describe the ground state of the material. In practice, DFT-LDA or GGA functionals are employed and yield

[1] Schröer *et al.* [201] noticed the accidental admixture of Zn $3d$ with O $2p$ states within DFT-LDA.

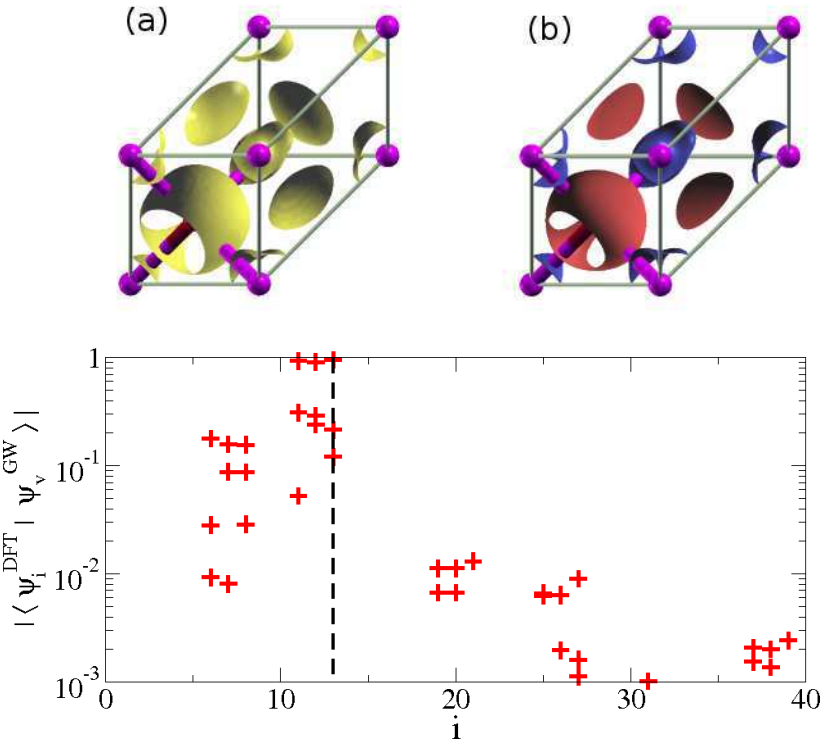


**Figure 5.2:** Top panel: (a) Isosurfaces of the DFT-LDA wave function at $\Gamma_{15v}$ (VBM) (b) Isosurface of the difference between the QP wave function and the DFT-LDA wave function at $\Gamma_{15v}$ (VBM). The value defining the isosurface corresponds to 0.4 of the value defining the isosurface in (a). The Blue isosurfaces are negative changes and red isosurfaces are positive changes. The magenta balls represent zinc atom while red balls represent oxygen ones. Bottom panel: Projection of the QP wave function at $\Gamma_{15v}$ in the DFT-LDA basis set.

particularly good results for *sp* materials with *standard* bonding. As stated in chapter 1, there are two different bonds in CIS, namely Cu-(S,Se) and In-(S,Se), whose difference gives rise to the anion displacement $u$. Bonding occurs under the condition of spatial overlap of the orbitals of two different atoms with the same symmetry. In CIS, the bonding states are hybridized states of (Cu, In) with (S, Se). In the case of Cu-(S, Se), the bonding is realized due to Cu 3*d*-(S, Se)*p* hybridized states. The origin of the bonding between In and the anion is on the other hand rather unclear as it involves two different kinds of hybridization: In 4*d*-(S, Se)*s* and In 5*s*-(S, Se)*p*. Jaffe *et al.* [4] favored the existence of the latter while they only mentioned the rather elongated form of the (S, Se)*s* towards the Indium atom. More recent

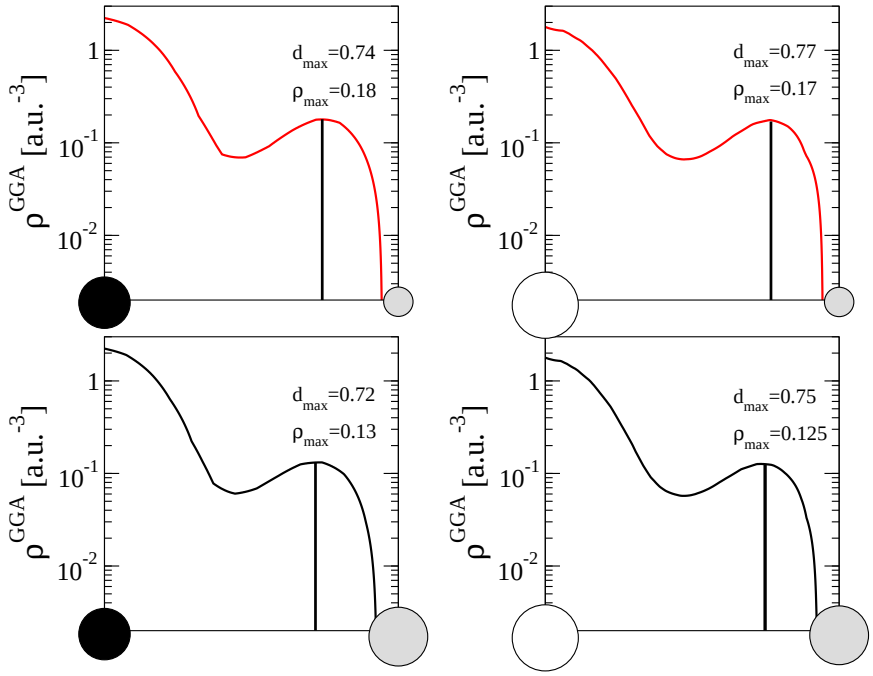


**Figure 5.3:** Electronic density along the Cu-Se bond (left panel) and the In-Se bond (right panel). Small black sphere refers to Cu, light grey ones refer to Se and large to In.

reports [202] acknowledge the participation of both bonding states though.

The participation of localized states such as Cu $3d$ and In $4d$ in the bonding can make the DFT treatment inappropriate. We will show that this observation is indeed relevant for the case of CIS. Figure 5.3 depicts the electronic density along Cu-(S,Se) and In-(S,Se) bond predicted in DFT-GGA. First, one can realize that both bonds are rather ionic as the maximum of density is shifted towards the anion site. Besides, the In-(S,Se) bond appears to be more ionic than Cu-(S,Se) because $d_{\max}^{\mathrm{In-(S,Se)}} > d_{\max}^{\mathrm{Cu-(S,Se)}}$ where $d_{\max}^{X-(S,Se)}$ is the distance between the maximum of density along X-(S,Se) bond and the cation X. High electronic density closed to Cu and In originates from $d$ states and seem to participate in the bonding density. Such assumption will be proved in the forthcoming discussion.
We performed lattice relaxation by means of energy minimization using different energy functionals. Results are summarized in Table 5.4.

The lattice parameter $a$ and the tetragonal distortion $\eta$ are well described by DFT with the usual deviation from experiments between LDA and GGA:

GGA overestimates the lattice parameter while LDA underestimates it. However, $u$ is systematically underestimated by DFT: GGA give a slightly better $u$ than LDA but it is attributed to the change in the lattice parameter. The anion displacement is systematically underestimated: it is related to the underestimation of the Cu-(S,Se) bond length relative to the In-(S,Se) bond length. If one considers the smallest experimental values available, the discrepancy is of the order of 2.2% for both type of anions. The deviation increases up to 7% if only single crystal samples are taken into account [203]. Besides, Jaffe *et al.* have already investigated the effect of $u$ on the bonding: the increase of $u$ results in a lowering of the charge on the stretched Cu-S bond and therefore in reducing the Cu $3d$ character of the bond. Inversely, one can think that the change in the Cu $3d$ character of the bonding states may impact on the value of $u$. Following this spirit, it is natural to relax the structure with techniques curing partially the localization problem of DFT when treating Cu $3d$ states: GGA+U and Hybrid functionals. For GGA+U, we use a value of U previously used for CIS [34], U$\approx$ 6 eV. One can notice an overall improvement of all the structural parameters. While LDA+U only slightly modifies the lattice constant, HSE06 predicts $a$ in agreement with experiments by less than 1%. The most striking improvement concern the anion displacement $u$: LDA+U significantly increases $u$ compared to standard DFT while Hybrid functionals reach a much better agreement with experiments.

A large review of the literature for $CuInSe_2$ shows that previous DFT calculations of CIS [211, 212, 213, 214] seem to fail systematically when based on local or semi local approximations (see Figure 5.4 bottom panel). However, our calculations using hybrid functionals (PBE0, HSE06 or B3LYP) display values for the anion displacement closer to experiments (see Figure 5.4 bottom panel). In order to understand the discrepancy between hybrids and

| | | LDA | GGA | GGA+U | HSE06 | Expt. |
|---|---|---|---|---|---|---|
| $CuInS_2$ | $a$ | 5.43 | 5.58 | 5.575 | 5.545 | 5.52 |
| | $\eta$ | 1.004 | 1.004 | 1.003 | 1.004 | 1.006 |
| | $u$ | 0.216 | 0.220 | 0.223 | 0.229 | 0.230 |
| $CuInSe_2$ | $a$ | 5.70 | 5.97 | 5.97 | 5.817 | 5.784 |
| | $\eta$ | 1.004 | 1.004 | 1.004 | 1.004 | 1.004 |
| | $u$ | 0.216 | 0.218 | 0.222 | 0.227 | 0.235 |

**Table 5.4:** Structure parameters of $CuIn(S,Se)_2$ within different approximations of the total energy functional.

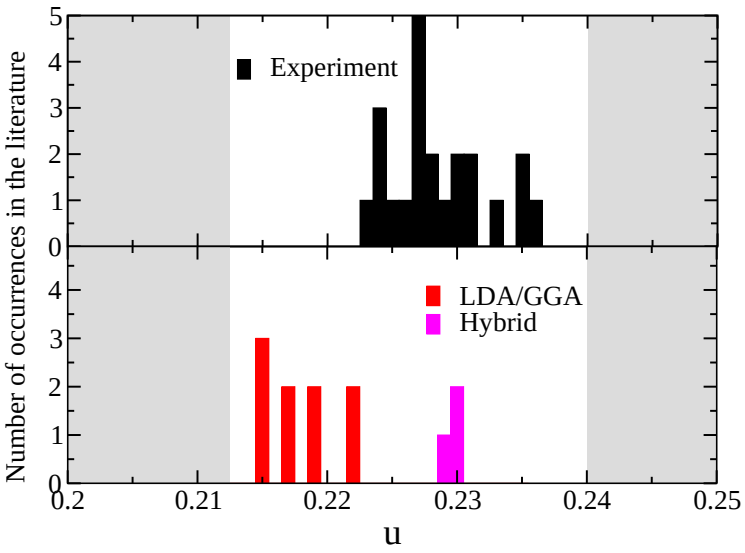


**Figure 5.4:** Histogram representing the experimental distribution of $u$ for $CuInSe_2$ (top panel) [203, 204, 205, 206, 207, 208, 209, 210] (see discussion in Appendix 2) and theoretical distribution of $u$ for the same compound (bottom panel). In the latter, red histograms correspond to LDA/GGA DFT calculation [211, 212, 213, 214] while magenta one to Hybrid calculations (HSE06,PBE0 and B3LYP) we performed.

local and semi-local functionals, let us focus on the electronic density calculated within these two methods. Figure 5.5 shows $\rho^{\text{HSE06}} - \rho^{\text{GGA}}$ along the Cu-S bond and the In-S bond for value of $u = u^{\text{GGA}}$. HSE06 acts similarly along both bonds: the cation density is significantly localized close to the nuclei, which results in a loss of electronic density at mid-bond distance. Inversely, HSE06 delocalizes the electronic density around the anion site towards mid-bond. The In-S bond seems to undergo a much more important correction than the Cu-S bond.

In the GGA+U calculation, we only consider the U correction to the Cu $3d$ electrons. Applying the GGA+U to In $4d$ electrons[2] improves significantly the structural parameters: the lattice paremeters are within 1%-agreement with experiments ($a = 5.52$ Å for $CuInS_2$ and $a = 5.818$ Å for $CuInSe_2$), comparable to those obtained with HSE06. The anion displacement $u$ is also dramatically impacted by the Hubbard correction applied to In $4d$ states: $u = 0.231$ for $CuInS_2$ and $u = 0.228$ for $CuInSe_2$. The correct treatment of the In $4d$ states are hence the key for a good description of parameter $u$.

[2] We choose $U_d = 7$ eV for In $4d$ as previously reported in the literature [215]

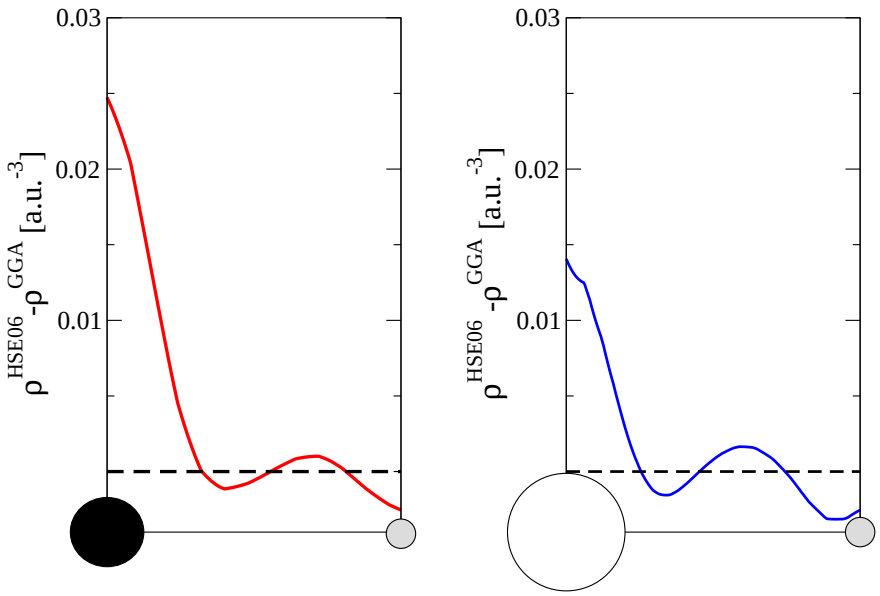


**Figure 5.5:** Difference of the electronic densities predicted within GGA and within HSE06 at $u = u^{\mathrm{GGA}}$ along Cu-S (left panel) and In-S bond(right panel). The medium black sphere refers to the Cu atom, the large white sphere to the In atom and the small grey spheres to the S atoms.

It also proves that the In $4d$ and (S, Se) $s$ states are mainly responsible for the In-S bonding. GGA+U outperforms hence significantly standard GGA as long as both Cu $3d$ and In $4d$ states are corrected and gives structural parameters very close to the ones predicted within HSE06 at a significantly lower computational cost.

**In summary**, our conclusions are

- Bonding in CIS occurs between Cu $3d$ and (S, Se) $p$ states for the Cu-(S, Se) bond and between In $4d$ and (S, Se) $s$ for the In-(S, Se) bond.

- Prediction of the internal anion displacement is hindered by the lack of localization of Cu $3d$ and In $4d$ states.

- Numerical methods that cure the self-interaction problems and therefore that localize correctly Cu $3d$ and In $4d$ states yield values of $u$ in good agreement with experimental results.

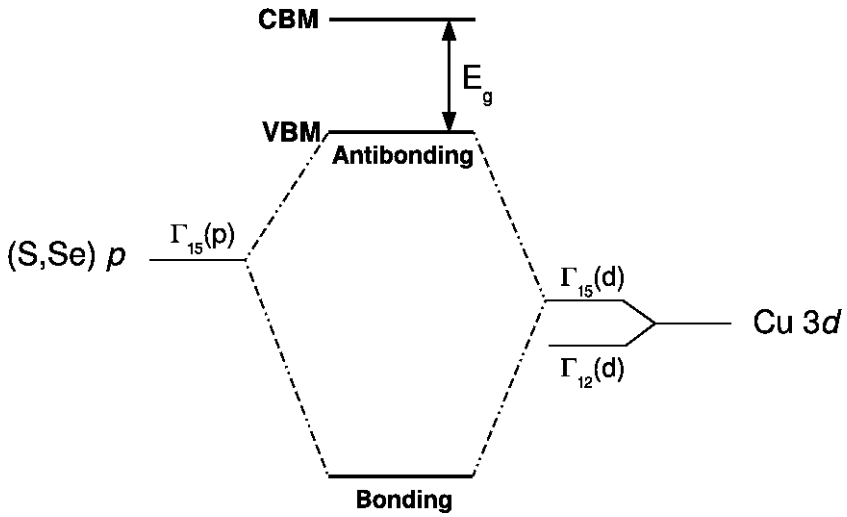


**Figure 5.6:** Schematic molecular orbital diagram of the bonding between Cu $3d$ and (S, Se) $p$ states in CIS.

## 5.3 Electronic structure of $CuIn(S,Se)_2$

As a ternary compound, CIS presents a complex electronic structure. Let us first focus on the upper part of the valence density of states. It involves the bonding between Cu $3d$ states and anion $p$ states. The fivefold degenerate $3d$ states of copper transform into a threefold degenerate $\Gamma_{15}$(d) combination and into a twofold degenerate $\Gamma_{12}$(d) combination. $\Gamma_{15}$(d) states are characterized by lobes pointing toward adjacent anions while $\Gamma_{12}$(d) are formed by lobes pointing between nearest-neighbors anions. Therefore, $\Gamma_{15}$(d) can hybridize with states of the same symmetry to form bonding states and antibonding states such as depicted in Fig. 5.6.

One can show by perturbation theory that these states will repel each other with a strength proportional to $\frac{|\langle p|V|d\rangle|^2}{\Delta\epsilon_{pd}}$ where $\langle p|V|d\rangle$ is the $p$-$d$ repulsion matrix element and $\Delta\epsilon_{pd}$ the energy separation between $\Gamma_{15}$(p) and $\Gamma_{15}$(d), $\Delta\epsilon_{pd} = \Gamma_{15}(p) - \Gamma_{15}(d)$. Therefore if $\Delta\epsilon_{pd}$ is small enough like in the case of CIS, the VBM will be shifted upward and as a result will close the bandgap: this is the so-called *bandgap anomaly* of I-III-VI ternary compounds [216].[3] It is explained partially by the $p$-$d$ repulsion often refered as the $d$-orbital hybridization contribution $\Delta E_g^d$ and by structural effects $\Delta E_g^S$ stemming from either the anion displacement $u$ or the tetragonal distortion

[3]The bandgap anomaly is the name given to the bandgap closing of the I-III-VI compound compared to their binary isoelectronic counterpart.

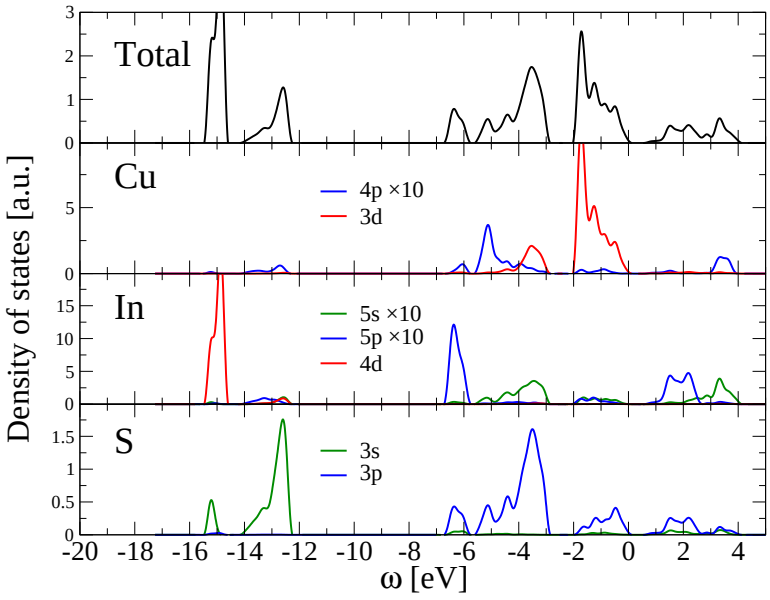


**Figure 5.7:** Total density of states (top panel) and partial density of states (bottom panels) of $CuInS_2$. Red lines refer to $d$ states, blue lines to $p$ states and green lines to $s$ states. Fermi level is at 0 eV.

$\eta$ [1]. It results that the upper part of the valence density of states is dominated by pure Cu 3$d$, (S,Se) $p$, Cu 3$d$-(S,Se)$p$ bonding and antibonding states. On the other hand, the lower part of the valence density of states (down to 20 eV) mainly originates from In-(S,Se) bonding states, pure In and (S,Se) states. In order to distinguish between all these states, we will introduce the partial density of states. The partial density of states, also refered to the projected density of states, corresponds to the local density of states around an atom of specific angular momentum $l$. The $l$-projected density of states of an atom $\alpha$ reads

$$\mathrm{PDOS}^{\alpha,l}(\omega) = \sum_{\mathbf{k}_i,i} \sum_{m} \int_{\mathbf{r}\in V} d\mathbf{r} Y_{l,m}(\theta,\phi)\psi^{*}_{\mathbf{k}_i,i}(\mathbf{r}) \times \delta(\omega - \epsilon_{\mathbf{k}_i,i}) \qquad (5.2)$$

where V is a volume around atom $\alpha$ and $Y_{l,m}$ are the spherical harmonics with $\{\theta,\phi\}$ the spherical angular coordinates. One can then identify which atoms and orbitals participate in the density of states in a particular energy range. Figure 5.7 shows the density of states of $CuInS_2$ calculated within DFT-GGA, together with its partial density of states.

Down to 6 eV below the Fermi level, the valence is formed exclusively

by Cu and (S,Se) states. The strong peak at 2 eV below the Fermi level originates from Cu 3*d* states. A small heteropolar gap of 1 eV separates the non bonding states of Cu 3*d* from the non bonding states of S 3*p* (or Se 4*p*). This is the first major disagreement with experiments as this gap is not present in photoemission experiments [217]. The structure at $\omega = -6.5$ eV is assigned to the bonding states In *p*-(S, Se) *s*. At much lower energy, two structures dominate, the anion *s* electrons around 13 eV below the Fermi level and the In 4*d* electrons 15 eV below the Fermi level. One can notice some admixture between In 4*d* and S 3*p* which confirms the hypothesis formulated previously about bonding occuring between In 4*d* and (S,Se) *s* states. The overall agreement with experiment is rather good with only exception of the heteropolar gap and the position of the In 4*d* states. The position of the valence features for the different methods are summarized in Tables 5.5.

Figure 5.8 shows the density of states calculated within GGA+U and HSE06. The geometry has been relaxed with the functional used to compute the DOS. GGA+U gives significantly different results than GGA: the Hubbard term acting on the Cu 3*d* states pushes the upper valence bands downwards, filling partially the heteropolar gap. The overall upper valence structure is improved and now agrees with experiments. However, the binding energies of states lying at lower energies are significantly worsened compared to GGA. This can be understood easily by the fact that GGA+U only acts on Cu 3*d*, which is believed to be the largest source of error of the GGA calculation. Such dissymetric treatment between Cu 3*d* states and states with different symmetry or originating from different atoms may result in a non physical solution when large hybrization is present (as is the case for CIS). On the other hand, HSE06 improves further the density of states: first, the heteropolar gap is recovered while slightly smaller than in GGA. The main difference with GGA occurs for the position of (S,Se)*s* and In 4*d* states that lies significantly lower in energy. This trend worsens the agreement with experiments for (S,Se)*s* while it improves the agreement for In 4*d*. Still, the

| | GGA | GGA+U | HSE06 | Expt. [218, 219] |
|---|---|---|---|---|
| $E_g$ | -0.02 | 0.6 | 1.3 | 1.54 |
| In-S | 6.3 | 5.9 | 6.7 | 6.9 |
| S 3*s* band | 12.5 | 12.2 | 13.6 | 13.0 |
| In 4*d* band | 14.9 | 14.5 | 16.4 | 18.2 |

**Table 5.5:** Comparison of calculated and observed band structure of $CuInS_2$ (eV).

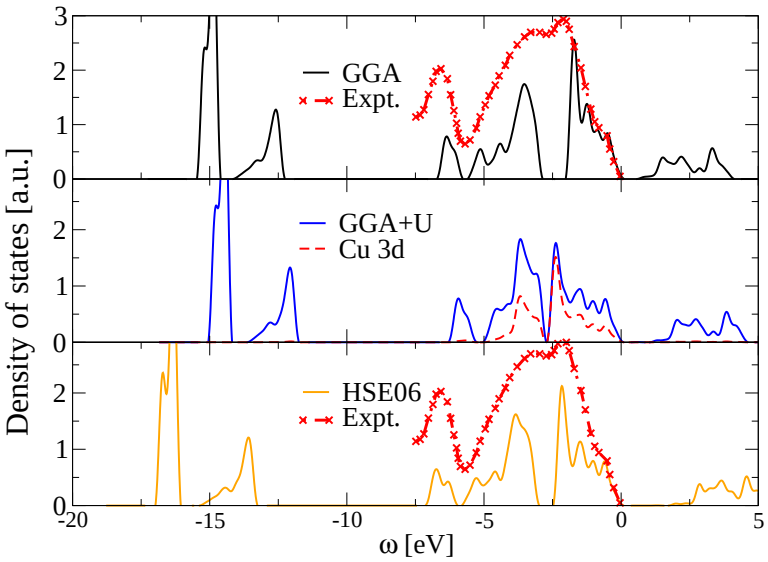


**Figure 5.8:** Total density of states of $CuInS_2$ calculated with DFT-LDA (top panel), GGA+U (middle panel) and HSE06 (bottom panel). Red dashed line with red crosses corresponds to photoemission experiments [217]. Fermi level is at 0 eV.

latter lies still 2 eV higher than observed in the photoemission experiments.

Prediction of the anion displacement $u$ is one of the largest disagreements between the different methods. It is therefore natural to see the effect of $u$ on the density of states. Figure 5.9 displays the density of states of $CuInS_2$ calculated within the GGA for different values of $u$, keeping $a$ and $\eta$ fixed. Some states are more impacted than others by the variations of $u$: the binding energy of In-S bonding states and In $4d$ states remain constant while those of S $3s$ are strongly dependent on $u$. This argument is in favor of the hypothesis stating that the bonding between In and the anion proceed more through the In $4d$ and (S,Se) $s$ orbitals rather than In $5p$ and (S,Se) $p$. The other astonishing effect of $u$ is the variation of the bandwidth of the upper part of the density of states. As a consequence, the heteropolar gap is significantly affected. The $p$-$d$ repulsion seems therefore particularly sensitive to the anion displacement. Parallely, we already know that it is responsible for the *bandgap anomaly*. Hence, $u$ might have a significant impact on the bandgap too.

Considering that experimental data are usually obtained from polycrystalline or thin film samples, one can think of $u$ as a distribution over the

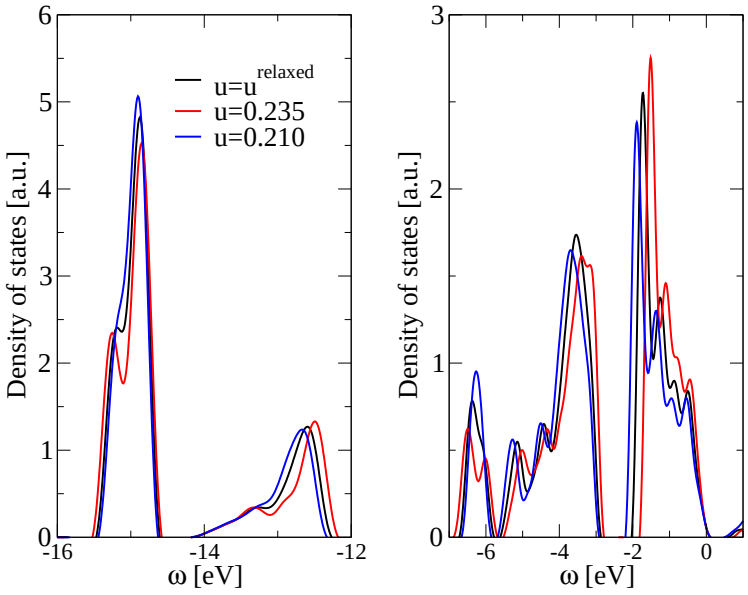


**Figure 5.9:** Density of states calculated within the LDA in different regions of the valence. The black curve refers to the relaxed value of $u$ in GGA ($u^{\text{relaxed}} = 0.219$), the red curve to $u = 0.235$ and the blue curve to $u = 0.21$

whole experimental range and therefore, filling of the heteropolar gap by Cu $3d$ and S $3p$ states might occur. This simple study of the effect on $u$ on the density of states shows that $u$ plays a central role in the determination of the electronic density of CIS. Furthermore, as theory significantly disagrees on the value of $u$, and experiments yield a large dispersion of measured values of $u$, we believe that a systematic study of the bandgap versus $u$ should be performed.

**In conclusion**,

- The electronic structure of CIS is extremely complex and involves many types of orbitals from three different atomic species.
- The top valence spectrum is formed by Cu $3d$ and (S,Se) $p$ bonding/antibonding/non bonding states.
- The lowest part of the valence spectrum is composed of In and (S,Se) states mainly.
- The $p - d$ repulsion has an important impact on the bandgap known as the *bandgap anomaly*.

## 5.4 Analysis of the bandgap

The quantity of interest for PV applications being the bandgap, let us turn to its theoretical treatment. This section will be divided in two subsections: one discussing the bandgap at the DFT level and beyond (GGA+U or Hybrid) and a second one discussing the bandgap at the GW level.

### 5.4.1 Prediction of the bandgap within DFT and Hybrid functional

DFT-KS is by no means able to reproduce experimental bandgaps as it has been shown in the theoretical part of this thesis. However, DFT is a rather common starting point for any advanced *ab initio* methods and is often a good indicator about which physics is present in the material. We performed different DFT calculations with different flavors of technicalities: LDA or GGA, pseudopotential versus PAW, inclusion of semicores (3s and 3p for Cu and 4s and 4p for In) in the pseudopotential.
DFT-LDA/GGA predicts a strong dependence of $E_g$ upon $u$: $dE_g/du \approx 13$ eV (see table 5.6 and Figure 5.10), which is in agreement with previous DFT-corrected [4]calculation [220, 1, 216].

If one considers $\Delta u = 0.02$ defined in Sec. 5.2, the change of bandgap will be $\approx 0.2$ eV which is inconsistent with the observation of the bandgap stability in CIS. The use of PAW instead of pseudopotentials shifts rigidly the curve while GGA tends to open up the gap more than LDA. They do not modify the slope of the dependence though. However, in most of the range of $u$, the bandgap remains *negative*. Similar prediction of a *negative bandgap* by DFT have been obtained for InN [221] or polymorphes of $Cu_2S$ [222]. The overestimation of the $p-d$ repulsion in DFT is responsible for such remarkable feature of the band structure. The case of $Cu_2S$ is of particular

[4]In [220], the authors used a constant scissor operator to open up the bandgaps. Such approach only modified the absolute position of the curve $E_g$ versus $u$, not its slope. On the other hand, in [1, 216], the authors scaled the exchange potential of the LDA functional in order to open up the bandgap.

| | GGA | GGA+U | HSE06 | $G_0W_0$ | scGW |
|---|---|---|---|---|---|
| $\frac{\partial E_g}{\partial u}$ [eV] | 13 | 17.7 | 20.6 | 15.0 | 37.8 |

**Table 5.6:** Slope of the dependence of bandgap of $CuInS_2$ on the anion displacement $u$

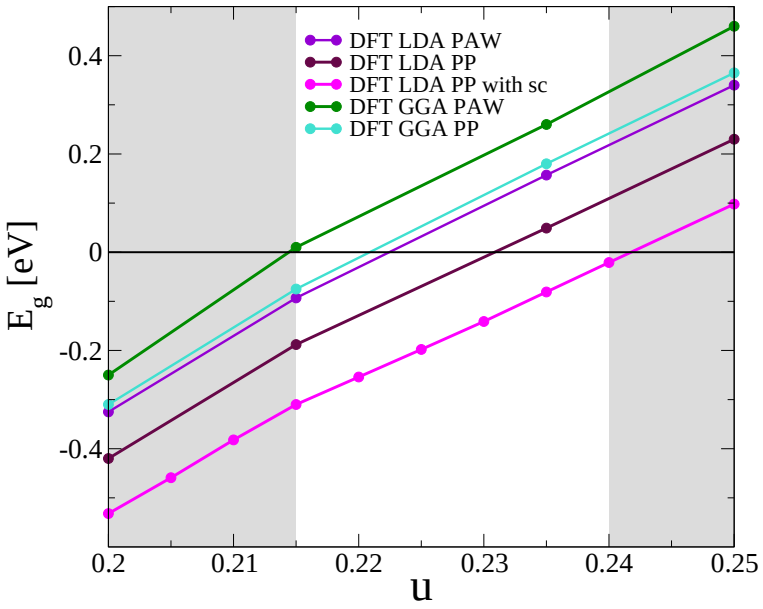


**Figure 5.10:** Bandgap of $CuInS_2$ calculated in DFT versus the anion displacement $u$: impact of the DFT functional and the treatment of core-valence interaction on the dependence of the bandgap on $u$

interest as its valence density of states shows important similarities with the one of CIS:

- the valence band maximum is formed by an hybridization of Cu $3d$ and S $2p$ states.
- a large peak in the density of states at 3 eV below the Fermi level originates exclusively from Cu $3d$ states.
- a 1.5 eV wide heteropolar gap separates this peak from the non bonding S $2p$ states at 4 eV.
- in DFT-KS, the conduction band minimum composed of Cu $4s$ states lies below the VBM, which results in a *negative bandgap*.

Similarly, the valence band maximum in CIS occurs at $\Gamma$. It is composed of hybridized states Cu $3d$ ($\sim 50\%$) and (S,Se)$p$($\sim 50\%$). Its form is characteristic of the chalcopyrite phase: the triply degenerate zinc blende state $\Gamma_v$ is split in the doubly degenerate $\Gamma_{5v}$ and the simply degenerate $\Gamma_{4v}$. The difference between $\Gamma_{5v}$ and $\Gamma_{4v}$ is called the crystal-field splitting $\Delta_{\mathrm{CF}}$. Spin-orbit coupling may further split the doubly degenerate $\Gamma_{5v}$ states.

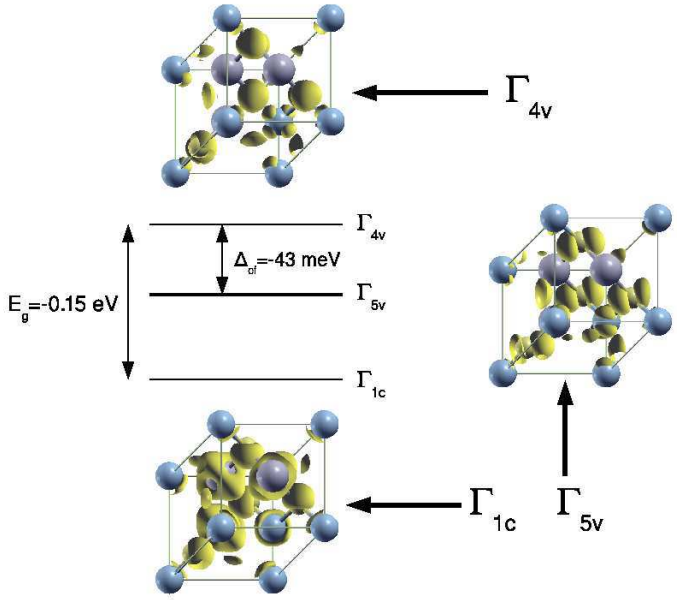


**Figure 5.11:** Schematic structure of the top of the valence bands and the bottom conduction band at Γ of $CuInS_2$ for $u = 0.230$. Isosurface plot represents the wavefunction of each states.

Inspecting the valence band structure close to the Γ point, one can notice the presence of a band lying just below the three states forming the valence band maximum at Γ. This band is depicted in Figure 5.11. It has mainly In $p$ character and a rather delocalized form, while the top valence is thought to be formed by localized antibonding states of Cu $3d$ and (S,Se) $p$. *GW* results will confirm that the band with In $p$ character is indeed the *true* CBM. We also used GGA+U and HSE06 in order to predict bandgap of CIS. Figure 5.12 depicts the bandgap calculated in GGA+U or HSE06 versus $u$. One can see a clear improvement of GGA+U and HSE06 compared to DFT: CIS is predicted as a semiconductor for a larger range of $u$. Furthermore, HSE06 reaches the experimental bandgap of 1.5 eV for $u \approx 0.235$. These two methods appear to be extremely powerful due to the relatively good agreement with experiments and the low computational workload. Besides, it was shown in the previous section that they significantly improve the structural parameters. The dependence of $E_g$ upon $u$ gets significantly stronger 18 eV (GGA+U) to 21 eV (HSE06) compared to DFT. Considering again $\Delta u$, it yields variations of the bandgap of 0.27 eV (GGA+U) and 0.315 eV (HSE06). In order to resolve the apparent contradiction between the exper-

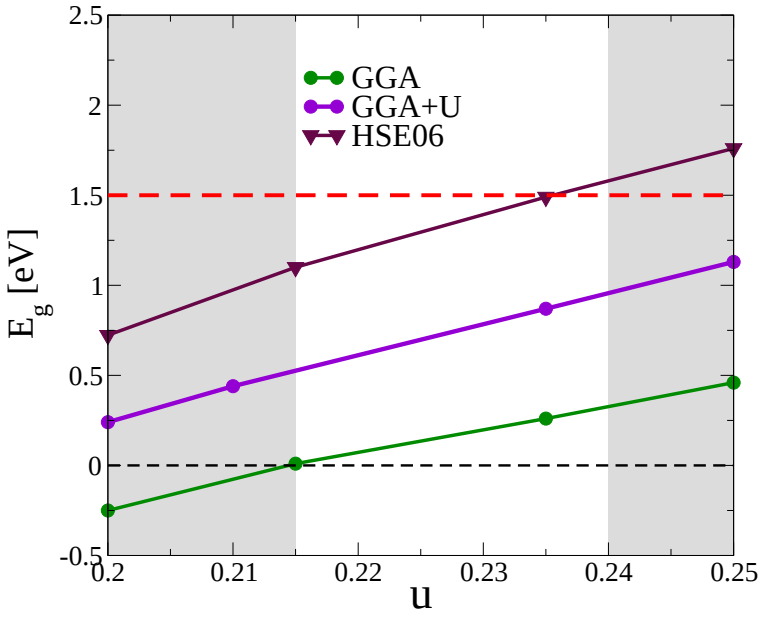


**Figure 5.12:** Bandgap of $CuInS_2$ versus the anion displacement $u$: impact of different functionals (Hybrid, GGA+U) on the dependence of the bandgap on $u$.

imentally observed stability of the bandgap and the theoretically predicted strong dependence of $E_g$ on $u$, we will present the GW results, which is the state-of-the-art method to treat the bandgap.

### 5.4.2 Prediction of bandgap by means of *GW* methods

DFT-LDA or -GGA seems to be a particularly *bad* starting point: it predicts a semi-metal while the real nature of CIS is semiconductor. Therefore, a simple first order perturbation correction will not be sufficient to correct the bandgap as the occupation number of the bands has to be modified too. Two alternatives are possible when such case is encountered: the choice of a different starting point [199, 167, 131] or the use of an iterative scheme of *GW* [178]. We have chosen an hybrid solution, *i.e.* we construct a better starting point with a simplified iterative *GW* before performing a first order perturbation $G_0W_0$, namely sc-COHSEX+$G_0W_0$ [155, 180] (see section 4.5.3).

We use as a starting point DFT-LDA with pseudopotentials including the whole $n = 3$ shell for Cu and the whole $n = 4$ shell for In. First, we

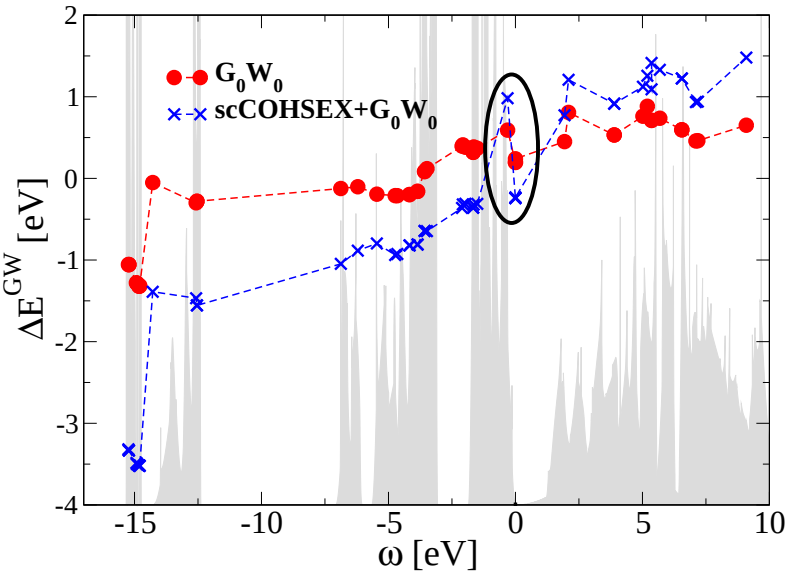


**Figure 5.13:** GW corrections versus the DFT-LDA energies. Red dots represent the $G_0W_0$ corrections on top of DFT-LDA while blue crosses represent sc-COHSEX+$G_0W_0$ corrections. The shaded pattern in the background represents the DFT-LDA density of states.

perform a perturbative $G_0W_0$. As already expressed in the previous section, the perturbative approach may seem inadequate due to the relatively bad starting point that represents DFT-LDA. However, the $G_0W_0$ results reveal some interesting points about the physics of CIS.
First, we get a confirmation of the presence of a *negative* bandgap in DFT. Figure 5.13 depicts the *GW* correction of each state as a function of the DFT energies of the states. Occupied and unoccupied states should present a fairly different *GW* correction: however one of the occupied states, the *true* CBM, despite its non zero occupation number displays a different GW correction compared to other occupied states. It originates from its rather delocalized character which make its correction ressemble those of the other conduction bands. While it is not very pronounced in the perturbative approach, it is emphasized by the self-consistency.

Already at this perturbative level, the occupation numbers of states close to the Fermi level at Γ should be modified accordingly to the new ordering of the energies induced by the *GW* correction. [5] Doing so, the bandgap opens

[5] One can consider the reorganisation of the band according to the new Fermi level as the first iteration of the self-consistency.

up on a wider region of $u$. Nevertheless, $G_0W_0$ still underestimates drastically the bandgap.
The second interesting point about $G_0W_0$ occurs for large values of $u$, *i.e.*,k $u > 0.240$ where the initial DFT bandgap is already positive. While the dependence upon $u$ of $E_g^{G_0W_0}$ is very similar to the one of DFT in the range where DFT predicts a negative bandgap, this dependence becomes suddenly stronger for $u > 0.240$ (cf Tab. 5.6). Such dramatic increase is again in disagreement with the experimentally observed stability of the gap: considering the range for $\Delta u$, then the variation of $E_g$ reaches $\sim 0.6$ eV.
In this spirit, we performed a scCOHSEX+$G_0W_0$ calculation. It opens up the bandgap significantly with respect to the perturbative approach, giving reasonable results compared to experiments for the experimental range of $u$. The scCOHSEX+$G_0W_0$ corrections are depicted in Fig. 5.13: they are strongly state-dependent and because of the rich structure of the CIS density of states, vary significantly from one state to the other. A large positive correction to one occupied state is observed close to the Fermi level, which is the fingerprint of the negative bandgap character predicted in DFT. Looking now to the variation with respect to $u$, on the region delimited by the two shaded areas of Fig. 5.12, HSE06 and scCOHSEX+$G_0W_0$ are in overall good agreement with each other, the only exception being the slope of the bandgap dependence. This slope is similar to the one we found in $G_0W_0$ when the starting point predicts a semiconductor character for CIS. The sc-COHSEX increases the slope while the perturbative $G_0W_0$ on top of it, retains the slope but closes the bandgap by 0.3 eV. Therefore, dynamical correlation effects do not play a major role in this dependence which is due exclusively to static many-body effects. The self-consistent COHSEX treatment catches therefore the physics of CIS in a much simpler formalism than the full dynamical self-consistent GW. On the other hand, neither DFT, DFT+U, nor hybrids can describe it successfully. In order to understand its origin, we will try to answer some of the questions that have been raised in the previous two sections.

**In conclusion**,

- The self-consistent *GW* scheme (, *i.e.*, sc-COHSEX+$G_0W_0$) predict bandgaps in agreement with experiments.

- The dependence of $E_g$ upon $u$ is strongly increased within the sc*GW* formalism with respect to the other methods.

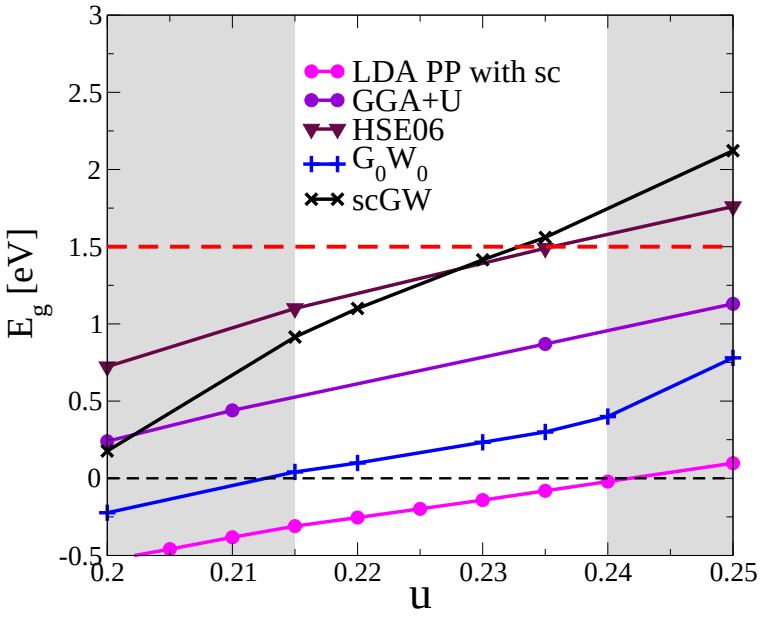


**Figure 5.14:** Bandgap of $CuInS_2$ versus the anion displacement $u$: impact of many-body effects and flavors of *GW* on the dependence of the bandgap on $u$

## 5.5 Many-body effects on the energy gap of CIS

### 5.5.1 Why is the bandgap so dependent on the value of $u$ ?

The answer to this question was partially given by Jaffe *et al.* [216]: the character of the states forming the VBM and the CBM are similar to the character of the states forming the bond. While this statement may look intuitively correct, it is however difficult to prove it within DFT as the exchange and correlation potential is density dependent and not orbital dependent. In order to get more insight into the physics of the problem, we choose to investigate the contribution of the Hartree-Fock self-energy to the VBM and CBM.

The exchange part of the self energy is given by expression (4.74). The quantity of interest will be the matrix element of $\Sigma_x$ for VBM and CBM. We can define the following quantity

$$\sigma^x_{i,\mathbf{k},(c,v)} = -\int d\mathbf{r}_2 d\mathbf{r}_1 \phi_{i,\mathbf{k}}(\mathbf{r}_1)\phi^*_{c,v}(\mathbf{r}_1)v_{\mathbf{k}}(\mathbf{r}_1,\mathbf{r}_2)\phi^*_{i,\mathbf{k}}(\mathbf{r}_2)\phi_{c,v}(\mathbf{r}_2) \tag{5.3}$$

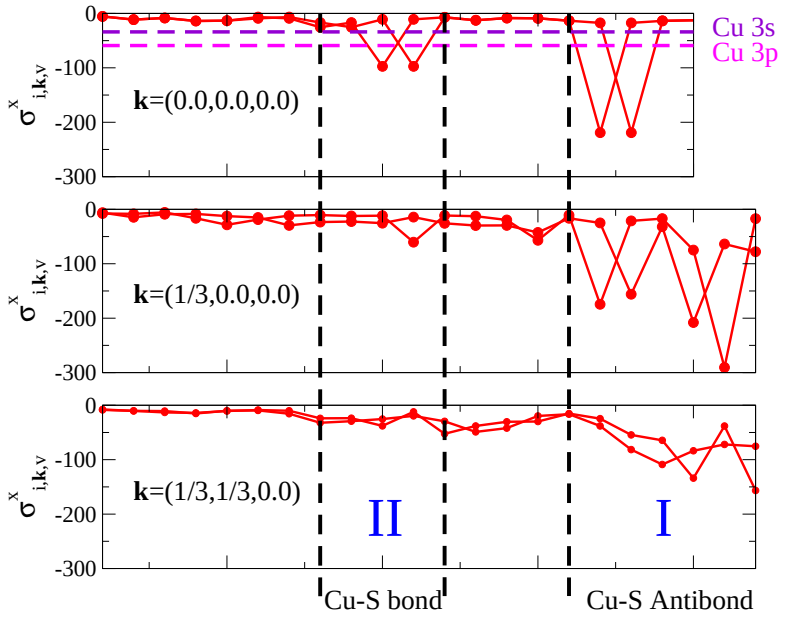


**Figure 5.15:** $\sigma^x_{i,\mathbf{k}}$ versus $i$ for $\Gamma_{4v}$. Magenta line refers to the contribution of Cu $3p$ and violet line refers to the contribution of Cu $3s$

Then, matrix element $\langle c,v|\Sigma_x|c,v\rangle$ is expressed by means of $\sigma^x_{i,\mathbf{k}_i,(c,v)}$

$$\langle c,v|\Sigma_x|c,v\rangle = \sum_{i\in \text{ occ}} \sigma^x_{i,\mathbf{k}_i,(c,v)} \tag{5.4}$$

where the sum over $i$ restricts to occupied states only.

Figure 5.15 shows $\sigma^x_{\mathbf{k},i}$ as a function of the index of states in the case of the valence band maximum $\Gamma_{4v}$. The self-interaction case *i.e.* $\phi_i = \phi_v$ has not been displayed as it is cancelled by the Hartree potential and it is order of magnitude larger than the other contributions. However, the contribution from the Cu $3s$ and Cu $3p$ states have been depicted: there are many reports in the literature about the importance of including the core contributions in order to treat correctly the exchange interaction [165, 164]. Besides, it gives an idea about the magnitude of the contributions originating from other states. Considering Fig. 5.15, one can see that there are many *significant* (*i.e.*, higher than the semicore ones) contribution to $\Sigma_x$. They can be categorized in two groups:

- group I situated down to 1 eV below $\Gamma_{4v}$ corresponding to anti-bonding states.

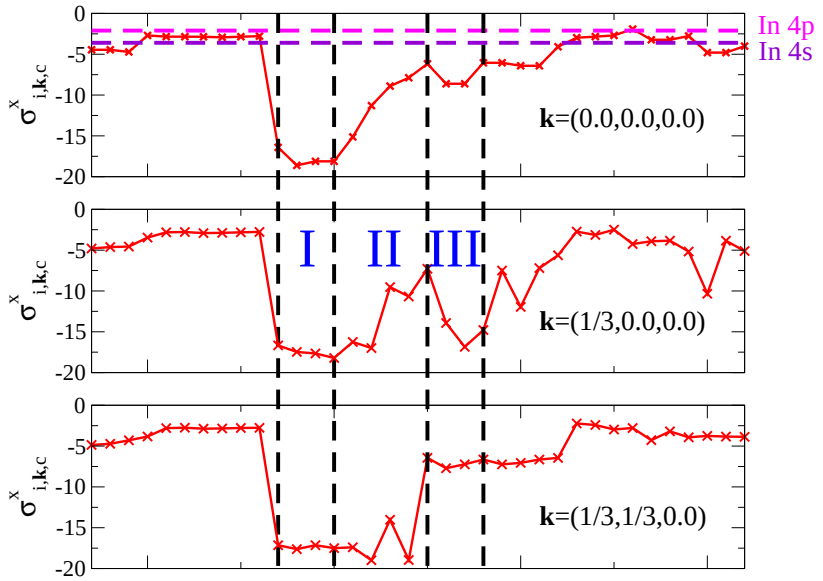


**Figure 5.16:** $\sigma^x_{i,\mathbf{k}}$ versus $i$ for $\Gamma_{1c}$. Magenta line refers to the contribution of In $4p$ and violet line refers to the contribution of In $4s$

- group II situated 3.5 eV below $\Gamma_{4v}$ at $\Gamma$ only corresponding to bonding states

Thus, the *GW* correction to the VBM is dependent not only on the anti-bonding states, as expected by the anti-bonding character of $\Gamma_{4v}$, but also on the bonding states. A similar analysis can be carried out for the conduction band maximum (Figure 5.16) and yields similar results.

First, the absolute value of the contribution is much smaller compared to $\Gamma_{4v}$. $\Gamma_{1c}$ is a conduction band with a delocalized character and therefore $\langle\Gamma_{1c}|\Sigma_x|\Gamma_{1c}\rangle$ is relatively small ($\approx -7$ eV) compared to $\langle\Gamma_{4v}|\Sigma_x|\Gamma\rangle \approx -32$ eV. $\Gamma_{1c}$ is formed by hybridized In $5p$ and S $3p$ states. In $4s$ and Cu $4p$ contributions have been displayed too. Despite the rather small overlap between Indium $4(s,p)$ and $5p$, core states from indium participate significantly in the matrix element $\langle\Gamma_{1c}|\Sigma_x|\Gamma_{1c}\rangle$. It gives an indication about the effect of semicore states on the GW corrections [163] and the adequacy of the pseudopotential approach versus all electron methods [158, 159, 223]. Three main groups of bands contribute significantly to $\langle\Gamma_{1c}|\Sigma_x|\Gamma_{1c}\rangle$.

- group I situated 13 eV below the Fermi level corresponding to In $4d$ and S $3s$ bonding states

- group II situated 6 eV below the Fermi level corresponding to In $5p$

and S 3$p$ bonding states.

- group III situated 4.5 eV below the Fermi level corresponding to pure S 3$p$ states.

The *GW* correction to CBM depends essentially on In-S bonding states. The main contribution arises from the In 4$d$-S 3$s$ which, as have shown in the previous section, constitute the main contribution to the electronic density between In and the anion.

**In conclusion**, the strong dependence of the bandgap calculated within *GW* can be explained by the large contribution of the bonding orbitals to some part of the self-energy. The treatment of the exchange and correlation by means of density-dependent potential hinders the interaction between VBM/CBM and In-(S,Se) or Cu-(S,Se) bonding states.

### 5.5.2 How do VBM and CBM depend on $u$?

While *GW* was initially used to evaluate bandgaps, recent works [188, 224] have drawn attention to the relative position of the band edge in order to calculate bandoffsets at heterojunctions or to correct defect formation energies. The *GW* correction to the band edge is however much more sensitive to many parameters than the bandgap: for example, the convergence of the QP band edge positions may be difficult to achieve [5] and the absolute positions may depend on the plasmon pole model [188]. In the following discussion, we will concentrate more on the $u$-dependence of the corrections to the VBM or CBM, which should not be affected by the aforementioned difficulties. Indeed, all calculations have been performed at the same degree of convergence and using the same technicalities. Figure 5.17 displays the GW correction to the CBM and the VBM for the different flavors of self-consistency: on energies only or on both energies and wave functions. The effect on the bandgap is also depicted in the left panel of Figure 5.17.

The *GW* corrections to the CBM and the VBM seem to be extremely sensitive to the type of self-consistency employed. On the contrary, the bandgap and its dependence on $u$ remain almost unaffected. The effects on the absolute values of CBM and VBM cancel each other, resulting in the innocuity of the self-consistency scheme on the bandgap. Therefore, full self-consistency is only mandatory if one is interested in the absolute *GW* correction to the band edges. Besides, the slope of the $u$-dependence of $\Delta E_v^{GW}$ and $\Delta E_c^{GW}$ appears to be very similar: indeed, the slope of $\Delta E_v^{GW}$ versus $u$ is -13.3 eV against 12 eV for $\Delta E_c^{GW}$. Both band edges participate equally in the large $u$

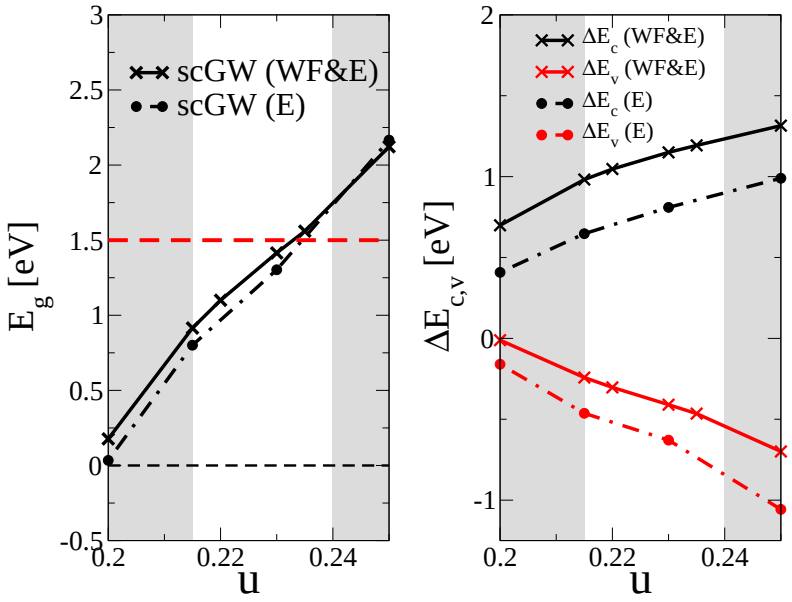


**Figure 5.17:** Left panel: Bandgap of $CuInS_2$ versus anion displacement $u$ for two flavors of self-consistency: dashed line, self-consistency on both wave functions and energies and solid line self-consistency on energies only. Right panel: GW corrections to the VBM and the CBM versus $u$ for different flavors of self-consistencies.

dependence of the bandgap. One should notice that the previously mentioned slope should be added up to the $u$-dependence predicted in DFT-LDA.

Let us focus now on the discrepancy between the two flavors of self-consistency. In order to investigate which role the change of wavefunctions plays in the self-energy operator, we introduce the notation $\delta E_{c,v} = \Delta E^{\mathrm{WF\&E}}_{v,c} - \Delta E^{\mathrm{E}}_{v,c}$ where WF&E and E refer to the type of self-consistency. Moreover, we introduce the partioning of the COHSEX hamiltonian

$$H^{\mathrm{COHSEX}} = H^{\mathrm{HF}} + \Sigma_c^{\mathrm{SEX}} + \Sigma^{\mathrm{COH}}\,, \tag{5.5}$$

where $H^{\mathrm{HF}}$ is the Hartree-Fock hamiltonian, $\Sigma_c^{SEX}$ the correlation part of the SEX self-energy $\Sigma_c^{SEX} = \Sigma^{SEX} - \Sigma_x^{\mathrm{HF}}$ and $\Sigma^{COH}$, the Coulomb-hole self-energy. We also introduce the notation $\Delta\langle\rangle$ such as for $H^{\mathrm{HF}}$

$$\Delta\langle H^{\mathrm{HF}}\rangle = \langle H^{\mathrm{HF}}_{\mathrm{WF\&E}}\rangle - \langle H^{\mathrm{HF}}_{\mathrm{E}}\rangle\,. \tag{5.6}$$

Figure 5.18 displays the decomposition of $\delta E_{c,v}$ according to $\Delta\langle H^{\mathrm{HF}}\rangle$,$\Delta\langle\Sigma_c^{\mathrm{SEX}}\rangle$ and $\Delta\langle\Sigma^{\mathrm{COH}}\rangle$. Let us first consider the conduction band: the difference in $\Delta E_c^{GW}$ is constant $\approx 0.3$ eV over the full range of $u$. It is explained by

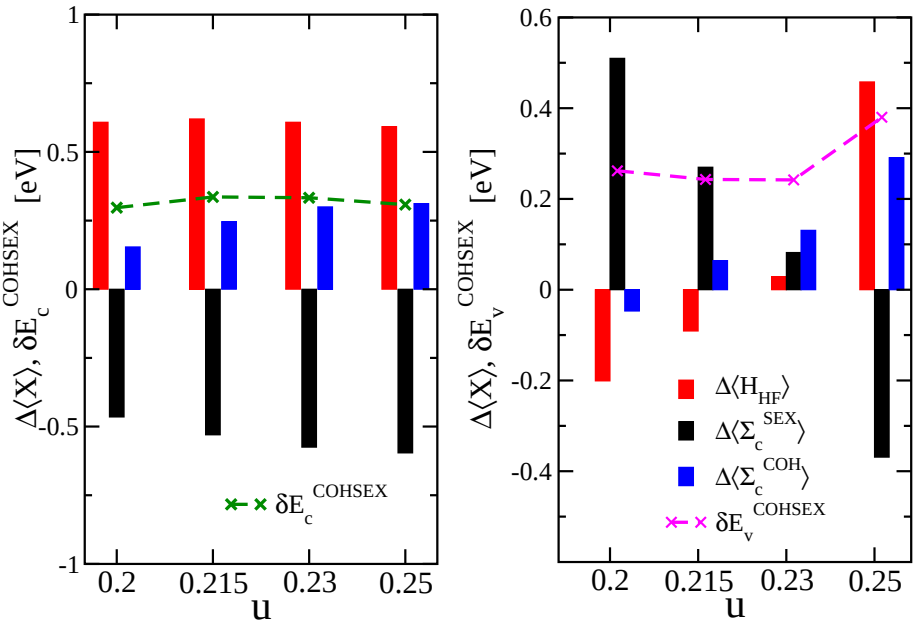


**Figure 5.18:** Left panel: Decomposition of the difference between the sc-COHSEX correction to the CBM obtained from tw flavors of self-consistency: WF&E and E only versus $u$. Right panel: the same the left panel but for the VBM.

an almost $u$-independent change in both three constituents of the COHSEX Hamiltonian. While the change in the wave functions significantly affects $\Delta\langle H^{\mathrm{HF}}\rangle$ and $\Delta\langle\Sigma_c^{\mathrm{SEX}}\rangle$ ($\sim 0.5$ eV), both quantities cancel each other, leaving $\delta E_c^{GW}$ mainly dependent on $\Delta\langle\Sigma^{\mathrm{COH}}\rangle$.

We can delineate two regions of $u$:

- the first one for small $u$ where $\delta E_c^{GW}$ mainly originates from $\Delta\langle\Sigma_c^{\mathrm{SEX}}\rangle$.
- the second one for large $u$ where it originates from $\Delta\langle\Sigma^{\mathrm{COH}}\rangle$.

On the other hand, the case of the valence band is more complex. One can see that the change in operators are extremely $u$-dependent. Again, important cross cancellations occur between changes in HF hamiltonian and in the correlation part of the screened exchange self-energy. The strong $u$-dependence of $\Sigma^{\mathrm{COH}}$ highlights the importance of the changes in the VBM wave function: indeed let us recall the form of Coulomb Hole self energy

$$\langle v|\Sigma^{\mathrm{COH}}|v\rangle = \frac{1}{2}\int d\mathbf{r}\left[W(\mathbf{r},\mathbf{r}) - v(\mathbf{r},\mathbf{r})\right]|\phi_v(\mathbf{r})|^2 \ . \tag{5.7}$$

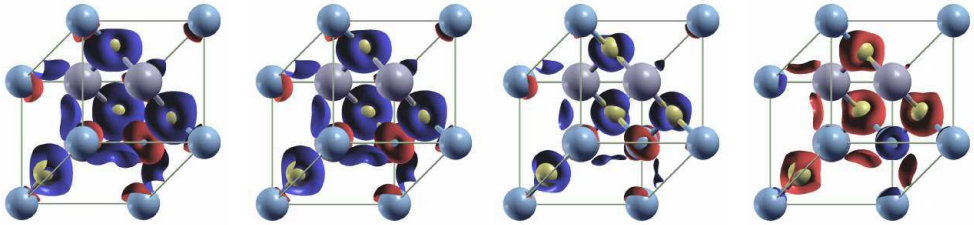

**Figure 5.19:** Isosurfaces of $\Delta\rho$ for $u = [0.200, 0.215.0.235, 0.250]$ from left to right. Red surfaces are the positive variations while blue surfaces are the negative ones

As Fig. 5.18 shows, the bandgap is only mildly influenced by the choice of the self consistent scheme, therefore the screening may be equivalent for both schemes. Therefore, changes in $\langle v|\Sigma^{\mathrm{COH}}|v\rangle$ can only be attributed to change in $|v\rangle$. Moreover, the variation of the latter will be weighted by $W_p$: changes in the value of the dielectric constant might affect $\Delta\langle\Sigma^{\mathrm{COH}}\rangle$. Figure 5.19 depicts the change of state $\Gamma_{4v}$ [c.f. (5.1)] as a function of $u$. For small $u$, there is an important electronic transfert from the S atom to the Cu atom, increasing the Cu $3d$ character of the VBM. However, for large $u$ (*i.e.*, for $u = 0.25$), the opposite transfer occurs: there is an electronic transfer from the Cu atom to the S atom. The changes in the wavefunction influences also the other operators but in a much more complex way as several states participate in their construction.

Let us now turn to another intriguing feature of Figure 5.18: the strong $u$-dependence of $\mathrm{H}^{\mathrm{HF}}$ and $\Sigma_c^{\mathrm{SEX}}$. In fact, decomposing further the Hartree-Fock Hamiltonian into a Hartree-Fock self energy and a Hartree Hamiltonian, one can see that big differences occur in both $\Sigma^{\mathrm{HF}}$ and $H^{\mathrm{Hartree}}$ but the changes in the latter cancel partially the changes in the former (see Fig. 5.20). One can think of $\Sigma_c^{\mathrm{SEX}}$ as a Fock exchange-like term where the bare Coulomb potential $v$ has been replaced by $W - v$. For small $u$, the bandgap is small, making the dielectric constant large and consequently $W - v \approx -v$. Therefore, it is not surprising that $\Delta\langle\Sigma_c^{\mathrm{SEX}}\rangle$ has a different sign than $\Delta\langle H^{\mathrm{HF}}\rangle$ and its absolute value is significantly larger than the one of $\Delta\langle H^{\mathrm{HF}}\rangle$. For large $u$, the bandgap becomes larger which results in a smaller dielectric constant. Therefore, $\Delta\langle\Sigma_c^{\mathrm{SEX}}\rangle$ will be softened in the same way the $\Delta\langle\Sigma^{\mathrm{HF}}\rangle$ is smoothed by $\Delta\langle H^{\mathrm{Hartree}}\rangle$. It explains the relative similarity of $\langle H^{\mathrm{HF}}\rangle$ and $\langle\Sigma_c^{\mathrm{SEX}}\rangle$ for $u = [0.235 - 0.25]$.

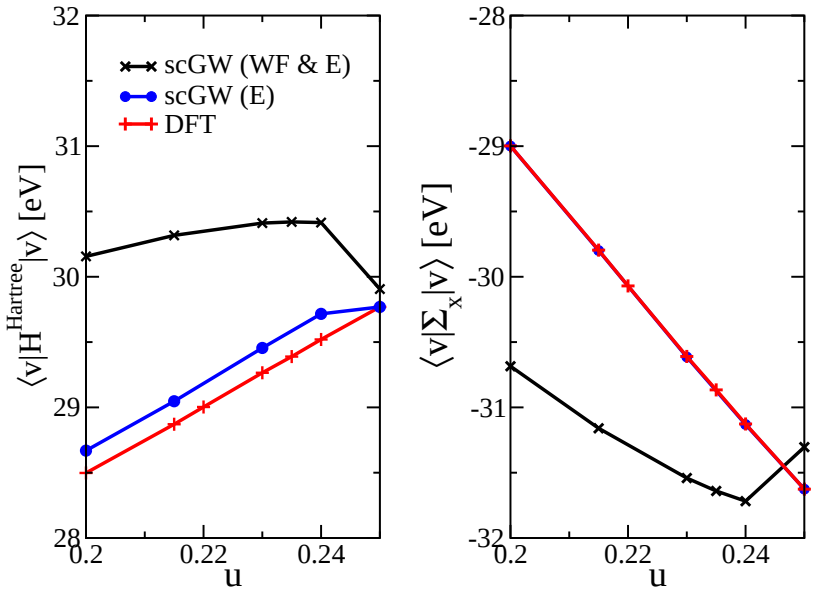


**Figure 5.20:** Matrix elements of the Hartree Hamiltonian (left panel) and the Fock exchange self-energy (right panel) at VBM. Red curve refers to the use of DFT-LDA wave functions and energies to construct $H^{\text{Hartree}}$ and $\Sigma_x$, blue curve refers to the use of DFT-LDA wave functions and energies, and black curve refers to the use of *GW* wave functions and energies.

**In conclusion**,

- The GW corrections to CBM and VBM account equally for the strong $u$-dependence of the bandgap.
- The effects of the QP wave functions are rather important on the band edge position but not on the bandgap.
- Changes in the wave functions induce an almost $u$-independent change of the operators for the CBM. It is not the case for the VBM.
- The VBM wave function changes significantly between DFT and sc-COHSEX.
- The screening plays a central role in the dependence of the valence edge upon $u$.

### 5.5.3 Why do GGA+U and Hybrid fail to predict such a strong dependence ?

Let us consider first the case of GGA+U. Anisimov *et al.* [93] established a connection between LDA+U and *GW*. They made the following approximation:

- the bandgap is formed by occupied and unoccupied $d$ states
- $d$ states are well separated from the other states, therefore only the contribution from $d$ states are retained in the calculation of the self-energy.

Then, the on-site interaction U is simply given by

$$U = \int d\mathbf{r}_1 d\mathbf{r}_2 \phi_d(\mathbf{r}_1)\phi_d^*(\mathbf{r}_1)W(\mathbf{r}_1,\mathbf{r}_2,\omega=0)\phi_d(\mathbf{r}_2)\phi_d^*(\mathbf{r}_2) \tag{5.8}$$

U depends strongly on the screening which may follow the variation of the bandgap. However, in Figure 5.12, the on-site interaction U was being kept fixed for all $u$. This explains why GGA+U cannot reproduce the slope of $E_g$ versus $u$.

The same analysis can be performed for the hybrids. While HSE06 and GGA+U do not seem to have much in common, they both share the use of fixed parameters. Considering HSE06, it contains two parameters: the mixing parameter $b = \frac{1}{4}$ and the range separation parameter $\mu$. Following Fuchs *et al.* [131], the exchange and correlation functional reads

$$E_{\text{xc}} = E_{\text{xc}}^{\text{GGA}} + (b \times E_x^{sr,HF} - b \times E_x^{sr,GGA})\,. \tag{5.9}$$

This expression ressembles the expression used in $G_0W_0$. The second term of the right hand side can be viewed as a correction to the GGA exchange and correlation functional. This quantity only corrects the short range part of the exchange potential. Then, $bE_x^{\text{HF}}$ can be considered as a model *GW* self energy, with a model screening $\tilde{W}$

$$\tilde{W} = b \times v \times \theta(\mu - \mathbf{r})\,, \tag{5.10}$$

where $\mu$ is the range separation parameter and $\theta$ is the function employed to realize the range separation[6]. For most hybrids, $b \approx 0.25$. Hence, it corresponds to a dielectric constant $\varepsilon_\infty = 4$ in the model GW. As already

[6]Here we have generalized the discussion, but for the case of the HSE hybrid, $\theta(r) = \text{erf}(r)$

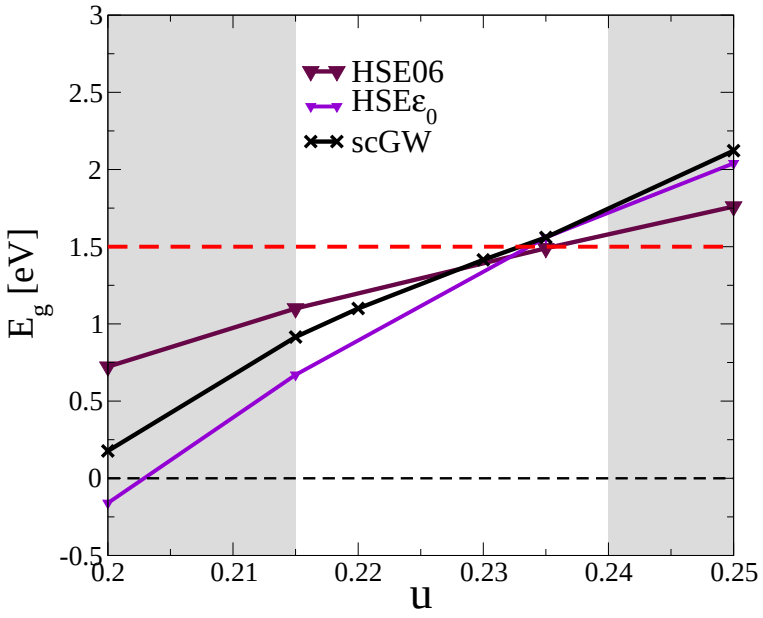


**Figure 5.21:** Bandgap of $CuInS_2$ versus the anion displacement $u$: impact of the parameter $b$ of HSE06 on the dependence of the bandgap on $u$.

pointed out in the discussion about LDA+U, the dielectric constant has a strong dependence on $u$. Employing a constant mixing parameter amounts to constraining the bandgap by enforcing the dielectric constant. It is therefore interesting to relax this constraint by setting $b = 1/\varepsilon_\infty^{RPA}$ with $\varepsilon_\infty^{RPA}$ calculated after the self-consistent COHSEX within the RPA approximation.

Figure 5.21 displays the bandgaps calculated within the modified HSE06 called HSE$\varepsilon_\infty$. For small $u$, the dielectric constant predicted in sc-COHSEX is rather large compared to the other value in the range of $u$. As a consequence, in this range of $u$, HSE$\varepsilon_\infty$ results are very close to GGA ones: indeed, $b$ for HSE$\varepsilon_\infty$ depends inversely on $\varepsilon_\infty$ so the contribution of the screened Fock exchange term (for HSE06) almost vanishes. Hence, we will exclude this point from the interpolation of the $E_g$ versus $u$ curves. The HSE$\varepsilon_\infty$ method yields surprisingly good agreement with scGW not only from the point of view of the slope $dE_g/du = 37.9$ but also with respect to the absolute position of the curves. We recover the *GW* results by varying the mixing parameter $b$ of HSE06 according to the dielectric constant $\epsilon_\infty$ calculated after the self-consistent COHSEX. The dielectric constant is rather underestimated due to the overestimation of the bandgap within COHSEX [170]. It can also be affected by the approximation used to estimate it: dynamical effects or low-

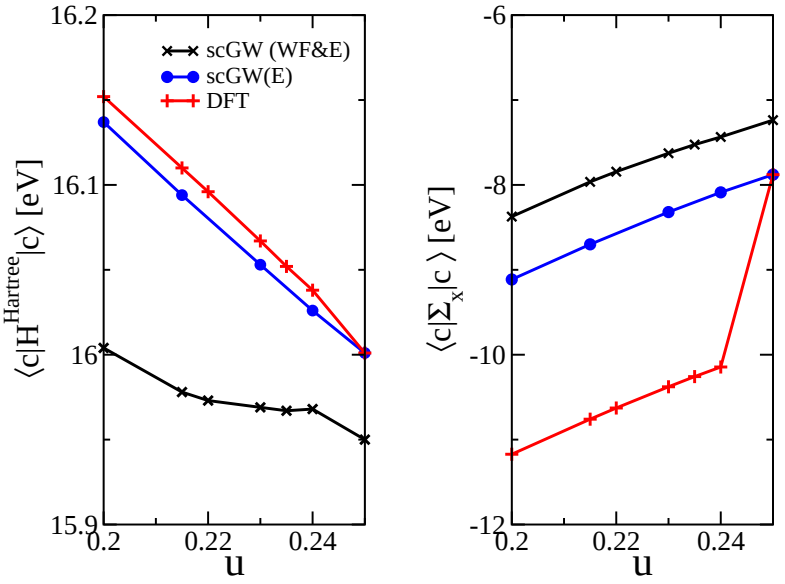


**Figure 5.22:** Matrix elements of the Hartree Hamiltonian (left panel) and the Fock exchange self-energy (right panel) at CBM. Red curve refers to the use of DFT-LDA wave functions and energies to construct $H^{\text{Hartree}}$ and $\Sigma_x$, blue curve refers to the use of DFT-LDA wave functions and *GW* energies, and black curve refers to the use of *GW* wave functions and energies.

ering of the screening at finite distance. Overall, the agreement in absolute value of the bandgap for HSE$\varepsilon_\infty$ might be incidental: the fact that HSE06 already contains a part of screening compensates for the underestimation of the dielectric constant. This fact is further confirmed by the insensitivity of the slope upon variations of the range separation parameter $\mu$ while the absolute position of the curve is strongly impacted.

### 5.5.4 Why is there a discontinuity in $G_0W_0$ ?

The last feature of Figure 5.14 that remains unexplained is the discontinuity in $G_0W_0$ at $u \approx 0.245$. First, it seems to occur when the DFT starting point predicts a positive bandgap. Second, it increases considerably the slope of the curve. Figure 5.22 displays the matrix elements of $H^{\text{Hartree}}$, $\Sigma_x$ for the CBM. Starting from DFT-LDA, the *real* conduction band is assigned a non zero occupation number.

Therefore, it will enter summation over states that are weighted by the respective occupation number of each bands: the electronic density that appears in the Hartree potential and the Fock exchange term. In order to

investigate this phenomenon, we will consider construction of the operator with DFT wave functions and energies which implies DFT occupation number (red line in Figure 5.22) or DFT wave functions and QP energies which implies just change in the occupation numbers (blue line in Figure 5.22) and finally QP wave functions and energies (black broken lines in Figure 5.22) Focusing on the Hartree term, we see little difference when the conduction state is included or excluded from the electronic density. However, there is a 2 eV upward shift of $\langle c|\Sigma_x|c\rangle$(difference between red and blue curve). It is self-interaction-like problem. The matrix element of the Fock term is particularly sensitive to the spatial overlap between the wave functions involved in $\Sigma_x$ and the wave functions $|c\rangle$. By changing the occupation number of the conduction state, it is included or excluded of the summation over states in the exchange part of the self energy.

## 5.6 Link between structural parameters and defects

The strong $u$-dependence of the bandgap is rather surprising if one considers the rather broad experimental range of $u$ on one side, and the observed stability of the bandgap on the other side. Besides, industrial PV applications require high stability of the electronic properties in order to obtain reproducibility. Therefore, our findings seem in total contradiction with the use of CIGS in the PV industry. Indeed, one question still remains unanswered from the experimental point of view: *why is CIGS so good for photovoltaic application despite large defect concentrations and presence of structural defects?*. Defects usually create deep levels inside the bandgap where carriers get trapped. CIGS presents also lots of structural defects like grain boundaries. These two facts are apparently totally disconnected from one another. We propose the existence of a hidden link between the concentration of defects and structural deviation in the lattice. In order to be realistic, we choose to consider only the copper vacancy as the predominant defect: this choice is driven by both experimental observation of copper poor samples and *ab initio* calculations that demonstrate the low formation energy of $\mathrm{V_{Cu}}$. Hence, let us assume that the bandgap is a function of $u$ and concentration of copper vacancy $[\mathrm{V_{Cu}}]$ so that $\mathrm{E_g} = \mathrm{E_g(u, [V_{Cu}])}$. A variation of the bandgap $\Delta\mathrm{E_g}$ can hence be expressed as

$$\Delta E_g = \frac{\partial E_g}{\partial u}\Delta u + \frac{\partial E_g}{\partial [V_{\mathrm{Cu}}]}\Delta[V_{\mathrm{Cu}}], \tag{5.11}$$

where $u$ is the anion displacement and $[\mathrm{V_{Cu}}]$ is the concentration of Cu vacancies. Moreover, it is possible to relate the concentration of $\mathrm{V_{Cu}}$ to the formation energy with a Boltzman distribution as

$$[\mathrm{V_{Cu}}] = \mathrm{N_{Cu}} e^{\frac{-\Delta \mathrm{H_f}}{\mathrm{k_B T}}} \tag{5.12}$$

with T is the growth temperature for which the diffusion of defects is quenched (T$\approx 500-600°$ C) and $\Delta H_f$ the formation enthalpy of $\mathrm{V_{Cu}}$. The latter quantity is the keystone of our reasoning as it allows one to vary the concentration of defects. We will see in the subsequent section how it relates to the structural parameter $u$.

### 5.6.1 Formation energy

First, we will introduce the concept of formation energy in a rather general way before focusing on the particular case of $\mathrm{V_{Cu}}$. Equation (5.12) shows that the formation energy is the energetic barrier necessary to overcome in order to create a defect. Intuitively, the formation energy can be simply thought as the difference of total energy between two systems: the system with one defect and the system without defect. However, in order to conserve the number of particle in our system, one should add or remove the energy of the single atom that has been removed, exchanged or added depending on the type of defects considered. This total energy can be calculated considering either an isolated atom or a more realistic reservoir like fcc metallic Cu. The formation energy of defect $\alpha$ is then expressed as

$$\Delta E_f(\alpha) = E_{\mathrm{tot}}(\mathrm{CIS}, \alpha) - \mathrm{E_{tot}}(\mathrm{CIS}) + \mathrm{n_{Cu}}\mu^{\mathrm{sol}}_{\mathrm{Cu}} + \mathrm{n_{In}}\mu^{\mathrm{sol}}_{\mathrm{In}} + \mathrm{n_{S,Se}}\mu^{\mathrm{sol}}_{\mathrm{S,Se}}\,, \tag{5.13}$$

where $\mathrm{E_{tot}}$ is the ground state energy of the supercell, $n$ are numbers of atoms that take part into the defects, $\mu$'s are the chemical potential of the reservoirs. Furthermore, the formation energy depends also on the chemical potential of the atomic species $\mu_{\mathrm{Cu,In,S,Se}}$: this potentials are bounded by the domain of stability of CIS and the domain of formation of possible binaries phase($\mathrm{Cu_2Se}$ or $\mathrm{In_2Se_3}$). Therefore, one can calculate the defect formations enthalpies for the specific deposition condition *i.e.* Cu-poor or Se-rich. The formation enthalpy then reads

$$\Delta H_f(\alpha) = \Delta E_f(\alpha) + n_{\mathrm{Cu}}\mu_{Cu} + n_{\mathrm{In}}\mu_{\mathrm{In}} + n_{\mathrm{S,Se}}\mu_{\mathrm{S,Se}}\,. \tag{5.14}$$

The case of a charged defect is more complex as an electron should be transferred to or from an electron reservoir. Therefore, the absolute energy of

this reservoir should be known as done previously for the reservoir of atoms. Intuitively , the previous expression can be extended by adding the term $\delta E(\alpha, q) = E_{\text{tot}}(\alpha, q) - E_{\text{tot}}(\alpha, q = 0)$ which corresponds to the energy necessary to add a charge $q$ on the defect site. Another term arises from the energy difference between the system with N electrons and the system $N - |q|$ electrons with $q$: it describes the relaxation of the host system under the inclusion or removal of one electron. The formation enthalpy reads

$$\begin{aligned} \Delta H_f(\alpha, q) =& \Delta E_f(\alpha) + \delta E(\alpha, q) + \delta E(CIS, -q) + \\ & n_{\text{Cu}}\mu_{Cu} + n_{\text{In}}\mu_{\text{In}} + n_{\text{S,Se}}\mu_{\text{S,Se}} + qE_f \, . \end{aligned} \tag{5.15}$$

The calculation of the defect formation energy involves many terms, each of them is subject to errors. Lany *et al.* [224] identified the two main sources of error for ZnO:

- the band edge corrections due to the DFT underestimation of the bandgap: other methods such as *GW* or EXX can significanly improve the treatment of the bandgap compared to DFT, however the quality of the total energy they can provide is still under debate. One can overcome this problem by using a smartly designed partitioning of the formation energy: introducing fictitious charge state, one can recast the formation energy expression and use the GW method to evaluate some quantities while DFT is used to estimate others [225].

- corrections due to supercell size: usually supercells up to 1000 atoms are used to calculate formation energy. It corresponds to very high concentration of defects ($10^{21} - 10^{20}$ cm$^{-3}$) and are rather far from the situation of an isolated defect ($\sim 10^{14}$ cm$^{-3}$). The supercell size is responsible for many spurious and undesired effects: Moss-Burstein-like band filling [226], supercell-supercell interaction of charged defects due to the periodic boundary condition [227] (like image charge), quadripole interaction or elastic contribution due to the induced pressure of the defect.

All those errors are usually corrected *ex post facto*. While it could be efficient if the effects previously mentioned are small, it fails in situation where the bandgap is strongly underestimated. Figure 5.23 shows the formation energy of $V_O$ in ZnO versus the Fermi energy calculated with different methods and corrections. One can see that the domain of stability of the neutral and charge defect varies considerably from one corrective scheme to the other.

In the case of interest, a simple scheme has been proposed to correct the formation energy for the case of a shallow acceptor like $V_{Cu}$ [224]. A defect

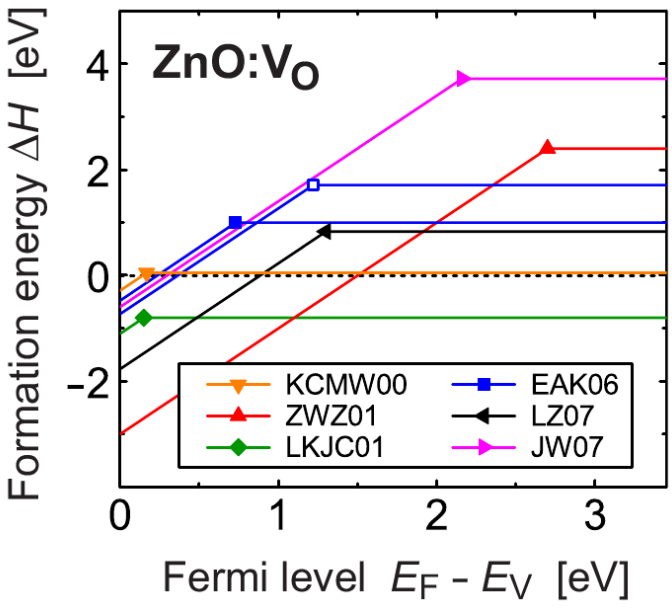


**Figure 5.23:** Formation energy $\Delta H$ of the O vacancy in ZnO under O-poor/Zn-rich conditions as calculated in theoretical works using different correction schemes. For more details about the legend, refer to [224].

usually creates a Defect Localized State (DLS). If this state occurs in the bandgap, then the defect is said to be deep. However, if the DLS occurs as a resonance in the continuum of the host bands, the electrons or holes localized on the DLS relaxes to a Perturbative Host State (PHS), which is the CBM or the VBM. Then, the defect is said to be shallow and in this particular case, the occupied shallow state is believed to follow the movement of the CBM or VBM. Therefore, changes in the VBM or CBM will have a direct effect on the formation energy. One can immediately see the connection between the formation energy and the change of the VBM position due to variations of $u$. Lany *et al.* proposed a scheme in order to correct the bandgap problem for shallow acceptor such as

$$\Delta E_{\mathrm{f}} = \Delta E_{\mathrm{f}}^{\mathrm{DFT}} - \Delta \mathrm{E}_{v}^{\mathrm{LDA+U}}\,, \tag{5.16}$$

where $\Delta E_{\mathrm{f}}^{\mathrm{DFT}}$ is the formation energy calculated within DFT and $\Delta \mathrm{E}_{v}^{\mathrm{LDA+U}}$ is the correction to the VBM obtained from a LDA+U calculation. While their goal was to correct the bandgap error inherited from the DFT, our goal is slightly different: not only get the correct bandgap but also the dependence of VBM as function of $u$. Therefore, we obtain the $u$-dependent formation

energy

$$\Delta E_{\mathrm{f}}(u) = \Delta E_{\mathrm{f}}^{\mathrm{DFT}} - \Delta \mathrm{E}_{v}^{\mathrm{scGW}}(u) \tag{5.17}$$

We assume that the main contribution to the total energy comes from the valence band maximum. We consider the evolution of the position of the VBM within DFT, taking as a reference the In $4d$ states: [7] the change in the position of the VBM is an order of magnitude smaller than $\Delta \mathrm{E}_{v}^{\mathrm{scGW}}$. Then, this corrective scheme is physically sound. We perform a GGA calculation on a 432-atom supercell, considering only the neutral copper vacancy. Indeed, the domain of predominance of the neutral defect with respect to the charged defect lies close to the VBM *i.e.* $E_f$ is close to the VBM which corresponds to experimental observations. Then, the formation energy $\Delta E_f(V_{\mathrm{Cu}}) = 0.75$ eV. In order to evaluate the formation enthalpy, we need to define the chemical potential of copper $\mu_{\mathrm{Cu}}$ by considering the growth conditions. We set $\mu_{\mathrm{Cu}}$ to -0.6 eV considering experimental evidences in the case of Cu-poor growth conditions [228]. As a consequence, the formation enthalpy is very low $\Delta H_f(V_{\mathrm{Cu}}) = 0.1$ eV, which will yield large concentration of defects. We have to make sure that the domain of concentration of $\mathrm{V_{Cu}}$ we are working with corresponds to the experimentally observed Cu-deficiency. It will also determine the size of the supercell that we should use to describe the influence of the copper vacancy on the bandgap. At $T = 600$°C, expression (5.12) yields concentration of copper vacancy such as $[\mathrm{V_{Cu}}] \approx 5 \times 10^{20}$ cm$^{-3}$ which corresponds to rather small supercell (16 to 64-atom supercell). From the experimental point of view, samples usually display a content ratio of the metallic cation of $0.88 <$ Cu/(In+Ga)$< 0.95$ [45]. In Ref. [203], the ratio goes down to 0.77. Hence, the Boltzmann distribution describes correctly the important copper deficiency observed in real samples. Besides, the rather high doping allows us to apply computationally expensive GW methods to compute the bandgap.

### 5.6.2 Evaluating the evolution of $\mathrm{E}_g$ as a function of $V_{\mathrm{Cu}}$

For very high doping, a shrinkage of the bandgap is usually observed. In the case of CIS, under high concentration of $\mathrm{V_{Cu}}$, any *ab initio* methods from DFT-KS to the most advanced *GW* techniques fail to reproduce such trend. Instead a large opening of the bandgap occurs. This has been attributed to the relaxation of the *p-d* repulsion which is the key mechanism behind the bandgap anomaly in CIS. We performed $\mathrm{G_0W_0}$ on supercells whose size ranges between 16 and 64 atoms. These supercell sizes correspond to 25% (16 atoms),12% (32 atoms) and 6% (64 atoms) deviations of stoichiometry. The

[7] We have already shown that In $4d$ states are not affected by variations of $u$

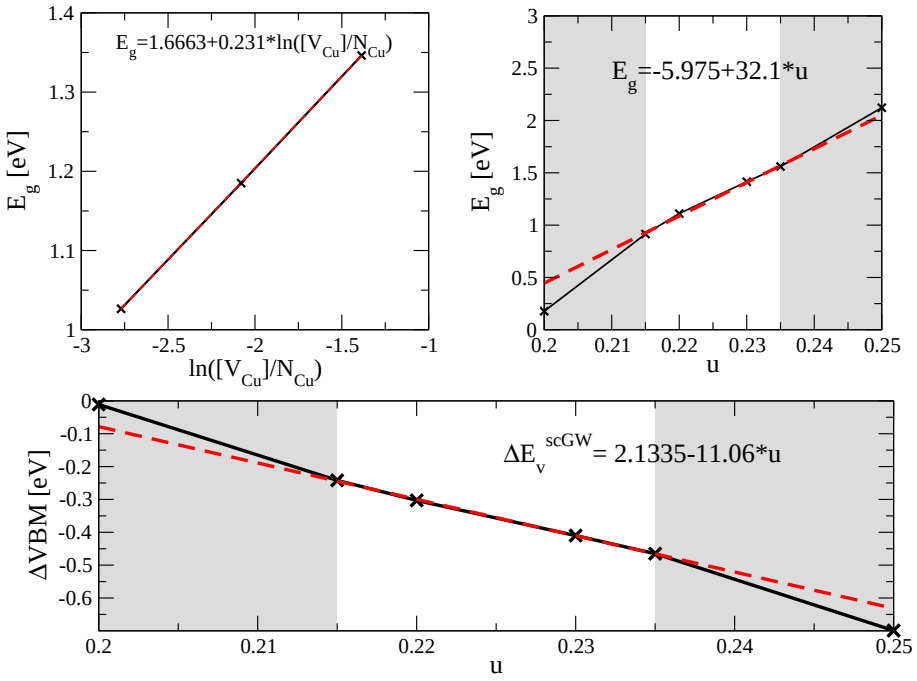


**Figure 5.24:** Top left panel: Value of the bandgap of $CuInS_2$ calculated in $G_0W_0$ versus the concentration of copper vacancy. Top right panel: Value of the bandgap of $CuInS_2$ calculated in scGW versus the anion displacement. Bottom panel: $\Delta E_v^{\mathrm{scGW}}$ of $CuInS_2$ versus the anion displacement.

use of the perturbative approach is motivated by the fact that the starting LDA gap of the system with vacancies is already positive. As shown in previous section for the case of the pure compound when $u < 0.235$, a perturbative approach can fully catch the physics of the material and is computationnally more tractable for such large size of supercells.

The exact relation between $E_g$ and $[V_{Cu}]$ is not known. However, using a logarithmic scale, we found a linear dependence of $E_g$ on $[V_{Cu}]$. We can perform a linear regression as shown in Figure 5.24

$$E_g = 0.231 \times \ln\left(\frac{[V_{Cu}]}{N_{Cu}}\right) + 1.67 \tag{5.18}$$

$$\frac{\partial E_g}{\partial [V_{Cu}]} = \frac{0.231}{[V_{Cu}]} \tag{5.19}$$

We have obtained a simple formulation of the partial derivative of $E_g$ with respect to $[V_{Cu}]$ in the range of concentrations $V_{Cu}$ that we are interested in. In order to simplify further expression (5.21), we should find a way to express $\Delta[V_{Cu}]$ as a function of $\Delta u$. Therefore, in the next section we will

investigate the effect of $V_{Cu}$ on the anion displacement.

### 5.6.3 Does $V_{Cu}$ have an influence on $u$ ?

We assume that $u$ can act on the concentration of copper vacancies. But *does* $V_{Cu}$ *induce changes in the value of* $u$ *?* Indeed, the presence of a copper vacancy produces a void in the crystal, that results in the inward relaxation of the neighbor anions. Besides, the crystal loses all the proper symmetries of the chalcopyrite phase which can ensue long range structural changes. In order to assess this problem, we extract the distribution of *local* $u$ from a relaxed 64-atom supercell with one $V_{Cu}$. The anion displacement is defined as a *local* quantity: each anion is surrounded by 4 cations (2 copper and 2 indium), resulting in two In-S and two Cu-S bonds. In the case of a perfect crystal, the lengths of the former is equal for each anion and the same for every anion resulting in one unique $u$. However, in a defect supercell, these bond lengths vary over the supercell. One can then define for each anion, 4 different values of $u$. Figure 5.25 shows the distribution obtained with a GGA relaxation for different sizes of supercell. The distribution presents a mean value close to that of the perfect crystal, $u^{GGA}$=0.2184, and the standard deviation decreases with increasing supercell size. Furthermore, additional distributions appear at small and large values of u ($u < 0.216$ and $u > 0.219$). They originate from the region close to the vacancy. The maximum variation of $u$ recorded with respect to the mean value $u^{GGA}$ is less than 0.003 which represents a change in bandgap of 0.1 eV.

We can conclude that $[V_{Cu}]$ has only a mild effect on the anion displacement except in a region localized around the copper vacancy. We can therefore simplify expression (5.11) by assuming that $u$ is not a function of $[V_{Cu}]$ and that $[V_{Cu}]$ is an exclusive function of $u$. Therefore, Eq. (5.11) can be rewritten as

$$\Delta E_g = \left\{ \frac{\partial E_g}{\partial u} + \frac{\partial E_g}{\partial [V_{Cu}]} \frac{d[V_{Cu}]}{du} \right\} \Delta u \,, \tag{5.20}$$

$$\Delta E_g = \left\{ \frac{\partial E_g}{\partial u} - \frac{\partial E_g}{\partial [V_{Cu}]} \frac{[V_{Cu}]}{k_B T} \frac{d\Delta E_v^{scGW}}{du} \right\} \Delta u \,, \tag{5.21}$$

and with Eq. (5.19), one obtains

$$\Delta E_g = \left\{ \frac{\partial E_g}{\partial u} - \frac{0.231}{k_B T} \frac{d\Delta E_v^{scGW}}{du} \right\} \Delta u \,. \tag{5.22}$$

If the second term of Eq. (5.22) is important, we can hope that it decreases the contribution of the first term. Physically speaking, it results in saying

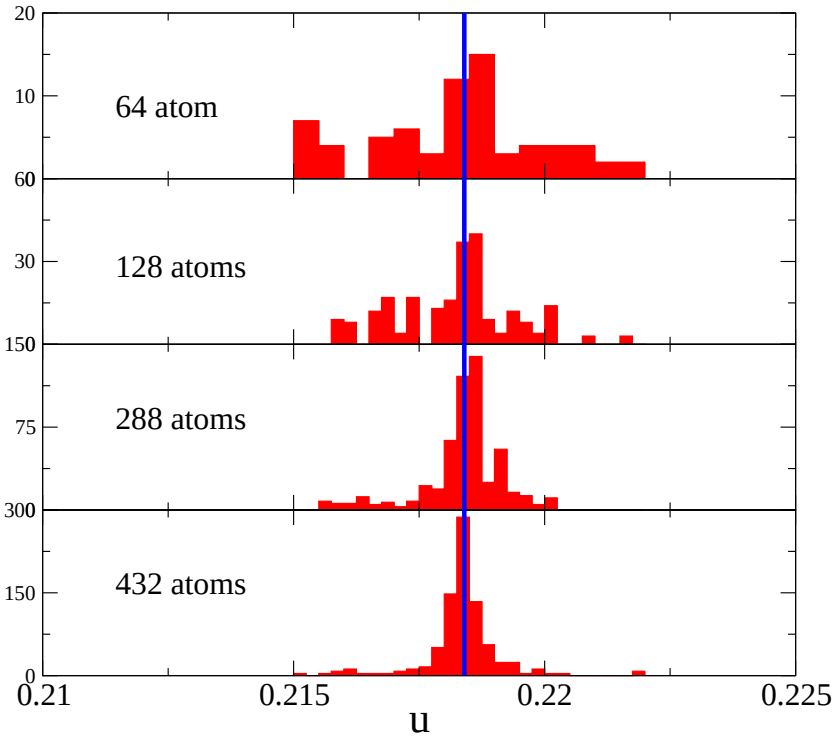


**Figure 5.25:** Distribution of *local* $u$ depending on the size of the supercell of $V_{Cu}$:$CuInS_2$. Blue line refers to the value of $u$ for the perfect crystal relaxed in GGA.

that there is an effective link between the anion displacement and the concentration of defects. In the next subsection, we will consider such compensating model quantitatively.

### 5.6.4 The feedback loop

The feedback loop process is recalled in Figure 5.26.

In order to evaluate the variations of $\Delta E_g$ with and without the feedback loop, we have to evaluate numerically the last two partial derivatives. The variations of both $E_g$ and $\Delta E_v^{\mathrm{scGW}}$ are almost linear with respect to $u$. To perform the linear regression, we choose a restricted range of values of $u$ that encompasses both the theoretical and the experimental values *i.e.* [0.215-0.235]. Doing so we obtain

$$\Delta E_g = 32.2 \times \Delta u \text{ without feedback loop}, \quad (5.23)$$

$$\Delta E_g = -1.9 \times \Delta u \text{ with feedback loop and T=600°C}. \quad (5.24)$$

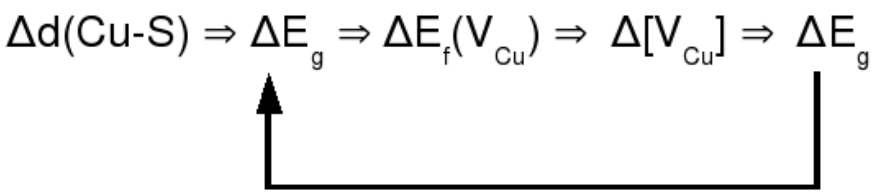


**Figure 5.26:** Scheme of the feedback-loop that stabilizes the gap.

Considering a variation of $\Delta u = 0.02$, the bandgap will change by 0.65 eV without feedback loop and by only 0.07 eV with the feedback loop (for T=600°C). Therefore, the effect of the concentration of copper vacancy counterbalances not only qualitatively but also quantitatively the variation of $u$. The linear dependence of the bandgap upon $[V_{Cu}]$ calculated within DFT is much stronger than the one calculated within GW. Thus, if one had used DFT bandgap instead of GW ones, one would have ended up with a $\Delta E_g$ much larger than the one obtained here.

**In conclusion**,we have demonstrated the strong link between defects and structural parameters. The experimentally observed stability of CIS with respect to variation of stoichiometry and morphology of thin films can be explained by such mechanism. The use of a method beyond standard DFT is needed in order to treat correctly the localization problem of $3d$ states and the underestimation of the bandgap. Moreover, it reveals the strong dependence of the bandgap on the anion displacement. Many of the results obtained for CIS might hence require to be revised using MBPT techniques.

# Chapter 6

# Conclusion

This work has led to various conclusions, ranging from understanding the physics of CIS in much more detail than previously to insights concerning new emerging *ab initio* technics such as hybrid functionals. The goal of the thesis was to investigate the physics of the electronic structure and of defects in CIS, concerning the latter in particular the energy levels that can appear inside the bandgap. Calculating defects still remains extraordinary CPU time-consuming even for simple materials such as silicon. Regarding the difficulties encountered to obtain a correct *ab initio* treatment of CIS, one may think the study of defects in CIS is a rather *impossible mission*. However, we expect this thesis could provide the necessary tools to tackle this problem in a very near future. From the point of view of the physics of CIS, we demonstrated that

- the geometry of CIS is not well described by KS-DFT using standard functionals. Usually, KS-DFT e.g. in the LDA is one of the standard methods to predict crystal structure: experience has shown that despite its simplicity it performs surprisingly well for a large class of materials. CIS is not the only material where problems arise, but previous *ab initio* calculations had not pointed to such an obstacle in the case of CIS.

- the wrong description of the anion displacement by KS-DFT stems from the lack of localization of the wave functions. The non proper treatment of the self-interaction problem due to a poor description of the exchange interaction results in the delocalization of the $3d$ and $4d$ states. It was known for a long time that Cu $3d$ were actively participating in the bonding, on the other hand, we prove the importance of the In $4d$ states in the In-(S, Se) bonding. The problem of localization has been resolved by the use of hybrid functionals or LDA+U. The latter method

has shown promising results at the same level as Hybrid functionals but at the same computational cost as KS-DFT.

- the experimental results display a broad range of values of $u$: by considering the only X-ray diffraction technics, we realize that obtaining the anion displacement was an incredibly difficult tasks. We conclude that the experimental range of $u$ was not only due to experimental accuracy but also to the material itself. Indeed, only few experiments study single crystal samples, the difficulty in obtaining good single crystal samples and the advent of CIS polycrystalline and thin film samples in the photovoltaic industry results in such a wide distribution of $u$.

- while KS-DFT already displays an *alarming* strong dependence of the bandgap on $u$, *GW* increases this trend. We reveal that the origin of this dependence is linked to the nature of the valence band maximum and the conduction band minimum and their space proximity with states participating in the bonding. The agreement with experiments considering this effect is however extremely bad, as CIS is famous for the stability of its bandgap against the morphology of thin films for example. It is one of the many reasons (and maybe the most important one) why CIS is such an invaluable material for photovoltaic applications.

- Hybrid functionals (and to some extent LDA+U) show their limits when screening is central to the physics of the material. Indeed, screened hybrids fail to reproduce the strong dependence of the bandgap versus the anion displacement, which was found to originate from the fixed mixing parameter for the Fock exchange part.

- the deformation of the structure induces changes in the defect concentration through a complicated process, resulting in compensation. The deviation of stoichiometry systematically observed experimentally should also produce a large effect on the bandgap: however, it is not seen experimentally. Therefore, we propose a mechanism based on the observed concurrent effects of variations of $u$ and variations of the concentration of $V_{Cu}$.

In summary, this work has given insight into the interplay between geometry, defects and electronic structure of CIS. The seeming contradiction between the large variety of samples and the observed stability of the bandgap could be resolved. Moreover, this work might leave more open issues than answers to existing questions. Many *ab initio* works have been published

for the last 20 years, relying on DFT only. Unfortunately, this thesis shows that not only is DFT *bad* at predicting bandgaps (as expected) but also, in the particular case of CIS and with standard functionals, *bad* at treating the ground state itself. The latter has many implications, mainly about the validity of previous studies: *what about the phonon calculations realized with DFT-relaxed lattice structures* [212, 229, 230], for instance. Most of them might still be valid, but doubt is permitted concerning some of them. For instance, we have obtained some preliminary results on defect formation energies which demonstrate that such doubts are legitimate. However, the *ab initio* treatment of CIS is not *hopeless* with respect to the complexity of the methods one should use to correctly describe the electronic properties of CIS: as we found, for instance LDA+U can efficiently replace hybrid functional. Such discovery opens the way to an improved treatment of larger systems such as surfaces, interfaces and grain boundaries. A careful check of the LDA+U formalism in the presence of intrinsic defects should be performed before moving to such systems. Indeed, it will not be surprising that the state charges of In and Cu change and consequently U itself. As one can see, there are lots of calculations still waiting to be performed to fully understand the physics of CIS. However, we wish that this thesis could open new paths in the *ab initio* study of CIS and could help to design better PV cells.

# Appendix A

# Pseudopotentials

Usually for most of the electronical properties of a solid, core electrons do not play any important role. We can estimate most of the properties just looking at the behavior of the valence states. Moreover, all electron methods, despite the use of localized basis that considerably decrease the computational charge, has restricted *ab initio* studies to small systems up to recently. The idea of the pseudopotential is simple: replace the effect of the core electrons and the nucleus on the valence electrons with an effective potential called pseudopotentials. Technically, the pseudopotential is enforced to be relatively smooth to limit the number of basis functions.

## A.1 How to construct a pseudopotential

The general method used independently of the type of pseudopotentials to generate is the following:

- use the frozen core approximation [231]: in order to do so, determine which electrons can be considered as valence and which as core electrons. By considering the energy level of each state, the separation can look obvious but in the particular case of $GW$, difficulties may arise when dealing with $d$ electrons.

- approximate the wavefunctions in the core region: the pseudized wavefunction should ressemble as much as possible the all-electron wavefunction, but being as smooth as possible in the core region. Multiple schemes have been developped in order to pseudize the wavefunctions. We will focus on two of them : the Hamman schem [232] and the Trouillier Martins scheme [233] which are both norm conserving.

- invert the Schrödinger equation to obtain the potential. The ultimate procedure then is to unscreen the potential from the Hartree and the exchange and correlation potential.

Starting from the following Schrödinger equation, assuming spherical condition as it is the case for the isolated atom:

$$-\frac{1}{2}\nabla^2 + \frac{l(l+1)}{r^2} + (v^{\text{eff}}(r) - \epsilon_i)u_{n,l}(\epsilon_i, r) = 0 \tag{A.1}$$

where $u_{n,l}(\epsilon_i, r)$ is the radial component of the wavefunction of state (n,l), n being the principal quantum number and l the azimuthal quantum number. Due to orthogonality constraints, $u_{n,l}$ presents an oscillatory behavior close to the radius, which is particularly difficult to describe when using plane waves as basis set. Therefore, the pseudized $u_{n,l}$ will be approximated by a smooth nodeless function of r in the core region and by the all electron $u_{n,l}$ outside this region. The core region is defined by a cutoff radius $r_{\text{cut}}$. Moreover, as an additional constraint, the pseudized eigenenergies should be equal to the all electron ones $\epsilon_i^{ps} = \epsilon_i^{ae}$. For both schemes precedently cited, the form of the pseudized wave function orginited from work by Kerker [234]: $r^{l+1}\exp[p(r)]$ if $r \leq r_{\text{cut}}$ with $p(r)$ a polynomial function of $r$ whose degree sets the number of conditions necessary to solve this non linear system of equations. In the case of Trouillier Martins scheme, the polynomial is of degree 6 in $r^2$, which yields 7 conditions to determine all the parameters of the function. Additional conditions can be added in order to get a complete defined set of equations:

- 5 of them are the continuity of $u_{n,l}$ at the $\text{r}_{\text{cut}}$ and the subsequent derivative.
- one of them is the conservation of the charge in the core region (guarantee the norm conservation).
- the last one being the the zero curvature of the pseudopotential at the origin (insure that your pseudopotential will be smooth).

Some of these conditions can be relaxed to get smoother pseudopotentials: the ultrasoft pseudopotential takes advantage of such idea by removing the norm-conservation constraint and correcting the resulting charge deficit by introducing atom-centered augmentation charges. The Hamann scheme uses a lower degree polynom $p(r)$ lowering the number of conditions to fulfill to three. Most notably, the last condition is removed making the Hammann scheme a little bit "harder", *i.e.*, leadiong to a deeper potential than the

Trouillier Martins ones. Besides, the matching of all electron and pseudized wave function is exponentially beyond $r_{\rm cut}$ while it is exact for TM. As a result, cut off radii used in Hamman scheme are usually smaller than the TM ones $r_{\rm cut}^{\rm H} \simeq 0.5 - 0.75 \times r_{\rm cut}^{\rm TM}$.

After generating the pseudized wavefunction, it is easy to see that inverting equation (A.1), we can express the pseudopotential as

$$v_{n,l}^{ps,scr} = \epsilon_l^{ps} - \frac{l(l+1)}{2r^2} + \frac{1}{2u_{n,l}^{ps}(r)} \frac{d^2}{dr^2} u_{n,l}^{ps}(r)\,. \tag{A.2}$$

One should notice that the potential is orbital dependent and that the nodeless characteristic of pseudized wavefunction is of primer importance when inverting the SE. Still, the pseudopotential should undergo a last manipulation before getting actually used in calculation. We are interested in the ionic component of the pseudopotential *ie* the interaction between the nucleus and the core electrons with the valence electrons. However the potential expressed in equation (A.2) contains the Hartree and the exchange and correlation potential, which are removed as follows

$$v_{n,l}^{ps}(r) = v_{n,l}^{ps,scr} - v^{\rm H}[\rho_0^{ps}, r] - v^{\rm xc}[\rho_0^{ps}, r]\,, \tag{A.3}$$

where $\rho_0^{ps}$ is the electronic density constructed with the pseudized wavefunction. The total ionic potential can be then rewritten in a compact way as

$$v_{\rm ion} = \sum_{l,n} |Y_{l,n}\rangle v_{n,l}^{\rm ps} \langle Y_{l,n}|\,. \tag{A.4}$$

Further simplication can be obtained by noticing that $v_{n,l}$ reduces at large $r$ to the ionic Coulomb potential $\mathrm{Z^{ion}/r}$ independent of n and l. Therefore the pseudopotential can be expressed as a local pseudopotential plus a few n,l-dependent short range potential.

$$v_{\rm ion} = v_{\rm loc}^{\rm ps} + \sum_{l,n} |Y_{l,n}\rangle \delta v_l^{\rm ps} \langle Y_{l,n}|\,, \tag{A.5}$$

where $\delta v_l^{\rm ps} = v_l^{\rm ps} - v_{\rm loc}^{\rm ps}$. While the first term is local in $r$-space, the second one is angularly semi-local while radially local. As a consequence, the expression of the latter term is the most time and computer consuming in a calculation. Assuming a $N$-dimensional basis set, the use of the semi-local form requires the evaluation and the storage of $\sim (N^2 + N)/2$ matrix elements. For very large system, evaluation of this matrix element becomes quickly prohibitive. Kleinman and Bylander proposed a scheme to transform the semi-local term into a fully separable non-local quantity [235].*ie*

$\delta v_l(r,r') = F_l(r) f_l F_l(r')$. The disconnected spacial variables $r$ and $r'$ permit to express $\delta v_l$ as a projector

$$\delta v_l^{\mathrm{KB}} = E_l^{\mathrm{KB}} |\xi_l\rangle\langle\xi_l| \,, \tag{A.6}$$

with the condition that $\delta v_l^{\mathrm{KB}} u_l^{\mathrm{ps}} = \delta v_l^{\mathrm{ps}} u_l^{\mathrm{ps}}$. Then the ernergy and the projector are expressed as

$$E_l^{\mathrm{KB}} = \frac{\langle u_l^{\mathrm{ps}} \delta v_l^{\mathrm{ps}} | \delta v_l^{\mathrm{ps}} u_l^{\mathrm{ps}} \rangle}{\langle u_l^{\mathrm{ps}} | \delta v_l^{\mathrm{ps}} | u_l^{\mathrm{ps}} \rangle} \,, \tag{A.7}$$

$$|\xi_l\rangle = \frac{|\delta v_l^{\mathrm{ps}} u_l^{\mathrm{ps}}\rangle}{(\langle u_l^{\mathrm{ps}} \delta v_l^{\mathrm{ps}} | \delta v_l^{\mathrm{ps}} u_l^{\mathrm{ps}} \rangle)^{1/2}} \,. \tag{A.8}$$

This procedure considerably reduces the computational cost of pseudopotential to simply N matrix elements. However, the major drawbacks of such method is the possible apparition of unphysical states called *ghost states*. While semilocal potentials are totally free of such problem, the fully non local potentials are not. The reason is that the latter will require an integrable differential Schrödinger equation while the former just a linear differentiable Schrödinger equation. The Wronskian theorem ensures that the eigenfunctions are energetically ordered according to their node structure. However this corollary does not hold for integrable differential equation and as a consequence *ghost states* can occur for fully non-local potentials.

## A.2 How to test pseudopotentials ?

In practice, pseudopotentials are generated considering the atomic system with a fixed electronic configuration. However, in solids, charge transfer and ionization occur and pseudopotentials should be able to handle such cases: it is the so called transferability property of pseudopotential. Careful check is required by performing pseudopotential and all electron calculations on atoms with different state occupations. Figure A.1 shows such tests in the case of the indium pseudopotential. While the promotion of one 5s electron to 5p gives rise the ionization of a the atom is much more tedious test

Another important concern is the use of Kleinman-Bylander approximation that might result in the appearance of ghost state. No systematic tests have been able to detect ghost state: however Gonze *et al.* introduced simple quantities like the KB cosine $\mathrm{C}^{\mathrm{KB}}$ defined as the cosine of the angle between $|u_l^{\mathrm{ps}}\rangle$ and $|\delta v_l^{\mathrm{ps}} u_l^{\mathrm{ps}}\rangle$, whose careful check allows one to detect the presence of a *ghost state* [236]. A more brute force method to detect ghost state is the

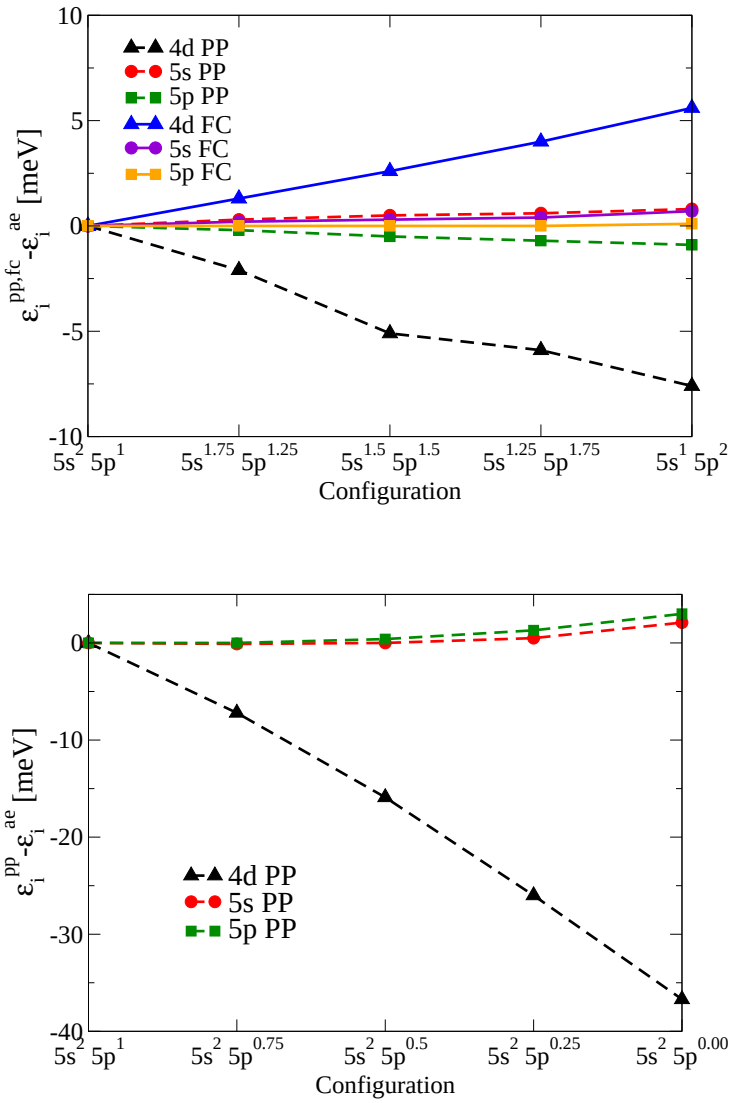


**Figure A.1:** Transferability test for the In pseudopotential. Top panel: promotion of one 5s electron to 5p states, PP refers to pseudopotential while FC refers to all electron frozen core approximation. Bottom panel: when ionizing the atom.

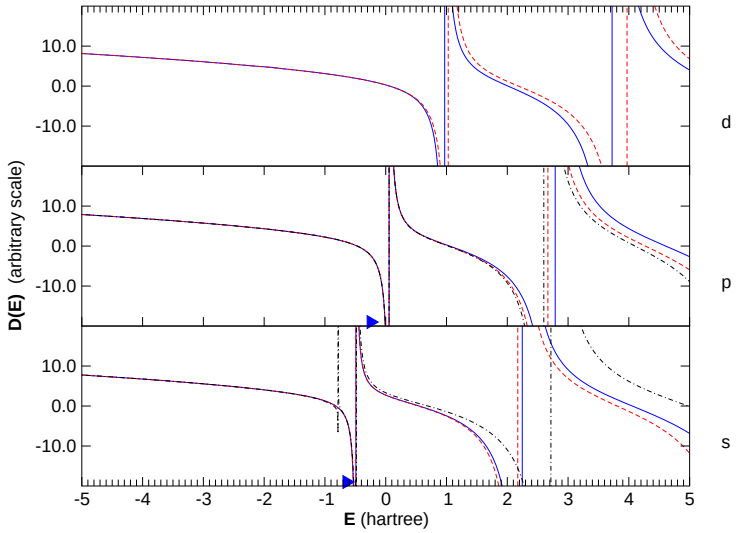


**Figure A.2:** Logarithmic derivative of the radial component of the different wavefunctions: all-electron wavefunction (blue curve), semi-local wavefunction (red curve) and fully non-local wavefunction (dashed black line) for Se.

inspection of the logarithmic derivative of the radial wavefunction $u_l(\epsilon, r)$ defined as

$$D(\epsilon, r^{\text{out}}) = \frac{d}{dr} \ln u_l(\epsilon, r) \Big|_{r=r^{\text{out}}}, \tag{A.9}$$

where $r^{\text{out}}$ is an arbitrary radius chosen outside the core region. One can immediately noticed that $\mathrm{D}(\epsilon, r^{\text{out}})$ will have poles for $\epsilon = \epsilon_l$. *Ghost states* will result in additional poles at lower energies. Besides, the logarithmic derivative should be correct to 1st order around the reference energy $\epsilon_l$ due to the norm-conserving scheme. One of the pathological cases of ghost states is the selenium pseudopotential. Let us consider a pseudopotential of Se generated using Hamann scheme with the following cutoff radius $r^s = 0.9$ bohr, $r^s = 1.06$ bohr and $r^d = 1.765$ bohr (the $d$ component refers to the 4d excited states). The channel $l = 2$ is taken as the local component. Figure A.2 shows the logarithmic derivative for the components $s$,$p$ and $d$. There is a ghost appearing just below the reference energy of the $s$ component. Several strategies can be undertaken in order to make the *ghost state* disappear: (i) change $r^s$ or (ii) change the local component from $l = 2$ to $l = 0$ (iii) switch from Hamann scheme to TM scheme. We choose the second option as it allows us to improve the scattering properties of the pseudopotential independently of the presence of ghost states. Figure A.3 shows the logartihmic derivative after modification of the pseudopotential: the ghost state disappears.

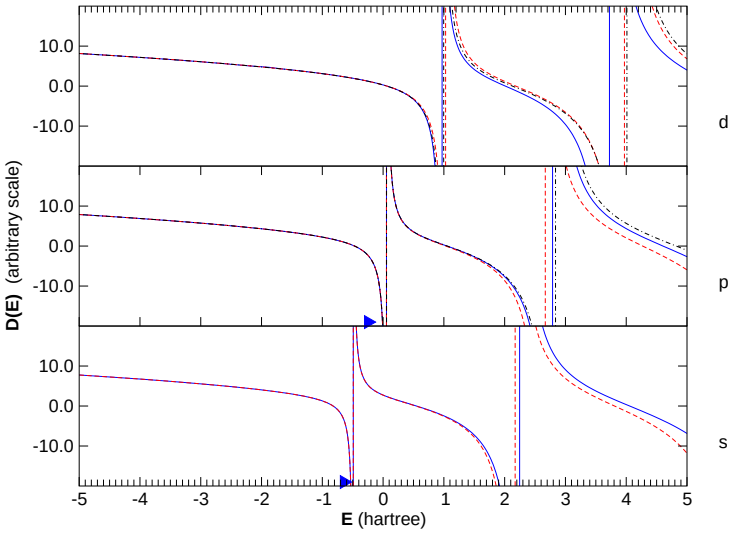


**Figure A.3:** Logarithmic derivative of the radial component of the different wavefunctions: all-electron wavefunction (blue curve), semi-local wavefunction (red curve) and fully non-local wavefunction (dashed black line) for Se.

Special attention has been drawn when building pseudopotentiel in order to get good scattering properties together with transferability for all of the pseudopotentials used in this thesis.

## A.3 Projected Augmented Wave (PAW)

Norm-conserving pseudopotentials encounter a certain success, however they require important number of basis functions to represent localized states. There is an example in this thesis with the use of a energy cutoff of 90 Hartree, yielding 35,000 plane waves to represent the wave functions of CIS. An efficient alternative to norm conserving PP is the projector augmented wave (PAW) method [237]. Being historically posterior to the UltraSoft pseudopotentials [238], it borrows one of the central ideas of the former: the use of atom-centered augmentation sphere. One can define a smooth part of the valence wave function $\tilde{\psi}_v$ and a linear transformation that relates the all electron valence wave functions to the smooth part $\psi_v = \mathcal{T}\tilde{\psi}_v$. The transformation $\mathcal{T}$ is supposed to be unity except within spheres centered on nuclei called augmentation spheres $\Omega_R$

$$\mathcal{T} = 1 + \sum_{\Omega_R} \tilde{\mathcal{T}}_R \,. \tag{A.10}$$

Let us consider we can expand the smooth wavefunction in partial waves inside each spheres

$$|\tilde{\psi}\rangle = \sum_i c_i |\tilde{\psi}_i\rangle \,, \tag{A.11}$$

then by acting with $\mathcal{T}$ on $|\tilde{\psi}\rangle$ we can construct the all electron wavefunction

$$|\psi\rangle = |\tilde{\psi}\rangle + \sum_i c_i \{|\psi_i\rangle - |\tilde{\psi}_i\rangle\} \,. \tag{A.12}$$

One can notice that the last term of equation (A.12) corrects the double counting of the smooth function in full space, *i.e* in both the augmentation spheres and the intertitial space between the spheres, and in the augmentation spheres. The general structure of equation (A.12) is recovered when expressing any operator of the original all-electron problem. For example, in the case of DFT calculation, one can express the electronic density as

$$n = \tilde{n} + \sum_{\Omega_R} n^R - \tilde{n}^R \,. \tag{A.13}$$

In addition to giving access to all-electron wavefunctions, the PAW permits to evaluate quantities at two different levels: using smooth functions in full space, or using localized functions in the augmentation spheres. One expects this particular feature of the PAW technic to decrease the computational load and to make the investigation of larger systems possible. Technical difficulties arise from the use of two different grids: the first one covers the full space and is used for the smooth functions while the second one is a radial grid in order to sample the augmentation sphere. In some cases (evaluation of Hartree potential for example), interpolation of functions should be realized from one grid to the other and further simplifications are used in order to keep the advantages of the scheme.

In this thesis, we have used the formalism of pseudopotentials except when dealing with large supercells with defects. The PAW implementation of *GW*, while certainly cumbersome [191, 239], allows one to treat larger systems, by reducing considerably the number of plane waves in order to represent the wave functions.

# Appendix B

# X-Ray Diffraction

One of the many ways to access the atomic structure of a solid is X-Ray Diffraction. The principle is very simple: the incoming X-ray electromagnetic radiation with momentum $\mathbf{k}_i$ is scattered by the electronic cloud surrounding the atoms and results in an outcoming radiation of the same frequency with a momentum $\mathbf{k}_f$. The intensity of the scattered radiation is the consequence of multiple interferences between the waves scattered by different atoms. This interference pattern can be then understood following Bragg's law expressed for lattice planes defined by their Miller indices *(h,k,l)*

$$2d_{hkl}sin(\theta) = n\lambda\,, \tag{B.1}$$

where $d_{hkl}$ is the distance between equivalent lattice planes in reciprocal space, $\theta$ the angle between the scattering plane and the incident wave, $n$ an integer and $\lambda$ the wavelengths of the X-ray. This relation gives the angular positions of the reflection peaks as seen in $\theta - 2\theta$ graph obtained from XRD experiments. Inversely, lattice parameter and tetragonal distortion can be extracted from the angular positions of the reflections. The intensity of the reflection is proportional to the square of the Fourier transform of the electronic density and depends only on the positions of the atoms inside the unit cell

$$I(hkl) \propto \left| \int \rho(\mathbf{r}) e^{-i(\mathbf{k}_f - \mathbf{k}_i).\mathbf{r}} \right|^2 . \tag{B.2}$$

Analysis of the extinction of Bragg peaks permits to estimate the internal structure parameter like anion displacement. One should notice that it can be strongly influenced by the presence of defects. An alternative to the calculation of the electronic density is the use of atomic scattering factor $f_\alpha$ ($\alpha$ =Cu,In,(S,Se)), which describes the density as a superposition of spherical

electronic density around each atoms.

$$I(hkl) \propto \left| \sum_{\alpha} f_{\alpha} e^{-i(\mathbf{k}_f - \mathbf{k}_i).\mathbf{r}_{\alpha}} \right|^2 . \tag{B.3}$$

Such a decomposition is very useful as it permits to assign the contribution of each type of atoms to the different Bragg peaks.

As stated before, the scattered waves formed an interference pattern made of constructive and destructive interferences. Therefore, not all the reciprocal lattice vectors will lead to a Bragg peak. For the chalcopyrite lattice structure, there are 3 categories of *allowed* reflections

- $(h, k, l/2)$ all even or odd yielding structure factor as

$$F_{h,k,l} = (f_{\mathrm{Cu}} + f_{\mathrm{In}} + f_{\mathrm{S,Se}}) \times g_1(h, k, l, u)\,, \tag{B.4}$$

  where $g$ is an arbitrary function. These reflections refer to the underlying zinc-blende structure of chalcopyrite.

- (h,k) even or (l/2) odd or vice and versa yielding structure factor as

$$F_{h,k,l} = f_{\mathrm{S,Se}} \times g_2(h, k, l, u)\,. \tag{B.5}$$

  The structure factor of these Bragg peaks depends only on the anion scattering factor and therefore on the anion displacement.

- (h) even and (k,l) odd or (k) even and (h,l) odd yielding structure factor as

$$F_{h,k,l} = \frac{1}{2}(f_{\mathrm{Cu}} - f_{\mathrm{In}}) g_3(h, k, l) + 2 f_{\mathrm{S,Se}} g_4(h, k, l, u)\,. \tag{B.6}$$

  The structure factor decomposition of this Bragg peaks have a cationic and anionic component making it sensitive to the anion displacement too.

Figure B.1 shows the $u$-dependence of the intensity of these peaks calculated using atomic scattering factors. Similar results were found by Jaffe *et al.*[4] using equation B.2. A strong dependence on the anion displacement appears for category (ii) and (iii) but their relative intensities compared to the (112) peak $I_{hkl}/I_{112}$ never reach the detection limit of XRD ($\approx 1\%$). Besides, the presence of defects might alter significantly the analysis. Therefore, other technics should be considered in order to obtain the value of the anion displacement: for example Extended X-Ray Absorption Fine Structure [240]. It is based on X-Ray absorption. Contrary to X-Ray diffraction, X-Ray absorption experiments consist of shooting photons in the x-ray range of energy.

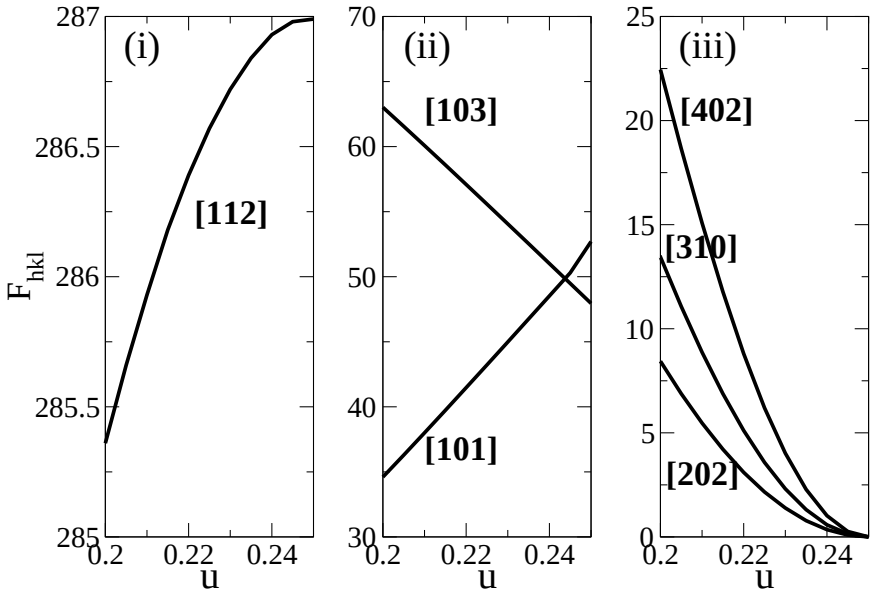


**Figure B.1:** Structure factor of some allowed reflections versus the anion displacement $u$. Left panel shows category (i), middle panel category (ii) and right panel category (iii)

The photon will be eventually absorbed by a core electron, which results in the creation of a core hole and an unbound electron. At the same time, one can think of this electron has a quantum wave that produces interferences with other electron quantum waves originating from different atoms. This interferences may be either constructive or destructive and result in wriggles in the X-ray spectra. The interpretation of such spectra can be studied and reveals interesting features of the atomic structure or the electronic and vibrational properties of the materials. However, the analysis of EXAFS is an exhausting task as a single electron scattering theory is not sufficient to explain the spectra. Multiple scattering effects must be included, which makes the interpretation of the spectra excessively difficult: thereby, structural changes together with the effects of defects may mix up and assignment to one feature of the spectra of the specific effect is very difficult.

Figure 5.4 depicts the distribution of values of $u$ for $CuInSe_2$ obtained experimentally with different methods ranging from x-ray diffraction to EXAFS. One can see a rather broad dispersion of $u$. It could be due to the samples: different growth technics, different morphologies or different doping properties. Merino *et al.* [203] proposed a possible explanation to such

experimental dispersion: low values of $u$ $u \approx 0.222$ seems to originate from polycrystalline samples while single crystal samples show larger values of $u$ ($u \approx 0.235$). Despite the multiplicity of the causes for such effect, the facts are there: the value of $u$ is extremely sample dependent and it is therefore difficult to attribute a precise value to $u$. The same holds for $CuInS_2$ but unfortunately, the lack of experimental data [241, 242, 243] does not permit us to visualize any distribution of $u$. Let us define the range of $u$ for $CuInS_2$ being the difference between $u^{\min}$ and $u^{\max}$ the two extremal experimental value. Thus defined, the experimental range of $CuInS_2$ is comparable to the one of $CuInSe_2$. Moreover, the shaded regions depicted in graphics representing $E_g$ versus $u$ is defined such the non-physical value of $u$ *i.e.* $u$ lies outside both the experimental range and the range predicted by means of *ab initio* methods.

# Appendix C

# *GW* study of delafossite CuAlO2

The knowledge obtained from the work on CIS has allowed us also to work on related calsses of materials. Therefore, I have proposed and carried out, in collaboration with Fabio Trani and Miguel Marques from Laboratoire de physique de la matiére condensée et nanostructures, University of Lyon, a study on the transparent conducting oxide $CuAlO_2$. In this appendix, we present the publication which results from this work:

*Effects of Electronic and Lattice Polarization on the Band Structure of Delafossite Transparent Conductive Oxides*, J.Vidal, J., F. Trani, F. Bruneval, M. A. L. Marques, and S. Botti, Physical Review Letters, **104**,136401,(2010).

We use hybrid functionals and restricted self-consistent *GW*, state-of-the-art theoretical approaches for quasiparticle band structures, to study the electronic states of delafossite $Cu(Al,In)O_2$, the first p-type and bipolar transparent conductive oxides. We show that self-consistent *GW* gives remarkably wider band gaps than all the other approaches used so far. Accounting for polaronic effects in the *GW* scheme we recovered a very nice agreement with experiments. Furthermore, the modifications with respect to the Kohn-Sham bands are strongly $\boldsymbol{k}$-dependent, which makes questionable the common practice of using a scissor operator. Finally, our results support the view that the low energy structures found in optical experiments, and initially attributed to an indirect transition, are due to intrinsic defects in delafossite samples.

Many high-technology devices, such as flat panel displays, touch screens, or even thin-film solar cells, require the use of thin transparent contacts. These contacts are usually built from insulating oxides that, for a certain range of doping, become conductive while retaining transparency in the visible spectrum. The most common examples of these so-called transparent

conductive oxides (TCOs) are electron (n-)doped $SnO_2$, $In_2O_3$, and ZnO. Hole (p-)doping of wide gap semiconductors was for long time very hard to obtain [244, 245]. It is therefore not surprising that the discovery of p-doping in $CuAlO_2$ thin films with a carrier mobility of about $10\,cm^2/(V\,s)$ attracted great interest [246, 247]. Other members of the delafossite family, like $CuGaO_2$ [248] and $CuInO_2$ [249], were discovered shortly after. The latter compound is particularly interesting as it exhibits bipolar (n- *and* p-type) conductivity by doping with appropriate impurities and tuning the film-deposition conditions [249]. This opens the way to the development of transparent p-n junctions, and therefore fully transparent optoelectronic devices, functional windows and stacked solar cells with improved efficiency.

$CuAlO_2$ is by far the most studied system of the family of delafossite TCOs, both theoretically and experimentally. However, there is still no agreement neither on the origin of the p-type conductivity, nor on the electronic bands of the pure crystal. Measurements of the direct *optical* band gap ($E_g^{dir}$) of $CuAlO_2$ fall in the range from 2.9 to 3.9 eV [246, 247, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260], with most values in the interval 3.4–3.7 eV. These experiments also yield a large dispersion of indirect gaps ($E_g^{ind}$), from 1.65 to 2.1 eV, with one experiment measuring 2.99 eV [260]. Unfortunately, there is only one photoemission experiment [250] that gives a value of 3.5 eV for the *quasiparticle* band gap. Note that the *optical* and *quasiparticle* gaps differ by the exciton binding energy. Concerning $CuInO_2$, optical experiments measured $E_g^{dir}$ between 3.9 and 4.45 eV [249, 261, 262], with only one estimation of $E_g^{ind}$ at 1.44 eV [262].

From the theoretical perspective, the situation is also quite complex, even if the full Cu 3*d* shell should exclude the strongly correlated electron regime. These materials are usually studied within density functional theory (DFT), using the standard local density (LDA) or generalized gradient approximations (GGA). However, it is well known that the Kohn-Sham band structures systematically underestimate the band gaps. For similar compounds, like $Cu_2O$ and $CuIn(S,Se)_2$, Kohn-Sham LDA calculations lead to unreasonable band structures, in particular due to the misrepresentation of the hybridization between the *d* electrons of the metal and *p* electrons of the anion [181, 263]. To overcome this situation, hybrid functionals have been recently proposed, with very promising results [264], especially for materials with small and intermediate band gaps [265, 266]. Other approaches include LDA+$U$, that tries to improve the description of electronic correlations through the introduction of a mean-field Hubbard-like term. This method has been quite successful in the study of the electronic structure of strongly correlated systems, but it relies on a parameter $U$, that is often adjusted to experiments.

Arguably the most reliable and used *ab initio* technique to obtain quasiparticle band structures is the many-body $GW$ approach [152]. The common practice within this framework is to start from a DFT calculation, and evaluate perturbatively the $GW$ energy corrections to the band structure. This procedure, which we will refer to as $G_0W_0$, is justified when the departure wave functions and band structure are already close to the quasiparticle ones. This is indeed the case in many systems, explaining why $G_0W_0$ has been extremely successful in describing electron addition and removal energies for metals, semiconductors and insulators [267]. However, it has been recently shown that $G_0W_0$ fails for many transition metal oxides [181, 180].

To solve this problem one can perform restricted self-consistent (sc) $GW$ [178, 192]. This technique has the advantage of being independent of the starting point at the price of large computational complexity. Fortunately, there is an alternative procedure that yields wavefunctions that are extremely close to those obtained in a full sc-$GW$ calculation, namely sc-COHSEX as explained in Ref. [155]. The dynamical effects that are absent in COHSEX calculations can then be accounted for by performing a final perturbative $GW$ step. This method, that we will refer to as sc-$GW$, has been applied to many oxide compounds, yielding excellent results for the band gaps and the quasiparticle band structure [181, 155, 180, 263].

Note that these theoretical techniques yield quasiparticle bands, and not optical gaps. To evaluate these latter quantities one mostly resorts to the solution of the Bethe-Salpeter equation. For the delafossite structures there is one such calculation starting from a GGA+$U$ band structure [184]. It yields for $CuAlO_2$ a very large exciton binding energy of about 0.5 eV for the first direct transition. The choice of the parameter $U$ was found to have significant consequences on the width of the band gap, but it did not affect significantly the exciton binding energy. We can thus assume that 0.5 eV is a reasonable estimate of this latter quantity.

In the following, we present calculations of the band structures of $CuAlO_2$ and $CuInO_2$ using some of the most accurate theoretical tools available in the community. These include the standard LDA, hybrid functionals (namely B3LYP and two flavors of HSE), LDA+$U$, $G_0W_0$ and sc-$GW$. As discussed above, we expect sc-$GW$ to be the most accurate *ab initio* approach. When the comparison was possible, we found our results in excellent agreement with previous calculations (Refs. [250, 259, 260, 268, 269, 270, 271] for LDA, Ref. [271] for B3LYP, and Ref. [272] for GGA+$U$).

The hybrid and LDA+$U$ calculations were performed with VASP [273] and ABINIT [274] respectively, using the PAW formalism and an energy cutoff of 44 Ha. The parameter $U$ was set to 8 eV as in Ref. [184]. Our $GW$ calculations were performed with ABINIT, starting from LDA band

structures. The energy cutoff was 120 Ha for the ground state calculation, and the **k**-point grid was a $4 \times 4 \times 4$ Monkhorst-Pack. Note that it was absolutely essential to use the method of Ref. [5], due to the extremely slow convergence with respect to the number of conduction states.

In Figs. C.2 and C.1 we show direct and indirect photoemission gaps and the band structures of $CuAlO_2$, obtained using different theoretical approaches. The minimum $E_g^{dir}$ of $CuAlO_2$ is always found at L, where the dipole transition between the band edge states is allowed [270]. All calculations, except sc-*GW*, give a fundamental $E_g^{ind}$ between the conduction band minimum at Γ and the valence band maximum along the Γ-F line. The experimental data for optical gaps are also presented with an error bar that reflects the dispersion of the most likely values found in literature. LDA exhibits, as expected, the smallest gaps. Basically every approach beyond it opens up the gap by different amounts and modifies the band dispersions. The direct and indirect gaps have similar behaviors in the different theories, and both increase when going from LDA<$G_0W_0$<HSE03<HSE06<B3LYP<sc-*GW*. On the other hand, the difference $E_g^{dir} - E_g^{ind}$ seems to decrease with the sophistication of the method, reaching nearly zero for the sc-*GW* calculation. This is a consequence of the drastic change of the conduction band dispersion, which displaces the conduction minimum from Γ to L when sc-*GW* is applied (see Fig. C.1). Only LDA+*U* does not follow the trend, as it is the only case in which $E_g^{dir} - E_g^{ind}$ gets significantly larger than in LDA.

Looking at the direct gap, we point out that most of the methods give results that are within the experimental range, when an exciton binding energy of around 0.5 eV [184] is considered. This is true for LDA+*U*, $G_0W_0$, the hybrids HSE03 and HSE06. However, for sc-*GW* and even for B3LYP, the theoretical gap is larger by about 1–1.5 eV than the experimental findings. For $CuInO_2$ (see Fig. C.3) we have to make the comparison with care, as the smallest $E_g^{dir}$ is located at Γ, where optical transitions are forbidden [270]. A meaningful comparison with experiments must consider the gap at L. Thus, we find that both trends and quantitative results are analogous to those for $CuAlO_2$. In particular, sc-*GW* yields again $E_g^{dir}$ larger by 1–1.5 eV than the experimental range.

We stress again that, to date, sc-*GW* is arguably the best method available to estimate band gaps of wide-gap semiconductors, and that it gives excellent results for compounds like $Cu_2O$ and $CuIn(S,Se)_2$ [181, 263]. It is unlikely that the presence of defects can lead to such a large shrinkage of $E_g^{dir}$. However, there is another effect that has been neglected up to now: the change of screening due to the polarization of the lattice. In fact, according to the experimental data [275], unfortunately available only for $CuAlO_2$,

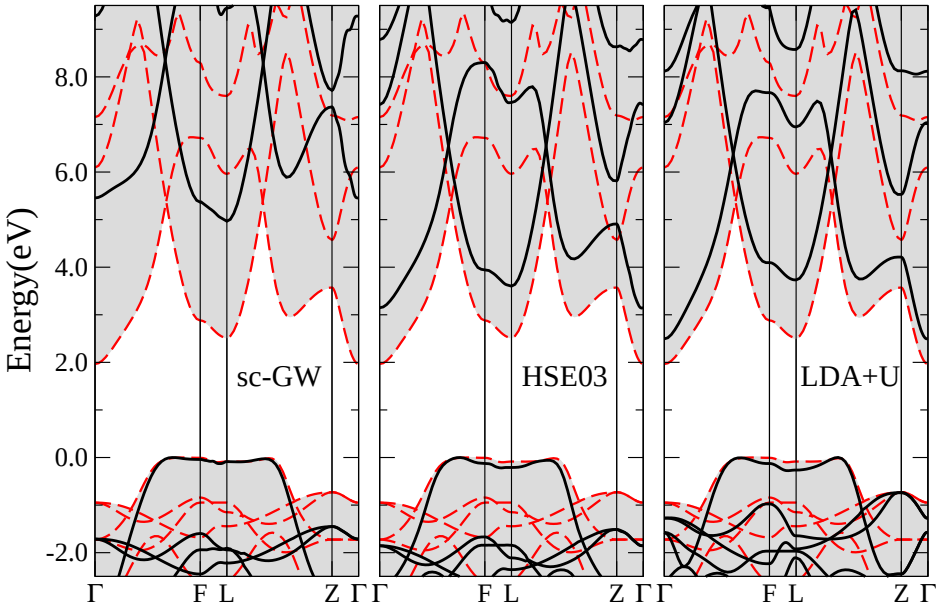


**Figure C.1:** (Color online) Band structures for $CuAlO_2$: comparison of LDA (red dashed lines) with sc-*GW* (left panel), HSE03 (central panel), and LDA+*U* (right panel).

the polaron constant [1] for this system is large ($\alpha_p \sim 1$), indicating a non-negligible contribution of the lattice polarization to the electronic screening. It is known that in other ionic compounds with similar polaron constants this can lead to a shrinkage to the band gap by about 1 eV [276]. A full sc-*GW* calculation including in an *ab initio* framework the effects of the lattice polarization is to date beyond reach. However, a reliable estimate can be obtained using the model proposed by Bechstedt *et al.* [276], which gives a static representation of the polaronic effects based on difference of experimental static dielectric constants. By performing a perturbative *GW* step including model polaronic effects on top of the sc-COHSEX, we found a uniform (**k**-independent) shrinkage of the band gap by 1.2 eV. As we can see in Fig. C.2, this correction brings our results for $E_g^{dir}$ well within the experimental range (once the excitonic correction of about 0.5 eV is also considered). As it is observed in Ref. [276], the polaronic model employed can only overestimate the correction. All these results point to the conclusion that the agreement of the other methods with experiment was fortuitous and due to a cancellation of errors.

[1] The polaron constant is defined as $\alpha_P = (1/\epsilon_\infty - 1/\epsilon_0)(\hbar/2ma_B^2\omega_{LO})^{1/2}$, where $a_B$ is the Bohr radius, $\omega_{LO}$ the longitudinal optical frequency of the highest $E_u$ phonon mode, $\epsilon_\infty$ and $\epsilon_0$ the low frequency electronic and the static dielectric constants, respectively.

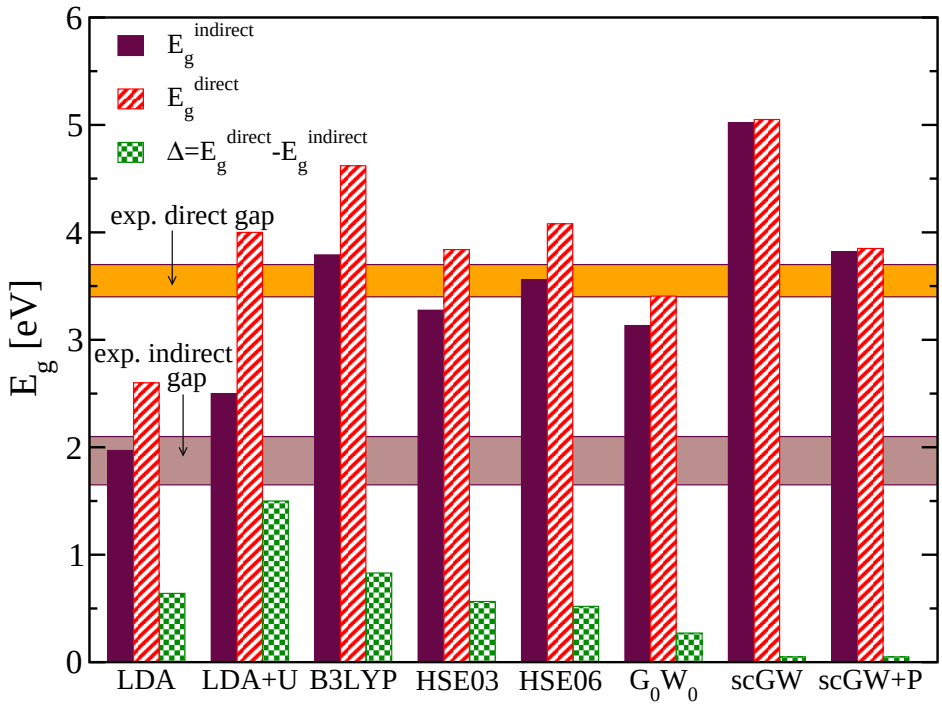


**Figure C.2:** (Color online) Band gaps of $CuAlO_2$ using: LDA, LDA+$U$, hybrid, $G_0W_0$, sc-$GW$, and sc-$GW$ including model polaronic corrections. The horizontal zones contain data extracted from optical experiments (see text).

Looking now at the indirect gap, we focus on Fig. C.2 as there are more experimental data for $CuAlO_2$. All the hybrids and $GW$ calculations yield indirect gaps much larger than the experimental range 1.65-2.1 eV, even taking into account any possible excitonic and polaronic effects. Moreover, sc-$GW$, the best method used in this work, yields the highest $E_g^{ind}$ at around 5 eV, while the difference $E_g^{dir} - E_g^{ind}$ is in general much smaller than the experimental value ($\approx$2 eV), and even vanishing for the sc-$GW$ calculation. From Fig. C.3 we realize that these conclusions are as well valid for $CuInO_2$, where the best estimates for the indirect band gap is much larger than the experimental value of 1.44 eV [262].

These are very strong arguments in favor of Robertson *et al.* [271] that suggested that the experimental "indirect gap" absorption was due to defects, and should not be present in the defect-free compound. Also Pellicer- Porres *et al.* [260] questioned the interpretation of the low energy peaks as indirect transitions, as the absorption coefficient is more than two orders of magnitude larger than in typical indirect absorption edges. The most promising defects that could be responsible are oxygen interstitials $O_i$, as DFT calculations

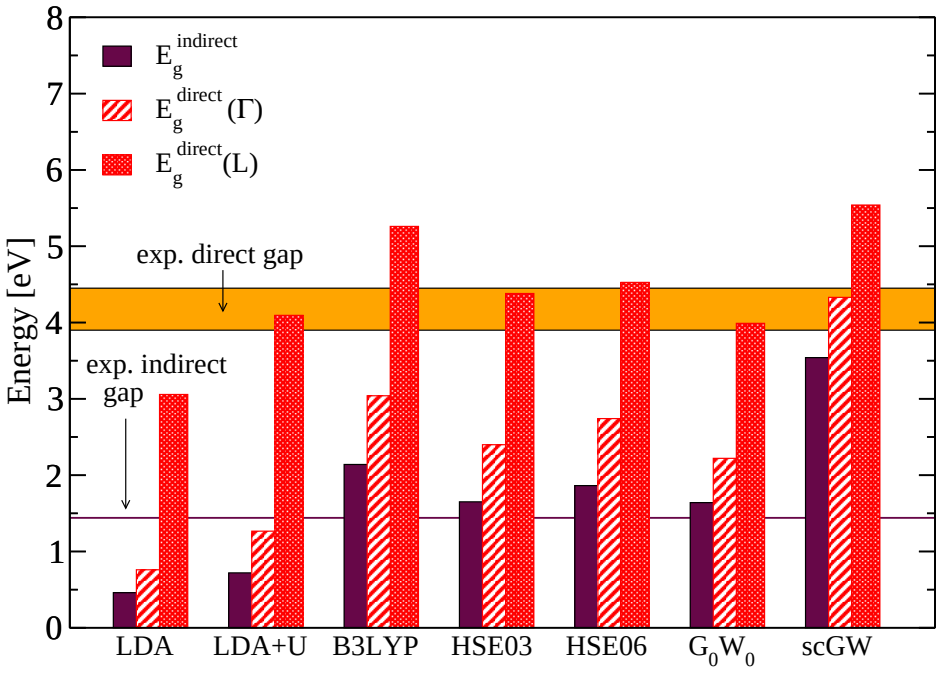


**Figure C.3:** (Color online) Band gaps of $CuInO_2$ using: LDA, LDA+$U$, hybrid, $G_0W_0$, sc-$GW$. The horizontal zones contain data extracted from optical experiments (see text).

within the LDA predict low formation energies and the introduction of states in the gap at 0.7 and 1.4 eV [277]. However, a full clarification of this issue will require sc-$GW$ or hybrid calculations for these, and other more complex defects.

Finally, we analyze more in detail the band structures of $CuAlO_2$ shown in Fig. C.1. LDA calculations (red dashed lines) are compared with sc-$GW$, HSE03, and LDA+$U$ calculations. The main effect of LDA+$U$ is to open the LDA gap by an amount that can be controlled by the parameter $U$. The difference $E_g^{dir} - E_g^{ind}$ is in this approximation enhanced, due to a change of the character of the lowest conduction band along the symmetry lines. Hybrid calculations using HSE03 give a comparable $E_g^{dir}$ and a modified dispersion of both valence and conduction states close to the Fermi energy, which reduces $E_g^{dir} - E_g^{ind}$. The conduction band minimum (CBM) within HSE03 is still located at Γ, but the difference between the CBM at L and Γ gets significantly smaller. For sc-$GW$, besides the further increase of the band gaps, the dispersion of the bands is strongly affected by the many-body effects. In fact, the $GW$ corrections exhibit an unusual dispersion of around 1 eV when looking at the different **k**-points, displacing the CBM from

Γ to L. We note that often in semiconductor physics one assumes that the quasiparticle corrections can be modeled by a rigid shift (the so-called scissor operator). From our results it follows that one should refrain from using this simple approximation for these important materials. We can also conclude that hybrid calculations give a better description of band dispersions than LDA+$U$, even if the two approaches yield similar band gaps.

In conclusion, it is clear that the delafossite family exhibits complex and unusual band gap physics that can not be captured by standard theoretical approximations. We found that the direct band gap is well reproduced by the best many-body approaches if polaronic effects are taken into account. We can expect that this situation, of a large gap that is reduced substantially by polaronic effects, is quite general and is present in many more materials that previously expected. In fact, the apparent good agreement between calculated gaps (with hybrid functionals or $G_0W_0$) and experimental gaps for materials as simple and widely studied as LiF can be accidental: the underestimation of the gap by these methods is compensated by the neglect of polaronic effects. Furthermore, the modifications with respect to the LDA Kohn-Sham bands are strongly $\boldsymbol{k}$-dependent, which makes questionable the common practice of using a scissor operator. The band dispersion obtained by hybrid functional calculations is in between the LDA and sc-$GW$ dispersion, while the LDA+$U$ calculations open up the gap but do not give a significant improvement of the band dispersion. Finally, our calculations rule out the interpretation of the low energy features in the absorption spectra as arising from a putative indirect band gap. These structures should rather come from intrinsic defects, as proposed in Refs. [260, 271]. However, a complete understanding of the electronic and excitation properties of these systems will only be achieved, in our opinion, by a high-level theoretical scheme (like sc-$GW$) including both defects and effects from the lattice polarization in an ab initio framework. Work along these lines is already in progress.

# Bibliography


[1] J. E. Jaffe and A. Zunger, Phys. Rev. B **27**, 5176 (1983).

[2] S. Zhang, S. Wei, A. Zunger, and H. Katayama-Yoshida, Phys. Rev. B **57**, 9642 (1998).

[3] W. Shockley and H. J. Queisser, J. Appl. Phys. **32**, 510 (1961).

[4] J. E. Jaffe and A. Zunger, Phys. Rev. B **28**, 5822 (1983).

[5] F. Bruneval and X. Gonze, Phys. Rev. B **78**, 85125 (2008).

[6] A. Einstein, Ann. Phys. Chem **17**, 132 (1905).

[7] D. Chapin, C. Fuller, and G. Pearson, J. Appl. Phys. **25**, 676 (1954).

[8] G. Pearson, American Journal of Physics **25**, 591 (1957).

[9] U. Rau and H. Schock, Applied Physics A: Materials Science & Processing **69**, 131 (1999).

[10] D. Staebler and C. Wronski, Appl. Phys. Lett. **31**, 292 (1977).

[11] R. Birkmire, (2008).

[12] C. Kaufmann, A. Neisser, R. Klenk, and R. Scheer, Thin Solid Films **480-481**, 515 (2005).

[13] U. Rau and H. Schock, Appl. Phys. A **69**, 131 (1999).

[14] A. Ennaoui, U. Blieske, and M. Ch Lux-Steiner, Progress in Photovoltaics Research and Applications **6**, 447 (1998).

[15] N. Naghavi, S. Spiering, M. Powalla, B. Canava, A. Taisne, J. Guillemoles, S. Taunier, A. Etcheberry, and D. Lincot, *MATERIALS RESEARCH SOCIETY SYMPOSIUM PROCEEDINGS* (PUBLISHER, ADDRESS, 2003), Vol. 763, pp. 465–470.

[16] S. Siebentritt, Solar Energy **77**, 767 (2004).

[17] W. Shockley and W. T. Read, Phys. Rev. **87**, 835 (1952).

[18] M. Contreras, K. Ramanathan, J. AbuShama, F. Hasoon, D. Young, B. Egaas, and R. Noufi, Progress in Photovoltaics Research and Applications **13**, 209 (2005).

[19] U. Rau, Appl. Phys. Lett. **74**, 111 (1999).

[20] M. A. Green, Progress in Photovoltaics: Research and Applications **4**, 375 (1996).

[21] C. Sah, R. Noyce, and W. Shockley, Proceedings of the IRE **45**, 1228 (1957).

[22] M. Gloeckler, J. R. Sites, and W. K. Metzger, J. Appl. Phys. **98**, 113704 (2005).

[23] P. Landsberg and G. Tonge, J. Appl. Phys. **51**, R1 (1980).

[24] J. Zhao, A. Wang, and M. Green, Prog. Photovolt: Res. Appl **7**, 144 (1999).

[25] M. Green, *Third generation photovoltaics: advanced solar energy conversion* (Springer Verlag, ADDRESS, 2003).

[26] M. A. Green, K. Emery, Y. Hishikawa, and W. Warta, Progress in Photovoltaics: Research and Applications **17**, 320 (2009).

[27] D. Su, W. Neumann, R. Hunger, P. Schubert-Bischoff, M. Giersig, H. Lewerenz, R. Scheer, and E. Zeitler, Appl. Phys. Lett. **73**, 785 (1998).

[28] M. H. Bode, J. Appl. Phys. **76**, 159 (1994).

[29] R. Herberholz, U. Rau, H. Schock, T. Haalboom, T. Gödecke, F. Ernst, C. Beilharz, K. Benz, and D. Cahen, Eur. Phys. J. AP **6**, 131 .

[30] A. Klein and W. Jaegermann, Appl. Phys. Lett. **74**, 2283 (1999).

[31] L. Stolt, J. Hedström, J. Kessler, M. Ruckh, K. Velthaus, and H. Schock, Appl. Phys. Lett. **62**, 597 (1993).

[32] J. Guillemoles, T. Haalboom, T. Gödecke, F. Ernst, and D. Cahen, *MATERIALS RESEARCH SOCIETY SYMPOSIUM PROCEEDINGS* (PUBLISHER, ADDRESS, 1998), Vol. 485, pp. 127–132.

[33] M. Contreras, H. Wiesner, J. Tuttle, K. Ramanathan, and R. Noufi, Solar Energy Materials and Solar Cells **49**, 239 (1997).

[34] S. Lany and A. Zunger, Phys. Rev. B **72**, 035215 (2005).

[35] S. Lany and A. Zunger, J. Appl. Phys. **100**, 113725 (2006).

[36] M. Igalson and R. Bacewicz, Cryst. Res. Technol. **31**, 445 (1996).

[37] M. Igalson and H. Schock, J. Appl. Phys. **80**, 5765 (1996).

[38] D. Braunger, S. Zweigart, and H. Schock, *Proc. 2nd. World Conf. Photovoltaic Solar Energy Conversion* (PUBLISHER, ADDRESS, 1998), pp. 1113–1116.

[39] D. Niles, M. Al-Jassim, and K. Ramanathan, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films **17**, 291 (1999).

[40] D. Niles, K. Ramanathan, F. Hasoon, R. Noufi, B. Tielsch, and J. Fulghum, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films **15**, 3044 (1997).

[41] M. Contreras, B. Egaas, P. Dippo, J. Webb, J. Granata, K. Ramanathan, S. Asher, A. Swartzlander, and R. Noufi, *CONFERENCE RECORD IEEE PHOTOVOLTAIC SPECIALISTS CONFERENCE* (PUBLISHER, ADDRESS, 1997), Vol. 26, pp. 359–362.

[42] U. Rau, M. Schmitt, F. Engelhardt, O. Seifert, J. Parisi, W. Riedl, J. Rimmasch, and F. Karg, Solid state communications **107**, 59 (1998).

[43] S. Wei, S. Zhang, and A. Zunger, Appl. Phys. Lett. **72**, 3199 (1998).

[44] G. Hanna, A. Jasenek, U. Rau, and H. Schock, Thin Solid Films **387**, 71 (2001).

[45] I. Repins, M. A. Contreras, E. Brian, *et al.*, Prog. Photovolt. **16**, 235 (2008).

[46] S. Hegedus, Prog. Photovolt: Res. Appl **14**, 393 (2006).

[47] *Proceedings of the 29th IEEE Photovoltaic Spectroscopy Conference* (IEEE, Piscataway, 2002).

[48] *Proceedings of the 11th International Photovoltaic Conference.* (North-Holland, Amsterdam, 2001).

[49] C. H. Champness, in *Proceedings of the 29th IEEE Photovoltaic Spectroscopy Conference* (IEEE, Piscataway, 2002), p. 732.

[50] L. S. Yip and I. Shih, in *Proceedings of the 1st World Conference on Photovoltaic Energy Conversion* (IEEE, Piscataway, 1994), p. 210.

[51] U. Rau, K. Taretto, and S. Siebentritt, Applied Physics A: Materials Science & Processing **96**, 221 (2009).

[52] I. Visoly-Fisher, S. Cohen, K. Gartsman, A. Ruzin, and D. Cahen, Advanced Functional Materials **16**, 649 (2006).

[53] J. Ziman, *Principles of the Theory of Solids* (Cambridge Univ Pr, ADDRESS, 1979).

[54] R. Martin, *Electronic structure: basic theory and practical methods* (Cambridge Univ Pr, ADDRESS, 2004).

[55] M. Born and R. Oppenheimer, Ann. Phys **84**, 457 (1927).

[56] R. Feynman, Phys. Rev. **56**, 340 (1939).

[57] L. Thomas, *Proceedings of the Cambridge Philosophical Society* (PUBLISHER, ADDRESS, 1927), Vol. 23, p. 542.

[58] E. Fermi, Rend. Accad. Naz. Lincei **6**, 602 (1927).

[59] E. Teller, Rev. Mod. Phys. **34**, 627 (1962).

[60] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[61] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[62] M. Seah and W. Dench, Surf. Interface Anal **1**, (1979).

[63] T. Koopmans, Physica **1**, 104 (1934).

[64] G. Onida *et al.*, Rev. Mod. Phys. **74**, 601 (2002), and references therein.

[65] M. Levy, J. Perdew, and V. Sahni, Phys. Rev. A **30**, 2745 (1984).

[66] L. J. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983).

[67] J. P. Perdew and M. Levy, Phys. Rev. Lett. **51**, 1884 (1983).

[68] D. Ceperley and B. Alder, Phys. Rev. Lett. **45**, 566 (1980).

[69] R. Hood, M. Chou, A. Williamson, G. Rajagopal, R. Needs, and W. Foulkes, Phys. Rev. Lett. **78**, 3350 (1997).

[70] O. Gunnarsson and B. I. Lundqvist, Phys. Rev. B **13**, 4274 (1976).

[71] M. SHANG-kENG and K. Brueckner, Phys Rev **165**, 18 (1968).

[72] J. Perdew and S. Kurth, *A Primer in Density Functional Theory* (PUBLISHER, ADDRESS, 2003), Vol. 620, pp. 1–55.

[73] F. Herman, J. Van Dyke, and I. Ortenburger, Phys. Rev. Lett. **22**, 807 (1969).

[74] J. P. Perdew and W. Yue, Phys. Rev. B **33**, 8800 (1986).

[75] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

[76] J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).

[77] M. Levy, Phys. Rev. A **43**, 4637 (1991).

[78] P. Bagno, O. Jepsen, and O. Gunnarsson, Phys. Rev. B **40**, 1997 (1989).

[79] D. J. Singh, W. E. Pickett, and H. Krakauer, Phys. Rev. B **43**, 11628 (1991).

[80] J. Zhu, X. W. Wang, and S. G. Louie, Phys. Rev. B **45**, 8887 (1992).

[81] R. Schmid, E. Engel, R. Dreizler, P. Blaha, and K. Schwarz, Advanced in Quantum Chemistry 209 (1998).

[82] S. Varga, B. Fricke, H. Nakamatsu, T. Mukoyama, J. Anton, D. Geschke, A. Heitmann, E. Engel, and T. Baştug, J. Chem. Phys. **112**, 3499 (2000).

[83] W. Liu and C. van Wüllen, J. Chem. Phys. **113**, 2506 (2000).

[84] E. Fermi and E. Amaldi, Accad. Ital. Rome **6**, 117 (1934).

[85] J. Perdew, Chemical Physics Letters **64**, 127 (1979).

[86] J. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[87] R. A. Heaton, J. G. Harrison, and C. C. Lin, Phys. Rev. B **28**, 5992 (1983).

[88] N. Hamada and S. Ohnishi, Phys. Rev. B **34**, 9042 (1986).

[89] S. Kurth and J. P. Perdew, Phys. Rev. B **59**, 10461 (1999).

[90] J. Hubbard, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **276**, 238 (1963).

[91] V. I. Anisimov, I. V. Solovyev, M. A. Korotin, M. T. Czyżyk, and G. A. Sawatzky, Phys. Rev. B **48**, 16929 (1993).

[92] D.-K. Seo, Phys. Rev. B **76**, 033102 (2007).

[93] V. Anisimov, F. Aryasetiawan, and A. Lichtenstein, Journal of Physics: Condensed Matter **9**, 767 (1997).

[94] A. I. Liechtenstein, V. I. Anisimov, and J. Zaanen, Phys. Rev. B **52**, R5467 (1995).

[95] C. Herring, *Magnetism* (Academic Press New York, ADDRESS, 1966).

[96] V. I. Anisimov and O. Gunnarsson, Phys. Rev. B **43**, 7570 (1991).

[97] P. H. Dederichs, S. Blügel, R. Zeller, and H. Akai, Phys. Rev. Lett. **53**, 2512 (1984).

[98] M. S. Hybertsen, M. Schlüter, and N. E. Christensen, Phys. Rev. B **39**, 9028 (1989).

[99] A. K. McMahan, R. M. Martin, and S. Satpathy, Phys. Rev. B **38**, 6650 (1988).

[100] M. Cococcioni and S. de Gironcoli, Phys. Rev. B **71**, 035105 (2005).

[101] A. Baldereschi, Phys. Rev. B **7**, 5212 (1973).

[102] D. Chadi and M. Cohen, Phys. Rev. B **8**, 5747 (1973).

[103] H. Monkhorst and J. Pack, Phys. Rev. B **13**, 5188 (1976).

[104] P. Pulay, Molecular Physics: An International Journal at the Interface Between Chemistry and Physics **17**, (1969).

[105] F. Manghi, E. Molinari, C. Bertoni, and C. Calandra, Journal of Physics C: Solid State Physics **15**, 1099 (1982).

[106] D. G. Anderson, J. ACM **12**, 547 (1965).

[107] C. G. Broyden, Math. Comp. **19**, 577 (1965).

[108] P. Pulay, Chemical Physics Letters **73**, 393 (1980).

[109] W. Pickett, Comp. Phys. Rep **9**, 15 (1989).

[110] J. Harris and R. O. Jones, Journal of Physics F: Metal Physics **4**, 1170 (1974).

[111] D. C. Langreth and J. P. Perdew, Phys. Rev. B **15**, 2884 (1977).

[112] M. Levy and J. P. Perdew, Phys. Rev. A **32**, 2010 (1985).

[113] A. Becke, J. Chem. Phys. **98**, 1372 (1993).

[114] J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss, J. Chem. Phys. **90**, 5622 (1989).

[115] A. D. Becke, J. Chem. Phys. **104**, 1040 (1996).

[116] J. Perdew, M. Ernzerhof, and K. Burke, The Journal of Chemical Physics **105**, 9982 (1996).

[117] C. Adamo and V. Barone, Chemical Physics Letters **274**, 242 (1997).

[118] C. Adamo and V. Barone, Chemical Physics Letters **298**, 113 (1998).

[119] C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999).

[120] M. Ernzerhof and G. E. Scuseria, J. Chem. Phys. **110**, 5029 (1999).

[121] D. Bylander and L. Kleinman, International Journal of Modern Physics B **10**, 399 (1996).

[122] E. Engel, in *A Primer in density functional theory*, edited by C. Fiolhais, F. Nogueira, and M. A. L. Marques (Springer-Verlag, Berlin, 2003).

[123] A. Görling and M. Levy, Phys. Rev. A **50**, 196 (1994).

[124] J. Krieger, Y. Li, and G. Iafrate, Physics Letters A **146**, 256 (1990).

[125] M. Städele, J. Majewski, P. Vogl, and A. Görling, Phys. Rev. Lett. **79**, 2089 (1997).

[126] M. Städele, M. Moukara, J. Majewski, P. Vogl, and A. Görling, Phys. Rev. B **59**, 10031 (1999).

[127] R. Magyar, A. Fleszar, and E. Gross, Phys. Rev. B **69**, 45111 (2004).

[128] M. Grüning, A. Marini, and A. Rubio, J. Chem. Phys. **124**, 154108 (2006).

[129] M. Grüning, A. Marini, and A. Rubio, Phys. Rev. B **74**, 161103 (2006).

[130] A. Seidl, A. Görling, P. Vogl, J. A. Majewski, and M. Levy, Phys. Rev. B **53**, 3764 (1996).

[131] F. Fuchs, J. Furthmüller, F. Bechstedt, M. Shishkin, and G. Kresse, Phys. Rev. B **76**, 115109 (2007).

[132] M. Marsman, J. Paier, A. Stroppa, and G. Kresse, Journal of Physics: Condensed Matter **20**, 064201 (2008).

[133] A. Savin and H. Flad, International Journal of Quantum Chemistry **56**, 327 (1995).

[134] T. Leininger, H. Stoll, H. Werner, and A. Savin, Chemical Physics Letters **275**, 151 (1997).

[135] J. Toulouse, F. m. c. Colonna, and A. Savin, Phys. Rev. A **70**, 062505 (2004).

[136] J. Toulouse, F. Colonna, and A. Savin, J. Chem. Phys. **122**, 014110 (2005).

[137] J. Heyd, G. Scuseria, and M. Ernzerhof, J. Chem. Phys. **118**, 8207 (2003).

[138] A. Krukau, O. Vydrov, A. Izmaylov, and G. Scuseria, J. Chem. Phys. **125**, 224106 (2006).

[139] A. Lee, S. Taylor, J. Dombroski, and P. Gill, Phys. Rev. A **55**, 3233 (1997).

[140] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).

[141] V. Galitskii and A. Migdal, Zhur. Eksptl. i Teoret. Fiz **34**, (1958).

[142] G. Strinati, La Rivista del Nuovo Cimento (1978-1999) **11**, 1 (1988).

[143] B. Farid, in *Electron correlation in the solid state*, edited by N. March (Imperial College Press, London, 1995).

[144] M. Stone and P. Goldbart, *Mathematics for physics: a guided tour for graduate students* (Cambridge Univ. Press, Cambridge, 2009).

[145] P. Morse and H. Feshbach, .

[146] G. Csanak, H. Taylor, and R. Yaris, Adv. At. Mol. Phys **7**, 287 (1971).

[147] L. Ahlfors, *Complex analysis* (McGraw-Hill, New York, 1953).

[148] F. Bruneval, Ph.D. thesis, Ecole Polytechnique, Palaiseau, France, 2005.

[149] P. C. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959).

[150] L. Kadanoff and G. Baym, *Quantum statistical mechanics* (W.A. Benjamin, New York, 1962).

[151] J. Hubbard, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 539 (1957).

[152] L. Hedin, Phys. Rev. **139**, A796 (1965).

[153] L. Hedin, Journal of Physics: Condensed Matter **11**, R489 (1999).

[154] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).

[155] F. Bruneval, N. Vast, and L. Reining, Phys. Rev. B **74**, 045102 (2006).

[156] L. Hedin and B. Lundqvist, Journal of Physics C: Solid State Physics **4**, 2064 (1971).

[157] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10159 (1988).

[158] W. Ku and A. G. Eguiluz, Phys. Rev. Lett. **89**, 126401 (2002).

[159] K. Delaney, P. Garcia-Gonzalez, A. Rubio, P. Rinke, and R. Godby, Phys. Rev. Lett. **93**, 249701 (2004).

[160] F. Aryasetiawan and O. Gunnarsson, Reports on Progress in Physics **61**, 237 (1998).

[161] L. Hedin and S. Lundqvist, Solid State Physics **23**, 1 (1969).

[162] G. Arbman and U. von Barth, Journal of Physics F: Metal Physics **5**, 1155 (1975).

[163] M. L. Tiago, S. Ismail-Beigi, and S. G. Louie, Phys. Rev. B **69**, 125212 (2004).

[164] A. Marini, G. Onida, and R. Del Sole, Phys. Rev. Lett. **88**, 16403 (2001).

[165] M. Rohlfing, P. Krüger, and J. Pollmann, Phys. Rev. Lett. **75**, 3489 (1995).

[166] P. Rinke, A. Qteish, J. Neugebauer, C. Freysoldt, and M. Scheffler, New J. Phys. **7**, 126 (2005).

[167] E. Kioupakis, P. Zhang, M. Cohen, and S. Louie, Phys. Rev. B **77**, 155114 (2008).

[168] H. Jiang, R. I. Gomez-Abal, P. Rinke, and M. Scheffler, Phys. Rev. Lett. **102**, 126403 (2009).

[169] A. Georges, G. Kotliar, W. Krauth, and M. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).

[170] M. van Schilfgaarde, T. Kotani, and S. Faleev, Phys. Rev. Lett. **96**, 226402 (2006).

[171] B. Holm and U. Von Barth, Phys. Rev. B **57**, 2108 (1998).

[172] E. L. Shirley, Phys. Rev. B **54**, 7758 (1996).

[173] W.-D. Schöne and A. G. Eguiluz, Phys. Rev. Lett. **81**, 1662 (1998).

[174] B. Holm and F. Aryasetiawan, Phys. Rev. B **62**, 4858 (2000).

[175] G. Baym and L. P. Kadanoff, Phys. Rev. **124**, 287 (1961).

[176] A. Schindlmayr, T. Pollehn, and R. Godby, Phys. Rev. B **58**, 12684 (1998).

[177] R. Ummels, P. Bobbert, and W. Van Haeringen, Phys. Rev. B **57**, 11962 (1998).

[178] S. V. Faleev, M. van Schilfgaarde, and T. Kotani, Phys. Rev. Lett. **93**, 126406 (2004).

[179] T. Kotani, M. van Schilfgaarde, and S. V. Faleev, Phys. Rev. B **76**, 165106 (2007).

[180] M. Gatti, F. Bruneval, V. Olevano, and L. Reining, Phys. Rev. Lett. **99**, 266402 (2007).

[181] F. Bruneval *et al.*, Phys. Rev. Lett. **97**, 267601 (2006).

[182] A. Chantis *et al.*, Phys. Rev. B **76**, 165126 (2007).

[183] A. N. Chantis, R. C. Albers, M. D. Jones, M. van Schilfgaarde, and T. Kotani, Phys. Rev. B **78**, 081101 (2008).

[184] R. Laskowski *et al.*, Phys. Rev. B **79**, 165209 (2006).

[185] R. W. Godby and R. J. Needs, Phys. Rev. Lett. **62**, 1169 (1989).

[186] W. von der Linden and P. Horsch, Phys. Rev. B **37**, 8351 (1988).

[187] G. Engel and B. Farid, Phys. Rev. B **47**, 15931 (1993).

[188] R. Shaltaf, G. Rignanese, X. Gonze, F. Giustino, and A. Pasquarello, Phys. Rev. Lett. **100**, 186401 (2008).

[189] P. Umari, G. Stenuit, and S. Baroni, Phys. Rev. B **79**, 201104 (2009).

[190] P. Trevisanutto, C. Giorgetti, L. Reining, M. Ladisa, and V. Olevano, Phys. Rev. Lett. **101**, 226405 (2008).

[191] S. Lebègue, B. Arnaud, M. Alouani, and P. Bloechl, Phys. Rev. B **67**, 155208 (2003).

[192] M. van Schilfgaarde, T. Kotani, and S. Faleev, Phys. Rev. Lett. **96**, 226402 (2006).

[193] M. van Schilfgaarde, T. Kotani, and S. V. Faleev, Phys. Rev. B **74**, 245125 (2006).

[194] L. Ley, R. A. Pollak, F. R. McFeely, S. P. Kowalczyk, and D. A. Shirley, Phys. Rev. B **9**, 600 (1974).

[195] C. J. Vesely, R. L. Hengehold, and D. W. Langer, Phys. Rev. B **5**, 2296 (1972).

[196] P. D. C. King, T. D. Veal, A. Schleife, J. Zúñiga Pérez, B. Martel, P. H. Jefferson, F. Fuchs, V. Muñoz Sanjosé, F. Bechstedt, and C. F. McConville, Phys. Rev. B **79**, 205205 (2009).

[197] A. R. H. Preston, B. J. Ruck, L. F. J. Piper, A. DeMasi, K. E. Smith, A. Schleife, F. Fuchs, F. Bechstedt, J. Chai, and S. M. Durbin, Phys. Rev. B **78**, 155114 (2008).

[198] R. A. Powell, W. E. Spicer, and J. C. McMenamin, Phys. Rev. B **6**, 3056 (1972).

[199] P. Rinke, A. Qteish, J. Neugebauer, C. Freysoldt, and M. Scheffler, New J. Phys. **7**, 126 (2005).

[200] A. Schleife, F. Fuchs, J. Furthmüller, and F. Bechstedt, Phys. Rev. B **73**, 245212 (2006).

[201] P. Schröer, P. Krüger, and J. Pollmann, Phys. Rev. B **47**, 6971 (1993).

[202] S. Laribi, Master's thesis, 2000.

[203] J. Merino, J. de Vidales, S. Mahanty, R. Díaz, F. Rueda, and M. León, J. Appl. Phys. **80**, 5610 (1996).

[204] M. Yu Rigan, V. I. Tkachenko, N. P. Stasiak, and L. G. Novikova, High-pure Substances, Russian Acad. Sci. **2**, 126 (1990).

[205] G. Zahn and P. Paufler, Crystal research and technology(1979) **23**, 499 (1988).

[206] H. Kitahara, K. Suda, and N. Ishizawa, Rep. Res. Lab. Eng. Mater, Tohoku Inst. Technol. **21**, 1 (1996).

[207] F. Frolow, L. Chernyak, D. Cahen, H. Hallak, J. Gabboun, A. Kvick, and H. Graafsma, *CONFERENCE SERIES-INSTITUTE OF PHYSICS* (PUBLISHER, ADDRESS, 1998), Vol. 152, pp. 67–70.

[208] L. Kaplan, G. Leitus, V. Lyakhovitskaya, F. Frolow, H. Hallak, Å. Kvick, and D. Cahen, Advanced Materials **12**, 366 (2000).

[209] W. Paszkowicz, R. Lewandowska, and R. Bacewicz, Journal of Alloys and Compounds **362**, 241 (2004).

[210] R. Reena Philip and B. Pradeep, Thin Solid Films **472**, 136 (2005).

[211] T. Maeda and T. Wada, Journal of Physics and Chemistry of Solids **66**, 1924 (2005).

[212] J. Lazewski, K. Parlinski, B. Hennion, and R. Fouret, Journal of Physics: Condensed Matter **11**, 9665 (1999).

[213] J. Lazewski, H. Neumann, K. Parlinski, G. Lippold, and B. Stanbery, Phys. Rev. B **68**, 144108 (2003).

[214] Ç. Kiliç and A. Zunger, Appl. Phys. Lett. **83**, 2007 (2003).

[215] P. Erhart, A. Klein, R. G. Egdell, and K. Albe, Phys. Rev. B **75**, 153205 (2007).

[216] J. E. Jaffe and A. Zunger, Phys. Rev. B **29**, 1882 (1984).

[217] H. Sommer, A. Weiss, H. Neumann, and R. Tomlinson, Crystal Research and Technology **25**, 1183 .

[218] W. Braun, A. Goldmann, and M. Cardona, Phys. Rev. B **10**, 5069 (1974).

[219] S. Kono and M. Okusawa, Journal of the Physical Society of Japan **37**, 1301 (1974).

[220] F. D. Jiang and J. Y. Feng, Semicond. Sci. Technol. **23**, 025001 (2008).

[221] L. F. J. Piper, T. D. Veal, P. H. Jefferson, C. F. McConville, F. Fuchs, J. Furthmüller, F. Bechstedt, H. Lu, and W. J. Schaff, Phys. Rev. B **72**, 245319 (2005).

[222] P. Lukashev, W. Lambrecht, T. Kotani, and M. van Schilfgaarde, Phys. Rev. B **76**, 195202 (2007).

[223] C. Friedrich, A. Schindlmayr, S. Blügel, and T. Kotani, Phys. Rev. B **74**, 45104 (2006).

[224] S. Lany and A. Zunger, Phys. Rev. B **78**, 235104 (2008).

[225] P. Rinke, A. Janotti, M. Scheffler, and C. Van de Walle, Phys. Rev. Lett. **102**, 26402 (2009).

[226] E. Burstein, Phys. Rev. **93**, 632 (1954).

[227] G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995).

[228] J. Guillemoles, Thin Solid Films **361**, 338 (2000).

[229] R. Eryiğit, C. Parlak, and R. Eryiğit, The European Physical Journal B **33**, 251 (2003).

[230] C. Parlak and R. Eryiğit, Phys. Rev. B **70**, 75210 (2004).

[231] U. Von Barth and C. Gelatt, Phys. Rev. B **21**, 2222 (1980).

[232] D. Hamann, Phys. Rev. B **40**, 2980 (1989).

[233] N. Troullier and J. Martins, Phys. Rev. B **43**, 1993 (1991).

[234] G. Kerker, Journal of Physics C: Solid State Physics **13**, L189 (1980).

[235] L. Kleinman and D. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[236] X. Gonze, R. Stumpf, and M. Scheffler, Phys. Rev. B **44**, 8503 (1991).

[237] P. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[238] D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).

[239] M. Shishkin and G. Kresse, Phys. Rev. B **74**, 35101 (2006).

[240] J. J. Rehr and R. C. Albers, Rev. Mod. Phys. **72**, 621 (2000).

[241] S. Schorr, M. Tovar, N. Stüßer, and K. Bente, Physica B: Physics of Condensed Matter **350**, 411 (2004).

[242] H. Spiess, U. Haeberlen, G. Brandt, A. Rauber, and J. Schneider, phys. stat. sol.(b) **62**, 183 (1974).

[243] S. C. Abrahams and J. L. Bernstein, J. Chem. Phys **59**, 5415 (1973).

[244] S. Nakamura, M. Senoh, and N. Iwasa, Jpn. J. Appl. Phys. **31**, L139 (1992).

[245] J. Neugebauer and C. G. Van de Walle, Phys. Rev. Lett. **75**, 4452 (1995).

[246] F. Benko and F. Koffyberg, J. Phys. Chem. Solids **45**, 57 (1984).

[247] H. Kawazoe, M. Yasukawa, H. Hyodo, M. Kurita, H. Yanagi, and H. Hosono, Nature **389**, 939 (1997).

[248] K. Ueda, T. Hase, H. Yanagi, H. Kawazoe, H. Hosono, H. Ohta, M. Orita, and M. Hirano, J. Appl. Phys. **89**, 1790 (2001).

[249] H. Yanagi, T. Hase, S. Ibuki, K. Ueda, and H. Hosono, Appl. Phys. Lett. **78**, 1583 (2001).

[250] H. Yanagi, S. Inoue, K. Ueda, H. Kawazoe, H. Hosono, and N. Hamada, J. Appl. Phys. **88**, 4159 (2000).

[251] A. Banerjee, S. Kundoo, and K. Chattopadhyay, Thin Solid Films **440**, 5 (2003).

[252] C. Ong and H. Gong, Thin Solid Films **445**, 299 (2003).

[253] T. Dittrich, L. Dloczik, T. Guminskaya, M. Lux-Steiner, N. Grigorieva, and I. Urban, Appl. Phys. Lett. **85**, 742 (2004).

[254] E. Alkoy and P. Kelly, Vacuum **79**, 221 (2005).

[255] A. Banerjee, R. Maity, and K. Chattopadhyay, Mat. Lett. **58**, 10 (2004).

[256] A. Banerjee and K. Chattopadhyay, J. Appl. Phys. **97**, (2005).

[257] A. Banerjee, C. Ghosh, S. Das, and K. Chattopadhyay, Physica B **370**, 264 (2005).

[258] R.-S. Yu, S.-C. Liang, C.-J. Lu, D.-C. Tasi, and F.-S. Shieu, Appl. Phys. Lett. **90**, 191117 (2007).

[259] S. Gilliland, J. Pellicer-Porres, A. Segura, A. Munoz, P. Rodriguez-Hernandez, D. Kim, M. S. Lee, and T. Y. Kim, Phys. Stat. Sol. (b) **244**, 309 (2007).

[260] J. Pellicer-Porres, A. Segura, A. Gilliland, A. Munoz, P. Rodriguez-Hernendez, D. Kim, M. Lee, and T. Kim, Appl. Phys. Lett. **88**, 181904 (2006).

[261] C. W. Teplin, T. Kaydanova, D. L. Young, J. D. Perkins, D. S. Ginley, A. Ode, and D. W. Readey, Appl. Phys. Lett. **85**, 3789 (2004).

[262] M. Sasaki and M. Shimode, J. Phys. Chem. Solids **64**, 1675 (2003).

[263] J. Vidal, S. Botti, P. Olsson, J.-F. m. c. Guillemoles, and L. Reining, Phys. Rev. Lett. **104**, 056401 (2010).

[264] E. N. Brothers, A. F. Izmaylov, J. O. Normand, V. Barone, and G. E. Scuseria, J. Chem. Phys. **129**, 011102 (2008).

[265] J. Paier, M. Marsman, and G. Kresse, Phys. Rev. B **78**, 121201 (2008).

[266] J. Hafner, J. Comput. Chem. **29**, 2044 (2008).

[267] W. G. Aulbur, L. Jönsson, and J. Wilkins, Solid State Phys. **54**, 1 (2000).

[268] B. J. Ingram, T. O. Mason, R. Asahi, K. T. Park, and A. J. Freeman, Phys. Rev. B **64**, 155114 (2001).

[269] L. Shi, Z. Fang, and J. Li, J. Appl. Phys. **104**, 073527 (2008).

[270] X. Nie, S. Wei, and S. Zhang, Phys. Rev. Lett. **88**, (2002).

[271] J. Robertson, P. Peacock, M. Towler, and R. Needs, Thin Solid Films **411**, 96 (2002).

[272] M. N. Huda *et al.*, Phys. Rev. B **80**, 035205 (2009).

[273] G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).

[274] X. Gonze *et al.*, Z. Kristallogr. **220**, 558 (2005).

[275] J. Pellicer-Porres, A. Segura, and D. Kim, Semicond. Sci. Technol. **24**, 015002 (2009).

[276] F. Bechstedt, K. Seino, P. H. Hahn, and W. G. Schmidt, Phys. Rev. B **72**, 245114 (2005).

[277] I. Hamada and H. Katayama-Yoshida, Physica B **376–377**, 808 (2006).

# List of Figures