2013

Analysis of defects in GaN using Hybrid Density Functional Theory

Ibrahima Castillo Diallo
Virginia Commonwealth University

Analysis of Defects in GaN using Hybrid Density Functional Theory

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Physics / Applied Physics at Virginia Commonwealth University.

By

I. Castillo Diallo

B.S. in Physics

Minor in Mathematics

Virginia Commonwealth University, 2011

M.S. in Physics/Applied Physics

Virginia Commonwealth University, 2013

Major Director:

Demchenko Denis O., Ph.D., Assistant Professor, Department of Physics

Virginia Commonwealth University

Richmond, Virginia, 23284

April 29, 2013

Table of Contents

List of Figures

Abstract

ANALYSIS OF DEFECTS IN GaN USING HYBRID DENSITY FUNCTIONAL THEORY

By Ibrahima Castillo Diallo

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2013

Major Director: Denis O. Demchenko, Ph.D., Assistant Professor, Department of Physics

In this thesis, we first present a brief overview of various theoretical approaches used to examine the electronic structure of defects in GaN. Using the recently developed hybrid density functional theory (HSE06) along with the experimental measurements, we propose a new explanation of the nature of the yellow luminescence band in carbon-doped GaN. We conduct a systematic study of electronic and optical properties of defects (Carbon, Oxygen, Silicon related) that are candidates for the origin of yellow luminescence. We show that the C_N-O_N complex is significantly more likely to form compared to isolated carbon configurations. In contrast to the properties of the isolated carbon acceptor, calculated defect levels and optical transitions involving deep level of the C_N-O_N complex agree quite well with our thermal luminescence quenching data as well as with the experimentally measured C-doped GaN luminescence spectra. Hence, the C_N-O_N complex, rather than isolated C impurity, is more likely to resolve a long-standing problem of the yellow luminescence in GaN.

0. Introduction

The intentional incorporation of impurities in small concentrations and formation of native defects, such as recombination centers or compensation defects[1], are essential to controlling the electrical and optical properties of wide-gap semiconductors[2]. However, achieving such control is far from being trivial and a comprehensive knowledge of the fundamental techniques that control doping and formation of defects and impurities are hence required. Due to recent breakthroughs in algorithmic developments as well as the unceasing improvement of large-scale computational tools, first principles calculations are now reaching good levels of precision[3] in describing the properties of impurities and native defects in semiconductors.

One of the first theoretical method used to describe localized defects in semiconductors involved Green's functions[4,5,6].These calculations depict the Hamiltonian of the perfect crystal and the Hamiltonian of the localized defect within the bulk[7]. Due to the extremely large number of particles in the system, the computation of such Hamiltonians becomes impossible. Another early theoretical method that describes the electronic structure of defects in solids is the effective mass theory (EMT). The EMT is an analytical theory that uses experimental effective masses as input parameters for the calculation of simple band structure. The defect is inserted in the calculation as a perturbing spherically symmetric potential. However, the results of EMT remain semi-quantitative since it only provides an approximate description of the defect levels and does not address the formation energy of the defect.

Kohn Sham Density Functional Theory (DFT)[8] has proven to be a prevailing tool for analyzing and understanding defect energetics and electronic structure in semiconductors. Good progress for approximating the crucial exchange-correlation (*xc*) energy from the Kohn-Sham approach has been made in the last decades. One of the most relevant formalisms for the analysis of the electronic structure of spin systems is the Local Spin Density Approximation (LSDA)[9]. In solids, for regions where the charge density is assumed to be slowly varying, the *xc*-energy is defined as the integral over all space with the *xc*-energy at each point being considered the same as in a uniform electron gas of the same charge density[10]. The LSDA has provided acceptable results in crystal structure, bond lengths and vibrational frequencies[10,11] in both homogenous and inhomogeneous systems. However, the LSDA leads to major drawbacks such as the inability to describe the magnetic configuration of transition metals, the lack of cancelation of self-interaction which is crucial for strongly localized states and the severe underestimation of the band gap in semiconductors and insulators. Such shortcomings have stimulated ideas for the

creation of improved functionals such as non-empirical Gradient Approximations[12] (GGA). Several sophisticated adaptations of GGA have been developed in the last decades[13,14,15,16], but the most commonly used version is the Perdew-Burke-Ernzerhof (PBE)[17] approximation that employs both the density and its gradient at each point in space. Both GGA and LDA were derived in the limits of the homogeneous electron gas theory and are therefore expected to be useful for systems with slowly varying charge densities[18]. These complex formalisms have provided satisfactory results for the computation of molecular binding energies, atomic ionization energies and geometrical structure of molecules and solids[19,20]. The partial error cancellation in the exchange and correlation energy parts integrated in both first-principles calculations methods provided the accuracy required for DFT to be used in Solid States physics as well as in chemistry[10,21]. Nevertheless, the underestimation of the band gap in solids remain one of the major drawbacks of both LDA and GGA formalisms.[22,23]

To remedy the band gap problem and several other unphysical results of LDA and GGA, much effort has been put into the improvement of the *xc* parameter[24,25,26]. One of the most fruitful approximations in the computation of band gap is described by a combination of Green Function and screened Coulomb interaction, often referred to as the GW method[27]. However, the GW method happens to be computationally expensive for complex systems. Alternative approach that rectified the band gap problem was the construction of a hybrid functional theory (HFT) that contains a mixture of a certain amount of non-local Fock exchange to a part of *local/semilocal* LDA/PBE exchange[28,29,30]. Due to the periodicity of the lattice that generates a long range Hartree-Fock (HF) exchange interactions, the use of hybrid functional in Solid States Physics has been inadequate[18,31]. Significant progress into reducing the computational effort of calculating long range Fock exchange has been achieved by the creation of the HSE03 (Heyd-Scuseria-Ernzherof) functional[32]. This newly derived formalism separates the Fock exchange into short-range and long range components. The short range exchange energy is made of 25% of HF and 75% of PBE, while the long range part is entirely represented by the PBE exchange energy.[32]24 Such modifications would cause major corrections to the electronic properties of the system and can therefore be used to compute precise band gaps, bulk moduli and atomization energies of solids including semiconductors and metals[32,11 ,33,34,35]. An in-depth study of electronic structure of solids has not yet reached its peak with the development of HSE03 formalism. More detailed analysis of energetics of defects in semiconductors[36,37], vibrational frequencies of lattice[38,39], phonons –dispersion relations of the group IV elemental solids[18] and optical properties of semiconductors[40] have been recently achieved with the creation of the HSE06[41] approximation. Nearly perfect agreement of HSE06 electronic structure of semiconductors and insulators with experiment can be obtained by tuning the fraction of Fock exchange. With such precise method, direct comparison of calculated defect properties with experiment is finally possible.

In this thesis, these new methods of calculation will be applied to the analysis of properties of defects in GaN. Technological advancements in GaN doping have made it a suitable material for applications in recently developed blue and green light-emitting diodes[42], the blue-emitting GaN-based lasers[43,44] and solar cells.[45] Understanding the optical and electronic properties of defects in GaN is of great importance for evaluating the degree to which they affect the devices' performance. The most notoriously controversial defect-induced optical transition in GaN is centered around 2.2-2.3 eV and is often referred to as the yellow luminescence (YL).[46,47,48] This YL band is usually observed in *n*-type GaN,[46,49] both for undoped samples[50] and samples containing carbon impurities.[51,52,53] Nevertheless, the microscopic origin of the YL band has been debated for almost 30 years, and the exact attribution of the YL band to specific defects in

GaN has been unclear. Early works attributed the YL band to the formation of the V_{Ga}-C_{Ga} complex.[47] With the development of DFT, it has been suggested that Ga vacancies, V_{Ga}, both isolated and bound into a complex with oxygen donors, O_N, (and possibly silicon donors, Si_{Ga}) are responsible for the observed YL.[54,55] These theoretical results have found some experimental support where positron annihilation experiments showed apparent correlation of the YL intensity with the concentration of Ga vacancies.[56] On the other hand, experimental findings have also indicated that lattice defects alone do not cause YL, rather H, C, and O, possibly bound into complexes, produce the observed YL band.[53] Most recently it has been suggested that carbon substituting for nitrogen C_N creates a deep acceptor in GaN, which may be responsible for the YL band.[57]

The goal of the first part of this thesis is to provide an overview of the methodology used to perform first-principles calculations for defects and impurities in the bulk. Subsequently, in Section II, we intend to use this formalism to investigate Photoluminescence (PL) in GaN and we finally demonstrate that the deep donor complex $(C_N\text{-}O_N)^0$ explains the microscopic mechanism of the YL in GaN.

1. Doping in Semiconductors

The introduction of defects in a host crystal modifies the characteristics of the material in various ways. Because of the multiplicity of imperfections that can occur within the lattice, we will first describe ways of classifying them and later discuss the main differences between shallow and deep impurities.

1.1 Creation of Native Point Defects

In this section, the brief analysis on some characteristics of defects in semiconductors is based on the detailed review written by S. T. Pantelides (1979)[58]. Native defects are inherent imperfections that are formed within the "pure" lattice and can be point defects, which correspond to the improper location of atoms, planar defects, which describe misplaced planes of atoms and line defects which correspond to misplaced line of atoms. Since we only investigated point defects as lattice-type defects in GaN, the characteristics of either planar or line defects shall not be discussed in this thesis. Native point defects usually occur in vacant and interstitial sites. In case of vacancies, atoms are missing from their regular atomic site.

Interstitial point defects describe extra-atoms that occupy interstitial sites in the crystal. In addition to native point defects, foreign impurities may occur inside the crystal and can be classified in terms of their physical crystal sites.

1.2 Incorporation of impurities in the crystal lattice

Foreign impurities may occur in interstitial locations and substitutional sites in which case the impurity replaces the host atom. Substitutional atoms that generally have more valence electrons than the host atoms are called donors since they must donate electrons to the host atoms in order to fulfill local bonding requirements. While on the other hand, substitutional atoms that possess less valence electrons are called acceptors since they must accept electrons from the host atoms in order to bond with their nearest neighbors. In the case of compound semi-conductors, the same foreign impurity could be characterized as a donor or acceptor depending on which host atom it substitutes. Impurities may be called shallow when their respective energy levels are very

close to the conduction or valence band. On the contrary impurities are called deep when their respective energy levels are far from the band edges. Shallow and deep impurities play an important role in the properties of a given material and will be discussed in the following section. In addition to isolated defects and impurities, one must also notice that it is sometimes possible for neighboring defects and impurities to interact and form complexes. The simplest situation is a complex pair consisting of two isolated impurities occupying neighboring sites, a vacancy defect and the nearest neighbor impurity, and two vacancies defects in neighboring sites. More details in swirl defect complexes in Si and dark-like defect complexes in GaAs-GaAlAs can be found in reviews written by De Kock (1973)[59] and Petrov and Hartmann (1973)[60], respectively.

1.3 Shallow and deep impurities in semiconductors

In this thesis, we use first-principles electronic structure calculations of defects in GaN bulk. In other words, we are solving the Schrodinger equation in GaN lattice containing a defect, using periodic boundary conditions. The use of periodically repeated supercells[61] provides a physically appealing description of the defect and its closest neighbors. These supercells are composed of numerous primitive unit cells which contain a single defect. Even though the supercell technique precisely describes the local arrangements of bonding between atoms and the defect crystal structure, it unfortunately introduces some drawbacks that need to be corrected such as the divergence of the Coulomb energy for charged defects[61], the band-filling error[62] and the potential alignment[62] for charged impurities.

The role that most impurities and defects play in a given semiconductor often depends on the concentration in which it can be incorporated in the material and the kind of localized states they create in the band gap. Materials can be grown as either p-type, where there are more acceptors than donors or as n-type in which the number of donors exceeds the number of acceptors. By setting up the concentrations of shallow donors and acceptors as non-uniform within a given semiconductor, one can create a p-n junction which consists of a region doped with acceptors (p-type) adjacent to a region doped with donors (n-type). As one approaches the junction, electrons from the conduction band (CBM) jump down to the neutral acceptor states of the p-type sample and thus become negatively charged. On the other hand, neutral donors from the n-type sample donate electrons to the valence band (VBM) and become positively charged. An electric field is created across the junction due to this charge distribution. A device created by such arrangement of defects has led to the creation of modern electronics.

There are fundamental differences in defects' properties depending on the proximity of the defects to the band edges. Shallow levels are characterized by their extreme closeness to the band edges at room temperature. At such shallow levels, impurities have ionization energies comparable to $k_B T$ and therefore will play a crucial role in controlling conductivity. The case of neutral shallow donor impurities in semiconductors requires careful investigation because of its weakly localized characteristic. Based on the effective mass theory, the wave functions of shallow defects are Hydrogen-like and thus relatively spread out in real space. However within the supercell formalism, our supercell is not large enough to completely encompass such widespread wave function. Supercells that could contain a weakly localized wavefunction of a shallow impurity contain tens of thousands atoms and are therefore computationally prohibitive. When a shallow defect is computed in relatively small supercell, the conduction band level shifts down and becomes the perturbed host state (PHS) while the impurity band level moves up into the conduction band. The electron located at the impurity level will drop to the perturbed host

state and such unphysical occurrence gives rise to completely delocalized charge densities. Our neutral system will therefore be composed of a positive ion surrounded by completely delocalized electrons. Such situation is quite similar to a positive ion sitting in uniform negatively charged compensating background. In order to correct the total energy of the finite supercell due to such unphysical interactions, one must include a special correction which will be later discussed in Section 5.3.2.2.

On the other hand, deep impurities have localized wave functions and therefore provide levels inside the band gap that could increase the probability of recombination between the electrons and holes. In addition to emitting phonons during the recombination process in the case of deep impurities, photons may also be produced and photoluminescence could therefore be measured.

2. Density Functional Theory (DFT)

2.1 Foundation and Importance

The understanding of the electronic structure of defects in semiconductors is based upon theoretical methods of statistical and quantum mechanics. If one wishes to discuss the properties of interacting defects within the bulk, it is natural to consider the time independent Schrödinger equation for N electrons with M ions,

$$\hat{H}_{Tot}\Psi_{Tot} = E_{Tot}\Psi_{Tot}, \tag{2.1.1}$$

where E_{Tot} represents the total energy of the system and the many-body wavefunction $\Psi_{Tot} = \Psi(\vec{r}_1,...,\vec{r}_N,\vec{s}_1,...,\vec{s}_N;\vec{R}_1,...,\vec{R}_M,\vec{S}_1,...,\vec{S}_M)$ gives all the necessary information about the system. The position and spin of the kth electron are respectively denoted by $\vec{r}_k,\vec{s}_k$ and the position and spin components of the Kth nuclei are represented by $\vec{R}_K,\vec{S}_K$. The Hamiltonian of our previous equation describes the correlated motion of the electrons and nuclei in our system and is represented by:

$$\hat{H}_{Tot} = -\frac{\hbar^2}{2m_e}\sum_{k=1}^{N}\nabla_k{}^2 - \sum_{K=1}^{M}\frac{\hbar^2}{2M_K}\nabla_K{}^2 + \sum_{k=1}^{N}\sum_{l>k}^{N}\frac{e^2}{|\vec{r}_k-\vec{r}_l|} + \sum_{\substack{k=1\\K=1}}^{N,M}\frac{Z_K e^2}{|\vec{r}_k-\vec{R}_K|} + \sum_{K=1}^{M}\sum_{L>K}^{M}\frac{Z_K Z_L e^2}{|\vec{R}_K-\vec{R}_L|} \tag{2.1.2}$$

Where electrons with charge $-e$ and mass m_e are represented by the lower case subscripts and nuclei with charge $+Z_K e$ and mass M_K, are denoted by upper case subscripts. In our molecular Hamiltonian, we suppose that the motion of both the electrons and the nuclei are treated strictly non-relativistically. Although the relativistic corrections of the kinetic energy are completely neglected, one must not forget that for heavy atoms, *relativistic properties modify the structure of the Hamiltonian to an extent which is even noticeable in molecular bonds*.[63] Furthermore, the description of $\hat{H}_{Tot}$ will be restricted to a zero temperature formalism.

The terms of this quite complex molecular Hamiltonian describe

- The kinetic energy operators for the electrons and ions, $\hat{T}_k$ and $\hat{T}_K$:

$$\hat{T}_k = -\frac{\hbar^2}{2m_e}\sum_{k=1}^{N}\nabla_k{}^2 \tag{2.1.3}$$

$$\hat{T}_K = -\sum_{K=1}^{M} \frac{\hbar^2}{2M_K} \nabla_K{}^2 \tag{2.1.4}$$

- The potential due the electron-electron interaction and the potential acting on the electrons due to the nuclei, $\hat{V}_{ee}$ and $\hat{V}_{ie}$ respectively:

$$\hat{V}_{ee} = \sum_{k=1}^{N} \sum_{l>k}^{N} \frac{e^2}{|\vec{r}_k - \vec{r}_l|} \tag{2.1.5}$$

$$\hat{V}_{ie} = \sum_{\substack{k=1 \\ K=1}}^{N,M} \frac{Z_K e^2}{|\vec{r}_k - \vec{R}_K|} \tag{2.1.6}$$

- And the ion-ion interaction potential, $\hat{V}_{ii}$:

$$\hat{V}_{ii} = \sum_{K=1}^{M} \sum_{L>K}^{M} \frac{Z_K Z_L e^2}{|\vec{R}_K - \vec{R}_L|} \tag{2.1.7}$$

A rather rough estimate of the computational complexity of such Schrodinger equation is to visualize the fairly vast scale of our resulting Hamiltonian operator. For a typical system, the number of electrons is approximately ten times greater than the number of ions and the total amount of M ions is quite close to Avogadro's number, where $M \approx N_A \approx 6.02 \times 10^{23} mol^{-1}$.[64] Hence the total number of variables is to the order of 10^{24}. For *N* electrons and *M* ions, the many-body wavefunction reaches the degree of freedom of *4N+4M* and therefore the computation of the full many-body wavefunction remains impossible for real systems with more than few electrons. Even though analytical solutions of the Schrödinger equations can be solved for a few very simple systems[65], our cases of interest involve optical properties of systems that contain tremendous amount of electrons and also thermodynamic transition levels associated with deep and shallow defects. A complete description of such systems with quantum mechanics is quite complex and thus one requires the use of an approximate, more simplified representation of our initial system.

A convenient way to reduce the scale of our Hamitonian is to make use of the Born-Oppenheimer approximation[66], where the nuclear motion can be separated from the electronic motion. Basically in our system, since the nuclei are much heavier than the electrons $(M_K \gg m_k)$, the inverse mass of the Kth nuclei $\frac{1}{M_K}$ becomes extremely small and hence the kinetic energy operator for ions $\hat{T}_K = -\sum_{K=1}^{M} \frac{\hbar^2}{2M_K} \nabla_K{}^2$ becomes the only negligible term in our many-body Hamiltonian. From such approximation, the electrons will configure themselves as if the ions were static and those fixed ions will not affect the states of the electrons except as a potential $\hat{V}_{ie}$.[64] Consequently, the interaction potential between "fixed" Kth and Lth ions will become a constant classical electrostatic potential E_{KL}. Hence, by neglecting the kinetic energy of ions and setting their potential as a constant electrostatic potential, the many-body Hamiltonian becomes the electronic Hamiltonian $\hat{H}_{elec}$, in which the position of the nuclei are only *parameters*[10] where:

$$\hat{H}_{elec} = \hat{T}_k + \hat{V}_{ee} + \hat{V}_{ie} + E_{KL}, \tag{2.1.8}$$

By a *parametric dependence*, we imply that the electronic wavefunction Ψ_{elec} of the new obtained Schrödinger equation $\hat{H}_{elec}\Psi_{elec} = E_{elec}\Psi_{elec}$, is a *different function of the electronic coordinates* for different locations of the nuclei.[67] In order to avoid the messiness of the units of $\hat{H}_{elec}$, we adopt Hartree atomic units $\hbar = e = m_e = 4\pi\varepsilon_0 = 1$, then the kinetic energy operator for electrons becomes:

$$\hat{T}_{ee} = -\frac{1}{2}\sum_{k=1}^{N}\nabla_k^2 \tag{2.1.9}$$

The electron-electron interaction potential $\hat{V}_{ee}$ and the potential acting on the electrons due to nuclei $\hat{V}_{ie}$,

$$\hat{V}_{ee} = \sum_{k=1}^{N}\sum_{l>k}^{N}\frac{1}{\left|\vec{r}_k - \vec{r}_l\right|} \tag{2.1.10}$$

$$\hat{V}_{ie} = \sum_{\substack{k=1 \\ K=1}}^{N,M}\frac{Z_K}{\left|\vec{r}_k - \vec{R}_K\right|} \tag{2.1.11}$$

Even though we have marginally reduced the number of variables in the general Hamiltonian, the obtained electronic Hamiltonian still achieves frightening proportions.

One of the earliest and traditional formalisms that approximates the many-body wavefunction Ψ_{Tot} was derived by Hartree[68] in 1928 who rewrote Ψ_{Tot} as a product of single particle functions, i.e,

$$\Psi_{Tot}(\vec{r}_1, \vec{r}_2, ..., \vec{r}_N) = \Psi(\vec{r}_1)\Psi(\vec{r}_2)....\Psi(\vec{r}_N)$$

Each one of the obtained wavefunctions $\Psi_k(\vec{r}_k)$ satisfies a one electron Schrodinger equation, and the Hartree Hamiltonian yields:

$$\hat{H}^{Har} = \hat{T}_k + \hat{V}_{ee}^{Har} + \hat{V}_{ie} \tag{2.1.12}$$

which can be rewritten as $\hat{H}^{Har} = -\frac{1}{2}\sum_{k=1}^{N}\nabla_k^2 + \sum_{k=1}^{N}\sum_{l>k}^{N}\int_{\Omega}\frac{\left|\Psi_k(\vec{r}_k)\right|^2\left|\Psi_l(\vec{r}_l)\right|^2}{\left|\vec{r}_k - \vec{r}_l\right|}d^3\vec{r}_k d^3\vec{r}_l + \sum_{\substack{k=1 \\ K=1}}^{N,M}\frac{Z_K}{\left|\vec{r}_k - \vec{R}_K\right|}$ (2.1.13)

where $\hat{V}_{ee}^{Har} = \sum_{k=1}^{N}\sum_{l>k}^{N}\iint\frac{\left|\Psi_k(\vec{r}_k)\right|^2\left|\Psi_l(\vec{r}_l)\right|^2}{\left|\vec{r}_k - \vec{r}_l\right|}d^3\vec{r}_k d^3\vec{r}_l$ (2.1.14)

is the newly derived potential term and arises solely from the electrostatic energy (Coulomb's integral). In 1930, Fermi statistics[69,70] became incorporated into the Hartree formalism by Fock and Slater[71], where the total wave function of the N-electron system can be approximated by an anti-symmetrized product of N orthonormal spin orbitals $\chi_\alpha(\vec{r}_k)$[72]. The newly derived formalism became known as the HF method and the spin orbitals $\chi_\alpha(\vec{r}_k)$ represent one-electron wavefunctions that result from the product of a spatial orbital $\chi_{a_\alpha}(\vec{r}_k)$ with a spin part $\left|m_s\right\rangle$ of the spin orbital χ_α. Further analysis of the HF formalism is postponed until it can be thoroughly discussed in Chapter II. Since the HF model describes a given system as the combination of anti-symmetric one-electron wave functions, the number of variables from the wave functions remains quite bothersome.

In the sixties, a fantastic breakthrough finally made it possible to explain an N-electrons system with the introduction of a three-dimensional electron density $n(\vec{r})$. Contrary to the electronic wavefunction, the electron density does not increase spatially as three times the number of electrons but rather remains invariant regardless the size of the system[73]. The very first approach based on the density of the electrons in the system $n(\vec{r})$, was done by Thomas[74] and Fermi[75] in the late twenties. In this new scheme, the motion of electrons is completely uncorrelated and their corresponding kinetic energies can be described as a "functional" of a local density *based on free electrons density* $n(\vec{r})^{5/3}$ in a homogeneous gas.[65] However, in the original Thomas-Fermi (TF) method, the exchange and correlation energies among electrons were completely neglected. In 1930, the local exchange effects[76] were proposed by Dirac and became incorporated into the TF original formalism. In the mid-fifties, Slater[77] proposed a simplification of the HF potential in the $X\alpha$ scheme by setting up an *average potential* in which all the electrons are in motion and argues that this potential in a system of varying densities could be approximated with *a local dependence proportional to* $n(\vec{r})^{1/3}$.[65] Countless improvements of the HF method have been made in the last decades and have been essential to the development to modern density functional theory. In fact, those traditional HF wavefunctions can be used to compute fairly precise results for smaller systems, providing benchmarks for further developing density functional, *which can hence be applied to larger systems*.[78] One must wait until the mid-sixties to finally obtain a formalism that does not start with too crude approximations[79], and yet provides a certain balance between accuracy and computational cost. The first thorough and complete proof of the existence of Density Functional Theory (DFT) was given by Hohenberg and Kohn in 1964.[80] These authors demonstrated that the ground state electron density of a system implicitly contains all the information included within its quite complex ground state many-electrons wavefunction.[73,80] In other words, all characteristics of the systems can be considered as *functionals* of the ground state density.[10] Furthermore, they argue that, for any given external potential $\hat{V}_{ext}$ (which corresponds to the potential acting on the electrons due to nuclei $\hat{V}_{ie}$) if a "universal functional" for the total energy $E_{HK}[n(\vec{r})]$ of our system is known, then by *minimizing the total energy of the system, with respect to the ground state density* $n_0(\vec{r})$, *one would determine the exact ground state energy*.[10] Such electron density based method simplifies the calculations of defects' electronic properties in the bulk. DFT allows much larger systems to be solved by ab-initio methods, while retaining much of their precision. Nonetheless, precision is a quite relative term. Even though, theoretically, DFT is an exact functional, its actual performance relies on the quality of the approximate density functionals employed. DFT is not just another method of *parametrizing empirical results*, but is rather a *meticulous way of analyzing any interacting system, by mapping exactly to a much easier-to-solve non-interacting system*.[78] To appreciate the full extent of HK theorems, and the beauty of the DFT method, one must understand the meaning of functionals, and a brief analysis of functionals is given in the following Section 2.2.

2.2 Functional and Functional derivatives

The mathematical and physical explanations of functional and functional derivatives are crucial if one needs to understand the quantum mechanics of interacting electrons in terms of DFT and Hybrid Functional Theory (HFT). In this chapter, we briefly discuss the central role of

functionals and functional derivatives in the formalisms we will be describing in the following chapters.

A functional is closely related to the more familiar concept of a function and one recalls that a function *f* is defined to be a mapping of a variable *y* to a number *f(y)*. From a mathematical perspective, one can define a functional as a function of a function. In other words, a functional $F[f(y)]$ assigns a unique number to an entire function and is therefore a mapping of a function *f* onto a value or number.[72] A fairly simple example of a functional can be conceptualized by looking at the total charge enclosed Q in a closed surface S bounded by a given volume Ω often used in classical electromagnetism,

$$Q = \int_{\Omega} n(\vec{r}) d^3 r = Q[n(\vec{r})] \tag{2.2.1}$$

where $n(\vec{r})$ is the total charge density of the system. From this functional, we notice that $Q[n(\vec{r})]$ is a rule for going from a function $n(\vec{r})$ to a number Q. The square bracket notation $Q[n(\vec{r})]$ indicates that Q depends on $n(\vec{r})$ everywhere in the volume Ω. Moreover, $Q[n(\vec{r})]$ is also called *local-functional* since the functional does not depend on its gradient, Laplacian or other higher-order derivatives. More detailed analysis of functionals are given by Volterra[81] (1959), and Parr and Yang[72] (1989).

Differentiation of functionals is an extension of the concept of partial differentiation for multi-variable functions[82]. Functional derivatives basically allow us to study how a functional changes with respect to variation of a function *f* at the point *y*. Let a function *f(y)* be defined over a specific interval $[y_{\min}, y_{\max}]$, and be a subject to an arbitrary small perturbation $\delta f(y)$ that is $f \to f + \delta f$. Then the perturbation $\delta f(y)$ is also defined over the same interval $[y_{\min}, y_{\max}]$. For some functional $F[f(y)]$, the value of $F[f + \delta f]$ can be therefore approximated by using a Taylor's[83] expansion in powers of the perturbation $\delta f(y)$,

$$F[f + \delta f] = F[f] + \int_{y_{\min}}^{y_{\max}} \Gamma_1(y)\delta f(y) dy + \frac{1}{2!}\int_{y_{\min}}^{y_{\max}} \Gamma_2(y, y_1)\delta f(y)\delta f(y_1) dy_1 dy_2 + \ldots \tag{2.2.2}$$

$$\text{Hence } F[f + \delta f] - F[f] = \int_{y_{\min}}^{y_{\max}} \Gamma_1(y)\delta f(y) dy + \frac{1}{2!}\int_{y_{\min}}^{y_{\max}} \Gamma_2(y, y_1)\delta f(y)\delta f(y_1) dy_1 dy_2 + \ldots \tag{2.2.3}$$

Where $\Gamma_1(y) = \dfrac{\delta F[f]}{\delta f(y)}$ is the first Taylor's expansion coefficient and describes the rate of change of the functional or the functional slope for a small variation of *f* at *y*. Correspondingly, $\Gamma_2(y, y_1) = \dfrac{\delta^2 F[f]}{\delta f(y)\delta f(y_1)}$ is the second Taylor's expansion coefficient and described the rate of change of the functional when *f* is simultaneously subject to small perturbations at *y* and y_1. Since the functional derivative only measures the first order change in a functional, the second integral of Eq. (2.2.3) is neglected; hence the quantity $\dfrac{\delta F[f]}{\delta f(y)}$ becomes the *functional derivative of F with respect to f* at the point *y*.[72]

Now that we have briefly described what is a functional and how to differentiate functionals, we can further analyze HK theorems that stipulate that the ground-state energy $E_{HK,0}[n(\vec{r})]$ of *any system of interacting particles* is a functional of the particle's density $n(\vec{r})$ and minimizing that energy will determine the *exact ground-state density* $n_0(\vec{r})$.[80]

2.3 Hohenberg-Kohn (HK) theorems

One may argue that the very core of the entire field of DFT rests on the two HK theorems. In the original HK paper, they proved that any characteristics of the many-electron system can be regarded *as a functional of the ground state density*.[10]
By definition, in a given electronic system, the total number of electrons *per unit volume in a given state*[72] is defined as the electron density for that state where

$$N = \int n(\vec{r}) d^3 r \tag{2.3.1}$$

Thus the probability of finding one electron within volume element $d^3 r_i ... d^3 r_N$ with arbitrary spin, while the other *N-1* electrons have arbitrary and spatial coordinates is given by:

$$n(\vec{r}_1) = N \int ... \int \left| \Psi(\vec{x}_1, \vec{x}_2, ..., \vec{x}_N \right|^2 ds_1 d^3 r_i ... d^3 r_N \tag{2.3.2}$$

Now let's consider N interacting electrons subjected to an external potential $\hat{V}_{ext}$, where the corresponding Hamiltonian can be regarded as:

$$\hat{H}^{HK} = \hat{T}_{ee} + \hat{V}_{ee} + \hat{V}_{ext} \tag{2.3.3}$$

Since $\hat{V}_{ext}$ completely determines the Hamiltonian and the total number of electrons is represented by *N*, we can therefore conclude that $\hat{V}_{ext}$ and *N* establish all properties for the ground-state of the system. The first theorem established by Hohenberg and Kohn substitutes $\hat{V}_{ext}$ and *N* for the use of the electronic density $n(\vec{r})$ as basic variable. The correlation between the external potential and the electronic density can be regarded as follow:
For any system of N interacting particles subjected to an external potential, the external potential is determined *uniquely*, *within a trivial additive constant*, *by the ground state electronic density* $n_0(\vec{r})$.[10,72,80] In other words, the ground state density determines all properties of the ground state. However, even though the first HK theorem rigorously proves the existence of a functional of the electron density that can be used to solve the many-electron system, the theorem says nothing about the actual functional.
Hence, we are still left with the initial problem of many interacting electrons subjected to an external potential.
Fortunately, the second HK theorem describes a crucial property of the functional. It states that by deriving the true or exact functional form, one can fluctuate the electron density until the energy of the functional is minimized, *hence giving a prescription for obtaining the relevant electron density*.[84] To put it differently, for any change in the external potential, the particles in a given system will react, so that their energy is minimized and this response would be unique. One can exemplify the corollary of the second HK theorem by looking at how specifically the orbitals contract if the nuclear charge is increased in a given atom. Another example would be the explicit positional shift of the wavefunction and density as atoms in a given molecule

separate or get closer to one another. By explicit positional shift, we are not referring to a mere additive constant but we are just noticing a direct correlation between the electronic density and the atom structure. For readers who are interested by the explicit proofs of such two groundbreaking theorems, they can refer to Parr and Yang (1989) and Engel and Dreizler (2009)[63]. A more intuitive approach to functional would be the introduction of a *two-step minimization* procedure described by Levy[85,86] and Lieb[87,88] and their approach is thoroughly explained by R. M. Martin (2004)[10]. In the following chapters, we will discuss various functionals that allows one to replace the quite complex N-electrons wavefunction by the simpler electron density.

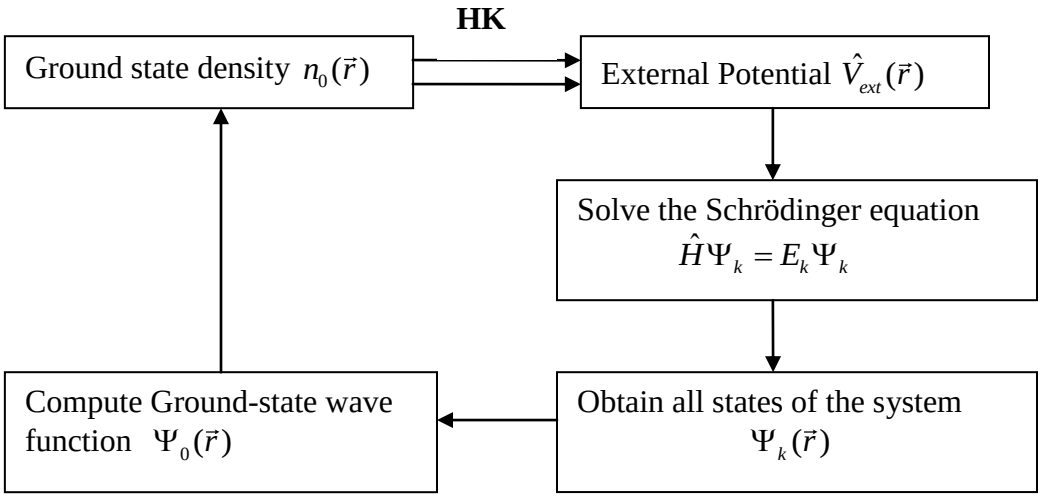

Figure 1: Schematic illustration of the first Hohenberg-Kohn (HK) Theorem by R. M. Martin (2004)[10].The double arrow labeled "HK" denotes the first HK theorem which states that for any system of interacting particles subjected to an external potential, the external potential is determined uniquely, within a trivial additive constant, by the ground state particle density $n_0(\vec{r})$.

2.4 Thomas-Fermi-Dirac approximation

Even though the Thomas-Fermi-Dirac (TFD)[74,75,76] approximation historically precedes the HK theorems, we decided to introduce the TFD approximation right after the brief description of the HK theorems. Such odd structured plan can be justified in view of the fact that even though one of the earliest calculation schemes that introduced the electron density instead of the many-electron wavefunction was given by Thomas and Fermi, one requires the HK theorems to justify the use of electron density functional.

In this section, the brief discussion of the initial TF[74,75] equations is heavily based on an excellent article written by Jones and Gunnarsson (1989)[65]. In the original TF equations, the electron-electron interaction potential is described by the electrostatic Coulomb integral where the *electrons are treated as completely independent particles*[65],

$$\hat{V}_{ee}^{TF}[n] = \frac{1}{2}\iint d^3r d^3r_1 \frac{n(\vec{r})n(\vec{r}_1)}{|\vec{r}-\vec{r}_1|} \tag{2.4.1}$$

Moreover, the total kinetic energy is approximated as an *explicit* functional of the density $T_{ee}^{TF}[n]$ for non-interacting electrons in a uniform gas,

$$\hat{T}_{ee}^{TF}[n] = C_l \int n(\vec{r})^{\frac{5}{3}} d^3r \text{ , where } C_l = \frac{3\hbar^2(3\pi^2)^{\frac{2}{3}}}{10m} \tag{2.4.2}$$

Such approximation is quite decent for slow variation of the electron density in an idealized homogeneous medium. However, one may argue that one of the most severe deficiencies in this model is the neglect of the exchange and correlation energies among the electrons. In 1930, Dirac[76] formulated a local approximation to the Hartree exchange energy where,

$$\hat{K}^D[n] = C_m \int_\Omega n(\vec{r})^{4/3} d^3r \text{ with } C_m = \frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} \tag{2.4.3}$$

By inserting the previously derived expressions for the electron-electron potential, kinetic energy and exchange energy in the original TF equation, one obtains a new energy functional for electrons subjected to an external potential,

$$E^{TFD}[n] = C_l \int_\Omega n(\vec{r})^{5/3} d^3r - C_m \int_\Omega n(\vec{r})^{4/3} d^3r + \int_\Omega n(\vec{r}) V_{ext} d^3r + \frac{1}{2}\iint d^3r d^3r_1 \frac{n(\vec{r})n(\vec{r}_1)}{|\vec{r}-\vec{r}_1|} \tag{2.4.4}$$

The ground state energy and density can be found by minimizing the functional $E^{TFD}[n]$ under the condition that the total number of electrons is given by $N = \int n(\vec{r}) d^3r$.

By incorporating this constraint by the method of Lagrange multipliers[10,22], the ground state density must therefore satisfy the variational principle

$$\delta\left\{E_{TFD}[n] - \beta_{TFD}\left(\int_\Omega n(\vec{r}) d^3r - N\right)\right\} = 0 \tag{2.4.5}$$

The corresponding Euler-Lagrange equation is the generalization of the Thomas-Fermi-Dirac equation:

$$\beta^{TFD} = \frac{\delta E^{TFD}}{\delta n} = \frac{5}{3} C_l n(\vec{r})^{2/3} - \frac{4}{3} C_m n(\vec{r})^{1/3} - \Theta(\vec{r}) \tag{2.4.6}$$

$$\text{Where } \Theta(\vec{r}) = \frac{\delta}{\delta n}\left\{\int_\Omega n(\vec{r}) d^3r + \frac{1}{2}\iint d^3r d^3r_1 \frac{n(\vec{r})n(\vec{r}_1)}{|\vec{r}-\vec{r}_1|}\right\} \tag{2.4.7}$$

And $\Theta(\vec{r}) = \frac{Z}{r} - \int d^3r_1 \frac{n(\vec{r}_1)}{|\vec{r}-\vec{r}_1|}$ is the classical electrostatic potential at $\vec{r}$.

Even though the introduction of a density functional theory would simplify the complexity of the full many-electron wave functions, the TFD approximation holds *severe deficiencies*. Such crude approximation lacks the precision to describe shell structure of atoms and binding of molecules.[89,90,91] Furthermore, one notices that as we get further from the nucleus, the charge density *does not decay exponentially* but rather in form of a power function and as we infinitesimally approach the nucleus, the charge density blows up to infinity.[65] In order to overcome those deficiencies, future works on improvements and modifications of the original TFD approximation have mainly been conveyed by Weisacker (1935)[92], Gross and Dreizler (1981)[93] and Perdew(1985a)[94]. Countless implementations into DFT have continued for many years but it should be realized that Density Functional Theory itself does not truly provide a way

to fully understand the properties of a material by *simply analyzing the form of the electronic density.*[10] Extracting directly from the electronic density any general characteristics of a system is not quite intuitive. One can conceptualize such difficulty by thinking on how to construct an accurate kinetic energy functional of N- non interacting particles subjected to an external potential from a given electronic density. This therefore leads us to the Kohn-Sham approach, in which an accurate kinetic energy functional is computed in terms of orbitals for corresponding non interacting electrons.

2.5 The Kohn-Sham (KS) approach

The basic idea behind the KS approach was to collect particles that were not interacting with each other, then to solve each electron individually subjected to the initial effective fictitious potential formed by the non-interacting particles.[64] After solving the Schrodinger equation for individual electrons, a new effective potential is obtained and then the electron density is recomputed which in turn give another effective potential. This loop is repeated until the minimum energy of the system is reached. A schematic representation of the self-consistent-field KS loop is represented by Figure 2 which is based on the *non-interacting-V-representability* assumption. This assumption relies on the fact that the ground-state density of our collection of non-local particles can represent the exact ground state density of the system. By constructing non-interacting electrons with the same density as the physical system, *the solution of the KS equations would produce the exact non-interacting kinetic energy, which includes almost all the true kinetic energy.*[78] The residual difference between the independent-particle kinetic energy $\hat{T}_S$ and the true kinetic energy is related to the famous exchange-correlation parameter E_{xc} which will shortly be discussed. Before explicitly explaining the KS ansatz, we should recall the definition of the total energy functional used in the first HK theorem where

$$E^{HK}[n] = F^{HK}[n] + \int_{\Omega} d^3 r \hat{V}_{ext} n(\vec{r}) + E_{KL} \tag{2.5.1}$$

Here E_{KL} represents the classical Coulomb interaction between fixed Kth and Lth ions described in Section 1.1 and $F^{HK}[n]$ is a functional independent of the external potential that describes all internal energies of the interacting electrons system:

$$\hat{F}^{HK}[n] = \hat{T}[n] + \hat{V}_{ee}{}^{HK}[n] \tag{2.5.2}$$

where the potential between electrons $\hat{V}_{ee}{}^{HK}[n]$ is defined as:

$$\hat{V}_{ee}{}^{HK}[n] = \frac{1}{2} \iint d^3 r d^3 r_1 \frac{n(\vec{r}) n(\vec{r}_1)}{|\vec{r} - \vec{r}_1|} + \text{“nonclassical terms”} \tag{2.5.3}$$

The first component of the previous equation is described in Section 2.1 and is denoted as the Hartree energy. The second component which is identified as the “non classical terms” is described by Parr and Wang as the most important part of the exchange-correlation energy. Based on the first HK theorem, Kohn and Sham introduce the concept of an auxiliary non interacting Hamiltonian in a non spin polarized system where:

$$\hat{H}^{KS} = \sum_{k=1}^{N} \left[-\frac{1}{2} \nabla^2 + \hat{V}_s^{KS}(\vec{r}) \right] \tag{2.5.4}$$

For a system of N independent electrons obeying this auxiliary Hamiltonian, the ground state has one electron in each of the N orbitals Ψ_i with the lowest eigenvalue ε_i of the Hamiltonian.[10] Hence the one electron Hamiltonian can be written as:

$$\left[-\frac{1}{2}\nabla^2 + \hat{V}_s^{KS}(\vec{r})\right]\Psi_k = \varepsilon_k \Psi_k \tag{2.5.5}$$

Which yield to the density of the auxiliary KS Hamiltonian,

$$n(\vec{r}) = \sum_{k=1}^{N} \left|\Psi_k(\vec{r})\right|^2 \tag{2.5.6}$$

In terms of KS orbitals, the HK expression for the ground state energy functional is given by:

$$E^{KS}[n] = \hat{T}_s[n] + E_{xc}[n] + \int_\Omega d^3r \hat{V}_{ext} n(\vec{r}) + E_{KL} \tag{2.5.7}$$

where the independent-particle kinetic energy $\hat{T}_S$ is

$$\hat{T}_S[n] = -\frac{1}{2}\sum_{k=1}^{N}\left\langle \Psi_k \left|\nabla^2\right| \Psi_k \right\rangle$$

Even though the kinetic energy $\hat{T}_S$ is unique by the HK theorem, it is still not the "exact" kinetic energy of the system. In order to create an exact kinetic energy component, Kohn and Sham defined a universal function $\hat{F}^{KS}[n]$ that includes all internal energies as:

$$\hat{F}^{KS}[n] = \hat{T}_s[n] + \frac{1}{2}\iint d^3r d^3r_1 \frac{n(\vec{r})n(\vec{r}_1)}{\left|\vec{r}-\vec{r}_1\right|} + E_{xc} \tag{2.5.8}$$

And equated this universal function $\hat{F}^{KS}[n]$ to the HK universal function $\hat{F}^{HK}[n]$ described in Eq. (2.5.1):

$$\hat{T}[n] + \hat{V}_{ee}^{HK}[n] = \hat{T}_s[n] + \frac{1}{2}\iint d^3r d^3r_1 \frac{n(\vec{r})n(\vec{r}_1)}{\left|\vec{r}-\vec{r}_1\right|} + E_{xc} \tag{2.5.9}$$

Which yields:

$$E_{xc} = \left(\hat{T}[n] - \hat{T}_s[n]\right) + \left(\hat{V}_{ee}^{HK}[n] - \frac{1}{2}\iint d^3r d^3r_1 \frac{n(\vec{r})n(\vec{r}_1)}{\left|\vec{r}-\vec{r}_1\right|}\right)$$

Or $E_{xc} = \left(\hat{T}[n] - \hat{T}_s[n]\right) + nonclassical\ terms$ (2.5.10)

Here, the exchange-correlation parameter is small in magnitude and represents the difference between the true kinetic energy and the independent-particle kinetic energy, and the non classical terms from the potential between electrons. Even though the magnitude of the independent-particle kinetic energy $\hat{T}_s[n]$ is slightly different from the true kinetic energy $\hat{T}[n]$ of the system, $\hat{T}_s[n]$ is treated exactly in this approach. The exact computation of $\hat{T}_s[n]$ circumvents many of the deficiencies of the TF approximation, that we have previously discussed in Section 2.4, such as the absence of chemical bondings in molecules and the lack of shell structure for atoms.[89,90,91] The Kohn-Sham equations represent a major breakthrough into solving the many-body problem because it offers an exact and unique ground state energy and density of the many-body electron problem, provided that the exchange and correlation parameter is explicitly defined.

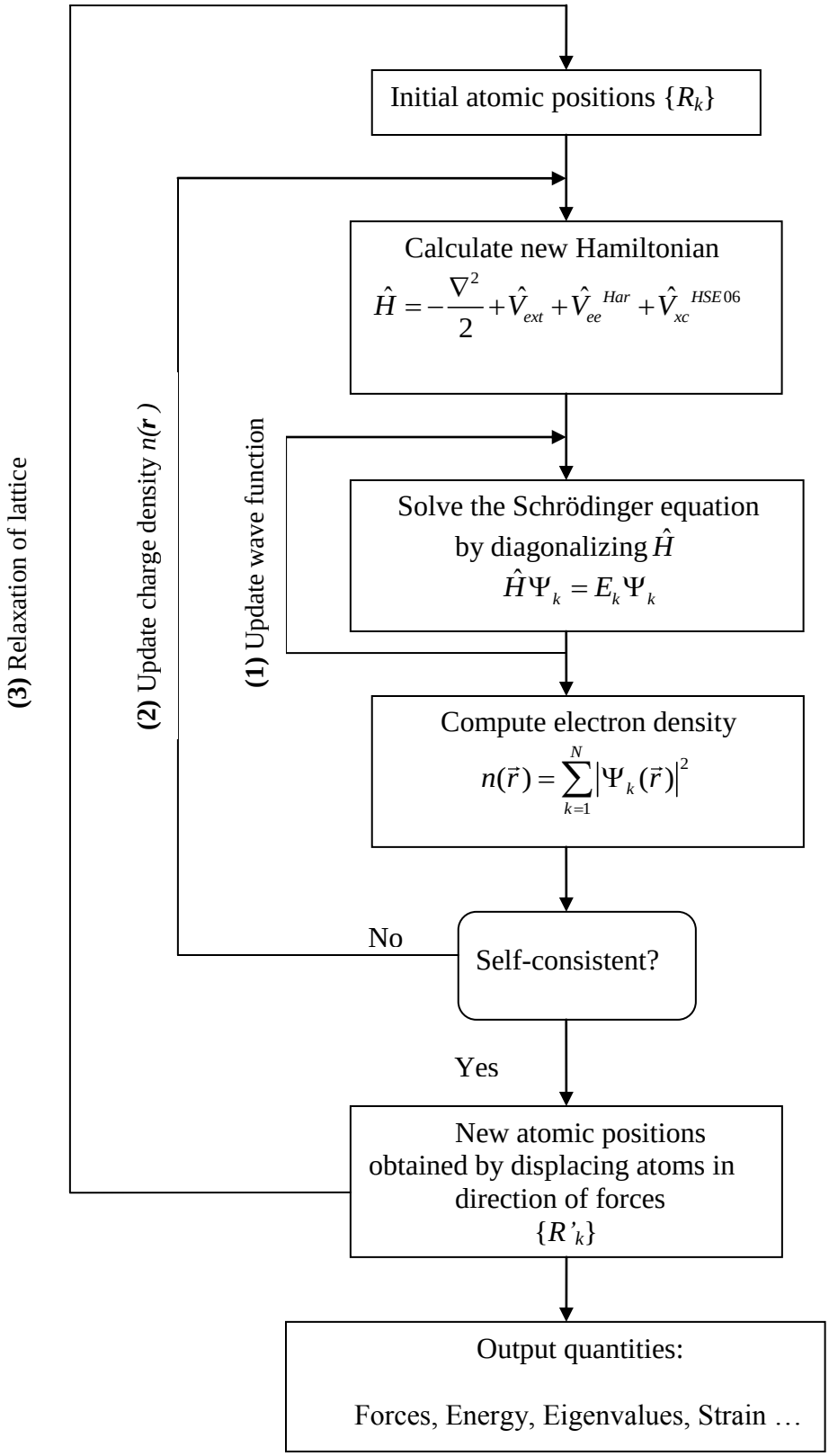

Figure 2: Schematic representation of the self-consistent loop where the charge density $n(\vec{r})$ and wave function $\Psi_k(\vec{r})$ are spin-dependent. The first (1) and second (2) loop must be iterated simultaneously for the two spins where the potential for each spin is calculated as a functional of the density of both spins.

2.5.1 Kohn-Sham variational equations

The Euler equation that is equivalent to the auxiliary single KS electron equation is given by:

$\hat{V}_s^{KS} + \frac{\delta F^{KS}[n]}{\delta n(\vec{r})} = \mu$ where μ is the Lagrange multiplier associated with the constraint $N = \int n(\vec{r}) d^3 r$. By varying the KS wave functions and applying the chain rule to Eq. (2.5.7), we obtain:

$$\hat{V}_s^{KS} = \frac{\delta}{\delta n(\vec{r})}\left[\int_\Omega d^3 r \hat{V}_{ext}(\vec{r}) n(\vec{r})\right] + \frac{\delta}{\delta n(\vec{r})}\left[\frac{1}{2}\iint d^3 r d^3 r_1 \frac{n(\vec{r}) n(\vec{r}_1)}{|\vec{r} - \vec{r}_1|}\right] + \frac{\delta E_{xc}}{\delta n(\vec{r})} \quad (2.5.1.1)$$

Hence $\hat{V}_s^{KS} = \hat{V}_{ext} + \int d^3 r_1 \frac{n(\vec{r}_1)}{|\vec{r} - \vec{r}_1|} + \hat{V}_{xc}$ (2.5.1.2)

where $\hat{V}_{xc} = \frac{\delta E_{xc}}{\delta n(\vec{r})}$ (2.5.1.3)

Here, $\hat{V}_{xc}$ is the obtained exchange correlation potential. By fully incorporating the exchange correlation effect of electrons, the *KS theory supersedes the HF theory, in providing an exact and unique molecular orbital theory.*[78] Unfortunately, an exact computation of the exchange and correlation potential remains complex and the physical interpretation of the exchange and correlation parameters are given in the following section.

2.5.2 Exchange and Correlation

The exchange and correlation parameter can be physically interpreted via an adiabatic approximation[19,95,96] or mathematically analyzed through a wave-vector[72] analysis. However the derivation of exchange and correlation energy in both methods requires an incredible amount of algebra. Roughly speaking, one can think of the exchange energy as the energy difference between different particles due to their spin characters while the correlation energy would be the difference between the total internal energy of the system minus the independent KS particle kinetic energy, the Coulomb energy and the exchange energy. Unfortunately, the relationship between the exchange and correlation is quite intimate and it is fairly complicated to "exactly" separate exchange from correlation.

Even though the KS equations exactly determine the previously unknown kinetic energy functionals by writing the electron density into N orbital components, the derivation of the exchange and correlation parameter is unclear. *By explicitly separating out the long-range Hartree components and the noninteracting particles kinetic energy*[10], the remaining exchange-correlation parameter can be approximated by using the uniform-electron gas formula. Hence, by treating the exchange-correlation functional locally or nearly local, it can be expressed as:

$$E_{xc}^{LDA} = \int d^3 r \varepsilon_{xc}(n, \vec{r}) n(\vec{r}) \quad (2.5.2.1)$$

Where $\varepsilon_{xc}(n)$ designates the exchange-correlation energy per electron at point *r* in a uniform electron gas of density *n*. According to Eq. (2.5.1.3), the corresponding exchange-correlation potential is thus given by:

$$\hat{V}_{xc}^{\ LDA} = \frac{\delta E_{xc}^{\ LDA}}{\delta n(\vec{r})} = \varepsilon_{xc}(n,\vec{r}) + n(\vec{r})\frac{\delta \varepsilon_{xc}}{\delta n(\vec{r})} \quad (2.5.2.2)$$

In order to make the previous equation more suitable for numerical calculations, the first term $\varepsilon_{xc}(n,\vec{r})$ can be divided into exchange and correlation components,

$$\varepsilon_{xc}(n,\vec{r}) = \varepsilon_x(n,\vec{r}) + \varepsilon_c(n,\vec{r}) \quad (2.5.2.3)$$

Hence Eq. (2.5.2.1) becomes:

$$E_{xc}^{\ LDA} = E_x^{\ LDA} + E_c^{\ LDA} = \int d^3r \varepsilon_x(n,\vec{r}) n(\vec{r}) + \int d^3r \varepsilon_c(n,\vec{r}) n(\vec{r}) \quad (2.5.2.4)$$

Where $E_x^{\ LDA} = \int d^3r \varepsilon_x(n,\vec{r}) n(\vec{r})$ (2.5.2.5)

and $E_c^{\ LDA} = \int d^3r \varepsilon_c(n,\vec{r}) n(\vec{r})$ (2.5.2.6)

The exchange energy $E_x^{\ LDA}$ can be easily computed since the *Kohn-Sham wave functions are simply Slater determinants of plane waves*.[78] Based on the Dirac exchange energy derived in Eq. (2.4.3), the exchange energy functional for a uniform gas is defined as:

$$E_x^{\ LDA} = -C_m \int n(\vec{r})^{4/3} d^3r \text{ with } C_m = \frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} \quad (2.5.2.5)$$

By comparing the previous equation with Eq. (2.5.2.5), the exchange component becomes:

$$\varepsilon_x = -C_m n(\vec{r})^{1/3} \text{ where } C_m = \frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} \quad (2.5.2.6)$$

Unfortunately, the second term of Eq. (2.5.2.3) or the so-called correlation parameter is far more complicated since it depends explicitly on the physical ground-state wavefunction of the uniform gas. The correlation energy can be vaguely conceptualized as the energy that is created from the correlated motion of the electrons due to electrostatic repulsion and attraction to the compensated positively charged background. Over the years, various approximations of the correlation energy have been made and one of the most accurate calculations was the Quantum Monte Carlo computations for uniform gas.[97]

By looking back at the derivation of the exchange and correlation potential described in Eq. (2.5.2.2), we define the second term $n(\vec{r})\frac{\delta \varepsilon_{xc}}{\delta n(\vec{r})}$ as the "response function". This "response function" or "response potential"is explicitly defined by Gritsenko et al. (1994)[98] as the response function of the exchange correlation hole subjected to density variations. The response potential can be physically interpreted as a discontinuity in the functional derivatives of ε_{xc} for finite systems[99] and bulk insulators[100]. In an excellent review written by Perdew and Levy (1982)[101], they argued that the band structure computed from the Kohn Sham ansatz underestimates the actual gap width by an amount equal to the derivative discontinuity. Hence, the difference between the *highest occupied and lowest unoccupied one-electron levels*[101] does not quite represent the actual band gap of a given system.

After locally approximating the exchange-correlation potential, the KS orbital equations derived in Eq. (2.5.5) can be rewritten as:

$$\left[-\frac{1}{2}\nabla^2 + \hat{V}_{ext} + \int d^3r_1 \frac{n(\vec{r}_1)}{|\vec{r} - \vec{r}_1|} + \hat{V}_{xc}^{\ LDA}\right]\Psi_k = \varepsilon_k \Psi_k \quad (1.5.2.7)$$

and the self consistent solutions of these local KS orbital equations represent the LDA formalism which is further investigated in the next section.

2.6 Functionals for Exchange and Correlation

Before introducing any approximation of the exchange-correlation parameter, we notice that the Kohn-Sham Density Functional Theory as it has been discussed up to this point is restricted to only unpolarized systems. By adding a magnetic field to the usual scalar external potential acting on the many-electrons systems, we are building more physical insight into the approximation of the exchange correlation functional. The next two sections are devoted into describing relevant functionals that approximate the spin dependent exchange correlation parameter.

2.6.1 Local Spin Density Functional (LSDA)

In addition to the local approximation of the exchange-correlation energy described in Section 2.5.2 by Kohn and Sham, we decide to insert spin dependence into the homogenous exchange correlation parameter, which yields:

$$E_{xc}^{\ LSDA} = E_{xc}^{\ LDA}\left[n^{\uparrow}, n^{\downarrow}\right] = \int d^3 r \varepsilon_{xc}\left[n^{\uparrow}, n^{\downarrow}\right] n(\vec{r}) \qquad (2.6.1.1)$$

This integral can be simplified or rather approximated by explicitly separating the exchange and correlation components:

$$E_{xc}^{\ LDA}\left[n^{\uparrow}, n^{\downarrow}\right] = \int d^3 r \varepsilon_x\left[n^{\uparrow}, n^{\downarrow}\right] n(\vec{r}) + \int d^3 r \varepsilon_c\left[n^{\uparrow}, n^{\downarrow}\right] n(\vec{r}) \qquad (2.6.1.2)$$

The LSDA can be described in terms of the up and down spin densities, but it is usually expressed in terms of the total charge density $n^{\uparrow}(\vec{r}) + n^{\downarrow}(\vec{r})$ and the local relative spin polarization:

$$\zeta(\vec{r}) = \frac{n^{\uparrow}(\vec{r}) - n^{\downarrow}(\vec{r})}{n^{\uparrow}(\vec{r}) + n^{\downarrow}(\vec{r})} \qquad (2.6.1.3)$$

For the far more sophisticated case of spin dependent correlation energy, there have been some serious attempts to parameterize the uniform gas correlation energy as a function of spin polarization, $\varepsilon_c(n, \zeta)$.[102,103] Even though the LSDA principles were funded on the basis of a homogenous gas approximation, its success for very inhomogeneous cases are not to be overlooked. One reason for such success might be the cancellation of errors between exchange and correlation. In 1966, Tong and Sham[104] noticed that in LSDA calculations, the total exchange energy is typically underestimated by about 10% while the correlation energy is overestimated by a factor of 2 or more. Since for many physical systems, the exchange energy is about 4 times greater than correlation,[78] the overestimation of the correlation energy greatly cancels the underestimation of the exchange energy. Due to the partial cancellation of errors, the LDSA gives excellent approximations of bond lengths, ionization and binding and dissociation energies. Unfortunately, LSDA tends to fail into describing weak bondings, systems with slowly varying densities, correct band gaps and magnetism of transition metals. To improve upon the local spin density formalism, one should take into account the problem of the unphysical self-interaction term in the approximation of the exchange and correlation parameter. In order to remove those spurious self interaction tems, Perdew and Zunger (1981)[105] suggested a Self

Interaction Correction (SIC) to LSDA. In such corrections, the newly obtained Kohn Sham orbitals vary for different potentials hence causing non-orthogonality of the orbitals. Further physical insight into the non orthogonality of the orbitals is given in the original review.[105]
Even after improving the LSDA, one observes that the exchange correlation potential has a quite short range and it only depends on the local density. Consequently the LSDA potential has the wrong asymptotic dependence as r tends to infinity and one might consider a *functional of the magnitude of the gradient of the density as well as the value of the density at each point.*[10]

2.6.2 Generalized Gradient Approximations (GGA)

The idea of constructing a functional that is not as computationally demanding as fully non local functionals and yet possesses a potential that does not diverge for exponentially decaying densities would be groundbreaking. One of the first suggestions into expanding the exchange correlation parameter in function of the magnitude of the gradient of the density was given in the original paper of Kohn and Sham[8] and it was referred to as a gradient expansion approximation (GEA). The insertion of gradient-dependent functionals into the local density approximation of the exchange correlation parameter was first thought to be very attractive for applications but unfortunately most gradient expansion approximations did not lead to consistent improvements over the LSDA. Such failure could be explained by the fact that the gradient in real materials is so large that a given gradient expansion would break down. Indeed, the major drawbacks of the GEA were that its corresponding exchange correlation hole was not physical, nor did it satisfy the normalization conditions of the exchange and correlation holes and the negativity state of the exchange hole.[12,106] By eliminating the spurious long-range term of the second order expansion of the GEA exchange-correlation hole, generalized gradient approximations[107,108,109] (GGAs) were created and their corresponding exchange correlation energy were approximated by:

$$E_{xc}^{GGA}\left[n^{\uparrow},n^{\downarrow}\right]=\int d^{3}r n(\vec{r})\varepsilon_{xc}\left(n^{\uparrow},n^{\downarrow},\left|\nabla n^{\uparrow}\right|,\left|\nabla n^{\downarrow}\right|...\right) \qquad (2.6.2.1)$$

GGA functionals have recently become quite popular in condensed matter physics. As a general trend, GGAs yield more accurate atomization energies[94,110,111], total energies[112] and barriers to chemical reactions[113] than LSDA. Furthermore GGAs can also correct the underestimation of bulk constant previously computed by LSDA. One of the arguably most crucial improvements of the GGAs over LSDA is the prediction of the correct ferromagnetism configuration of bcc ground state metallic iron.[114] However, except for the case of Hydrogen,[115,116] GGA bond lengths computations are not better to corresponding LSDA calculations for arbitrarily chosen molecules. Even though some GGAs provide some improvements over the LSDA, one should keep a realistic perspective of the predictive power of the GGAs and further understanding on the properties of each GGA functional is therefore required. One of the first serious attempts into writing a gradient functional that would be subjected to further improvements in the near future was performed by Perdew and Wang in 1986 (PW86).[107] New physical insight[117] was brought upon the exchange and correlation parameter when $E_{xc}^{GGA}\left[n^{\uparrow},n^{\downarrow}\right]$ was rewritten as:

$$E_{xc}^{GGA}\left[n^{\uparrow},n^{\downarrow}\right]=\int d^{3}r\varepsilon_{x}^{unpol}(n)F_{xc}^{unpol}\left[s(\vec{r}),r_{s}(\vec{r})\right] \qquad (2.6.2.2)$$

here $\varepsilon_x^{unpol}(n) = -\frac{3}{4}\left(\frac{3}{\pi}n(\vec{r})\right)^{1/3}$ is the exchange energy in an unpolarized system previously computed from the LDA formalism and F_{xc}^{unpol} is the dimensionless *enhancement factor over local exchange*.[117] A thorough investigation of the physical meaning of every single component from Eq. (1.6.2.2) is given in the original review.[117] In order to distinguish a GGA that would offer a consistent improvement over the previous LSDA, several authors[117,118] decided to plot the GGA dimensionless enhancement factor F_x^{GGA} in function of reduced density gradient

$$s(\vec{r}) = \frac{|\nabla n(\vec{r})|}{2k_F n(\vec{r})}, \tag{2.6.2.3}$$

for various values of Wigner-Seitz radius $r_s(\vec{r}) = \left(\frac{3}{4\pi n(\vec{r})}\right)^{1/3}$ (2.6.2.4)

Here, $k_F = \left(3\pi^2 n(\vec{r})\right)^{1/3}$ is the local Fermi wavevector (2.6.2.5)

A non monotonic behavior[118] from various famous functionals such as Becke (B88)[108], Perdew and Wang (PW91)[112] and the famous Perdew-Burke-Enzherof (PBE)[119]is observed after plotting the exchange part of the enhancement factor F_x in function of the reduced density gradient $s(\vec{r})$. Huge discrepancies among the functionals appear for large regions of $s(\vec{r})$ and such differences might be explained by the inaptitude of well describing regions that contain large gradient by the density gradient expansion.[10] Although some widely used gradient formalism such as the PBE-GGA and PW91 predict several correct physical properties, one cannot guarantee their absolute superiority in the entire condensed matter field.

As one would have expected, even though the correlation energy is a lot smaller than the exchange energy, rewriting the correlation parameter as a functional is quite complex. In generalized gradient approximations, some interesting works have been done on estimating the correlation parameter[112,119,120] but further improvements are still required in this area. Unfortunately, regardless of how precise these gradient functionals are constructed, DFT calculations are still not exact solutions of the full Schrodinger equation. Even though the true ground state energy of a given system can be well approximated by pure DFT, one can still not compute excited state energies. In addition to the non-time dependence of DFT, there exist important situations for which DFT computations yield unphysical results. The underestimation of bandgaps in semiconductors and insulators and the inaccurate weak Van der Waals attractions are one of the main disadvantages in the density functional formalism. Hence, instead of attacking the many-body problem from a pure electron density perspective, one might reconsider computing wave functions that would converge to an exact solution of the Schrodinger equation.

3. The Hartree-Fock (HF) method

The wavefunctions-based method provides a different approach to the treatment of the exchange-correlation interactions. One of the very first approaches to the many-particle problem was proposed by Hartree (1928)[68], where he assumed that the many-electron wave function can be expressed as a simple product of one-electron orbitals. Even though this approach is not quite realistic enough for general electronic systems, it is included in this section to illustrate the basic features of one-electron approaches. In the original review, Hartree proposed the self-consistent

field (SCF) approach in which electrostatic field felt by an electron was due to the potential created by the nucleus together with the field produced by other electrons.[121]

3.1 The Hartree approximation

For simplifications purposes, we will first assume that electrons are completely independent of each other and the electronic Hamiltonian can be described by:

$$\hat{H}_{elec} = \sum_{k=1}^{N} h_k \tag{3.1.1}$$

Where N is the total number of electrons and h_k describes the kinetic and potential energy for the *kth* electron. By solving the Schrodinger equation for a single electron, we obtain:

$$h_k \chi_\alpha = \varepsilon_k \chi_\alpha \tag{3.1.2}$$

The eigenfunctions χ_α of the previous one-particle Schrodinger equation represent the α th spin orbital of the kth electron and can be rewritten as:

$$\chi_\alpha(\vec{r}_k) = \chi_{a_\alpha}(\vec{r}_k)\left|m_s\right\rangle \tag{3.1.3}$$

where $\chi_{a_\alpha}(\vec{r}_k)$ and $\left|m_s\right\rangle$ respectively represent the spatial orbital and spin part of the spin orbital χ_α. Since the total electronic Hamiltonian $\hat{H}_{elec}$ is a sum of one electron operators h_k, the corresponding eigenfunctions of $\hat{H}_{elec}$ will therefore be products of the one-electron spin orbitals:

$$\Psi_k(\vec{r}_1, \vec{r}_2, ..., \vec{r}_N) = \chi_1(\vec{r}_1)\chi_2(\vec{r}_2)...\chi_\alpha(\vec{r}_N) \tag{3.1.4}$$

Initially, Hartree thought about using the variational principle for the ground state wavefunction by adopting this simple product trial function that we have derived in Eq. (3.1.4). However, electrons are indistinguishable spin particles or fermions, and by exchanging two electrons, the wavefunction must correspondingly switch sign according to the antisymmetry principle. By incorporating a Slater determinant[71,122] into the Hartree approximation, Fock proposed an antisymmetric N-electron wavefunction which can be described as:

$$\Psi_k^{\,HF}(\vec{r}_1, \vec{r}_2, ..., \vec{r}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\vec{r}_1) & \chi_2(\vec{r}_1)\cdots & \chi_\alpha(\vec{r}_1) \\ \chi_1(\vec{r}_2) & \chi_2(\vec{r}_2)\cdots & \chi_\alpha(\vec{r}_2) \\ \vdots & \vdots & \vdots \\ \chi_1(\vec{r}_N) & \chi_2(\vec{r}_N) & \chi_\alpha(\vec{r}_N) \end{vmatrix} \tag{3.1.5}$$

The coefficient in front of the matrix represents a normalization factor and within the matrix, we notice that N electrons can occupy α orbitals without specifying which electron is in which orbital. By investigating the nature of the rows and columns of this newly obtained determinantal expression, we notice that the conditions of the Pauli Exclusion Principle are completely satisfied. By obtaining an antisymmetric Hartree product, we consequently obtain exchange components since the *exchange effects arise from the requirements of the invariance of* $|\Psi|^2$ *to interchange space and spin coordinates of any two electrons*.[67] More details on the nature of the exchange energy and the antisymmetry principle can be qualitatively described by Slater (1960)[123] and Szabo and Ostlund (1996).[67] Although we have obtained a wavefunction that satisfies the antisymmetry principle, we are still not yet at the level of exactly solving the full interacting Hamiltonian.

3.2 The Hartree-Fock equations

By using the antisymmetric wave function that we have derived in Eq. (3.1.5), we can approximate the ground state of an N electron system by:

$$\left|\Psi_0\right\rangle = \left|\chi_1(\vec{r}_1)\chi_2(\vec{r}_2)...\chi_\alpha(\vec{r}_N)\right\rangle \tag{3.2.1}$$

By applying the variational principle to the electronic Hamiltonian, we obtain the lowest energy which is described by:

$$E_0 = \left\langle \Psi_0 \left| \hat{H}_{elec} \right| \Psi_0 \right\rangle \tag{3.2.2}$$

By minimizing E_0 with respect to the choice of spin orbitals, we obtain the so-called Hartree-Fock equation:

$$\left[-\frac{1}{2}\nabla^2 + V_{eff}(\vec{r}) \right] \chi_\alpha(\vec{r}_k) = \varepsilon_k \chi_\alpha(\vec{r}_k) \tag{3.2.3}$$

For electrons respectively located at $\vec{r}_1$ and $\vec{r}_2$, the effective potential $\hat{V}_{eff}(\vec{r})$ is described as:

$$\hat{V}_{eff}(\vec{r}) = \hat{V}_{ext}(\vec{r}) + \sum_{\alpha,\beta} \int\int \left[d^3r_1 d^3r_2 \left|\chi_\alpha{}^*(\vec{r}_1)\right|^2 \left|\chi_\beta{}^*(\vec{r}_2)\right|^2 \frac{1}{\left|\vec{r}_1 - \vec{r}_2\right|} \right]$$
$$- \sum_{\alpha,\beta} \int\int \left[d^3r_1 d^3r_2 \chi_\alpha{}^*(\vec{r}_1)\chi_\beta(\vec{r}_1)\chi_\beta{}^*(\vec{r}_2)\chi_\alpha(\vec{r}_2)\delta m_\alpha \delta m_\beta \frac{1}{\left|\vec{r}_1 - \vec{r}_2\right|} \right] \tag{3.2.4}$$

The second term in the left hand-side is the purely classical Coulomb repulsion between electrons that we derived in Eq. (2.1.14) and the third component is the famous exchange energy. The negativity of the exchange terms corresponds to an odd permutation in the Slater determinant. Basically, the exchange components prevent spin-like electrons from being too close to each other or to occupy the same state which hence reduces the electrostatic Coulomb repulsion. Furthermore, the Kronecker Delta terms from the exchange operator are required since the potential does not flip the spins of the electrons, and hence in case of different spin orbitals, the exchange term should completely vanish. In order to complete our brief analysis of the HF approximation, we must define how the spin orbitals are computed and how should it be combined to obtain a quite precise many-body wavefunction. A good approximation of spin orbitals χ_α can be obtained by linear combinations of basis functions $\Phi_j(\vec{r})$:

$$\chi_\alpha = \sum_{j=1}^{J} a_{\alpha,j} \Phi_j(\vec{r}) \tag{3.2.5}$$

where $a_{\alpha,j}$ are the expansion coefficients. From such combination, one can hence conclude that the accuracy of the HF formalism relies solely on the specification of the expansion coefficients. Like the previously discussed KS equations, we notice an iterative process from the HF formalism since

- One must first approximate the spin orbitals based on the size of the basis set.
- Then define the probability density $\sum_\alpha \left|\chi_\alpha{}^*(\vec{r}_1)\right|^2$ or electron density $n(\vec{r}_1)$ from the previous estimate of the spin orbitals.
- From the obtained electron density, one must again solve the single electron HF equation for the spin orbitals.

- Finally we can compare the obtained spin orbitals with the ones we initially estimated. In case that the obtained spin orbitals correspond to our initial guess, the total energy and other characteristics of the system can therefore be calculated. On the other hand, if the obtained spin orbitals are inconsistent with the initial ones, the trial spin orbitals must be updated by using the same iterative process.

The HF approximation has been for a quite long time the most renowned formalism to calculate the electronic structure of molecular systems. Unfortunately, even though the exact exchange was treated exactly from the HF wavefunction, the many-body correlation effects were completely absent. In molecular calculations, improving the HF method can be obtained by expensive post-HF methods such as the configuration interaction (CI), coupled cluster (CC), Moller-Plesset perturbation theory (MP) and the quadratic configuration interaction (QCI) formalism.

Unfortunately, several of these post HF methods are only possible for relatively small number of atoms due to the exponential increase of computational expense associated with the size of the system and the number of basis functions. A circumvention of the computational expense of such calculations can be found in the previously discussed DFT that uses periodic and spatially localized functions. Nevertheless, density functional formalism carries major drawbacks that are explicitly described by Jones and Gunnarsson (1989).[65] A way forward into finally obtaining an excellent approximation of the solution of the many-body Schrodinger equation might be a density functional theory in which there would be a combination of the exact exchange functional based on wave-function-based methods and approximations of the correlation component. In the next section, we will focus on giving a brief overview of this new class of mixed functionals called “hybrid functionals”.

4. Hybrid Functionals

4.1 Importance and definition

In the last decade, wave-functions-based methods and DFT have proven to be very powerful tools in describing various properties in a wide range of materials. Nevertheless, the explicit derivation of the exchange correlation parameter in DFT and the computational cost of post-HF methods are still methodological barriers that have not been overcome. First-principles method based on DFT such as LSDA has been quite successful, especially for those where the electronic density is quite uniform. In order to address the main limitations of LSDA, an expansion of the density in terms of the gradient and higher order derivatives has been carried out. Unfortunately, the improvement of GGAs over LSDA are not yet substantial since GGAs still do not completely correct the band gap in many systems and do not satisfy known asymptotic behaviors for isolated atoms. The main reasons of such limitations is probably due to the fact that self-interaction are still present in the Hartree term, the non-locality of the exchange component is not fully taken into account and the complexity of computing the correlation term. The next step beyond first-principles methods based on DFT and post-HF formalisms might be the introduction of so-called hybrid functionals which are obtained by an admixture of a non-local fixed amount of Fock exchange to GGA-type functionals. Recently developed hybrid functionals such as Perdew-Burke-Ernzherof-zero-parameter (PBE0)[124,125] and Becke-three parameter-Lee-Yang-Parr (B3LYP)[109,126,127] have correctly predicted band gaps, bulk moduli and lattice constants and substantially improved the thermochemical[28] properties of many systems.

Nevertheless, the computational cost of calculating the exact Fock exchange under periodic boundary conditions made the use of these hybrid formalisms almost unattainable. In order to make the Fock exchange parameter tractable in periodic systems, Heyd (2006)[31] argues that the spatial decay of the exchange interactions must be accelerated or one might artificially cut off or truncate part of the exchange interactions. The first alternative, which is the acceleration of spatial decay, might well predict the total energy of the system but would still neglect long range exchange-correlation. The second approach, truncating the exchange interactions are known to work quite well in localized systems. Nevertheless, in case of delocalized charge distribution where the HF exchange does not rapidly decay over distance, such truncation methods appear completely unphysical and create some severe convergence problems in the self-consistent-field process. Hence, in order to accelerate the HF decay without neglecting long range interactions, the *1/r* part of the exchange interaction can be replaced with a Coulomb screened potential. In addition to the intractability of the Fock exchange parameter, one should notice in metallic systems the *logarithmic divergence*[63] *of the partial derivative* $\left.\frac{\partial \varepsilon}{\partial \vec{k}}\right|_{k=k_F}$ of *the single particle Hartree-Fock* $\varepsilon\left(\vec{k}\right)$, *with respect to the crystal momentum k at the Fermi level* k_F, where Aschcroft and Mermin (1976, Chap. 17)[128] defined the single particle HF energy $\varepsilon\left(\vec{k}\right)$ as:

$$\varepsilon\left(\vec{k}\right)=\frac{\hbar^2 k^2}{2m}-2\frac{e^2}{\pi}k_F F\left(\frac{k}{k_F}\right) \tag{4.1.1}$$

$$\text{where } F\left(\frac{k}{k_F}\right)=\frac{1}{2}+\frac{{k_F}^2-k^2}{4kk_F}\ln\left|\frac{k_F+k}{k_F-k}\right| \tag{4.1.2}$$

Such logarithmic divergence can be explained by *the divergence of the Fourier transform* $\frac{4\pi e^2}{k^2}$ *of the Coulomb interaction* $\frac{e^2}{r}$ *at* $k=0$.[128] By screening the *1/r* part of the exchange interaction, we obtain a potential that has a shorter range than *1/r* and we can therefore eliminate the unphysical singularity of the anomalous divergence of the derivative of the one-electron HF energy.

4.2 Screened Coulomb potential hybrid functionals (HSE03 and HSE06)

The starting point of this new class of hybrid screened functionals or the so-called Heyd-Scusseria-Ernzherof (HSE03)[32] screened hybrid functionals is the partitioning of the Coulomb interaction operator into short-range (SR) and long-range (LR) components[129], respectively:

$$\frac{1}{r}=\underbrace{\frac{erfc(wr)}{r}}_{Short-Range}+\underbrace{\frac{erf(wr)}{r}}_{Long-Range} \tag{4.2.1}$$

where *w* is an adjustable parameter that describes the range of short-range interactions. Choosing the error function *erf(wr)* and its complement $erfc(wr)=1-erf(wr)$ to achieve the Coulomb partition ensures that the short-range component $\frac{erfc(wr)}{r}$ is singular and rapidly

decays as a Gaussian, while the long-range part $\frac{erf(wr)}{r}$ is nonsingular and smoothly decays as r tends to infinity. Figure 3 illustrates the behavior of *1/r* as *r* increases, the fast decay of $\frac{erfc(wr)}{r}$ and the smooth decay of $\frac{erf(wr)}{r}$ for *w=1*.

The construction of HSE screened hybrid functionals is based on the admixture of both HF and PBE-GGA exchange in the short range parts. The exact exchange mixing in the short range for short-range interactions allows the *PBE exchange hole to be delocalized only among the closest neighbors of a reference point. The long range component is represented by the PBE-GGA functional.*[11]

Before further describing the HSE hybrid functional, we initially start by writing the exchange correlation parameter from the free-parameter-Perdew-Burke-Ernzherof (PBE0)[124,125] hybrid functional:

$$E_{xc}^{PBE0} = bE_x^{HF} + (1-b)E_x^{PBE} + E_c^{PBE} \quad (4.2.2)$$

Where *b = 1/4* is the exchange coefficient that is determined by an adiabatic expansion calculated by Perdew et al. (2006).[29] In order to make the previous equation more suitable for numerical calculations, the exchange-correlation energy is divided into exchange and correlation components, where the exchange part is described as:

$$E_x^{PBE0} = bE_x^{HF} + (1-b)E_x^{PBE} \quad (4.2.3)$$

By splitting all terms of the PBE0 exchange energy into short range and long-range components and assuming that HF and PBE long range contributions tend to cancel each other, the HSE exchange-correlation parameter is expressed by:

$$E_{xc}^{HSE} = bE_x^{HF,SR}(w) + (1-b)E_x^{PBE,SR}(w) + E_x^{PBE,LR}(w) + E_c^{PBE} \quad (4.2.4)$$

From a quantitative perspective, the HSE exchange-correlation energy can be rewritten as:

$$E_{xc}^{HSE} = \frac{1}{4}E_x^{HF,SR}(w) + \frac{3}{4}E_x^{PBE,SR}(w) + E_x^{PBE,LR}(w) + E_c^{PBE} \quad (4.2.5)$$

Where the screening parameters for the HF and PBE are respectively given by $w^{HF} = \frac{0.15}{\sqrt{2}}$ and $w^{PBE} = 0.15 \times 2^{1/3}$. The screened approximation of the exchange correlation energy becomes the original hybrid PBE0 exchange-correlation as the adjustable parameter *w* goes to 0 and is *asymptotically* equivalent to pure PBE as *w* approaches infinity. In real space, the HSE03 is less computationally demanding than any hybrid functionals and can therefore be applied to solids and large molecules. The elimination of the long-range component of the Fock exchange drastically *reduces the range over which the real space integrals are computed*.[11] In addition to being less untractable in real space formalism, in reciprocal space, from a band structure perspective, one notices *an increased locality* of the HSE03 Fock exchange which permits a computation of the Fock exchange interaction in a much coarser mesh of points in the Brillouin zone (BZ).[11] The precision of the HSE03 method is illustrated in Figure 4 where the theoretical band gaps of various materials are compared with their respective experimental band gaps. The band gaps obtained by the PBE formalism (blue circles) are largely underestimated due to the approximation of the exchange and correlation energy. The hybrid functional PBE0 (green squares) yields better results than the PBE but it still tends to overestimate the band gaps of some materials because of the exact treatment of the Fock exchange in their formalism. The

HSE03 functional computed with the standard amount of Fock exchange (25 %) provides almost perfect agreement with the experimental results. Even though HSE03 functional allows some improvements in the correct prediction of thermochemical energies, band-gaps and atomization energies, its precision is yet to be desired. Paier et al (2006)[11] argue that HSE03 functional is not mature enough to fully replace well-known semi-local functionals because of the severe underestimation of the cohesive energy in several systems, the overestimation of the magnetic moment of transition metals and incorrect prediction of large gaps materials. Such inaccuracy might be traced back to the use of the standard amount Fock exchange (25 %) or imprecise computed screening parameter. In order to eliminate the drawbacks of the HSE03 formalism, one can think about further tuning the amount of Fock exchange and reexamining the choice of the adjustable parameter *w* for HF and PBE exchange.

The construction of a new screened hybrid functional (HSE06)[41] is based on the modification of the adjustable parameter w in both HF and PBE. With such adjustment, the previous defaults of the HSE03 are eliminated and certain equilibrium between computational cost and accurate physical results is finally reached. The HSE06 equations can be rewritten as:

$$\left[-\frac{1}{2}\nabla^2 + \hat{V}_{ext} + \hat{V}_{ee}^{Har} + \hat{V}_{xc}^{HSE06}\right]\Psi_k = \varepsilon_k \Psi_k \qquad (4.2.6)$$

The accuracy of the HSE06 formalism in the computation of band structure of wide-gap semiconductor material such as Gallium Nitride (GaN) is illustrated in Figure 5. The electronic band structure (red color) described by the GGA formalism, provides an adequate description for ground state properties determined by local ground state charge density that is derived from filled bands. However, if one is interested to properties related to the band gap or empty bands, GGA falls short of providing accurate results. The calculated HSE06 band gap (3.49 eV) is in excellent agreement with experimental band gap (3.50 eV).[130]

Now that we have provided a brief overview of the state-of-the-art methodology for performing first-principles calculations in perfect crystal structures, we will introduce in the next section various methods that help investigating the electronics of defects and impurities in semiconductors.

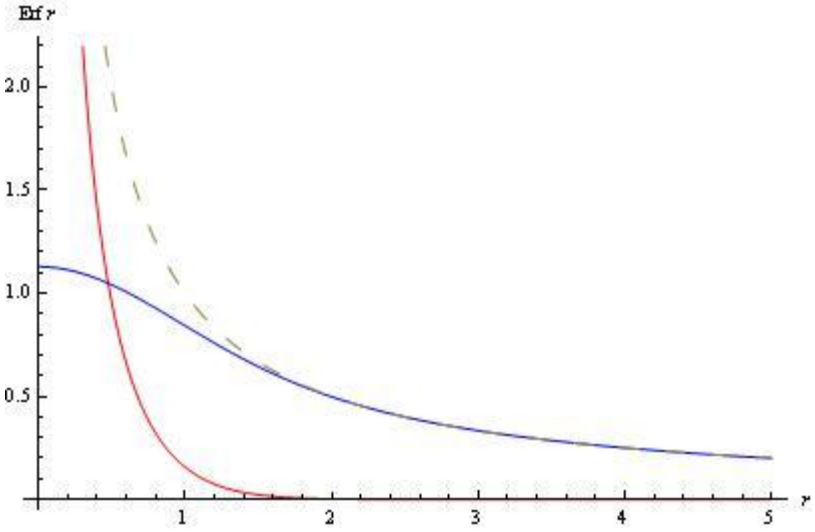

Figure 3: Graphs of Inverse function $\frac{1}{r}$, $\frac{erf(wr)}{r}$ and $\frac{erfc(wr)}{r}$ from Eq. (4.2.1) for $w=1$. Here $\frac{1}{r}$ is represented by the dashed line, $\frac{erfc(r)}{r}$ is rapidly decaying (red color) while $\frac{erf(r)}{r}$ is smoothly decreasing (blue color).

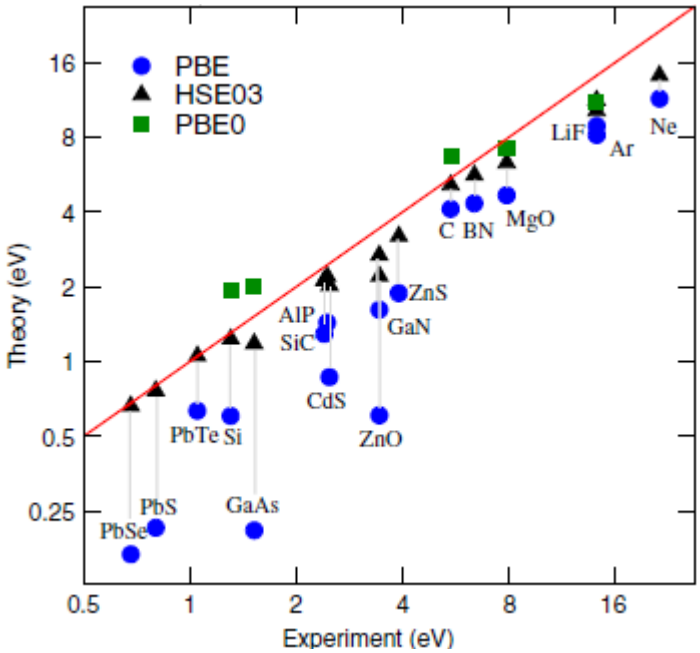

Figure 4: Illustrative comparison of band gaps done by Marsman et al.[150] where the theoretical band gaps obtained from PBE, PBE0 and HSE03 calculations are plotted against the experimental band gaps.

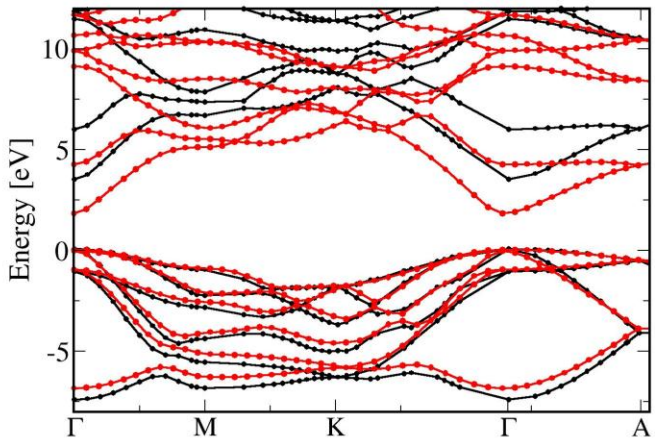

Figure 5: Electronic band structure of primitive unit cell of GaN obtained by the GGA (red color) and the HSE06 (black color) method. Energy is measured relative to the top of the valence band (0 eV). The Greek letters Γ-A represent high-symmetry points in the first Brillouin zone of hexagonal lattice.

5. Techniques for estimating supercell defects calculations

In the preceding KS chapter, it was discussed that several observables of the many-body system can be mapped into corresponding observables in a single-particle problem subjected to an effective potential. Nevertheless, the heavy task of computing the wave function of each of the vast number of electrons in a system that undergoes periodically repeated boundaries conditions still remains. Such problems can be overcome by applying Bloch's theorem to the HSE06 formalism which will give rise to a basis set consisting of plane waves. Plane waves are aesthetically appealing for periodic crystals since they possess the advantage of completely spanning the Hilbert space and they also provide mathematical simplicity for practical calculations. The following section is organized first to describe the derivation of plane wave basis by incorporating Bloch's theorem into the HSE06 method. The remaining sections are devoted to explaining relevant concepts that enter in the computation of formation energies of defects in semiconductors.

5.1 Plane waves basis sets in HSE06 formalism

According to Ashcroft and Mermin (1976, Chap. 8)[128], Bloch's theorem states that within a perfectly periodic potential, each electronic wave function can be rewritten as a product of a wavelike part and a cell-periodic part where:

$$\Psi_{n,\vec{k}}(\vec{r}) = e^{i\vec{k}\cdot\vec{r}} u_{n,\vec{k}}(\vec{r}) \qquad (5.1.1)$$

Here $u_{n,\vec{k}}(\vec{r})$ possesses the periodicity of the potential and verifies the relation $u_{n,\vec{k}}(\vec{r}) = u_{n,\vec{k}}(\vec{r}+\vec{R})$ for all $\vec{R}$ in a Bravais lattice. In Eq. (5.1.1), $\vec{k}$ represents the wave vector confined within the first Brillouin zone (BZ) and *n* describes the band index. Since $u_{n,\vec{k}}(\vec{r})$ is a periodic function, it can be expanded using a basis set of plane waves whose wave vectors are reciprocal lattice vectors of the crystal,

$$u_{n,\vec{k}} = \sum_{\vec{G}} C_{n,\vec{k}}(\vec{G}) e^{i\vec{G}\cdot\vec{r}} \tag{5.1.2}$$

Here, $\vec{G}$ are the reciprocal lattice vectors defined by $\vec{G}\cdot\vec{R} = 2\pi m$ for all $\vec{R}$, *m* is any integer and $C_{n,\vec{k}}(\vec{G})$ are the plane wave coefficients. By inserting Eq. (4.1.2) into Eq. (4.1.1), each electronic wave function in a cell of volume Ω can be described by:

$$\Psi_{n,\vec{k}}(\vec{r}) = \frac{1}{\sqrt{\Omega}} \sum_{\vec{G}=0}^{\infty} C_{n,\vec{k}}(\vec{G}) e^{i(\vec{k}+\vec{G})\cdot\vec{r}} \tag{5.1.3}$$

where the plane waves basis functions are defined by

$$\psi_{\vec{G}}(\vec{r}) = \frac{1}{\sqrt{\Omega}} e^{i\vec{G}\cdot\vec{r}} \tag{5.1.4}$$

which satisfy the orthonormality conditions:

$$\left\langle \psi_{\vec{G}} \middle| \psi_{\vec{G}'} \right\rangle = \frac{1}{\Omega}\int_{\Omega} e^{i(\vec{G}-\vec{G}')\cdot\vec{r}} d^3r = \frac{1}{\Omega}\left(\Omega\delta_{\vec{G},\vec{G}'}\right) = \delta_{\vec{G},\vec{G}'} \tag{5.1.5}$$

Now Eq. 4.1.3 can be rewritten as:

$$\Psi_{n,\vec{k}}(\vec{r}) = \sum_{\vec{G}=0}^{\infty} C_{n,\vec{k}}(\vec{G}) e^{i\vec{k}\cdot\vec{r}} \psi_{\vec{G}}(\vec{r}) \tag{5.1.6}$$

From the previous equation, one remarks that except for $\vec{G} = 0$, the reciprocal lattice vectors $\vec{G}$ that describe the plane wave expansion always lie outside the BZ, while the wavelike part $e^{i\vec{k}\cdot\vec{r}}$ involves a wave vector ***k*** in the first BZ.[121] Unfortunately, in our HSE06 formalism, the electronic density is expressed as a BZ average and thus one must incorporate the wavelike part into our previous plane wave basis functions which gives rise to:

$$\psi_{\vec{k},\vec{G}} = \frac{1}{\sqrt{\Omega}} e^{i(\vec{k}+\vec{G})\cdot\vec{r}} \tag{5.1.7}$$

Now the electronic wave function can be expressed as:

$$\Psi_{n,\vec{k}}(\vec{r}) = \sum_{\vec{G}=0}^{\infty} C_{n,\vec{k}}(\vec{G}) \psi_{\vec{k},\vec{G}}(\vec{r}) \text{, anywhere in the BZ.} \tag{5.1.8}$$

In principle, provided that each electron occupies a state corresponding to a specific ***k***, one requires an infinite number of reciprocal lattice vectors $\vec{G}$ to represent the wave functions with tremendous accuracy. Nevertheless, in practice, the Fourier coefficients $C_{n,\vec{k}}(\vec{G})$ of the plane wave function decrease with increasing $\left|\vec{k}+\vec{G}\right|$ and the plane wave expansion can hence be efficiently shortened at a finite number of terms.[121] Such truncation seems less complicated in case of the kinetic energy in reciprocal space. The expansion of the kinetic energy term from the HSE06 approximation in reciprocal formalism gives rise to a diagonal kinetic energy operator which is given by:

$$\hat{T}_{\vec{G},\vec{G}'} = \left\langle \psi_{\vec{k},\vec{G}} \left| -\frac{1}{2}\nabla^2 \right| \psi_{\vec{k},\vec{G}'} \right\rangle = \frac{1}{2}\left|\vec{k}+\vec{G}\right|^2 \delta_{\vec{G},\vec{G}'} \tag{4.1.9}$$

Since the amount of Fourier coefficients is inversely proportional to $\left|\vec{k}+\vec{G}\right|$, the coefficients for plane waves with small kinetic energy are typically more important than those with high kinetic energy and one may thus introduce a kinetic energy cut-off in order to achieve a finite basis set. However, the truncation of the basis set at finite cut-off energy will lead to an error in the calculated physical quantities but fortunately this error can be reduced by increasing the value of the cut-off energy. In addition to the kinetic energy, the potential energies can also be computed in reciprocal space where they are expressed in terms of their Fourier transforms:

$$\hat{V}_{\vec{G},\vec{G}'} = \left\langle \psi_{\vec{k},\vec{G}} \left| \hat{V} \right| \psi_{\vec{k},\vec{G}'} \right\rangle = \frac{1}{\sqrt{\Omega}} \int_{\Omega} e^{-i(\vec{G}-\vec{G}')\vec{r}} V(\vec{r}) d^3r = \hat{V}\left(\vec{G}-\vec{G}'\right) \tag{4.1.10}$$

Now, the expansion of the electronic wave functions in terms of plane waves allows the HSE06 equations to take on the secular form:

$$\sum_{\vec{G}'} \left[\frac{1}{2}\left|\vec{k}+\vec{G}\right|^2 + \hat{V}_{ext}\left(\vec{G}-\vec{G}'\right) + \hat{V}_{ee}^{\ Har}\left(\vec{G}-\vec{G}'\right) + \hat{V}_{xc}^{\ HSE06}\left(\vec{G}-\vec{G}'\right) \right] C_{n,\vec{k}}(\vec{G}') = \varepsilon_{n,\vec{k}} C_{n,\vec{k}}(\vec{G}) \tag{4.1.11}$$

Unfortunately, for a system containing valence and core electrons, the size of the previous Hamiltonian matrix will be gigantic and therefore intractable. In fact, at the center of the atom and at the valence region, the spatial variation of wave functions is quite fast and one must require an infinitely large number of plane waves. To overcome these difficulties, the *bare nuclear potential*[121] is substituted by a smooth pseudopotential. Basically, by using a pseudopotential, the core states will no longer exist and the valence pseudopotential wave functions become extremely smooth near the nuclei. Furthermore, the energy cut-off can directly be approximated from the atomic pseudopotentials but it usually needs to be adjusted by trials for a given system. For the purpose of our research, we will be using Projected- Augmented-Wave (PAW) method to compute our soft pseudopotentials and more details on their characteristics can be found in a review written by C. Roostgard[131] in 2009.Although technically appealing, the application of Bloch's theorem is not valid in crystals containing impurities since calculations using plane waves basis sets can only be performed on systems subjected to periodic boundary conditions. By incorporating impurities inside a bulk solid, the perfect periodicity becomes broken and one might require creating an artificial cell (supercell) that would be periodically reproduced throughout space.

5.2 Relaxation of supercell using HSE

Before creating an artificial supercell, one might be interested into calculating the theoretical lattice constant of the host crystal which is obtained by relaxing atoms in the primitive unit cell. A relaxation is obtained by voluntarily displacing atoms in such way that energy, forces, stresses and any other output quantities of the HSE self-consistent loop (Figure 2) are minimized. Once we obtain a converged lattice constant, we create an artificial supercell geometry in which the defect is surrounded by a region of bulk crystal that is subjected to periodic boundary conditions. Such method allows the calculation *of the energy per unit cell of a crystal enclosing an array of defects* instead of the computation of the energy of an entire crystal containing a single defect.[132] The artificial supercell method is quite convenient since it allows

the use of Bloch's theorem which requires translational periodicity of the system. However, it is crucial to include enough bulk solid in the supercell such that the defects are sufficiently separated and the properties of the isolated defects can thus be computed. More details on the characteristics of the periodicity of the surface of the supercell are given by M. C. Payne et al. (1992)[132]. In case of a charged system, the supercell approximation introduces the concept of uniform, neutralizing jellium background charge that usually circumvents the divergence of the Coulomb energy between the periodic charged defect images. Even though the supercell approximation provides a good description of the crystal structure and local stable *rearrangement of bonding between atoms*[133] in a given system, it suffers from major problems that will be discussed in Section 5.5. Although imperfect, the supercell method is still quite accurate and can be used within the HSE06 formalism to investigate the total energy of a system containing defects and impurities.

5.3 Defect Formation Energy

The probalibity of realizing a particular defect configuration *D* in a host lattice containing α type atoms in the charge state *q* within the supercell formalism is given by:

$$E_f[D^q] = E_{tot}[D^q] - E_{tot}[bulk] + q\times(E_v + \Delta E_F) - \sum_\alpha n_\alpha \mu_\alpha + \Delta E_{bf}[D^q] + \Delta E_{MP} + \Delta E_{PA} \quad (5.3.1)$$

Where $E_{tot}[D^q]$ and $E_{tot}[bulk]$ are the total energies of the bulk+defect and bulk-only supercell, respectively. The third term $q\times(E_v + \Delta E_F) = qE_F$ is the amount of energy it cost to charge the neutral impurity assuming that the exchange of electrons occurs at the Fermi level E_F. The number of atoms of type α (bulk atoms or defect atoms) that have been removed from or added to the host lattice are represented by the parameter n_α and μ_α indicates the corresponding chemical potential of the α atoms.

5.3.1 Chemical potentials

The chemical potentials depend on the source of the impurities and growth conditions and can be approximated from the HSE06 calculations. For instance, in case of "pure" GaN bulk, the Gallium(Ga)-rich conditions are present when Ga chemical potential equals that of metal Ga, where $\mu_{Ga} = \mu_{Ga[metal]}$. Similarly, extreme N-rich conditions occur when the chemical potential of N equals that of gas N_2, where $\mu_N = \mu_{N[N_2]}$. According to Van de Walle et al. (2004)[3], these extreme environmental growths correspond to placing upper bonds on the chemical potentials of Ga and N, respectively. Subsequently, we can also impose lower bounds by calculating the equilibrium constraint for the atomic chemical potential of the GaN bulk:

$$E_{GaN}^{Bulk} = \mu_{Ga} + \mu_N \quad (5.3.1.1)$$

where E_{GaN}^{Bulk} is the total energy of a two-atom unit of bulk GaN. By imposing the lower limit on μ_{Ga}, we obtain:

$$\mu_{Ga}^{\min} = E_{GaN}^{Bulk} - \mu_{N[N_2]}. \quad (5.3.1.2)$$

Analogically, the upper limit on μ_{Ga} results in a lower limit on μ_N which yields to:

$$\mu_N^{\min} = E_{GaN}^{Bulk} - \mu_{Ga[metal]}. \quad (5.3.1.3)$$

The chemical potential of GaN can therefore be expressed as

$$E_{GaN}^{Bulk} = \mu_{Ga[metal]} + \mu_{N[N_2]} + \Delta H_f[GaN], \quad (5.3.1.4)$$

where $\Delta H_f[GaN]$ is the enthalpy of formation or the energy gain in forming the crystal bulk GaN. After imposing the lower and upper bounds in the chemical potentials of both Nitrogen and Gallium, one can thus say that the Ga-rich environment is achieved by putting $\mu_{Ga} = \mu_{Ga[metal]}$ and $\mu_N = \mu_{N[N_2]} + \Delta H_f[GaN]$

Similarly, N-rich condition is also obtained by putting $\mu_N = \mu_{N[N_2]}$ and $\mu_{Ga} = \mu_{Ga[metal]} + \Delta H_f[GaN]$. Nevertheless, breaking down GaN bulk into two different components and computing their accurate corresponding chemical potentials is far from being trivial and every single computation that we have provided so far are HSE06 approximations. In fact, one must not forget that chemical potentials highly depend on temperature and pressure and an exact computation of such ambiguously defined physical quantities is impossible.

The values of the chemical potential have a tremendous affect on the defect formation energies and should therefore be considered as variables in our formalism.

5.3.2 Adjustment of finite-size effects in supercell calculations

Describing defects and impurities in a supercell method from a band structure perspective seems quite intuitive. Interactions between defects in neighboring supercells usually lead to a dispersed impurity band instead of a single localized eigenstate. In case of an infinitely large supercell in which the impurity would be close to being totally isolated, the defect—induced band would be completely flat. According to Van der Walle et al. (2004)[3], using special k-points to plot the band structure provides a way of averaging over the defect band. This would essentially compute the band's center of mass whose band level is quite similar to the completely isolated defect level. Consequently, one might avoid choosing the Γ point as one of the sampling points since at such high symmetry point, defect-defect interactions reaches its maximum which would thus lead to a very mediocre description of the band structure of a given system. One may counter argue that in case that the supercell is large enough (more than 100 atoms), interactions between neighboring defects is almost negligible and including the Γ point for Brillouin-zone integration results to fine accuracy and further numerical simplicity for the first-principles computations. Furthermore, provided that the electrons from the defect level interact with the valence or conduction band electrons, averaging over the defect band might not be the most appropriate method to approximate the total energy of the supercell. The valence band-defect band interactions might be so strong that the actual level of the defect would be shifted and would never correspond to the approximated center of mass that is computed from the special k-points method.

The use of the supercell method within the HSE06 self-consistent calculations includes several physical errors that need to be corrected. While various approaches for such corrections have been suggested in previous literature[62,134,135], we will be briefly discussing three different sets of corrections methods used in our GaN research.

5.3.2.1 Band-filling correction energy

One of the most ambiguously defined drawbacks of the supercell approximation is the possible hybridization of the impurity band with the host conduction (or valence) band. Such hybridization usually occurs in case that the impurity band is energetically close to the host band which would therefore create an overpopulation of the host band energy levels. In case of shallow donor impurities, if we plot its corresponding band structure, we would immediately notice an excess of electrons in the conduction band for different k-points. In order to eliminate those Moss-Burnstein-type filling effects[136], we apply a band filling correction $\nabla E[D^q]$ for shallow donors where:

$$\Delta E[D^q] = -\sum_{i \geq CBM, \vec{k}} (w_{\vec{k}} \eta_{i\vec{k}} \varepsilon_{i\vec{k}} - \varepsilon_{CBM,0}) \qquad (5.3.2.1.1)$$

here $\varepsilon_{i\vec{k}}$ are the single-particle eigenstates of the ith donor electron, $\varepsilon_{CBM,0}$ is the lowest occupied eigenstate of the CBM, usually located at the Γ point, $w_{\vec{k}}$ is the k-point weight derived from the k-mesh, and $\eta_{i\vec{k}}$ are the corresponding weights of the ith electron located at the kth point which is obtained from the tetrahedron k-space integration[137]. Corrections related to shallow donor impurities are actually a lot more complex than it appears to be and some physical insight was given in Section 1.3. In addition to band-filling correction, one must take into account the unphysical presence of the jellium compensating background for charged defects in a host crystal within the supercell formalism.

5.3.2.2 Image-charge correction

Computations of formation energies of a periodically repeated charged system requires careful investigation since neutrality of the system is crucial to avoid divergence of long-range Coulomb terms. In fact, the treatment of the $\vec{G} = 0$ terms from Eq. (5.1.10) leads to a divergence of the external, Hartree and exchange-correlation potentials. Since neutrality of the supercell is a necessary requirement, all computations should be rearranged in such way that the total number of electrons exactly matches the number of positive charges in the unit cell. Such neutrality condition is met by the incorporation of a uniform compensating (jellium) background within the supercell method. Even though this artificial background takes care of maintaining the charge neutrality for the calculation of the $\vec{G} = 0$ terms, interactions between the jellium background and the charged crystal (Madelung-type interactions) still occur and hence needs to be corrected. The integration of image-charge corrections in the treatment of charged supercell has been quite a debatable issue in the last decade.[134,138,139] One of the most serious approach into correcting these unwanted interactions was proposed by Makov and Payne (1995).[140] In their review, they first analyzed the charge density of a localized point defect in a solid and later derived the multipole correction of the total energy of a finite supercell *with respect to the total energy of an ideally infinite cell of a non periodic charged system as*[140]*:*

$$\Delta E_{MP} = \frac{q^2 \alpha}{2\varepsilon (V_{SC})^{1/3}} + \frac{2\pi q Q}{3\varepsilon V_{SC}} + O\left(V_{SC}^{-5/3}\right) \qquad (5.3.2.2.1)$$

Here the correction is expanded in powers of the reciprocal linear supercell dimension $V_{SC}^{1/3}$ where the first order (monopole) and third order (quadrupole) components dominate. The first component describes an array of specific point charges and a jellium background in a uniform

dielectric media ε. In this case, α is the crystal structure-dependent Madelung constant and V_{SC} is the volume of the supercell. The second term is generated from the *shape-dependent charge distribution within the artificial supercell with the uniform compensating background*[141]. Here, *Q* is the second radial moment of the defect charge density which can be derived from the electron density difference between the host with defect *(h+d)* and pure-crystal *(h)* lattice:

$$Q = \int r^2 \left[\rho_{h+d}(\vec{r}) - \rho_h(\vec{r})\right]. \qquad (5.3.2.2.2)$$

Even though the last component $O\left(V_{SC}^{-5/3}\right)$ of the multipole correction described in Eq. (5.3.2.2.1) is much smaller than the monopole and quadrupole terms and therefore negligible in most calculations, it was argued by Persson et al. (2005)[62] that correction of the 5th order are required for calculations of very shallow charged defect states. Another issue regarding calculations for charged states is the downshifting of the electronic single-particle energies from the $\vec{G} = 0$ terms.

4.2.2.3 Potential alignment correction for neutral and charged supercells

The true values of the Valence band maximum (VBM) or Conduction Band maximum (CBM) for the pure crystal lattice cannot be directly applied to the supercell with defect because of the energy contribution due to the interaction between the additional charge and the jellium background in the $\vec{G} = 0$ terms. Consequently, the *spectrum* of HSE06 eigenvalues is defined up to an unknown constant which depends on the average potential of the cell and the choice of the pseudopotentials.[62] In order to obtain consistency in the potentials, we decide to examine the potential in the supercell far from the impurity and align it with the average electrostatic potential of the pure host crystal. Such alignment gives rise to a shift or potential alignment ΔE_{PA} which is expressed as:

$$\Delta E_{PA} = q\left[V(D^q) - V(0)\right]_R$$

Here $\left[V(D^q) - V(0)\right]_R$ is the difference of potentials between the host + defect and pure-crystal lattice at a specific reference point.

5.4 Defects Transition Levels

5.4.1 Thermodynamic and Optical Levels

The incorporation of neutral and charged impurities in a given semiconductor introduces levels in the band gap or near the band edges. The Fermi-level position in which two different charge states q_1 and q_2 have equal formation energy, describes the thermodynamic transition energy level $\varepsilon_{The}(q_1/q_2)$. In the thermodynamic transition energy calculations, the atomic structure of each charge state must correspond to its relaxed equilibrium configuration. In fact, the relaxed atomic configuration for different charge states might not necessary be alike and this dissimilarity in atomic structure gives rise to the difference between the thermodynamic transition levels and the optical levels.[3]

The optical transition levels $\varepsilon_{Opt}(q_1/q_2)$ are defined in the same basis of the thermodynamic transition levels, apart from the fact that atomic positions of the initial state are kept during the vertical (optical) excitation according to Franck-Condon principle. In case of a deep donor X, the optical absorption energy $\varepsilon_{Opt}\left[q_1/\left(q_1+ne^-\right)\right]$ due to the excitation $n=+1$ of an electron from the defect level into the CBM is calculated as:

$$\varepsilon_{Opt}\left[q_1/\left(q_1+ne^-\right)\right]=E^f\left[\mathrm{X}^{q_1+1}\right]_{q_1}-E^f\left[\mathrm{X}^{q_1}\right]_{q_1}+E_g \qquad (5.4.1.1)$$

where $E^f\left[\mathrm{X}^{q_1+1}\right]_{q_1}$ is the formation energy of the deep defect X in the q_1+1 state from the q_1 state atomic configuration and $E^f\left[\mathrm{X}^{q_1}\right]_{q_1}$ is the formation energy of the X defect in the q_1 state from the q_1 relaxed atomic structure. E_g is the amount of energy it costs to add one electron to the CBM during the transition which should correspond to the band gap of the system where the transition occurs. However, the atomic configuration of the newly obtained charged defect $\left[\mathrm{X}^{q_1+1}\right]_{q_1}$ is not the most stable structure and loses the excess energy through phonons scattering and therefore ends up in the relaxed structure of the q_1+1 state. The optical transition that occurs due to the recombination of the electron from the CBM (or shallow donor states in some cases) with the valence band hole of the q_1+1 state is the emission energy and its corresponding energy $\varepsilon_{Opt}\left[\left(q_1+ne^-\right)/q_1\right]$ is given by:

$$\varepsilon_{Opt}\left[\left(q_1+ne^-\right)/q_1\right]=E^f\left[\mathrm{X}^{q_1+1}\right]_{q_1+1}-E^f\left[\mathrm{X}^{q_1}\right]_{q_1+1}+E_g \qquad (5.4.1.2)$$

where $E^f\left[\mathrm{X}^{q_1}\right]_{q_1+1}$ is the formation energy of the X defect in the q_1 state from the q_1+1 relaxed atomic structure and $E^f\left[\mathrm{X}^{q_1+1}\right]_{q_1+1}$ is the formation energy of the defect X in the q_1+1 state from the q_1+1 state atomic configuration. The atomic configuration of the initial state are always kept during the vertical transition because optical transition occurs at a very fast rate due to the huge difference of mass between nuclei and electrons.

Now that we have briefly discussed the methodology used to describe defects in semiconductors, we propose a solution to a long-standing problem of the GaN yellow luminescence.

6. Yellow luminescence of GaN generated by Carbon defect complexes

In the last chapter of the thesis, we demonstrate that the C_N-O_N complex is responsible for the observed carbon related YL in GaN. Using hybrid density functional theory and experimental photoluminescence (PL) measurements, we show that the calculated emission, absorption, zero phonon line, and thermodynamic transition level for the C_N-O_N complex are all in excellent agreement with the PL data. Furthermore, the formation energy of the carbon bound into the C_N-O_N complex is significantly lower than that of any other carbon defect. A systematic study has been performed in order to eliminate possible alternative explanations, including isolated defects and complexes, thus offering a solution to the YL problem.

6.1 Experimental methods

One of the first thorough experimental analyses on the mechanism of the YL in GaN was performed by T. Ogino and M. Aoki (1980).[47] The effects of intentional doping of foreign impurities possibly related to YL in bulk GaN were investigated, and it was found that Carbon-doped GaN has a drastic influence on the YL characteristic excitation band. Figure 6 and 7 show the PL intensity of "pure" GaN and C-Doped GaN at 4.2 ^{o}K, respectively. Based on the configurate coordinate model and the band model, the source of the yellow band was first thought to be a complex impurity consisting of a substitutional carbon replacing a nearest neighbor gallium atom and gallium vacancy. The unceasingly development of first-principles calculations and a better understanding of properties of impurities in semiconductors in the last decades has lead us to believe that the C_{Ga}-V_{Ga} complex might not truly explain the microscopic origin of YL in GaN. Figure 8 shows the spectral dependence (inset) and the temperature dependence of the PL intensity for the YL band in GaN. The YL band has an abrupt onset at 2.6 eV (which can be identified as the zero-phonon energy), a maximum at 2.20 eV and the full-width at half maximum of 410 meV at 15 K. From the fit of the temperature dependence of the YL intensity (Fig. 8) we have determined the activation energy (thermodynamic transition level) of E_A = 850 meV for the defect responsible for the YL band. The hole-capture coefficient, C_{pA}, obtained from this fit (6×10^{-7} cm^3/s) is close to the values reported in the literature,[46] and other parameters are very similar to the parameters of the YL band in C-doped GaN.[47] We observed a moderate shift to higher energies by only 8±2 meV with increasing excitation power density from 10^{-6} to 0.1 W/cm^2. This shift is commonly explained by the donor-acceptor-pair (DAP) type transition, where an electron bound to a shallow donor recombines with a hole bound to a deep acceptor. Due to the DAP interpretation, most defects suggested in the literature as sources of YL in GaN have been deep acceptors. However, as we show below, this is not necessarily the case. The sample used to obtain these results was grown by metalorganic chemical vapor deposition, and contained Si, C, and O atoms with concentrations of 3×10^{16}, 4×10^{16}, and 5×10^{16} cm^{-3}, respectively.[142] However, the concentration of the defects responsible for the YL band in this sample has been estimated to be 3.3×10^{15} cm^{-3}.[143] Therefore the concentration of defects responsible for the YL band is much lower than the concentrations of elemental point defects, suggesting that defect complexes, which have concentrations lower than that of the elemental defects, can explain this discrepancy.

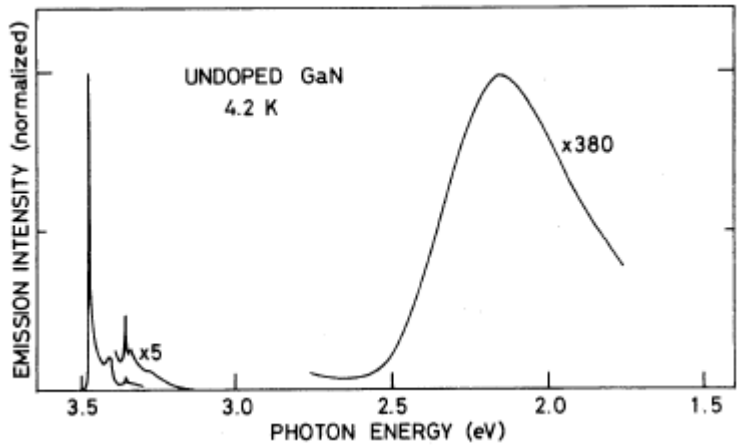

Figure 6: Photoluminescence spectrum of pure GaN at 4.7°K by Ogino and Aoki (1980)[47]. The dominant emission line located at 3.4719 eV is attributed to excitation recombination at neutral donor sites.

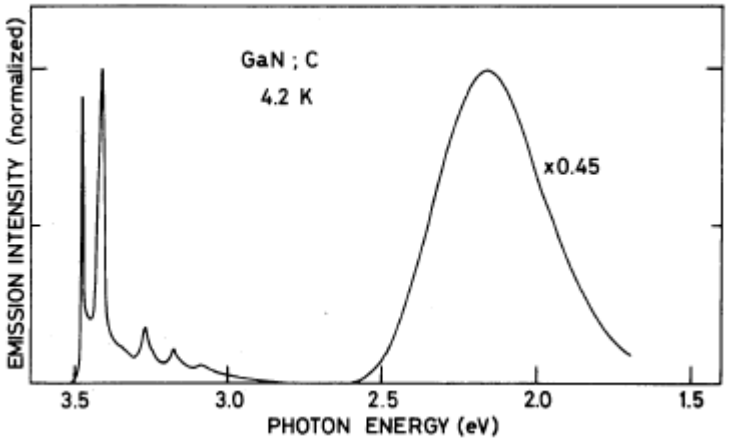

Figure 7: Photoluminescence spectrum of C-doped GaN at at 4.7°K by Ogino and Aoki (1980)[47]. The intensity of the characteristic excitation band is much larger than that of pure samples.

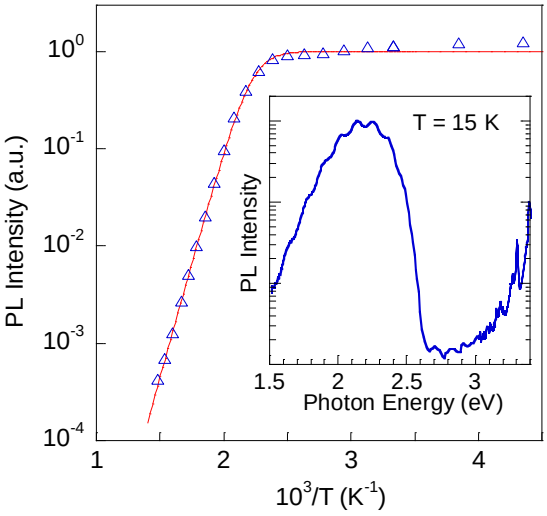

Figure 8: Temperature dependence of the YL band intensity. Excitation power density is 0.3 mW/cm^2. The line is a fit with Eq. (6) from Ref.46 with the following parameters: $\tau_R = 3.7\times10^{-4}$ s (determined from the time-resolved PL), $E_A = 850$ meV, $C_{pA} = 6\times10^{-7}$ cm^3/s. The inset shows the PL spectrum at 15 K.

6.2 Computational details

Our theoretical approach is based on the hybrid functional method which in recent years has become a preferred approach for the analysis of defects and their properties in semiconductors.[144] It offers a practical compromise between the semi/local approximations to DFT[3] and the computationally demanding many-body methods (GW).[145,146] Our calculations are based on the Heyd-Scuseria-Ernzerhof (HSE) hybrid functional[32] as implemented in the VASP program,[147] with the projector augmented wave method (PAW).[148,149] In an exchange tuned hybrid functional calculation of defects in semiconductors, the semi/local density exchange-correlation part of the density functional is mixed with a Fock-type exchange part in a ratio adjusted to match the band gap of the host material. Compared to semi/local functionals, this also improves the host lattice properties,[150] which is important to capture the defect relaxation properties. We use the HSE hybrid functional with the fraction of exact exchange of 0.31, and the screening parameter of 0.2 $Å^{-1}$. These parameters accurately reproduce both the band gap and the lattice properties of bulk GaN.[151] The resulting band gap of 3.49 eV is in good agreement with the low-temperature experimental value of 3.50 eV.[152] We used the value of the band gap renormalized by zero-point motion[153] to incorporate these effects into the gap fitting. Wurtzite GaN is illustrated in Figure 9 where the computed relaxed lattice parameters of the bulk (*a*=3.210 Å, *c*=5.198 Å, and *u*=0.377) have good agreements with experimental values (*a*=3.189 Å, *c*=5.185 Å).[154] The 128 atom supercells were used with atomic structures relaxed using HSE hybrid functional calculations to yield forces of 0.05 eV/Å or less. Figure 10 describes a 128 atoms bulk GaN supercell that contains 4 GaN primitive unit cells in the *a* and *b* directions and 2 primitive unit cells in the *c* direction. The plane-wave basis sets with 400 eV cut-off at the Γ–point were used in all electronic structure calculations. Spin polarized calculations were performed in all cases. Systematic tests were performed in order to identify and evaluate the sources of error, and the influence of any of these parameters on the defect thermodynamic and optical transition levels.

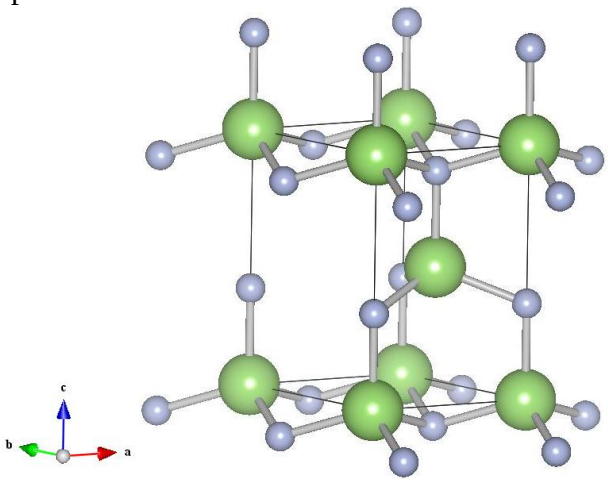

Figure 9: Crystal Structure of wurtzite GaN. The theoretical relaxed lattice parameters are computed to be a=*3.210 Å*, c=*5.198 Å*, *and* u=*0.377*.

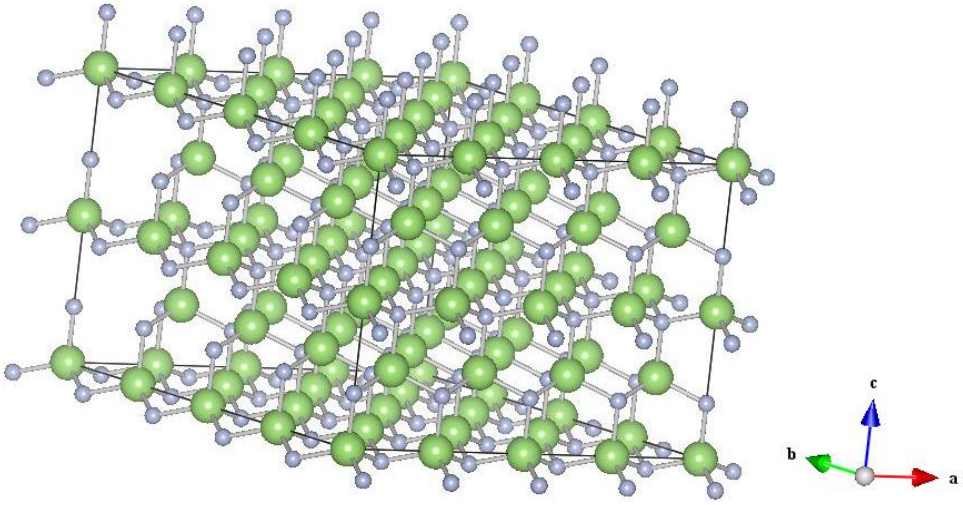

Figure 10: Relaxed atomic configuration of bulk GaN containing 128 atoms.

6.2.1 Formation Energy

The defect formation energy E_f determines the probability of a particular defect configuration to be realized. It is defined[3] $E_f = E_{tot}^{def} - E_{tot}^{bulk} - \sum_i \Delta n_i \mu_i + qE_F + \Delta V + \Delta E_{MP}$, as the total energy difference of the supercell with the defect and the bulk supercell, minus their difference in the chemical potentials μ_i for the number of atoms difference Δn_i in the two cells, adding the energy cost of charging the defect qE_F, assuming the exchange of electrons with the Fermi level E_F. The two remaining terms correct for the electrostatic errors of two different origins. The potential alignment ΔV arises from dropping the diverging $\vec{G} = 0$ term in the Fourier energy expansion in a charged supercell.[155] This term is usually small (0.05 to 0.15 eV) and proportional to the defect formal charge q. The last term is the spurious electrostatic interaction correction for charged defects, following Makov and Payne.[140,156] Here the Madelung energy was used along with the 3rd order corrections analyzed in detail by Lany et al.[1] Both terms scale as q^2 and depend on the supercell geometry. The Madelung energy for 128 atom GaN $\pm 1e$ charged cell is ~0.20 eV, while the 3rd order term is ~ -0.073 eV. Following Ref. 157, we also applied Madelung corrections to neutral impurities where electrons (holes) occupy the conduction (valence) band, i.e. forming a charged ion in delocalized compensating charge density.

For the 128 atom cells used in this work, the use of the Γ–point only rather than a k-point mesh was found to cause negligible errors. The total energy errors between the Γ–point and the 222 k-point either Monkhorst-Pack or Γ–centered mesh did not exceed 0.05 eV. A more significant source of error was found to be the plane-wave energy cutoff. In the literature for the typical HSE calculations of the defects in supercells, it is often set to 300 eV.[57,151] However, formation energies computed with a 400 eV energy cutoff were found to differ by about 0.1-0.2 eV from those of 300 eV. This error is not the same for different charge states of a given defect configuration, and therefore does not cancel out in optical transition calculations.

The remaining error related to the size of the cell includes several different error sources, i.e. elastic interactions and errors related to the supercell band structure. For example, between the

72 and 128 atom cells, the error in formation energy reaches 0.2 eV for isolated defects and up to 0.55 eV for some of the complexes. We tested hybrid functional calculations for supercells containing up to 300 atoms, and found that this error is reduced to about 0.15-0.2 eV for complexes and about 0.05 eV for isolated defects when using 128 atom cells. In some cases, this error can be estimated as the energy difference between the impurity band center of mass and the Γ–point eigenvalue.[157] We estimated these errors using GGA approximation for supercell sizes ranging from 128 to 572 atoms, and found them to be 0.1 to 0.2 eV. These values are very similar for different charge states of the same defect, leading to error cancellation in the computed transition energies.

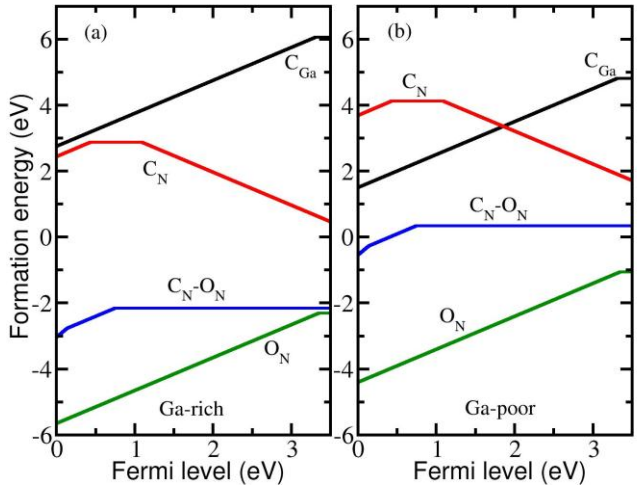

Figure 11: (Color online) Defect formation energies as a function of the Fermi energy in Ga-rich and Ga-poor growth conditions. Defect thermodynamic transition levels in the GaN band gap correspond to the intersections of different slopes (charge states) of each line.

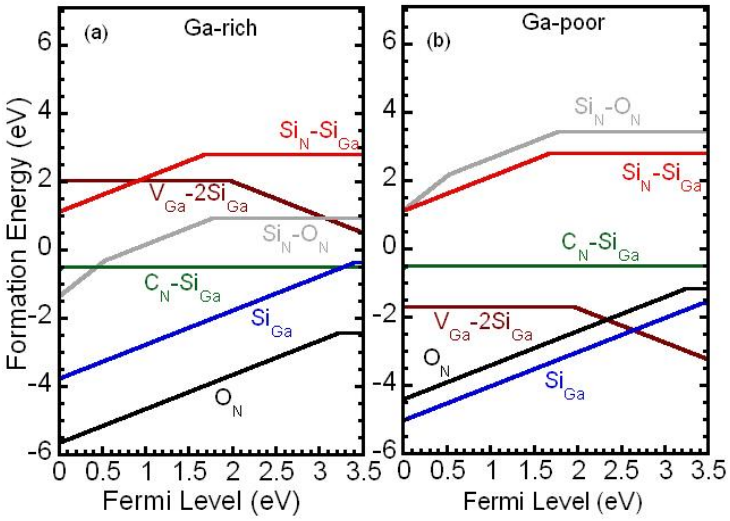

Figure 9: (Color online) Defect formation energies as a function of the Fermi energy in both Ga-rich and N-rich environments. The zero of Fermi level corresponds to the top of the valence band and only segments of the lowest-energy charge states are being displayed. The charge states or slopes of each segment correspond to the thermodynamic transition levels between different charge states.

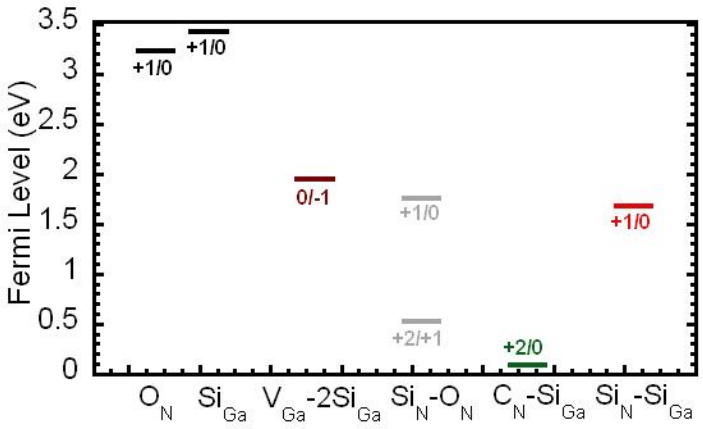

Figure 10: Thermodynamic Transition levels $\varepsilon(q_1, q_2)$ of defects in GaN in the energy bandgap between the VBM and CBM, where *(q_1/q_2)* describes the position at which charge states of the defects have equal energy.

6.2.2 Formation energies of defect complexes

The formation energies of different carbon defect configurations are presented in Figure 11, where in Ga-rich and Ga-poor growth conditions, Ga and N chemical potentials are separated by the GaN formation enthalpy, computed to be ΔH_{GaN} = -1.25 eV. The only carbon-related defect complex that is found to be energetically favorable is the C_N-O_N complex. For *n*-type GaN, its formation energy is more than 2.5 eV lower than that of the isolated C_N. The binding energy of this complex is 0.32 eV in both Ga-rich and Ga-poor conditions. However, in Ga-poor conditions the isolated oxygen donor has formation energy 1.4 eV lower than that of the complex, implying that in these conditions, complex concentrations will be low compared to those of the isolated impurities. In Ga-rich conditions, the complex formation energy is almost the same as that of O_N and ~2.6 eV lower than that of the C_N, thus the complex concentrations are expected to be large. The C_N-O_N complex is a deep donor, with a 0/1+ thermodynamic transition level at 0.75 eV above the VBM, and a deeper 1+/2+ level at 0.14 eV above the VBM.
It has been suggested that Ga vacancy complexes could be responsible for the YL.[47,54,55] However, the PL band produced by the V_{Ga}-O_N complex is computed here to be infrared, with a maximum at 1.42 eV. The V_{Ga}-C_{Ga} complex is unlikely to form, due to a high formation energy (~9.2 eV in *n*-type GaN), leading to a negative binding energy. The donor-acceptor complex C_N-C_{Ga} is also found here have a of high formation energy, 5.32 eV for *n*-type GaN, regardless of the growth conditions and is unlikely to form.
In close agreement with previously published results,[57] we also find that the isolated C_N is a deep acceptor with a transition energy of 1.09 eV. We also find a deep 1+/0 transition level at 0.43 eV

above the valence band maximum (VBM). A substitutional C_{Ga} donor was also found to have a relatively high formation energy, in both Ga-rich and Ga-poor conditions, with the shallow donor level at 0.18 eV below the conduction band minimum (CBM).

In addition to Carbon-related defects, we also investigated the electronic structure of isolated and complex Si impurities. Figure 12 presents the formation energies of Si defect configurations in both Ga-rich and N-rich environments using the HSE06 formalism. For each geometrical impurity, only the most energetically favorable charge states are being displayed. The Fermi energies at which the slopes change correspond to the thermodynamic transition levels that are illustrated in Figure 13. Amongst the isolated Si defects, the donor impurity Si_{Ga} shows the lowest formation energy in both growing environments and exhibits a particularly shallow transition level 0/+1 at 3.42 eV above the VBM, confirming previously found experimental results. The stable configuration of the substitutional impurity Si_{Ga} can be explained by the fact that Si atom can easily fit in the Ga site due to their similar atomic radii. Furthermore, Van de Walle et al.[3] also noticed that substitutional Si_N and interstitial Si usually cause large strain in the host crystal and are therefore energetically unfavorable. In Ga-rich conditions, the deep complex C_N-Si_{Ga} appears to be quite stable. The possession of the 0/+2 transition level extremely near the VBM (0.046 eV) indicates that C_N-Si_{Ga} is a very deep donor in the HSE06 band gap and cannot be a potential candidate to explain YL in GaN. Various discussions on Ga vacancy complexes have been previously suggested[3,46] and it was noted that V_{Ga} would diffuse pretty fast and would agglomerate with neighboring isolated defects to form complexes. A positively charged donor isolated impurity is usually more likely to form with the negatively charged donor Ga vacancy. A previous investigation on the electronic structure of the complex defect V_{Ga}-Si_{Ga} shows that such configuration exhibits a pretty low binding energy[3] (0.23 eV) and is therefore less probable to form. Hence, in our calculations, instead of analyzing the electronic structure of V_{Ga} with its single second nearest neighbors Si_{Ga}, we decided to attach two instersitial Si_{Ga} atoms with a Ga vacancy. With such structure, we expect a higher binding energy of the obtained complex and therefore a more energetically stable configuration. In fact, in N-rich environment, we notice that the complex V_{Ga}-$2Si_{Ga}$ exhibits the most favorable configuration amongst complexes in both p-type and n-type samples. As shown in Figure 12, the stable charge states of the V_{Ga}-$2Si_{Ga}$ complex are 0 and -1, and the transition level occurs at 1.96 eV above the VBM, meaning that V_{Ga}-$2Si_{Ga}$ acts as a deep acceptor. The low formation energy of -1 charge state is accompanied by a significant outward relaxation of the nearest N ions from the vacant Ga site by approximately 15%, in contrast to an inward relaxation of the nearest N ions to the 2 substitutional Si sites by roughly 10%. In the n-type sample, the formation energy of the complex V_{Ga}-$2Si_{Ga}$ in Ga-poor condition is -3.25 eV, which turns out to be much smaller than the formation energy of any other isolated or complex defects in GaN bulk. Such low formation energy shows that the complex V_{Ga}-$2Si_{Ga}$ may be the dominant compensation acceptor in n-type GaN. Even though extremely stable in the N-rich environment, the complex V_{Ga}-$2Si_{Ga}$ does not theoretically yield any interesting optical transition.

The shallow donor O_N is found to have a thermodynamic transition level 0.14 eV below the CBM. This defect has negative formation energy, implying that all available oxygen atoms will readily form the substitutional donors. The negative formation energy originates from the fact that all gallium oxides have a much larger magnitude of the formation enthalpy compared to that of GaN.[158] For example, the computed enthalpy of formation for common Ga_2O_3 is -10.5 eV, compared to -1.25 eV for GaN.

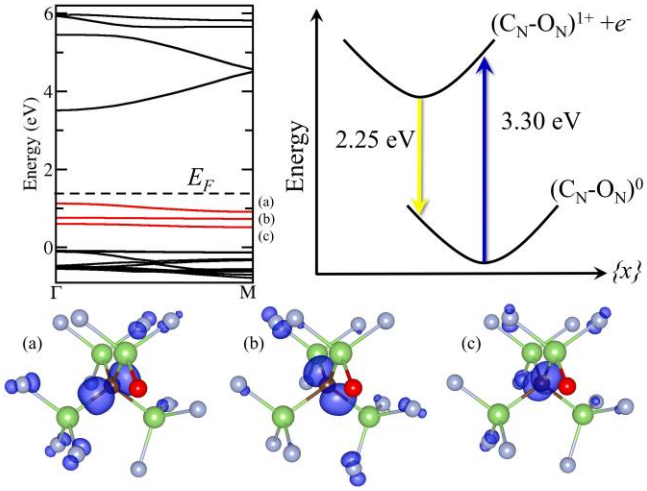

Figure 14: (Color online) Band structure of the $(C_N\text{-}O_N)^0$ defect complex ground state. The three localized defect states are plotted in red. The configuration coordinate diagram illustrates the absorption and emission energies. The charge densities of the three localized defect states are plotted at 15% of its maximum. Brown, red, green, and grey atoms are C, O, Ga, and N, respectively.

6.2.3 Optical transitions of complexes versus isolated defects

The calculated optical transitions using configuration coordinate diagrams are presented in Figure 14. Initially, the ground state of the $C_N\text{-}O_N$ complex in the *n*-type GaN is neutral. As a result of the optical excitation producing an electron-hole pair, the $C_N\text{-}O_N$ complex captures the hole which transfers the complex into a $(C_N\text{-}O_N)^{1+}$ charge state. The excitation energy for this transition is calculated to be 3.30 eV, which agrees with the experimental values of 3.19 eV[47] and 3.32 eV.[159] Loosing the excess energy through the lattice relaxation, the $(C_N\text{-}O_N)^{1+}$ complex relaxes into the minimum energy structure of the excited state. Figure 14 also shows the computed electronic band structure in the Γ–M direction, with the localized defect electronic levels (shown in red) slightly broadened by the spurious interactions due to the use of periodic boundary conditions. The crystal structure used in these band structure calculations is fixed at the relaxed $(C_N\text{-}O_N)^{1+}$ charged state, to represent the electronic structure related to the optical emission. The highest defect state will be occupied by the captured hole, allowing the subsequent radiative recombination of an electron in the CBM and the hole localized on the $C_N\text{-}O_N$ complex (the localized charge density of the hole is also shown in Fig. 14 (a)). This emission energy for the maximum of the YL band is computed to be 2.25 eV. The subsequent $(C_N\text{-}O_N)^0$ lattice relaxation energy (Franck-Condon shift) is computed to be 0.48 eV, yielding the zero-phonon transition at 2.7 eV. As a result, the complex returns to its ground state $(C_N\text{-}O_N)^0$. The difference between the absorption and emission peaks (Stokes shift) is found to be 1.05 eV. These results are in excellent agreement with the calculated results and our measurements, where the PL emission peak is found at 2.20 eV (Figure 8), supporting our proposed YL source as the $C_N\text{-}O_N$ complex. They also agree very well with configuration diagrams deduced from early experiments in Ref. 47.

Recently, Lyons et al.[57] have suggested that the YL band can be explained by the transitions to the C_N deep acceptor, obtaining 2.14 eV for the optical transition from the conduction band to the $C_N^{0/1-}$ level. Using the 300 eV energy cutoff in our hybrid functional calculations, we reproduce these results. However, increasing the cutoff to 400 eV results in a deeper thermodynamic transition energy of 1.09 eV (0.9 eV in Ref. 57), which does not agree well with a measured ionization energy of 0.85 eV for the YL-related acceptor obtained here and in early works on the subject.[47] This shift changes the calculated optical transition energy to a *red* optical transition with a maximum at 1.88 eV. The absorption band maximum also lowers and is found to be at 2.76 eV. Therefore, the calculated properties of the isolated C_N acceptor do not agree well with the experimental data for the carbon-related YL in GaN. On the other hand, the more energetically favorable C_N-O_N complex yields optical transitions in agreement with the measured PL spectrum (Figure 8).

The proposed explanation of the YL band by the C_N-O_N complex is also consistent with previously published experimental data. In particular, a blue shift of a PL band with increasing excitation intensity is commonly attributed to the DAP-type optical transitions involving a shallow donor and a deep acceptor.[46] However, the blue shift for a PL band can also be caused by transitions from shallow donors to a deep donor. For example, the presence of several types of shallow donors with different ionization energies would cause the same effect as the DAP with random distribution of pair separations. Moreover, the broadening of a shallow donor level due to the interaction of impurities is identical to the presence of several types of shallow donors. Hitherto, the attribution of the YL-related defect to a deep acceptor rather than a deep donor historically always appeared more reasonable, because the capture of holes by negatively charged acceptors is more efficient than the capture by a neutral donor. Nevertheless, our estimates for deep-level defects in GaN indicate that the hole-capture efficiency for a neutral donor is only an order of magnitude lower than that for a negatively charged acceptor,[160] which is sufficient to cause the observed YL.

6.3 Conclusion

In conclusion, we have demonstrated that the deep donor complex $(C_N\text{-}O_N)^0$ explains the microscopic mechanism of the YL in GaN. This complex has a low formation energy and therefore should be present in sufficient concentrations to cause the observed PL. Calculated optical transitions via the localized defect states of this complex are in excellent agreement with the measured PL data (experimental values are given in brackets): thermodynamic transition level of 0.75 eV (0.85 eV), absorption energy 3.30 eV (3.32 eV), emission energy 2.25 eV (2.20 eV), and zero phonon transition 2.70 eV (2.60 eV). This complex has not been proposed as a source of the yellow band in GaN, while all other defects previously suggested being sources of this band, exhibit high formation energies and would produce red or infrared PL. Thus, we resolve a 30 year-old problem of microscopic origin of yellow luminescence in GaN.

This work used the computational facilities of the VCU Center for High Performance Computing.

References

[1] S. Lany and A. Zunger, Phys. Rev. B **78**, 235104 (2008)
[2] F. A. Kroger, *The Chemistry of Imperfect crystals* (North-Holland, Amsterdam, 1974)
[3] C. G. Van de Walle and J. Neugebauer, J. Appl. Phys. **95**, 3851 (2004)
[4] J. Callaway, J. Math. Phys. **5**, 783 (1964)
[5] G. F. Koster, J. C. Slater, Phys. Rev. **95**, 1167 (1954)
[6] A. M. Stoneham, *Theory of Defects in Solids* (Clarendon, Oxford 1985)
[7] D. A. Drabold and S. K. Estreicher, *Theory of Defects in Semiconductors*, (Springer-Verlag, Heidelberg 2007)
[8] W. Kohn and L. J. Sham, Phys. Rev. Lett. 140, **1133** (1965)
[9] R. O. Jones, *Introduction to Density Functional Theory and Exchange-Correlation Energy Functionals* Computational Nanoscience: Do It Yourself!, Jülich, NIC Series Vol. 31. (2006), pp. 45-70
[10] R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods*, (Cambridge University Press, U.K) , 2004
[11] J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber et al., J. Chem. Phys. **124**, 154709 (2006)
[12] J. P. Perdew, Phys. Rev. Let. **55**, 1665 (1985)
[13] F. Herman, J. P. Dyke, and I. P. Ortenburger, Phys, Rev. Lett. **22**, 807 (1969)
[14] P. S. Svendsen and U. von Barth, Phys. Rev. B **54**, 17402 (1996)
[15] J. Tao, J.P. Perdew, V.N. Staroverov and G.E. Scuseria,, Phys. Rev. Lett. **91**, 14640 (2003).
[16] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke, Phys. Rev. Lett. **100**, 136406 (2008)
[17] J. P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996)
[18] K. Hummer, J. Harl, and G. Kresse, Phys. Rev. B **80**, 115205 (2009)
[19] O. Gunnarsson and B. I. Lundqvist, Phys. Rev. B **13**, 4274 (1976)(check the literature, its from Kerstin Hummer,* Judith Harl, and Georg Kresse)
[20] R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989) (check the literature, its from Kerstin Hummer,* Judith Harl, and Georg Kresse)
[21] W. Kohn, Rev. Mod. Phys. **71**, 1253 (1999)
[22] R. G. Parr and W. Yang, Phys. Chem. **46**, 701(1995)
[23] P. G.-G, M. Seidl, Phys. Chem. Chem. Phys. **12**, 14405 (2010)
[24] R. Armiento and A. E. Mattsson, Phys. Rev. B **72**, 085108 (2005)
[25] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E.Scuseria, L. A. Constantin, X. Zhou, and K. Burke, Phys. Rev.Lett. **100**, 136406 (2008)
[26] A. E. Mattsson, R. Armiento, and T. R. Mattsson, Phys. Rev.Lett. **101**, 239701 (2008)
[27] F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998)
[28] A. D. Becke, J. Chern. Phys. **96**, 2155 (1992); A. D. Becke, J. Chem. Phys. **98**, 1372 (1993); A. D. Becke, J. Chem. Phys. **107**, 8554 (1997);
[29] J. P. Perdew, M. Ernzerhof, and K. Burke, J. Chem. Phys. **105**, 9982 (1996)
[30] M. Ernzerhof, J. P. Perdew, and K. Burke, Int. J. Quantum Chem. **64**, 285 (1997)
[31] J. Heyd, Thesis for PHD, Rice University, (2004)
[32] J. Heyd, G. E. Scuseria, and M. Ernzerhof, J. Chem. Phys. **118**, 8207 (2003); J. Heyd, G. E. Scuseria, and M. Ernzerhof, J. Chem. Phys. **124**, 219906 (2006)
[33] J. E. Peralta, J. Heyd, G. E. Scuseria, and R. L. Martin, Phys. Rev. B **74**, 073101 (2006)

[34] J. Heyd and G. E. Scuseria, J. Chem. Phys. **121**, 1187 (2004)
[35] K. Hummer, A. Grüneis, and G. Kresse, Phys. Rev. B **75**, 195211 (2007)
[36] D. O. Demchenko and I. C. Diallo, soon to be published, Cu impurities in CdS
[37] P. Deák, B. Aradi, T. Frauenheim, E. Janzén, and A. Gali, Phys. Rev. B **81**, 153203 (2010)
[38] A. V. Krukau, O. A. Vydrov, A. F. Izmaylov, and G. E. Scuseria, J. Chem Phys. **125**, 224106 (2006)
[39] A. F. Izmaylov and G. E. Scuseria, J. Chem. Phys. **127**, 144106 (2007)
[40] D. O. Demchenko, I. C. Diallo and Reschikov, C impurities in GaN, soon to be published
[41] A.V. Krukau, O. A. Vydrov, A. F. Izmaylov, and G. E. Scuseria, J. Chem. Phys. **125**, 224106 (2006)
[42] S. Nakamura, T. Mukai, and M. Senoh, Appl. Phys. Lett. **64**, 1687 (1994)
[43] S. Nakamura and G. Fasol, *The Blue Laser-Diode—GaN Based Light Emitters and Lasers* (Springer-Verlag, Berlin, 1997).
[44] F. A. Ponce and D. P. Bour, Nature (London) **386**, 351 (1997)
[45] R. Dahal, J. Li, K. Aryal, J. Y. Lin, and H. X. Jiang, Appl. Phys. Lett. **97**, 073115 (2010)
[46] M. A. Reshchikov and H. Morkoç, J. Appl. Phys. **97**, 061301 (2005)
[47] T. Ogino and M. Aoki, Jpn. J. Appl. Phys. **19**, 2395 (1980)
[48] T. Suski, P. Berlin, H. Teisseyre, M. Leszcyn´sky, I. Grzegory, J. Jun, M. Boc´kowski, S. Porowski, and T. D. Moustakas, Appl. Phys. Lett. **67**, 2188 (1995)
[49] M. O. Manasreh and H. X. Jiang, *Three-nitride Semiconductors: Optical Properties, Volume 1* (CRC Press, New York, 2002).
[50] E. R. Glaser, T. A. Kennedy, K. Doverspike, L. B. Rowland, D. K. Gaskill, J. A. Freitas Jr., M. Asif Khan, D. T. Olson and J. N. Kuznia, D. K. Wickenden, Phys. Rev. B **51**, 13326 (1995).
[51] C. H. Seager, A. F. Wright, J. Yu, and W. Götz, J. Appl. Phys. **92**, 6553 (2002).
[52] A. Armstrong, A. R. Arehart, D. S. Green, U. K. Mishra, J. S. Speck, and S. A. Ringel, J. Appl. Phys. **98**, 053704 (2005).
[53] S. O. Kucheyev, M. Toth, M. R. Phillips, J. S. Williams, C. Jagadish, and G. Li, J. Appl. Phys. **91**, 5867 (2002).
[54] J. Neugebauer and C. G. Van de Walle, Appl. Phys. Lett. **69**, 503 (1996)
[55] T. Mattila and R. M. Nieminen, Phys. Rev. B 55, 9571 (1997).
[56] K. Saarinen, T. Laine, S. Kuisma, J. Nissilä, P. Hautojärvi, L. Dobrzynski, J. M. Baranowski, K. Pakula, R. Stepniewski, M. Wojdak, A. Wysmolek, T. Suski, M. Leszczynski, I. Grzegory, and S. Porowski, Phys. Rev. Lett. **79**, 3030 (1997).
[57] J. L. Lyons, A. Janotti, and C. G. Van de Walle, Appl. Phys. Lett. **97**, 152108 (2010)
[58] S. T. Pandelides, Rev. Mod. Phys. 50, 797 (1979)
[59] A. J. R. deKock, , Philips Res. Rep. Suppl. **1**, 1 (1973)
[60] P. M. Petroff and B. L. Hartmann, Appl. Phys. Lett. **23**, 469 (1973)
[61] J. Lento, J-L. Mozos, and R. Nieminen, J. Phys.: Condens. Matter **14**, 2637 (2002)
[62] C. Persson, Y. –J. Zhao, S. Lany, and A. Zunger, Phys. Rev. B **72**, 035211, (2005)
[63] E. Engel and R. M. Dreizler, *Density Functional Theory: An Advanced Course*, (Springer-Verlag, New York, 2011)
[64] S. Ask, E. Elfgren, and I. Jonsson, Density Functional Theory and Defects in semiconductors, Project Report (2000).
[65] R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989)
[66] M. Born and J. R. Oppenheimer, Ann. Physik **84**, 458 (1927)

[67] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Dover Publications Inc., New York, 1996)
[68] D. R. Hartree, Math. Proc. Cam. Phil. Soc. **24**, 89-110 (1928)
[69] E. Fermi, Zeitschrift f¸r Physik, **36**, 902-912 (1926)
[70] P.A.M. Dirac, Proc. Roy. Soc. **A112**, 281-305 (1926)
[71] J. C. Slater, Phys. Rev. **81**, 385 (1951)
[72] R. G. Parr and W. Yang, *Density Functional Theory of Atoms ad Molecules* (Oxford University Press, New York, 1989)
[73] M. Levy, "The Constrained Search Approach, Mappings to External Potentials, and Virial-Like Theorems for Electron-Density and One-Matrix Energy-Functional Theories", in *Lecture Notes in Physics:Density Functional Theory*, edited by J. Keller and J. L. Gasquez, (Springer-Verlag, New York, 1983)
[74] L. H. Thomas, Proc. Cambridge Phil. Roy. Soc. **23**, 542 (1927)
[75] E. Fermi, Rend. Accad. Naz. Lincei **6**, 602 (1927)
[76] P. A. M Dirac , Proc. Cambridge Philos. Soc. **26**, 376 (1930)
[77] J. C Slater, Phys. Rev. **81** (1951)
[78] K. Burke and friends, The ABC of DFT, 2007
[79] L. H. Thomas and K. Umeda, J. Chem. Phys. 26, **293** (1957)
[80] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964)
[81] V. Volterra, *Theory of Functionals and of integral and integro-differential equations* (Dover Publications Inc., New York, 1959)
[82] http://www.mrl.ucsb.edu/~ghf/che230a/ghf_monog_appx_C.pdf
[83] G. Arfken, *Mathematical Methods for Physicists*, 3rd ed., 303-313 (Academic Press, Orlando, 1985)
[84] D. S. Sholl, J. A. Steckel, Density Functional Theory: A Practical Introduction, (John Wiley & Sons, New Jersey, 2009)
[85] M. Levy, Proc. Nat. Acad. Sci. **76**, 6062 (1979)
[86] M. Levy, Phys. Rev. A **26**, 1200 (1982)
[87] E. Lieb, Int. J. Qunt. Chem. **24**, 243 (1983)
[88] E. Lieb, *Density Functional Methods in Physics*, edited by R. M. Dreizler and J. da Provindencia, (Plenum, New York, 1985)
[89] N. Balazs, Phys. Rev 156, **42** (1967)
[90] E. H. Lieb and B. Simon, Phys. Rev. Lett. **31**, 681 (1973)
[91] E. Teller, Rev. Mod. Phys. **34**, 627 (1962)
[92] C. F. V. Weizsacker, Z. Phys. **96**, 431, (1935)
[93] B. Jacob, R. M. Dreizler, and E. K. U. Gross, J. Phys. B. **14**, 2753 (1981)
[94] J. P. Perdew, Phys. Rev. Lett. **55**, 1665 (1985a)
[95] D. C. Langreth and J. P. Perdew, Phys. Rev. B **15**, 2884 (1977)
[96] J. Harris, Int. J. Quantum. Chem. **13**, 189 (1984)
[97] W. M. C. Foulkes, L. Mitas, R. J. Needs and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001)
[98] Oleg Gritsenko, Robert van Leeuwen, and Evert Jan Baerends, J. Chem. Phys. **101**, 8955 (1994)
[99] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Phys. Rev. Lett. **49**, 1691 (1982)
[100] L.J. Sham and M. Schluter, Phys. Rev. Lett. **51**, 1888 (1983)
[101] J. P. Perdew and Mel Levy, Phys. Rev. Lett . **51**, 1884 (1983)

[102] Y. Wang and J. P. Perdew, Phys. Rev. B **43**, 8911 (1991); J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992)
[103] C. Attaccalite, S. Moroni, P. G.-Giorgi, and G. B. Bachelet , Phys. Rev. Lett **88**, 256601(2002)
[104] B. Y. Tong and L. J. Sham, Phys. Rev. **144**, 1-4 (1966)
[105] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981)
[106] K. Burke, J. P. Perdew, and Y. Wang, in *Electronic Density Functional Theory: Recent Progress and New Directions*, edited by J. F. Dobson, G. Vignale, and M. P. Das (Plenum, NY, 1998)
[107] J. P. Perdew and Y. Wang, Phys. Rev. B **33**, 8800 (1986) ; J. P. Perdew, Phys. Rev. B **33**, 8822 (1986)
[108] A. D. Becke, Phys. Rev. A **38**, 3098 (1988)
[109] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988)
[110] B. G. Johnson, P. M. W. Gill, and J. A Pople, J. Chem. Phys. 98, 5612 (1993)
[111] C. W. Murray, N. C. Handy, and R. D. Amos, J. Chem. Phys. 98, 7145 (1993)
[112] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Sigh, and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992)
[113] C. Sosa and C. Lee, J. Chem. Phys. **98**, 8004 (1993)
[114] P. Bagbo, O, Jepsen, O. Gunnarsson, Phys. Rev. B. **40**, 1997 (1989)
[115] C. Lee, D. Vanderbilt, K. Laasonen, R. Car, M. Parrinello, Phys. Rev. B **47**, 4863 (1993)
[116] R. N. Barnett, U. Landman, Phys. Rev. B **48**, 2081 (1993)
[117] J. P. Perdew and K. Burke, Int. J. Quant. Chem. **57**, 309 (1996)
[118] Y. –H. Kim, I. –H. Lee, S. Nagaraja, J. –P. Leburton, R. Q. Hood, and R. M. Martin, Phys. Rev. B **61**, 5202 (2000)
[119] J. P. Perdew, K. Burke, and M. Ernzherof, Phys. Rev. Lett. **77**, 3865, (1996)
[120] S. –K. Ma and K. A. Brueckner, Phys. Rev. **165**, 18-31 (1968)
[121] J. Kohanoff, *Electronic Structure Calculations for Solids and Molecules*, (Cambridge University Press, New York, 2006)
[122] C. C. J. Roothan, Rev. Mod. Phys. **23**, 69 (1951)
[123] J. C. Slater, *Quantum Theory of Molecules and Solids*, Vol. 2: Symmetry and Energy Bands in Crystals (Mc Graw-Hill, New York, 1963-74)
[124] M. Ernzerhof and G. E. Scuseria, J. Chem. Phys. **110**, 5029 (1999)
[125] C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999)
[126] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993)
[127] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M.J. Frisch, J. Phys. Chem. **98**, 11623-11627 (1994)
[128] N. W. Ashcroft and N. D. Mermin, *Solid States Physics*, (Saunders College, Philadelphia, 1976)
[129] J.P. Dombroski, S.W. Taylor and P. M. W. Gill, J. Phys. Chem. **100**, 6272 (1996)
[130] R. Dingle, D. D. Sell, S. E. Stokowski, and M. Ilegems, Phys.Rev.B **4**, 1211 (1971); B. Monemar, Phys. Rev. B **10**, 676 (1974).
[131] C. Rostgaard, Cond. Mat. Matr-Sci. **0910.1921v2**, 1-25 (2009)
[132] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045, (1992)
[133] J. Lento, J. –L. Mozos, and R. M. Nieminen, J. Phys.: Condens. Matter **14**, 2637-2645 (2002)

[134] S. B. Zhang, J. Phys.: Condens. Matter **14**, R881–R903 (2002)
[135] P. Erhart, K. Albe, and A. Klein, Phys. Rev. B **73**, 205203, (2006)
[136] T. S. Moss, Proc. Phys. Soc. London, Sec. B **67**, 775 (1954); E. Burstein, Phys. Rev. **93**, 632 (1954)
[137] G. Kresse and J. Hafner, Phys. Rev. B **47**, R558 (1993); G. Kresse and J. Furthmüller, *ibid.* **54**, 11169 (1996)
[138] P. Erhart, K. Albe, and A. Klei, Phys. Rev. B **73**, 205203 (2006)
[139] T. R. Paudel and W. R. L. Lambrecht, Phys. Rev. B **77**, 205202 (2008)
[140] G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995)
[141] R. M. Nieminen in Supercell Methods for Defect Calculations edited by D. A. Drabold and S. K. Estreicher, *Theory of Defects in Semiconductors*, (Springer-Verlag, Heidelberg 2007)
[142] The GaN sample was grown and the secondary mass-spectrometry data were provided by S.-P. Guo and D. W. Gotthold (EMCORE).
[143] M. A. Reshchikov, Appl. Phys. Lett. **88**, 202104 (2006).
[144] A. Alkauskas, P. Broqvist, and A. Pasquarello, Phys. Rev. Lett. **101,** 046405 (2008).
[145] S. Lany and A. Zunger. Phys. Rev. B **81,** 113201 (2010).
[146] P. Rinke, A. Janotti, M. Scheffler, C. G. Van de Walle, Phys. Rev. Lett. **102,** 026402 (2009).
[147] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
[148] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[149] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[150] M. Marsman, J. Paier, A. Stroppa. and G. Kresse, J Phys: Condensed Matter **20,** 064201 (2008).
[151] J.L. Lyons, A.Janotti, and C.G. Van de Walle, Phys. Rev. Lett. **108**, 156403 (2012).
[152] R. Dingle, D. D. Sell, S. E. Stokowski, and M. Ilegems, Phys.Rev.B **4**, 1211 (1971); B. Monemar, Phys. Rev. B **10**, 676 (1974).
[153] F. J. Manjon, M. A. Hernandez-Fenollosa, B. Mari, S.F. Li, C. D. Poweleit, A. Bell, J. Menendez, and M. Cardona, Eur. Phys. J. B **40**, 453 (2004).
[154] H. Morkoç, *Handbook of Nitride Semiconductors and Devices*, Wiley, New York, 2008, Vol. 1–3.
[155] S. Lany and A. Zunger. J Appl Phys **100,** 113725 (2006).
[156] M. Leslie and M. J. Gillan, J. Phys. C **18**, 973 (1985).
[157] F. Oba, A. Togo, I. Tanaka, J. Paier, and G. Kresse, Phys. Rev. B **77,** 245202 (2008).
[158] M. Zinkevich and F. Aldinger, J Am Ceram Soc **87,** 683 (2004).
[159] M. A. Reshchikov, H. Morkoç, S. S. Park, and K. Y. Lee, Mat. Res. Soc. Symp. Proc. **693**, I6.19 (2002).
[160] M. A. Reshchikov and R. Y. Korotkov, Phys. Rev. B **64**, 115205 (2001).