UNIVERSITY OF THE PHILIPPINES

Bachelor of Science in Applied Physics

Christian Loer T. Llemit

First principle calculations of defect structures in Zinc Oxide

Thesis Adviser:

Roland V. Sarmago, Ph.D.

National Institute of Physics

University of the Philippines Diliman

Date of Submission:
July 2020

Thesis classification:
P

This thesis is not available to the public. Please ask the library for assistance.

National Institute of Physics
College of Science
University of the Philippines
Diliman, Quezon City

ENDORSEMENT

This is to certify that this undergraduate thesis entitled **First principle calculations of defect structures in Zinc Oxide** prepared and submitted by Christian Loer T. Llemit in partial fulfillment of the requirements for the degree of Bachelor of Science in Applied Physics, is hereby accepted.

ROLAND V. SARMAGO, Ph.D.
Thesis Adviser

The National Institute of Physics endorses acceptance of this undergraduate thesis as partial fulfillment of the requirements for the degree of Bachelor of Science in Applied Physics.

WILSON O. GARCIA, Ph.D.
Director
National Institute of Physics

The thesis is hereby officially accepted as partial fulfillment of the requirements for the degree of Bachelor of Science in Applied Physics.

GIOVANNI A. TAPANG, Ph.D.
Dean, College of Science

Acknowledgments

Nulla ac nisl. Nullam urna nulla, ullamcorper in, interdum sit amet, gravida ut, risus. Aenean ac enim. In luctus. Phasellus eu quam vitae turpis viverra pellentesque. Duis feugiat felis ut enim. Phasellus pharetra, sem id porttitor sodales, magna nunc aliquet nibh, nec blandit nisl mauris at pede. Suspendisse risus risus, lobortis eget, semper at, imperdiet sit amet, quam. Quisque scelerisque dapibus nibh. Nam enim. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Nunc ut metus. Ut metus justo, auctor at, ultrices eu, sagittis ut, purus. Aliquam aliquam.

Etiam pede massa, dapibus vitae, rhoncus in, placerat posuere, odio. Vestibulum luctus commodo lacus. Morbi lacus dui, tempor sed, euismod eget, condimentum at, tortor. Phasellus aliquet odio ac lacus tempor faucibus. Praesent sed sem. Praesent iaculis. Cras rhoncus tellus sed justo ullamcorper sagittis. Donec quis orci. Sed ut tortor quis tellus euismod tincidunt. Suspendisse congue nisl eu elit. Aliquam tortor diam, tempus id, tristique eget, sodales vel, nulla. Praesent tellus mi, condimentum sed, viverra at, consectetuer quis, lectus. In auctor vehicula orci. Sed pede sapien, euismod in, suscipit in, pharetra placerat, metus. Vivamus commodo dui non odio. Donec et felis.

Etiam suscipit aliquam arcu. Aliquam sit amet est ac purus bibendum congue. Sed in eros. Morbi non orci. Pellentesque mattis lacinia elit. Fusce molestie velit in ligula. Nullam et orci vitae nibh vulputate auctor. Aliquam eget purus. Nulla auctor wisi sed ipsum. Morbi porttitor tellus ac enim. Fusce ornare. Proin ipsum enim, tincidunt in, ornare venenatis, molestie a, augue. Donec vel pede in lacus sagittis porta. Sed hendrerit ipsum quis nisl. Suspendisse quis massa ac nibh pretium cursus. Sed sodales. Nam eu neque quis pede dignissim ornare. Maecenas eu purus ac urna tincidunt congue.

Donec et nisl id sapien blandit mattis. Aenean dictum odio sit amet risus. Morbi purus. Nulla a est sit amet purus venenatis iaculis. Vivamus viverra purus vel magna. Donec in justo sed odio malesuada dapibus. Nunc ultrices aliquam nunc. Vivamus facilisis pellentesque velit. Nulla nunc velit, vulputate dapibus, vulputate id, mattis ac, justo. Nam mattis elit dapibus purus. Quisque enim risus, congue non, elementum ut, mattis quis, sem. Quisque elit.

Maecenas non massa. Vestibulum pharetra nulla at lorem. Duis quis quam id lacus dapibus interdum. Nulla lorem. Donec ut ante quis dolor bibendum condimentum. Etiam egestas tortor vitae lacus. Praesent cursus. Mauris bibendum pede at elit. Morbi et felis a lectus interdum facilisis. Sed suscipit gravida turpis. Nulla at lectus. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Praesent nonummy luctus nibh. Proin turpis nunc, congue eu, egestas ut, fringilla at, tellus. In hac habitasse platea dictumst.

ABSTRACT

First principle calculations of defect structures in Zinc Oxide

Christian Loer T. Llemit
University of the Philippines, 2020

Adviser:
Roland V. Sarmago, Ph.D.

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetuer odio sem sed wisi.

Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetuer eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetuer tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Zinc Oxide (ZnO) has remained the focus of rigorous theoretical and experimental studies owing to its unique electronic and optical properties, and its application as an energy sources [1–5]. ZnO is a wide direct band gap with an experimental value ranging from 3.44 eV at 4 K to 3.37 eV at 300 K emitting at ultraviolet region [6–8]. In addition, this ionic semiconductor is known to host excitons with a large binding energy of 60 meV that persists even above room temperature [2], rendering its utility as a UV laser source [9], tunable UV photodetectors [10], and scintillator [11, 12].

However, realization of a perfect crystal is very difficult or impossible to achieve in experimental conditions due to presence of impurities even at ultra-high vacuum environment [13] and possibility of forming native defects due to imperfections in the crystal lattice [14]. These defects often directly or indirectly control the material properties such as luminescence efficiency and carrier lifetime [15]. In addition, presence of defects may lead to formation of midgap states which causes visible emissions other than the band to band emission [16]. The nonstoichiometry as well as unintenional n-type conductivity in ZnO were also attributed to defects [17, 18]. Many experiments have been devoted to characterize these native defects and impurities, in the hope that it will improve the properties of ZnO to its intended application [19–23]. Thus, understanding the behavior of defects in ZnO crystal is important to its successful application as a semiconductor.

First principles or ab-initio calculations have offered various insights into the nature of defects without the need for experimentation. Density Functional Theory (DFT) [24, 25] have been used in most first-principles calculations which proved to be an indispensable tool in probing the energetics, atomic and electronic structures of defects in semiconductors, including but not limited to native defects, impurities, dopants, and surface defects [26–28].

There are many published papers on first-principles calculations based on DFT with exchange-correlation functionals of Local Density Approximation (LDA) and Generalized Gradient Approximation (GGA) for native point defects in wurtzite ZnO [29–36]. However, standard DFT calculations have led to severe underestimation of the band gap which results to uncertainties in the formation energies and transition levels of the defect states. These uncertainties have led to significant discrepancies in the conclusions drawn from various published reports using such calculations. Hence, various correction schemes have been employed to improve the band gap. One such corrective approach is the use of Hubbard-U method to treat the over-delocalization of $3d$ Zn orbitals and $2p$ O orbitals, thereby increasing the band gap by shifting the valence band downwards and conduction band upwards [37–40].

1.1 Purpose and Motivation

In order to characterize and validate the effectiveness of ZnO as a potential semiconductor in its own niche of applications, it is important to first gain knowledge of the energetics of defects formation. The properties of semiconductors in general are dependent on the processes that are occurring on the atomic level inside the solid crystal. Introducing defects such as vacancies, interstitials, substitutions, antisites, and impurity atoms cause the formation of mid-gap states. These states are populated by charge carriers and serve as luminescent centers, thereby producing emissions in the visible region. It is therefore imperative to know the underlying mechanisms of these defects so that it can be possible to tune its properties according to its intended purpose. First principle calculations, in particular Density Functional Theory (DFT), will enable first order exploration on the characteristics of these defects without the need of doing physical experimentation. Investigations become more focused as soon as simulations indicate most likely pathway. It is very useful to have *a priori* knowledge of defects in order to optimize the appropriate parameters used in the experiment, thereby saving manhours and materials consumption.

1.2 Objectives

The main goal of this thesis is to understand the physics of defect formation through the study of its energetics and electronic structures. Specifically, the study aims to

- obtain the bandstructure and density of states of native point defects in wurtzite ZnO
- determine which atomic orbitals contribute to the defect energy level through projected density of states calculation
- calculate the defect formation energies and quantitatively describe the stabilities of defects
- calculate the transition level of charged defects

1.3 Scope and Limitations

This study considers only the native point defects in ZnO: oxygen and zinc vacancies (V_O and V_{Zn}), interstitials (O_i and Zn_i), antisites (O_{Zn} and Zn_O) and charged defects (V_O^{1+} and V_O^{2+}; V_{Zn}^{1-} and V_{Zn}^{2-}; ${O_i}^{1-}$ and ${O_i}^{2-}$; ${Zn_i}^{1+}$ and ${Zn_i}^{2+}$). Doping of impurity atoms are not considered in this study. Surface states that exist in vacuum-surface boundaries are not considered also since the study is limited to simulation of bulk solids. The correction scheme used for underestimation of band gap is Hubbard-U method and no additional corrections to address spurious interactions between charged defects since this study is not rigorous enough to investigate these interactions.

1.4 Thesis Outline

This thesis is organized as follows. Chapter 2 will discuss related literatures about ZnO. This includes pertinent properties of ZnO such as crystal structure and native point defects. Chapter 3 will discuss the theoretical framework in doing Density Functional Theory (DFT) calculations. Included in this chapter are various approximations used to simplify calculations of many-body problem into an independent non-interacting body problem. Chapter 4 will discuss the application of DFT in bulk solids. The bulk solid will be modeled based on supercell approach and necessary data sampling techniques needed to have a computationally tractable simulation are discussed in this chapter. Chapter **??** will discuss the technical details of running DFT in a simulation software. Chapter **??** will focus on the results of the simulation of the native point defects in ZnO. This chapter will report the energetics, atomic and electronic structures of the defects. Lastly, Chapter **??** will the summarize the results and will offer some recommendations.

Chapter 2

Review of Related Literature

2.1 Crystal Structure of ZnO

There are three possible crystal structures (phases) of ZnO, namely: wurtzite, zinc blende, and rocksalt as schematically shown in Figure 2.1. The zinc-blende phase can be only stabilized in cubic substrates while rocksalt phase forms at relatively high pressure. The wurtzite phase is the most stable phase in ambient conditions [2]. The wurtzite structure belongs to the space group C_{6v}^4 in Schoenflies notation, $P6_3mc$ in the Hermann–Mauguin notation, and space group number 186 in International Tables for Crystallography (ITA) notation [41]. The crystal structure of wurtzite ZnO unit cell is shown in Figure 2.1a. Each zinc atom is surrounded by four oxygen atoms, which are located at the corners of an irregular tetrahedron as visualized in Figure 2.2. In a similar vein, the oxygen atom is surrounded by four zinc atoms but with different bond lengths. The wurtzite structure belongs to hexagonal Bravais lattice with two variable lattice constants a and c with a ratio of $\sqrt{8/3} \approx 1.633$ for an ideal crystal. The wurtzite ZnO unit cell has four basis atoms located at $(0, 0, 0)$, $a(1/2, \sqrt{3}/6, c/2a)$ for the Zn atoms and $(0, 0, uc)$, $a(1/2, \sqrt{3}/6, [2u + 1]c/2a)$ for the O atoms. Here u is internal parameter which denotes the shortest bond length between Zn and O atoms expressed as a fraction of c ($u = 0.375$ in ideal crystal). However, in real ZnO crystal, the wurtzite deviates from the ideal structure either by changing the u parameter or the c/a ratio.

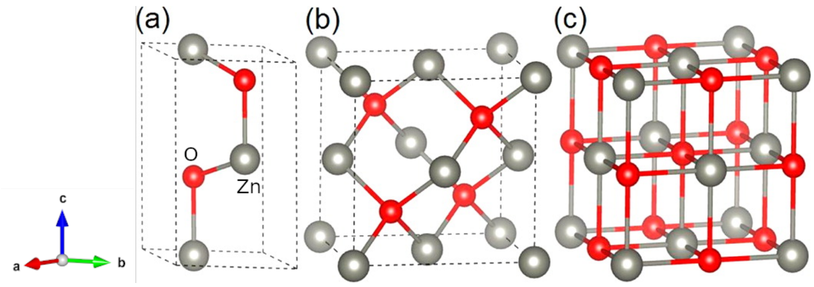

Figure 2.1: Crystal structure of (a)wurtzite (b) zinc-blende and (c) rocksalt ZnO unit cell. Figure taken from [42].

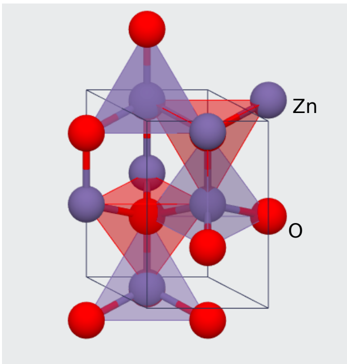

Figure 2.2: The wurtzite crystal structure where oxygen atoms are located at the corners of the tetrahedron enclosing the central zinc atom and vice versa. Figure taken from [43].

2.2 Brillouin Zone Symmetry

The primitive vectors of the direct (real) hexagonal lattice are

$$\vec{\mathbf{a}}_1 = a\hat{\boldsymbol{x}} \tag{2.1}$$

$$\vec{\mathbf{a}}_2 = \frac{a}{2}\hat{\boldsymbol{x}} + \frac{\sqrt{3}a}{2}\hat{\boldsymbol{y}} \tag{2.2}$$

$$\vec{\mathbf{a}}_3 = c\hat{\boldsymbol{z}} \tag{2.3}$$

where a and c are the lattice parameters of hexagonal ZnO. On the other hand, the primitive vectors of the reciprocal lattice can be derived using the formula (**??**) in Appendix **??**

$$\vec{\mathbf{b}}_1 = \frac{2\pi}{a}\left(\hat{\boldsymbol{k}}_x - \frac{1}{\sqrt{3}}\hat{\boldsymbol{k}}_y\right) \tag{2.4}$$

$$\vec{\mathbf{b}}_2 = \frac{2\pi}{a}\left(\frac{2}{\sqrt{3}}\hat{\boldsymbol{k}}_y\right) \tag{2.5}$$

$$\vec{\mathbf{b}}_3 = \frac{2\pi}{a}\left(\frac{a}{c}\hat{\boldsymbol{k}}_z\right) \tag{2.6}$$

The first Brillouin Zone of an hexagonal lattice is also an hexagonal. For more details about reciprocal lattice and Brillouin zones, see Appendix ??. Figure 2.3 shows the symmetry points inside the first Brillouin zone. The capital letters represent the high symmetry (HS) points inside the first Brillouin zone where their notations were traditionally used in the solid state physics literature [44]. Their values are shown in Table 2.1. The different symmetry points of wavevectors correspond to the different kinds of irreducible representations of the space group [45–48].

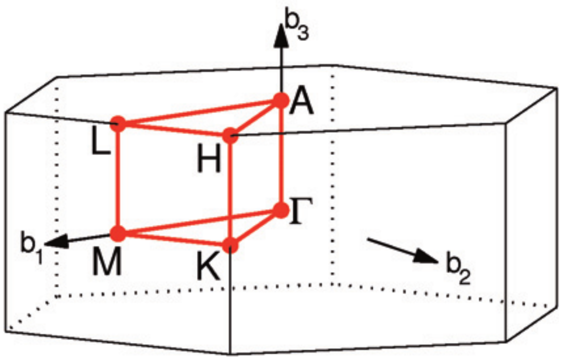

Figure 2.3: The first Brillouin zone of a typical hexagonal Bravais lattice. Figure taken from [49].

Table 2.1: High symmetry points of an hexagonal Bravais lattice

HS	$\times\vec{\mathbf{b}}_1$	$\times\vec{\mathbf{b}}_2$	$\times\vec{\mathbf{b}}_3$
Γ	0	0	0
A	0	0	1/2
H	1/3	1/3	1/2
K	1/3	1/3	0
L	1/2	0	1/2
M	1/2	0	0

2.3 Electronic band structure

The band structure of a semiconductor is very important for determining its potential utility. By just looking at the band structure, one can predict whether a semiconductor will have direct or indirect band gap. The band gap can be quantitatively measured and the wavevectors (or k-points) k for which the valence band maximum (vbm) and conduction band minimum (cbm) will occur can be determined also. Effective masses of holes and electrons can be obtained from the fitted parabola of the dispersion of the top of the valence band and the bottom of the conduction band. Several theoretical approaches have been employed to calculate the band structure of wurtzite ZnO. These theoretical approaches are accurate enough that they agree with experimental spectroscopy measurements such as photoelectron spectroscopy (PES) and angle-resolved photoelectron spectroscopy (ARPES). Density Functional Theory (DFT) is the usual choice for the theoretical band calculations of metals, semiconductors, and insulators in general. An open online initiative by Materials Project [43, 50] have databases on material properties ranging from elasticity, dielectric properties, X-ray diffraction, phonon dispersion, density of states to band structure of more than 120,000 compounds. Figure 2.4 shows the band structure obtained from the Materials Project database. It can be shown that the top of the valence band and the bottom of the conduction band coincides at the Γ point. Thus, the calculation predicted that wurtzite ZnO has a direct band gap. However, the calculated band gap is 0.7317 eV which is more than 75% underestimated from the experimental band gap of 3.37 eV. This is inherent band gap problem that is caused by the formulation of DFT itself. However, various corrections have carried out over the last several decades. A significant improvement in band gap was obtained by Slassi et al. [51] using GGA functional with approximation from Tran–Blaha modified Becke–Johnson (TB-mBJ) where they obtain a value of 2.70 eV. Another report [52] obtained a value of 2.49 eV using hybrid functional GGA-PBE-HSE06. The works of Ma et al. [42] using Hubbard-U correction have calculated a band gap value of 3.40 eV. The GW calculations, which is considered most accurate but the most computationally expensive, of Kim et al. obtained a value of 3.3 eV [53].

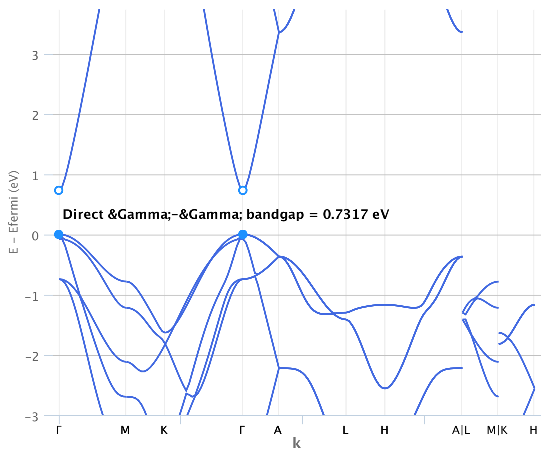

Figure 2.4: Band structure of wurtzite ZnO. Figure obtained from [50].

2.4 Defects

A perfect crystal never exists in nature. Atom arrangements do not follow perfect crystalline patterns since various factors such as temperature, pressure and chemical composition substantially affect the preferred structure. In addition, atoms are relatively immobile in a solid, hence, it is difficult to eliminate whatever imperfections introduced during crystal growth. It is this reason that defects or imperfections exist at stable conditions. Crystal defects can be classified according to the dimension. The 0-dimensional defects affect localized points in the crystal site, thus they are called point defects. The one-dimensional defects are called dislocations. They are lines along crystal lattice where the pattern is broken. The two-dimensional defects are surface defects, which include the external surface boundary and the grain boundaries along which crystals are joined together. Lastly, three-dimensional defects changes the lattice site at a finite volume. This includes precipitates, voids, and inclusion of second-phase particles.

Among these types of defects, the point defects play a vital role in semiconductor engineering. Material properties are substantially altered upon modifications of point defects. Point defects can be classified into native (or intrinsic) and extrinsic defects. Native point defects are formed from the atoms of the host crystal. Extrinsic defects consist of impurity or foreign atoms. This includes doping, addition, and substitution. The point defect can be further subdivided to its type which includes the following

- Vacancies - the absence of an atom from its normal location in a perfect crystal structure
- Interstitials - an atom is occupying an interstitial site, a small void space that under ordinary circumstances is not occupied. Self-interstitials are form when the atom of a host crystal occupy the interstitial sites.
- Substitutionals - formed when an extra atom replaces a host atom
- Antisites - a specific kind of substitutionals in which a host atom occupies the site which was originally occupied by another type of host atom
- Frenkel Defects - an atom displaced from its position to a nearby interstitial site
- Schottky Defects - an equal number of cations and anions are missing from their lattice sites. Hence, electrical neutrality of the crystal is conserved.

Insert figure of defects

Vacancies, interstitals, Schottky, Frenkel, and antisite which do not involve the foreign atoms are called the native point defects. Otherwise, interstitials and substitutionals involving foreign atoms are called extrinsic point defects. A useful notation for dealing with point defects is the use of Kröger-Vink notation [54]. The general notation can be expressed as X_Y^Z where X is element symbol for an atomic species located on a site, else it is written V for vacancy. The Y is the type of site occupied by X; i for an interstitial, otherwise, element symbol for site normally occupied by this element. Lastly Z is the charge of defect normally written as $'$ or $-$ for negative charge; $\cdot$ or $+$ for positive charge; while x, 0 or sometimes omitted for neutral charge. For instance, an oxygen antisite in ZnO crystal with a $2-$ charge state can be written as $\mathrm{O}_{\mathrm{Zn}}^{2-}$. A $+1$ oxygen vacancy is expressed as V_{O}^{+}.

There has been a great deal of interest in studying the point defects in ZnO due to its promising features in optoelectronics. For instance, the photoluminescence properties of ZnO can be controlled by doping an impurity atom [20]. However, some questions were remained unanswered or not yet convincingly addressed. Janotti et al. [14] have shown through DFT calculations that native donor defects such as oxygen vacancy V_{O}, zinc interstitial Zn_i, and zinc antisite Zn_{O} are unlikely to form in n-type ZnO since they have low formation energies, contrary to the conventional wisdom that these defects are the source of observed n-type conductivity in ZnO [55–58]. However, they deduced that native donors (acceptors) serve as a compensation carrier of the predominant acceptor (donor) dopants. The non-stoichiometry of ZnO can be explained by the formation of stable native defects.

The well known green luminescence (GL) band, manifesting as a broad peak around 500-530 nm, observed in nearly all ZnO samples regardless of the fabrication techniques, have been the topic of debate whether this is caused by the native defects or due to uncontrollable impurities during growth. Since the peak is broad, there a great likelihood that it is composed of multiple defects that are possibly interacting with each other to form a defect band.

Chapter 3

Theoretical Framework

3.1 Electronic Structure

The problem of electronic structure methods begins with the attempt to solve the general non-relativistic time-independent Schrödinger equation given as [59]

$$\hat{\mathcal{H}}\Psi = E\Psi \tag{3.1}$$

where $\hat{\mathcal{H}}$ is the Hamiltonian operator for a system of electrons, Ψ is the electronic wavefunction and E is the energy of the system. Consider a single electron in three dimensional system, the Schrödinger equation can be expressed as

$$\hat{\mathcal{H}}\Psi_n = -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\Psi_n + V\Psi_n = \epsilon_n\Psi_n \tag{3.2}$$

where m is the mass of electron, V is the effective potential energy and ϵ_n is the energy of electron in the orbital. The term orbital denotes the solution of the Schrödinger equation for a system of only one electron. This will be useful in later sections because this will allow to distinguish between the exact quantum state of a system of N interacting electrons from the approximate quantum state of N electrons in N orbitals, where each orbital is a solution to one-electron wavefunction in (3.2). If V is zero for the case of free electrons (i.e. non-interacting), then the orbital model is exact.

Since electrons are restricted by the potential inside the atom, the simplest way of solving (3.2) is by considering an infinite potential well. The electrons are confined inside a cube of length L where the potential V inside is zero and infinite at outside must satisfy the boundary condition

$$\Psi_n(L_x, L_y, L_z) = 0 \tag{3.3}$$

where L_x, L_y, L_z can be either 0 or L. The solution will have a sine dependence

$$\Psi_n(x, y, z) = \sqrt{\left(\frac{2}{L}\right)^3} \sin\left(\frac{n_x \pi}{L} x\right) \sin\left(\frac{n_y \pi}{L} y\right) \sin\left(\frac{n_z \pi}{L} z\right) \tag{3.4}$$

where n_x, n_y, n_z are integer quantum states. Provided that $k_i = n_i \pi / L$ where $i = x, y, \text{or } z$; then the energy dispersion relation can be expressed as

$$\epsilon_k = \frac{\hbar^2}{2m}(k_x^2 + k_y^2 + k_z^2) = \frac{\hbar^2}{2m} k^2 \propto k^2 \tag{3.5}$$

Note that energy levels are discretized by the quantum states which arises from imposing the boundary conditions.

3.1.1 Electronic Band structure

Inside the crystal lattice, the periodic arrangement of atoms or ions causes the potential to be periodic which eventually gives rise to the formation of energy bands. The wavefunction Ψ will become periodic in space with a period L and must obey the Born-von Karman boundary condition [60]

$$\Psi_k(x, y, z) = \Psi_k(x + L, y, z) \tag{3.6}$$

and similarly for the y and z coordinates. It can be shown that wavefunctions satisfying (3.2) and (3.6) are the Bloch form of a travelling plane wave

$$\Psi_k(\vec{\mathbf{r}}) = u_k(\vec{\mathbf{r}}) \exp\left(i\vec{\mathbf{k}} \cdot \vec{\mathbf{r}}\right) \tag{3.7}$$

where $u_k(\vec{\mathbf{r}})$ has the period of the crystal lattice with $u_k(\vec{\mathbf{r}}) = u_k(\vec{\mathbf{r}} + \vec{\mathbf{R}})$. Here $\vec{\mathbf{R}}$ is the translation vector which can be simply thought as the periodicity expressed as vector. The Bloch expression can be written as

$$\begin{aligned}
\Psi_k(\vec{\mathbf{r}} + \vec{\mathbf{R}}) &= u_k(\vec{\mathbf{r}} + \vec{\mathbf{R}}) \exp\left(i\vec{\mathbf{k}} \cdot (\vec{\mathbf{r}} + \vec{\mathbf{R}})\right) \\
\Psi_k(\vec{\mathbf{r}} + \vec{\mathbf{R}}) &= u_k(\vec{\mathbf{r}}) \exp\left(i\vec{\mathbf{k}} \cdot \vec{\mathbf{r}}\right) \exp\left(i\vec{\mathbf{k}} \cdot \vec{\mathbf{R}}\right) \\
\Psi_k(\vec{\mathbf{r}} + \vec{\mathbf{R}}) &= \Psi_k(\vec{\mathbf{r}}) \exp\left(i\vec{\mathbf{k}} \cdot \vec{\mathbf{R}}\right)
\end{aligned} \tag{3.8}$$

Notice that the wavefunction differs from the plane wave of free electrons only by a periodic modulation given by the new phase factor. This means that the electrons in the crystal lattice are treated as perturbed weakly by the periodic potential of the ion cores.

Band structure of free electron

A special case of periodicity is where the potential is set to zero, which is applicable for the free electrons. The wavefunction will be a plane wave

$$\Psi_k(\vec{\mathbf{r}}) = \exp\Big(i\vec{\mathbf{k}} \cdot \vec{\mathbf{r}}\Big) \tag{3.9}$$

that represents travelling wave with a momentum $\vec{\mathbf{p}} = \hbar\vec{\mathbf{k}}$. The energy dispersion relation is still given by (3.5) but this time the allowed energy values are distributed essentially from zero to infinity. Figure 3.1 shows the parabolic dependence of energy with the wavevector k. Since the system is periodic in real space, it must be true for the reciprocal space, in this case by $2\pi/a$ where a is some lattice constant. Figure 3.1a shows the extended zone scheme where there are no restrictions on the values of wavevector $\vec{\mathbf{k}}$. When wavevectors are outside the first Brillouin zone (BZ), they can be translated back to the first zone by subtracting a suitable reciprocal lattice vector. In mathematical sense [61]

$$\vec{\mathbf{k}} + \vec{\mathbf{G}} = \vec{\mathbf{k'}} \tag{3.10}$$

where $\vec{\mathbf{k'}}$ is the unrestricted wavevector, $\vec{\mathbf{k}}$ is in the first Brillouin zone, and $\vec{\mathbf{G}}$ is the translational reciprocal lattice vector. The energy dispersion relation can always be written as

$$\begin{aligned}\epsilon(k_x, k_y, k_z) &= \frac{\hbar^2}{2m}(\vec{\mathbf{k}} + \vec{\mathbf{G}})^2 \\ &= \frac{\hbar^2}{2m}[(k_x + G_x)^2 + (k_y + G_y)^2 + (k_z + G_z)^2]\end{aligned} \tag{3.11}$$

Figure 3.1b shows the reduced zone scheme where the bands are folded into the first BZ by applying (3.10). Any energy state beyond the first BZ is the same to a state inside the first BZ with a different band index n.

Band structure of electrons in solids

When atoms are very far from each other with no interaction, each electron occupies specific discrete orbitals such as 1s, 2p, 3d, etc. When they are bring closer enough, the outermost (valence) electrons interact with each other and will result in the energy level splitting. The innermost (core) electrons remain as they are, since they are closer to the nuclei and bounded by a deep potential well. For a solid containing a large N atoms, there will be N orbitals (i.e. N 3d-orbitals) trying to occupy the same energy level. Pauli's exclusion principle will prevent this from happening, hence what

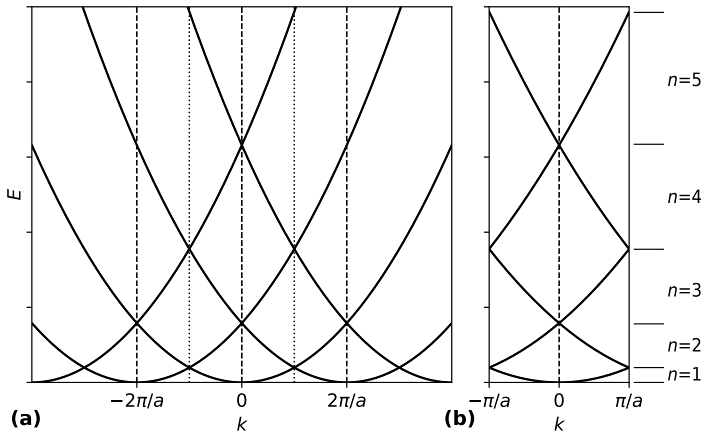

Figure 3.1: Free electron band structure where (a) is in the extended zone scheme and (b) in the reduced zone scheme. The dotted lines in (a) lies the first BZ.

happens is there will be splitting of the energy level that are closely spaced and this will eventually form a continuous band of energy levels. Figure 3.2 summarizes the evolution of energy levels as the atoms are brought together.

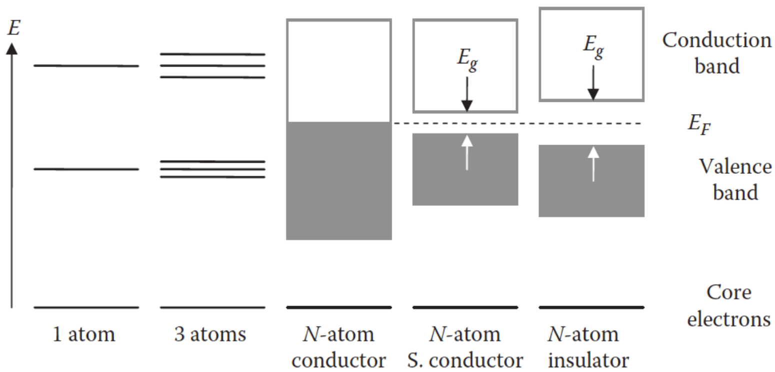

Figure 3.2: Formation of bands and band gaps when isolated atoms are bring closer together. Figure taken from [62]

Another interesting property of band structure is the formation of energy band gaps. This happens when the valence electrons interact with the periodic potential of the nuclei. Assuming a weak periodic potential, most of the band structure will not changed very much, except possibly at the Brillouin zone boundaries with a wavevector of $\vec{\mathbf{k}} = n\pi/a$. The orbitals with the wavevector at zone boundaries, chosen to be at high symmetry points, follows the Bragg diffraction condition and thus are diffracted. The valence electrons are scattered (or reflected) at the zone boundary in which the wavefunction are made up of equal plane waves travelling from the left and from the right. The wavefunction becomes a standing wave that resembles more of those bound states. Hence, there will be a forbidden region where travelling waves are not allowed. If sufficient energy is provided to the electron, they can overcome the binding potential.

The band gap is generally referred to the energy difference between the top of valence band, Valence band maximum (VBM), and the bottom of the conduction band, Conduction band minimum (CBM). If VBM and CBM coincides with each other, the material is said to be a conductor. Electrons can easily occupy the conduction band without any excitation, hence electrons are highly mobile that will lead to high current. For band gaps with a value comparable to the quantity k_BT, where k_B is the Boltzmann constant and T is the absolute temperature near room temperature, then the material is semiconductor. If band gap is much larger than k_BT, then the material is insulator. However, this criterion is very loose because there are materials with large band gaps such as ZnO, $SrIn_2O_4$, and diamond that are categorized as semiconductors. These materials are generally called wide-band gap semiconductors. If the VBM and CBM are located in the same wavevector k, then the gap is direct. Otherwise, it is indirect.

3.1.2 Density of States

Another useful quantity in describing the electronic structure is the density of states (DOS). In general, the density of states can be defined as [63]

$$D(\epsilon) = 2\sum_n \sum_k \delta(\epsilon - \epsilon_n(k)) \tag{3.12}$$

where for each band index n, the sum is over all allowed values of k lying inside the first Brillouin zone. The factor 2 comes from the allowed values of the spin quantum number for each allowed value of k. In the limit of large crystal, the k points are very close together, and the sum can be replaced by an integral. Since each allowed states will take up a volume of $(\Delta k)^3 = \pi^3/V$ where V is the volume of the solid in real space, it is convenient to write (3.12) as

$$D(\epsilon) = 2\,\frac{V}{\pi^3}\sum_n \sum_k \delta(\epsilon - \epsilon_n(k))(\Delta k)^3 \tag{3.13}$$

for in the limit of $V \to \infty$, $\Delta k \to 0$, so that

$$g(\epsilon) = \lim_{V\to\infty}\frac{1}{V}\,D(\epsilon) = \frac{2}{\pi^3}\sum_n \int \delta(\epsilon - \epsilon_n(k))\,\mathrm{d}^3k \tag{3.14}$$

Usually, the total DOS is set to be the number of states per unit energy per unit volume.

The DOS can be projected in terms of the orbital contribution of each atoms. This can be expanded in a complete orthonormal basis as [64]

$$D(\epsilon) = \sum_i D_i(\epsilon) \tag{3.15}$$

$$= \sum_i \sum_n \int \langle\psi_n| \alpha |\psi_n\rangle \, \delta(\epsilon - \epsilon_n(k)) \, \mathrm{d}^3k \tag{3.16}$$

where $D_i(\epsilon)$ is the projected density of states (PDOS) of orbital i with state α.

3.2 Many-body Physics

Despite the simplicity of Schrödinger equation in (3.1), solving it is a formidable task when dealing with many-electron systems. Analytical solutions to this equation only exist for the very simplest systems (i.e. hydrogenic atoms). Solving beyond '2 particle' system (electron and nucleus) is already intractable. In addition, solid state systems typically contains more than hundreds of particles, resulting in hundreds of simultaneous equations. Even the use of computational methods relies on a number of approximations just to make computations feasible enough. Hence, this section will discuss various levels of approximations without neglecting the parameter-free of first-principles calculations.

3.2.1 Many-particle Hamiltonian Operator

The exact many-particle Hamiltonian is consist of five operators which can be expressed as

$$\hat{\mathcal{H}} = \hat{\mathcal{T}}_n + \hat{\mathcal{T}}_e + \hat{\mathcal{V}}_{en} + \hat{\mathcal{V}}_{ee} + \hat{\mathcal{V}}_{nn} \tag{3.17}$$

where the $\hat{\mathcal{T}}$ and $\hat{\mathcal{V}}$ refer to kinetic energy and potential energy, respectively, and the labels e and n denotes the electronic and nuclear coordinates and their derivatives, respectively. This equation can be expanded as

$$\begin{aligned}\hat{\mathcal{H}} = &-\frac{\hbar^2}{2}\sum_I \frac{\nabla^2_{\vec{\boldsymbol{R}}_I}}{M_I} - \frac{\hbar^2}{2}\sum_i \frac{\nabla^2_{\vec{\boldsymbol{r}}_i}}{m_e} \\ &- \frac{1}{4\pi\epsilon_0}\sum_{I,i}\frac{e^2 Z_I}{\left|\vec{\boldsymbol{R}}_I - \vec{\boldsymbol{r}}_i\right|} + \frac{1}{8\pi\epsilon_0}\sum_{i\neq j}\frac{e^2}{|\vec{\boldsymbol{r}}_i - \vec{\boldsymbol{r}}_j|} + \frac{1}{8\pi\epsilon_0}\sum_{I\neq J}\frac{e^2 Z_I Z_J}{\left|\vec{\boldsymbol{R}}_I - \vec{\boldsymbol{R}}_J\right|}\end{aligned} \tag{3.18}$$

where M_I is the mass of the Ith nuclei (or usually ions) with charge Z_I located at site $\vec{\boldsymbol{R}}_I$, and electrons have mass m_e located at site $\vec{\boldsymbol{r}}_i$. The first and second terms are the kinetic energy of the atomic nuclei and electrons, respectively. The last three terms describe the Coulomb interaction between electrons and nuclei, between electrons and other electrons, and between nuclei and other nuclei.

3.2.2 Simplifying Assumptions

Solving (3.18) exactly is very impractical and not worth the effort. Hence, we resort to approximations in order to find acceptable eigenstates.

The first level of approximation is the Born-Oppenheimer approximation or the Adiabatic approximation [65]. It begins with the observation that the mass of nuclei is much larger compared to the electron, as such one can assume that electrons moving in a potential much faster than the nuclei and that the nuclei can be treated as fixed or 'frozen' with respect to motion. As a consequence, the nuclear kinetic energy will be zero and the nuclear interaction with the electron cloud can be treated as an external parameter. Hence, the first term in (3.18) will vanish and the last term reduces to a constant which can be neglected. The third term will become the external potential. The Hamiltonian reduces to

$$\hat{\mathcal{H}} = \hat{\mathcal{T}} + \hat{\mathcal{V}} + \hat{\mathcal{V}}_{ext} \tag{3.19}$$

and using Hartree atomic units $\hbar = m_e = e = 4\pi/\epsilon_0 = 1$ for simplicity

$$\hat{\mathcal{H}} = -\frac{1}{2}\sum_i \nabla^2_{\vec{r}_i} + \frac{1}{2}\sum_{i\neq j}\frac{1}{|\vec{\boldsymbol{r}}_i - \vec{\boldsymbol{r}}_j|} + \sum_i V_I(\left|\vec{\boldsymbol{R}}_I - \vec{\boldsymbol{r}}_j\right|) \tag{3.20}$$

3.2.3 Hartree Method

Since the second term in (3.20) includes electron-electron interaction which is difficult to evaluate, Hartree (1928) had proposed a simplified model where he treated each electrons to be independent and interacts with others in an averaged way [66].This implies that each electron does not recognize others as single entities but rather as a mean Coulomb field. The second term will be replaced by Hartree energy given as

$$\hat{\mathcal{V}}_H = \frac{1}{2}\iint \frac{\rho(\vec{\boldsymbol{r}})\rho(\vec{\boldsymbol{r}}')}{|\vec{\boldsymbol{r}} - \vec{\boldsymbol{r}}'|}\,\mathrm{d}^3r\,\mathrm{d}^3r' \tag{3.21}$$

where $\rho(\vec{\boldsymbol{r}})$ is the electron density. The total energy will be sum of N numbers of one-electron energies

$$E = E_1 + E_2 + \cdots + E_N \tag{3.22}$$

then, the N-electron wavefunction can be approximated as a product of one-electron wavefunctions

$$\Psi = \Psi_1 \times \Psi_2 \times \cdots \times \Psi_N \tag{3.23}$$

Hartree model successfully predicts the ground-state energy of Hydrogen atom to be around -13.6 eV. However, for other systems, Hartree model produced crude estimations because it does not take into account the quantum mechanical effects such as antisymmetry principle and the Pauli's exclusion principle. Moreover, the model does not include the exchange and correlation energies of every interacting electrons in the actual systems.

3.2.4 Hartree-Fock Method

Due to the limitations of Hartree Model, Fock (1930) has taken into account the anti-symmetric property of electron wavefunctions [67]. Pauli's exclusion principle posits that no two fermions can occupy the same quantum state because the wavefunction is antisymmetric upon particle exchange [68]. The many-electron wavefunction will be expressed in terms of Slater determinant [69]

$$\boldsymbol{\Psi} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \Psi_1(\vec{\boldsymbol{r}}_1) & \Psi_2(\vec{\boldsymbol{r}}_1) & \cdots & \Psi_N(\vec{\boldsymbol{r}}_1) \\ \Psi_1(\vec{\boldsymbol{r}}_2) & \Psi_2(\vec{\boldsymbol{r}}_2) & \cdots & \Psi_N(\vec{\boldsymbol{r}}_2) \\ \vdots & \vdots & \vdots & \vdots \\ \Psi_1(\vec{\boldsymbol{r}}_N) & \Psi_2(\vec{\boldsymbol{r}}_N) & \cdots & \Psi_N(\vec{\boldsymbol{r}}_N) \end{vmatrix} \tag{3.24}$$

Using the Slater determinant form of the wavefunction, the Hamiltonian can be written as before with the addition of exchange term

$$\hat{\mathcal{H}}_{HF} = \hat{\mathcal{T}} + \hat{\mathcal{V}}_{ext} + \hat{\mathcal{V}}_H + \hat{\mathcal{V}}_x \tag{3.25}$$

where

$$\hat{\mathcal{V}}_x = -\sum_j \int \frac{\psi_j^*(\vec{\boldsymbol{r}}')\psi(\vec{\boldsymbol{r}}')}{|\vec{\boldsymbol{r}} - \vec{\boldsymbol{r}}'|} \frac{\psi_j(\vec{\boldsymbol{r}})}{\psi(\vec{\boldsymbol{r}})} \,\mathrm{d}\vec{\boldsymbol{r}} \tag{3.26}$$

$\hat{\mathcal{V}}_H$ comes from the Hartree approximation of electron-elecron interaction and $\hat{\mathcal{V}}_x$ comes from the antisymmetric nature of wave function.

3.3 Density Functional Theory (DFT)

Density Functional Theory reframes the problem of calculating electronic properties in terms of the ground state electron density instead of the traditional electronic wave-functions [70]. The incredible success of DFT in predicting ground state properties have led to widespread applications in materials modelling research.

3.3.1 Hohenberg-Kohn (HK) Formalism

The modern formulations of DFT started in the seminal work of Hohenberg and Kohn in 1964 [24]. Hohenberg and Kohn have shown that the ground state properties can be written as unique functional of the ground state electron density. This statement has large implication because the problem of solving $3n$-dimensional equation simultaneously can be replaced by n separate three-dimensional equations with the use of electron density, $\rho(x, y, z)$.

First HK Theorem

The first theorem shows that electron density is a unique functional of the external potential. It states that there is a one-to-one correspondence between the ground state density $\rho_0(r)$ of a many-electron system and the external potential V_{ext}, to within an additive constant. Alternatively, it is impossible to have two external potentials, $V_{ext}(r)$ and $V'_{ext}(r)$, acting on an electron whose difference is not a constant, that give rise to the same ground state electron density, $\rho_0(r)$. That is,

$$\rho(r) = \rho'(r) \qquad \Longleftrightarrow \qquad V'_{ext}(r) - V_{ext}(r) = \text{constant} \tag{3.27}$$

If the external potential is known beforehand, then the ground state electron density can be obtained and vice versa. As the ground state electron density uniquely determines the Hamiltonian of the system, it follows that all measurable properties of the system can be expressed as a functional of the electron density.

Second HK Theorem

The second theorem proves the existence of the energy as a functional of the electron density. It states that there exists a universal functional for the energy $E[\rho]$ such that for any given $V_{ext}(r)$, the exact ground-state energy is the global minimum of this functional, and the ground-state density $\rho_0(r)$ is the density $\rho(r)$ that minimizes the functional. Note that the total energy in HK formulation gives an exact form and not approximate ones. The form of the energy functional can be expressed as

$$E_{HK}[\rho(r)] = \langle\psi|\,\hat{\mathcal{T}} + \hat{\mathcal{V}} + \hat{\mathcal{V}}_{ext}\,|\psi\rangle \tag{3.28}$$

$$= \langle\psi|\,\hat{\mathcal{T}} + \hat{\mathcal{V}}\,|\psi\rangle + \langle\psi|\,\hat{\mathcal{V}}_{ext}\,|\psi\rangle \tag{3.29}$$

$$= F[\rho(r)] + \int V_{ext}(r)\rho(r)\,\mathrm{d}^3r \tag{3.30}$$

where $F[\rho(r)]$ is the unknown functional that includes all internal energies, kinetic, and potential, that are independent of the external potential. The HK theorems only asserts the existence of energy functional but it does not provide a practical solution on solving the energy functional.

3.3.2 Kohn Sham (KS) Formulation

Kohn and Sham (1965) introduced an artificial system of non-interacting electrons with the same ground state electron density as the many-body Schrödinger equation [25]. Instead of using the fully interacting multi-electron wavefunctions, the KS formulation resorts to single-particle wavefunctions for solving the many-body problem. The Kohn-Sham Hamiltonian is just an extension of Hartree-Fock Hamiltonian described in (3.25). However, it was implicitly assumed that $\hat{\mathcal{T}}$ is the kinetic energy operator of non-interacting electrons. This assumption neglects the correlation of the interacting system, hence a correction factor must be added. The kinetic energy of the real interacting system can be rewritten as

$$\hat{\mathcal{T}} = \hat{\mathcal{T}}_{KS} + \hat{\mathcal{V}}_c \tag{3.31}$$

where $\hat{\mathcal{T}}_{KS}$ is kinetic energy of the non-interacting electron, and $\hat{\mathcal{V}}_c$ is the correlation energy that measures how much movement of one electron is influenced by the presence of other electrons. The total KS Hamiltonian has the form

$$\begin{aligned}\hat{\mathcal{H}}_{KS} &= (\hat{\mathcal{T}}_{KS} + \hat{\mathcal{V}}_c) + \hat{\mathcal{V}}_{ext} + \hat{\mathcal{V}}_H + \hat{\mathcal{V}}_x \\ &= \hat{\mathcal{T}}_{KS} + \hat{\mathcal{V}}_{ext} + \hat{\mathcal{V}}_H + \hat{\mathcal{V}}_{xc}\end{aligned} \tag{3.32}$$

where $\hat{\mathcal{V}}_{xc} = \hat{\mathcal{V}}_x + \hat{\mathcal{V}}_c$ is the combined exchange-correlation energy. It is instructive to see that the difference between Hartree Hamiltonian from Hartree-Fock Hamiltonian gives the exchange term while the difference between Hartree-Fock Hamiltonian and Kohn-Sham Hamiltonian gives the correlation term [71]. The theorem of Kohn and Sham can be formally formulated as follows:

The exact ground state density $\rho(\vec{\boldsymbol{r}})$ of an N-electron system is

$$\rho(\vec{\boldsymbol{r}}) = \sum_{i=1}^{N} \phi_i(\vec{\boldsymbol{r}})^* \phi_i(\vec{\boldsymbol{r}}) \tag{3.33}$$

where the single-particle KS orbitals $\phi_i(\vec{\boldsymbol{r}})$ are the N lowest energy solutions of the Kohn-Sham equation

$$\hat{\mathcal{H}}_{KS}\, \phi_i(\vec{\boldsymbol{r}}) = \epsilon_i\, \phi_i(\vec{\boldsymbol{r}}) \tag{3.34}$$

3.3.3 Self Consistent Field Calculation

In order to solve the KS equation (3.34), the Hamiltonian $\hat{\mathcal{H}}_{KS}$ must be known beforehand. However, the Hamiltonian depends entirely on the electron density $\rho(\vec{\boldsymbol{r}})$ that can only be solved from single-particle KS orbital $\phi_i(\vec{\boldsymbol{r}})$ given in (3.33). The orbital $\phi_i(\vec{\boldsymbol{r}})$ are in turn calculated from the KS equation and the cycle continues on. This infinite loop is visualized in Figure 3.3.

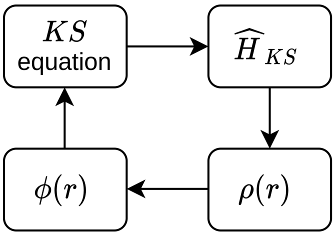

Figure 3.3: Solving Kohn Sham equation leads to a circular argument

To circumvent this, an iterative scheme was developed in which a trial electron density is introduced and the KS equation is iteratively solved to achieve convergence. This iterative process is often referred as Self Consistent Field (SCF) calculation [72]. Specific steps are illustrated in Figure 3.4. First, an initial trial electron density is provided. The trial electron density is usually derived from the superposition of known atomic potentials. Second, the KS equation is solved using the trial electron density. The resulting eigenfunction, in this case the orbital $\phi_i(\vec{\boldsymbol{r}})$, will then be used to calculate the new electron density. The new electron density is compared to the previous electron density and if the error is less than some acceptable deviation, then this will be the ground state density. Otherwise, the electron density is updated and the iteration is repeated kth times until convergence is achieved. Factors that affect the rate of convergence will be discussed on the next chapter.

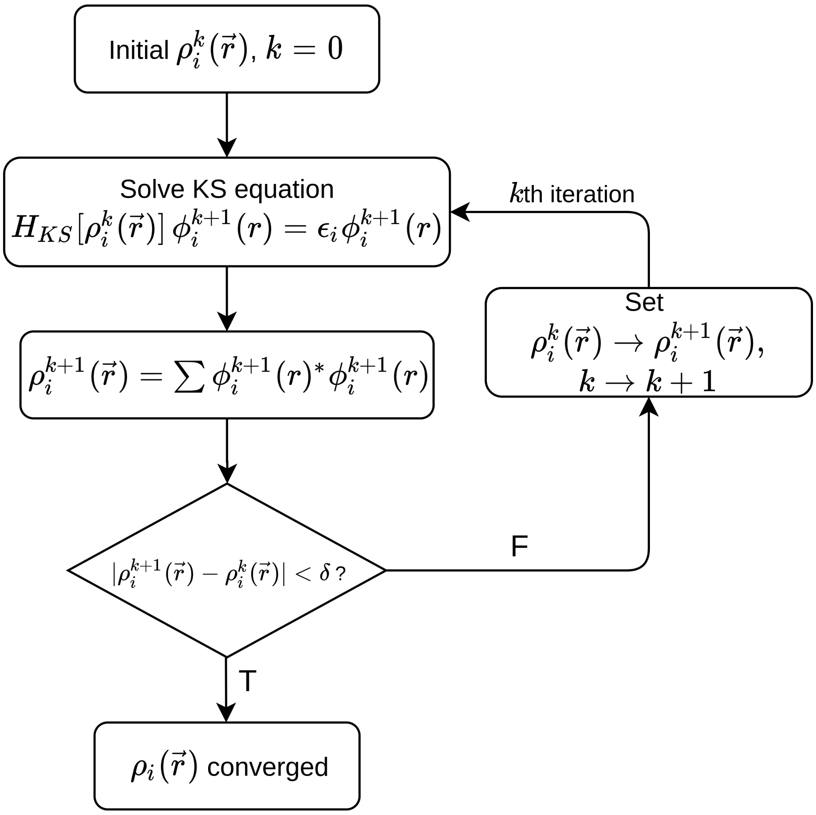

Figure 3.4: Convergence of electron density and other observable quantities using Self Consistent Field calculation

3.4 Exchange-Correlation Functional

So far, no analytical form for the exchange-correlation functional has been found yet that perfectly describes any interacting system [73–75]. The success of DFT depends on the improvement and refinement of the exchange-correlation functional and how it enables to predict many observable properties. Hence, the search for the universal functional is a hot topic of ongoing research. The choice of XC functional varies from different applications of DFT. Thus, there is no one particular functional in the literature which universally performs better than others across all applications.

3.4.1 Local Density Approximation (LDA)

The simplest commonly used exchange-correlation functional is the so called Local Density Approximation (LDA). LDA assumes that the electronic contribution to the exchange-correlation energy from each point in space is the same as to what it would be for a homogeneous electron gas with the uniform density throughout the whole system. This approximation was originally introduced by Kohn and Sham, and holds for a slowly varying density [25]. Using the approximation, the XC energy functional is given by

$$E_{xc}^{\mathrm{LDA}}[\rho] = \int \rho(\vec{\boldsymbol{r}})\, \epsilon_{xc}[\rho(\vec{\boldsymbol{r}})]\, \mathrm{d}^3 r \tag{3.35}$$

where $\epsilon_{xc}[\rho(\vec{\boldsymbol{r}})]$ is the exchange-correlation energy per particle of a uniform electron gas of density $\rho(\vec{\boldsymbol{r}})$. The quantity $\epsilon_{xc}[\rho(\vec{\boldsymbol{r}})]$ can be further split into exchange and corelation contributions

$$\epsilon_{xc}[\rho(\vec{\boldsymbol{r}})] = \epsilon_X[\rho(\vec{\boldsymbol{r}})] + \epsilon_C[\rho(\vec{\boldsymbol{r}})] \tag{3.36}$$

The exchange part was expressed analytically by Dirac [76]

$$\epsilon_x[\rho(\vec{\boldsymbol{r}})] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} \rho(\vec{\boldsymbol{r}}) \tag{3.37}$$

while the correlation part has been found numerically by Ceperley and Alder [77] using a stochastic quantum Monte Carlo method [78]. Later, an accurate parametrization of this data was published Perdew and Zunger (LDA-PZ) which is still used in DFT calculations [79]. LDA was expected to be best for solids with slowly varying densities like a nearly-free-electron metals and worst for inhomogeneous systems such as atoms where the density must go continuously to zero just outside the atom. The partial success of LDA in inhomogeneous systems is due to systematic error cancellation in which the correlation is underestimated but the exchange is overestimated resulting to a good value of E_{xc}^{LDA} [80, 81]. However, LDA tends to overestimate cohesive energies and binding energies for metals and insulators [82–84]. Errors in LDA are severely exaggerated for weakly bonded systems such as van der Waals and H-bond systems [85–87]. Nevertheless, LDA is fairly accurate in predicting elastic properties, such as bulk modulus [88, 89].

3.4.2 Generalized Gradient Approximation (GGA)

Attempts to improve the shortcomings of LDA has led to the use of gradient corrections. These so called Generalized Gradient Approximations (GGA) systematically calculate gradient corrections of the form $|\boldsymbol{\nabla}\rho(\vec{\boldsymbol{r}})|$, $|\boldsymbol{\nabla}\rho(\vec{\boldsymbol{r}})|^2$, $|\nabla^2\rho(\vec{\boldsymbol{r}})|$, etc. to the LDA. Such functionals can be generalized as

$$E_{xc}^{\mathrm{GGA}}[\rho] = \int f^{GGA}[\rho(\vec{\boldsymbol{r}}), \boldsymbol{\nabla}\rho(\vec{\boldsymbol{r}})]\, \mathrm{d}^3 r \tag{3.38}$$

where f^{GGA} is some arbitrary function of electron density and its gradient. GGA functionals are often term as semi-local because of their $\boldsymbol{\nabla}\rho(\vec{\boldsymbol{r}})$ dependence. Because of the flexibility in choosing f^{GGA}, a plethora of functionals have been developed and depending on the system under study, various results can be obtained. A more specific form of the GGA functional can be written as [83]

$$E_{xc}^{\mathrm{GGA}}[\rho] = \int \rho(\vec{\boldsymbol{r}})\, \epsilon_{xc}[\rho(\vec{\boldsymbol{r}})]\, F_{xc}[s]\, \mathrm{d}^3 r \tag{3.39}$$

where $\epsilon_{xc}[\rho(\vec{\boldsymbol{r}})]$ is the exchange-correlation energy per particle of an electron gas in a uniform electron density $\rho(\vec{\boldsymbol{r}})$ (i.e. similar to LDA). F_{xc} is the enhancement factor that tells how much XC energy is enhanced over its LDA value for a given $\rho(\vec{\boldsymbol{r}})$. Note the resemblance of GGA functional in (3.39) to the LDA functional in (3.35) which differ only by an enhancement factor. Here s is a dimensionless reduced gradient

$$s = \frac{|\boldsymbol{\nabla}\rho(\vec{\boldsymbol{r}})|}{2(3\pi^2)^{1/3}\rho(\vec{\boldsymbol{r}})^{4/3}} \tag{3.40}$$

The most popular GGA functionals used in the literature are Perdew-Burke-Ernzerhof (PBE) [90], PBEsol [91],Becke88 (B88) [92], Perdew-Wang (PW91) [93], Lee-Yang-Parr (LYP) [94], OptX (O) [95] and Xu (X) [96]. Among the functionals, PBE is the simplest and has exchange enhancement factor of the form

$$F_x^{\mathrm{PBE}}(s) = 1 + \kappa - \frac{\kappa}{1 + \mu s^2/\kappa} \tag{3.41}$$

where κ and μ are parameters obtained from physical constraints. When the density gradient approaches to zero ($|\boldsymbol{\nabla}\rho(\vec{\boldsymbol{r}})| \to 0, s \to 0$), $F_{xc}^{PBE}(s)$ will become unity and (3.39) reduces to LDA formulation. The form of the correlation functional is a complicated function of s and its discussion is beyond the scope of this thesis.

GGA functionals retained most of the correct features of LDA with much greater accuracy [97]. In addition, GGA tends to give better total energies, atomization energies, and energy barriers [92, 98–100]. However, GGA-based schemes typically fail on the region of weak interatomic interactions such as weak Hydrogen bonds, van der Waals interaction, and charge-transfer complexes [101–103].

3.5 Corrections to DFT

One important limitation of DFT that matters most in solid-state physics is the underestimation of band gap in semiconductors. A good theory must successfully predict the properties of wide range of materials including those novel materials as it is critical for applications in optoelectronics and nanotechnology. Hence, this section will discuss the inherent band gap problem and existing methods on improving the band gap.

3.5.1 Band Gap Problem

Given a set of eigenvalues, the band gap E_g is the difference in energy between the lowest unoccupied and highest occupied states

$$E_g^{KS} = \epsilon_{\text{CBM}} - \epsilon_{\text{VBM}} \tag{3.42}$$

where the CBM and VBM refer to conduction band minimum and valance band maximum, respectively. Note that E_g^{KS} is obtained from the calculation of Kohn-Sham band structure. The CBM and VBM can be approximated as

$$\epsilon_{\text{CBM}} \approx E_{N+1} - E_N \tag{3.43}$$

$$\epsilon_{\text{VBM}} \approx E_N - E_{N-1} \tag{3.44}$$

where E_N and $E_{N\pm1}$ are the ground-state total energies of the neutral system and with one electron added or removed, respectively. By combining equations (3.42)-(3.44), the band gap can be calculated as [104]

$$\begin{aligned} E_g^{KS} &\approx (E_{N-1} - E_N) - (E_N - E_{N+1}) \\ &\approx I - A \end{aligned} \tag{3.45}$$

The first term is precisely the ionization energy. In the case of solids, the same quantity is referred as work function which can be measured directly from photoelectron spectroscopy (PES) experiments. The second term is the electron affinity and can similarly be measured.

The origin of the band gap problem stems from the fact that (3.45) is an approximation to the 'quasiparticle gap' or 'electrical gap' [105]

$$E_g^{qp} = (E_{N-1} - E_N) - (E_N - E_{N+1}) \tag{3.46}$$

For the case of molecules and atoms, the calculation of total energies under DFT in the neutral state (E_N), cationic state (E_{N-1}), and anionic state (E_{N+1}) is possible.

Therefore, E_g^{qp} can be calculated directly from differences in total energies without invoking to Kohn-Sham eigenvalues. However, in the case of extended systems such as solids, the change in electron density upon the addition or removable of one electron is extremely small ($\Delta\rho \sim 10^{-20}\rho$). By taking the limit $\Delta\rho \to 0$, it can be shown that [106, 107]

$$\lim_{\Delta\rho\to 0} E_g^{qp} = E_g^{KS} + \Delta_{xc} \tag{3.47}$$

where the correction factor Δ_{xc} is given by

$$\Delta_{xc} = \lim_{\Delta\rho\to 0} V_{xc}[\rho + \Delta\rho] - V_{xc}[\rho - \Delta\rho] \tag{3.48}$$

This implies that quasiparticle gap and the Kohn–Sham band gap differs by a constant, Δ_{xc}. This also means that Δ_{xc} must not be zero, suggesting Δ_{xc} has a discontinuity at the specified limit [108, 109]. The problem with this formulation is that the exact exchange-correlation functional is not yet known. If LDA or GGA functional is used instead, V_{xc} will be a continuous function by construction and therefore there is no discontinuity (i.e. $\Delta_{xc} = 0$). The band gap problem of DFT is a result of the Kohn-Sham formulation of DFT, and in particular to the approximations made in exchange-correlation functional.

3.5.2 GW Approximation

The most suitable method for studying single particle excitation spectra such as ionization energies and electron affinities of extended systems is the Green's function. The Green's function relies on the calculation of the self-energy operator which is non-local, energy dependent, and non-Hermitian [110]. The self-energy is best approximated by the so called quasiparticle *GW* approximation, after the pioneering works of Hedin and Lundqvist [111, 112]. *GW* stands for the single-particle Green's function (G) and the dynamically screened Coulomb interaction (W). In practice, the exchange-correlation functional is replaced by the self-energy Σ [113]

$$\hat{\mathcal{V}}_{xc}\phi_i(\vec{\boldsymbol{r}}) \to \int \Sigma(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}', \epsilon_i)\phi_i(\vec{\boldsymbol{r}}')\ \mathrm{d}r' \tag{3.49}$$

so that the Kohn-Sham equation in (3.34) is modified as

$$(\hat{\mathcal{T}}_{KS} + \hat{\mathcal{V}}_{ext} + \hat{\mathcal{V}}_H)\phi_i(\vec{\boldsymbol{r}}) + \int \Sigma(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}', \epsilon_i)\phi_i(\vec{\boldsymbol{r}}')\ \mathrm{d}r' = \epsilon_i\phi_i(\vec{\boldsymbol{r}}) \tag{3.50}$$

The GW approximation for Σ is [114]

$$\Sigma(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}', \omega) = \frac{i}{4\pi} \int G(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}', \omega + \omega') W(\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}', \omega') \, d\omega' \tag{3.51}$$

where ω is the angular frequency related to energy as $\epsilon = \hbar\omega$. The precise meaning of G and W can be found in the seminal work of Hedin and Lundqvist [112] which involves the use of six coupled equations that are solved self-consistently. The self-energy Σ takes into account the finite discontinuity of Δ_{xc} in (3.48), thus yielding the correct quasiparticle band gap. Figure 3.5 illustrates the effectiveness of GW Approximation in improving the band gaps of semiconductors. Clearly, the band gaps calculated using LDA are greatly underestimated. The price to pay in using GW Approximation is that such calculations are considerably more computationally expensive, due partly to complications in convergence of total energies and unfavorable scaling with respect to the system size [115–117].

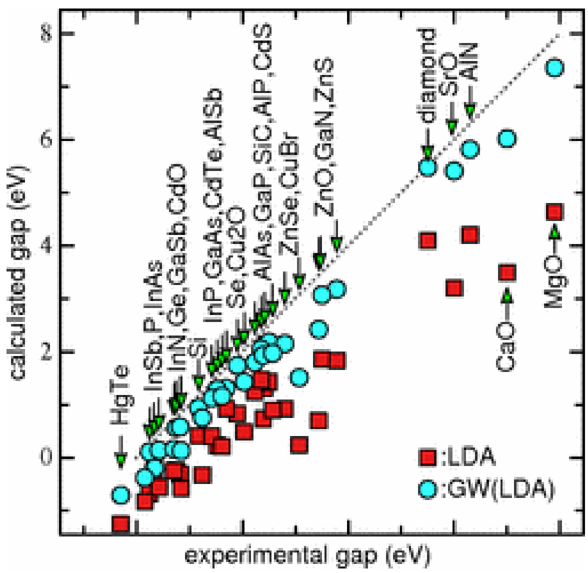

Figure 3.5: Improvement of band gap of semiconductors under GW Approximation. Squares correspond to band gaps calculated using LDA while circles correspond to GW Approximation. If the data point is below the dotted line, the calculated band gap is underestimated. Otherwise, it is overestimated. Illustration taken from [118].

3.5.3 Hybrid Functionals

Hybrid functional admixes a fixed amount of non-local Hartree-Fock exchange with the local or semi-local DFT exchange. The exchange functional in standard DFT was only approximated under LDA and GGA functional (i.e. eqtn (3.37)). However, the

HF exchange is given in exact form in eqtn (3.26). The simplest hybrid functional is the linear combination of the two exchange

$$E_{xc}^{\text{hybrid}} = a_0 E_x^{\text{HF}} + (1 - a_0) E_x^{\text{DFT}} + E_c, \quad 0 \leq a_0 \leq 1 \tag{3.52}$$

The hybrid functional used by Becke [119] has the form $E_x^{\text{HF}} = E_x^{\text{DFT}}$ with $a_0 = 0.5$ using LDA formulation. The PBE0 [120, 121] hybrid functional is constructed by a rational mixing of 25% HF exchange and 75% PBE exchange, with 100% PBE correlation having the form

$$E_{xc}^{\text{PBE0}} = \frac{1}{4} E_x^{\text{HF}} + \frac{3}{4} E_x^{\text{PBE}} + E_c^{\text{PBE}} \tag{3.53}$$

By far the most commonly used functional is the B3LYP (Becke,3-parameter,Lee,Yang,Par) [92, 94] hybrid functional which has the form [122]

$$E_{xc}^{\text{B3LYP}} = E_x^{\text{LDA}} + a_0 (E_x^{\text{HF}} - E_x^{\text{LDA}}) + a_x (E_x^{\text{B88}} - E_x^{\text{LDA}}) + E_c^{\text{LDA}} + a_c (E_c^{\text{LYP}} - E_c^{\text{LDA}}) \tag{3.54}$$

where $a_0 = 0.20$,$a_x = 0.72$, and $a_c = 0.81$. The parameters were determined by fitting to a data of measured atomization energies [120].

A critical feature of Hartree-Fock exchange is that it is nonlocal, that is, it cannot be evaluated at one particular spatial location, unless the electron density is known for all locations. Introducing nonlocality greatly increases the computational cost in solving Kohn-Sham equation. These type of functional are very difficult to apply in bulk and spatially extended systems. As a result, HF exact exchange find almost its use in quantum chemistry calculations involving molecules. However, progress is being done in developing the screened hybrid functionals in which the exchange interaction is split into two regions, a long-range (i.e. interstitial region) and a short-range (i.e. core region) interaction. The HF exchange is only incorporated to the short-range portion while standard DFT exchange acts on all portion. The Heyd, Scuseria, and Ernzerhof (HSE) functional is based on this approach which is calculated as [123, 124]

$$E_{xc}^{\text{HSE}} = \frac{1}{4} E_x^{\text{HF,SR}}(\omega) + \frac{3}{4} E_x^{\text{PBE,SR}}(\omega) + E_x^{\text{PBE,LR}}(\omega) + E_c^{\text{PBE}} \tag{3.55}$$

where the screening parameter ω defines the separation range, SR and LR refer to short range and long range, respectively.

3.5.4 Meta-GGA

Meta-GGA is an extension of the GGA in which the local kinetic energy density is included in the input to the functional. The GGA exchange-correlation functional in

(3.39) is modified to include the non-interacting kinetic energy density τ [82]

$$E_{xc}^{\mathrm{MGGA}}[\rho] = \int \rho(\vec{\boldsymbol{r}})\, \epsilon_{xc}[\rho(\vec{\boldsymbol{r}})]\, F_{xc}[\rho(\vec{\boldsymbol{r}}), \boldsymbol{\nabla}\rho(\vec{\boldsymbol{r}}), \tau(\vec{\boldsymbol{r}})]\ \mathrm{d}^3 r \tag{3.56}$$

where $\tau(\vec{\boldsymbol{r}})$ is defined as

$$\tau(\vec{\boldsymbol{r}}) = \frac{1}{2}\sum_{i=1}^{N} |\boldsymbol{\nabla}\phi_i(\vec{\boldsymbol{r}})|^2 \tag{3.57}$$

The implementation of Tao, Perdew, Staroverov, and Scuseria (TPSS) functional is based on meta-GGA functional [125]. Its exchange enhancement factor has a similar form as the PBE-GGA functional in (3.41) [82]

$$F_x^{\mathrm{TPSS}}(s) = 1 + \kappa - \frac{\kappa}{1 + \chi/\kappa} \tag{3.58}$$

where χ is a complicated function of $\rho(\vec{\boldsymbol{r}})$, $\boldsymbol{\nabla}\rho(\vec{\boldsymbol{r}})$, and $\tau(\vec{\boldsymbol{r}})$. Other meta-GGAs have also been proposed recently such as Tran, Blaha-modified Becke, Johnson (TB-mBJ) functional [126]; Perdew, Kurth, Zupan, and Blaha functional[127]; and Strongly Constrained and Appropriately Normed Density functional (SCAN) [128].

Since there are kinetic energy density corrections incorporated in the functional, the accuracy of meta-GGAs can compete with the computationally expensive hybrid or GW calculations. It is as cheap as the standard DFT such as LDA or GGA and hence can be scaled to large systems efficiently [126]. It also improves the band gaps of various insulators, semiconductors, oxides and halides but it fails sometimes on materials containing d and f orbitals [129, 130].

3.5.5 Hubbard-U Correction

One of the corrective approach used in the DFT electronic band gap problem is the DFT+U method. One pertinent problem in DFT is the description of strongly correlated systems in which the exchange-correlation functional tends to over-delocalize valence electrons. This problem is more pronounced to systems whose ground state energies are characterized by localized valence electrons such as d orbitals, f orbitals, and Mott insulators [131]. The inability of XC functionals to fully cancel the self-interaction in the Hartree term leads to an excessive delocalization, hence, creating a larger dispersion (i.e. larger valence bandwidth) and smaller band gap than what expected. The strong correlation stems from the Coulomb repulsion between electrons that forces them to localize [132]. This Coulomb potential is described by the term "U". Various models have been proposed to treat correlated systems, and one of the simplest is the "Hubbard-U" model which takes into account the "on-site" repulsion of electrons at the same atomic orbitals [133]. The total energy can be written as [134–136]

$$E_{\text{DFT}+U}[\rho(\vec{\boldsymbol{r}})] = E_{\text{DFT}}[\rho(\vec{\boldsymbol{r}})] + E_{\text{HUB}}[n^I_{mm'}] - E_{\text{dc}}[n^I] \tag{3.59}$$

where $E_{\text{DFT}}[\rho(\vec{\boldsymbol{r}})]$ is the eigenvalue of the Kohn-Sham equation in (3.34), E_{HUB} is the energy of Hubbard functional that describes the correlated systems, and E_{dc} is the double-counting correction when treating electronic interactions as a mean field. Based on this formulation, DFT+U energy can be expanded as

$$E_{\text{DFT}+U}[\rho(\vec{\boldsymbol{r}})] = E_{\text{DFT}}[\rho(\vec{\boldsymbol{r}})] + \sum_I \left[\frac{U^I}{2} \sum_{m \neq m'} n^I_m n^I_{m'} - \frac{U^I}{2} n^I (n^I - 1) \right] \tag{3.60}$$

where n^I_m are the occupation numbers of localized orbitals identified by the atomic site index I and state index m (magnetic quantum number for a particular angular quantum number l). The occupations are computed from the projection of Kohn-Sham orbitals onto the localized basis set such as the atomic orbital states [137]

$$n^I_{mm'} = \sum_{k,i} f_{ki} \left\langle \phi_{ki} \middle| \psi^I_m \right\rangle \left\langle \psi^I_{m'} \middle| \phi_{ki} \right\rangle \tag{3.61}$$

where ϕ_{ki} are the Kohn-Sham states (labeled by k-point k and band index i), f_{ki} represents their occupations according to Fermi-Dirac distribution, and ψ^I_m are the atomic orbitals. In some published works, the on-site Coulomb term U is replaced by an effective potential $U_{eff} = U + J$, where J is the site exchange term that accounts for Hund's rule coupling [134, 135]. The effective potential is proved to be crucial in describing strong spin-orbit coupling.

Hubbard-U calculations depend on the values of U, where it can be either formulated from first principles or achieved empirically by tuning it such that it agrees with the experimental results. The former can be achieved through linear response method in which the response of the localized states to a small perturbation is calculated [136, 138]. The latter is usually compared to the experimental band gap. Nevertheless, the empirical tuning is much preferred because of the significant computational cost in doing linear response calculations, and also the calculated U is not necessarily better than the empirical one [139]. As stated earlier, Hubbard-U calculation can be used to correct the band gaps of strongly correlated systems. This is applicable to semiconductors containing of d and f orbitals such as ZnO, CeO_2, TiO_2, etc. Figure 3.6 shows the improvement of band gaps of transition-metal oxides using

Hubbard-U correction. Note that the band gap of MnO is underestimated while FeO is incorrectly predicted as metallic when LDA is used.

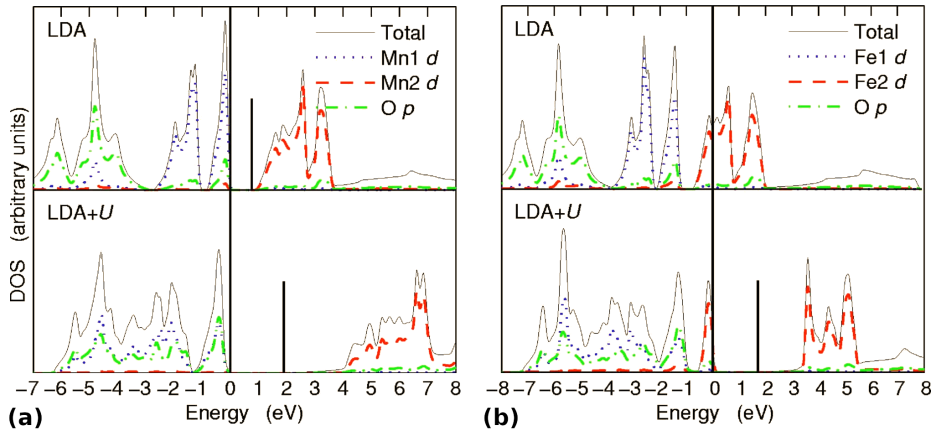

Figure 3.6: Comparison of density of states calculated by LDA and LDA+U for (a) MnO and (b) FeO. The solid vertical bars indicate the end of fundamental band gap. Fermi energy is set at 0 eV. Illustration taken from [140].

Chapter 4

DFT Calculation of Solids

4.1 Basis Sets

Solving the Kohn-Sham equation in (3.34) requires the use of mathematical representations to describe the single-particle Kohn-Sham orbitals $\phi_i(\vec{\boldsymbol{r}})$. One possibility is to express these orbitals as a basis sets that are known and numerically solvable. One starts by expressing orbitals as a linear combination of generic basis set

$$\phi_i(\vec{\boldsymbol{r}}) = \sum_{\alpha}^{M} c_{\alpha}^{i} \left|\chi_{\alpha}\right\rangle \tag{4.1}$$

where i is the band index, the sum runs over all the basis functions up to the dimension M, and c_{α}^{i} is the expansion coefficient of a known basis function $|\chi_{\alpha}\rangle$. Since $\phi_i(\vec{\boldsymbol{r}})$ spans the whole infinite space, M must be in principle infinite. However, in practice the basis set is truncated just enough for an accurate description of the orbital. The choice of the basis set depends on several factors such as (a) efficiency and (b) unbiased [141]. A basis set is efficient if it resembles $\phi_i(\vec{\boldsymbol{r}})$ closely, hence requiring less expansion coefficients and smaller dimension size. However, this assumes that the solution to the problem must be known beforehand. Such basis set can never be general because it will quickly yield a solution for a specific problem but will poorly perform for other cases. The problem is that optimizing a basis set for a specific system can cause bias. This means that if a property of a system is calculated, but the basis set is optimized for only one particular system, the result will be biased towards that one [142, 143]. It is the goal of theoretical condensed matter physics to find a basis set that is simultaneously efficient and unbiased. There are three types of basis sets that are commonly used for expansions, namely: local, nonlocal, and augmented basis.

4.1.1 Local Basis Set

A local basis has its peak centered on a local point and is well applicable to orbitals around individual atoms in real space. Gaussian basis sets or any atom-centered basis orbitals are examples of this type. It is a popular choice for atoms and molecules whose orbitals are highly localized around each atom. Hence, less than 20 basis functions per atom are sufficient enough to achieve acceptable accuracy. As an example, the Slater type basis orbitals (STO) are written as [144]

$$\left|\chi_{\alpha}^{\mathrm{STO}}\right\rangle = Ae^{-\alpha r} \tag{4.2}$$

where A is some normalization constant. Note that STO exponentially decays away from an atom centered at $\vec{\boldsymbol{r}}$. On the other hand, Gaussian type basis orbitals (GTO) are written as [145]

$$\left|\chi_{\alpha}^{\mathrm{GTO}}\right\rangle = Ae^{-\alpha r^2} \tag{4.3}$$

GTO has the advantage that all integrals associated with it can be performed analytically. Since these basis sets are localized, they cannot properly described the long-range interaction of metals and the periodicity of crystalline solids.

4.1.2 Nonlocal Basis Set

Nonlocal basis set span the whole space. An important class of basis orbital under this category is the plane wave basis (PW) described as

$$\left|\chi_{\alpha}^{\mathrm{PW}}\right\rangle = Ae^{i\alpha r} \tag{4.4}$$

which can be generalized as a dot product of wavevector $\vec{\mathbf{k}}$ and position vector $\vec{\mathbf{r}}$

$$\left|\chi_{\vec{\mathbf{k}}}^{\mathrm{PW}}\right\rangle = Ae^{i\vec{\mathbf{k}}\cdot\vec{\mathbf{r}}} \tag{4.5}$$

PWs are the most commonly used in DFT because of the following reasons: PWs are already solutions to periodic systems satisfying the Bloch condition; PWs are convenient in taking gradients and integrals because of their exponential form; changing the domain of PWs from real space to reciprocal space are easily executed using Fourier transformation; PWs are orthogonal which simplifies calculation; and lastly, PWs are independent of the atomic positions because of its nonlocal nature. However, there are also disadvantages of using it. It requires an enormous amount of PWs to properly describe the rapid fluctuation of orbital wavefunctions near the core region of an atom or ion. A direct fix to this problem is the application of pseudopotentials to smoothen the strong Coulomb potential of the nucleus, which will be the topic of later section [146].

4.1.3 Augmented Basis Set

Augmented basis sets are combinations of local and nonlocal basis sets. Under this category is the Augmented plane waves (APW) basis. The APW divides the space into two regions: the core region,where the orbitals are atomic-like; and the interstital region, where the orbitals resemble plane waves [147]. The basis orbitals are taken to be [148]

$$\left|\chi_{\vec{\mathbf{k}}}^{\mathrm{APW}}\right\rangle = \begin{cases} \text{atomic basis} & , \left|\vec{\mathbf{r}} - \vec{\mathbf{R}}\right| \leq r_c \\ Ae^{i\vec{\mathbf{k}}\cdot\vec{\mathbf{r}}} & , \left|\vec{\mathbf{r}} - \vec{\mathbf{R}}\right| > r_c \end{cases} \tag{4.6}$$

where $\vec{\mathbf{R}}$ is the center of the atom and r_c is the core radius. Outside the core, the orbital wavefunction is a plane wave because the potential is constant there. Inside the core, the orbital wavefunction is atomic-like and can be solved by the appropriate Schrödinger equation. The potential involved in this type of basis is usually called muffin-tin potential due its resemblance to muffin tins. APWs must satisfy the boundary conditions at $\left|\vec{\mathbf{r}} - \vec{\mathbf{R}}\right| = r_c$. That is, the basis orbital must be continuous at the boundary value and its slope exists [149]. Augmented plane waves are very accurate because it describes both core electrons and valence electrons well. However, accuracy is always associated with computational costs.

4.2 Matrix Formulation for KS equation

The use of basis sets transforms the Kohn-Sham equation into an ordinary matrix algebra that can be solved numerically. The Kohn-Sham equation in (3.34) is expanded in terms of basis sets using (4.1)

$$\hat{\mathcal{H}}_{KS} \sum_{\alpha}^{M} c_{\alpha}^{i} \left|\chi_{\alpha}\right\rangle = \epsilon_i \sum_{\alpha}^{M} c_{\alpha}^{i} \left|\chi_{\alpha}\right\rangle \tag{4.7}$$

Left "multiply" with $\langle\chi_\beta|$:

$$\sum_{\alpha}^{M} \left\langle\chi_\beta\right| \hat{\mathcal{H}}_{KS} \left|\chi_\alpha\right\rangle c_{\alpha}^{i} = \sum_{\alpha}^{M} \left\langle\chi_\beta|\chi_\alpha\right\rangle c_{\alpha}^{i}\epsilon_i \tag{4.8}$$

which can be simplified as

$$\sum_{\alpha}^{M} H_{\beta\alpha} c_{\alpha}^{i} = \sum_{\alpha}^{M} S_{\beta\alpha} c_{\alpha}^{i} \epsilon_i \tag{4.9}$$

where $H_{\beta\alpha}$ and $S_{\beta\alpha}$ are the energy-independent Hamiltionian and the overlap matrix, respectively [150]. The elements of these matrices are defined as

$$H_{\beta\alpha} = \langle\chi_\beta|\, \hat{\mathcal{H}}_{KS}\, |\chi_\alpha\rangle = \int \chi_\beta^*(\vec{\boldsymbol{r}}) \hat{\mathcal{H}}_{KS} \chi_\alpha(\vec{\boldsymbol{r}})\, \mathrm{d}^3 r \tag{4.10}$$

$$S_{\beta\alpha} = \langle\chi_\beta|\chi_\alpha\rangle = \int \chi_\beta^*(\vec{\boldsymbol{r}}) \chi_\alpha(\vec{\boldsymbol{r}})\, \mathrm{d}^3 r \tag{4.11}$$

The overlap matrix $S_{\beta\alpha}$ takes into account the possible non-orthogonality of the basis functions [72]. Note that for plane wave (PW) basis sets, which are orthonormal, the overlap matrix $S_{\beta\alpha}$ becomes an unit matrix. The general matrix eigenvalue problem can be recast into a compact form [151]

$$\boldsymbol{Hc} = \boldsymbol{Sc\Lambda} \tag{4.12}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix containing energy eigenvalues and $\boldsymbol{c}$ has the eigenfunction (expansion coefficients of the KS orbital) as columns. In solving (4.12), the normalization condition must be taken into account

$$\int \phi_i(\vec{\boldsymbol{r}}) \phi_i^*(\vec{\boldsymbol{r}}) = \int \sum_{\alpha}^{M} \sum_{\beta}^{M} c_\alpha^{i*} \chi_\alpha^* c_\beta^i \chi_\beta \, \mathrm{d}^3 r = 1 \tag{4.13}$$

$$= \sum_{\alpha}^{M} \sum_{\beta}^{M} c_\alpha^{i*} c_\beta^i \int \chi_\alpha^* \chi_\beta \; \mathrm{d}^3 r = 1 \tag{4.14}$$

$$= \sum_{\alpha}^{M} \sum_{\beta}^{M} c_\alpha^{i*} c_\beta^i S_{\alpha\beta} = 1 \tag{4.15}$$

There are $M \times N$ elements of $\boldsymbol{c}$ needed to be solved, where M is the total number of basis functions used and N is the total number of lowest-energy orbitals. In addition, there are N unknown energy eigenvalues to be solved. Fortunately, there are $M \times N$ independent equations in (4.12) and N equations coming from the normalization condition (4.15) so that $N(M+1)$ equations are simultaneously solved [141]. It is obvious that increasing either M or N will increase the computational power needed. This does not inlude yet the iterative self consistent field calculation, as shown in Figure 3.3, needed to have converged electron density. Common numerical algorithms used in matrix diagonalization are Davidson iterative diagonalization [152, 153], Residual minimization/direct inversion in the iterative subspace (RMM–DIIS) [154, 155], and Conjugate-gradient-like band-by-band diagonalization [156, 157]. Numerical algorithms must be efficient and optimized since most DFT codes spend substantial amount of time in matrix diagonalization.

4.3 Pseudopotential (PP) Approach

The idea behind the use of pseudopotentials is to replace the strong Coulomb potential of the nucleus by an effective potential acting on the valence electrons [158–160]. When atoms bond together to form a solid, the core electrons are so localized in a deep potential well that they remain invariant. Thus, their contribution to bonding is negligible and the replacement of its potential by a simple fictitious potential is justified. Figure 4.1 illustrates the action of pseudopotential on the wavefunction and potential of an atomic orbital. The all-electron wavefunction contains nodes which are computationally difficult to solve. On the other hand, pseudo wavefunction is nodeless everywhere, and therefore it greatly reduces the number of plane waves required for the calculation by a significant amount. Note that at large distances away from the nucleus, both potential becomes constant and the wavefunction is expected to be a plane wave. By effectively neglecting the core electrons from the calculation, the Kohn-Sham orbitals needed is dramatically reduced. This will substantially reduce the computational time required to calculate orbital-dependent quantities.

There are two criteria for choosing a good pseudopotential, namely: softness and transferability [162, 163]. A pseudopotential is soft if it requires few plane waves to model the system. This is similar to efficiency of basis sets. A pseudopotential is transferable if it can be used in whatever environment (e.g. molecule, solid, cluster, surface, metal, insulator, etc). The choice depends on which pseudopotential is advantageous to use and the type of calculation being done. The common pseudopotentials used in DFT codes are Norm-Conserving pseudopotential, Ultrasoft pseudopotential, and Projector Augmented Wave.

4.3.1 Norm-Conserving Pseudopotential (NCPP)

In norm-conserving pseudopotentials, the pseudopotential and all-electron charge densities are set equal so that the norm is conserved in both potentials [164, 165]. Pseu-

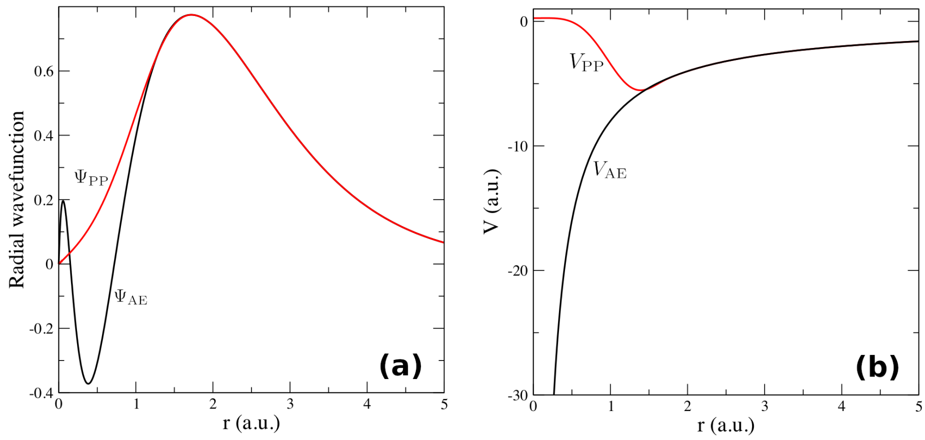

Figure 4.1: Schematic illustration of a (a) pseudo wavefunction pseudized from a 3s wavefunction of Si orbital and the (b) corresponding pseudo- (PP) and all-electron (AE) potentials. The all-electron approach takes into account all electrons including core and valence electrons. Illustration taken from [161].

dopotentials are generated to meet this criterion

$$\int_0^{r_c} |\phi_{\mathrm{PP}}(\vec{\boldsymbol{r}})|^2 \, \mathrm{d}\vec{\boldsymbol{r}} = \int_0^{r_c} |\phi_{\mathrm{AE}}(\vec{\boldsymbol{r}})|^2 \, \mathrm{d}\vec{\boldsymbol{r}} \tag{4.16}$$

where r_c is the chosen cutoff radius that separates the core region from the valence region. The constraint imposed on this pseudopotential leads to an improvement in transferability of potentials to different chemical environments. In addition, reducing r_c improves the transferability because in this way the pseudo wavefunction becomes closer to the all-electron result. However, the cutoff radius should be chosen outside the location of the maximum node of the all-electron wavefunction. Note that this pseudopotential gives only the valence charge density and not the total charge density. Other norm-conserving schemes were proposed by Troullier and Martin (TM) [165], and by Rappe, Rabe, Kaxiras, and Joannopoulos (RRKJ) [166].

4.3.2 Ultrasoft Pseudopotential (USPP)

Ultrasoft pseudopotentials were introduced in order for the calculations to have lowest numbers of plane waves basis set used since it was shown that the norm of the all-electron and pseudo wavefunction was not necessary requirement for transferability. Hence, this was done by Vanderbilt [167] who showed that smoother but highly transferable pseudopotentials are possible. The cutoff radius r_c is situated farther than the equivalent norm-conserving pseudopotential and the pseudo wavefunction is flatter. This leads to fewer plane waves that gives significant reduction in computational time.

Similar to norm-conserving pseudopotentials, the ultrasoft pseudopotential only gives valence charge densities, not total charge densities.

4.3.3 Projector Augmented Wave (PAW)

The Projector Augmented Wave takes into account both all-electron and pseudo wavefunction into calculations. It aims for both efficiency of using pseudopotential and the accuracy of using all-electron potential [168, 169]. However, the all-electron wavefunction is limited only on the core region and will be truncated beyond the cut-off radius r_c. A correction factor is added to subtract the overlapping part of the pseudo wavefunction in the core region [170]. Hence, the PAW wavefunction involves three terms

$$\psi_{\mathrm{PAW}} = \psi_{\mathrm{AE}} + \psi_{\mathrm{PP}} - \psi_{\mathrm{net}} \tag{4.17}$$

The actions of the terms in the equation above are visualized in Figure 4.2. The ψ_{PP} is expanded in plane wave basis sets while ψ_{AE} is only defined within the cutoff radius r_c. The ψ_{net} subtracts the overlapping part of ψ_{PP} in the core region.

Note that in Kohn-Sham formulation, these wavefunctions ψ become the independent Kohn-Sham orbitals ϕ. PAW calculations are accurate as all-electron calculations with much less computational effort. Unlike the two pseudopotentials mentioned before, PAW pseudopotential returns both the core and valence charge densities.

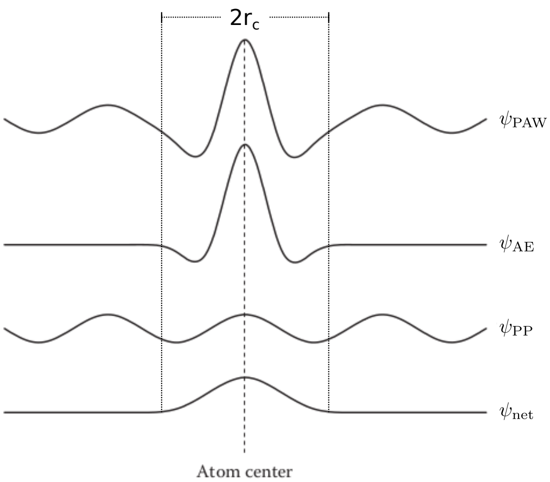

Figure 4.2: Schematic illustration of the wavefunctions used in PAW pseudopotential. Illustration taken from [62].

4.4 Supercells

Most solids are characterized by its regular repeating three-dimensional structure called a crystal lattice. Hence, it is possible to study solids by just looking at the building block, which is ordinarily called the unit cell. In order to model solids that are feasible for computational simulation, repeating unit cells that are stack together must be needed. These stacked unit cells are collectively called supercell. When implementing DFT, the periodic boundary conditions must be taken into account. In this case, the supercell is duplicated periodically throughout the whole space. However, the actual calculation is applied only on a single supercell while the rest (called images) simply copies it with no significant computational cost.

When defects are introduced into the supercell, it forms a periodic array of defects across all the images of the supercell. A supercell must be large enough so that the calculation is independent of the location of the defect inside the supercell and also to reduce the interaction with its images. Thus, the Kohn-Sham equation and other pertinent calculations are solved only within a single supercell.

4.5 DFT Calculation in Reciprocal Space

Working in reciprocal space is greatly convenient if functions are expressed in terms of plane waves. Plane waves propagate in real space but they become point in the reciprocal space, wherein each point corresponds to a particular wavevector $\vec{\mathbf{k}}$. The lattice points in real space will define the allowed wavevectors $\vec{\mathbf{G}}$, the reciprocal lattice vector which is subset of the reciprocal space. See Appendix **??** for discussions about reciprocal lattice and Brillouin zones. Since real space and reciprocal space have inverse relationship, increasing the supercell size by a certain factor will cause the supercell in reciprocal space to shrink by a same factor, and vice versa. Figure 4.3 illustrates this relationship. Note that no information is lost when transforming between the two spaces. In addition, bigger supercells require fewer k-points but in the expense of many atoms included in calculations.

In previous discussions, $\vec{\mathbf{k}}$ was defined as any wavevector in the reciprocal space. However, it can always be transform to $\vec{\mathbf{k}} \rightarrow \vec{\mathbf{k}} + \vec{\mathbf{G}}$ so that the new $\vec{\mathbf{k}}$ is in the first Brillouin zone and that any wavevectors are equivalent to the new one by a reciprocal lattice vector $\vec{\mathbf{G}}$. This transformation will limit DFT calculations inside the first

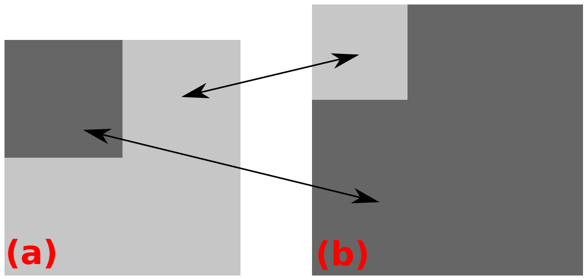

Figure 4.3: Relationship between a supercell in (a) real space and the corresponding (b) reciprocal space.

Brillouin zone instead of the whole reciprocal space.

Furthermore, one can take advantage of the symmetry of the solid to reduce the first Brillouin zone into what is called irreducible Brillouin zone (IBZ). Hence, DFT calculations will be narrowed down further into this Brillouin zone. Note that each k-points will have a weight factor that depends on how many times it was folded during symmetry operations (e.g. rotation and inversion). Figure 4.4 summarizes the various techniques employed in simplifying DFT calculations in solids starting from a bulk solid, then a supercell simulation, transformed to reciprocal space confined in first Brillouin zone, and further reduced to a irreducible zone by symmetry operations. All these techniques and the pseudopotential approximations made DFT calculations computationally feasible.

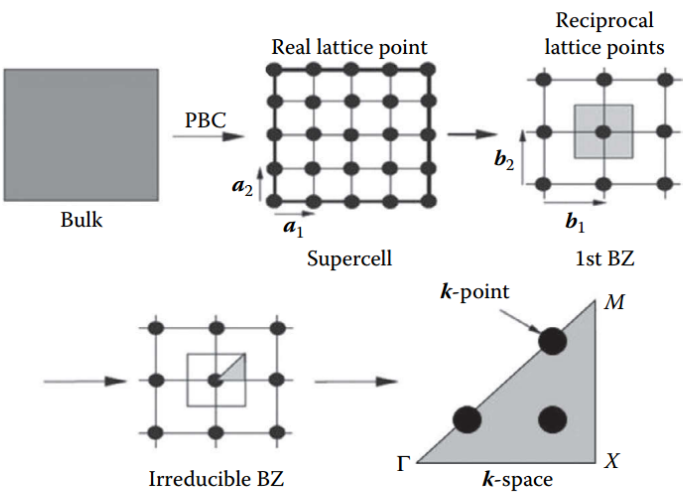

Figure 4.4: Various techniques used in treating solids in DFT calculations. Figure taken from [62].

4.6 k-point Sampling

It was shown that DFT calculations can be solved within the irreducible Brillouin zone. However, there are infinite numbers of k-points inside IBZ that are well qualified for a plane wave. One way to deal with this problem is to sample finite number of k-points that represent each region well. This sampling technique is justified by the fact that orbitals and other quantities vary smoothly in the IBZ. Sampling in IBZ must satisfy two goals: select few k-points as possible to reduce computational time, and select enough so that they represent the actual quantities well. Convergence tests in which the number of k-points are varied until quantities such as total energy does not change anymore must be conducted. The quantity is said to be converged with respect to k-points. Convergence tests are very helpful in finding the optimum number of k-points with minimum error. Any integrated function $f(\vec{\mathbf{r}})$ can be written over the Brillouin zone

$$f(\vec{\mathbf{r}}) = \int_{BZ} F(\vec{\mathbf{k}})\,\mathrm{d}\vec{\mathbf{k}} \tag{4.18}$$

where $F(\vec{\mathbf{k}})$ is the Fourier transform of $f(\vec{\mathbf{r}})$. To evaluate computationally, the integral can be approximated by weighted sum over special k-points

$$\int_{BZ} F(\vec{\mathbf{k}})\,\mathrm{d}\vec{\mathbf{k}} \approx \sum_j w_j F(\vec{\mathbf{k}}_j) \tag{4.19}$$

where w_j are the weighted factors. There are standard schemes for generating k-points grid mesh and probably the most popular one is the Monkhorst-Pack method.

4.6.1 Monkhorst-Pack method

The Monkhorst-Pack method generates a grid of uniform special k-points along the three lattice vectors in the reciprocal space [171]. The construction of the special points is based on the formula

$$u_r = \frac{2r - q_r - 1}{2q_r} \qquad , r = 1, \ldots, q_r \tag{4.20}$$

where q_r determines the number of k-points used along one of the axis in $r = x, y,$ or z, and r varies from 1 to q_r. The k-point is given by

$$\vec{\mathbf{k}} = u_x\vec{\mathbf{b}}_1 + u_y\vec{\mathbf{b}}_2 + u_z\vec{\mathbf{b}}_3 \tag{4.21}$$

wherein $\vec{\mathbf{b}}_1, \vec{\mathbf{b}}_2, \vec{\mathbf{b}}_3$ are the primitive lattice vectors of the reciprocal lattice. u_r gives the fractional part of the corresponding component of the reciprocal lattice vector.

For instance, in a $4 \times 4 \times 4$ grid this will correspond to $q_x = q_y = q_z = 4$ with a total number of $q_x \times q_y \times q_z = 64$ k-points. The number of k-points will be further reduced by symmetry operations inside the irreducible Brillouin zone.

The center of the mesh can be centered on the origin ($\vec{\mathbf{k}} = 0$ or Γ point) or shifted by a fixed amount away from the origin. The former is important if one needs to know the electronic states at the Γ point. The latter generally breaks the symmetry, hence, it must be use with care.

4.6.2 Gamma Point Sampling

For very large supercell, the associated first Brillouin zone are very small that it approaches the zone center or the Γ point. Thus, it would be practical to use only one k-point with high weight factor. Calculations based on the sampling at the Γ point reduce significant computational cost because the real and reciprocal space coincide with each other at the origin and the KS orbital will real quantity so that any consideration for complex numbers is not necessary. Γ point calculation are routinely used in massive calculations.

4.7 Bloch Representations in DFT

The Bloch expression for the Kohn-Sham orbital is similar to the many-particle wave-function derived from Schrödinger equation. See section 3.1.1 for the discussion about periodicity. Here, the Bloch form of KS orbital is

$$\phi_{n,k}(\vec{\mathbf{r}}) = u_{n,k}(\vec{\mathbf{r}}) e^{i\vec{\mathbf{k}}\cdot\vec{\mathbf{r}}} \tag{4.22}$$

where $\phi_{n,k}(\vec{\mathbf{r}})$ now depends both on the band index n and the wavevector $\vec{\mathbf{k}}$ confined in the first Brillouin zone. Expanding the periodic function $u_{n,k}$ in terms of plane waves whose wavevectors are reciprocal lattice vector $\vec{\mathbf{G}}$

$$u_{n,k}(\vec{\mathbf{r}}) = \sum_{\vec{G}} C_{n,k}(\vec{\mathbf{G}})\, e^{i\vec{\mathbf{G}}\cdot\vec{\mathbf{r}}} \tag{4.23}$$

where $C_{n,k}(\vec{\mathbf{G}})$ is the expansion coefficient of plane wave basis sets. The phase factor $\exp\Big(i\vec{\mathbf{G}}\cdot\vec{\mathbf{r}}\Big)$ represents a plane wave travelling in space, perpendicular to $\vec{\mathbf{G}}$. Thus, the KS orbital can be rewritten as

$$\phi_{n,k}(\vec{\mathbf{r}}) = \sum_{\vec{G}} C_{n,k}(\vec{\mathbf{G}})\, e^{i(\vec{\mathbf{k}}+\vec{\mathbf{G}})\cdot\vec{\mathbf{r}}} \tag{4.24}$$

The coefficients $C_{n,k}(\vec{\mathbf{G}})$ can be solved by taking the inverse Fourier transform of $\phi_{n,k}(\vec{\mathbf{r}})$

$$C_{n,k}(\vec{\mathbf{G}}) = \mathcal{F}^{-1}[\phi_{n,k}(\vec{\mathbf{r}})] \tag{4.25}$$
$$= \int \phi_{n,k}(\vec{\mathbf{r}}) e^{-i(\vec{\mathbf{k}}+\vec{\mathbf{G}})\cdot\vec{\mathbf{r}}}\,\mathrm{d}\vec{\boldsymbol{r}} \tag{4.26}$$
$$= \phi_{n,k}(\vec{\mathbf{G}}) \tag{4.27}$$

Similarly, the electron density in real space and reciprocal space are Fourier transform of each other

$$\rho(\vec{\mathbf{r}}) = \sum_{\vec{\boldsymbol{G}}} \rho(\vec{\mathbf{G}}) e^{i\vec{\mathbf{G}}\cdot\vec{\mathbf{r}}} \tag{4.28}$$
$$\rho(\vec{\mathbf{G}}) = \int \rho(\vec{\mathbf{r}}) e^{-i\vec{\mathbf{G}}\cdot\vec{\mathbf{r}}}\,\mathrm{d}\vec{\boldsymbol{r}} \tag{4.29}$$

4.8 Energy Operators in Reciprocal Space

Since DFT calculations take place in the reciprocal space, the Hamiltonian of the KS equation in (3.34) must be transform from the real space to reciprocal space. In the Kohn-Sham formulation, both non-interacting kinetic energy and Hartree potential are easily evaluated because they are local in reciprocal space. The action of the kinetic energy operator on the KS orbital is

$$\hat{\mathcal{T}}_{KS}\left|\phi_{n,k}(\vec{\mathbf{r}})\right\rangle = -\frac{1}{2}\nabla^2\left(\sum_{\vec{\boldsymbol{G}}} C_{n,k}(\vec{\mathbf{G}})\, e^{i(\vec{\mathbf{k}}+\vec{\mathbf{G}})\cdot\vec{\mathbf{r}}}\right) \tag{4.30}$$
$$= \frac{1}{2}\sum_{\vec{\boldsymbol{G}}}(\vec{\mathbf{k}}+\vec{\mathbf{G}})^2 C_{n,k}(\vec{\mathbf{G}})\, e^{i(\vec{\mathbf{k}}+\vec{\mathbf{G}})\cdot\vec{\mathbf{r}}} \tag{4.31}$$

Hence, the effect of the kinetic energy operator in the reciprocal space is to multiply each coefficient by one-half times the square of its wavevector. Its matrix representation is

$$\hat{\mathcal{T}}_{KS}(\vec{\mathbf{G}},\vec{\mathbf{G}}') = \left\langle\phi_{n,k}(\vec{\mathbf{r}})\right|\hat{\mathcal{T}}_{KS}\left|\phi_{n,k}(\vec{\mathbf{r}})\right\rangle \tag{4.32}$$
$$= \frac{1}{2}\left|\vec{\mathbf{k}}+\vec{\mathbf{G}}\right|^2\delta_{\vec{\mathbf{G}},\vec{\mathbf{G}}'} \tag{4.33}$$

In the above equation, the bra-term is expanded in terms of $\vec{\mathbf{G}}$ while the ket-term is expanded in $\vec{\mathbf{G}}'$. The Hartree potential in reciprocal space is given by

$$\hat{\mathcal{V}}_H = \frac{1}{2}\sum_{\vec{\boldsymbol{G}}}\left|\rho(\vec{\mathbf{G}})\right|^2 \tag{4.34}$$

which is more simple compared to its real space counterpart given in (3.21). Its matrix representation is given by

$$\hat{\mathcal{V}}_H(\vec{\mathbf{G}},\vec{\mathbf{G}}') = \langle\phi_{n,k}(\vec{\mathbf{r}})|\,\hat{\mathcal{V}}_H\,|\phi_{n,k}(\vec{\mathbf{r}})\rangle \tag{4.35}$$

$$= \hat{\mathcal{V}}_H(\vec{\mathbf{G}}-\vec{\mathbf{G}}') \tag{4.36}$$

The remaining external potential and exchange-correlation energy can be obtained from their Fourier transform, respectively

$$\hat{\mathcal{V}}_{ext}(\vec{\mathbf{G}}) = \int \hat{\mathcal{V}}_{ext}(\vec{\mathbf{r}})e^{-i\vec{\mathbf{G}}\cdot\vec{\mathbf{r}}}\,\mathrm{d}\vec{\boldsymbol{r}} \tag{4.37}$$

$$\hat{\mathcal{V}}_{xc}(\vec{\mathbf{G}}) = \int \hat{\mathcal{V}}_{xc}(\vec{\mathbf{r}})e^{-i\vec{\mathbf{G}}\cdot\vec{\mathbf{r}}}\,\mathrm{d}\vec{\boldsymbol{r}} \tag{4.38}$$

Their matrix representations are similar to (4.36). Taking all the derivations above, the complete Kohn-Sham equation in reciprocal space is

$$\sum_{\vec{G}'}\left[\frac{1}{2}\left|\vec{\mathbf{k}}+\vec{\mathbf{G}}\right|^2\delta_{\vec{\mathbf{G}},\vec{\mathbf{G}}'} + \hat{\mathcal{V}}_{ext}(\vec{\mathbf{G}}-\vec{\mathbf{G}}') + \hat{\mathcal{V}}_H(\vec{\mathbf{G}}-\vec{\mathbf{G}}') + \hat{\mathcal{V}}_{xc}(\vec{\mathbf{G}}-\vec{\mathbf{G}}')\right] C_{n,k}(\vec{\mathbf{G}}') = \epsilon_{n,k}\,C_{n,k}(\vec{\mathbf{G}}) \tag{4.39}$$

which can be rewritten as a general matrix equation

$$\hat{\mathcal{H}}_{\vec{\mathbf{G}},\vec{\mathbf{G}}'}C_{n,k}(\vec{\mathbf{G}}') = \epsilon_{n,k}\,C_{n,k}(\vec{\mathbf{G}}) \tag{4.40}$$

This is similar to (4.12) but the overlapping matrix is set to identity matrix.

4.9 Cutoff Energy

The plane wave basis expansion of both KS orbital and the electron density in (4.24) and (4.28) is evaluated in the complete set of reciprocal lattice vector $\vec{\mathbf{G}}$, which is infinite. This means that it will take infinitely long to compute desired properties. Nevertheless, orbitals and electron densities tend to become smoothly varying at large $\vec{\mathbf{G}}$-vectors. Thus, their plane wave components become negligible for large $\vec{\mathbf{G}}$. The expansion can be truncated by introducing kinetic energy cut-off E_{cut} defined as

$$E_{cut} = \left|\vec{\mathbf{k}}+\vec{\mathbf{G}}_{cut}\right|^2 \tag{4.41}$$

the $\vec{\mathbf{G}}_{cut}$ serves as the upper bound for the expansion series of KS orbital in (4.24). This means that plane waves whose kinetic energy is less than this cut-off energy are the only ones included in DFT calculations. The cut-off radius of electron densities is usually a multiple of E_{cut} quantified by

$$mE_{cut} = \left|\vec{\mathbf{k}} + \vec{\mathbf{G}}_{rho}\right|^2 \tag{4.42}$$

where m is called a dual, a multiplier of E_{cut}, and $\vec{\mathbf{G}}_{rho}$ serves as the upper bound for the expansion series of electron density in (4.28). The number of plane waves can be estimated as

$$N_{\mathrm{PW}} \approx \frac{1}{2\pi^2} V E_{cut}^{3/2} \tag{4.43}$$

where V is the volume of the supercell. The cut-off energy that is appropriate for a given calculation is not usually known in advance, as it varies on the configuration of the system. However, convergence tests can be conducted where the cut-off energies is increased until the desired properties stop changing. Note that wavevectors $\vec{\mathbf{k}}$ were discretized by using k-point sampling in the IBZ, $\vec{\mathbf{G}}$-vectors become finite by energy cutoff, and the band index n depends on the number of orbital states, which is also finite. Hence, solving Kohn-Sham in (4.39) becomes computationally tractable.

4.10 Ionic Relaxation

For every DFT calculations, the system must be fully relaxed both electronically and structurally. The electronic relaxation is given by the self-consistent field calculation in section 3.3.3. The structural relaxation or ionic relaxation, also known as geometric optimization, computes the forces of each atom and are moved to directions of minimum forces for the next electronic relaxation. The force on the Ith atom positioned at $\vec{\mathbf{R}}_I$ can be calculated from Hellman-Feynman theorem as [172, 173]

$$\vec{\mathbf{F}}_I = -\frac{\partial E}{\partial \vec{\mathbf{R}}_I} = -\langle\phi|\frac{\partial \hat{\mathcal{H}}}{\partial \vec{\mathbf{R}}_I}|\phi\rangle - \frac{\partial E_{II}}{\partial \vec{\mathbf{R}}_I} \tag{4.44}$$

$$= -\int \rho(\vec{\mathbf{r}}) \frac{\partial \hat{\mathcal{V}}_{ext}(\vec{\mathbf{r}})}{\partial \vec{\mathbf{R}}_I} \mathrm{d}\vec{\boldsymbol{r}} - \frac{\partial E_{II}}{\partial \vec{\mathbf{R}}_I} \tag{4.45}$$

Note that the calculation of forces is given strictly in terms of the electron density and the external potential, independent of electron kinetic energy, Hartree potential, and exchange-correlation terms. Thus, forces can be calculated by taking simple derivative operations on two potential terms. This is why force calculations are very fast that they are almost unnoticed in a DFT calculation. The atoms will move in the direction of least force and the process is repeated again until the total force of the system is negligible. The schematic diagram of the complete relaxation in DFT calculations is shown in Figure 4.5. The inner loop bounded by dashed lines is the self-consistent field calculation that was shown previously in Figure 3.4. The outer loop is the ionic relaxation. If the total force of the system is almost zero, then desired material properties such total energy, pressure, stress, etc. can be calculated.

Common numerical algorithms used to implement ionic relaxation are the quasi-Newton method [174], the conjugate gradient (CG) method [175], and the damped molecular dynamics method [176].

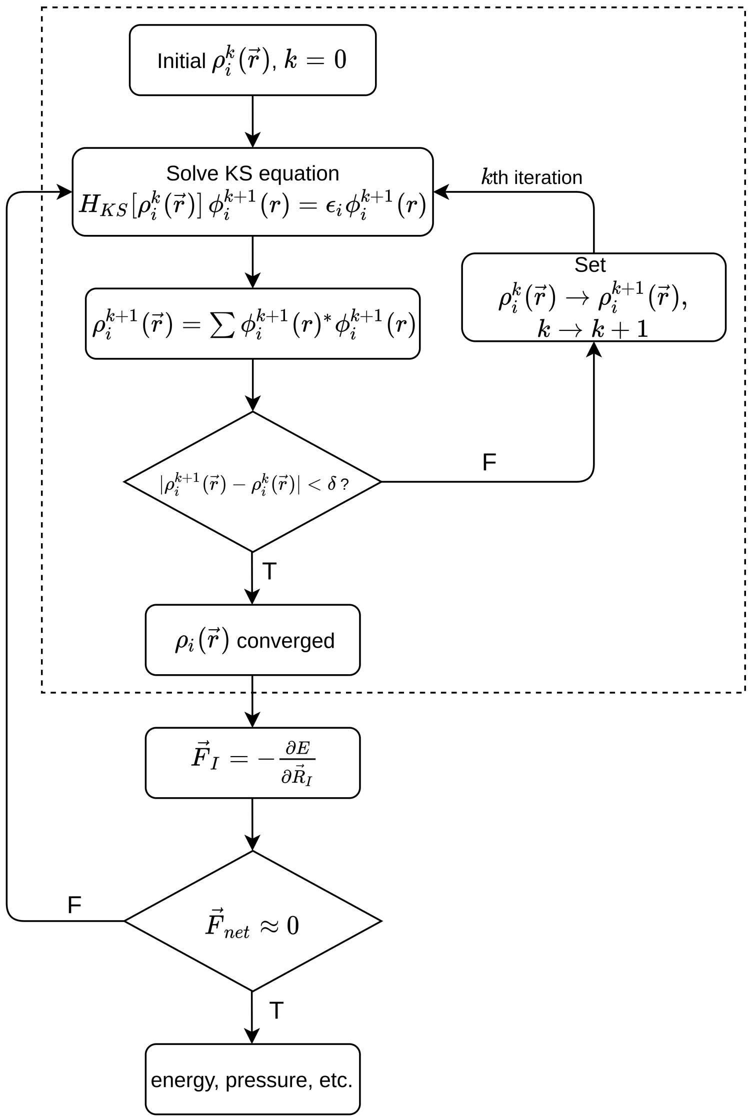

Figure 4.5: Schematic diagram of the complete relaxation in DFT simulations.

4.11 Density Mixing Schemes

In each iteration of the self consistent field calculation, it starts with a given electron density $\phi_i(\vec{r})$ and obtain the corresponding Kohn-Sham Hamiltonian and its eigenstates (see section 3.3.3). A new electron density can be computed from the occupied eigenstates. Afterwards, the new input density is updated from the old ones and will be used for the next iteration. The end goal is to reach the self consistent solution, i.e. $\rho_{out}(\vec{\mathbf{r}}) = \rho_{in}(\vec{\mathbf{r}})$, which can be thought as fixed point problem of the form $x_{n+1} = f(x_n)$. If the procedure converges, then the final value is the fixed point $\bar{x} = f(\bar{x})$. The simplest strategy is the linear mixing scheme [177]

$$\rho_{in}^{n+1}(\vec{\mathbf{r}}) = \alpha\rho_{out}^{n}(\vec{\mathbf{r}}) + (1-\alpha)\rho_{in}^{n}(\vec{\mathbf{r}}) \tag{4.46}$$

where α is the empirical mixing parameter adjusted to minimize the number of iterations needed for self consistency. The larger the α, the more is contributed from the output density. The purpose of density mixing is to prevent the charge sloshing. Charge sloshing is the consistent charge overshooting, or the large charge redistribution, that occurs from one iteration to the next. Density mixing will damp out these charge displacements leading to a better convergence [178]. Other advanced mixing schemes commonly used are Broyden mixing [179], Thomas-Fermi charge mixing [180], and Pulay mixing [157, 181]. The detailed explanations of these mixing schemes will not be discussed and interested readers are referred to the corresponding references.

4.12 Smearing

For materials that have band gap such as insulators and semiconductors, the electron densities decay smoothly near the band gap. However, metals have abrupt change of occupations from 0 to 1 at the Fermi level. This implies that any integration of functions that are discontinuous at the Fermi level will require a dense grid of k-points to achieve an acceptable accuracy. This will slow down the speed of convergence for a given set of k-points. The best way to deal with this problem is to use smearing. For instance, consider the total energy

$$E = \sum_i \int_{BZ} \epsilon_{ik}\, \Theta(\epsilon_{ik} - \mu)\, \mathrm{d}\vec{\mathbf{k}} \tag{4.47}$$

where $\Theta(\epsilon_{ik} - \mu)$ is the Dirac step function defined as

$$\Theta(x) = \begin{cases} 1 & , x \leq 0 \\ 0 & , x > 1 \end{cases} \tag{4.48}$$

Here, ϵ_{ik} is the energy of the ith band state located at wavevector $\vec{\mathbf{k}}$. Due to finite computer resources, the integral can be approximated by weighted sum over special k-points similar to (4.19)

$$E = \sum_{i} \sum_{k \in \text{IBZ}} w_k \epsilon_{ik} \, \Theta(\epsilon_{ik} - \mu) \tag{4.49}$$

where w_k are the weighted factors of each sampled k-points. The next step is to replace the Dirac step function by a smearing function. This will result to a much faster convergence speed without destroying the accuracy. The final approximate form will be

$$E = \sum_{k \in \text{IBZ}} w_k \sum_{i} f_{ik} \epsilon_{ik} \tag{4.50}$$

where f_{ik} is the smearing function which is also known as the partial occupancy. In similar vein, the electron density in (3.33) is reformulated to include the partial occupancy of the KS orbitals

$$\rho(\vec{\boldsymbol{r}}) = \sum_{k \in \text{IBZ}} w_k \sum_{i=1}^{N} f_{ik} \, \phi_{ik}(\vec{\boldsymbol{r}})^* \phi_{ik}(\vec{\boldsymbol{r}}) \tag{4.51}$$

The equation above implies that for a general system, in principle, there should be a set orbitals for every possible value of k-points. There are different smearing methods used in the literature. For example, the Fermi smearing method replaces the smearing function using Fermi-Dirac distribution [182, 183]

$$f_{ik} = \frac{1}{\exp[(\epsilon_{ik} - E_f)/k_B T] + 1} \tag{4.52}$$

where E_f is the Fermi energy, k_B is the Boltzmann's constant, and T is the absolute temperature. The $k_B T$ is also called the broadening parameter that quantifies the degree of broadening of the Fermi-Dirac distribution. Figure 4.6 illustrates the partial occupancies as a function of energy. Observe that there are eigenstates above the Fermi energy when smearing is applied. This means that eigenstates are partially filled inside the Fermi surface but the discontinuity at E_f is removed by a smooth function. Note that T has no physical meaning in DFT, unless the system under study is really at finite electronic temperature. Gaussian smearing is also possible which is given by [184]

$$f_{ik} = \frac{1}{2}\left[1 - erf\left(\frac{\epsilon_{ik} - E_f}{k_B T}\right)\right] \tag{4.53}$$

where $erf(x)$ is the Gauss error function. The Methfessel–Paxton smearing method [185] approximates the smearing function f_{ik} by a hierarchy of increasingly accurate smooth functions based on Hermite polynomials. Smaller k-points are adequate to accurately describe DFT quantities.

Lastly, the linear tetrahedron method divides the irreducible Brillouin zone (IBZ) into many small tetrahedrons [186]. The energy eigenvalues ϵ_{ik} inside each tetrahedron are linearly interpolated and integrated within these tetrahedrons. This method is especially suited for transition metals and rare earths whose delicate details of the Fermi surface requires a finer resolution. It is also preferred method for band structure and DOS calculations of semiconductors. The smearing method can also be used to broaden the Dirac delta function. For instance, the δ function is replaced by a Gaussian distribution

$$\delta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-x^2/2\sigma^2\right) \tag{4.54}$$

where σ is the broadening parameter. This broadening is very useful for calculations requiring delta functions such as density of states (DOS) in (3.14).

As a final note, the amount of broadening of smearing function must be optimized. Too large smearing might result in large error in calculation, whereas too small smearing requires a much finer k-point mesh, a computationally demanding task.

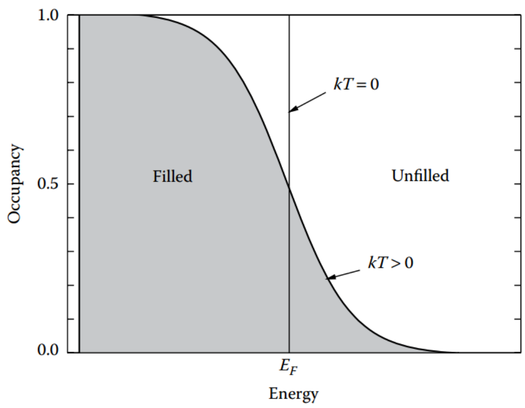

Figure 4.6: Partial occupancies near the Fermi energy using Fermi smearing. Illustration taken from [62].

4.13 Defect Formation Energies

One of the keys in probing point defects is the Gibbs free energy of defect formation, ΔG^f. It determines the stability of point defects at a given temperature, pressure, and composition. However, ΔG^f cannot be computed directly from first-principles calculations, but its electronic contribution denoted as ΔH^f is obtained from the total energy of a supercell model with appropriate approximations. Formally, under dilute defects, the Gibbs formation energy can be evaluated as [187**?** , 188]

$$\Delta G^f = G[\text{def}] - G[\text{bulk}] - \sum_i n_i \mu_i + q(E_{\text{vbm}} + E_f) \tag{4.55}$$

where $G[\text{def}]$ and $G[\text{bulk}]$ denote the Gibbs free energy of a simulation model containing a defect in charge state q and that of the perfect-crystal supercell, respectively; n_i and μ_i are the number and the chemical potential of constituent atoms of type i (i = Zn, O for ZnO), respectively; E_{vbm} is the absolute energy of the valence band maximum and E_f is the chemical potential of electrons (i.e. Fermi level) measured with respect to the top of the valence band. The Fermi level is strictly not an independent parameter since it is determined from the charge neurality principle. However, for practical considerations, it is informative to examine the dependence of defect formation energy with varying E_f in order to examine the behavior of defects when the doping or its charge changes. Hence, E_f is varied from zero to fundamental band gap E_g. By thermodynamic principles, the Gibbs free energy is often decomposed as [17]

$$G = H^{\text{el}} + H^{\text{vib}} - TS^{\text{vib}} + PV \tag{4.56}$$

where H^{el} denotes the electronic contribution to the total energy; H^{vib} and S^{vib} are the vibrational contributions to the total energy and entropy, respectively. T, P and V are the usual temperature, pressure and volume, respectively. For most calculations in solids, the vibrational contributions and the PV term are negligible at moderate temperatures and atmospheric pressure. Hence, the Gibbs free energy can be approximated in terms of H^{el}, which is obtained from supercell simulations. The Gibbs free energy of defect formation is in turn given by the approximation

$$G \approx H^{\text{el}} \Longrightarrow \Delta G^f = \Delta H^f \tag{4.57}$$

where ΔH^f, shortly called the defect formation energy, is given by the general formula [189]

$$\Delta H^f[X^q] = E_{\mathrm{tot}}[X^q] - E_{\mathrm{tot}}[\mathrm{bulk}] - \sum_i n_i\mu_i + q(E_{\mathrm{vbm}} + E_f) + E_{\mathrm{corr}} \tag{4.58}$$

where the first term is the total energy of a supercell containing a defect X at charge q and the second term is the total energy of the defect-free supercell. The other terms retain their usual meanings given in equation (4.55). Note that $n_i > 0$ ($n_i < 0$) if the atom of type i is added (removed) to form the defect. The last term E_{corr} is the correction term that accounts for errors in finite k-point sampling and the electrostatic interactions between the supercell and its images [190–192]. For instance, the defect formation energy, neglecting the correction term, of a positively charged Oxygen vacancy in ZnO crystal lattice is given by

$$\Delta H^f[V_{\mathrm{O}}^{1+}] = E_{\mathrm{tot}}[V_{\mathrm{O}}^{1+}] - E_{\mathrm{tot}}[\mathrm{ZnO}] - (-1)\mu_{\mathrm{O}} + (E_{\mathrm{vbm}} + E_f) \tag{4.59}$$

The formation energy of a point defect also depends on the chemical potentials of its constituent atoms. However, chemical potentials strongly depends on the experimental conditions, such as growth and annealing environments. For the case of ZnO, the chemical potentials μ_{O} and μ_{Zn} can vary in Zn rich, O rich, or in between, conditions. Ultimately, different experimental scenarios can be investigated. Hence, chemical potentials should explicitly regarded as variables in the defect calculations. However, by thermodynamic equilibrium, they are subject to bounds. These bounds are the transitions of forming new secondary phases. For ZnO, the chemical potentials are linked together by [189]

$$\mu_{\mathrm{Zn}} + \mu_{\mathrm{O}} = \mu_{\mathrm{ZnO}} \approx E_{\mathrm{tot}}[\mathrm{ZnO}_{\mathrm{unit}}] \tag{4.60}$$

where μ_{ZnO} is the chemical potential of ZnO taken to be the total energy of a ZnO unit cell. Usually, the atomic chemical potentials are referenced to their standard phase

$$\Delta\mu_{\mathrm{Zn}} = \mu_{\mathrm{Zn}} - \mu_{\mathrm{Zn}}(\mathrm{metal}) \tag{4.61}$$

$$\Delta\mu_{\mathrm{O}} = \mu_{\mathrm{O}} - \mu_{\mathrm{Zn}}(\mathrm{O}_2) \tag{4.62}$$

such that

$$\begin{aligned}
\Delta\mu_{\mathrm{ZnO}} &= \Delta\mu_{\mathrm{Zn}} + \Delta\mu_{\mathrm{O}} && (4.63)\\
&= (\mu_{\mathrm{Zn}} + \mu_{\mathrm{O}}) - \mu_{\mathrm{Zn}}(\mathrm{metal}) - \mu_{\mathrm{Zn}}(\mathrm{O}_2) && (4.64)\\
&= \mu_{\mathrm{ZnO}} - \mu_{\mathrm{Zn}}(\mathrm{metal}) - \mu_{\mathrm{O}}(\mathrm{O}_2) && (4.65)\\
&= \Delta H^f(\mathrm{ZnO}) && (4.66)
\end{aligned}$$

where $\Delta H^f(\mathrm{ZnO})$ is the enthalpy of formation of bulk ZnO which is negative for a stable compound. The term μ_{ZnO} is calculated from DFT total energy of ZnO unit cell, $\mu_{\mathrm{Zn}}(\mathrm{metal})$ from bulk hcp Zn and lastly, $\mu_{\mathrm{O}}(\mathrm{O}_2)$ taken from half the total energy of isolated O_2 molecule. The upper bound of μ_{Zn} and μ_{O} is constrained by their corresponding chemical potential at the standard phase. That is,

$$\begin{aligned}
\mu_{\mathrm{Zn}} \leq \mu_{\mathrm{Zn}}(\mathrm{metal}) &\implies \Delta\mu_{\mathrm{Zn}} \leq 0 && (4.67)\\
\mu_{\mathrm{O}} \leq \mu_{\mathrm{O}}(\mathrm{O}_2) \quad &\implies \Delta\mu_{\mathrm{O}} \leq 0 && (4.68)
\end{aligned}$$

The lower bound of μ_{O} (O-poor) is the upper bound of μ_{Zn} (Zn-rich) and vice versa

$$\begin{aligned}
\Delta\mu_{\mathrm{O}} \geq \Delta H^f(\mathrm{ZnO}), \Delta\mu_{\mathrm{Zn}} \leq 0 \quad &[\text{O-poor,Zn-rich}] && (4.69)\\
\Delta\mu_{\mathrm{Zn}} \geq \Delta H^f(\mathrm{ZnO}), \Delta\mu_{\mathrm{O}} \ \leq 0 \quad &[\text{O-rich,Zn-poor}] && (4.70)
\end{aligned}$$

Using (4.61) and (4.62), the above equations can be generalized as [193]

$$\begin{aligned}
\mu_{\mathrm{O}} &= \mu_{\mathrm{O}}(\mathrm{O}_2) + \lambda\,\Delta H^f(\mathrm{ZnO}) && (4.71)\\
\mu_{\mathrm{Zn}} &= \mu_{\mathrm{Zn}}(\mathrm{metal}) + (1-\lambda)\Delta H^f(\mathrm{ZnO}) && (4.72)
\end{aligned}$$

where the parameter λ varies from zero (O-rich, Zn-poor) to unity (O-poor, Zn-rich). The parametrization form is useful since the calculation of defect formation energies in (4.58) depends on the nonstandard chemical potentials μ_{O} and μ_{Zn}.

When two distinct charged defects of the same atom have the same formation energies, the position of the Fermi energy E_f is defined to be the defect transition level. This transition level is defined as [14]

$$\epsilon(q/q') = \frac{\Delta H^f[X^q; E_f = 0] - \Delta H^f[X^{q'}; E_f = 0]}{q' - q} \tag{4.73}$$

where $\Delta H^f[X^q; E_f = 0]$ is the formation energy of defect X at charge state q measured when the Fermi level is at the valence band maximum. The significance of transition levels is to determine the range of Fermi levels where a given charged defect is stable. That is, for Fermi levels below $\epsilon(q/q')$, charge state q is the most stable, otherwise, charge state q' is the stable one. Figure 4.7 illustrates the stabilities of a system with three different charge defects. The transition level defined in the above equation is the thermodynamic transition level, and should not be confused with the Kohn-Sham states resulted from the bandstructure calculations nor to be confused with the optical transition level. The thermodynamic transition level arises from the atomic relaxation of the lattice, while the optical transition level arises from optical excitations and the the atomic geometry are frozen during the transition. In addition, the thermodynamic transition is independent of the charge transfer, i.e. it is invariant whether electron is added or removed. However, for optical transitions, adding an electron results to a higher energy level while removing an electron results to a lower energy level. The energy difference is related to Stokes shifts (absorption vs. emission) [189].

Conventionally, if a defect transition level is situated such that it is likely to be thermally ionized at room temperature, then the transition level is called a shallow level, otherwise it is deep level. The shallow levels can occur in two scenarios: first, the transition level is near to the band edges (VBM for acceptor, CBM for donor); second, the transition level is in resonance with the valence or conduction bands (located inside in one of the bands). For resonant states, the defect charge carrier will find a lower energy state by transferring to the band edges [14].

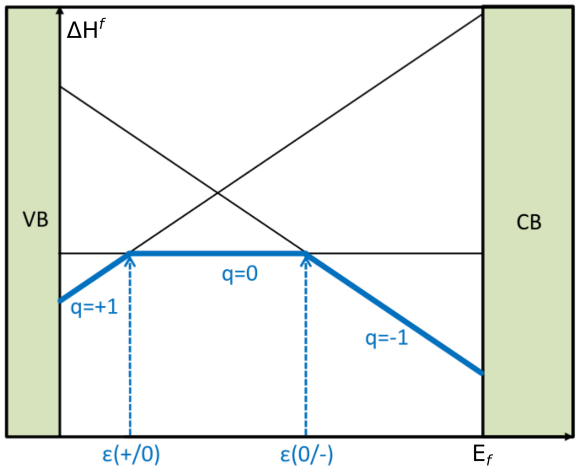

Figure 4.7: Schematic illustration of defect formation energy ΔH^f dependence on the Fermi level E_f for the three charge defects q: +1, 0, and -1. The solid lines correspond to the defect formation energy whose slope is the defect charge value as defined in Eq. (4.58). Two transition levels can be observed: a deep acceptor at $\epsilon(+/0)$ and a deep donor at $\epsilon(0/-)$. The thick solid lines indicate the energetically most stable charge state for a given Fermi level. Illustration taken from [189].

References

[1] D.C. Look. Recent advances in ZnO materials and devices. *Materials Science and Engineering: B*, 80(1-3):383–387, mar 2001.

[2] Ü. Özgür, Ya. I. Alivov, C. Liu, A. Teke, M. A. Reshchikov, S. Doğan, V. Avrutin, S.-J. Cho, and H. Morkoç. A comprehensive review of ZnO materials and devices. *Journal of Applied Physics*, 98(4):041301, aug 2005.

[3] C. Klingshirn. ZnO: From basics towards applications. *Physica Status Solidi B*, 244(9):3027–3073, sep 2007.

[4] Anderson Janotti and Chris G Van de Walle. Fundamentals of zinc oxide as a semiconductor. *Reports on Progress in Physics*, 72(12):126501, oct 2009.

[5] Agnieszka Kołodziejczak-Radzimska and Teofil Jesionowski. Zinc oxide—from synthesis to application: A review. *Materials*, 7(4):2833–2881, apr 2014.

[6] D. C. Reynolds, D. C. Look, B. Jogai, C. W. Litton, G. Cantwell, and W. C. Harsch. Valence-band ordering in ZnO. *Physical Review B*, 60(4):2340–2344, jul 1999.

[7] Pietro Camarda, Fabrizio Messina, Lavinia Vaccaro, Simonpietro Agnello, Gianpiero Buscarino, Reinhard Schneider, Radian Popescu, Dagmar Gerthsen, Roberto Lorenzi, Franco Mario Gelardi, and Marco Cannas. Luminescence mechanisms of defective ZnO nanoparticles. *Physical Chemistry Chemical Physics*, 18(24):16237–16244, 2016.

[8] M. Chitra, G. Mangamma, K. Uthayarani, N. Neelakandeswari, and E.K. Girija. Band gap engineering in ZnO based nanocomposites. *Physica E: Low-dimensional Systems and Nanostructures*, 119:113969, may 2020.

[9] D. M. Bagnall, Y. F. Chen, Z. Zhu, T. Yao, S. Koyama, M. Y. Shen, and T. Goto. Optically pumped lasing of ZnO at room temperature. *Applied Physics Letters*, 70(17):2230–2232, apr 1997.

[10] P. A. Rodnyi and I. V. Khodyuk. Optical and luminescence properties of zinc oxide (review). *Optics and Spectroscopy*, 111(5):776–785, nov 2011.

[11] E. I. Gorokhova, G. V. Anan'eva, V. A. Demidenko, P. A. Rodnyĭ, I. V. Khodyuk, and E. D. Bourret-Courchesne. Optical, luminescence, and scintillation properties of ZnO and ZnO:ga ceramics. *Journal of Optical Technology*, 75(11):741, nov 2008.

[12] J. Ji, A. M. Colosimo, W. Anwand, L. A. Boatner, A. Wagner, P. S. Stepanov, T. T. Trinh, M. O. Liedke, R. Krause-Rehberg, T. E. Cowan, and F. A. Selim. ZnO luminescence and scintillation studied via photoexcitation, x-ray excitation and gamma-induced positron spectroscopy. *Scientific Reports*, 6(1), aug 2016.

[13] Nobuhiro Konishi, Tadashi Shibata, and Tadahiro Ohmi. Impurity back diffusion through an ultrahigh vacuum turbomolecular pump under large gas throughput. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 14(5):2958–2962, sep 1996.

[14] Anderson Janotti and Chris G. Van de Walle. Native point defects in ZnO. *Physical Review B*, 76(16), oct 2007.

[15] Bhupendra K. Sharma, Neeraj Khare, and D. Haranath. Photoluminescence lifetime of al-doped ZnO films in visible region. *Solid State Communications*, 150(47-48):2341–2345, dec 2010.

[16] Ping Wang, Jingfang He, Lixin Guo, Yintang Yang, and Shukai Zheng. The electronic structures and optical properties of yttrium-doped zinc oxide with zinc interstitial defects calculated by first-principles. *Materials Science in Semiconductor Processing*, 36:36–42, aug 2015.

[17] Fumiyasu Oba, Minseok Choi, Atsushi Togo, and Isao Tanaka. Point defects in ZnO: an approach from first principles. *Science and Technology of Advanced Materials*, 12(3):034302, jun 2011.

[18] Lishu Liu, Zengxia Mei, Aihua Tang, Alexander Azarov, Andrej Kuznetsov, Qi-Kun Xue, and Xiaolong Du. Oxygen vacancies: The origin of n-type conductivity in ZnO. *Physical Review B*, 93(23), jun 2016.

[19] Verdad C. Agulto, Melvin John F. Empizo, Keisuke Kawano, Yuki Minami, Kohei Yamanoi, Nobuhiko Sarukura, Allan Christopher C. Yago, and Roland V. Sarmago. Two-step fabrication of ZnO-PVP composites with tunable visible emissions. *Optical Materials*, 76:317–322, feb 2018.

[20] E. Musavi, M. Khanlary, and Z. Khakpour. Red-orange photoluminescence emission of sol-gel dip-coated prepared ZnO and ZnO:al nano-crystalline films. *Journal of Luminescence*, 216:116696, dec 2019.

[21] Kamal Rudra and Y.K. Prajapati. Effect of mn doping on defect-related photoluminescence and nanostructure of dense 3-d nano-root network of ZnO. *Ceramics International*, 46(8):10135–10141, jun 2020.

[22] Xiangliu Chen, Qingshuang Xie, and Jitao Li. Significantly improved photoluminescence properties of ZnO thin films by lithium doping. *Ceramics International*, 46(2):2309–2316, feb 2020.

[23] Mingjing Bian, Hongyan Zhang, Jun Zhang, and Zhijun Li. Effects of post-annealing on photoluminescence of eu-doped ZnO microsphere for single-component white-light materials. *Optik*, 209:164607, may 2020.

[24] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

[25] Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965.

[26] Zhi Yang and Shi-Jie Xiong. Adsorption of ag and au atoms on wurtzite ZnO (0001) surface. *Surface Science*, 605(1-2):40–45, jan 2011.

[27] Saeed Masoumi, Ebrahim Nadimi, and Faramarz Hossein-Babaei. Electronic properties of ag-doped ZnO: DFT hybrid functional study. *Physical Chemistry Chemical Physics*, 20(21):14688–14693, 2018.

[28] Ivan I Vrubel, Anastasiia A Pervishko, Dmitry Yudin, Biplab Sanyal, Olle Eriksson, and Piotr A Rodnyi. Oxygen vacancy in ZnO- w phase: pseudohybrid hubbard density functional study. *Journal of Physics: Condensed Matter*, 32(31):315503, may 2020.

[29] A. F. Kohan, G. Ceder, D. Morgan, and Chris G. Van de Walle. First-principles study of native point defects in ZnO. *Physical Review B*, 61(22):15019–15027, jun 2000.

[30] Chris G. Van de Walle. Defect analysis and engineering in ZnO. *Physica B: Condensed Matter*, 308-310:899–903, dec 2001.

[31] S. B. Zhang, S.-H. Wei, and Alex Zunger. Intrinsic n-type versus p-type doping asymmetry and the defect physics of ZnO. *Physical Review B*, 63(7), jan 2001.

[32] Fumiyasu Oba, Shigeto R. Nishitani, Seiji Isotani, Hirohiko Adachi, and Isao Tanaka. Energetics of native defects in ZnO. *Journal of Applied Physics*, 90(2):824–828, jul 2001.

[33] Eun-Cheol Lee, Y.-S. Kim, Y.-G. Jin, and K. J. Chang. Compensation mechanism for n acceptors in ZnO. *Physical Review B*, 64(8), aug 2001.

[34] Paul Erhart, Andreas Klein, and Karsten Albe. First-principles study of the structure and stability of oxygen defects in zinc oxide. *Physical Review B*, 72(8), aug 2005.

[35] Paul Erhart and Karsten Albe. First-principles study of migration mechanisms and diffusion of oxygen in zinc oxide. *Physical Review B*, 73(11), mar 2006.

[36] Stephan Lany and Alex Zunger. Dopability, intrinsic conductivity, and nonstoichiometry of transparent conducting oxides. *Physical Review Letters*, 98(4), jan 2007.

[37] Liping Qiao, Changchun Chai, Yintang Yang, Xinhai Yu, and Chunlei Shi. Strain effects on band structure of wurtzite ZnO: a GGA + U study. *Journal of Semiconductors*, 35(7):073004, jul 2014.

[38] M. K. Yaakob, N. H. Hussin, M. F. M. Taib, T. I. T. Kudin, O. H. Hassan, A. M. M. Ali, and M. Z. A. Yahya. First principles LDA+U calculations for ZnO materials. *Integrated Ferroelectrics*, 155(1):15–22, may 2014.

[39] Sara Sadat Parhizgar and Javad Beheshtian. Effect of nitrogen doping on electronic and optical properties of ZnO sheet: DFT+U study. *Computational Condensed Matter*, 15:1–6, jun 2018.

[40] Kausar Harun, Nor Azmira Salleh, Bahri Deghfel, Muhamad Kamil Yaakob, and Ahmad Azmin Mohamad. DFT+U calculations for electronic, structural, and optical properties of ZnO wurtzite structure: A review. *Results in Physics*, 16:102829, mar 2020.

[41] Theo Hahn. *International Tables for Crystallography, Vol A*, volume A. Springer, Dordrecht, Netherlands, 2005.

[42] Xinguo Ma, Ying Wu, Yanhui Lv, and Yongfa Zhu. Correlation effects on lattice relaxation and electronic structure of ZnO within the GGA+U formalism. *The Journal of Physical Chemistry C*, 117(49):26029–26039, nov 2013.

[43] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. The materials project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, jul 2013.

[44] L. P. Bouckaert, R. Smoluchowski, and E. Wigner. Theory of brillouin zones and symmetry properties of wave functions in crystals. *Physical Review*, 50(1):58–67, jul 1936.

[45] U Shmueli. *International Tables for Crystallography, Vol B.* Kluwer Academic, Dordrecht, Netherlands, 2001.

[46] Mois I. Aroyo, Asen Kirov, Cesar Capillas, J. M. Perez-Mato, and Hans Wondratschek. Bilbao crystallographic server. II. representations of crystallographic point groups and space groups. *Acta Crystallographica Section A Foundations of Crystallography*, 62(2):115–128, mar 2006.

[47] J. Perez-Mato, D Orobengoa, Emre Tasci, Gemma De la Flor Martin, and A Kirov. Crystallography online: Bilbao crystallographic server. *Bulgarian Chemical Communications*, 43:183–197, 01 2011.

[48] Mois I. Aroyo, Danel Orobengoa, Gemma de la Flor, Emre S. Tasci, J. Manuel Perez-Mato, and Hans Wondratschek. Brillouin-zone database on theBilbao crystallographic server. *Acta Crystallographica Section A Foundations and Advances*, 70(2):126–137, feb 2014.

[49] Wahyu Setyawan and Stefano Curtarolo. High-throughput electronic band structure calculations: Challenges and tools. *Computational Materials Science*, 49(2):299–312, aug 2010.

[50] Shyue Ping Ong, Shreyas Cholia, Anubhav Jain, Miriam Brafman, Dan Gunter, Gerbrand Ceder, and Kristin A. Persson. The materials application programming interface (API): A simple, flexible and efficient API for materials data based on REpresentational state transfer (REST) principles. *Computational Materials Science*, 97:209–215, feb 2015.

[51] A. Slassi, S. Naji, A. Benyoussef, M. Hamedoun, and A. El Kenz. On the transparent conducting oxide al doped ZnO: First principles and boltzmann equations study. *Journal of Alloys and Compounds*, 605:118–123, aug 2014.

[52] Jin-Hua Luo, Qun Liu, Li-Na Yang, Zhu-Zhu Sun, and Ze-Sheng Li. First-principles study of electronic structure and optical properties of (zr–al)-codoped ZnO. *Computational Materials Science*, 82:70–75, feb 2014.

[53] Maengsuk Kim and Chul Hong Park. Self-consistent GW calculation of the electronic structure of co-doped ZnO. *Journal of the Korean Physical Society*, 60(2):292–296, jan 2012.

[54] Ferdinand Anne Kröger. The chemistry of imperfect crystals. Technical report, North-Holland Pub. Co., 1964.

[55] Sol E. Harrison. Conductivity and hall effect of ZnO at low temperatures. *Physical Review*, 93(1):52–62, jan 1954.

[56] D.G. Thomas. Interstitial zinc in zinc oxide. *Journal of Physics and Chemistry of Solids*, 3(3-4):229–237, 1957.

[57] A. Hausmann and W. Teuerle. Halleffekt-messungen an reinen und kupfer dotierten zinkoxidkristallen. *Zeitschrift für Physik A Hadrons and nuclei*, 259(2):189–194, apr 1973.

[58] K.I. Hagemark. Defect structure of zn-doped ZnO. *Journal of Solid State Chemistry*, 16(3-4):293–299, jan 1976.

[59] Erwin Schrödinger. Quantisierung als eigenwertproblem. *Annalen der physik*, 385(13):437–490, 1926.

[60] Frank Herman. Lattice vibrational spectrum of germanium. *Journal of Physics and Chemistry of Solids*, 8:405–418, jan 1959.

[61] Charles Kittel. *Introduction to Solid State Physics*. John Wiley & Sons Inc, 2004.

[62] June Gunn Lee. *Computational Materials Science:An Introduction*. Taylor & Francis Inc, 2016.

[63] N. Mermin Neil W. Ashcroft. *Solid State Physics*. Cengage Learning, Inc, 1976.

[64] J Enkovaara, C Rostgaard, J J Mortensen, J Chen, M Dułak, L Ferrighi, J Gavnholt, C Glinsvad, V Haikola, H A Hansen, H H Kristoffersen, M Kuisma, A H Larsen, L Lehtovaara, M Ljungberg, O Lopez-Acevedo, P G Moses, J Ojanen, T Olsen, V Petzold, N A Romero, J Stausholm-Møller, M Strange, G A Tritsaris, M Vanin, M Walter, B Hammer, H Häkkinen, G K H Madsen, R M Nieminen, J K Nørskov, M Puska, T T Rantala, J Schiøtz, K S Thygesen, and K W Jacobsen. Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *Journal of Physics: Condensed Matter*, 22(25):253202, jun 2010.

[65] Max Born and Robert Oppenheimer. Zur quantentheorie der molekeln. *Annalen der physik*, 389(20):457–484, 1927.

[66] D. R. Hartree. The wave mechanics of an atom with a non-coulomb central field. part i. theory and methods. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(1):89–110, jan 1928.

[67] V. Fock. Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. *Zeitschrift für Physik*, 61(1-2):126–148, jan 1930.

[68] Wolfgang Pauli. Über den zusammenhang des abschlusses der elektronengruppen im atom mit der komplexstruktur der spektren. *Zeitschrift für Physik*, 31(1):765–783, 1925.

[69] John C Slater. The theory of complex spectra. *Physical Review*, 34(10):1293, 1929.

[70] Walter Kohn. Nobel lecture: Electronic structure of matter—wave functions and density functionals. *Reviews of Modern Physics*, 71(5):1253, 1999.

[71] Stefaan Cottenier et al. Density functional theory and the family of (l) apw-methods: a step-by-step introduction. *Instituut voor Kern-en Stralingsfysica, KU Leuven, Belgium*, 4(0):41, 2002.

[72] N D Woods, M C Payne, and P J Hasnip. Computing the self-consistent field in kohn–sham density functional theory. *Journal of Physics: Condensed Matter*, 31(45):453001, aug 2019.

[73] Pragya Verma and Donald G Truhlar. Status and challenges of density functional theory. *Trends in Chemistry*, 2(4):302–318, 2020.

[74] Miguel A.L. Marques, Micael J.T. Oliveira, and Tobias Burnus. Libxc: A library of exchange and correlation functionals for density functional theory. *Computer Physics Communications*, 183(10):2272–2281, oct 2012.

[75] Maximiliano Segala and Delano P. Chong. An evaluation of exchange-correlation functionals for the calculations of the ionization energies for atoms and molecules. *Journal of Electron Spectroscopy and Related Phenomena*, 171(1-3):18–23, apr 2009.

[76] P. A. M. Dirac. Note on exchange phenomena in the thomas atom. *Mathematical Proceedings of the Cambridge Philosophical Society*, 26(3):376–385, jul 1930.

[77] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Physical Review Letters*, 45(7):566–569, aug 1980.

[78] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal. Quantum monte carlo simulations of solids. *Reviews of Modern Physics*, 73(1):33–83, jan 2001.

[79] J. P. Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Physical Review B*, 23(10):5048–5079, may 1981.

[80] O. Gunnarsson and B. I. Lundqvist. Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism. *Physical Review B*, 13(10):4274–4298, may 1976.

[81] O. Gunnarsson, M. Jonson, and B.I. Lundqvist. Exchange and correlation in inhomogeneous electron systems. *Solid State Communications*, 24(11):765–768, dec 1977.

[82] Viktor N. Staroverov, Gustavo E. Scuseria, Jianmin Tao, and John P. Perdew. Tests of a ladder of density functionals for bulk solids and surfaces. *Physical Review B*, 69(7), feb 2004.

[83] Gábor I. Csonka, John P. Perdew, Adrienn Ruzsinszky, Pier H. T. Philipsen, Sébastien Lebègue, Joachim Paier, Oleg A. Vydrov, and János G. Ángyán. Assessing the performance of recent density functionals for bulk solids. *Physical Review B*, 79(15), apr 2009.

[84] Judith Harl, Laurids Schimka, and Georg Kresse. Assessing the quality of the random phase approximation for lattice constants and atomization energies of solids. *Physical Review B*, 81(11), mar 2010.

[85] Changyol Lee, David Vanderbilt, Kari Laasonen, R. Car, and M. Parrinello. Ab initio studies on the structural and dynamical properties of ice. *Physical Review B*, 47(9):4863–4872, mar 1993.

[86] D. R. Hamann. H2o hydrogen bonding in density-functional theory. *Physical Review B*, 55(16):R10157–R10160, apr 1997.

[87] Peter J. Feibelman. Lattice match in density functional calculations: ice ih vs. β-AgI. *Physical Chemistry Chemical Physics*, 10(32):4688, 2008.

[88] Sverre Froyen and Marvin L. Cohen. Structural properties of III-v zinc-blende semiconductors under pressure. *Physical Review B*, 28(6):3258–3265, sep 1983.

[89] Jiajin Tan, Ying Li, and Guangfu Ji. Elastic constants and bulk modulus of semiconductors: Performance of plane-wave pseudopotential and local-density-approximation density functional theory. *Computational Materials Science*, 58:243–247, jun 2012.

[90] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865–3868, oct 1996.

[91] John P. Perdew, Adrienn Ruzsinszky, Gábor I. Csonka, Oleg A. Vydrov, Gustavo E. Scuseria, Lucian A. Constantin, Xiaolan Zhou, and Kieron Burke. Restoring the density-gradient expansion for exchange in solids and surfaces. *Physical Review Letters*, 100(13), apr 2008.

[92] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*, 38(6):3098–3100, sep 1988.

[93] John P. Perdew and Yue Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Physical Review B*, 45(23):13244–13249, jun 1992.

[94] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2):785–789, jan 1988.

[95] Nicholas C. Handy and Aron J. Cohen. Left-right correlation energy. *Molecular Physics*, 99(5):403–412, mar 2001.

[96] X. Xu and W. A. Goddard. The x3lyp extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties. *Proceedings of the National Academy of Sciences*, 101(9):2673–2677, feb 2004.

[97] Kieron Burke, John P. Perdew, and Matthias Ernzerhof. Why the generalized gradient approximation works and how to go beyond it. *International Journal of Quantum Chemistry*, 61(2):287–293, 1997.

[98] David C. Langreth and M. J. Mehl. Beyond the local-density approximation in calculations of ground-state electronic properties. *Physical Review B*, 28(4):1809–1834, aug 1983.

[99] John P. Perdew, J. A. Chevary, S. H. Vosko, Koblar A. Jackson, Mark R. Pederson, D. J. Singh, and Carlos Fiolhais. Atoms, molecules, solids, and surfaces:

Applications of the generalized gradient approximation for exchange and correlation. *Physical Review B*, 46(11):6671–6687, sep 1992.

[100] E. I. Proynov, E. Ruiz, A. Vela, and D. R. Salahub. Determining and extending the domain of exchange and correlation functionals. *International Journal of Quantum Chemistry*, 56(S29):61–78, feb 1995.

[101] K. Kim and K. D. Jordan. Comparison of density functional and MP2 calculations on the water monomer and dimer. *The Journal of Physical Chemistry*, 98(40):10089–10094, oct 1994.

[102] JoséM. Pérez-Jordá and A.D. Becke. A density-functional study of van der waals forces: rare gas diatomics. *Chemical Physics Letters*, 233(1-2):134–137, feb 1995.

[103] Eliseo Ruiz, Dennis R. Salahub, and Alberto Vela. Defining the domain of density functionals: Charge-transfer complexes. *Journal of the American Chemical Society*, 117(3):1141–1142, jan 1995.

[104] Paula Mori-Sánchez, Aron J. Cohen, and Weitao Yang. Localization and delocalization errors in density functional theory and implications for band-gap prediction. *Physical Review Letters*, 100(14), apr 2008.

[105] John P. Perdew, Weitao Yang, Kieron Burke, Zenghui Yang, Eberhard K. U. Gross, Matthias Scheffler, Gustavo E. Scuseria, Thomas M. Henderson, Igor Ying Zhang, Adrienn Ruzsinszky, Haowei Peng, Jianwei Sun, Egor Trushin, and Andreas Görling. Understanding band gaps of solids in generalized kohn–sham theory. *Proceedings of the National Academy of Sciences*, 114(11):2801–2806, mar 2017.

[106] John P. Perdew, Robert G. Parr, Mel Levy, and Jose L. Balduz. Density-functional theory for fractional particle number: Derivative discontinuities of the energy. *Physical Review Letters*, 49(23):1691–1694, dec 1982.

[107] John P. Perdew and Mel Levy. Physical content of the exact kohn-sham orbital energies: Band gaps and derivative discontinuities. *Physical Review Letters*, 51(20):1884–1887, nov 1983.

[108] L. J. Sham and M. Schlüter. Density-functional theory of the energy gap. *Physical Review Letters*, 51(20):1888–1891, nov 1983.

[109] E. J. Baerends. From the kohn–sham band gap to the fundamental gap in solids. an integer electron approach. *Physical Chemistry Chemical Physics*, 19(24):15639–15656, 2017.

[110] M. J. van Setten, F. Weigend, and F. Evers. The GW-method for quantum chemistry applications: Theory and implementation. *Journal of Chemical Theory and Computation*, 9(1):232–246, dec 2012.

[111] Lars Hedin. New method for calculating the one-particle green's function with application to the electron-gas problem. *Physical Review*, 139(3A):A796–A823, aug 1965.

[112] Lars Hedin and Stig Lundqvist. Effects of electron-electron and electron-phonon interactions on the one-electron states of solids. *Solid State Physics*, 23:1–181, 1970.

[113] Mark S. Hybertsen and Steven G. Louie. Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. *Physical Review B*, 34(8):5390–5413, oct 1986.

[114] R. W. Godby, M. Schlüter, and L. J. Sham. Accurate exchange-correlation potential for silicon and its discontinuity on addition of an electron. *Physical Review Letters*, 56(22):2415–2418, jun 1986.

[115] Georgy Samsonidze, Manish Jain, Jack Deslippe, Marvin L. Cohen, and Steven G. Louie. Simple approximate physical orbitals for GW Quasiparticle calculations. *Physical Review Letters*, 107(18), oct 2011.

[116] Jack Deslippe, Georgy Samsonidze, Manish Jain, Marvin L. Cohen, and Steven G. Louie. Coulomb-hole summations and energies for GW calculations with limited number of empty orbitals: A modified static remainder approach. *Physical Review B*, 87(16), apr 2013.

[117] Weiwei Gao, Weiyi Xia, Xiang Gao, and Peihong Zhang. Speeding up GW calculations to meet the challenge of large scale quasiparticle predictions. *Scientific Reports*, 6(1), nov 2016.

[118] M. van Schilfgaarde, Takao Kotani, and S. Faleev. Quasiparticle self-Consistent GW Theory. *Physical Review Letters*, 96(22), jun 2006.

[119] Axel D. Becke. A new mixing of hartree–fock and local density-functional theories. *The Journal of Chemical Physics*, 98(2):1372–1377, jan 1993.

[120] John P. Perdew, Matthias Ernzerhof, and Kieron Burke. Rationale for mixing exact exchange with density functional approximations. *The Journal of Chemical Physics*, 105(22):9982–9985, dec 1996.

[121] Carlo Adamo and Vincenzo Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of Chemical Physics*, 110(13):6158–6170, apr 1999.

[122] Joachim Paier, Martijn Marsman, and Georg Kresse. Why does the b3lyp hybrid functional fail for metals? *The Journal of Chemical Physics*, 127(2):024103, jul 2007.

[123] Jochen Heyd, Gustavo E. Scuseria, and Matthias Ernzerhof. Hybrid functionals based on a screened coulomb potential. *The Journal of Chemical Physics*, 118(18):8207–8215, may 2003.

[124] Aliaksandr V. Krukau, Oleg A. Vydrov, Artur F. Izmaylov, and Gustavo E. Scuseria. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *The Journal of Chemical Physics*, 125(22):224106, dec 2006.

[125] Jianmin Tao, John P. Perdew, Viktor N. Staroverov, and Gustavo E. Scuseria. Climbing the density functional ladder: Nonempirical meta–generalized gradient approximation designed for molecules and solids. *Physical Review Letters*, 91(14), sep 2003.

[126] Fabien Tran and Peter Blaha. Accurate band gaps of semiconductors and insulators with a semilocal exchange-correlation potential. *Physical Review Letters*, 102(22), jun 2009.

[127] John P. Perdew, Stefan Kurth, Aleš Zupan, and Peter Blaha. Accurate density functional with correct formal properties: A step beyond the generalized gradient approximation. *Physical Review Letters*, 82(12):2544–2547, mar 1999.

[128] Jianwei Sun, Adrienn Ruzsinszky, and John P. Perdew. Strongly constrained and appropriately normed semilocal density functional. *Physical Review Letters*, 115(3), jul 2015.

[129] David J. Singh. Structure and optical properties of high light output halide scintillators. *Physical Review B*, 82(15), oct 2010.

[130] David J. Singh. Electronic structure calculations with the tran-blaha modified becke-johnson density functional. *Physical Review B*, 82(20), nov 2010.

[131] Burak Himmetoglu, Andrea Floris, Stefano de Gironcoli, and Matteo Cococcioni. Hubbard-corrected DFT energy functionals: The LDA+U description of correlated systems. *International Journal of Quantum Chemistry*, 114(1):14–49, jul 2013.

[132] Z.-X. Shen, R. S. List, D. S. Dessau, B. O. Wells, O. Jepsen, A. J. Arko, R. Barttlet, C. K. Shih, F. Parmigiani, J. C. Huang, and P. A. P. Lindberg. Electronic structure of NiO: Correlation and band effects. *Physical Review B*, 44(8):3604–3626, aug 1991.

[133] Sarah A. Tolba, Kareem M. Gameel, Basant A. Ali, Hossam A. Almossalami, and Nageh K. Allam. The DFT+U: Approaches, accuracy, and applications. In *Density Functional Calculations - Recent Progresses of Theory and Application*. InTech, may 2018.

[134] A. I. Liechtenstein, V. I. Anisimov, and J. Zaanen. Density-functional theory and strong interactions: Orbital ordering in mott-hubbard insulators. *Physical Review B*, 52(8):R5467–R5470, aug 1995.

[135] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton. Electron-energy-loss spectra and the structural stability of nickel oxide:an LSDA+u study. *Physical Review B*, 57(3):1505–1509, jan 1998.

[136] Matteo Cococcioni and Stefano de Gironcoli. Linear response approach to the calculation of the effective interaction parameters in the LDA+U method. *Physical Review B*, 71(3), jan 2005.

[137] Burak Himmetoglu, Renata M. Wentzcovitch, and Matteo Cococcioni. First-principles study of electronic and structural properties of CuO. *Physical Review B*, 84(11), sep 2011.

[138] Heather J. Kulik, Matteo Cococcioni, Damian A. Scherlis, and Nicola Marzari. Density functional theory in transition-metal chemistry: A self-consistent Hubbard U Approach. *Physical Review Letters*, 97(10), sep 2006.

[139] M. E. Arroyo de Dompablo, A. Morales-Garcia, and M. Taravillo. DFT + U calculations of crystal lattice, electronic structure, and phase stability under pressure of TiO_2 polymorphs. *The Journal of Chemical Physics*, 135(5):054503, aug 2011.

[140] Fabien Tran, Peter Blaha, Karlheinz Schwarz, and Pavel Novák. Hybrid exchange-correlation energy functionals for strongly correlated electrons: Applications to transition-metal monoxides. *Physical Review B*, 74(15), oct 2006.

[141] Stefaan Cottenier. *Density Functional Theory and the family of (L) APW-methods: a step-by-step introduction.* Instituut voor Kern-en Stralingsfysica, KU Leuven, Belgium, 2013.

[142] Javier Junquera, Óscar Paz, Daniel Sánchez-Portal, and Emilio Artacho. Numerical atomic orbitals for linear-scaling calculations. *Physical Review B*, 64(23), nov 2001.

[143] Manuel J. Louwerse and Gadi Rothenberg. Transferable basis sets of numerical atomic orbitals. *Physical Review B*, 85(3), jan 2012.

[144] J. C. Slater. Analytic atomic wave functions. *Physical Review*, 42(1):33–43, oct 1932.

[145] S. F. Boys and Alfred Charles Egerton. Electronic wave functions - I. A general method of calculation for the stationary states of any molecular system. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 200(1063):542–554, feb 1950.

[146] M D Segall, Philip J D Lindan, M J Probert, C J Pickard, P J Hasnip, S J Clark, and M C Payne. First-principles simulation: ideas, illustrations and the

CASTEP code. *Journal of Physics: Condensed Matter*, 14(11):2717–2744, mar 2002.

[147] J. C. Slater. An augmented plane wave method for the periodic potential problem. *Physical Review*, 92(3):603–608, nov 1953.

[148] E Sjöstedt, L Nordström, and D.J Singh. An alternative way of linearizing the augmented plane-wave method. *Solid State Communications*, 114(1):15–20, mar 2000.

[149] O. Krogh Andersen. Linear methods in band theory. *Physical Review B*, 12(8):3060–3083, oct 1975.

[150] Jorge Kohanoff. *Electronic structure calculations for solids and molecules: theory and computational methods.* Cambridge University Press, Cambridge, 2006.

[151] Ryan Requist and Oleg Pankratov. Generalized kohn-sham system in one-matrix functional theory. *Physical Review B*, 77(23), jun 2008.

[152] Ernest R. Davidson. The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. *Journal of Computational Physics*, 17(1):87–94, jan 1975.

[153] M. Crouzeix, B. Philippe, and M. Sadkane. The davidson method. *SIAM Journal on Scientific Computing*, 15(1):62–76, jan 1994.

[154] D M Wood and A Zunger. A new method for diagonalising large matrices. *Journal of Physics A: Mathematical and General*, 18(9):1343–1359, jun 1985.

[155] M.J. Rayson and P.R. Briddon. Rapid iterative method for electronic-structure eigenproblems using localised basis functions. *Computer Physics Communications*, 178(2):128–134, jan 2008.

[156] Henk A. Van der Vorst and Kees Dekker. Conjugate gradient type methods and preconditioning. *Journal of Computational and Applied Mathematics*, 24(1-2):73–87, nov 1988.

[157] G. Kresse and J. Furthmüller. Efficient iterative schemes forab initiototal-energy calculations using a plane-wave basis set. *Physical Review B*, 54(16):11169–11186, oct 1996.

[158] James C. Phillips. Energy-band interpolation scheme based on a pseudopotential. *Physical Review*, 112(3):685–695, nov 1958.

[159] James C. Phillips and Leonard Kleinman. New method for calculating wave functions in crystals and molecules. *Physical Review*, 116(2):287–294, oct 1959.

[160] Marvin L. Cohen and Volker Heine. The fitting of pseudopotentials to experimental data and their subsequent application. In *Solid State Physics*, pages 37–248. Elsevier, 1970.

[161] Dominik Bogdan Jochym. *Development of non-local density functional methods.* phdthesis, Durham University, 2008.

[162] N. Troullier and JoséLuís Martins. A straightforward method for generating soft transferable pseudopotentials. *Solid State Communications*, 74(7):613–616, may 1990.

[163] Martin Fuchs and Matthias Scheffler. Ab initio pseudopotentials for electronic structure calculations of poly-atomic systems using density-functional theory. *Computer Physics Communications*, 119(1):67–98, jun 1999.

[164] D. R. Hamann, M. Schlüter, and C. Chiang. Norm-conserving pseudopotentials. *Physical Review Letters*, 43(20):1494–1497, nov 1979.

[165] N. Troullier and José Luriaas Martins. Efficient pseudopotentials for plane-wave calculations. *Physical Review B*, 43(3):1993–2006, jan 1991.

[166] Andrew M. Rappe, Karin M. Rabe, Efthimios Kaxiras, and J. D. Joannopoulos. Optimized pseudopotentials. *Physical Review B*, 41(2):1227–1230, jan 1990.

[167] David Vanderbilt. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Physical Review B*, 41(11):7892–7895, apr 1990.

[168] P. E. Blöchl. Projector augmented-wave method. *Physical Review B*, 50(24):17953–17979, dec 1994.

[169] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59(3):1758–1775, jan 1999.

[170] Carsten Rostgaard. The Projector Augmented-wave Method.

[171] Hendrik J. Monkhorst and James D. Pack. Special points for brillouin-zone integrations. *Physical Review B*, 13(12):5188–5192, jun 1976.

[172] Hans Hellman. Einführung in die quantenchemie. *Franz Deuticke, Leipzig*, page 285, 1937.

[173] R. P. Feynman. Forces in molecules. *Physical Review*, 56(4):340–343, aug 1939.

[174] Frank E. Curtis and Xiaocun Que. A quasi-newton algorithm for nonconvex, nonsmooth optimization with global convergence guarantees. *Mathematical Programming Computation*, 7(4):399–428, may 2015.

[175] Y. H. Dai and Y. Yuan. A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization*, 10(1):177–182, jan 1999.

[176] M.I.J. Probert. Improved algorithm for geometry optimisation using damped molecular dynamics. *Journal of Computational Physics*, 191(1):130–146, oct 2003.

[177] G. P. Kerker. Efficient iteration scheme for self-consistent pseudopotential calculations. *Physical Review B*, 23(6):3082–3084, mar 1981.

[178] D. D. Johnson. Modified broyden's method for accelerating convergence in self-consistent calculations. *Physical Review B*, 38(18):12807–12813, dec 1988.

[179] C. G. Broyden. A class of methods for solving nonlinear simultaneous equations. *Mathematics of Computation*, 19(92):577–577, 1965.

[180] D. Raczkowski, A. Canning, and L. W. Wang. Thomas-fermi charge mixing for obtaining self-consistency in density functional calculations. *Physical Review B*, 64(12), sep 2001.

[181] Péter Pulay. Convergence acceleration of iterative sequences. the case of scf iteration. *Chemical Physics Letters*, 73(2):393–398, jul 1980.

[182] Paul Adrien Maurice Dirac. On the theory of quantum mechanics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 112(762):661–677, oct 1926.

[183] Enrico Fermi. Sulla quantizzazione del gas perfetto monoatomico. *Rendiconti Lincei*, 3:145–149, 1926.

[184] C. L. Fu and K. M. Ho. First-principles calculation of the equilibrium ground-state properties of transition metals: Applications to nb and mo. *Physical Review B*, 28(10):5480–5486, nov 1983.

[185] M. Methfessel and A. T. Paxton. High-precision sampling for brillouin-zone integration in metals. *Physical Review B*, 40(6):3616–3621, aug 1989.

[186] Peter E. Blöchl, O. Jepsen, and O. K. Andersen. Improved tetrahedron method for brillouin-zone integrations. *Physical Review B*, 49(23):16223–16233, jun 1994.

[187] S. Zhang and John Northrup. Chemical potential dependence of defect formation energies in GaAs: Application to ga self-diffusion. *Physical Review Letters*, 67(17):2339–2342, oct 1991.

[188] Chris G. Van de Walle, D. B. Laks, G. F. Neumark, and S. T. Pantelides. First-principles calculations of solubilities and doping limits: Li, na, and n in ZnSe. *Physical Review B*, 47(15):9425–9434, apr 1993.

[189] Christoph Freysoldt, Blazej Grabowski, Tilmann Hickel, Jörg Neugebauer, Georg Kresse, Anderson Janotti, and Chris G. Van de Walle. First-principles calculations for point defects in solids. *Reviews of Modern Physics*, 86(1):253–305, mar 2014.

[190] M Leslie and N J Gillan. The energy and elastic dipole tensor of defects in ionic crystals calculated by the supercell method. *Journal of Physics C: Solid State Physics*, 18(5):973–982, feb 1985.

[191] G. Makov and M. C. Payne. Periodic boundary conditions in ab initio calculations. *Physical Review B*, 51(7):4014–4022, feb 1995.

[192] Ismaila Dabo, Boris Kozinsky, Nicholas E. Singh-Miller, and Nicola Marzari. Electrostatics in periodic boundary conditions and real-space corrections. *Physical Review B*, 77(11), mar 2008.

[193] Xu Sun, You song Gu, Xue qiang Wang, and Yue Zhang. Defects energetics and electronic properties of li doped ZnO: A hybrid hartree-fock and density functional study. *Chinese Journal of Chemical Physics*, 25(3):261–268, jun 2012.